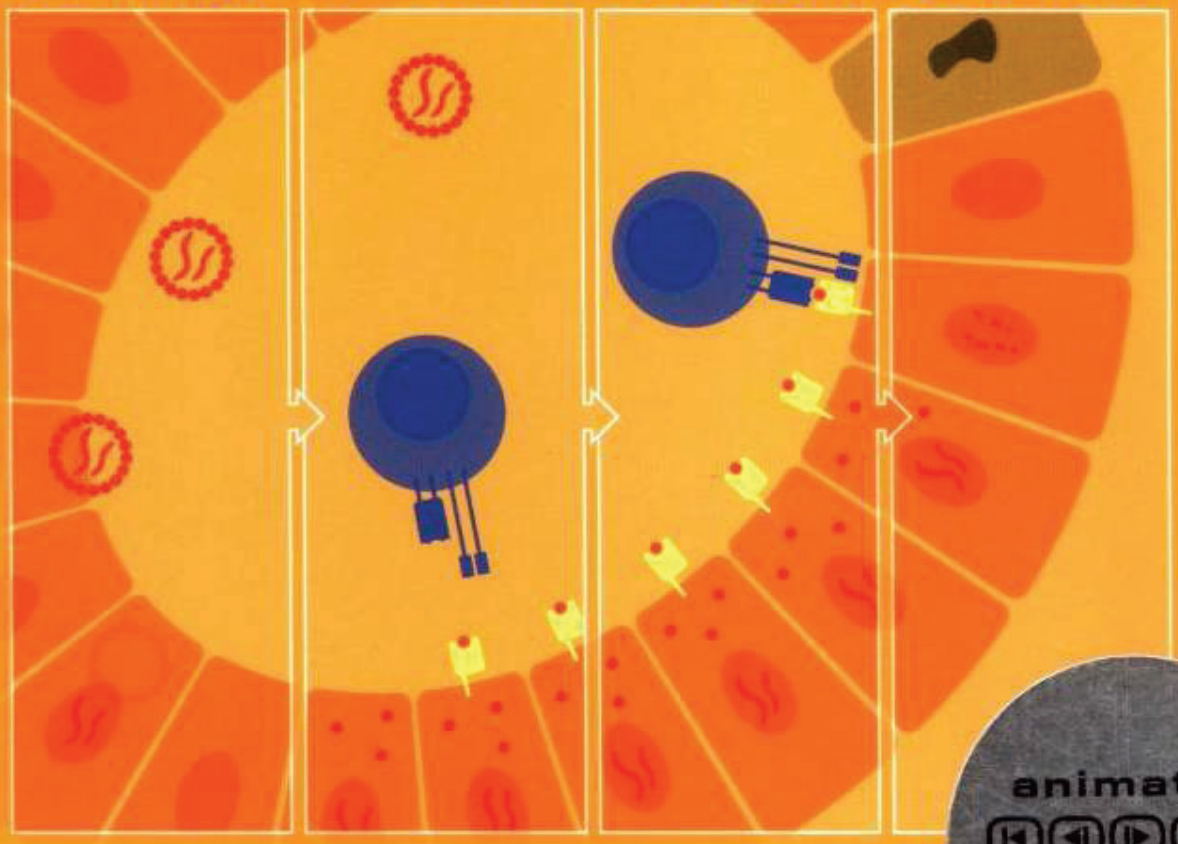
immuno biology 5

THE IMMUNE SYSTEM IN HEALTH AND DISEASE



FIFTH EDITION

animated
cd-rom inside

CHARLES A JANEWAY ▪ PAUL TRAVERS
MARK WALPORT ▪ MARK SHLOMCHIK

Bicara Ex. 1058, Page 1 of 85
Bicara v. JHU - PGR2026-00025

Short contents

Preface to the Fifth Edition

Acknowledgments

Part I. An Introduction to Immunobiology and Innate Immunity

1. Basic Concepts in Immunology

2. Innate Immunity

Part II. The Recognition of Antigen

3. Antigen Recognition by B-cell and T-cell Receptors

4. The Generation of Lymphocyte Antigen Receptors

5. Antigen Presentation to T Lymphocytes

Part III. The Development of Mature Lymphocyte Receptor Repertoires

6. Signaling Through Immune System Receptors

7. The Development and Survival of Lymphocytes

Part IV. The Adaptive Immune Response

8. T Cell-Mediated Immunity

9. The Humoral Immune Response

10. Adaptive Immunity to Infection

Part V. The Immune System in Health and Disease

11. Failures of Host Defense Mechanisms

12. Allergy and Hypersensitivity

13. Autoimmunity and Transplantation

14. Manipulation of the Immune Response

Afterword. Evolution of the Immune System: Past, Present, and Future, by Charles A. Janeway, Jr.

Appendix I. Immunologists' Toolbox

Immunization.

The detection, measurement, and characterization of antibodies and their use as research and diagnostic tools.

Isolation of lymphocytes.

Characterization of lymphocyte specificity, frequency, and function.

Detection of immunity *in vivo*.

Manipulation of the immune system.

Appendix II. CD Antigens.

Appendix III. Cytokines and Their Receptors.

Appendix IV. Chemokines and Their Receptors.

Appendix V. Immunological Constants.

Biographies

Glossary

Full contents

Preface to the Fifth Edition

Acknowledgments

Part I. An Introduction to Immunobiology and Innate Immunity

1. Basic Concepts in Immunology

The components of the immune system.

Principles of innate and adaptive immunity.

The recognition and effector mechanisms of adaptive immunity.

Summary to Chapter 1.

References

2. Innate Immunity

The front line of host defense.

The complement system and innate immunity.

Receptors of the innate immune system.

Induced innate responses to infection.

Summary to Chapter 2.

References

Part II. The Recognition of Antigen

3. Antigen Recognition by B-cell and T-cell Receptors

The structure of a typical antibody molecule.

The interaction of the antibody molecule with specific antigen.

Antigen recognition by T cells.

Summary to Chapter 3.

References

4. The Generation of Lymphocyte Antigen Receptors

The generation of diversity in immunoglobulins.

T-cell receptor gene rearrangement.

Structural variation in immunoglobulin constant regions.

Summary to Chapter 4.

References

5. Antigen Presentation to T Lymphocytes

The generation of T-cell receptor ligands.

The major histocompatibility complex and its functions.

Summary to Chapter 5.

References

Part III. The Development of Mature Lymphocyte Receptor Repertoires

6. Signaling Through Immune System Receptors

General principles of transmembrane signaling.

Antigen receptor structure and signaling pathways.

Other signaling pathways that contribute to lymphocyte behavior.

Summary to Chapter 6.

References

7. The Development and Survival of Lymphocytes

Generation of lymphocytes in bone marrow and thymus.

The rearrangement of antigen-receptor gene segments controls lymphocyte development.

Interaction with self antigens selects some lymphocytes for survival but eliminates others.

Survival and maturation of lymphocytes in peripheral lymphoid tissues.

Summary to Chapter 7.

References

Part IV. The Adaptive Immune Response

8. T Cell-Mediated Immunity

The production of armed effector T cells.

General properties of armed effector T cells.

T cell-mediated cytotoxicity.

Macrophage activation by armed CD4 T_H1 cells.

Summary to Chapter 8.

References

9. The Humoral Immune Response

B-cell activation by armed helper T cells.

The distribution and functions of immunoglobulin isotypes.

The destruction of antibody-coated pathogens via Fc receptors.

Summary to Chapter 9.

References

10. Adaptive Immunity to Infection

Infectious agents and how they cause disease.

The course of the adaptive response to infection.

The mucosal immune system.

Immunological memory.

Summary to Chapter 10.

References

Part V. The Immune System in Health and Disease

11. Failures of Host Defense Mechanisms

Pathogens have evolved various means of evading or subverting normal host defenses.

Inherited immunodeficiency diseases.

Acquired immune deficiency syndrome.

Summary to Chapter 11.

References

12. Allergy and Hypersensitivity

The production of IgE.

Effector mechanisms in allergic reactions.

Hypersensitivity diseases.

Summary to Chapter 12.

References

13. Autoimmunity and Transplantation

Autoimmune responses are directed against self antigens.

Responses to alloantigens and transplant rejection.

Self-tolerance and its loss.

Summary to Chapter 13.

References

14. Manipulation of the Immune Response

Extrinsic regulation of unwanted immune responses.

Using the immune response to attack tumors.

Manipulating the immune response to fight infection.

Summary to Chapter 14.

References

Afterword. Evolution of the Immune System: Past, Present, and Future, by Charles A. Janeway, Jr.

Evolution of the innate immune system.

Evolution of the adaptive immune response.

The importance of immunological memory in fixing adaptive immunity in the genome.

Future directions of research in immunobiology.

Summary of the Afterword.

Appendix I. Immunologists' Toolbox

Immunization.

The detection, measurement, and characterization of antibodies and their use as research and diagnostic tools.

Isolation of lymphocytes.

Characterization of lymphocyte specificity, frequency, and function.

Detection of immunity *in vivo*.

Manipulation of the immune system.

Appendix II. CD Antigens.

Appendix III. Cytokines and Their Receptors.

Appendix IV. Chemokines and Their Receptors.

Appendix V. Immunological Constants.

Biographies

Glossary

Immunobiology

Charles A. Janeway Jr.
Yale University School of Medicine

Paul Travers
Anthony Nolan Research Institute, London

Mark Walport
Imperial College School of Medicine, London

Mark J. Shlomchik
Yale University School of Medicine

Vice President: Denise Schanck

Text Editors: Penelope Austin, Eleanor Lawrence

Managing Editor: Sarah Gibbs

Editorial Assistant: Mark Ditzel

Managing Production Editor: Emma Hunt

Production Assistant: Angela Bennett

New Media Editor: Michael Morales

Copyeditor: Len Cegielka

Indexer: Liza Furnival

Illustration and Layout: Blink Studio, London

Manufacturing: Marion Morrow, Rory MacDonald

Garland Publishing, New York

ISBN 0 8153 3642 X (paperback) Garland

ISBN 0 4430 7098 9 (paperback) Churchill Livingstone

ISBN 0 4430 7099 7 (paperback) International Student Edition

© 2001 by Garland Publishing

Library of Congress Cataloging-in-Publication Data

Immunobiology : the immune system in health and disease / Charles A. Janeway, Jr. ... [et al.].--
5th ed.

p. cm.

Includes bibliographical references and index.

ISBN 0-8153-3642-X (pbk.) 1. Immunology. 2. Immunity. I. Janeway, Charles. II. Title.

QR181 .I454 2001

616.07'9--dc21 2001016039

Acknowledgments

Text

We would like to thank the following experts who read parts or the whole of the fourth edition chapters indicated and provided us with invaluable advice in developing this fifth edition.

Chapter 2: Ivan Lefkovits, Basel Institute for Immunology, Switzerland; Anthony T. Vella, Oregon State University.

Chapter 3: Sherie Morrison, University of California, Los Angeles; Michael S. Neuberger, MRC Laboratory of Molecular Biology, Cambridge.

Chapter 4: Ian A. Wilson, The Scripps Research Institute, La Jolla; Peter Cresswell, Yale University School of Medicine; Mark M. Davis, Stanford University School of Medicine; Paul M. Allen, Washington University School of Medicine, St. Louis; John Trowsdale, Cambridge University.

Chapter 5: John C. Cambier, National Jewish Medical and Research Center, Denver; Dan R. Littman, Skirball Institute of Biomolecular Medicine, New York; Arthur Weiss, The University of California, San Francisco.

Chapter 6: Richard R. Hardy, Fox Chase Cancer Center, Philadelphia; John G. Monroe, University of Pennsylvania Medical Center; Max D. Cooper, Comprehensive Cancer Center, University of Alabama; David Nemazee, The Scripps Research Institute, La Jolla; Michel C. Nussenzweig, Rockefeller University, New York.

Chapter 7: Alexander Y. Rudensky, University of Washington School of Medicine; Johnathan Sprent, The Scripps Research Institute, La Jolla; Leslie J. Berg, University of Massachusetts Medical School; Adrian C. Hayday, Guy's King's St Thomas' Medical School, University of London; Mike Owen, Imperial Cancer Research Fund, London; Robert H. Swanborg, Washington State University; Steve C. Jameson, University of Minnesota.

Chapter 8: Donna Paulnock, University of Wisconsin; Tim Springer, Center for Blood Research, Harvard Medical School; Marc K. Jenkins, University of Minnesota; Jürg Tschopp, University of Lausanne; Ralph Steinman, The Rockefeller University, New York.

Chapter 9: Michael C. Carroll, The Center for Blood Research, Harvard Medical School; E. Sally Ward, University of Texas; Jeffrey Ravetch, Rockefeller University, New York; Garnett Kelsoe, Duke University Medical Center, Durham; Douglas Fearon, University of Cambridge.

Chapter 10: Alan Ezekowitz, Massachusetts General Hospital, Harvard Medical School; Eric Pamer, Yale University School of Medicine; Adrian C. Hayday, Guy's King's St Thomas' Medical School, University of London.

Chapter 11: Fred Rosen, Center for Blood Research, Harvard Medical School; Robin A. Weiss, Royal Free and University College Medical School, London.

Chapter 12: Raif S. Geha, Children's Hospital, Harvard Medical School; Hugh A. Sampson, Mount Sinai Medical Center, New York; Philip W. Askenase, Yale University School of Medicine; Jeffrey Ravetch, The Rockefeller University, New York.

Chapter 13: Diane Mathis, Harvard Medical School; Christopher C. Goodnow, John Curtin School of Medical Research, Canberra; Jeffrey Ravetch, The Rockefeller University, New York; Kathryn Wood, University of Oxford; Hugh Auchincloss, Massachusetts General Hospital, Harvard Medical School; Joseph E. Craft, Yale University School of Medicine; Jan Erikson, The Wistar Institute, University of Pennsylvania; Keith Elkon, Cornell University, New York; Fiona Powrie, University of Oxford.

Chapter 14: Thierry Boon, Ludwig Institute for Cancer Research, Brussels; Gerry Crabtree, Stanford University School of Medicine; Jeffrey A. Bluestone, University of Chicago.

Appendix II: Joost J. Oppenheim, National Cancer Institute–Frederick Cancer Research and Development Center, Maryland.

Appendix III: Jason Cyster, University of California, San Francisco; Craig Gerard, Children's Hospital, Harvard Medical School.

Immunobiology Animations

We would like to thank Hung-Sia Teh of the University of British Columbia and David A. Lawlor of the Rochester Institute of Technology, for reviewing these animations.

Photographs

The following photographs have been reproduced with the kind permission of the journal in which they originally appeared.

Chapter 1

Fig. 1.1 courtesy of Yale University Harvey Cushing/John Hay Whitney Medical Library.

Fig. 1.9 photo from *The Journal of Experimental Medicine* 1972, **135**:200-214. © 1972 The Rockefeller University Press.

Chapter 2

Fig. 2.10 photo from *FEBS Letters* 1989, **250**:78-84. © 1989 Elsevier Science.

Fig. 2.13 photo from *The Journal of Immunology* 1990, **144**:2287-2294. © 1990 The American Association of Immunologists.

Fig. 2.24 photos from *Blut* 1990, **60**:309-318. © 1990 Springer-Verlag.

Fig. 2.39 photo from *Nature* 1994, **367**:338-345. © 1994 Macmillan Magazines Limited.

Chapter 3

Fig. 3.1 photo from *Nature* 1992, **360**:369-372. © 1992 Macmillan Magazines Limited.

Fig. 3.4 photo from *Advances in Immunology* 1969, **11**:1-30. © 1969 Academic Press.

Fig. 3.8 panel a from *Science* 1990, **248**:712-719. © 1990 American Association for the Advancement of Science; panel b from *Structure* 1993, **1**:83-93 © 1993 Current Biology.

Fig. 3.10 from *Science* 1986, **233**:747-753. © 1986 American Association for the Advancement of Science.

Fig. 3.13 photos from *Science* 1996, **274**:209-219. © 1996 American Association for the Advancement of Science.

Fig. 3.14 panel a from *Journal of Biological Chemistry* 1998, **263**:10541-10544. © 1998 American Society for Biochemistry and Molecular Biology.

Fig. 3.18 from *Nature* 1997, **387**:630-634. © 1997 Macmillan Magazines Limited.

Fig. 3.27 from *Science* 1996, **274**:209-219. © 1996 American Association for the Advancement of Science.

Fig 3.28 from *Science* 1999, **286**:1913-1921. © 1999 American Association for the Advancement of Science.

Chapter 4

Fig. 4.23 top photo from the *European Journal of Immunology* 1988, **18**:1001-1008. © 1988 Wiley-VCH.

Chapter 5

Fig. 5.4 from *Science* 1995, **268**:533-539. © 1995 American Association for the Advancement of Science.

Fig. 5.7 model structure from *Cell* 1996, **84**:505-507. © 1996 Cell Press.

Fig. 5.18 photo from *Nature* 1996, **384**:188-192. © 1996 Macmillan Magazines Limited.

Chapter 7

Fig. 7.3 panel b from the *European Journal of Immunology* 1987, **17**:1473-1484. © 1987 VCH Verlagsgesellschaft mbH.

Fig. 7.10 photos from *Nature* 1994, **372**:100-103. © 1994 Macmillan Magazines Limited.

Fig. 7.32 photos from *International Immunology* 1996, **8**:1537-1548. © 1996 Oxford University Press.

Chapter 8

Fig. 8.2 bottom panel from *Nature* 1997, **388**:787-792. © 1997 Macmillan Magazines Limited.

Fig. 8.29 panel c from *Second International Workshop on Cell Mediated Cytoxicity*. Eds. P.A. Henkart, and E. Martz. © 1985 Plenum Press.

Fig. 8.37 panels a and b from *Second International Workshop on Cell Mediated Cytoxicity* . Eds. P.A. Henkart, and E. Martz. © 1985 Plenum Press; panel c from *Immunology Today* 1985, **6**:21-27. © 1985 Elsevier Science.

Chapter 9

Fig. 9.15 left panel from *The Journal of Immunology* 1989, **134**:1349-1359. © 1989 The American Association of Immunologists. Middle and right panels from *Annual Reviews of Immunology* 1989, **7**:91-109. © 1989 Annual Reviews.

Fig. 9.21 from *Nature* 1994, **372**:336-343. © 1994 Macmillan Magazines Limited.

Fig. 9.27 planar conformation from the *European Journal of Immunology* 1988, **18**:1001-1008. © 1988 Wiley-VCH.

Chapter 11

Fig. 11.6 top panels from *International Reviews of Experimental Pathology* 1986, **28**:45-78, edited by M.A. Epstein and G.W. Richter. © 1986, Academic Press.

Fig. 11.26 from *Cell* 1998, **93**:665-671. © 1998 Cell Press Limited.

Fig. 11.27 from the *Nature* 1995, **373**:117-122. © 1995 Macmillan Magazines Limited.

Chapter 13

Fig. 13.20 photo from *Cell* 1989, **59**:247-255. © Cell Press.

Fig. 13.34 photos from *The Journal of Experimental Medicine* 1992, **176**:1355-1364. © 1992 The Rockefeller University Press.

Chapter 14

Fig. 14.16 photos from *Mechanisms of Cytoxicity by Natural Killer Cells*, edited by R.B. Herberman and D.M. Callewaert © 1985 Academic Press.

Appendix I

Fig. A.39 from *Nature* 2000, **403**:503-511. © 2000 Macmillan Magazines Limited.

3. Antigen Recognition by B-cell and T-cell Receptors

→ **Introduction to Chapter 3**

→ **The structure of a typical antibody molecule**

→ **The interaction of the antibody molecule with specific antigen**

→ **Summary to Chapter 3.**

→ **References to Chpater 3**

Introduction to Chapter 3

We have learned in Chapter 2 that the body is defended by innate immune responses, but these will only work to control pathogens that have certain molecular patterns or that induce interferons and other secreted yet non-specific defenses. Most crucially, they do not allow memory to form as they operate by receptors that are coded in the genome. Thus, innate immunity is good for preventing pathogens from growing freely in the body, but it does not lead to the most important feature of adaptive immunity, which is long-lasting memory of specific pathogen.

To recognize and fight the wide range of pathogens an individual will encounter, the lymphocytes of the adaptive immune system have evolved to recognize a great variety of different antigens from bacteria, viruses, and other disease-causing organisms. The antigen-recognition molecules of B cells are the immunoglobulins, or Ig. These proteins are produced by B cells in a vast range of antigen specificities, each B cell producing immunoglobulin of a single specificity (see Sections 1-8 to 1-10). Membrane-bound immunoglobulin on the B-cell surface serves as the cell's receptor for antigen, and is known as the **B-cell receptor** (**BCR**). Immunoglobulin of the same antigen specificity is secreted as antibody by terminally differentiated B cells—the plasma cells. The secretion of antibodies, which bind pathogens or their toxic products in the extracellular spaces of the body, is the main effector function of B cells in adaptive immunity.

Antibodies were the first molecules involved in specific immune recognition to be characterized and are still the best understood. The antibody molecule has two separate functions: one is to bind specifically to molecules from the pathogen that elicited the immune response; the other is to recruit other cells and molecules to destroy the pathogen once the antibody is bound to it. For example, binding by antibody neutralizes viruses and marks pathogens for destruction by phagocytes and complement, as described in Section 1-14. These functions are structurally separated in the antibody molecule, one part of which specifically recognizes and binds to the pathogen or antigen whereas the other engages different effector mechanisms. The antigen-binding region varies extensively between antibody molecules and is thus known as the variable region or **V region**. The variability of antibody molecules allows each antibody to bind a different specific antigen, and the total repertoire of antibodies made by a single individual is large enough to ensure that virtually any structure can be recognized. The region of the antibody molecule that engages the effector functions of the immune system does not vary in the same way and is thus known as the constant region or C region. It comes in five main forms, which are each specialized for activating different effector mechanisms. The membrane-bound B-cell receptor does not have these effector functions, as the C region remains inserted in the membrane of the B cell. Its function is as a receptor that recognizes and binds antigen by the V regions exposed on the surface of the cell, thus transmitting a signal that causes B-cell activation leading to clonal expansion and specific antibody production.

The antigen-recognition molecules of T cells are made solely as membrane-bound proteins and only function to signal T cells for activation. These T-cell receptors (**TCR**s) are related to immunoglobulins both in their protein structure—having both V and C regions—and in the genetic mechanism that produces their great variability (see Section 1-10 and Chapter 4). However, the T-cell receptor differs from the B-cell receptor in an important way: it does not recognize and bind antigen directly, but instead recognizes short peptide fragments of pathogen protein antigens, which are bound to MHC molecules on the surfaces of other cells.

The MHC molecules are glycoproteins encoded in the large cluster of genes known as the major histocompatibility complex (**MHC**) (see Sections 1-16 and 1-17). Their most striking structural feature is a cleft running across their outermost surface, in which a variety of peptides can be bound. As we shall discuss further in Chapter 5, MHC

molecules show great genetic variation in the population, and each individual carries up to 12 of the possible variants, which increases the range of pathogen-derived peptides that can be bound. T-cell receptors recognize features both of the peptide antigen and of the MHC molecule to which it is bound. This introduces an extra dimension to antigen recognition by T cells, known as **MHC restriction**, because any given T-cell receptor is specific not simply for a foreign peptide antigen, but for a unique combination of a peptide and a particular MHC molecule. The ability of T-cell receptors to recognize MHC molecules, and their selection during T-cell development for the ability to recognize the particular MHC molecules expressed by an individual, are topics we shall return to in Chapters 5 and 7.

In this chapter we focus on the structure and antigen-binding properties of immunoglobulins and T-cell receptors. Although B cells and T cells recognize foreign molecules in two distinct fashions, the receptor molecules they use for this task are very similar in structure. We will see how this basic structure can accommodate great variability in antigen specificity, and how it enables immunoglobulins and T-cell receptors to carry out their functions as the antigen-recognition molecules of the adaptive immune response.

The structure of a typical antibody molecule.

Antibodies are the secreted form of the B-cell receptor. An antibody is identical to the B-cell receptor of the cell that secretes it except for a small portion of the C-terminus of the heavy-chain constant region. In the case of the B-cell receptor the C-terminus is a hydrophobic membrane-anchoring sequence, and in the case of antibody it is a hydrophilic sequence that allows secretion. Since they are soluble, and secreted in large quantities, antibodies are easily obtainable and easily studied. For this reason, most of what we know about the B-cell receptor comes from the study of antibodies.

Antibody molecules are roughly Y-shaped molecules consisting of three equal-sized portions, loosely connected by a flexible tether. Three schematic representations of antibody structure, which has been determined by X-ray crystallography, are shown in Fig. 3.1. The aim of this part of the chapter is to explain how this structure is formed and how it allows antibody molecules to carry out their dual tasks—binding on the one hand to a wide variety of antigens, and on the other hand to a limited number of effector molecules and cells. As we will see, each of these tasks is carried out by separable parts of the molecule. The two arms of the Y end in regions that vary between different antibody molecules, the V regions. These are involved in antigen binding, whereas the stem of the Y, or the C region, is far less variable and is the part that interacts with effector cells and molecules.

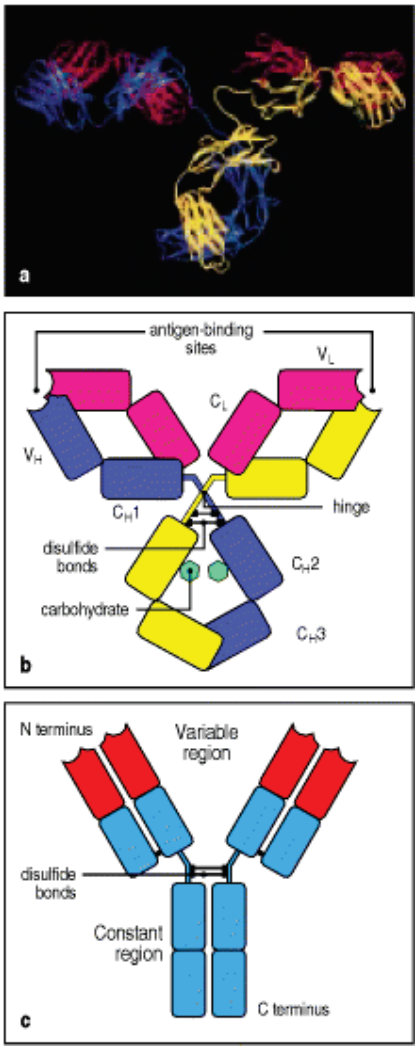


Figure 3.1. Structure of an antibody molecule. Panel a illustrates a ribbon diagram based on the X-ray crystallographic structure of an IgG antibody, showing the course of the backbones of the polypeptide chains. Three globular regions form a Y. The two antigen-binding sites are at the tips of the arms, which are tethered to the trunk of the Y by a flexible hinge region. A schematic representation of the structure in a is given in panel b, illustrating the four-chain composition and the separate domains comprising each chain. Panel c shows a simplified schematic representation of an antibody molecule that will be used throughout this book. Photograph courtesy of A. McPherson and L. Harris, reprinted with permission from Nature 360:369-372, ©1992 Macmillan Magazines Limited.

All antibodies are constructed in the same way from paired heavy and light polypeptide chains, and the generic term immunoglobulin is used for all such proteins. Within this general category, however, five different classes of immunoglobulins—IgM, IgD, IgG, IgA, and IgE—can be distinguished by their C regions, which will be described more fully in Chapter 4. More subtle differences confined to the V region account for the specificity of antigen binding. We will use the IgG antibody molecule as an example to describe the general structural features of immunoglobulins.

3-1. IgG antibodies consist of four polypeptide chains.

IgG antibodies are large molecules, having a molecular weight of approximately 150 kDa, composed of two different kinds of polypeptide chain. One, of approximately 50 kDa, is termed the **heavy** or **H chain**, and the other, of 25 kDa, is termed the **light** or L chain (Fig. 3.2). Each IgG molecule consists of two heavy chains and two light chains. The two heavy chains are linked to each other by disulfide bonds and each heavy chain is linked to a light chain by a disulfide bond. In any given immunoglobulin molecule, the two heavy chains and the two light chains are identical, giving an antibody molecule two identical antigen-binding sites (see Fig. 3.1), and thus the ability to bind simultaneously to two identical structures.

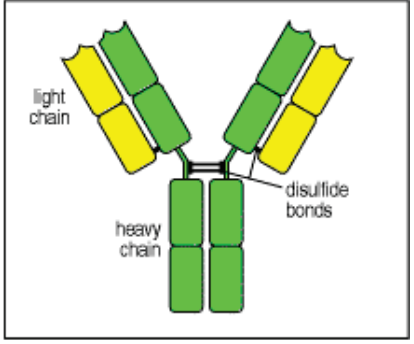


Figure 3.2. Immunoglobulin molecules are composed of two types of protein chain: heavy chains and light chains. Each immunoglobulin molecule is made up of two heavy chains (green) and two light chains (yellow) joined by disulfide bonds so that each heavy chain is linked to a light chain and the two heavy chains are linked together.

Two types of light chain, termed **lambda** (λ) and **kappa** (κ), are found in antibodies. A given immunoglobulin either has κ chains or λ chains, never one of each. No functional difference has been found between antibodies having λ or κ light chains, and either type of light chain may be found in antibodies of any of the five major classes. The ratio of the two types of light chain varies from species to species. In mice, the average κ to λ ratio is 20:1, whereas in humans it is 2:1 and in cattle it is 1:20. The reason for this variation is unknown. Distortions of this ratio can sometimes be used to detect the abnormal proliferation of a clone of B cells. These would all express the identical light chain, and thus an excess of λ light chains in a person might indicate the presence of a B-cell tumor producing λ chains.

By contrast, the class, and thus the effector function, of an antibody, is defined by the structure of its heavy chain. There are five main **heavy-chain classes** or isotypes, some of which have several subtypes, and these determine the functional activity of an antibody molecule. The five major classes of immunoglobulin are immunoglobulin M (IgM), immunoglobulin D (IgD), immunoglobulin G (IgG), immunoglobulin A (IgA), and immunoglobulin E (IgE). Their heavy chains are denoted by the corresponding lower-case Greek letter (μ, δ, γ, α, and ε, respectively). IgG is by far the most abundant immunoglobulin and has several subclasses (IgG1, 2, 3, and 4 in humans). Their distinctive functional properties are conferred by the carboxy-terminal part of the heavy chain, where it is not associated with the light chain. We will describe the structure and functions of the different heavy-chain isotypes in Chapter 4. The general structural features of all the isotypes are similar and we will consider IgG, the most abundant isotype in plasma, as a typical antibody molecule.

3-2. Immunoglobulin heavy and light chains are composed of constant and variable regions.

The amino acid sequences of many immunoglobulin heavy and light chains have been determined and reveal two important features of antibody molecules. First, each chain consists of a series of similar, although not identical, sequences, each about 110 amino acids long. Each of these repeats corresponds to a discrete, compactly folded region of protein structure known as a protein domain. The light chain is made up of two such immunoglobulin domains, whereas the heavy chain of the IgG antibody contains four (see Fig. 3.1a). This suggests that the immunoglobulin chains have evolved by repeated duplication of an ancestral gene corresponding to a single domain.

The second important feature revealed by comparisons of amino acid sequences is that the amino-terminal sequences of both the heavy and light chains vary greatly between different antibodies. The variability in sequence is limited to approximately the first 110 amino acids, corresponding to the first domain, whereas the remaining domains are constant between immunoglobulin chains of the same isotype. The amino-terminal variable or V domains of the heavy and light chains (V_H and V_L, respectively) together make up the V region of the antibody and confer on it the ability to bind specific antigen, while the constant domains (C domains) of the heavy and light chains (C_H and C_L, respectively) make up the C region (see Fig. 3.1b, c). The multiple heavy-chain C domains are numbered from the amino-terminal end to the carboxy terminus, for example C_H1, C_H2, and so on.

3-3. The antibody molecule can readily be cleaved into functionally distinct fragments.

The protein domains described above associate to form larger globular domains. Thus, when fully folded and assembled, an antibody molecule comprises three equal-sized globular portions joined by a flexible stretch of polypeptide chain known as the hinge region (see Fig. 3.1b). Each arm of this Y-shaped structure is formed by the

association of a light chain with the amino-terminal half of a heavy chain, whereas the trunk of the Y is formed by the pairing of the carboxy-terminal halves of the two heavy chains. The association of the heavy and light chains is such that the V_H and V_L domains are paired, as are the C_H1 and C_L domains. The C_H3 domains pair with each other but the C_H2 domains do not interact; carbohydrate side chains attached to the C_H2 domains lie between the two heavy chains. The two antigen-binding sites are formed by the paired V_H and V_L domains at the ends of the two arms of the Y (see Fig. 3.1b).

Proteolytic enzymes (proteases) that cleave polypeptide sequences have been used to dissect the structure of antibody molecules and to determine which parts of the molecule are responsible for its various functions. Limited digestion with the protease papain cleaves antibody molecules into three fragments (Fig. 3.3). Two fragments are identical and contain the antigen-binding activity. These are termed the Fab fragments, for **F**ragment **a**ntigen **b**inding. The Fab fragments correspond to the two identical arms of the antibody molecule, which contain the complete light chains paired with the V_H and C_H1 domains of the heavy chains. The other fragment contains no antigen-binding activity but was originally observed to crystallize readily, and for this reason was named the Fc fragment, for **F**ragment **c**rystallizable. This fragment corresponds to the paired C_H2 and C_H3 domains and is the part of the antibody molecule that interacts with effector molecules and cells. The functional differences between heavy-chain isotypes lie mainly in the Fc fragment.

The protein fragments obtained after proteolysis are determined by where the protease cuts the antibody molecule in relation to the disulfide bonds that link the two heavy chains. These lie in the hinge region between the C_H1 and C_H2 domains and, as illustrated in Fig. 3.3, papain cleaves the antibody molecule on the amino-terminal side of the disulfide bonds. This releases the two arms of the antibody as separate Fab fragments, whereas in the Fc fragment the carboxy-terminal halves of the heavy chains remain linked.

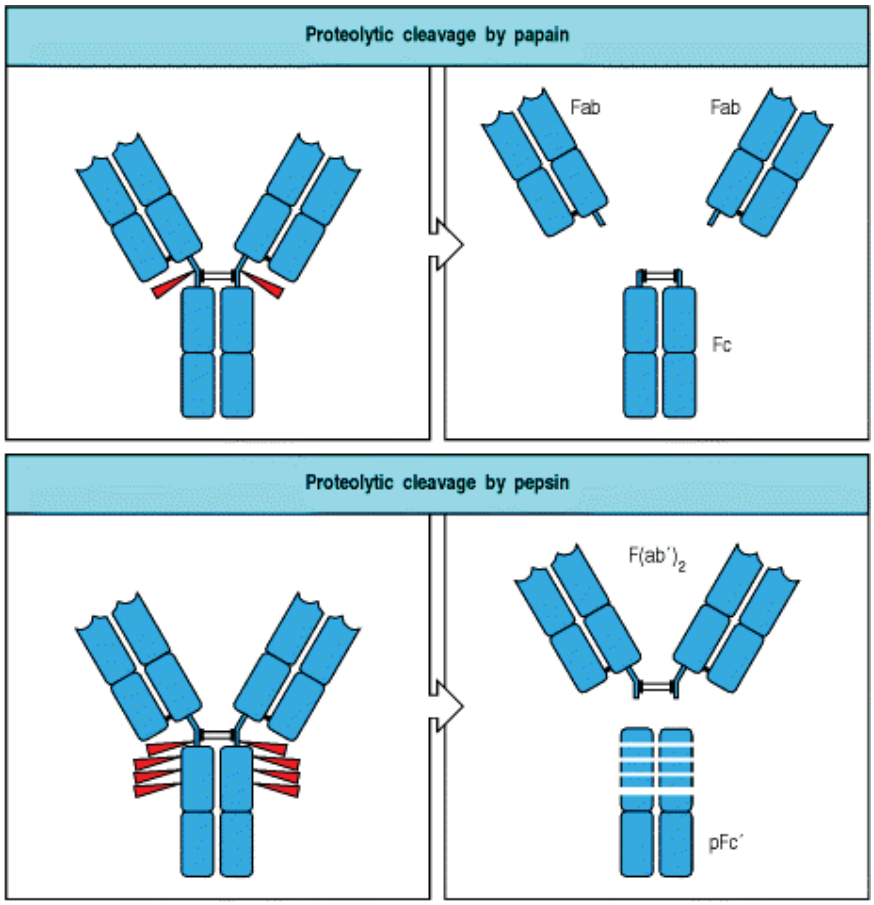


Figure 3.3. The Y-shaped immunoglobulin molecule can be dissected by partial digestion with proteases. Papain cleaves the immunoglobulin molecule into three pieces, two Fab fragments and one Fc fragment (upper panels). The Fab fragment contains the V regions and binds antigen. The Fc fragment is crystallizable and contains C regions. Pepsin cleaves immunoglobulin to yield one $F(ab')_2$ fragment and many small pieces of the Fc fragment, the largest of which is called the pFc′ fragment (lower panels). $F(ab')_2$ is written with a prime because it contains a few more amino acids than Fab, including the cysteines that form the disulfide bonds.

Another protease, pepsin, cuts in the same general region of the antibody molecule as papain but on the carboxy-terminal side of the disulfide bonds (see Fig. 3.3). This produces a fragment, the **$F(ab')_2$ fragment**, in which the two antigen-binding arms of the antibody molecule remain linked. In this case the remaining part of the heavy chain is cut into several small fragments. The $F(ab')_2$ fragment has exactly the same antigen-binding characteristics as the original antibody but is unable to interact with any effector molecule. It is thus of potential value in therapeutic applications of antibodies as well as in research into the functional role of the Fc portion.

Genetic engineering techniques also now permit the construction of many different antibody-related molecules. One important type is a truncated Fab comprising only the V domain of a heavy chain linked by a stretch of synthetic peptide to a V domain of a light chain. This is called single-chain Fv, named from **F**ragment **v**ariable. Fv molecules may become valuable therapeutic agents because of their small size, which allows them to penetrate tissues readily. They can be coupled to protein toxins to yield immunotoxins with potential application, for example, in tumor therapy in the case of a Fv specific for a tumor antigen (see Chapter 14).

3-4. The immunoglobulin molecule is flexible, especially at the hinge region.

The hinge region that links the Fc and Fab portions of the antibody molecule is in reality a flexible tether, allowing independent movement of the two Fab arms, rather than a rigid hinge. This has been demonstrated by electron microscopy of antibodies bound to haptens. These are small molecules of various sorts, typically about the size of a tyrosine side chain. They can be recognized by antibody but are only able to stimulate production of antihapten antibodies when linked to a larger protein **carrier**(see Appendix I, Section A-1). An antigen made of two identical hapten molecules joined by a short flexible region can link two or more anti-hapten antibodies, forming dimers, trimers, tetramers, and so on, which can be seen by electron microscopy (Fig. 3.4). The shapes formed by these complexes demonstrate that antibody molecules are flexible at the hinge region. Some flexibility is also found at the junction between the V and C domains, allowing bending and rotation of the V domain relative to the C domain. For example, in the antibody molecule shown in Fig. 3.1a, not only are the two hinge regions clearly bent differently, but the angle between the V and C domains in each of the two Fab arms is also different. This range of motion has led to the junction between the V and C domains being referred to as a 'molecular balland-socket joint.' Flexibility at both the hinge and V-C junction enables the binding of both arms of an antibody molecule to sites that are various distances apart, for example, sites on bacterial cell-wall polysaccharides. Flexibility at the hinge also enables the antibodies to interact with the antibody-binding proteins that mediate immune effector mechanisms.

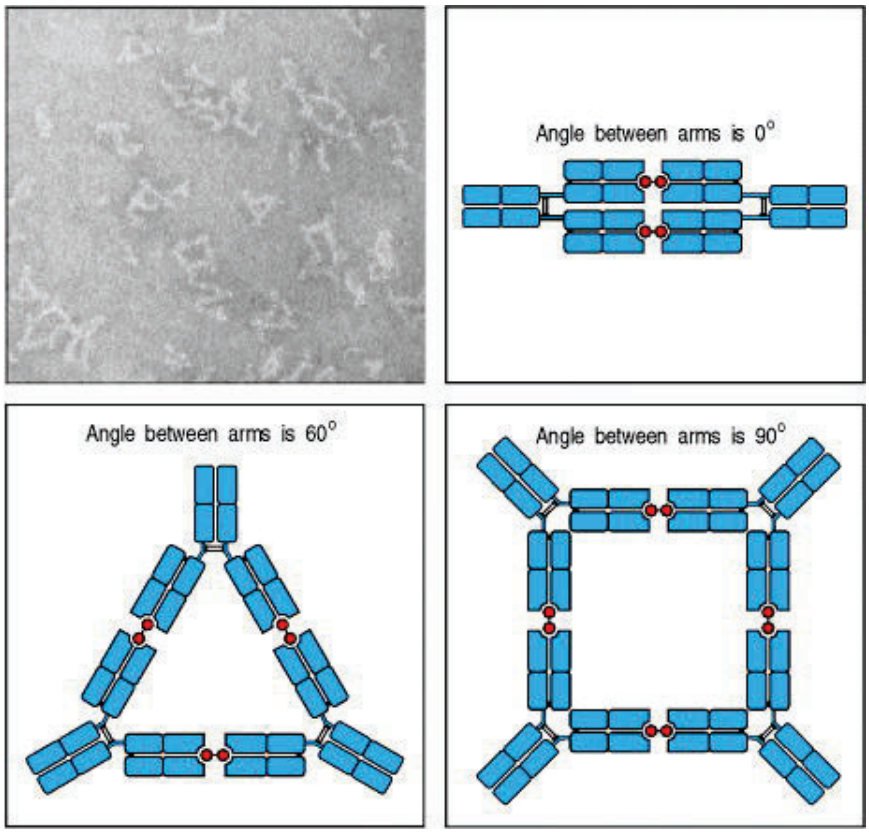


Figure 3.4. Antibody arms are joined by a flexible hinge. An antigen consisting of two hapten molecules (red balls

in diagrams) that can cross-link two antigen-binding sites is used to create antigen:antibody complexes, which can be seen in the electron micrograph. Linear, triangular, and square forms are seen, with short projections or spikes. Limited pepsin digestion removes these spikes (not shown in the figure), which therefore correspond to the Fc portion of the antibody; the $F(ab')_2$ pieces remain cross-linked by antigen. The interpretation of the complexes is shown in the diagrams. The angle between the arms of the antibody molecules varies, from 0° in the antibody dimers, through 60° in the triangular forms, to 90° in the square forms, showing that the connections between the arms are flexible. Photograph (× 300,000) courtesy of N.M. Green.

3-5. The domains of an immunoglobulin molecule have similar structures.

As we saw in Section 3-2, immunoglobulin heavy and light chains are composed of a series of discrete protein domains. These protein domains all have a similar folded structure. Within this basic three-dimensional structure, there are distinct differences between V and C domains. The structural similarities and differences can be seen in the diagram of a light chain in Fig. 3.5. Each domain is constructed from two β sheets, which are elements of protein structure made up of strands of the polypeptide chain (β strands) packed together; the sheets are linked by a disulfide bridge and together form a roughly barrel-shaped structure, known as a β barrel. The distinctive folded structure of the immunoglobulin protein domain is known as the **immunoglobulin fold**.

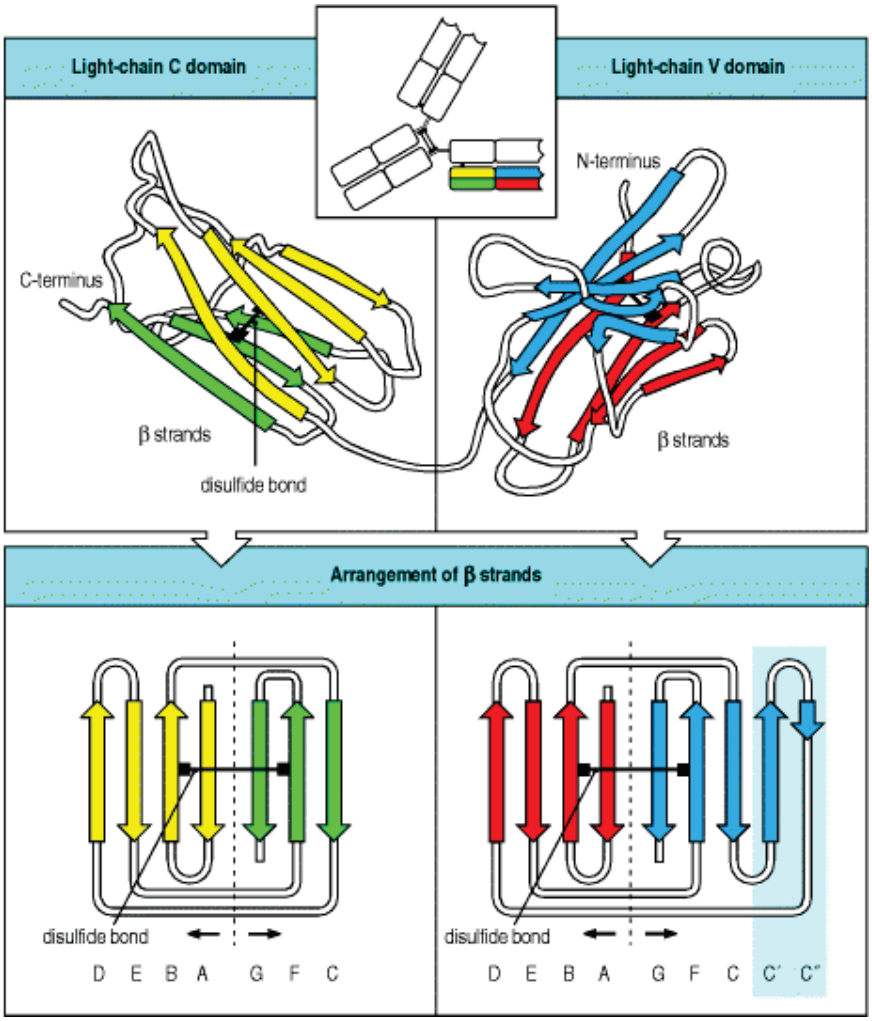


Figure 3.5. The structure of immuno-globulin constant and variable domains. The upper panels show schematically the folding pattern of the constant (C) and variable (V) domains of an immunoglobulin light chain. Each domain is a barrel-shaped structure in which strands of polypeptide chain (β strands) running in opposite directions (antiparallel) pack together to form two β sheets (shown in yellow and green in the diagram of the C domain), which are held together by a disulfide bond. The way the polypeptide chain folds to give the final structure can be seen more clearly when the sheets are opened out, as shown in the lower panels. The β strands are lettered sequentially with respect to the order of their occurrence in the amino acid sequence of the domains; the order in each β sheet is characteristic of immunoglobulin domains. The β strands C' and C'' that are found in the V domains but not in the C domains are indicated by a blue shaded background. The characteristic four-strand plus three-strand (C-

region type domain) or four-strand plus five-strand (V-region type domain) arrange-ments are typical immuno-globulin superfamily domain building blocks, found in a whole range of other proteins as well as antibodies and T-cell receptors.

Both the essential similarity of V and C domains and the critical difference between them are most clearly seen in the bottom panels of Fig. 3.5, where the cylindrical domains are opened out to reveal how the polypeptide chain folds to create each of the β sheets and how it forms flexible loops as it changes direction. The main difference between the V and C domains is that the V domain is larger, with an extra loop. We will see in Section 3-6 that the flexible loops of the V domains form the antigen-binding site of the immunoglobulin molecule.

Many of the amino acids that are common to C and V domains of immuno-globulin chains lie in the core of the immunoglobulin fold and are critical to its stability. For that reason, other proteins having sequences similar to those of immunoglobulins are believed to form domains of similar structure, and in many cases this has been demonstrated by crystallography. These **immunoglobulin-like domains** are present in many other proteins of the immune system, and in proteins involved in cell-cell recognition in the nervous system and other tissues. Together with the immunoglobulins and the T-cell receptors, they make up the extensive immunoglobulin superfamily.

Summary.

The IgG antibody molecule is made up of four polypeptide chains, comprising two identical light chains and two identical heavy chains, and can be thought of as forming a flexible Y-shaped structure. Each of the four chains has a variable (V) region at its amino terminus, which contributes to the antigen-binding site, and a constant (C) region, which determines the isotype. The isotype of the heavy chain determines the functional properties of the antibody. The light chains are bound to the heavy chains by many noncovalent interactions and by disulfide bonds, and the V regions of the heavy and light chains pair in each arm of the Y to generate two identical antigen-binding sites, which lie at the tips of the arms of the Y. The possession of two antigen-binding sites allows antibody molecules to cross-link antigens and to bind them much more stably. The trunk of the Y, or Fc fragment, is composed of the carboxy-terminal domains of the heavy chains. Joining the arms of the Y to the trunk are the flexible hinge regions. The Fc fragment and hinge regions differ in antibodies of different isotypes, thus determining their functional properties. However, the overall organization of the domains is similar in all isotypes.

The interaction of the antibody molecule with specific antigen.

We have described the structure of the antibody molecule and how the V regions of the heavy and light chains fold and pair to form the antigen-binding site. In this part of the chapter we will look at the antigen-binding site in more detail. We will discuss the different ways in which antigens can bind to antibody and address the question of how variation in the sequences of the antibody V domains determines the specificity for antigen.

3-6. Localized regions of hypervariable sequence form the antigenbinding site.

The V regions of any given antibody molecule differ from those of every other. Sequence variability is not, however, distributed evenly throughout the V regions but is concentrated in certain segments of the V region. The distribution of variable amino acids can be seen clearly in what is termed a variability plot (Fig. 3.6), in which the amino acid sequences of many different antibody V regions are compared. Three segments of particular variability can be identified in both the V_H and V_L domains. They are designated **hypervariable regions** and are denoted HV1, HV2, and HV3. In the light chains these are roughly from residues 28 to 35, from 49 to 59, and from 92 to 103, respectively. The most variable part of the domain is in the HV3 region. The regions between the hypervariable regions, which comprise the rest of the V domain, show less variability and are termed the framework regions. There are four such regions in each V domain, designated FR1, FR2, FR3, and FR4.

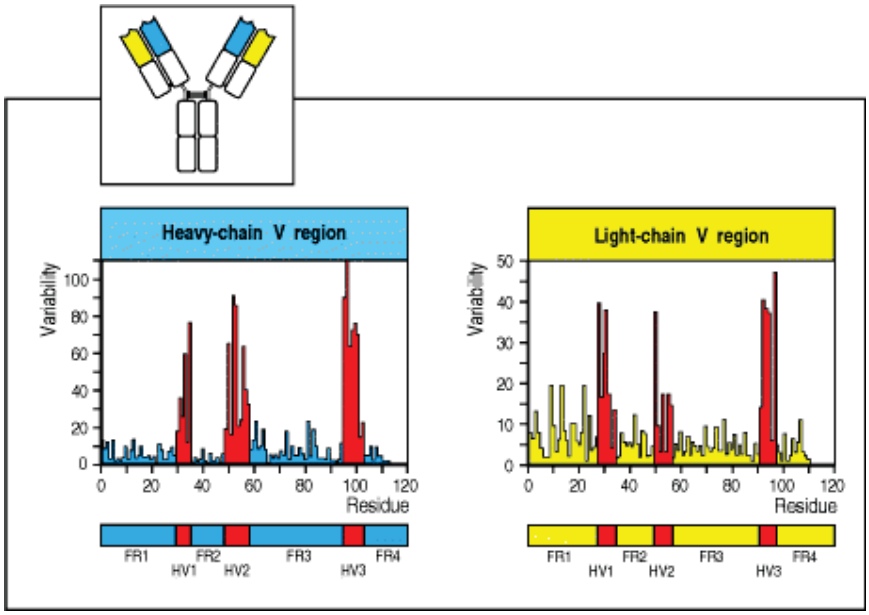


Figure 3.6. There are discrete regions of hypervariability in V domains. A variability plot derived from comparison of the amino acid sequences of several dozen heavy-chain and light-chain V domains. At each amino acid position the degree of variability is the ratio of the number of different amino acids seen in all of the sequences together to the frequency of the most common amino acid. Three hypervariable regions (HV1, HV2, and HV3) are indicated in red and are also known as the complementarity-determining regions, CDR1, CDR2, and CDR3. They are flanked by less variable framework regions (FR1, FR2, FR3, and FR4, shown in blue or yellow).

The framework regions form the β sheets that provide the structural framework of the domain, whereas the hypervariable sequences correspond to three loops at the outer edge of the β barrel, which are juxtaposed in the folded domain (Fig. 3.7). Thus, not only is sequence diversity concentrated in particular parts of the V domain but it is localized to a particular region on the surface of the molecule. When the V_H and V_L domains are paired in the antibody molecule, the hypervariable loops from each domain are brought together, creating a single hypervariable site at the tip of each arm of the molecule. This is the binding site for antigen, the antigen-binding site or antibody combining site. The three hypervariable loops determine antigen specificity by forming a surface complementary to the antigen, and are more commonly termed the **complementarity-determining regions**, or CDRs (**CDR1**, **CDR2**, and **CDR3**). Because CDRs from both V_H and V_L domains contribute to the antigen-binding site, it is the combination of the heavy and the light chain, and not either alone, that determines the final antigen specificity. Thus, one way in which the immune system is able to generate antibodies of different specificities is by generating different combinations of heavy- and light-chain V regions. This means of producing variability is known as combinatorial diversity; we will encounter a second form of combinatorial diversity when we consider in Chapter 4 how the genes encoding the heavy- and light-chain V regions are created from smaller segments of DNA.

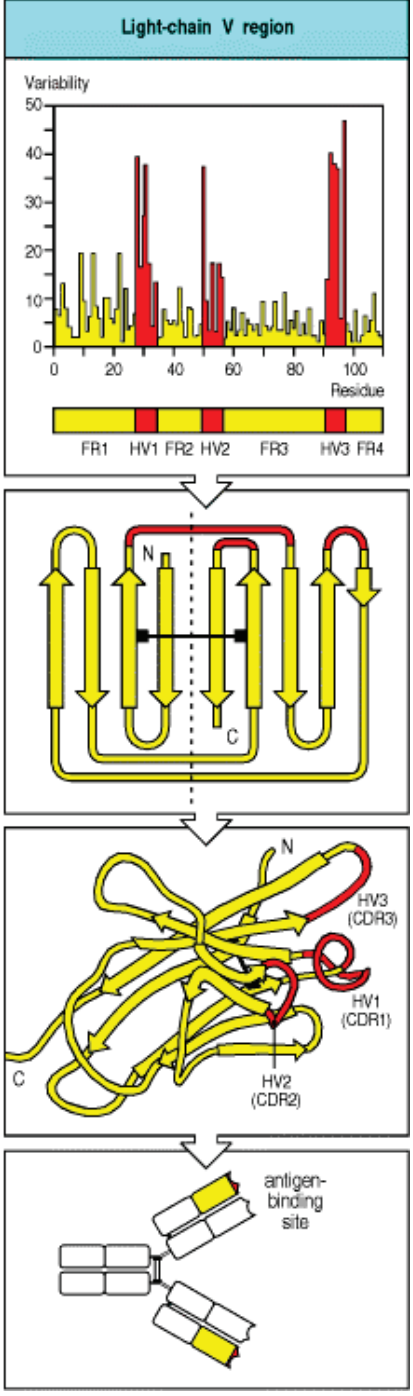


Figure 3.7. The hypervariable regions lie in discrete loops of the folded structure. When the hypervariable regions (CDRs) are positioned on the structure of a V domain it can be seen that they lie in loops that are brought together in the folded structure. In the antibody molecule, the pairing of a heavy and a light chain brings together the hypervariable loops from each chain to create a single hypervariable surface, which forms the antigen-binding site at the tip of each arm.

3-7. Antibodies bind antigens via contacts with amino acids in CDRs, but the details of binding depend upon the size and shape of the antigen.

In early investigations of antigen binding to antibodies, the only available sources of large quantities of a single type of antibody molecule were tumors of antibody-secreting cells. The antigen specificities of the tumor-derived antibodies were unknown, so many compounds had to be screened to identify ligands that could be used to study antigen binding. In general, the substances found to bind to these antibodies were haptens (see Section 3-4) such as phosphorylcholine or vitamin K_1. Structural analysis of complexes of antibodies with their hapten ligands provided the first direct evidence that the hypervariable regions form the antigen-binding site, and demonstrated the structural basis of specificity for the hapten. Subsequently, with the discovery of methods of generating monoclonal antibodies (see Appendix I, Section A-12), it became possible to make large amounts of pure antibodies specific for many

different antigens. This has provided a more general picture of how antibodies interact with their antigens, confirming and extending the view of antibody-antigen interactions derived from the study of haptens.

The surface of the antibody molecule formed by the juxtaposition of the CDRs of the heavy and light chains creates the site to which an antigen binds. Clearly, as the amino acid sequences of the CDRs are different in different antibodies, so are the shapes of the surfaces created by these CDRs. As a general principle, antibodies bind ligands whose surfaces are complementary to that of the antibody. A small antigen, such as a hapten or a short peptide, generally binds in a pocket or groove lying between the heavy- and light-chain V domains (Fig. 3.8, left and center panels). Other antigens, such as a protein molecule, can be of the same size as, or larger than, the antibody molecule itself, and cannot fit into a groove or pocket. In these cases, the interface between the two molecules is often an extended surface that involves all of the CDRs and, in some cases, part of the framework region of the antibody (Fig. 3.8, right panel). This surface need not be concave but can be flat, undulating, or even convex.

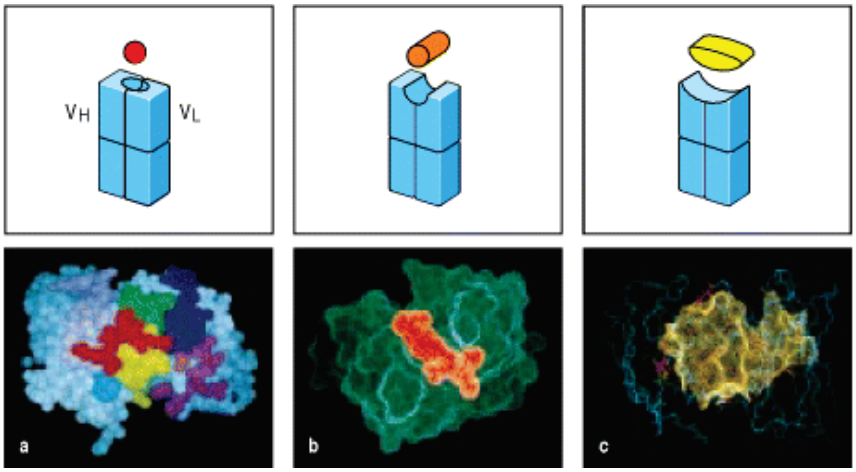


Figure 3.8. Antigens can bind in pockets or grooves, or on extended surfaces in the binding sites of antibodies. The panels in the top row show schematic representations of the different types of binding site in a Fab fragment of an antibody: left, pocket; center, groove; right, extended surface. Below are examples of each type. Panel a: space-filling representation of the interaction of a small peptide antigen with the complementarity-determining regions (CDRs) of a Fab fragment as viewed looking into the antigen-binding site. Seven amino acid residues of the antigen, shown in red, are bound in the antigen-binding pocket. Five of the six CDRs (H1, H2, H3, L1, and L3) interact with the peptide, whereas L2 does not. The CDR loops are colored as follows: L2, magenta; L3, green; H1, blue; H2, pale purple; H3, yellow. Panel b: in a complex of an antibody with a peptide from the human immunodeficiency virus, the peptide (orange) binds along a groove formed between the heavy- and light-chain V domains (green). Panel c: complex between hen egg-white lysozyme and the Fab fragment of its corresponding antibody (HyHel5). Two extended surfaces come into contact, as can be seen from this computer-generated image, where the surface contour of the lysozyme molecule (yellow dots) is superimposed on the antigen-binding site. Residues in the antibody that make contact with the lysozyme are shown in full (red); for the rest of the Fab fragment only the peptide backbone is shown (blue). All six CDRs of the antibody are involved in the binding. Photographs a and b courtesy of I.A. Wilson and R.L. Stanfield, reprinted with permission from Science 248:712-719. ©1990 American Association for the Advancement of Science. Photograph c courtesy of S. Sheriff.

3-8. Antibodies bind to conformational shapes on the surfaces of antigens.

The biological function of antibodies is to bind to pathogens and their products, and to facilitate their removal from the body. An antibody generally recognizes only a small region on the surface of a large molecule such as a polysaccharide or protein. The structure recognized by an antibody is called an antigenic determinant or epitope. Some of the most important pathogens have polysaccharide coats, and antibodies that recognize epitopes formed by the sugar subunits of these molecules are essential in providing immune protection from such pathogens. In many cases, however, the antigens that provoke an immune response are proteins. For example, protective antibodies against viruses recognize viral coat proteins. In such cases, the structures recognized by the antibody are located on the surface of the protein. Such sites are likely to be composed of amino acids from different parts of the polypeptide chain that have been brought together by protein folding. Antigenic determinants of this kind are known as **conformational** or discontinuous epitopes because the structure recognized is composed of segments of the protein that are discontinuous in the amino acid sequence of the antigen but are brought together in the three-dimensional structure. In contrast, an epitope composed of a single segment of polypeptide chain is termed a **continuous** or linear

epitope. Although most antibodies raised against intact, fully folded proteins recognize discontinuous epitopes, some will bind peptide fragments of the protein. Conversely, antibodies raised against peptides of a protein or against synthetic peptides corresponding to part of its sequence are occasionally found to bind to the natural folded protein. This makes it possible, in some cases, to use synthetic peptides in vaccines that aim at raising antibodies against a pathogen protein.

3-9. Antigen-antibody interactions involve a variety of forces.

The interaction between an antibody and its antigen can be disrupted by high salt concentrations, extremes of pH, detergents, and sometimes by competition with high concentrations of the pure epitope itself. The binding is therefore a reversible noncovalent interaction. The forces, or bonds, involved in these noncovalent interactions are outlined in Fig. 3.9.

Noncovalent forces	Origin	
Electrostatic forces	Attraction between opposite charges	$-\overset{\oplus}{N}H_3 \quad \overset{\ominus}{O}OC-$
Hydrogen bonds	Hydrogen shared between electronegative atoms (N,O)	$>N^{\delta-}-H^{\delta+}\cdots O^{\delta-}=C<$
Van der Waals forces	Fluctuations in electron clouds around molecules oppositely polarize neighboring atoms	$\delta^+\ \delta^- \rightleftharpoons \delta^-\ \delta^+$
Hydrophobic forces	Hydrophobic groups interact unfavorably with water and tend to pack together to exclude water molecules. The attraction also involves van der Waals forces	

Figure 3.9. The noncovalent forces that hold together the antigen:antibody complex. Partial charges found in electric dipoles are shown as δ^+ or δ^-. Electrostatic forces diminish as the inverse square of the distance separating the charges, whereas van der Waals forces, which are more numerous in most antigen-antibody contacts, fall off as the sixth power of the separation and therefore operate only over very short ranges. Covalent bonds never occur between antigens and naturally produced antibodies.

Electrostatic interactions occur between charged amino acid side chains, as in salt bridges. Interactions also occur between electric dipoles, as in hydrogen bonds, or can involve short-range van der Waals forces. High salt concentrations and extremes of pH disrupt antigen-antibody binding by weakening electrostatic interactions and/or hydrogen bonds. This principle is employed in the purification of antigens using affinity columns of immobilized antibodies, and vice versa for antibody purification (see Appendix I, Section A-5). Hydrophobic interactions occur when two hydrophobic surfaces come together to exclude water. The strength of a hydrophobic interaction is proportional to the surface area that is hidden from water. For some antigens, hydrophobic interactions probably account for most of the binding energy. In some cases, water molecules are trapped in pockets in the interface between antigen and antibody. These trapped water molecules may also contribute to binding, especially between polar amino acid residues.

The contribution of each of these forces to the overall interaction depends on the particular antibody and antigen involved. A striking difference between antibody interactions with protein antigens and most other natural protein-protein interactions is that antibodies possess many aromatic amino acids in their antigen-binding sites. These amino acids participate mainly in van der Waals and hydrophobic interactions, and sometimes in hydrogen bonds. In general, the hydrophobic and van der Waals forces operate over very short ranges and serve to pull together two surfaces that are complementary in shape: hills on one surface must fit into valleys on the other for good binding to occur. In contrast, electrostatic interactions between charged side chains, and hydrogen bonds bridging oxygen and/or nitrogen atoms, accommodate specific features or reactive groups while strengthening the interaction overall.

For example, in the complex of hen egg-white lysozyme with the antibody D1.3 (Fig. 3.10), strong hydrogen bonds are formed between the antibody and a particular glutamine in the lysozyme molecule that protrudes between the V_H and V_L domains. Lysozymes from partridge and turkey have another amino acid in place of the glutamine and do not bind to the antibody. In the high-affinity complex of hen egg-white lysozyme with another antibody, HyHel5 (see Fig. 3.8c), two salt bridges between two basic arginines on the surface of the lysozyme interact with two glutamic acids, one each from the V_H CDR1 and CDR2 loops. Again, lysozymes that lack one of the two arginine residues show a 1000-fold decrease in affinity. Although overall surface complementarity must play an important part in antigen-antibody inter-actions, specific electrostatic and hydrogen-bonding interactions appear to determine antibody affinity. In most antibodies that have been studied at this level of detail, only a few residues make a major contribution to the binding energy. Genetic engineering by site-directed mutagenesis can further tailor an antibody's binding to its complementary epitope.

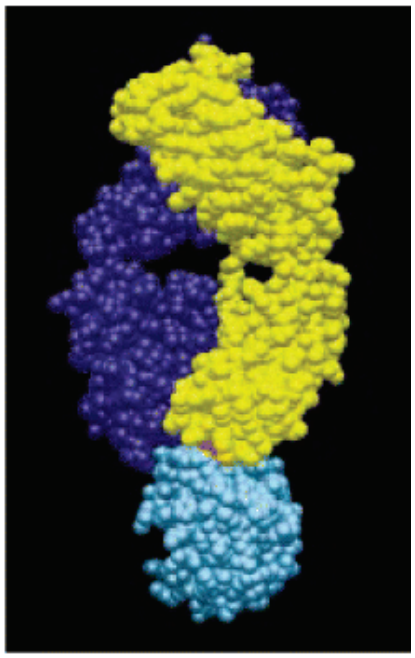


Figure 3.10. The complex of lysozyme with the antibody D1.3. The interaction of the Fab fragment of D1.3 with hen egg-white lysozyme is shown, with the lysozyme in blue, the heavy chain in purple and the light chain in yellow. A glutamine residue of lysozyme, shown in red, protrudes between the two V domains of the antigen-binding site and makes hydrogen bonds important to the antigen-antibody binding. Original photograph courtesy of R.J. Poljak, reprinted with permission from Science 233:747-753. ©1986 American Association for the Advancement of Science.

Summary.

X-ray crystallographic analysis of antigen:antibody complexes has demonstrated that the hypervariable loops (complementarity-determining regions) of immunoglobulin V regions determine the specificity of antibodies. With protein antigens, the antibody molecule contacts the antigen over a broad area of its surface that is complementary to the surface recognized on the antigen. Electrostatic interactions, hydrogen bonds, van der Waals forces, and hydrophobic interactions can all contribute to binding. Amino acid side chains in most or all of the hypervariable loops make contact with antigen and determine both the specificity and the affinity of the interaction. Other parts of the V region play little part in the direct contact with the antigen but provide a stable structural framework for the hypervariable loops and help determine their position and conformation. Antibodies raised against intact proteins usually bind to the surface of the protein and make contact with residues that are discontinuous in the primary structure of the molecule; they may, however, occasionally bind peptide fragments of the protein, and antibodies raised against peptides derived from a protein can sometimes be used to detect the native protein molecule. Peptides binding to antibodies usually bind in the cleft between the V regions of the heavy and light chains, where they make specific contact with some, but not necessarily all, of the hypervariable loops. This is also the usual mode of binding for carbohydrate antigens and small molecules such as haptens.

Antigen recognition by T cells.

In contrast to the immunoglobulins, which interact with pathogens and their toxic products in the extracellular spaces of the body, T cells only recognize foreign antigens that are displayed on the surfaces of the body's own cells. These antigens can derive from pathogens such as viruses or intracellular bacteria, which replicate within cells, or from pathogens or their products that cells have internalized by endocytosis from the extracellular fluid.

T cells can detect the presence of an intracellular pathogen because infected cells display on their surface peptide fragments derived from the pathogen's proteins. These foreign peptides are delivered to the cell surface by specialized host-cell glycoproteins. These are encoded in a large cluster of genes that were first identified by their powerful effects on the immune response to transplanted tissues. For that reason, the gene complex was called the **major histocompatibility complex (MHC)**, and the peptide-binding glycoproteins are still known as MHC molecules. The recognition of antigen as a small peptide fragment bound to an MHC molecule and displayed at the cell surface is one of the most distinctive features of T cells, and will be the focus of this part of the chapter. How peptide fragments of antigen become complexed with MHC molecules will be considered in Chapter 5.

In this part of the chapter we will describe the structure and properties of the T-cell antigen receptor, T-cell receptor, or TCR for short. As might be expected from their function as highly variable antigen-recognition structures, T-cell receptors are closely related to antibody molecules in the structure of their genes. There are, however, important differences between T-cell receptors and immunoglobulins that reflect the special features of antigen recognition by the T-cell receptor, and its lack of effector functions.

3-10. The antigen receptor on T cells is very similar to a Fab fragment of immunoglobulin.

T-cell receptors were first identified using monoclonal antibodies that bound only one cloned T-cell line but not others and that could specifically inhibit antigen recognition by that clone of T cells, or specifically activate them (see Appendix I, Section A-19). These clonotypic antibodies were then used to show that each T cell bears about 30,000 antigen-receptor molecules on its surface, each receptor consisting of two different polypeptide chains, termed the **T-cell receptor α** (**TCRα**) and **β** (**TCRβ**) **chains**, linked by a disulfide bond. These α:β heterodimers are very similar in structure to the Fab fragment of an immunoglobulin molecule (Fig. 3.11), and they account for antigen recognition by most T cells. A minority of T cells bear an alternative, but structurally similar, receptor made up of a different pair of polypeptide chains designated γ and δ. γ:δ T-cell receptors appear to have different antigen-recognition properties from the α:β T-cell receptors, and the function of γ:δ T cells in immune responses is not yet entirely clear. In the rest of this chapter, we shall use the term T-cell receptor to mean the α:β receptor, except where specified otherwise. Both types of T-cell receptor differ from the membrane-bound immunoglobulin that serves as the B-cell receptor: a T-cell receptor has only one antigen-binding site, whereas a B-cell receptor has two, and T-cell receptors are never secreted, whereas immunoglobulin can be secreted as antibody.

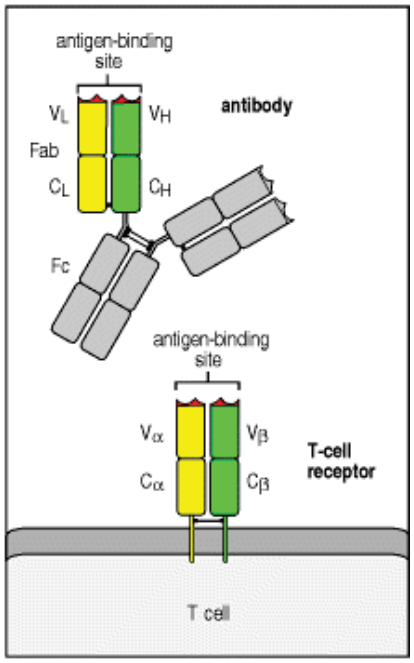


Figure 3.11. The T-cell receptor resembles a membrane-bound Fab fragment. The Fab fragment of antibody molecules is a disulfide-linked heterodimer, each chain of which contains one immunoglobulin C domain and one V

domain; the juxtaposition of the V domains forms the antigen-binding site (see Section 3-6). The T-cell receptor is also a disulfide-linked heterodimer, with each chain containing an immunoglobulin C-like domain and an immunoglobulin V-like domain. As in the Fab fragment, the juxtaposition of the V domains forms the site for antigen recognition.

Our initial insights into the structure and function of the α:β T-cell receptor came from studies of cloned cDNA encoding the receptor chains. The amino acid sequences predicted from T-cell receptor cDNAs show clearly that both chains of the T-cell receptor have an amino-terminal variable (V) region with homology to an immunoglobulin V domain, a constant (C) region with homology to an immunoglobulin C domain, and a short hinge region containing a cysteine residue that forms the interchain disulfide bond (Fig. 3.12). Each chain spans the lipid bilayer by a hydrophobic transmembrane domain, and ends in a short cytoplasmic tail. These close similarities of T-cell receptor chains to the heavy and light immunoglobulin chains first enabled prediction of the structural resemblance of the T-cell receptor heterodimer to a Fab fragment of immunoglobulin.

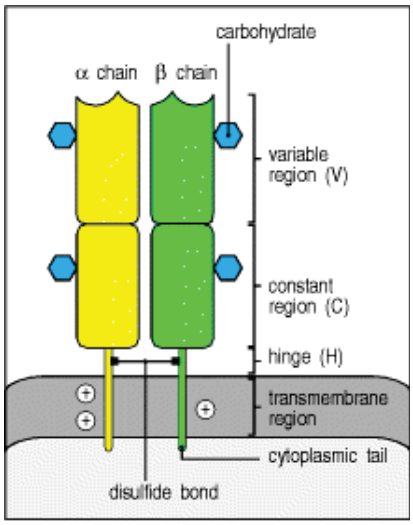


Figure 3.12. Structure of the T-cell receptor. The T-cell receptor heterodimer is composed of two trans-membrane glycoprotein chains, α and β. The extracellular portion of each chain consists of two domains, resembling immunoglobulin V and C domains, respectively. Both chains have carbohydrate side chains attached to each domain. A short segment, analogous to an immunoglobulin hinge region, connects the immunoglobulin-like domains to the membrane and contains the cysteine residue that forms the interchain disulfide bond. The trans-membrane helices of both chains are unusual in containing positively charged (basic) residues within the hydrophobic transmembrane segment. The α chains carry two such residues; the β chains have one.

Recently, the three-dimensional structure of the T-cell receptor has been determined. The structure is indeed similar to that of an antibody Fab fragment, as was suspected from earlier studies on the genes that encoded it. The T-cell receptor chains fold in much the same way as those of a Fab fragment (Fig. 3.13a), although the final structure appears a little shorter and wider. There are, however, some distinct differences between T-cell receptors and Fab fragments. The most striking difference is in the C_α domain, where the fold is unlike that of any other immunoglobulin-like domain. The half of the domain that is juxtaposed with the C_β domain forms a β sheet similar to that found in other immunoglobulin-like domains, but the other half of the domain is formed of loosely packed strands and a short segment of α helix (Fig. 3.13b). The intramolecular disulfide bond, which in immunoglobulin-like domains normally joins two β strands, in a C_α domain joins a β strand to this segment of α helix.

There are also differences in the way in which the domains interact. The interface between the V and C domains of both T-cell receptor chains is more extensive than in antibodies, which may make the hinge joint between the domains less flexible. And the interaction between the C_α and C_β domains is distinctive in being assisted by carbohydrate, with a sugar group from the C_α domain making a number of hydrogen bonds to the C_β domain (see Fig. 3.13b). Finally, a comparison of the variable binding sites shows that, although the complementarity-determining region (CDR) loops align fairly closely with those of antibody molecules, there is some displacement relative to those of the antibody molecule (Fig. 3.13c). This displacement is particularly marked in the V_α CDR2 loop, which is oriented at roughly right angles to the equivalent loop in antibody V domains, as a result of a shift in the β strand that anchors one end of the loop from one face of the domain to the other. A strand displacement also causes a change in

the orientation of the V_β CDR2 loop in two of the seven V_β domains whose structures are known. As yet, the crystallographic structures of only seven T-cell receptors have been solved to this level of resolution, so it remains to be seen to what degree all T-cell receptors share these features, and whether there is more variability to be discovered.

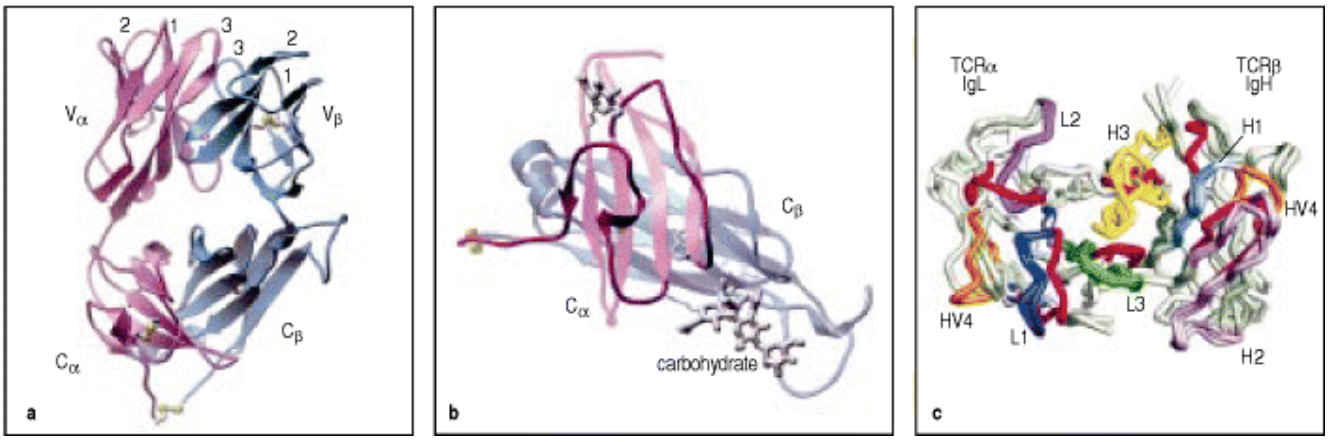


Figure 3.13. The crystal structure of an α:β T-cell receptor resolved at 2.5 Å. In panels a and b the α chain is shown in pink and the β chain in blue. Disulfide bonds are shown in green. In panel a, the T-cell receptor is viewed from the side as it would sit on a cell surface, with the CDR loops that form the antigen-binding site (labeled 1, 2, and 3) arrayed across its relatively flat top. In panel b, the C_α and C_β domains are shown. The C_α domain does not fold into a typical immunoglobulin-like domain; the face of the domain away from the C_β domain is mainly composed of irregular strands of polypeptide rather than β sheet. The intramolecular disulfide bond joins a β strand to this segment of α helix. The interaction between the C_α and C_β domains is assisted by carbohydrate (colored grey and labeled on the figure), with a sugar group from the C_α domain making hydrogen bonds to the C_β domain. In panel c, the T-cell receptor is shown aligned with the antigen-binding sites from three different antibodies. This view is looking down into the binding site. The V_α domain of the T-cell receptor is aligned with the V_L domains of the antigen-binding sites of the antibodies, and the V_β domain is aligned with the V_H domains. The CDRs of the T-cell receptor and immunoglobulin molecules are colored, with CDRs 1, 2, and 3 of the TCR shown in red and the HV4 loop in orange. For the immunoglobulin V domains, the CDR1 loops of the heavy chain (H1) and light chain (L1) are shown in light and dark blue, respectively, and the CDR2 loops (H2, L2) in light and dark purple, respectively. The heavy-chain CDR3 loops (H3) are in yellow; the light-chain CDR3s (L3) are in bright green. The HV4 loops of the TCR (orange) have no hypervariable counterparts in immunoglobulins. Photographs courtesy of I.A. Wilson, reprinted with permission from Science 274:209-219. ©1996 American Association for the Advancement of Science.

3-11. A T-cell receptor recognizes antigen in the form of a complex of a foreign peptide bound to an MHC molecule.

Antigen recognition by T-cell receptors clearly differs from recognition by B-cell receptors and antibodies. Antigen recognition by B cells involves direct binding of immunoglobulin to the intact antigen and, as discussed in Section 3-8, antibodies typically bind to the surface of protein antigens, contacting amino acids that are discontinuous in the primary structure but are brought together in the folded protein. T cells, on the other hand, were found to respond to short contiguous amino acid sequences in proteins. These sequences were often buried within the native structure of the protein and thus could not be recognized directly by T-cell receptors unless some unfolding of the protein antigen and its 'processing' into peptide fragments had occurred (Fig. 3.14).

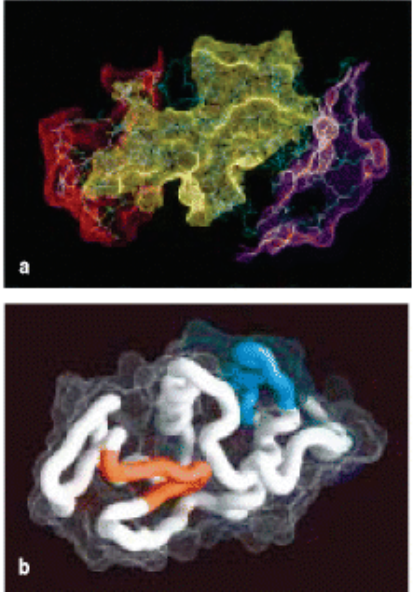


Figure 3.14. Differences in the recognition of hen egg-white lysozyme by immunoglobulins and T-cell receptors. Antibodies can be shown by X-ray crystallography to bind epitopes on the surface of proteins, as shown in panel a, where the epitopes for three antibodies are shown on the surface of hen egg lysozyme (see also Fig. 3.10). In contrast, the epitopes recognized by T-cell receptors need not lie on the surface of the molecule, as the T-cell receptor recognizes not the antigenic protein itself but a peptide fragment of the protein. The peptides corresponding to two T-cell epitopes of lysozyme are shown in panel b, one epitope, shown in blue, lies on the surface of the protein but a second, shown in red, lies mostly within the core and is inaccessible in the folded protein. For this residue to be accessible to the T-cell receptor, the protein must be unfolded and processed. Panel a courtesy of S. Sheriff.

The nature of the antigen recognized by T cells became clear with the realization that the peptides that stimulate T cells are recognized only when bound to an MHC molecule. These cell-surface glycoproteins are encoded by genes within the major histocompatibility complex (MHC). The ligand recognized by the T cell is thus a complex of peptide and MHC molecule. The evidence for involvement of the MHC in T-cell recognition of antigen was at first indirect, but it has recently been proved conclusively by stimulating T cells with purified peptide:MHC complexes. The T-cell receptor interacts with this ligand by making contacts with both the MHC molecule and the antigen peptide. TOP

3-12. T cells with different functions are distinguished by CD4 and CD8 cell-surface proteins and recognize peptides bound to different classes of MHC molecule.

T cells fall into two major classes that have different effector functions. The two classes are distinguished by the expression of the cell-surface proteins CD4 and CD8. These two types of T cell differ in the class of MHC molecule they recognize. There are two classes of MHC molecule—**MHC class I** and **MHC class II**—which differ in their structure and expression pattern on tissues of the body (see Section 3-13). CD4 and CD8 were known as markers for different functional sets of T cells for some time before it became clear that they play an important part in the direct recognition of MHC class II and MHC class I molecules, respectively. CD4 binds to the MHC class II molecule and CD8 to the MHC class I molecule. During antigen recognition, depending on the type of T-cell, CD4 or CD8 molecules associate on the T-cell surface with the T-cell receptor and bind to invariant sites on the MHC portion of the composite MHC:peptide ligand. This binding is required for the T cell to make an effective response, and so CD4 and CD8 are called co-receptors.

CD4 is a single-chain molecule composed of four immunoglobulin-like domains (Fig. 3.15). The first two domains (D_1 and D_2) of the CD4 molecule are packed tightly together to form a rigid rod some 60 Å long, which is joined by a flexible hinge to a similar rod formed by the third and fourth domains (D_3 and D_4). CD4 binds MHC class II molecules through a region that is located mainly on a lateral face of the first domain, D_1. Because CD4 binds to a site on the β_2 domain of the MHC class II molecule that is well away from the site where the T-cell receptor binds (Fig. 3.16a), the CD4 molecule and the T-cell receptor can bind the same peptide:MHC class II complex. CD4 interacts strongly with a cytoplasmic tyrosine kinase called Lck, and can deliver this tyrosine kinase into close proximity with the signaling components of the T-cell receptor complex. This results in enhancement of the signal that is generated when the T-cell receptor binds its peptide:MHC class II ligand, as we will discuss further in Chapter 6. When CD4 and the T-cell receptor can simultaneously bind to the same MHC class II:peptide complex, the

sensitivity of a T cell to antigen presented by MHC class II molecules is markedly increased; the T-cell in this case requires 100-fold less antigen for activation.

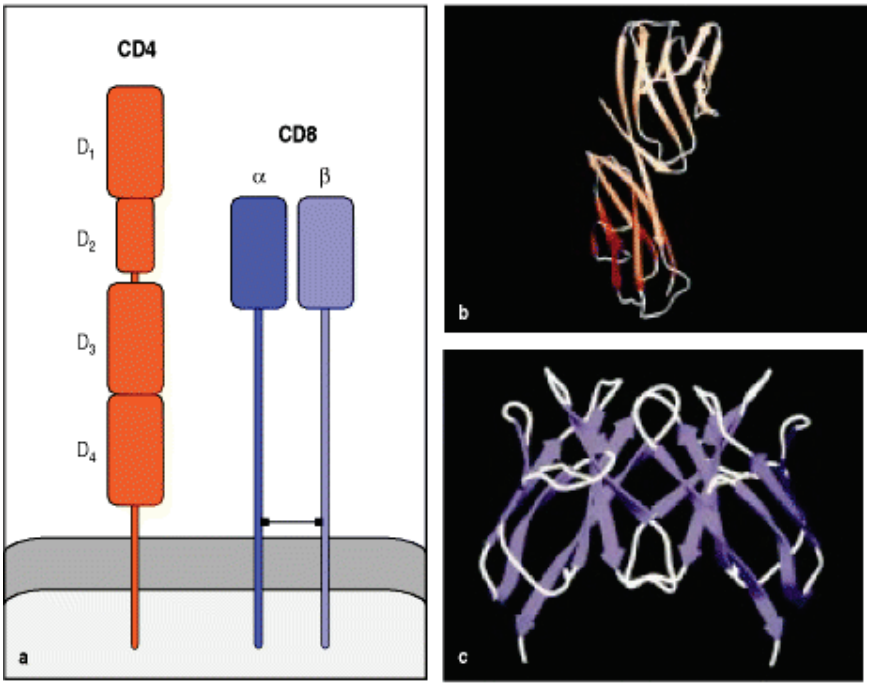


Figure 3.15. The outline structures of the CD4 and CD8 co-receptor molecules . The CD4 molecule contains four immunoglobulin-like domains, as shown in diagrammatic form in panel a, and as a ribbon diagram of the structure in panel b. The amino-terminal domain, D_1, is similar in structure to an immuno-globulin V domain. The second domain, D_2, although clearly related to an immunoglobulin domain, is different from both V and C domains and has been termed a C2 domain. The first two domains of CD4 form a rigid rodlike structure that is linked to the two carboxy-terminal domains by a flexible link. The binding site for MHC class II molecules is thought to involve both the D_1 and D_2 domains. The CD8 molecule (panels a and c) is a heterodimer of an α and a β chain covalently linked by a disulfide bond; an alternative form of CD8 exists as a homodimer of α chains. The heterodimer is depicted in panel a, while the ribbon diagram in panel c is of the homodimer. CD8α and CD8β chains have very similar structures, each having a single domain resembling an immunoglobulin V domain and a stretch of polypeptide chain, believed to be in a relatively extended conformation, that anchors the V-like domain to the cell membrane.

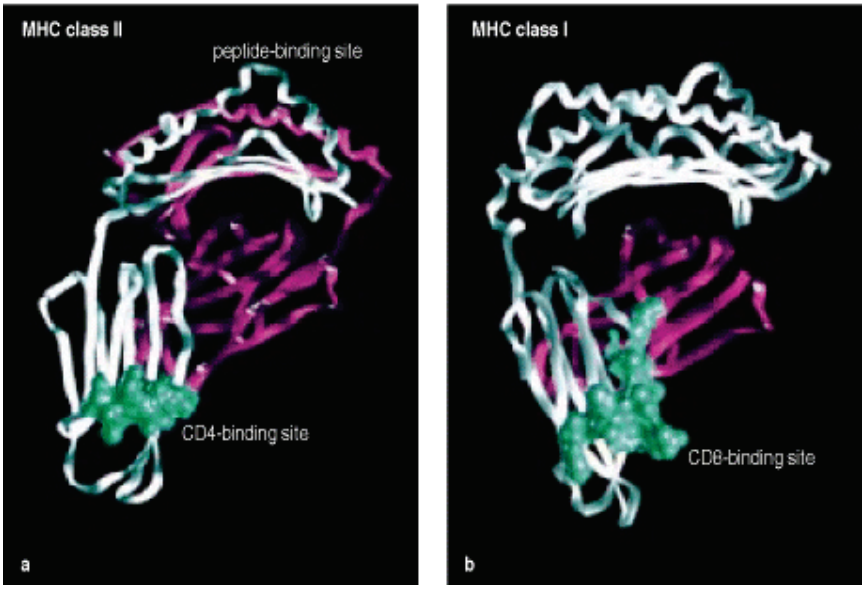


Figure 3.16. The binding sites for CD4 and CD8 on MHC class II and class I molecules lie in the immunoglobulin-like domains. The binding sites for CD4 and CD8 on the MHC class II and class I molecules, respectively, lie in the immunoglobulin-like domains nearest to the membrane and distant from the peptide-binding cleft. In panel a, the binding site for CD4 is shown as a bright green surface. It lies at the base of the β_2 domain of an

MHC class II molecule and is distant from the peptide-binding site at the top of the molecule. The α chain of the MHC class II molecule is purple, and the β chain is white. In panel b, the binding site for CD8 is also shown as a green surface, at the base of the α_3 domain of the MHC class I molecule. The α chain of the class I molecule is white, and the β_2-microglobulin is purple. Photographs courtesy of C. Thorpe.

CD4 binding to an MHC class II molecule on its own is weak, and it is not clear whether such binding would be able to transmit a signal to the interior of the T cell. As shown in Fig. 3.17, CD4 can form homodimers through a site in the D_4 domain, which leaves the MHC-binding site free to interact with an MHC class II molecule. Thus, the CD4 dimer could cross-link two MHC class II molecules and thus the two T-cell receptors bound to them. Whether the dimerization of CD4 is important in its co-receptor function is not known at present.



Figure 3.17. CD4 is capable of forming dimers. The structure of the extra-cellular domains of the CD4 molecule has been determined by X-ray crystallography. Two molecules of CD4 can interact with each other through their D_4 domains, forming homodimers. The site that binds MHC class II molecules remains available in such dimers.

Although CD4 and CD8 both function as co-receptors, their structures are quite distinct. The CD8 molecule is a disulfide-linked heterodimer consisting of an α and a β chain, each containing a single immunoglobulin-like domain linked to the membrane by a segment of extended polypeptide chain (see Fig. 3.15). This segment is extensively glycosylated, which is thought to be important in maintaining this polypeptide in an extended conformation and protecting it from cleavage by proteases. CD8α chains can also form homo-dimers, although these are not found when the CD8β chains are present.

CD8 binds weakly to an invariant site in the α_3 domain of an MHC class I molecule (Fig. 3.16b), which is equivalent to the site in MHC class II molecules to which CD4 binds. Although only the interaction of the CD8α homodimer with MHC class I is so far known in detail, it is clear from this that the MHC class I binding site of the CD8 α:β heterodimer will be formed by the interaction of the CD8α and β chains. In addition, CD8 (most probably through the α chain) interacts with residues in the base of the α_2 domain of the MHC class I molecule. Binding in this way, CD8 leaves the upper surface of the MHC class I molecule exposed and free to interact simultaneously with a T-cell receptor, as shown in Fig. 3.18. Like CD4, CD8 also binds Lck through the cytoplasmic tail of the α chain and brings it into close proximity with the T-cell receptor. And as with CD4, the presence of CD8 increases the sensitivity of T cells to antigen presented by MHC class I molecules by about 100-fold. Thus, CD4 and CD8 have similar functions and bind to the same approximate location in MHC class I and MHC class II molecules even though the structures of the two co-receptor proteins are only distantly related.

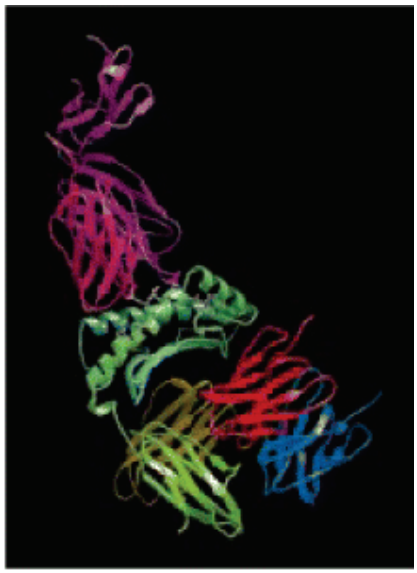


Figure 3.18. CD8 binds to a site on MHC class I molecules distant from that to which the T-cell receptor binds. The relative positions of the T-cell receptor and CD8 molecules bound to the same MHC class I molecule can be seen in this hypothetical reconstruction of the interaction of an MHC class I molecule (the α chain is shown in green; β_2-microglobulin (dull yellow) can be seen faintly in the background) with a T-cell receptor and CD8. The α and β chains of the T-cell receptor are shown in pink and purple, respectively. The CD8 structure is that of a CD8α homodimer, but is colored to represent the likely orientation of the subunits in the heterodimer, with the CD8α subunit in red and the CD8β subunit in blue. Photograph courtesy of G. Gao, reprinted with permission from Nature 387:630-634, ©1997 Macmillan Magazines Limited.

3-13. The two classes of MHC molecule are expressed differentially on cells.

MHC class I and MHC class II molecules have a distinct distribution among cells that reflects the different effector functions of the T cells that recognize them (Fig. 3.19). MHC class I molecules present peptides from pathogens, commonly viruses, to CD8 cytotoxic T cells, which are specialized to kill any cell that they specifically recognize. As viruses can infect any nucleated cell, almost all such cells express MHC class I molecules, although the level of constitutive expression varies from one cell type to the next. For example, cells of the immune system express abundant MHC class I on their surface, whereas liver cells (hepatocytes) express relatively low levels (see Fig. 3.19). Nonnucleated cells, such as mammalian red blood cells, express little or no MHC class I, and thus the interior of red blood cells is a site in which an infection can go undetected by cytotoxic T cells. As red blood cells cannot support viral replication, this is of no great consequence for viral infection, but it may be the absence of MHC class I that allows the *Plasmodium* species that cause malaria to live in this privileged site.

Tissue	MHC class I	MHC class II
Lymphoid tissues		
T cells	+++	+*
B cells	+++	+++
Macrophages	+++	++
Other antigen-presenting cells (eg Langerhans' cells)	+++	+++
Epithelial cells of the thymus	+	+++
Other nucleated cells		
Neutrophils	+++	–
Hepatocytes	+	–
Kidney	+	–
Brain	+	–†
Non-nucleated cells		
Red blood cells	–	–

Figure 3.19. The expression of MHC molecules differs between tissues. MHC class I molecules are expressed on all nucleated cells, although they are most highly expressed in hematopoietic cells. MHC class II molecules are normally expressed only by a subset of hematopoietic cells and by thymic stromal cells, although they may be expressed by other cell types on exposure to the inflammatory cytokine interferon-γ. *In humans, activated T cells express MHC class II molecules, whereas in mice, all T cells are MHC class II-negative. †In the brain, most cell types are MHC class II-negative but microglia, which are related to macrophages, are MHC class II-positive.

In contrast, the main function of the CD4 T cells that recognize MHC class II molecules is to activate other effector cells of the immune system. Thus MHC class II molecules are normally found on B lymphocytes, dendritic cells, and macrophages—cells that participate in immune responses—but not on other tissue cells (see Fig. 3.19). When CD4 T cells recognize peptides bound to MHC class II molecules on B cells, they stimulate the B cells to produce antibody Likewise, CD4 T cells recognizing peptides bound to MHC class II molecules on macrophages activate these cells to destroy the pathogens in their vesicles We shall see in Chapter 8 that MHC class II molecules are also expressed on specialized antigen-presenting cells in lymphoid tissues where naive T cells encounter antigen and are first activated. The expression of both MHC class I and MHC class II molecules is regulated by cytokines, in particular interferons, released in the course of immune responses. Interferon-γ (IFN-γ), for example, increases the expression of MHC class I and MHC class II molecules, and can induce the expression of MHC class II molecules on certain cell types that do not normally express them. Interferons also enhance the antigenpresenting function of MHC class I molecules by inducing the expression of key components of the intracellular machinery that enables peptides to be loaded onto the MHC molecules.

3-14. The two classes of MHC molecule have distinct subunit structures but similar three-dimensional structures.

The two classes of MHC molecule differ from each other in their structure and also have different distributions on the cells of the body. Their different structures enable the two classes of MHC molecules to serve distinct functions in antigen presentation, binding peptides from different intracellular sites and activating different subsets of T cells, as we will see in Chapter 5. Despite their differences in subunit structure, however, MHC class I and class II molecules are closely related in overall structure. In both classes, the two paired protein domains nearest the membrane resemble immunoglobulin domains, whereas the two domains distal to the membrane fold together to create a long

cleft, or groove, which is the site at which a peptide binds. Purified peptide:MHC class I and peptide:MHC class II complexes have been characterized structurally, allowing us to describe in detail both the MHC molecules themselves and the way in which they bind peptides.

MHC class I structure is outlined in Fig. 3.20. MHC class I molecules consist of two polypeptide chains, a larger α chain encoded in the MHC genetic locus, and a smaller noncovalently associated chain, **β_2-microglobulin**, which is not polymorphic and is not encoded in the MHC locus. Only the class I α chain spans the membrane. The complete molecule has four domains, three formed from the MHC-encoded α chain, and one contributed by β_2-microglobulin. The α_3 domain and β_2-microglobulin have a folded structure that closely resembles that of an immunoglobulin domain. The most remarkable feature of MHC class I molecules is the structure of the folded α_1 and α_2 domains. These two domains form the walls of a cleft on the surface of the molecule; this is the site of peptide binding. They also are sites of polymorphisms that determine T-cell antigen recognition (see Chapter 5).

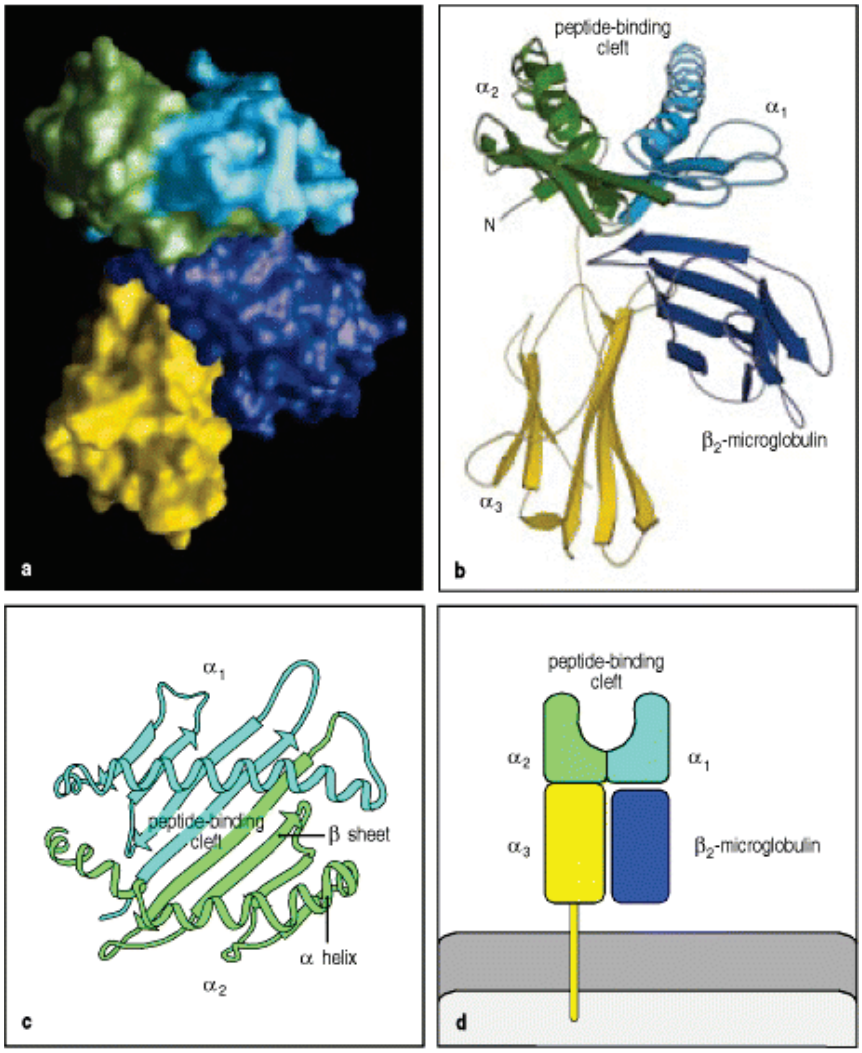


Figure 3.20. The structure of an MHC class I molecule determined by X-ray crystallography. Panel a shows a computer graphic representation of a human MHC class I molecule, HLA-A2, which has been cleaved from the cell surface by the enzyme papain. The surface of the molecule is shown, colored according to the domains shown in panels b-d and described below. Panels b and c show a ribbon diagram of that structure. Shown schematically in panel d, the MHC class I molecule is a heterodimer of a membrane-spanning α chain (molecular weight 43 kDa) bound noncovalently to β_2-microglobulin (12 kDa), which does not span the membrane. The α chain folds into three domains: α_1, α_2, and α_3. The α_3 domain and β_2-microglobulin show similarities in amino acid sequence to immunoglobulin C domains and have similar folded structures, whereas the α_1 and α_2 domains fold together into a single structure consisting of two segmented α helices lying on a sheet of eight antiparallel β strands. The folding of the α_1 and α_2 domains creates a long cleft or groove, which is the site at which peptide antigens bind to the MHC molecules. The transmembrane region and the short stretch of peptide that connects the external domains to the cell surface are not seen in panels a and b as they have been removed by the papain digestion. As can be seen in panel c, looking down on the molecule from above, the sides of the cleft are formed from the inner faces of the two α helices;

the β-pleated sheet formed by the pairing of the α_1 and α_2 domains creates the floor of the cleft. We shall use the schematic representation in panel d throughout this text.

An MHC class II molecule consists of a noncovalent complex of two chains, α and β, both of which span the membrane (Fig. 3.21). The MHC class II α and β chains are both encoded within the MHC. The crystallographic structure of the MHC class II molecule shows that it is folded very much like the MHC class I molecule. The major differences lie at the ends of the peptide-binding cleft, which are more open in MHC class II molecules compared with MHC class I molecules. The main consequence of this is that the ends of a peptide bound to an MHC class I molecule are substantially buried within the molecule, whereas the ends of peptides bound to MHC class II molecules are not. Again, the sites of major polymorphism are located in the peptide-binding cleft, which in the case of an MHC class II molecule are formed by the α_1, and β_1 domains (see Chapter 5).

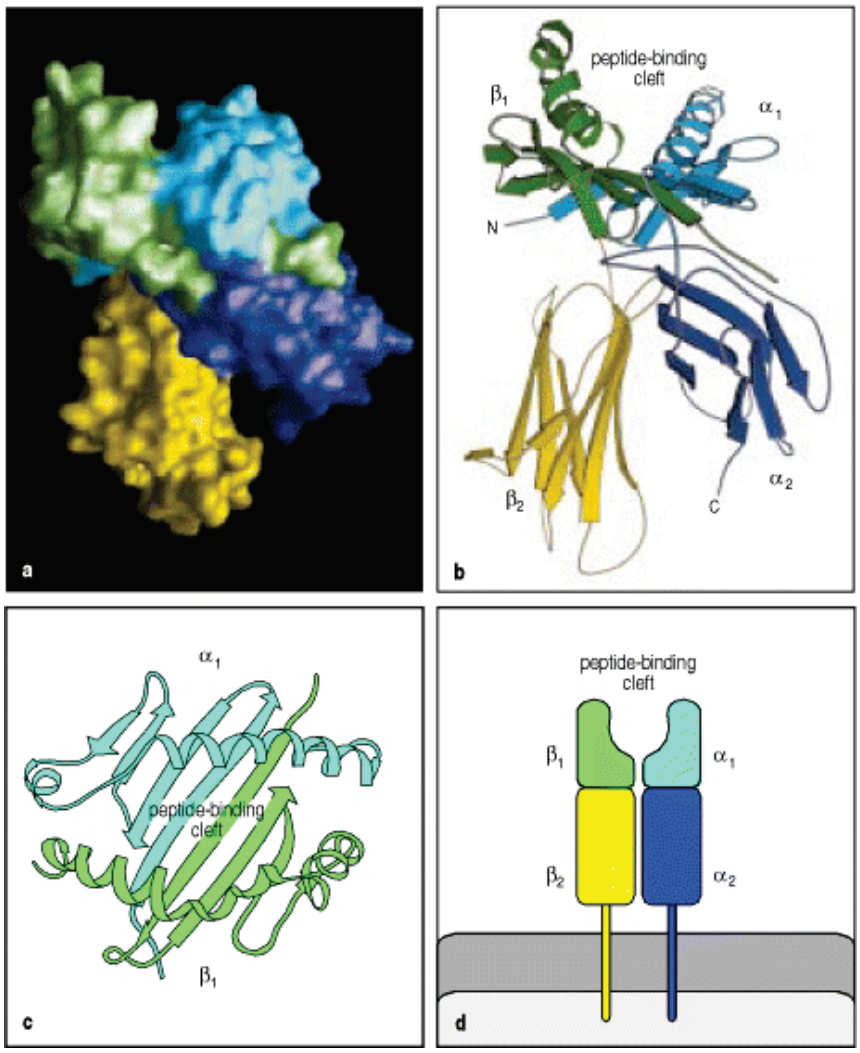


Figure 3.21. MHC class II molecules resemble MHC class I molecules in overall structure. The MHC class II molecule is composed of two trans-membrane glycoprotein chains, α (34 kDa) and β (29 kDa), as shown schematically in panel d. Each chain has two domains, and the two chains together form a compact four-domain structure similar to that of the MHC class I molecule (compare with panel d of Fig. 3.20). Panel a shows a computer graphic representation of the surface of the MHC class II molecule, in this case the human protein HLA-DR1, and panel b shows the equivalent ribbon diagram. The α_2 and β_2 domains, like the α_3 and β_2-microglobulin domains of the MHC class I molecule, have amino acid sequence and structural similarities to immunoglobulin C domains; in the MHC class II molecule, the two domains forming the peptide-binding cleft are contributed by different chains and are therefore not joined by a covalent bond (see panels c and d). Another important difference, not apparent in this diagram, is that the peptide-binding groove of the MHC class II molecule is open at both ends.

In both MHC class I and class II molecules, bound peptides are sandwiched between the two α-helical segments of the MHC molecule (Fig. 3.22). The T-cell receptor interacts with this compound ligand, making contacts with both the MHC molecule and with the peptide fragment of antigen.

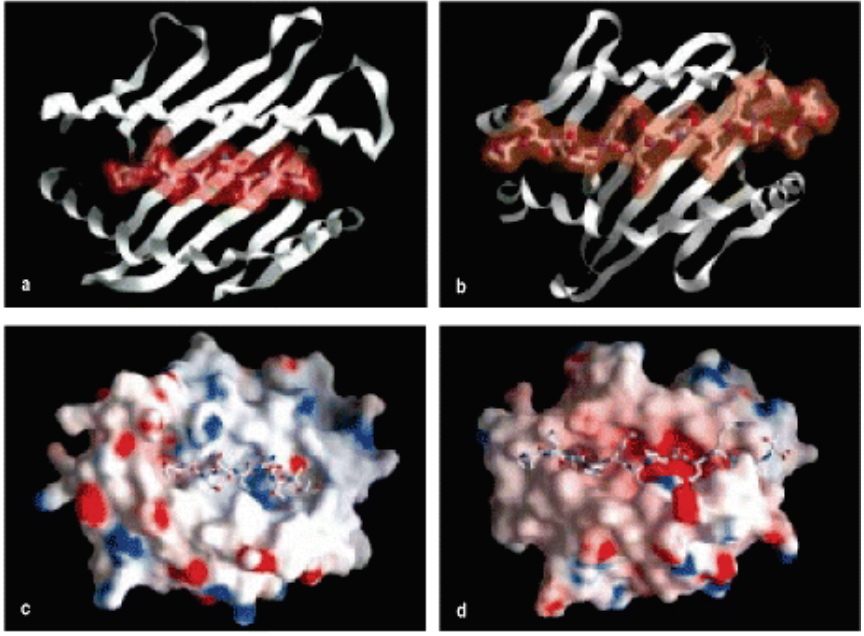


Figure 3.22. MHC molecules bind peptides tightly within the cleft. When MHC molecules are crystallized with a single synthetic peptide antigen, the details of peptide binding are revealed. In MHC class I molecules (panels a and c) the peptide is bound in an elongated conformation with both ends tightly bound at either end of the cleft. In the case of MHC class II molecules (panels b and d), the peptide is also bound in an elongated conformation but the ends of the peptide are not tightly bound and the peptide extends beyond the cleft. The upper surface of the peptide:MHC complex is recognized by T cells, and is composed of residues of the MHC molecule and the peptide. In representations c and d, the electrostatic potential of the MHC molecule surface is shown, with blue areas indicating a positive potential and red a negative potential.

3-15. Peptides are stably bound to MHC molecules, and also serve to stabilize the MHC molecule on the cell surface.

An individual can be infected by a wide variety of different pathogens the proteins of which will not generally have peptide sequences in common. If T cells are to be alerted to all possible infections, then the MHC molecules on each cell (both class I and class II) must be able to bind stably to many different peptides. This behavior is quite distinct from that of other peptide-binding receptors, such as those for peptide hormones, which usually bind only a single type of peptide. The crystal structures of peptide:MHC complexes have helped to show how a single binding site can bind peptides with high affinity while retaining the ability to bind a wide variety of different peptides.

An important feature of the binding of peptides to MHC molecules is that the peptide is bound as an integral part of the MHC molecule's structure, and MHC molecules are unstable when peptides are not bound. The stability of peptide binding is important, because otherwise, peptide exchanges occurring at the cell surface would prevent peptide:MHC complexes from being reliable indicators of infection or of uptake of specific antigen. As a result of this stability, when MHC molecules are purified from cells, their bound peptides co-purify with them, and this has enabled the peptides bound by specific MHC molecules to be analyzed. The peptides are released from the MHC molecules by denaturing the complex in acid, and can then be purified and sequenced. Pure synthetic peptides can also be incorporated into previously empty MHC molecules and the structure of the complex determined, revealing details of the contacts between the MHC molecule and the peptide. From the sequences of peptides bound to specific MHC molecules, combined with structural analysis of the peptide:MHC complex, a detailed picture of the binding interactions has been built up. We will first discuss the peptide-binding properties of MHC class I molecules.

3-16. MHC class I molecules bind short peptides of 8–10 amino acids by both ends.

The binding of a peptide in the peptide-binding cleft of an MHC class I molecule is stabilized at both ends by contacts between atoms in the free amino and carboxy termini of the peptide and invariant sites that are found at each end of the cleft of all MHC class I molecules (Fig. 3.23). These contacts are thought to be the main stabilizing contacts for peptide:MHC class I complexes because synthetic peptide analogues lacking terminal amino and carboxyl groups fail to bind stably to MHC class I molecules. Other residues in the peptide serve as additional anchors. Peptides that bind to MHC class I molecules are usually 8–10 amino acids long. The peptide lies in an

elongated conformation along the groove; variations in peptide length appear to be accommodated, in most cases, by a kinking in the peptide backbone. However, two examples of MHC class I molecules where the peptide is able to extend out of the groove at the carboxy terminus suggest that some length variation may also be accommodated in this way.

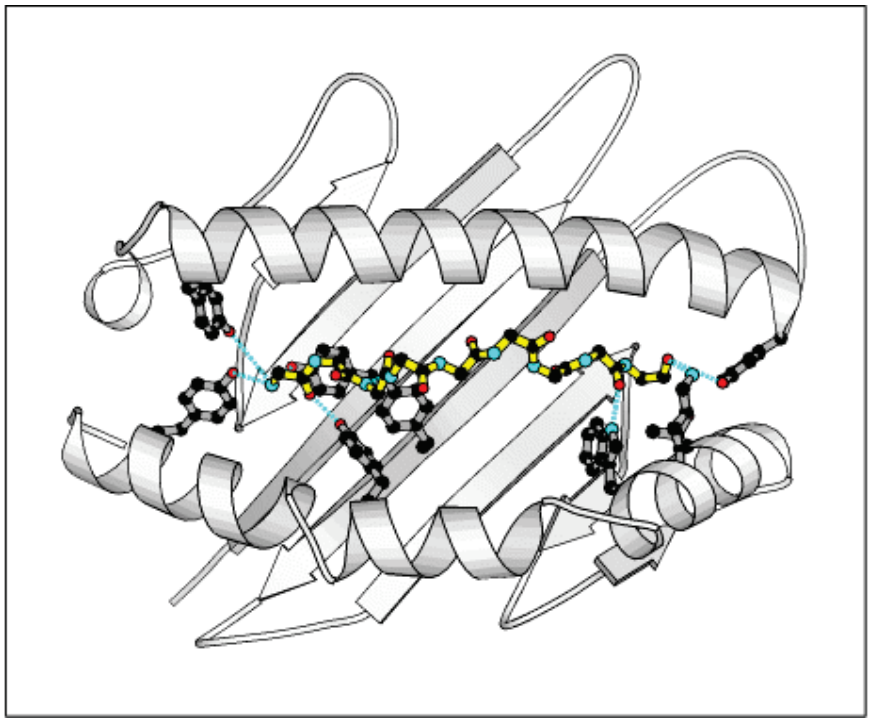


Figure 3.23. Peptides are bound to MHC class I molecules by their ends. MHC class I molecules interact with the back-bone of a bound peptide (shown in yellow) through a series of hydrogen bonds and ionic interactions (shown as dotted blue lines) at each end of the peptide. The amino terminus of the peptide is to the left; the carboxy terminus to the right. Black circles are carbon atoms; red are oxygen; blue are nitrogen. The amino acid residues in the MHC molecule that form these bonds are common to all MHC class I molecules and their side chains are shown in full (in gray) upon a ribbon diagram of the MHC class I groove. A cluster of tyrosine residues common to all MHC class I molecules forms hydrogen bonds to the amino terminus of the bound peptide, while a second cluster of residues forms hydrogen bonds and ionic interactions with the peptide backbone at the carboxy terminus and with the carboxy terminus itself.

These interactions give all MHC class I molecules their broad peptide-binding specificity. In addition, MHC molecules are highly polymorphic. There are hundreds of different versions, or alleles, of the MHC class I genes in the human population as a whole, and each individual carries only a small selection of them. The main differences between the allelic MHC variants are found at certain sites in the peptide-binding cleft, resulting in different amino acids in key peptide interaction sites in the different MHC variants. The consequence of this is that the different MHC variants preferentially bind different peptides. The peptides that can bind to a given MHC variant have the same or very similar amino acid residues at two or three particular positions along the peptide sequence. The amino acid side chains at these positions insert into pockets in the MHC molecule that are lined by the polymorphic amino acids. Because the binding of these side chains anchors the peptide to the MHC molecule, the peptide residues involved have been called anchor residues. Both the position and identity of these anchor residues can vary, depending on the particular MHC class I variant that is binding the peptide. However, most peptides that bind to MHC class I molecules have a hydrophobic (or sometimes basic) anchor residue at the carboxy terminus (Fig. 3.24). Changing an anchor residue can prevent the peptide from binding and, conversely, most synthetic peptides of suitable length that contain these anchor residues will bind the appropriate MHC class I molecule, in most cases irrespective of the amino acids at other positions in the peptide. These features of peptide binding enable an individual MHC class I molecule to bind a wide variety of different peptides, yet allow different MHC class I allelic variants to bind different sets of peptides.

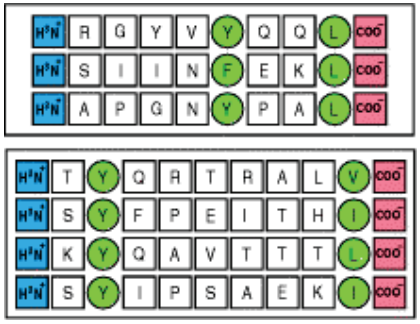


Figure 3.24. Peptides bind to MHC molecules through structurally related anchor residues. Peptides eluted from two different MHC class I molecules are shown. The anchor residues (green) differ for peptides that bind different alleles of MHC class I molecules but are similar for all peptides that bind to the same MHC molecule. The upper and lower panels show peptides that bind to two different alleles of MHC class I molecules. The anchor residues that bind a particular MHC molecule need not be identical, but are always related (for example, phenylalanine (F) and tyrosine (Y) are both aromatic amino acids, whereas valine (V), leucine (L), and isoleucine (I) are all large hydro-phobic amino acids). Peptides also bind to MHC class I molecules through their amino (blue) and carboxy (red) termini.

3-17. The length of the peptides bound by MHC class II molecules is not constrained.

Peptide binding to MHC class II molecules has also been analyzed by elution of bound peptides and by X-ray crystallography, and differs in several ways from peptide binding to MHC class I molecules. Peptides that bind to MHC class II molecules are at least 13 amino acids long and can be much longer. The clusters of conserved residues that bind the two ends of a peptide in MHC class I molecules are not found in MHC class II molecules, and the ends of the peptide are not bound. Instead, the peptide lies in an extended conformation along the MHC class II peptide-binding groove. It is held in this groove both by peptide side chains that protrude into shallow and deep pockets lined by polymorphic residues, and by interactions between the peptide backbone and side chains of conserved amino acids that line the peptide-binding cleft in all MHC class II molecules (Fig. 3.25). Although there are fewer crystal structures of MHC class II-bound peptides than of MHC class I, the available data show that amino acid side chains at residues 1, 4, 6, and 9 of an MHC class II-bound peptide can be held in these binding pockets.

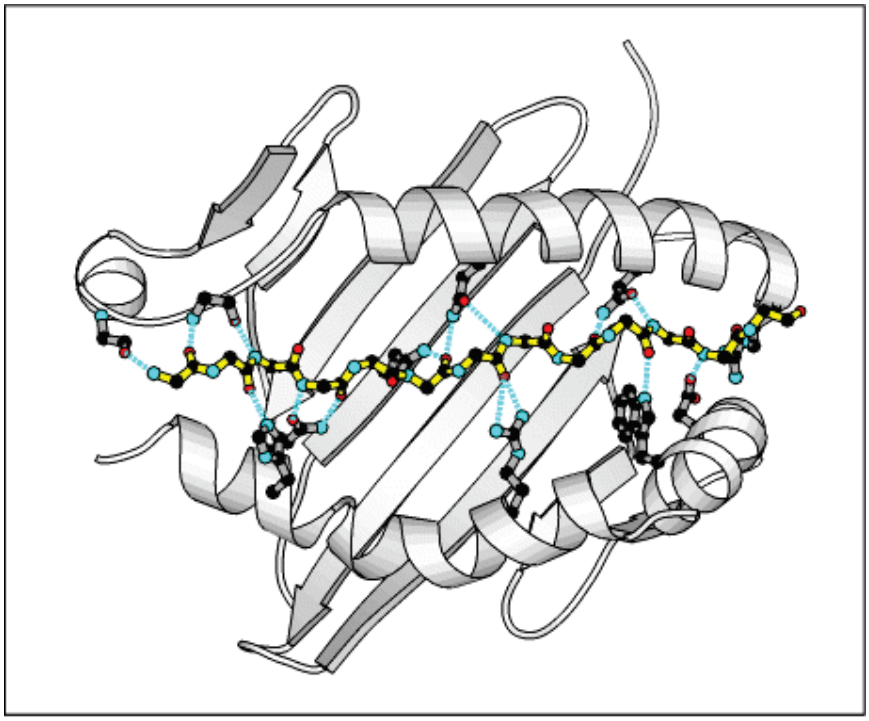


Figure 3.25. Peptides bind to MHC class II molecules by interactions along the length of the binding groove. A peptide (yellow; shown as the peptide backbone only, with the amino terminus to the left and the carboxy terminus to the right), is bound by an MHC class II molecule through a series of hydrogen bonds (dotted blue lines) that are distributed along the length of the peptide. The hydrogen bonds toward the amino terminus of the peptide are made

with the backbone of the MHC class II polypeptide chain, whereas throughout the peptide's length bonds are made with residues that are highly conserved in MHC class II molecules. The side chains of these residues are shown in gray upon the ribbon diagram of the MHC class II groove.

The binding pockets of MHC class II molecules are more permissive in their accommodation of different amino acid side chains than are those of the MHC class I molecule, making it more difficult to define anchor residues and predict which peptides will be able to bind particular MHC class II molecules (Fig. 3.26). Nevertheless, by comparing the sequences of known binding peptides, it is usually possible to detect a pattern of permissive amino acids for each of the different alleles of MHC class II molecules, and to model how the amino acids of this peptide sequence motif will interact with the amino acids that make up the peptide-binding cleft in the MHC class II molecule. Because the peptide is bound by its backbone and allowed to emerge from both ends of the binding groove there is, in principle, no upper limit to the length of peptides that could bind to MHC class II molecules. However, it appears that longer peptides bound to MHC class II molecules are trimmed by peptidases to a length of 13–17 amino acids in most cases. Like MHC class I molecules, MHC class II molecules that lack bound peptide are unstable, but the critical stabilizing interactions that the peptide makes with the MHC class II molecule are not yet known.

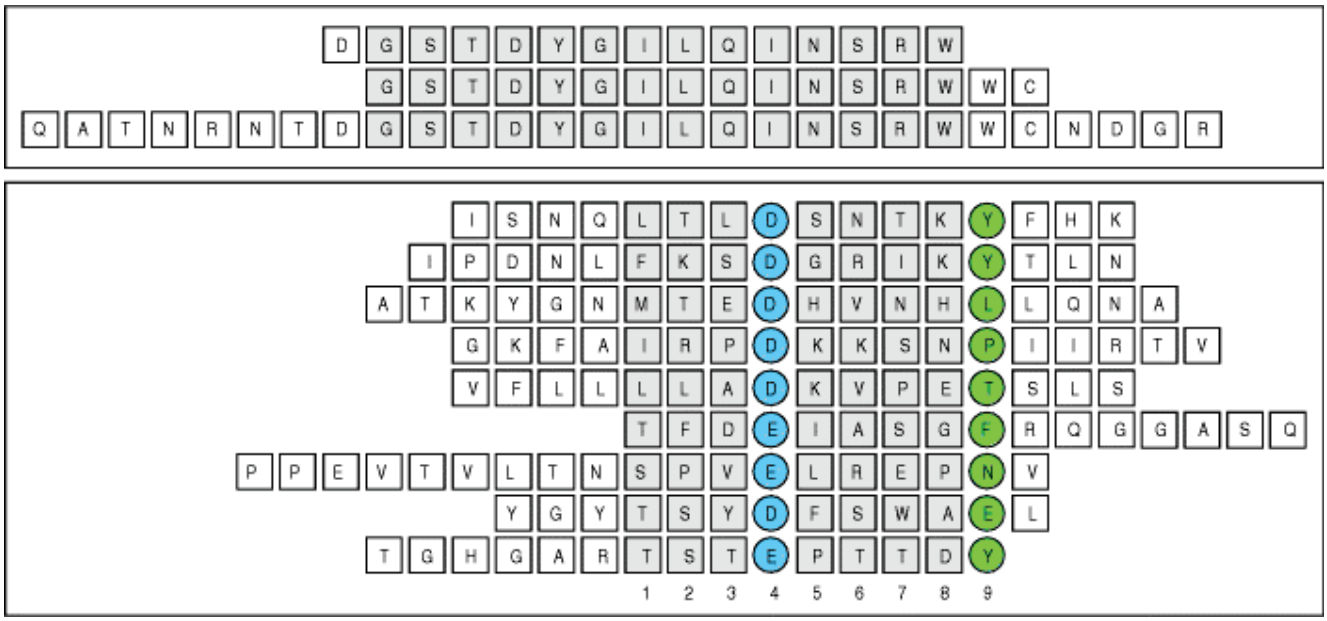
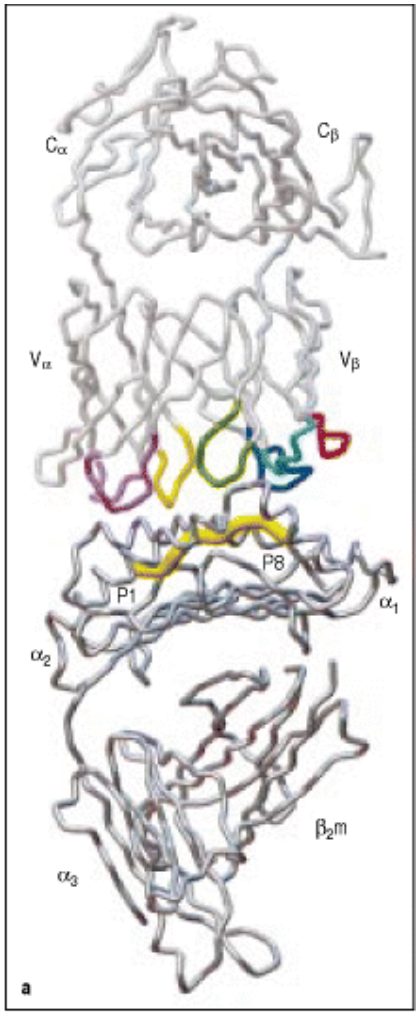


Figure 3.26. Peptides that bind MHC class II molecules are variable in length and their anchor residues lie at various distances from the ends of the peptide. The sequences of a set of peptides that bind to the mouse MHC class II A^k allele are shown in the upper panel. All contain the same core sequence but differ in length. In the lower panel, different peptides binding to the human MHC class II allele HLA-DR3 are shown. The lengths of these peptides can vary, and so by convention the first anchor residue is denoted as residue 1. Note that all of the peptides share a negatively charged residue (aspartic acid (D) or glutamic acid (E)) in the P4 position (blue) and tend to have a hydrophobic residue (for example, tyrosine (Y), leucine (L), proline (P), phenylalanine (F)) in the P9 position (green).

3-18. The crystal structures of several MHC:peptide:T-cell receptor complexes all show the same T-cell receptor orientation over the MHC:peptide complex.

At the time that the first X-ray crystallographic structure of a T-cell receptor was published, a structure of the same T-cell receptor bound to a peptide:MHC class I ligand was also produced. This structure (Fig. 3.27), which had been forecast by site-directed mutagenesis of the MHC class I molecule, showed the T-cell receptor aligned diagonally over the peptide and the peptide-binding groove, with the T-cell receptor α chain lying over the α_2 domain and the amino-terminal end of the bound peptide, the T-cell receptor β chain lying over the α_1 domain and the carboxy-terminal end of the peptide, with the CDR3 loops of both T-cell receptor α and T-cell receptor β meeting over the central amino acids of the peptide. The T-cell receptor is threaded through a valley between the two high points on the two surrounding α helices that form the walls of the peptide-binding cleft.



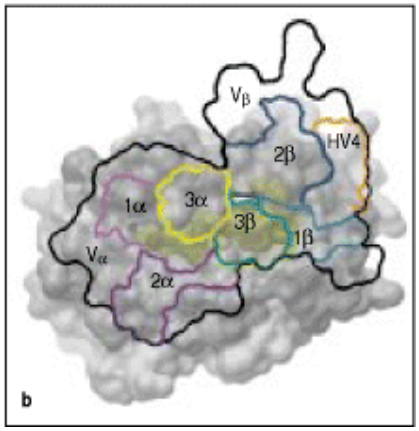


Figure 3.27. The T-cell receptor binds to the MHC:peptide complex. Panel a: the T-cell receptor binds to the top of the MHC:peptide complex, straddling, in the case of the class I molecule shown here, both the α_1 and α_2 domain helices. The CDRs of the T-cell receptor are indicated in color; the CDR1 and CDR2 loops of the β chain in light and dark blue, respectively; and the CDR1 and CDR2 loops of the α chain in light and dark purple, respectively. The α chain CDR3 loop is in yellow while the β chain CDR3 loop is in green. The β chain HV4 loop is orange. Panel b: the outline of the T-cell receptor antigen-binding site (thick black line) is superimposed upon the top surface of the MHC:peptide complex (the peptide is shaded dull yellow). The T-cell receptor lies diagonally across the MHC:peptide complex, with the α and β CDR3 loops of the T-cell receptor (3α, 3β, yellow and green, respectively) contacting the center of the peptide. The α chain CDR1 and CDR2 loops (1α, 2α, light and dark purple, respectively) contact the MHC helices at the amino terminus of the bound peptide, whereas the β chain CDR1 and CDR2 loops (1β, 2β, light and dark blue, respectively) make contact with the helices at the carboxy terminus of the bound peptide. Courtesy of I.A. Wilson, reprinted with permission from Science 274:209-219. ©1996 American Association for the Advancement of Science.

Analysis of other MHC class I:peptide:T-cell receptor complexes and of the single example so far of an MHC class II:peptide:T-cell receptor complex (Fig. 3.28) shows that all of them have a very similar orientation, particularly for the V_α domain, although some variability does occur in the location and orientation of the V_β domain. In this orientation, the V_α domain makes contact primarily with the amino terminus of the bound peptide, whereas the V_β domain contacts primarily the carboxy terminus of the bound peptide. Both chains also interact with the α helices of the MHC class I molecule (see Fig. 3.27). The T-cell receptor contacts are not symmetrically distributed over the MHC molecule, so whereas the V_α CDR1 and CDR2 loops are in close contact with the helices of the MHC:peptide complex around the amino terminus of the bound peptide, the β-chain CDR1 and CDR2 loops, which interact with the complex at the carboxy terminus of the bound peptide, have variable contributions to the binding. This suggests that the V_α contacts are responsible for the conserved orientation of the T-cell receptor on the MHC:peptide complex.

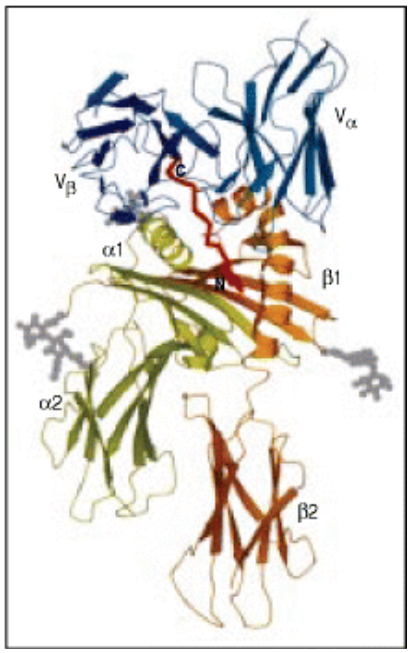


Figure 3.28. The T-cell receptor interacts with MHC class I and MHC class II molecules in a similar fashion. The structure of a T-cell receptor binding to an MHC class II molecule has been determined, and shows the T-cell receptor binding to an equivalent site, and in an equivalent orientation, to the way that TCRs bind to MHC class I molecules (see Fig. 3.27). The structure of the molecules is shown in a cartoon form, with the MHC class II α and β chains shown in light green and orange respectively. Only the V_α and V_β domains of the T-cell receptor are shown, colored in blue. The peptide is colored red, while carbohydrate residues are indicated in gray. The TCR sits in a shallow saddle formed between the MHC class II α and β chain α-helical regions, at roughly 90° to the long axis of the MHC class II molecule and the bound peptide. Courtesy of E.L. Reinherz, reprinted with permission from Science 286:1913-1921. ©1999 American Association for the Advancement of Science.

Comparison of the three-dimensional structure of the T-cell receptor to that of the same T-cell receptor complexed to its MHC-peptide ligand could address the question of whether the T-cell receptor, like some other receptors, undergoes a conformational change, or 'induced fit,' in its three-dimensional structure when it binds its specific ligand. To date, there is no certain answer, owing to the limitations of the available structures. For one thing, all the T-cell receptors analyzed to date have either been bound to ligands that do not produce activation, or are bound to ligands that can activate them, but the comparable unliganded receptor structures are not available. Also, the crystals of the T-cell receptor are all formed at 0 °C or below, which locks the receptor into a single conformation. However, subtly different peptides can have strikingly different effects when the same T cell recognizes either of the two peptides complexed with MHC. This could be due to differences in how T-cell receptor conformation is altered by binding the two related yet different ligands. Recent evidence also suggests that the temperature at which the T-cell receptor binds to a particular peptide:MHC complex makes a large difference in the extent of T-cell receptor aggregation; protein conformation is affected by temperature, and so these differences may well result from a conformational change.

From an examination of these structures it is hard to predict whether the main binding energy is contributed by T-cell receptor contacts with the bound peptide, or by T-cell receptor contacts with the MHC molecule. It is known that alterations as simple as changing a leucine to isoleucine in the peptide are sufficient to alter the T-cell response from

strong killing to absolutely no response at all. Studies show that mutations of single residues in the presenting MHC molecules can have the same effect. Thus, the specificity of T-cell recognition involves both the peptide and its presenting MHC molecule. This dual specificity underlies the MHC restriction of T-cell responses, a phenomenom that was observed long before the peptidebinding properties of MHC molecules were known. We will recount the story of how MHC restriction was discovered when we return to the issue of how MHC polymorphism affects antigen recognition by T cells in Chapter 5. Another consequence of this dual specificity is a need for T-cell receptors to be able to interact appropriately with the antigen-presenting surface of MHC molecules. It appears that there is some inherent specificity for MHC molecules encoded in the T-cell receptor genes, as well as selection during T-cell development for a repertoire of receptors able to interact appropriately with the particular MHC molecules present in that individual. We will be discussing the evidence for this in Chapter 7.

3-19. A distinct subset of T cells bears an alternative receptor made up of γ and δ chains.

During the search for the gene for the T-cell receptor α chain, another T-cell receptorlike gene was unexpectedly discovered. This gene was named T-cell receptor γ, and its discovery led to a search for further T-cell receptor genes. Another receptor chain was identified using antibody to the predicted sequence of the γ chain and was called the δ chain. It was soon discovered that a minority population of T cells bore a distinct type of T-cell receptor made up of γ:δ heterodimers rather than α:β heterodimers. The development of these cells is described in Sections 7-13 and 7-14.

To date, there is no crystallographic structure of a γ:δ T-cell receptor, although it is expected to be similar in shape to α:β T-cell receptors. γ:δ T-cell receptors may be specialized to bind certain kinds of ligands, including heat-shock proteins and nonpeptide ligands such as mycobacterial lipid antigens. It seems likely that γ:δ T-cell receptors are not restricted by the 'classical' MHC class I and class II molecules. They may bind the free antigen, much as immunoglobulins do, and/or they may bind to peptides or other antigens presented by nonclassical MHC-like molecules. These are proteins that resemble MHC class I molecules but are relatively nonpolymorphic. We still know little about how γ:δ T-cell receptors bind antigen and thus how these cells function, and what their role is in immune responses. The structure and rearrangement of the genes for γ:δ T-cell receptors is covered in Sections 4-13 and 7-13 and the functions of γ:δ T cells are considered in Chapter 8. TOP

Summary.

The receptor for antigen on most T cells, the α:β T-cell receptor, is composed of two protein chains, T-cell receptor α and T-cell receptor β, and resembles in many respects a single Fab fragment of immunoglobulin. T-cell receptors are always membrane-bound. α:β T-cell receptors do not recognize antigen in its native state, as do the immunoglobulin receptors of B cells, but recognize a composite ligand of a peptide antigen bound to an MHC molecule. MHC molecules are highly polymorphic glycoproteins encoded by genes in the major histocompatibility complex (MHC). Each MHC molecule binds a wide variety of different peptides, but the different variants each preferentially recognize sets of peptides with particular sequence and physical features. The peptide antigen is generated intracellularly, and bound stably in a peptide-binding cleft on the surface of the MHC molecule. There are two classes of MHC molecules and these are bound in their nonpolymorphic domains by CD8 and CD4 molecules that distinguish two different functional classes of α:β T cells. CD8 binds MHC class I molecules and can bind simultaneously to the same class I MHC:peptide complex being recognized by a T-cell receptor, thus acting as a co-receptor and enhancing the T-cell response; CD4 binds MHC class II molecules and acts as a co-receptor for T-cell receptors that recognise class II MHC:peptide ligands. T-cell receptors interact directly with both the antigenic peptide and polymorphic features of the MHC molecule that displays it, and this dual specificity underlies the MHC restriction of T-cell responses. A second type of T-cell receptor, composed of a γ and a δ chain, is structurally similar to the α:β T-cell receptor but appears to bind different ligands, including nonpeptide ligands. It is not thought to be MHC restricted. It is found on a minority population of T cells, the γ:δ T cells, whose biological function is still not clear.

Summary to Chapter 3.

B cells and T cells use different, but structurally similar, molecules to recognize antigen. The antigen-recognition molecules of B cells are immunoglobulins, and are made both as a membrane-bound receptor for antigen, the B-cell receptor, and as secreted antibodies that bind antigens and elicit humoral effector functions. The antigen-recognition molecules of T cells, on the other hand, are made only as cell-surface receptors. Immunoglobulins and T-cell receptors are highly variable molecules, with the variability concentrated in that part of the molecule, the variable (V)

region, that binds to antigen Immunoglobulins bind a wide variety of chemically different antigens, whereas the major α:β form of T-cell receptors will recognize only peptide fragments of foreign proteins bound to the MHC molecules that are ubiquitous on cell surfaces.

Binding of antigen by immunoglobulins has chiefly been studied using antibodies. The binding of antibody to its corresponding antigen is highly specific, and this specificity is determined by the shape and physicochemical properties of the antigen-binding site. The part of the antibody that elicits effector functions, once the variable part has bound an antigen, is located at the other end of the molecule from the antigen-binding sites, and is termed the constant region. There are five main functional classes of antibody, each encoded by a different type of constant region. As we will see in Chapter 9, these in turn interact with different components of the immune system to incite an inflammatory response and eliminate the antigen.

The T-cell receptor differs in several respects from the B-cell immunoglobulins. Among the most important of these differences is the absence of a secreted form of the receptor. This reflects the functional differences between T cells and B cells. B cells deal with pathogens and their protein products circulating within the body; secretion of a soluble antigen-recognition molecule by the activated B cell after antigen has been encountered enables them to mop up antigen effectively throughout the extracellular spaces of the body. T cells, on the other hand, are specialized for cell-cell interactions. They either kill cells that are infected with intracellular pathogens and that bear foreign antigenic peptides on their surface, or interact with cells of the immune system that have taken up foreign antigen and are displaying it on the cell surface. They thus have no requirement for a soluble, secreted receptor.

An additional distinctive feature of the T-cell receptor compared with immunoglobulins is that it recognizes a composite ligand made up of the foreign peptide bound to a self MHC molecule. This forces T cells to interact with infected body cells to become activated. Each T-cell receptor is specific for a particular combination of peptide and a self MHC molecule.

MHC molecules are encoded by a family of highly polymorphic genes and, although each individual expresses several, this represents only a small selection of all possible variants. During T-cell development, the T-cell receptor repertoire is selected so that the T cells of each individual recognize antigen only in conjunction with their own MHC molecules. Expression of multiple variant MHC molecules each with a different peptide-binding repertoire helps ensure that T cells from an individual will be able to recognize at least some peptides generated from nearly every pathogen.

References to Chapter 3

General references.

Ager, A., Callard, R., Ezine, S., Gerard, C., and Lopez-Botet, M.: Immune receptor supplement. *Immunol. Today* 1996, 17.

D.R. Davies and S. Chacko. Antibody structure *Acc. Chem. Res.* 1993. 26: 421-427.

Frazer, K., and Capra, J.D.: Immunoglobulins: structure and function, in Paul W.E. (ed): *Fundamental Immunology* , 4th edn. New York, Raven Press, 1998.

K.C. Garcia, L. Teyton, and I.A. Wilson. Structural basis of T cell recognition *Annu. Rev. Immunol.* 1999. 17: 369-397. (PubMed)

R.N. Germain. MHC-dependent antigen processing and peptide presentation: providing ligands for T lymphocyte activation *Cell* 1994. 76: 287-299. (PubMed)

Honjo, T., and Alt, F.W. (eds): *Immunoglobulin Genes* , 2nd edn. London, Academic Press, 1996.

G.(.e.d.). Moller. Origin of major histocompatibility complex diversity *Immunol. Rev.* 1995. 143: 5-292. (PubMed)

R.J. Poljak. Structure of antibodies and their complexes with antigens *Mol. Immunol.* 1991. 28: 1341-1345. (PubMed)

Section references.

3-1 IgG antibodies consist of four polypeptide chains.

G.M. Edelman. Antibody structure and molecular immunology *Scand. J. Immunol.* 1991. 34: 1-22. (PubMed)

C. Faber, L. Shan, Z. Fan, L.W. Guddat, C. Furebring, M. Ohlin, C.A.K. Borrebaeck, and A.B. Edmundson. Three-dimensional structure of a human Fab with high affinity for tetanus toxoid *Immunotechnology* 1998. 3: 253-270. (PubMed)

L.J. Harris, S.B. Larson, K.W. Hasel, J. Day, A. Greenwood, and A. McPherson. The 3-dimensional structure of an intact monoclonal antibody for canine lymphoma *Nature* 1992. 360: 369-372. (PubMed)

3-2 Immunoglobulin heavy and light chains are composed of constant and variable regions.

W.H. Han, J.X. Mou, J. Sheng, J. Yang, and Z.F. Shao. Cryo-atomic force microscopy—a new approach for biological imaging at high resolution *Biochemistry* 1995. 34: 8215-8220. (PubMed)

3-3 The antibody molecule can readily be cleaved into functionally distinct fragments.

R.R. Porter. Structural studies of immunoglobulins *Scand. J. Immunol.* 1991. 34: 381-389. (PubMed)

Y. Yamaguchi, H. Kim, K. Kato, K. Masuda, I. Shimada, and Y. Arata. Proteolytic fragmentation with high specificity of mouse IgG—mapping of proteolytic cleavage sites in the hinge region *J. Immunol. Methods* 1995. 181: 259-267. (PubMed)

3-4 The immunoglobulin molecule is flexible, especially at the hinge region.

M. Gerstein, A.M. Lesk, and C. Chothia. Structural mechanisms for domain movements in proteins *Biochemistry* 1994. 33: 6739-6749. (PubMed)

J.K. Kim, M.F. Tsen, V. Ghetie, and E.S. Ward. Evidence that the hinge region plays a role in maintaining serum levels of the murine IgG1 molecule *Mol. Immunol.* 1995. 32: 467-475. (PubMed)

3-5 The domains of an immunoglobulin molecule have similar structures.

Barclay, A.N., Brown, M.H., Law, S.K., McKnight, A.J., Tomlinson, M.G., and van der Merwe, P.A. (eds): *The Leukocyte Antigen Factsbook* , 2nd edn. London, Academic Press, 1997.

E. Hsu and L.A. Steiner. Primary structure of immunoglobulin through evolution *Curr. Opin. Struct. Biol.* 1992. 2: 422-430.

3-6 Localized regions of hypervariable sequence form the antigen-binding site.

V. Chitarra, P.M. Alzari, G.A. Bentley, T.N. Bhat, J.L. Eisele, A. Houdusse, J. Lescar, H. Souchon, and R.J. Poljak. 3-dimensional structure of a heteroclitic antigen-antibody cross reaction complex *Proc. Nat. Acad. Sci. USA* 1993. 90: 7711-7715. (PubMed) (Full Text in PMC)

L.K. Gilliland, N.A. Norris, H. Marquardt, T.T. Tsu, M.S. Hayden, M.G. Neubauer, D.E. Yelton, R.S. Mittler, and J.A. Ledbetter. Rapid and reliable cloning of antibody variable regions and generation of recombinant single-chain antibody fragments *Tissue Antigens* 1996. 47: 1-20. (PubMed)

G. Johnson and T.T. Wu. Kabat Database and its applications: 30 years after the first variability plot *Nucleic Acids Res.* 2000. 28: 214-218. (PubMed) (Full Text in PMC)

T.T. Wu and E.A. Kabat. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity *J. Exp. Med.* 1970. 132: 211-250. (PubMed)

3-7 Antibodies bind antigens via contacts with amino acids in CDRs, but the details of binding depend upon the size and shape of the

antigen. & 3-8 Antibodies bind to conformational shapes on the surfaces of antigens.

D.R. Davies and G.H. Cohen. Interactions of protein antigens with antibodies *Proc. Natl. Acad. Sci. USA* 1996. 93: 7-12. (PubMed) (Full Text in PMC)

E.A. Padlan. Anatomy of the antibody molecule *Mol. Immunol.* 1994. 31: 169-217. (PubMed)

R.L. Stanfield and I.A. Wilson. Protein-peptide interactions *Curr. Opin. Struct. Biol.* 1995. 5: 103-113. (PubMed)

I.A. Wilson and R.L. Stanfield. Antibody-antigen interactions: new structures and new conformational changes *Curr. Opin. Struct. Biol.* 1994. 4: 857-867. (PubMed)

3-9 Antigen-antibody interactions involve a variety of forces.

B.C. Braden and R.J. Poljak. Structural features of the reactions between antibodies and protein antigens *FASEB J.* 1995. 9: 9-16. (PubMed)

B.C. Braden, E.R. Goldman, R.A. Mariuzza, and R.J. Poljak. Anatomy of an antibody molecule: structure, kinetics, thermodynamics and mutational studies of the antilysozyme antibody D1 *Immunol. Rev.* 1998. 163: 45-57. (PubMed)

R. Ros, F. Schwesinger, D. Anselmetti, M. Kubon, R. Schäfer, A. Plückthun, and L. Tiefenauer. Antigen binding forces of individually addressed single-chain Fv antibody molecules *Proc. Natl. Acad. Sci. USA* 1998. 95: 7402-7405. (PubMed) (Full Text in PMC)

3-10 The antigen receptor on T cells is very similar to a Fab fragment of immunoglobulin.

K.C. Garcia, M. Degano, R.L. Stanfield, A. Brunmark, M.R. Jackson, P.A. Peterson, L. Teyton, and I.A. Wilson. An αβ T cell receptor structure at 2 5 Å and its orientation in the TCR-MHC complex. *Science* 1996. 274: 209-219. (PubMed)

D. Housset, G. Mazza, C. Gregoire, C. Piras, B. Malissen, and J.C. Fontecilla-Camps. The three-dimensional structure of a T cell antigen receptor VαVβ domain reveals a novel arrangement of the Vβ domain *EMBO J.* 1997. 16: 4205-4216. (PubMed)

3-11 A T-cell receptor recognizes antigen in the form of a complex of a foreign peptide bound to an MHC molecule.

M.M. Davis, J.J. Boniface, Z. Reich, D. Lyons, J. Hampl, B. Arden, and Y. Chien. Ligand recognition by alpha beta T cell receptors *Annu. Rev. Immunol.* 1998. 16: 523-544. (PubMed)

D.N. Garboczi, P. Ghosh, U. Utz, Q.R. Fan, W.E. Biddison, and D.C. Wiley. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2 *Nature* 1996. 384: 134-141. (PubMed)

3-12 T cells with different functions are distinguished by CD4 and CD8 cell-surface proteins and recognize peptides bound to different classes of MHC molecule.

G.F. Gao, J. Tormo, U.C. Gerth, J.R. Wyer, A.J. McMichael, D.I. Stuart, J.I. Bell, E.Y. Jones, and B.Y. Jakobsen. Crystal structure of the complex between human CD8αα and HLA-A2 *Nature* 1997. 387: 630-634. (PubMed)

L.A. Morrison, A.E. Lukacher, V.L. Braciale, D.P. Fan, and T.J. Braciale. Differences in antigen presentation to MHC class I- and class II-restricted influenza virus-specific cytolytic T-lymphocyte clones *J. Exp. Med.* 1986. 163: 903. (PubMed)

H. Wu, P.D. Kwong, and W.A. Hendrickson. Dimeric association and segmental variability in the structure of human CD4 *Nature* 1997. 387: 527-530. (PubMed)

R. Zamoyska. CD4 and CD8: modulators of T cell receptor recognition of antigen and of immune responses? *Curr. Opin. Immunol.* 1998. 10: 82-86. (PubMed)

3-13 The two classes of MHC molecule are expressed differentially on cells.

V. Steimle, C.A. Siegrist, A. Mottet, B. Lisowska-Grospierre, and B. Mach. Regulation of MHC class II expression by interferon-γ mediated by the transactivator gene CIITA *Science* 1994. 265: 106-109. (PubMed)

3-14 The two classes of MHC molecule have distinct subunit structures but similar three-dimensional structures. / 3-15 Peptides are stably bound to MHC molecules, and also serve to stabilize the MHC molecule on the cell surface.

A. Dessen, C.M. Lawrence, S. Cupo, D.M. Zaller, and D.C. Wiley. X-ray crystal structure of HLA-DR4 (DRA*0101, DRB1*0401) complexed with a peptide from human collagen II *Immunity* 1997. 7: 473-481. (PubMed)

D.H. Fremont, W.A. Hendrickson, P. Marrack, and J. Kappler. Structures of an MHC class II molecule with covalently bound single peptides *Science* 1996. 272: 1001-1004. (PubMed)

D.H. Fremont, M. Matsumura, E.A. Stura, P.A. Peterson, and I.A. Wilson. Crystal structures of two viral peptides in complex with murine MHC class 1 H-2Kb *Science* 1992. 257: 919-927. (PubMed)

D.H. Fremont, D. Monnaie, C.A. Nelson, W.A. Hendrickson, and E.R. Unanue. Crystal structure of I-Ak in complex with a dominant epitope of lysozyme *Immunity* 1998. 8: 305-317. (PubMed)

D.R. Madden, J.C. Gorga, J.L. Strominger, and D.C. Wiley. The three- dimensional structure of HLA-B27 at 2.1Å resolution suggests a general mechanism for tight peptide binding to MHC. *Cell* 1992. 70: 1035-1048. (PubMed)

V.L. Murthy and L.J. Stern. The class II MHC protein HLA-DR1 in complex with an endogenous peptide: implications for the structural basis of the specificity of peptide binding *Structure* 1997. 5: 1385-1396. (PubMed)

L.J. Stern and D.C. Wiley. Antigenic peptide binding by class I and class II histocompatibility proteins *Structure* 1994. 2: 245-251. (PubMed)

3-16 MHC class I molecules bind short peptides of 8–10 amino acids by both ends.

M. Bouvier and D.C. Wiley. Importance of peptide amino and carboxyl termini to the stability of MHC class I molecules *Science* 1994. 265: 398-402. (PubMed)

G.A. Weiss, E.J. Collins, D.N. Garboczi, D.C. Wiley, and S.L. Schreiber. A tricyclic ring system replaces the variable regions of peptides presented by three alleles of human MHC class I molecules *Chem. Biol.* 1995. 2: 401-407. (PubMed)

3-17 The length of the peptides bound by MHC class II molecules is not constrained.

H.G. Rammensee. Chemistry of peptides associated with MHC class I and class II molecules *Curr. Opin. Immunol.* 1995. 7: 85-96. (PubMed)

A.Y. Rudensky, P. Preston-Hurlburt, S.C. Hong, A. Barlow, and C.A. Janeway Jr. Sequence analysis of peptides bound to MHC class II molecules *Nature* 1991. 353: 622. (PubMed)

3-18 The crystal structures of several MHC:peptide:T-cell receptor complexes all show the same T-cell receptor orientation over the MHC:peptide complex.

Y.H. Ding, K.J. Smith, D.N. Garboczi, U. Utz, W.E. Biddison, and D.C. Wiley. Two human T cell receptors bind in a similar diagonal mode to the HLA-A2/Tax peptide complex using different TCR amino acids *Immunity* 1998. 8: 403-411. (PubMed)

K.C. Garcia, M. Degano, L.R. Pease, M. Huang, P.A. Peterson, L. Leyton, and I.A. Wilson. Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen *Science* 1998. 279: 1166-1172. (PubMed)

D.B. Sant'Angelo, G. Waterbury, P. Preston-Hurlburt, S.T. Yoon, R. Medzhitov, S.C. Hong, and C.A. Janeway Jr. The specificity and orientation of a TCR to its peptide-MHC class II ligands *Immunity* 1996. 4: 367-376. (PubMed)

M.K. Teng., A. Smolyar, A.G.D. Tse, J.H. Liu, J. Liu, R.E. Hussey, S.G. Nathenson, H.C. Chang, E.L. Reinherz, and

J.H. Wang. Identification of a common docking topology with substantial variation among different TCR-MHC-peptide complexes *Curr. Biol.* 1998. 8: 409-412. (PubMed)

3-19 A distinct subset of T cells bears an alternative receptor made up of γ and δ chains.

J.F. Bukowski, C.T. Morita, and M.B. Brenner. Human gamma delta T cells recognize alkylamines derived from microbes, edible plants, and tea: implications for innate immunity *Immunity* 1999. 11: 57-65. (PubMed)

Y.H. Chien, R. Jores, and M.P. Crowley. Recognition by gamma/delta T cells *Annu. Rev. Immunol.* 1996. 14: 511-532. (PubMed)

H. Li, M.I. Lebedeva, A.S. Llera, B.A. Fields, M.B. Brenner, and M.B. Mariuzza. Structure of the Vδ domain of a human γ:δ T-cell antigen receptor *Nature* 1998. 391: 502-506. (PubMed)

F.M. Spada, E.P. Grant, P.J. Peters, M. Sugita, A. Melian, D.S. Leslie, H.K. Lee, E. van Donselaar, D.A. Hanson, A.M. Krensky, O. Majdic, S.A. Porcelli, C.T. Morita, and M.B. Brenner. Self-recognition of CD1 by gamma/delta T cells: implications for innate immunity *J. Exp. Med.* 2000. 191: 937-948. (PubMed)

4. The Generation of Lymphocyte Antigen Receptors

➔ **Introduction to Chapter 4**

➔ **The generation of diversity in immunoglobulins**

➔ **T-cell receptor gene rearrangement**

➔ **Structural variation in immunoglobulin constant regions**

➔ **Summary to Chapter 4**

➔ **References to Chapter 4**

Introduction to Chapter 4

Lymphocyte antigen receptors, in the form of immunoglobulins on B cells and T-cell receptors on T cells, are the means by which lymphocytes sense the presence of antigens in their environment. The receptors produced by each lymphocyte have a unique antigen specificity, which is determined by the structure of their antigen-binding site, as described in Chapter 3. Because each person possesses billions of lymphocytes, these cells collectively provide the individual with the ability to respond to a great variety of antigens. The wide range of antigen specificities in the antigen receptor repertoire is due to variation in the amino acid sequence at the antigen-binding site, which is made up from the variable (V) regions of the receptor protein chains. In each chain the V region is linked to an invariant constant (C) region, which provides effector or signaling functions.

Given the importance of a diverse repertoire of lymphocyte receptors in the defense against infection, it is not surprising that a complex and elegant genetic mechanism has evolved for generating these highly variable proteins. Each receptor chain variant cannot be encoded in full in the genome, as this would require more genes for antigen receptors than there are genes in the entire genome. Instead, we will see that the V regions of the receptor chains are encoded in several pieces—so-called gene segments. These are assembled in the developing lymphocyte by somatic DNA recombination to form a complete V-region sequence, a mechanism known generally as **gene rearrangement**. Each type of gene segment is present in multiple copies in the germline genome. The selection of a gene segment of each type during gene rearrangement occurs at random, and the large number of possible different combinations accounts for much of the diversity of the receptor repertoire.

In the first two parts of this chapter we will describe the gene rearrangement mechanism that generates the V regions of immunoglobulin and T-cell receptor genes. The basic mechanism is common to both B cells and T cells, and involves many if not all of the same enzymes. We will describe the details of the enzymology of this recombination process, the evolution of which was probably critical to the evolution of the vertebrate adaptive immune system.

In B cells, but not T cells, the rearranged V region undergoes additional modification, known as somatic hypermutation. This does not occur until after B cells encounter and become activated by antigen. In these cells, the V regions of the assembled immunoglobulin genes undergo a high rate of point mutation that creates additional diversity within the expanding clone of B cells responding to antigen.

In the third part of the chapter we consider the limited, but functionally important, diversity of immunoglobulin C regions. The C regions of T-cell receptors do not show such diversity as they function only as part of a membrane-bound antigen receptor. Their role is to anchor and support the V regions at the cell surface as well as linking the binding of antigen by the V regions to the receptor-associated intracellular signaling complex. The C regions of immunoglobulins also serve these functions but in addition the C regions of the heavy chain are responsible for the effector functions of the secreted immunoglobulins, or antibodies, made by activated B cells. These C_H regions come in several different versions, or isotypes, each of which has a different effector function. In B cells that have become activated by antigen, the heavy-chain V region can become associated with a different C_H region by a further somatic recombination event, in the process known as isotype switching. This enables the different heavy-chain C regions, each with a different function, to be represented among antibodies of the same antigen specificity.

The generation of diversity in immunoglobulins.

Virtually any substance can elicit an antibody response. Furthermore, the response even to a simple antigen bearing a single antigenic determinant is diverse, comprising many different antibody molecules each with a unique affinity, or binding strength, for the antigen and a subtly different specificity. The total number of antibody specificities available to an individual is known as the antibody repertoire, or immunoglobulin repertoire, and in humans is at least 10^{11}, perhaps many more. The number of antibody specificities present at any one time is, however, limited by the total number of B cells in an individual, as well as by each individual's encounters with antigens.

Before it was possible to examine the immunoglobulin genes directly, there were two main hypotheses for the origin of this diversity. The germline theory held that there is a separate gene for each different immunoglobulin chain and that the antibody repertoire is largely inherited. By contrast, **somatic diversification theories** proposed that the observed repertoire is generated from a limited number of inherited V-region sequences that undergo alteration within B cells during the individual's lifetime. Cloning of the immunoglobulin genes revealed that the antibody repertoire is, in fact, generated by DNA rearrangements during B-cell development. As we will see in this part of the chapter, a DNA sequence encoding a V region is assembled at each locus by selection from a relatively small group of inherited gene segments. Diversity is further enhanced by the process of somatic hypermutation in mature activated B cells. Thus the somatic diversification theory was essentially correct, although the concept of multiple germline genes embodied in the germline theory also proved true.

4-1. Immunoglobulin genes are rearranged in antibody-producing cells.

In nonlymphoid cells, the gene segments encoding the greater part of the V region of an immunoglobulin chain are some considerable distance away from the sequence encoding the C region. In mature B lymphocytes, however, the assembled V-region sequence lies much nearer the C region, as a consequence of gene rearrangement. Rearrangement within the immunoglobulin genes was originally discovered 25 years ago, when it first became possible to study the organization of the immunoglobulin genes in both B cells and nonlymphoid cells using restriction enzyme analysis and Southern blotting. In this procedure, chromosomal DNA is first cut with a restriction enzyme, and the DNA fragments containing particular V- and C-region sequences are identified by hybridization with radiolabeled DNA probes specific for the relevant DNA sequences. In germline DNA, from nonlymphoid cells, the V- and C-region sequences identified by the probes are on separate DNA fragments. However, in DNA from an antibody-producing B cell these V- and C-region sequences are on the same DNA fragment, showing that a rearrangement of the DNA has occurred. A typical experiment using human DNA is shown in Fig. 4.1.

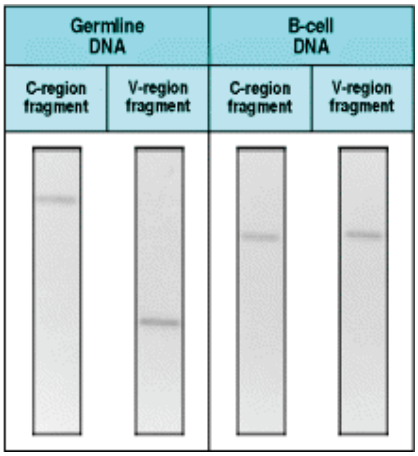


Figure 4.1. Immunoglobulin genes are rearranged in B cells. The two photographs on the left (germline DNA) show a Southern blot of a restriction enzyme digest of DNA from nonlymphoid cells from a normal person. The locations of immunoglobulin DNA sequences are identified by hybridization with V- and C-region probes. The V and C regions are found in distinct DNA fragments in the nonlymphoid DNA. The two photographs on the right (B-cell DNA) are of the same restriction digest of DNA from peripheral blood lymphocytes from a patient with chronic lymphocytic leukemia (see Chapter 7), in which a single clone of B cells is greatly expanded. The malignant B cells express the V region from which the V-region probe was obtained and, owing to their predominance in the cell population, this unique rearrangement can be detected. In this DNA, the V and C regions are found in the same fragment, which is a different size from either the C- or the V-region germline fragments. Although not shown in this

figure, a population of normal B lymphocytes has many different rearranged genes, so they yield a smear of DNA fragment sizes, which are not visible as a crisp band. Photograph courtesy of S. Wagner and L. Luzzatto.

This simple experiment showed that segments of genomic DNA within the immunoglobulin genes are rearranged in cells of the B-lymphocyte lineage, but not in other cells. This process of rearrangement is known as somatic recombination, to distinguish it from the meiotic recombination that takes place during the production of gametes.

4-2. The DNA sequence encoding a complete V region is generated by the somatic recombination of separate gene segments.

The V region, or V domain, of an immunoglobulin heavy or light chain is encoded by more than one gene segment. For the light chain, the V domain is encoded by two separate DNA segments. The first segment encodes the first 95–101 amino acids of the light chain and is termed a **V gene segment** because it encodes most of the V domain. The second segment encodes the remainder of the V domain (up to 13 amino acids) and is termed a **joining** or **J gene segment**.

The rearrangements that lead to the production of a complete immunoglobulin light-chain gene are shown in Fig. 4.2 (center panel). The joining of a V and a J gene segment creates a continuous exon that encodes the whole of the light-chain V region. In the unrearranged DNA, the V gene segments are located relatively far away from the C region. The J gene segments are located close to the C region, however, and joining of a V segment to a J gene segment also brings the V gene close to a C-region sequence. The J gene segment of the rearranged V region is separated from a C-region sequence only by an intron. In the experiment shown in Fig. 4.1, the germline DNA fragment identified by the 'V-region probe' contains the V gene segment, and that identified by the 'C-region probe' actually contains both the J gene segment and the C-region sequence. To make a complete immunoglobulin light-chain messenger RNA, the V-region exon is joined to the C-region sequence by RNA splicing after transcription (see Fig. 4.2).

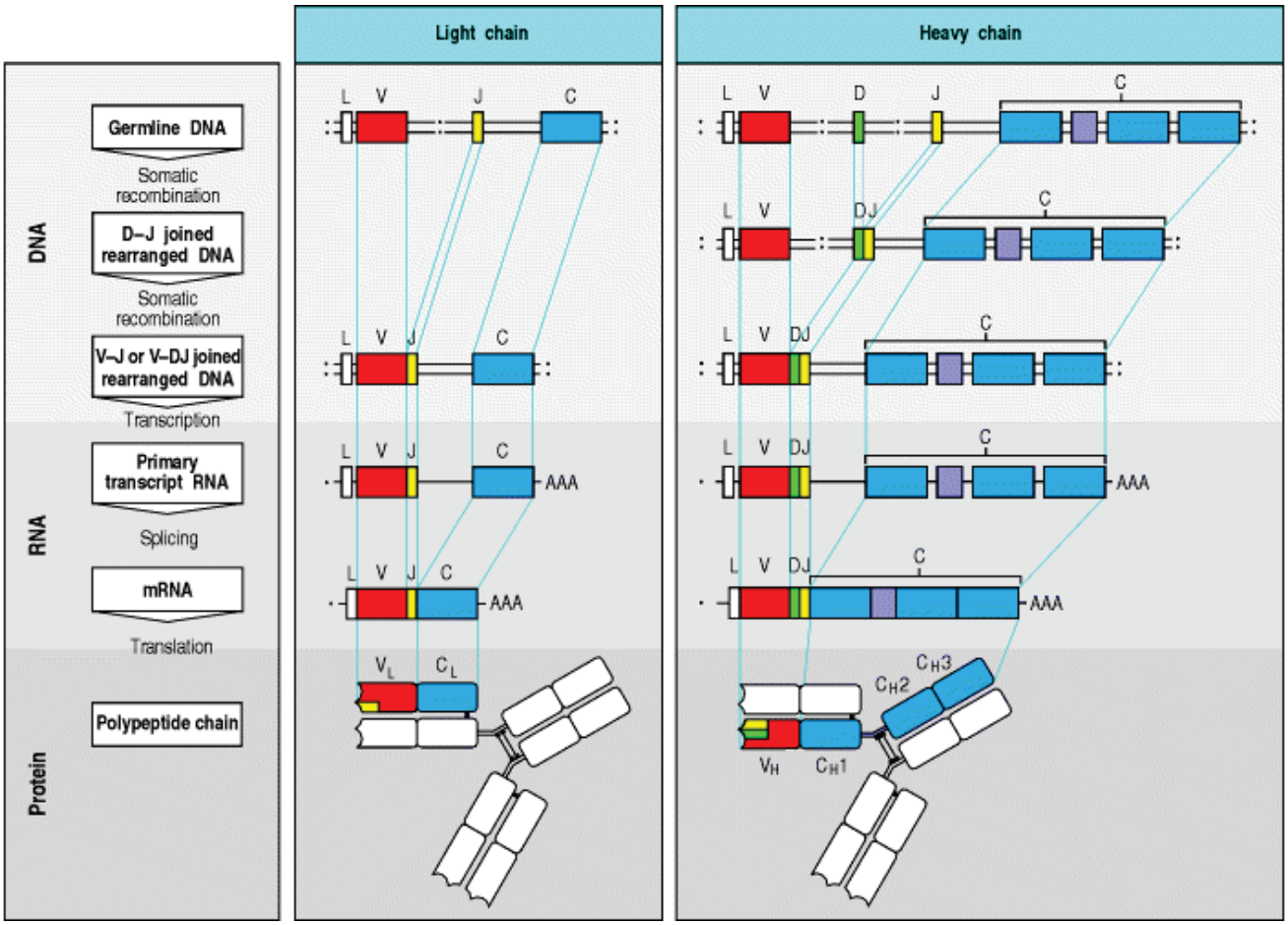


Figure 4.2. V-region genes are constructed from gene segments. Light-chain V-region genes are constructed from two segments (center panel). A variable (V) and a joining (J) gene segment in the genomic DNA are joined to form a complete light-chain V-region exon. Immunoglobulin chains are extra-cellular proteins and the V gene segment is preceded by an exon encoding a leader peptide (L), which directs the protein into the cell's secretory pathways and is

then cleaved. The light-chain C region is encoded in a separate exon and is joined to the V-region exon by splicing of the light-chain RNA to remove the L-to-V and the J-to-C introns. Heavy-chain V regions are constructed from three gene segments (right panel). First, the diversity (D) and J gene segments join, then the V gene segment joins to the combined DJ sequence, forming a complete V_H exon. A heavy-chain C-region gene is encoded by several exons. The C-region exons, together with the leader sequence, are spliced to the V-domain sequence during processing of the heavy-chain RNA transcript. The leader sequence is removed after translation and the disulfide bonds that link the polypeptide chains are formed. The hinge region is shown in purple.

A heavy-chain V region is encoded in three gene segments. In addition to the V and J gene segments (denoted V_H and J_H to distinguish them from the light-chain V_L and J_L), there is a third gene segment called the **diversity** or **D_H gene segment**, which lies between the V_H and J_H gene segments. The process of recombination that generates a complete heavy-chain V region is shown in Fig. 4.2 (right panel), and occurs in two separate stages. In the first, a D_H gene segment is joined to a J_H gene segment; then a V_H gene segment rearranges to DJ_H to make a complete V_H-region exon. As with the light-chain genes, RNA splicing joins the assembled V-region sequence to the neighboring C-region gene.

4-3. There are multiple different V-region gene segments.

For simplicity, we have so far discussed the formation of a complete immunoglobulin V-region sequence as though there were only a single copy of each gene segment. In fact, there are multiple copies of all of the gene segments in germline DNA. It is the random selection of just one gene segment of each type to assemble a V region that makes possible the great diversity of V regions among immunoglobulins. The numbers of functional gene segments of each type in the human genome, as determined by gene cloning and sequencing, are shown in Fig. 4.3. Not all the gene segments discovered are functional, as a proportion have accumulated mutations that prevent them from encoding a functional protein. These are termed 'pseudogenes.' Because there are many V, D, and J gene segments in germline DNA, no single one is essential. This reduces the evolutionary pressure on each gene segment to remain intact, and has resulted in a relatively large number of pseudogenes. Since some of these pseudogenes can undergo rearrangement just like a normal functional gene segment, a significant proportion of rearrangements will incorporate a pseudogene and thus be nonfunctional.

Number of functional gene segments in human immunoglobulin loci

Segment	Light chains		Heavy chain
	κ	λ	H
Variable (V)	40	30	65
Diversity (D)	0	0	27
Joining (J)	5	4	6

Figure 4.3. The numbers of functional gene segments for the V regions of human heavy and light chains. These numbers are derived from exhaustive cloning and sequencing of DNA from one individual and exclude all pseudogenes (mutated and nonfunctional versions of a gene sequence). Owing to genetic polymorphism, the numbers will not be the same for all people.

The immunoglobulin gene segments are organized into three clusters or genetic loci—the κ, λ, and heavy-chain loci. These are on different chromosomes and each is organized slightly differently, as shown in Fig. 4.4 for humans. At the λ light-chain locus, located on chromosome 22, a cluster of V_λ gene segments is followed by four sets of J_λ gene segments each linked to a single C_λ gene. In the κ light-chain locus, on chromosome 2, the cluster of V_κ gene segments is followed by a cluster of J_κ gene segments, and then by a single C_κ gene. The organization of the heavy-chain locus, on chromosome 14, resembles that of the κ locus, with separate clusters of V_H, D_H, and J_H gene segments and of C_H genes. The heavy-chain locus differs in one important way: instead of a single C-region, it contains a series of C regions arrayed one after the other, each of which corresponds to a different isotype. Generally, a cell expresses

only one at a time, beginning with IgM. The expression of other isotypes, such as IgG, can occur through isotype switching, as will be described in Section 4-16.

The human V gene segments can be grouped into families in which each member shares at least 80% DNA sequence identity with all others in the family. Both the heavy-chain and κ-chain V gene segments can be subdivided into seven such families, whereas there are eight families of V_λ gene segments. The families can be grouped into clans, made up of families that are more similar to each other than to families in other clans. Human V_H gene segments fall into three such clans. All of the V_H gene segments identified from amphibians, reptiles, and mammals also fall into the same three clans, suggesting that these clans existed in a common ancestor of these modern animal groups.

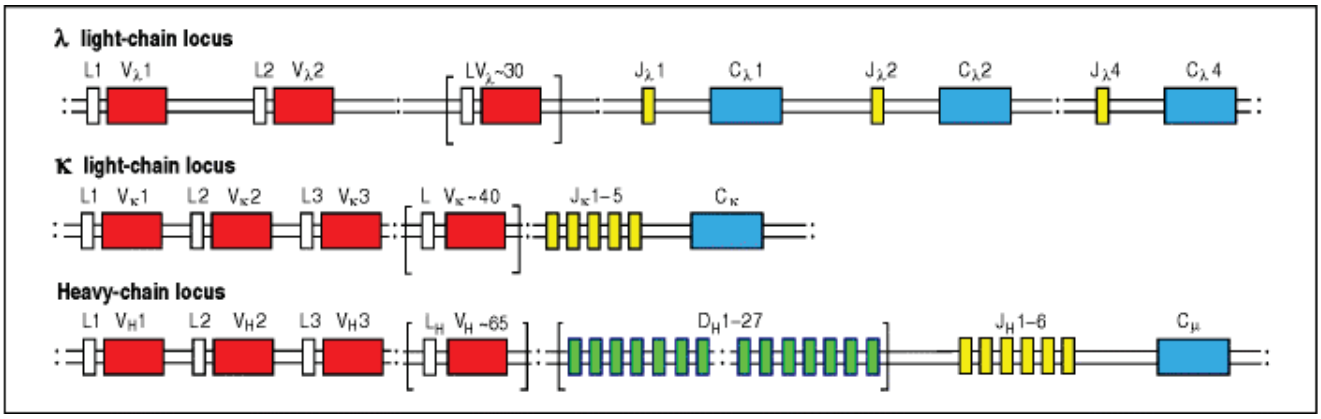


Figure 4.4. The germline organization of the immunoglobulin heavy- and light-chain loci in the human genome. The genetic locus for the λ light chain (chromosome 22) has about 30 functional V_λ gene segments and four pairs of functional J_λ gene segments and C_λ genes. The κ locus (chromosome 2) is organized in a similar way, with about 40 functional V_κ gene segments accompanied by a cluster of five J_κ gene segments but with a single C_κ gene. In approximately 50% of individuals, the entire cluster of κ V gene segments has undergone an increase by duplication (not shown for simplicity). The heavy-chain locus (chromosome 14) has about 65 functional V_H gene segments and a cluster of around 27 D segments lying between these V_H gene segments and six J_H gene segments. The heavy-chain locus also contains a large cluster of C_H genes that are described in Fig. 4.18. For simplicity we have shown only a single C_H gene in this diagram without illustrating its separate exons, have omitted pseudogenes, and have shown all V gene segments in the same orientation. L, leader sequence. This diagram is not to scale: the total length of the heavy-chain locus is over 2 megabases (2 million bases), whereas some of the D segments are only six bases long.

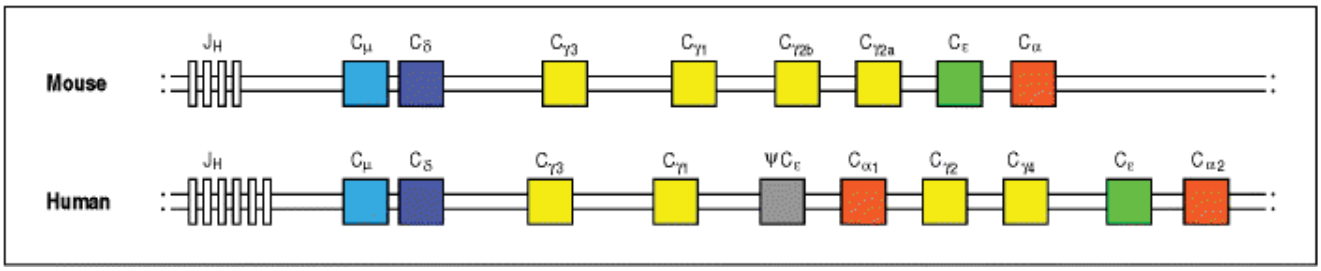


Figure 4.18. The organization of the immunoglobulin heavy-chain C-region genes in mice and humans (not to scale). In humans, the cluster shows evidence of evolutionary duplication of a unit consisting of two γ genes, an ε gene and an α gene. One of the εgenes has become inactivated and is now a pseudogene (ψ); hence only one subtype of IgE is expressed. For simplicity, other pseudogenes are not illustrated, and the exon details within each C gene are not shown. The classes of immunoglobulins found in mice are called IgM, IgD, IgG1, IgG2a, IgG2b, IgG3, IgA, and IgE.

4-4. Rearrangement of V, D, and J gene segments is guided by flanking DNA sequences.

A system is required to ensure that DNA rearrangements take place at the correct locations relative to the V, D, or J gene segment coding regions. In addition, joins must be regulated such that a V gene segment joins to a D or J and not to another V. DNA rearrangements are in fact guided by conserved noncoding DNA sequences that are found adjacent to the points at which recombination takes place. These sequences consist of a conserved block of seven

nucleotides—the **heptamer** 5′CACAGTG3′—which is always contiguous with the coding sequence, followed by a nonconserved region known as the spacer, which is either 12 or 23 nucleotides long. This is followed by a second conserved block of nine nucleotides—the **nonamer** 5′ACAAAAACC3′ (Fig. 4.5). The spacer varies in sequence but its conserved length corresponds to one or two turns of the DNA double helix. This brings the heptamer and nonamer sequences to the same side of the DNA helix, where they can be bound by the complex of proteins that catalyzes recombination. The heptamer-spacer-nonamer is called a **recombination signal sequence** (**RSS**).

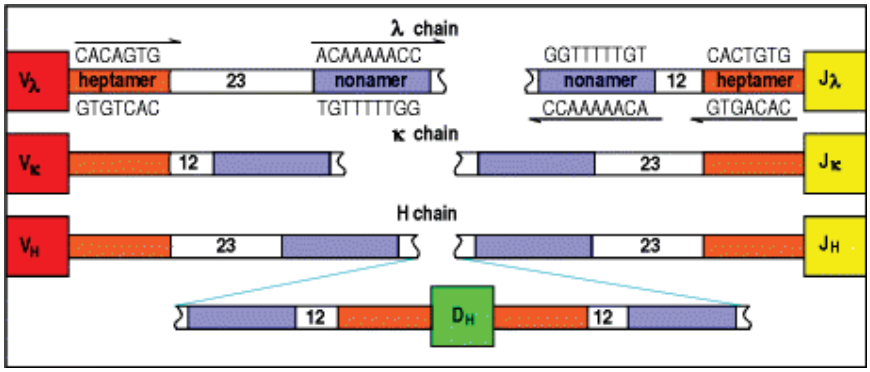


Figure 4.5. Conserved heptamer and nonamer sequences flank the gene segments encoding the V regions of heavy (H) and light (λ and κ) chains. The spacer (white) between the heptamer (orange) and nonamer (purple) sequences is always either approximately 12 bp or approximately 23 bp, and joining almost always involves a 12 bp and a 23 bp recombination signal sequence.

Recombination only occurs between gene segments located on the same chromosome. It generally follows the rule that only a gene segment flanked by a RSS with a 12-base pair (bp) spacer can be joined to one flanked by a 23 bp spacer RSS. This is known as the 12/23 rule. Thus, for the heavy chain, a D_H gene segment can be joined to a J_H gene segment and a V_H gene segment to a D_H gene segment, but V_H gene segments cannot be joined to J_H gene segments directly, as both V_H and J_H gene segments are flanked by 23 bp spacers and the D_H gene segments have 12 bp spacers on both sides (see Fig. 4.5).

It is now apparent, however, that, even though it violates the 12/23 rule, direct joining of one D gene segment to another can occur in most species. In humans, D-D fusion is found in approximately 5% of antibodies and is the major mechanism accounting for the unusually long CDR3 loops found in some heavy chains. By creating extra-long CDR3s and unusual amino acid combinations, these D-D fusions add further to the diversity of the antibody repertoire.

The mechanism of DNA rearrangement is similar for the heavy- and light-chain loci, although only one joining event is needed to generate a light-chain gene whereas two are needed to generate a complete heavy-chain gene. The commonest mode of rearrangement (Fig. 4.6, left panels) involves the looping-out and deletion of the DNA between two gene segments. This occurs when the coding sequences of the two gene segments are in the same orientation in the DNA. A second mode of recombination can occur between two gene segments that have opposite transcriptional orientations. This mode of recombination is less common, although such rearrangements account for about half of all V_κ to J_κ joins; the transcriptional orientation of half of the human V_κ gene segments is opposite to that of the J_κ gene segments. The mechanism of recombination is essentially the same, but the DNA that lies between the two gene segments meets a different fate (Fig. 4.6, right panels). When the RSSs in such cases are brought together and recombination takes place, the intervening DNA is not lost from the chromosome but is retained in an inverted orientation.

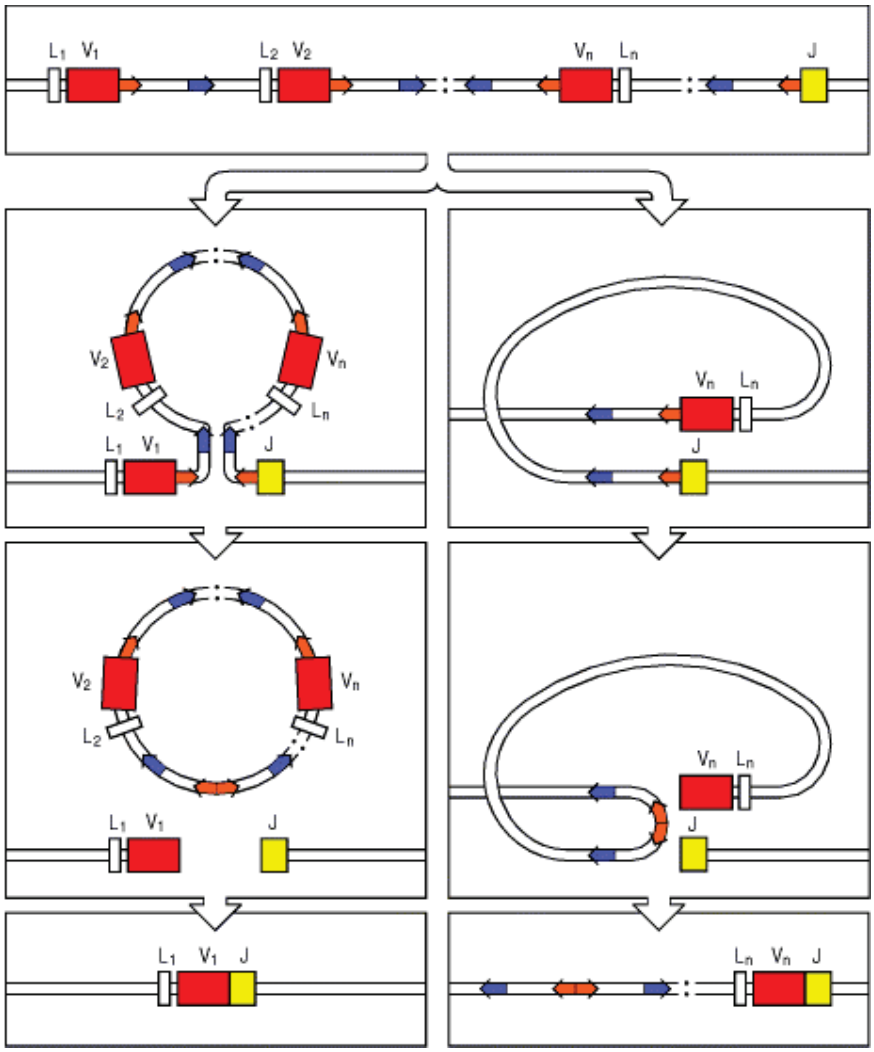


Figure 4.6. V-region gene segments are joined by recombination. In every V-region recombination event, the signals flanking the gene segments are brought together to allow recombination to take place. For simplicity, the recombination of a light-chain gene is illustrated; for the heavy-chain gene, two separate recombination events are required to generate a functional V region. In some cases, as shown in the left panels, the V and J gene segments have the same transcriptional orientation. Juxtaposition of the recombination signal sequences results in the looping out of the intervening DNA. Heptamers are shown in orange, nonamers in purple, and the arrows represent the directions of the heptamer and nonamer recombination signals (see Fig. 4.5). Recombination occurs at the ends of the heptamer sequences, creating a signal joint and releasing the intervening DNA in the form of a closed circle. Subsequently, the joining of the V and J gene segments creates the coding joint. In other cases, illustrated in the right panels, the V and J gene segments are initially oriented in opposite transcriptional directions. Bringing together the signal sequences in this case requires a more complex looping of the DNA. Joining the ends of the two heptamer sequences now results in the inversion and integration of the intervening DNA. Again, the joining of the V and J segments creates a functional V-region exon.

4-5. The reaction that recombines V, D, and J gene segments involves both lymphocyte-specific and ubiquitous DNA-modifying enzymes.

The molecular mechanism of V-region DNA rearrangement, or **V(D)J recombination**, is illustrated in Fig. 4.7. The 12 bp spaced and 23 bp spaced RSSs are brought together by interactions between proteins that specifically recognize the length of spacer and thus enforce the 12/23 rule for recombination. The DNA molecule is then broken in two places and rejoined in a different configuration. The ends of the heptamer sequences are joined precisely in a head-to-head fashion to form a signal joint in a circular piece of extrachromosomal DNA, which is lost from the genome when the cell divides. The V and J gene segments, which remain on the chromosome, join to form what is called the coding joint. This junction is imprecise, and consequently generates much additional variability in the V-region sequence.

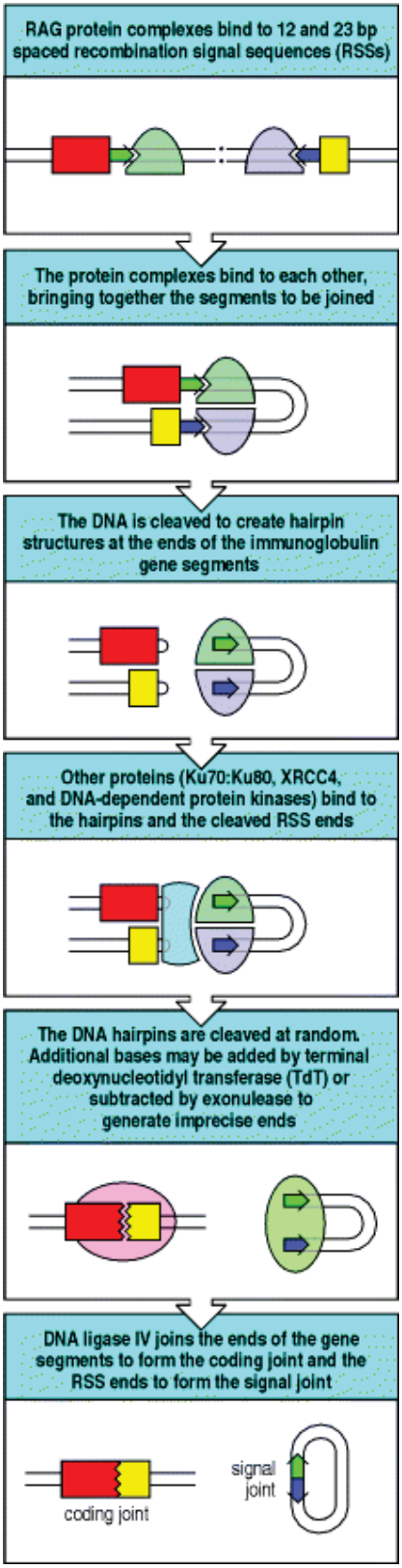


Figure 4.7. Enzymatic steps in the rearrangement of immunoglobulin gene segments. Rearrangement begins with the binding of RAG-1, RAG-2, and high mobility group (HMG) proteins (not shown). These RAG-1:RAG-2 complexes (domes, colored green or purple for clarity although they are identical at each recombination site) recognize the recombination signal sequences (arrows) flanking the coding sequences to be joined (red and yellow rectangles). These are then brought together (second panel), following which the RAG complex is activated to cut one strand of the double-stranded DNA precisely at the end of the heptamer sequences (third panel). The 5′ cut end of this DNA strand then reacts with the complementary uncut strand, breaking it to leave a double-stranded break at the end of the heptamer sequence, and forming a hairpin by joining to the cut end of its complementary strand on the other side of the break. Subsequently, through the action of additional essential proteins such as Ku70:Ku80 (indicated in blue) that join the complex (fourth panel) along with the RAG proteins, the DNA hairpin is cleaved at a

random site to yield a single-stranded DNA end. This end is then modified by the action of TdT and exonuclease (indicated in pink, fifth panel), which randomly creates diverse, imprecise ends. Finally (sixth panel) the two heptamer sequences, which are not modified, are ligated to form the precise signal joint, while the coding joint is also ligated, both by the action of DNA ligase IV.

The complex of enzymes that act in concert to effect somatic V(D)J recombination is termed the **V(D)J recombinase**. The products of the two genes RAG-1 and ***RAG-2*** (**recombination-activating genes**) comprise the lymphoid-specific components of the recombinase. This pair of genes is only expressed in developing lymphocytes while they are engaged in assembling their antigen receptors, as is described in more detail in Chapter 7. They are essential for V(D)J recombination. Indeed, these genes, when expressed together, are sufficient to confer on nonlymphoid cells such as fibroblasts the capacity to rearrange exogenous segments of DNA that contain appropriate RSSs; this is how *RAG-1* and *RAG-2* were initially discovered.

Although the RAG proteins are required for V(D)J recombination, they are not the only enzymes in the recombinase. The remaining enzymes are ubiquitously expressed DNA-modifying proteins that are involved in double-stranded DNA repair, DNA bending, or the modification of the ends of the broken DNA strands. They include the enzyme DNA ligase IV, the enzyme DNA-dependent protein kinase (DNA-PK), and Ku, a well-known autoantigen, which is a heterodimer (Ku 70:Ku 80) that associates tightly with DNA-PK.

V(D)J recombination is a multistep enzymatic process in which the first reaction is an endonucleolytic cleavage requiring the coordinated activity of both RAG proteins. Initially, two RAG protein complexes, each containing RAG-1, RAG-2, and high-mobility group proteins, recognize and align the two RSSs that are guiding the join (see Fig. 4.7). RAG-1 is thought to specifically recognize the nonamer of the RSS. At this stage, the 12/23 rule is established through mechanisms that are still poorly understood. The endonuclease activity of the RAG protein complexes then makes two single-strand DNA breaks at sites just 5′ of each bound RSS, leaving a free 3′-OH group at the end of each coding segment. This 3′-OH group then hydrolyzes the phosphodiester bond on the other strand, sealing the end of the double-stranded DNA to create a DNA 'hairpin' out of the gene segment coding region. This process simultaneously creates a flush double-stranded break at the ends of the two heptamer signal sequences. The DNA ends do not float apart, however, but are held tightly in a complex by the RAG proteins and other associated DNA repair enzymes until the join is completed. The two RSSs are precisely joined to form the signal joint. Coding joint formation is more complex. First, the DNA hairpin is nicked open by a single-stranded break, again by the RAG proteins. The nicking can happen at various points along the hairpin, which leads to sequence variability in the eventual joint. The DNA repair enzymes in the complex then modify the opened hairpins by removing nucleotides (by exonuclease activity) and by randomly adding nucleotides (by terminal deoxynucleotidyl transferase, TdT). It is not known if addition and deletion of nucleotides at the ends of coding regions occurs simultaneously or in a defined order. Finally, ligases such as DNA ligase IV join the processed ends together to generate a continuous double-stranded DNA, thus reconstituting a chromosome that includes the rearranged gene. This enzymatic process seems to create diversity in the joint between gene segments, while ensuring that the RSS ends are ligated without modification, and that unintended genetic damage such as a chromosomal break is avoided.

The recombination mechanism controlled by the RAG proteins shares many interesting features with the mechanism by which retroviral integrases catalyze the insertion of retroviral DNA into the genome, and also with the transposition mechanism used by transposons (mobile genetic elements that encode their own transposase, allowing them to excise and reinsert themselves in the genome). Even the structure of the *RAG* genes themselves, which lie close together in the chromosome and lack the usual mammalian introns, is reminiscent of a transposon. Indeed, it has recently been shown that the RAG complex can act as a transposase *in vitro*. These features have provoked speculation that the RAG complex originated as a transposase whose function was adapted by vertebrates to allow V gene segment recombination, thus leading to the advent of the vertebrate adaptive immune system. Consistent with this idea, no genes homologous to the *RAG* genes have been found in nonvertebrates.

The *in vivo* roles of the enzymes involved in V(D)J recombination have been established through natural or artificially induced mutations. Mice in which either of the *RAG* genes is knocked out suffer a complete block in lymphocyte development at the gene rearrangement stage. Mice lacking TdT do not add extra nucleotides to the joints between gene segments. A mutation that was discovered some time ago results in mice that make only trivial amounts of immunoglobulins or T-cell receptors. Such mice suffer from a **s**evere **c**ombined **i**mmune **d**eficiency—hence the name ***scid*** for this mutation. These mice have subsequently been found to have a mutation in the enzyme DNA-PK that prevents the efficient rejoining of DNA at gene segment junctions. Mutations of other proteins that are involved in DNA joining also give the *scid* phenotype.

4-6. The diversity of the immunoglobulin repertoire is generated by four main processes.

Antibody diversity is generated in four main ways. Two of these are consequences of the recombination process just discussed (see Sections 4-4 and 4-5) which creates complete immunoglobulin V-region exons during early B-cell development. The third is due to the different possible combinations of a heavy and a light chain in the complete immunoglobulin molecule. The fourth is a mutational process that occurs in mature B cells, acting only on rearranged DNA encoding the V regions.

The gene rearrangement that combines two or three gene segments to form a complete V-region exon generates diversity in two ways. First, there are multiple different copies of each type of gene segment, and different combinations of gene segments can be used in different rearrangement events. This combinatorial diversity is responsible for a substantial part of the diversity of the heavy- and light-chain V regions. Second, junctional diversity is introduced at the joints between the different gene segments as a result of addition and subtraction of nucleotides by the recombination process. A third source of diversity is also combinatorial, arising from the many possible different combinations of heavy- and light-chain V regions that pair to form the antigen-binding site in the immunoglobulin molecule. The two means of generating combinatorial diversity alone could give rise, in theory, to approximately 3.5×10^6 different antibody molecules (see Section 4-7). Coupled with junctional diversity, it is estimated that as many as 10^{11} different receptors could make up the repertoire of receptors expressed by naive B cells. Finally, somatic hypermutation introduces point mutations into the rearranged V-region genes of activated B cells, creating further diversity that can be selected for enhanced binding to antigen. We will discuss these mechanisms at greater length in the following sections.

4-7. The multiple inherited gene segments are used in different combinations.

There are multiple copies of the V, D, and J gene segments, each of which is capable of contributing to an immunoglobulin V region. Many different V regions can therefore be made by selecting different combinations of these segments. For human κ light chains, there are approximately 40 functional V_κ gene segments and five J_κ gene segments, and thus potentially 200 different V_κ regions. For λ light chains there are approximately 30 functional V_λ gene segments and four J_λ gene segments, yielding 120 possible V_λ regions. So, in all, 320 different light chains can be made as a result of combining different light-chain gene segments. For the heavy chains of humans, there are 65 functional V_H gene segments, approximately 27 D_H gene segments, and 6 J_H gene segments, and thus around 11,000 different possible V_H regions ($65 \times 27 \times 6 \approx 11{,}000$). During B-cell development, rearrangement at the heavy-chain gene locus to produce any one of the possible heavy chains is followed by several rounds of cell division before light-chain gene rearrangement takes place. The particular combination of gene segments used to produce a heavy chain does not appear to restrict the choice of gene segments that can be recombined to assemble a light-chain variable region. Thus, in theory any one possible heavy chain can be produced together with any one possible light chain in a single B cell. As both the heavy- and the light-chain V regions contribute to antibody specificity, each of the 320 different light chains could be combined with each of the approximately 11,000 heavy chains to give around 3.5×10^6 different antibody specificities. This theoretical estimate of combinatorial diversity is based on the number of germline V gene segments contributing to functional antibodies (see Fig. 4.3); the total number of V gene segments is larger, but the additional gene segments are pseudogenes and do not appear in expressed immunoglobulin molecules.

In practice, combinatorial diversity is likely to be less than one might expect from the theoretical calculations above. One reason for this is that not all V gene segments are used at the same frequency; some are common in antibodies, while others are found only rarely. It is also clear that not every heavy chain can pair with every light chain; certain combinations of V_H and V_L regions result in failure to assemble a stable immunoglobulin molecule. Cells that have heavy and light chains that cannot pair may continue to undergo light-chain gene rearrangement until a suitable light chain is produced, or may be eliminated, but in both cases a heavy- and light-chain combination that does not pair is lost from the repertoire. Nevertheless, it is thought that most heavy and light chains can pair with each other, and that this type of combinatorial diversity has a major role in the formation of an immunoglobulin repertoire with a wide range of specificities. In addition, two further processes add greatly to repertoire diversity—imprecise joining of V, D, and J gene segments and somatic hypermutation.

4-8. Variable addition and subtraction of nucleotides at the junctions between gene segments contributes to diversity in the third hypervariable region.

Of the three hypervariable loops in the protein chains of immunoglobulins, two are encoded within the V gene segment DNA. The third (HV3 or CDR3, see Fig. 3.6) falls at the joint between the V gene segment and the J gene

segment, and in the heavy chain is partially encoded by the D gene segment. In both heavy and light chains, the diversity of CDR3 is significantly increased by the addition and deletion of nucleotides at two steps in the formation of the junctions between gene segments. The added nucleotides are known as P-nucleotides and N-nucleotides and their addition is illustrated in Fig. 4.8.

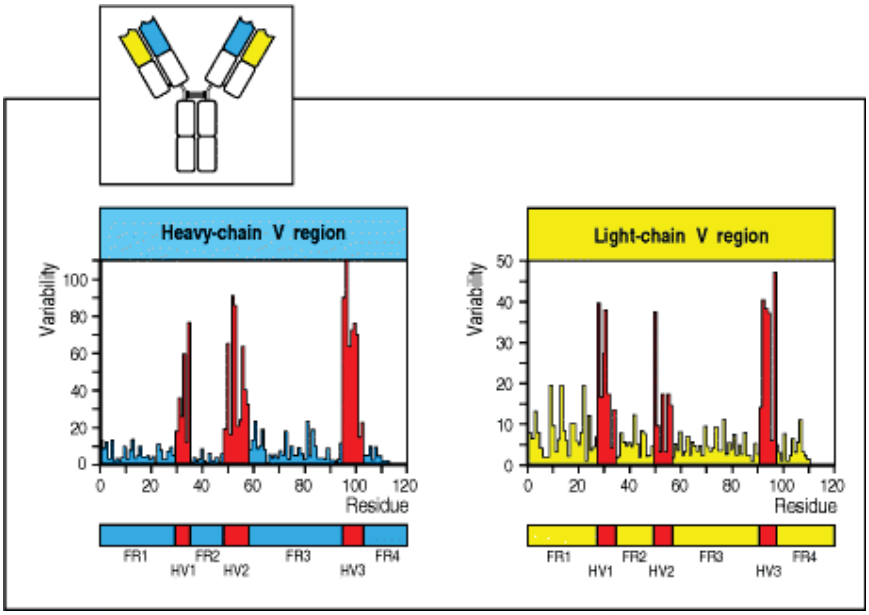


Figure 3.6. There are discrete regions of hypervariability in V domains. A variability plot derived from comparison of the amino acid sequences of several dozen heavy-chain and light-chain V domains. At each amino acid position the degree of variability is the ratio of the number of different amino acids seen in all of the sequences together to the frequency of the most common amino acid. Three hypervariable regions (HV1, HV2, and HV3) are indicated in red and are also known as the complementarity-determining regions, CDR1, CDR2, and CDR3. They are flanked by less variable framework regions (FR1, FR2, FR3, and FR4, shown in blue or yellow).

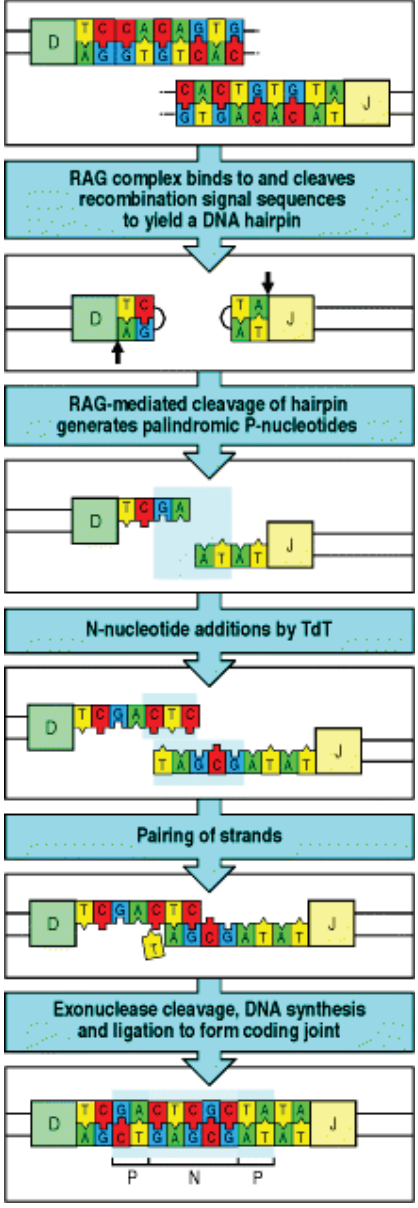


Figure 4.8. The introduction of P- and N-nucleotides at the joints between gene segments during immunoglobulin gene rearrangement. The process is illustrated for a D_H to J_H rearrangement; however, the same steps occur in V_H to D_H and in V_L to J_L rearrangements. After formation of the DNA hairpins (see Fig. 4.7), the two heptamer sequences, as indicated by the outline, are ligated to form the signal joint (not shown here), while RAG proteins cleave the DNA hairpin at a random site to yield a single-stranded DNA end. Depending on the site of cleavage, this single-stranded DNA may contain nucleotides that were originally complementary in the double-stranded DNA and which therefore form short DNA palindromes, as indicated by the shaded box in the third panel. Such stretches of nucleotides that originate from the complementary strand are known as P-nucleotides. For example, the sequence GA at the end of the D segment shown is complementary to the preceding sequence TC. Where the enzyme terminal deoxynucleotidyl transferase (TdT) is present, nucleotides are added at random to the ends of the single-stranded segments (fourth panel), indicated by the shaded box surrounding these nontemplated, or N, nucleotides. The two single-stranded ends then pair (fifth panel). Exonuclease trimming of unpaired nucleotides and repair of the coding joint by DNA synthesis and ligation leaves both the P- and N-nucleotides present in the final coding joint (indicated by shading in the bottom panel). The randomness of insertion of P- and N-nucleotides makes an individual P-N region a valuable marker for following an individual B-cell clone as it develops, for instance in studies of somatic hypermutation (see Fig. 4.9).

P-nucleotides are so called because they make up palindromic sequences added to the ends of the gene segments. After the formation of the DNA hairpins as described in Section 4-5, the RAG protein complex catalyzes a single-stranded cleavage at a random point within the coding sequence but near the original point at which the hairpin was first formed. When this cleavage occurs at a different point from the initial break, a single-stranded tail is formed from a few nucleotides of the coding sequence plus the complementary nucleotides from the other DNA strand (see

Fig. 4.8). In most light-chain gene rearrangements, DNA repair enzymes then fill in complementary nucleotides on the single-stranded tails which would leave short palindromic sequences at the joint, if the ends are rejoined without any further exonuclease activity (see below). In heavy-chain gene rearrangements and in some human light-chain genes, however, N-nucleotides are first added by a quite different mechanism.

N-nucleotides are so called because they are nontemplate-encoded. They are added by the enzyme terminal deoxynucleotidyl transferase (TdT) to single-stranded ends of the coding DNA after hairpin cleavage. After the addition of up to 20 nucleotides by this enzyme, the two single-stranded stretches at the ends of the gene segments form base pairs over a short region. Repair enzymes then trim off any nonmatching bases, synthesize complementary bases to fill in the remaining single-stranded DNA, and ligate it to the P-nucleotides (see Fig. 4.8). N-nucleotides are found especially in the V-D and D-J junctions of the assembled heavy-chain gene; they are less common in light-chain genes because TdT is expressed for only a short period in B-cell development, during the assembly of the heavy-chain gene, which occurs before that of the light-chain gene.

Nucleotides can also be deleted at gene segment junctions. This is accomplished by as yet unidentified exonucleases. Thus, the length of heavy-chain CDR3 can be even shorter than the smallest D segment. In some instances it is difficult, if not impossible, to recognize the D segment that contributed to CDR3 formation because of the excision of most of its nucleotides. Deletions may also erase the traces of P-nucleotide palindromes introduced at the time of hairpin opening. For this reason, many completed V(D)J joins do not show obvious evidence of P-nucleotides.

As the total number of nucleotides added by these processes is random, the added nucleotides often disrupt the reading frame of the coding sequence beyond the joint. Such frameshifts will lead to a nonfunctional protein, and DNA rearrangements leading to such disruptions are known as nonproductive rearrangements. As roughly two in every three rearrangements will be nonproductive, many B cells never succeed in producing functional immunoglobulin molecules, and junctional diversity is therefore achieved only at the expense of considerable wastage. We will discuss this further in Chapter 7.

4-9. Rearranged V genes are further diversified by somatic hypermutation.

The mechanisms for generating diversity described so far all take place during the rearrangement of gene segments in the initial development of B cells in the central lymphoid organs. There is an additional mechanism that generates diversity throughout the V region and that operates on B cells in peripheral lymphoid organs after functional immunoglobulin genes have been assembled. This process, known as somatic hypermutation, introduces point mutations into the V regions of the rearranged heavy- and light-chain genes at a very high rate, giving rise to mutant B-cell receptors on the surface of the B cells (Fig. 4.9). Some of the mutant immunoglobulin molecules bind antigen better than the original B-cell receptors, and B cells expressing them are preferentially selected to mature into antibody-secreting cells. This gives rise to a phenomenon called affinity maturation of the antibody population, which we will discuss in more detail in Chapters 9 and 10.

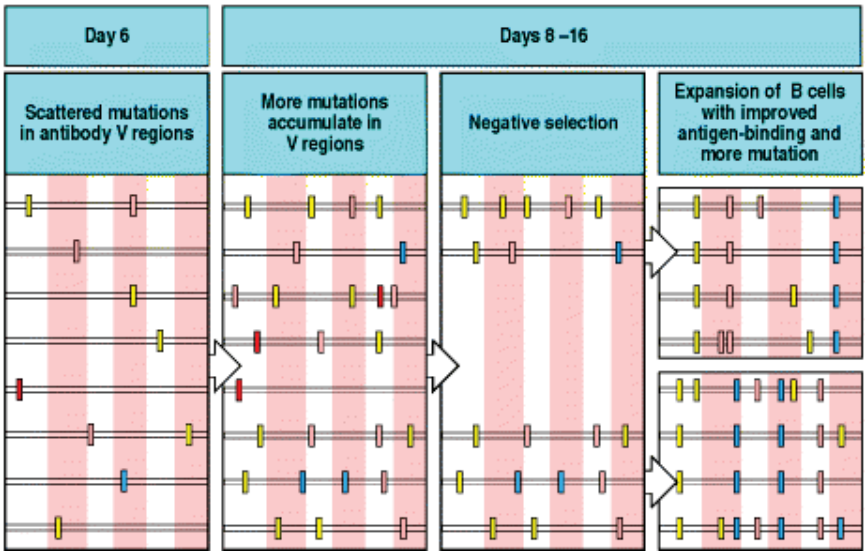


Figure 4.9. Somatic hypermutation introduces variation into the rearranged immunoglobulin variable region that is subject to negative and positive selection to yield improved antigen binding. In some circumstances it is

possible to follow the process of somatic hypermutation by sequencing immunoglobulin variable regions at different time points after immunization. The result of one such experiment is depicted here. Within a few days of immunization, it is found that the variable regions within a particular clone of responding B cells have begun to acquire mutations (first panel). Each variable region is represented by a horizontal line, on which the positions of the mutations are represented by vertical bars. These may be silent (yellow bars), neutral (pink bars), deleterious (red bars), or positive (blue bars). Over the course of the next week, more mutations accumulate (second panel). Those B cells whose variable regions have accumulated deleterious mutations and can no longer bind antigen die, a process of negative selection (third panel). Those B cells whose variable regions have acquired mutations that result in improved antigen binding are able to compete effectively for binding to the antigen, and receive signals that drive their proliferation and expansion, along with continued mutation (fourth panel). This process of mutation and selection can actually go through multiple cycles (not shown for simplicity) during the second and third weeks of the germinal center reaction. In this way, over time, the antigen-binding efficiency of the antibody response is improved.

Somatic hypermutation occurs when B cells respond to antigen along with signals from activated T cells. The immunoglobulin C-region gene, and other genes expressed in the B cell, are not affected, whereas the rearranged V_H and V_L genes are mutated even if they are nonproductive rearrangements and are not expressed. The pattern of nucleotide base changes in nonproductive V-region genes illustrates the result of somatic hypermutation without selection for enhanced binding to antigen. The base changes are distributed throughout the V region, but not completely randomly: there are certain 'hotspots' of mutation that indicate a preference for characteristic short motifs of four to five nucleotides, and perhaps also certain ill-defined secondary structural features. The pattern of base changes in the V regions of expressed immunoglobulin genes is different. Mutations that alter amino acid sequences in the conserved framework regions will tend to disrupt basic antibody structure and are selected against. In contrast, the result of selection for enhanced binding to antigen is that base changes that alter amino acid sequences, and thus protein structure, tend to be clustered in the CDRs, whereas silent mutations that preserve amino acid sequence and do not alter protein structure are scattered throughout the V region.

The mechanism of somatic hypermutation is poorly defined, but there have been several new discoveries that shed some light. It is known that mutation requires the presence of enhancers, DNA sequences that enhance the transcription of immunoglobulin genes in B cells, as well as a transcriptional promoter. The promoter, and the sequences that are the target of mutation need not derive from immunoglobulin V genes, however. The generation of new mutations in V regions in mutating B cells has recently been shown to be accompanied by double-stranded breaks in the DNA which are thought to then be repaired in an error-prone way. In addition, it has recently been discovered that deficiency in an RNA editing enzyme called Activation Induced Cytidine Deaminase, blocks the accumulation of somatic hypermutations. The mechanism by which this enzyme contributes to hypermutation is unknown. Interestingly, deficiency of this enzyme also abrogates the rearrangement of C-region genes that underlies the immunoglobulin class switching seen in activated B cells (see Section 4-16).

4-10. In some species most immunoglobulin gene diversification occurs after gene rearrangement.

As we have seen in the preceding sections, a proportion of the immunoglobulin diversity in an adult human derives from the existence of a variety of germline gene segments, and a proportion from somatic alterations acquired during the lifetime of the individual. This particular combination of heritable and acquired components of diversity operates in several mammalian immune systems, including those of humans and mice. Other species achieve a mix of inherited and acquired diversity by different means. Overall, it would appear that there is strong selective pressure to generate sufficient diversity in the immune system to protect the organism from common pathogens, and several different mechanisms have evolved toward this end.

In birds, rabbits, cows, pigs, sheep, and horses there is little or no germline diversity in the V, D, and J gene segments that are rearranged to form the genes for the initial B-cell receptors, and the rearranged V-region sequences are identical or similar in most immature B cells. These B cells then migrate to specialized microenvironments, the best known of which is the bursa of Fabricius in chickens. Here, B cells proliferate rapidly, and their rearranged immunoglobulin genes undergo further diversification. In birds and rabbits this occurs by a process that includes gene conversion, in which an upstream V segment pseudogene exchanges short sequences with the expressed rearranged V-region gene (Fig. 4.10). In sheep and cows, diversification is the result of somatic hypermutation, which occurs in an organ known as the ileal Peyer's patch. Somatic hypermutation probably also contributes to immunoglobulin diversification in birds and rabbits.

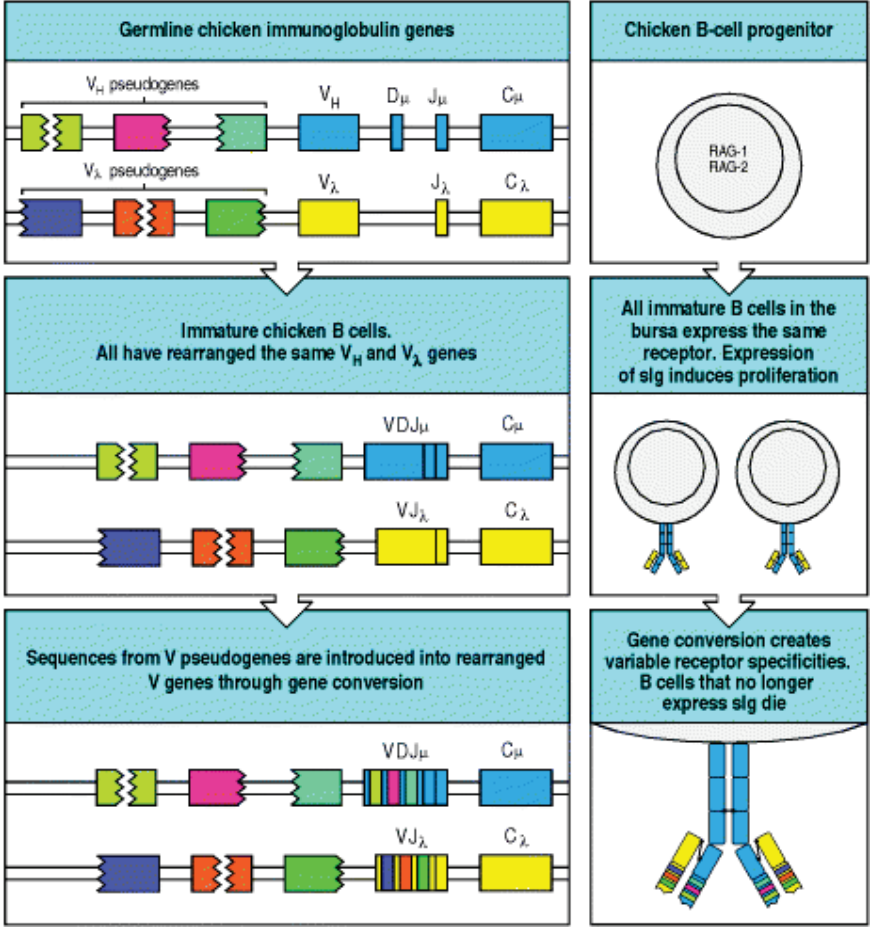


Figure 4.10. The diversification of chicken immunoglobulins occurs through gene conversion. In chickens, all B cells express the same surface immunoglobulin (sIg) initially; there is only one active V, D, and J gene segment for the chicken heavy-chain gene and one active V and J gene segment for each of the light-chain genes (top left panel). Gene rearrangement can thus produce only a single receptor specificity. Immature B cells expressing this receptor migrate to the bursa of Fabricius, where the expression of sIg induces cell proliferation (center panels). Gene conversion events introduce sequences from adjacent V pseudogenes into the expressed gene, creating diversity in the receptors (bottom panels). Some of these gene conversions will inactivate the previously expressed gene (not shown). If a B cell can no longer express sIg after such a gene conversion, it is eliminated.

Summary.

Diversity within the immunoglobulin repertoire is achieved by several means. Perhaps the most important factor that enables this extraordinary diversity is that V regions are encoded by separate gene segments, which are brought together by somatic recombination to make a complete V-region gene. Many different V-region gene segments are present in the genome of an individual, and thus provide a heritable source of diversity. Additional diversity, termed combinatorial diversity, results from the random recombination of separate V, D, and J gene segments to form a complete V-region exon. Variability at the joints between segments is increased by the insertion of random numbers of P- and N-nucleotides and by variable deletion of nucleotides at the ends of some coding sequences. The association of different light- and heavy-chain V regions to form the antigen-binding site of an immunoglobulin molecule contributes further diversity. Finally, after an immunoglobulin has been expressed, the coding sequences for its V regions are modified by somatic hypermutation upon stimulation of the B cell by antigen. The combination of all these sources of diversity generates a vast repertoire of antibody specificities from a relatively limited number of genes.

T-cell receptor gene rearrangement.

The mechanism by which B-cell antigen receptors are generated is such a powerful means of creating diversity that it is not surprising that the antigen receptors of T cells bear structural resemblances to immunoglobulins and are generated by the same mechanism. In this part of the chapter we describe the organization of the T-cell receptor loci and the generation of the genes for the individual T-cell receptor chains.

4-11. The T-cell receptor loci comprise sets of gene segments and are rearranged by the same enzymes as the immunoglobulin loci.

Like immunoglobulin heavy and light chains, T-cell receptor α and β chains each consist of a variable (V) amino-terminal region and a constant (C) region (see Section 3-10). The organization of the TCRα and TCRβ loci is shown in Fig. 4.11. The organization of the gene segments is broadly homologous to that of the immunoglobulin gene segments (see Sections 4-2 and 4-3). The TCRα locus, like those for the immunoglobulin light chains, contains V and J gene segments (V_α and J_α). The TCRβ locus, like that for the immunoglobulin heavy-chain, contains D gene segments in addition to V_β and J_β gene segments.

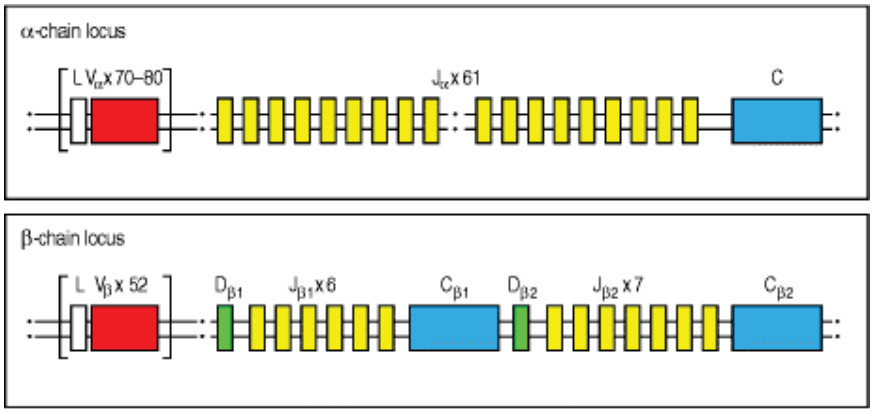


Figure 4.11. The germline organization of the human T-cell receptor α and β loci. The arrangement of the gene segments resembles that at the immunoglobulin loci, with separate variable (V), diversity (D), joining (J) gene segments, and constant (C) genes. The TCRα locus (chromosome 14) consists of 70–80 V_α gene segments, each preceded by an exon encoding the leader sequence (L). How many of these V_α gene segments are functional is not known exactly. A cluster of 61 J_α gene segments is located a considerable distance from the V_α gene segments. The J_α gene segments are followed by a single C gene, which contains separate exons for the constant and hinge domains and a single exon encoding the transmembrane and cytoplasmic regions (not shown). The TCRβ locus (chromosome 7) has a different organization, with a cluster of 52 functional V_β gene segments located distantly from two separate clusters each containing a single D gene segment, together with six or seven J gene segments and a single C gene. Each TCRβ C gene has separate exons encoding the constant domain, the hinge, the transmembrane region, and the cytoplasmic region (not shown). The TCRα locus is interrupted between the J and V gene segments by another T-cell receptor locus—the TCRδ locus (not shown here; see Fig. 4.15).

The T-cell receptor gene segments rearrange during T-cell development to form complete V-domain exons (Fig. 4.12). T-cell receptor gene rearrangement takes place in the thymus; the order and regulation of the rearrangements will be dealt with in detail in Chapter 7. Essentially, however, the mechanics of gene rearrangement are similar for B and T cells. The T-cell receptor gene segments are flanked by heptamer and nonamer recombination signal sequences (RSSs) that are homologous to those flanking immunoglobulin gene segments (see Section 4-4 and Fig. 4.5) and are recognized by the same enzymes. All known defects in genes that control V(D)J recombination affect T cells and B cells equally, and animals with these genetic defects lack functional lymphocytes altogether (see Section 4-5). A further shared feature of immunoglobulin and T-cell receptor gene rearrangement is the presence of P- and N-nucleotides in the junctions between the V, D, and J gene segments of the rearranged TCRβ gene. In T cells, P- and N-nucleotides are also added between the V and J gene segments of all rearranged TCRα genes, whereas only about half the V-J joints in immunoglobulin light-chain genes are modified by N-nucleotide addition and these are often left without any P-nucleotides as well (see Section 4-8 and Fig. 4.13).

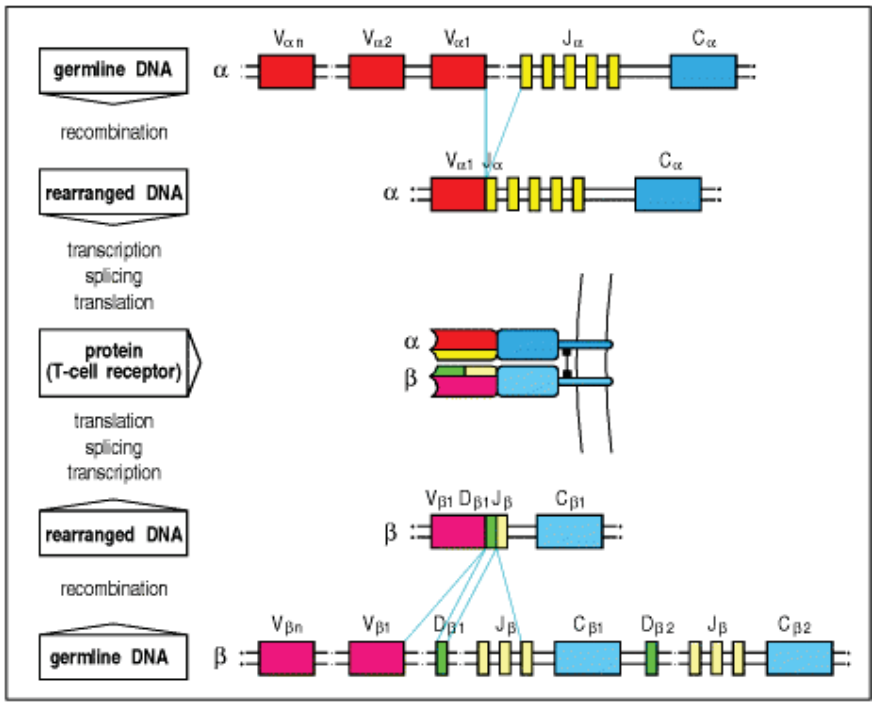


Figure 4.12. T-cell receptor α- and β-chain gene rearrangement and expression. The TCRα- and β-chain genes are composed of discrete segments that are joined by somatic recombination during development of the T cell. Functional α- and β-chain genes are generated in the same way that complete immunoglobulin genes are created. For the α chain (upper part of figure), a V_α gene segment rearranges to a J_α gene segment to create a functional V-region exon. Transcription and splicing of the VJ_α exon to C_α generates the mRNA that is translated to yield the T-cell receptor α-chain protein. For the β chain (lower part of figure), like the immunoglobulin heavy chain, the variable domain is encoded in three gene segments, V_β, D_β, and J_β. Rearrangement of these gene segments generates a functional VDJ_β V-region exon that is transcribed and spliced to join to C_β; the resulting mRNA is translated to yield the T-cell receptor β chain. The α and β chains pair soon after their biosynthesis to yield the α:β T-cell receptor heterodimer. Not all J gene segments are shown, and the leader sequences preceding each V gene segment are omitted for simplicity.

Element	Immunoglobulin		α:β receptors	
	H	κ+λ	β	α
Variable segments (V)	65	70	52	~70
Diversity segments (D)	27	0	2	0
D segments read in 3 frames	rarely	–	often	–
Joining segments (J)	6	5(κ) 4(λ)	13	61
Joints with N- and P-nucleotides	2	50% of joints	2	1
Number of V gene pairs	3.4×10^6		5.8×10^6	
Junctional diversity	$\sim 3 \times 10^7$		$\sim 2 \times 10^{11}$	
Total diversity	$\sim 10^{14}$		$\sim 10^{18}$	

Figure 4.13. The numbers of human T-cell receptor gene segments and the sources of T-cell receptor diversity compared with those of immunoglobulins. Note that only about half of human κ chains contain N-nucleotides. Somatic hypermutation as a source of diversity in immunoglobulins is not included in this figure.

The main differences between the immunoglobulin genes and those encoding T-cell receptors reflect the fact that all the effector functions of B cells depend upon secreted antibodies whose different heavy-chain C-region isotypes trigger distinct effector mechanisms. The effector functions of T cells, in contrast, depend upon cell-cell contact and are not mediated directly by the T-cell receptor, which serves only for antigen recognition. Thus, the C regions of the TCRα and TCRβ loci are much simpler than those of the immunoglobulin heavy-chain locus. There is only one C_α gene and, although there are two C_β genes, they are very closely homologous and there is no known functional distinction between their products. The T-cell receptor C-region genes encode only transmembrane polypeptides.

4-12. T-cell receptors concentrate diversity in the third hypervariable region.

The extent and pattern of variability in T-cell receptors and immunoglobulins reflect the distinct nature of their ligands. Whereas the antigen-binding sites of immunoglobulins must conform to the surfaces of an almost infinite variety of different antigens, and thus come in a wide variety of shapes and chemical properties, the ligand for the T-cell receptor is always a peptide bound to an MHC molecule. The antigen-recognition sites of T-cell receptors would therefore be predicted to have a less variable shape, with most of the variability focused on the bound antigenic peptide occupying the center of the surface in contact with the receptor.

In spite of differences in the sites of variability, the three-dimensional structure of the antigen-recognition site of a T-cell receptor looks much like that of an antibody molecule (see Sections 3-11 and 3-7, respectively). In an antibody, the center of the antigen-binding site is formed by the CDR3s of the heavy and light chains. The structurally equivalent third hypervariable loops (CDR3s) of the T-cell receptor α and β chains, to which the D and J gene segments contribute, also form the center of the antigen-binding site of a T-cell receptor; the periphery of the site consists of the equivalent of the CDR1 and CDR2 loops, which are encoded within the germline V gene segments for the α and β chains.

T-cell receptor loci have roughly the same number of V gene segments as do the immunoglobulin loci, but only B cells diversify rearranged V-region genes by somatic hypermutation. Thus, diversity in the CDR1 and CDR2 loops that comprise the periphery of the antigen-binding site will be far greater among antibody molecules than among T-cell receptors. This is in keeping with the fact that the CDR1 and CDR2 loops of a T-cell receptor will mainly contact the relatively less variable MHC component of the ligand rather than the highly variable peptide component (Fig. 4.14).

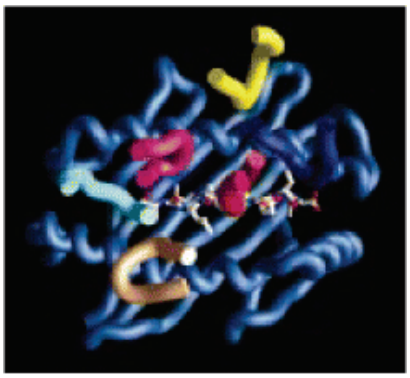


Figure 4.14. The most variable parts of the T-cell receptor interact with the peptide bound to an MHC molecule. The positions of the CDR loops of a T-cell receptor are shown as the colored tubes in this figure superimposing onto the MHC:peptide complex. The CDR1 loops of the α and β chains of the TCR are shown in pale and dark blue respectively, the CDR2 loops shown in fawn and yellow respectively, and the CDR3 loops shown in pale and dark red. The CDR3 loops lie in the center of the interface between the TCR and the MHC:peptide complex, and make direct contacts with the antigenic peptide

The structural diversity of T-cell receptors is mainly attributable to combinatorial and junctional diversity generated during the process of gene rearrangement. It can be seen from Fig. 4.13 that the variability in T-cell receptor chains is focused on the junctional region encoded by V, D, and J gene segments and modified by P- and N-nucleotides. The TCRα locus contains many more J gene segments than either of the immunoglobulin light-chain loci: in humans, 61 J_α gene segments are distributed over about 80 kb of DNA, whereas immunoglobulin light-chain loci have only five J gene segments at most (see Fig. 4.13). Because the TCRα locus has so many J gene segments, the variability generated in this region is even greater for T-cell receptors than for immunoglobulins. This region encodes the CDR3

loops in immunoglobulins and T-cell receptors that form the center of the antigen-binding site. Thus, the center of the T-cell receptor will be highly variable, whereas the periphery will be subject to relatively little variation. TOP

4-13. γ:δ T-cell receptors are also generated by gene rearrangement.

A minority of T cells bear T-cell receptors composed of γ and δ chains (see Section 3-19). The organization of the TCRγ and TCRδ loci (Fig. 4.15) resembles that of the TCRα and TCRβ loci, although there are important differences. The cluster of gene segments encoding the δ chain is found entirely within the TCRα locus, between the V_α and the J_α gene segments. Because all V_α gene segments are oriented such that rearrangement will delete the intervening DNA, any rearrangement at the α locus results in the loss of the δ locus. There are substantially fewer V gene segments at the TCRγ and TCRδ loci than at either the TCRα or TCRβ loci or at any of the immunoglobulin loci. Increased junctional variability in the δ chains may compensate for the small number of V gene segments and has the effect of focusing almost all of the variability in the γ:δ receptor in the junctional region. As we have seen, the amino acids encoded by the junctional regions lie at the center of the T-cell receptor binding site.

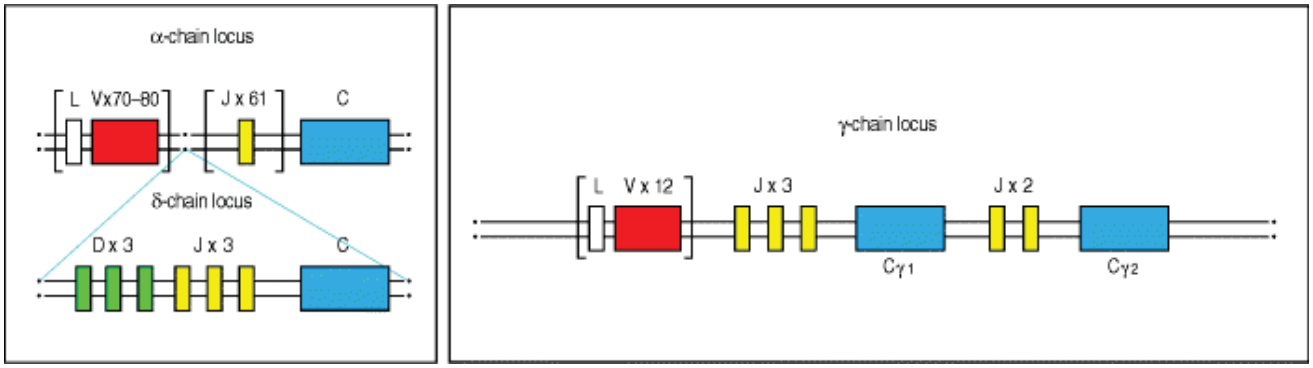


Figure 4.15. The organization of the T-cell receptor γ- and δ-chain loci in humans. The TCRγ and TCRδ loci, like the TCRα and TCRβ loci, have discrete V, D, and J gene segments, and C genes. Uniquely, the locus encoding the δ chain is located entirely within the α-chain locus. The three D_δ gene segments, three J_δ gene segments, and the single δ C gene lie between the cluster of V_α gene segments and the cluster of J_α gene segments, whereas the V_δ gene segments are interspersed among the V_α gene segments; it is not known exactly how many V_δ gene segments there are, but there are at least four. The human TCRγ locus resembles the TCRβ locus in having two C genes each with its own set of J gene segments. The mouse γ locus (not shown) has a more complex organization and there are three functional clusters of γ gene segments, each containing V and J gene segments and a C gene. Rearrangement at the γ and δ loci proceeds as for the other T-cell receptor loci, with the exception that during TCRδ rearrangement both D segments can be used in the same gene. The use of two D segments greatly increases the variability of the δ chain, mainly because extra N-region nucleotides can be added at the junction between the two D gene segments as well as at the V-D and D-J junctions.

T cells bearing γ:δ receptors are a distinct lineage of T cells whose functions are at present unknown. The ligands for these receptors are also largely unknown (see Section 3-19). Some γ:δ T-cell receptors appear to be able to recognize antigen directly, much as antibodies do, without the requirement for presentation by an MHC molecule or processing of the antigen. Detailed analysis of the rearranged V regions of γ:δ T-cell receptors shows that they resemble the V regions of antibody molecules more than they resemble the V regions of α:β T-cell receptors.

4-14. Somatic hypermutation does not generate diversity in T-cell receptors.

When we discussed the generation of antibody diversity in Section 4-9, we saw that somatic hypermutation increases the diversity of all three complementarity-determining regions of both immunoglobulin chains. Somatic hypermutation does not occur in T-cell receptor genes, so that variability of the CDR1 and CDR2 regions is limited to that of the germline V gene segments. All the diversity in T-cell receptors is generated during rearrangement and is consequently focused on the CDR3 regions.

Why T-cell and B-cell receptors differ in their abilities to undergo somatic hypermutation is not clear, but several explanations can be suggested on the basis of the functional differences between T and B cells. Because the central role of T cells is to stimulate both humoral and cellular immune responses, it is crucially important that T cells do not

react with self proteins. T cells that recognize self antigens are rigorously purged during development (see Chapter 7) and the absence of somatic hypermutation helps to ensure that somatic mutants recognizing self proteins do not arise later in the course of immune responses. This constraint does not apply with the same force to B-cell receptors, as B cells usually require T-cell help to secrete antibodies. A B cell whose receptor mutates to become self reactive would, under normal circumstances, fail to make antibody for lack of self-reactive T cells to provide this help (see Chapter 9).

A further argument is that T cells already interact with a self component, namely the MHC molecule that makes up the major part of the ligand for the receptor, and thus might be unusually prone to developing self-recognition capability through somatic hypermutation. In this case, the converse argument can also be made: because T-cell receptors must be able to recognize self MHC molecules as part of their ligand, it is important to avoid somatic mutation that might result in the loss of recognition and the consequent loss of any ability to respond. However, the strongest argument for this difference between immunoglobulins and T-cell receptors is the simple one that somatic hypermutation is an adaptive specialization for B cells alone, because they must make very high-affinity antibodies to capture toxin molecules in the extracellular fluids. We will see in Chapter 10 that they do this through somatic hypermutation followed by selection for antigen binding.

Summary.

T-cell receptors are structurally similar to immunoglobulins and are encoded by homologous genes. T-cell receptor genes are assembled by somatic recombination from sets of gene segments in the same way as are the immunoglobulin genes. Diversity is distributed differently in immunoglobulins and T-cell receptors; the T-cell receptor loci have roughly the same number of V gene segments but more J gene segments, and there is greater diversification of the junctions between gene segments during gene rearrangement. Moreover, functional T-cell receptors are not known to diversify their V genes after rearrangement through somatic hypermutation. This leads to a T-cell receptor in which the highest diversity is in the central part of the receptor, which contacts the bound peptide fragment of the ligand.

Structural variation in immunoglobulin constant regions.

So far we have focused on the structural variation inherent in the assembly of the V regions of the antibody molecule and T-cell receptor. We have seen how this variation creates a diverse repertoire of antigen-specificities, and we have also considered how these variable regions are attached to constant regions in the monovalent heterodimeric T-cell receptor, and the Y-shaped four-chain structure of the divalent immunoglobulin molecule. However, we have discussed only the general structural features of the immunoglobulin C region as illustrated by IgG, the most abundant type of antibody in plasma (see Section 3-1). Immunoglobulins can be made in several different forms, or isotypes, and we now consider how this structural variation is generated by linking different heavy-chain constant regions to the same variable region. The C_H regions, which determine the class or isotype of the antibody and thus its effector functions, are encoded in separate genes located downstream of the V genes at the heavy-chain locus. Initially only the first of these genes, the C_μ gene, is expressed in conjunction with an assembled V gene. However, during the course of an antibody response activated B cells can switch to the expression of a different downstream C_H gene by a process of somatic recombination known as isotype switching. In this part of the chapter we consider the structural features that distinguish the C_H regions of antibodies of the five major isotypes and confer on them their specialized functional properties as well as the mechanism of isotype switching. We also look at how alternative mRNA splicing allows the production of both membrane-bound and secreted forms of each isotype, and the simultaneous production of surface IgM and IgD in mature but naive B cells.

The use of isotype switching and alternative mRNA splicing to generate structural and functional variation is unique to the immunoglobulin heavy-chain locus and does not occur in T-cell receptor genes. This reflects the fact that immunoglobulins act as soluble molecules that must both bind antigen and recruit a variety of other effector cells and molecules to deal with it appropriately, whereas the T-cell receptor functions only as a membrane-bound receptor to activate an appropriate cellular immune response.

4-15. The immunoglobulin heavy-chain isotypes are distinguished by the structure of their constant regions.

The five main isotypes of immunoglobulin are IgM, IgD, IgG, IgE, and IgA. In humans, IgG antibodies can be further subdivided into four subclasses (IgG1, IgG2, IgG3, and IgG4), whereas IgA antibodies are found as two subclasses (IgA1 and IgA2). The IgG isotypes in humans are named in order of their abundance in serum, with IgG1 being the most abundant. The heavy chains that define these isotypes are designated by the lower-case Greek letters μ, δ, γ, ε, and α, as shown in Fig. 4.16, which also lists the major physical properties of the different human isotypes. IgM forms pentamers in serum, which accounts for its high molecular weight. Secreted IgA can occur as either a monomer or as a dimer.

	Immunoglobulin								
	IgG1	IgG2	IgG3	IgG4	IgM	IgA1	IgA2	IgD	IgE
Heavy chain	γ_1	γ_2	γ_3	γ_4	μ	α_1	α_2	δ	ε
Molecular weight (kDa)	146	146	165	146	970	160	160	184	188
Serum level (mean adult mg ml^{-1})	9	3	1	0.5	1.5	3.0	0.5	0.03	5×10^{-5}
Half-life in serum (days)	21	20	7	21	10	6	6	3	2
Classical pathway of complement activation	++	+	+++	−	+++	−	−	−	−
Alternative pathway of complement activation	−	−	−	−	−	+	−	−	−
Placental transfer	+++	+	++	−/+	−	−	−	−	−
Binding to macrophages and other phagocytes	+	−	+	−/+	−	+	+	−	+
High-affinity binding to mast cells and basophils	−	−	−	−	−	−	−	−	+++
Reactivity with staphylococcal Protein A	+	+	−/+	+	−	−	−	−	−

Figure 4.16. The properties of the human immunoglobulin isotypes. IgM is so called because of its size: although monomeric IgM is only 190 kDa, it normally forms pentamers, known as macroglobulin (hence the M), of very large molecular weight (see Fig. 4.23). IgA dimerizes to give a molecular weight of around 390 kDa in secretions. IgE antibody is associated with immediate-type hypersensitivity. When fixed to tissue mast cells, IgE has a much longer half-life than its half-life in plasma shown here. The activation of the alternative pathway of complement by IgA1 is caused not by its Fc portion but by its Fab portion. A subset of human V_H regions (clan 3, see Section 4-3) also binds staphylococcal Protein A through the Fab portion.

Sequence differences between immunoglobulin heavy chains cause the various isotypes to differ in several characteristic respects. These include the number and location of interchain disulfide bonds, the number of attached oligosaccharide moieties, the number of C domains, and the length of the hinge region (Fig. 4.17). IgM and IgE heavy chains contain an extra C domain that replaces the hinge region found in γ, δ, and α chains. The absence of the hinge region does not imply that IgM and IgE molecules lack flexibility; electron micrographs of IgM molecules binding to ligands show that the Fab arms can bend relative to the Fc portion. However, such a difference in structure may have functional consequences that are not yet characterized. Different isotypes and subtypes also differ in their ability to engage various effector functions, as will be described in Section 4-18.

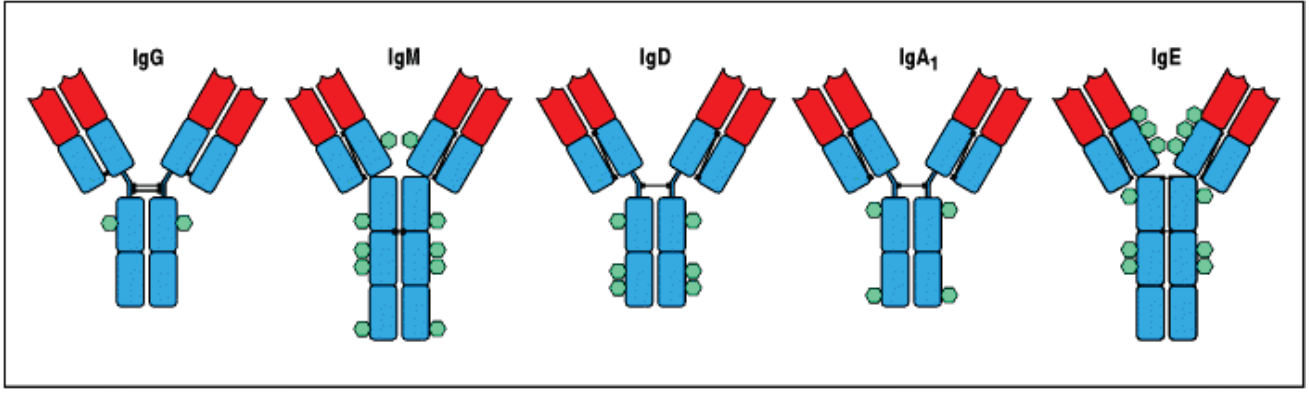


Figure 4.17. The structural organization of the main human immunoglobulin isotype monomers. Both IgM and IgE lack a hinge region but each contains an extra heavy-chain domain. Note the differences in the numbers and locations of the disulfide bonds (black lines) linking the chains. The isotypes also differ in the distribution of N-linked carbohydrate groups, shown as turquoise hexagons

4-16. The same V_H exon can associate with different C_H genes in the course of an immune response.

The V-region exons expressed by any given B cell are determined during its early differentiation in the bone marrow and, although they may subsequently be modified by somatic hypermutation, no further V(D)J recombination occurs. All the progeny of that B cell will therefore express the same assembled V genes. By contrast, several different C-region genes can be expressed in the B cell's progeny as the cells mature and proliferate in the course of an immune response. Every B cell begins by expressing IgM as its B-cell receptor, and the first antibody produced in an immune response is always IgM. Later in the immune response, however, the same assembled V region may be expressed in IgG, IgA, or IgE antibodies. This change is known as isotype switching. It is stimulated in the course of an immune response by external signals such as cytokines released by T cells or mitogenic signals delivered by pathogens, as we will discuss further in Chapter 9. Here we are concerned with the molecular basis of the isotype switch.

The immunoglobulin C_H genes form a large cluster spanning about 200 kb to the 3′ side of the J_H gene segments (Fig. 4.18). Each C_H gene is split into several exons (not shown in the figure), each corresponding to an individual immunoglobulin domain in the folded C region. The gene encoding the μ C region lies closest to the J_H gene segments, and therefore closest to the assembled V-region exon after DNA rearrangement. A complete μ heavy-chain transcript is produced from the newly rearranged gene. Any J_H gene segments remaining between the assembled V gene and the C_μ gene are removed during RNA processing to generate the mature mRNA. μ heavy chains are therefore the first to be expressed and IgM is the first immunoglobulin isotype to be expressed during B-cell development.

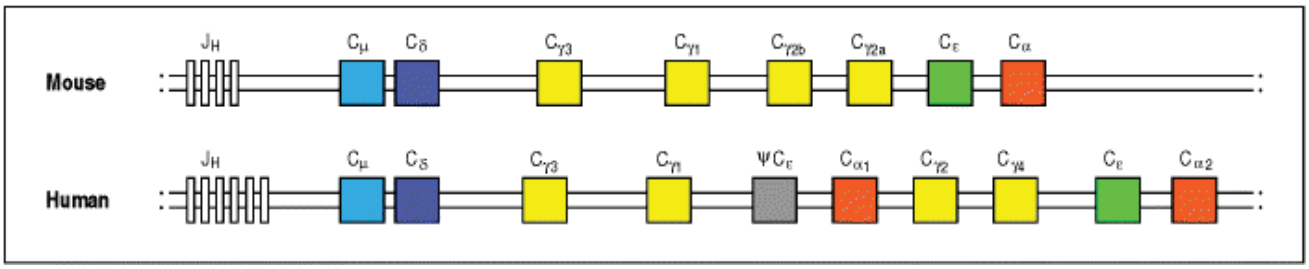


Figure 4.18. The organization of the immunoglobulin heavy-chain C-region genes in mice and humans (not to scale). In humans, the cluster shows evidence of evolutionary duplication of a unit consisting of two γ genes, an ε gene and an α gene. One of the εgenes has become inactivated and is now a pseudogene (ψ); hence only one subtype of IgE is expressed. For simplicity, other pseudogenes are not illustrated, and the exon details within each C gene are not shown. The classes of immunoglobulins found in mice are called IgM, IgD, IgG1, IgG2a, IgG2b, IgG3, IgA, and IgE.

Immediately 3′ to the μ gene lies the δ gene, which encodes the C region of the IgD heavy chain. IgD is coexpressed with IgM on the surface of almost all mature B cells, although this isotype is secreted in only small amounts and its function is unknown. Indeed, mice lacking the C_δ exons seem to have essentially normal immune systems. B cells expressing IgM and IgD have not undergone isotype switching, which, as we will see, entails an irreversible change

in the DNA. Instead, these cells produce a long primary transcript that is differentially cleaved and spliced to yield one of two distinct mRNA molecules. In one of these, the VDJ exon is linked to the C_μ exons to encode a μ heavy chain, and in the other the VDJ exon is linked to the C_δ exons to encode a δ heavy chain (Fig. 4.19). The differential processing of the long mRNA transcript is developmentally regulated, with immature B cells making mostly the μ transcript and mature B cells making mostly the δ form along with some of the μ transcript.

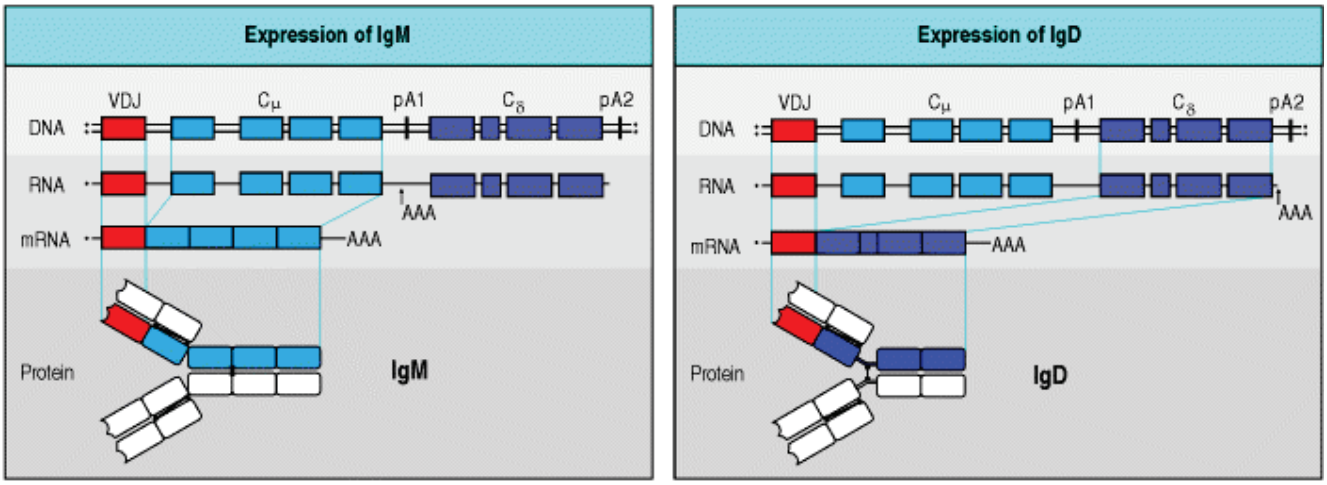


Figure 4.19. Co-expression of IgD and IgM is regulated by RNA processing. In mature B cells, transcription initiated at the V_H promoter extends through both C_μ and C_δ exons. This long primary transcript is then processed by cleavage and polyadenylation (AAA), and by splicing. Cleavage and polyadenylation at the μ site (pA1) and splicing between C_μ exons yields an mRNA encoding the μ heavy chain (left panel). Cleavage and polyadenylation at the δ site (pA2) and a different pattern of splicing that removes the C_μ exons yields mRNA encoding the δ heavy chain (right panel). For simplicity we have not shown all the individual C-region exons.

Switching to other isotypes occurs only after B cells have been stimulated by antigen. It occurs through a specialized nonhomologous DNA recombination mechanism guided by stretches of repetitive DNA known as switch regions. Switch regions lie in the intron between the J_H gene segments and the C_μ gene, and at equivalent sites upstream of the genes for each of the other heavy-chain isotypes, with the exception of the δ gene (Fig. 4.20, top panel). The μ switch region (S_μ) consists of about 150 repeats of the sequence [$(GAGCT)_n$ (GGGGGT)], where *n* is usually 3 but can be as many as 7. The sequences of the other switch regions (S_γ, S_α, and S_ε) differ in detail but all contain repeats of the GAGCT and GGGGGT sequences. Exactly how these repetitive sequences promote switch recombination is unclear because the enzyme(s) that promote switch recombination have not been identified; however, it is thought that the repetitive sequences might promote short stretches of homologous alignment that in turn promote DNA recombination.

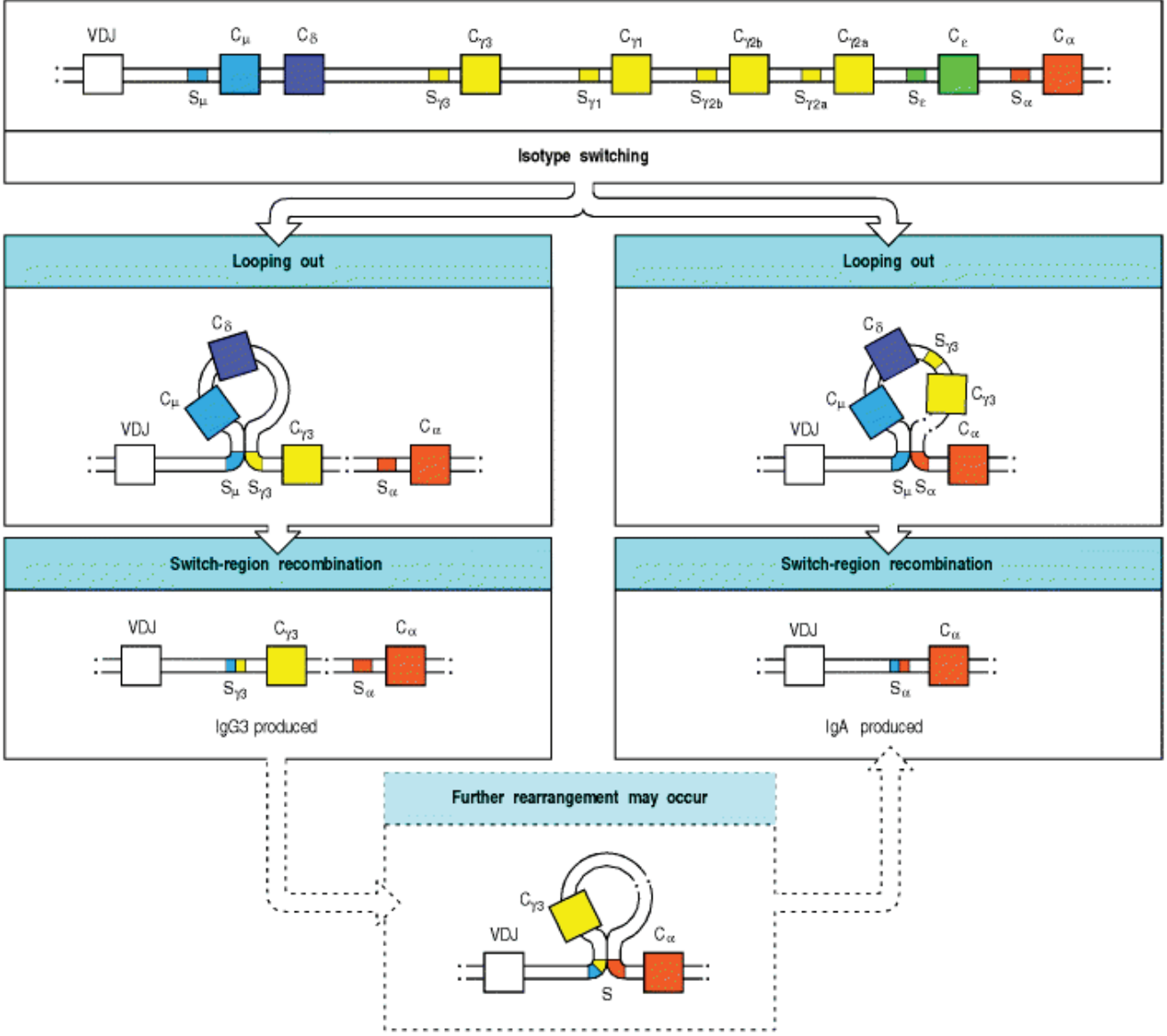


Figure 4.20. Isotype switching involves recombination between specific switch signals. Repetitive DNA sequences that guide isotype switching are found upstream of each of the immunoglobulin C-region genes, with the exception of the δ gene. The figure illustrates switching in the mouse heavy-chain genes. Switching occurs by recombination between these repetitive sequences, or switch signals, with deletion of the intervening DNA. The initial switching event takes place from the μ switch region; switching to other isotypes can take place subsequently from the recombinant switch region formed after μ switching. S, switch region.

When a B cell switches from coexpression of IgM and IgD to expression of an IgG subtype, DNA recombination occurs between S_μ and the S_γ of that IgG subtype. The C_μ and C_δ coding regions are deleted, and γ heavy-chain transcripts are made from the recombined gene. Figure 4.20 (left panels) illustrates switching to γ3 in the mouse. Some of the progeny of this IgG-producing cell may subsequently undergo a further switching event to produce a different isotype, for example IgA, as shown in the bottom panel of Fig. 4.20. Alternatively, as shown in the right panels of Fig. 4.20, the switch recombination may occur between S_μ and one of the switch regions downstream of the C_γ genes so that the cell switches from IgM to IgA or IgE (illustrated for IgA only). All switch recombination events produce genes that can encode a functional protein because the switch sequences lie in introns and therefore cannot cause frame shift mutations.

The enzymes that carry out isotype switching have not been clearly defined. However, we do know that DNA repair enzymes are involved since switching is markedly reduced in Ku protein knockouts; Ku proteins are also essential for the rejoining of DNA during V(D)J joining (see Section 4-5). Recently, it was discovered that deficiency in Activation Induced Cytidine Deaminase completely blocks isotype switching. As mentioned in Section 4-9, this deficiency also blocks somatic hypermutation. Activation Induced Cytidine Deaminise is thought to be an RNA editing enzyme and how it works to enable both hypermutation and switching is unknown at present. A deficiency in

this enzyme in humans has now been associated with a form of immuno-deficiency known as Hyper IgM type 2 syndrome, which is characterized by an absence of immunoglobulins other than IgM. A failure of T cells to activate isotype switching leads to a similar syndrome now classified as Hyper IgM type 1 (see Section 11-9).

Isotype switch recombination is unlike V(D)J recombination in several ways. First, all isotype switch recombination is productive; second, it uses different recombination signal sequences and enzymes; third, it happens after antigen stimulation and not during B-cell development in the bone marrow; and fourth, the switching process is not random but is regulated by external signals such as those provided by T cells, as will be discussed in Chapter 9.

4-17. Transmembrane and secreted forms of immunoglobulin are generated from alternative heavy-chain transcripts.

Immunoglobulins of all heavy-chain isotypes can be produced either in secreted form or as a membrane-bound receptor. All B cells initially express the transmembrane form of IgM; after antigen stimulation, some of their progeny differentiate into plasma cells producing the secreted form of IgM, whereas others undergo isotype switching to express transmembrane immunoglobulins of a different isotype before switching to the production of secreted antibody of the new isotype.

The membrane forms of all isotypes are monomers comprised of two heavy and two light chains: IgM and IgA polymerize only when they are secreted. In its membrane-bound form the immunoglobulin heavy chain has a hydrophobic transmembrane domain of about 25 amino acid residues at the carboxy terminus, which anchors it to the surface of the B lymphocyte. This transmembrane domain is absent from the secreted form, whose carboxy terminus is a hydrophilic secretory tail. The two different carboxy termini of the transmembrane and secreted forms of immunoglobulin heavy chains are encoded in separate exons and production of the two forms is achieved by alternative RNA processing (Fig. 4.21). The last two exons of each C_H gene contain the sequences encoding the secreted and the transmembrane regions respectively; if the primary transcript is cleaved and polyadenylated at a site downstream of these exons, the sequence encoding the carboxy terminus of the secreted form is removed by splicing and the cell-surface form of immunoglobulin is produced. Alternatively, if the primary transcript is cleaved at the polyadenylation site located before the last two exons, only the secreted molecule can be produced. This differential RNA processing is illustrated for C_μ in Fig. 4.21, but occurs in the same way for all isotypes.

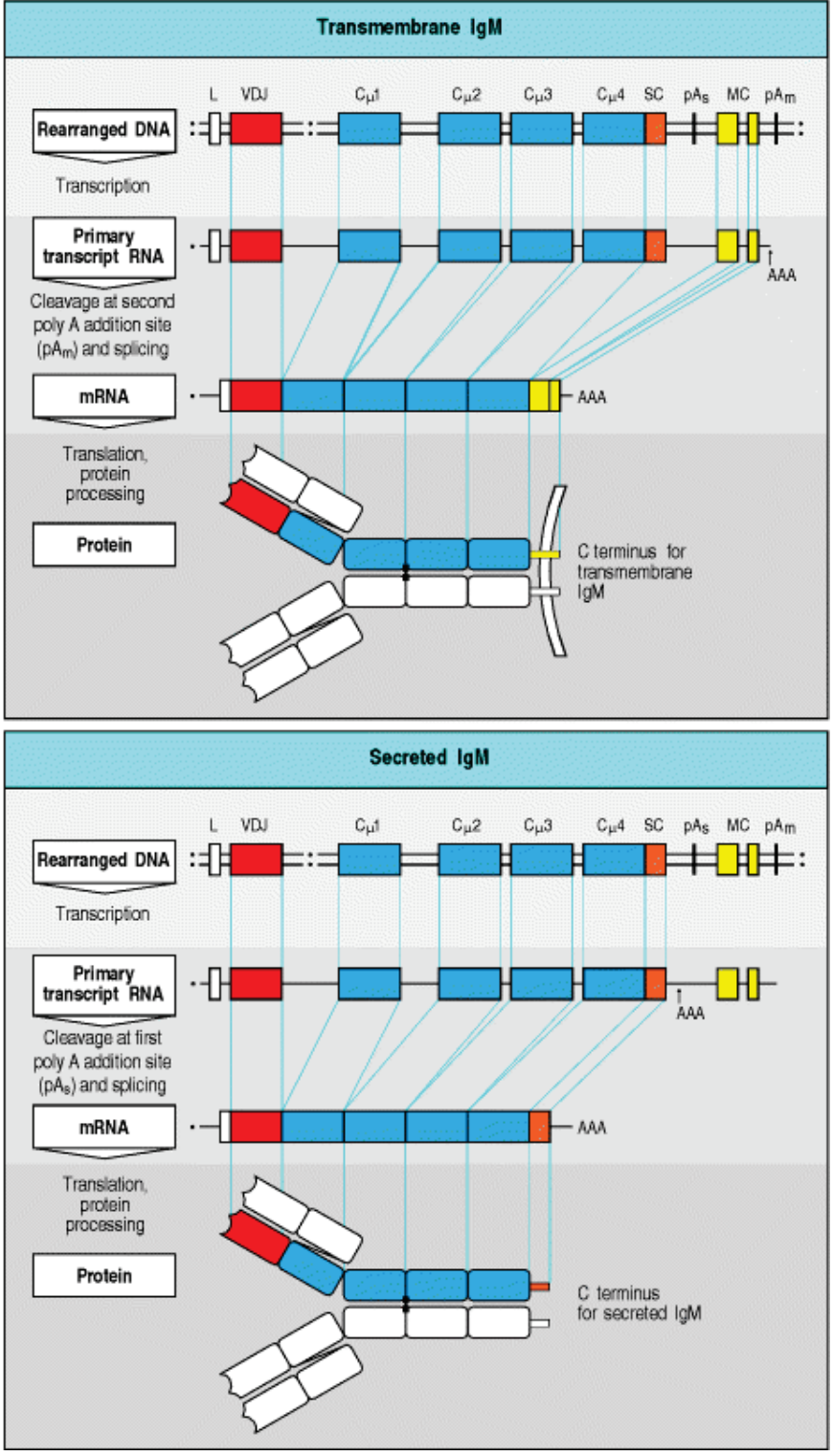


Figure 4.21. Transmembrane and secreted forms of immunoglobulins are derived from the same heavy-chain sequence by alternative RNA processing. Each heavy-chain C gene has two exons (membrane-coding (MC) yellow) that encode the transmembrane region and cytoplasmic tail of the transmembrane form, and a secretion-coding (SC) sequence (orange) that encodes the carboxy terminus of the secreted form. In the case of IgD, the SC sequence is present on a separate exon, but for the other isotypes, including IgM as shown here, the SC sequence is contiguous with the last C-domain exon. The events that dictate whether a heavy-chain RNA will result in a secreted or transmembrane immunoglobulin occur during processing of the initial transcript. Each heavy-chain C gene has two potential polyadenylation sites (shown as pA_S and pA_m). In the upper panel, the transcript is cleaved and polyadenylated (AAA) at the second site (pA_m). Splicing between a site located between the $C_\mu4$ exon and the SC sequence, and a second site at the 5′ end of the MC exons, results in removal of the SC sequence and joining of the MC exons to the $C_\mu4$ exon. This generates the transmembrane form of the heavy chain. In the lower panel, the primary transcript is cleaved and polyadenylated at the first site (pA_S), eliminating the MC exons and giving rise to the secreted form of the heavy chain.

Although the production of membrane-bound and secreted versions of the heavy chain is achieved by similar mechanisms to those that allow the co-expression of surface IgM and IgD (see Fig. 4.19), these two instances of alternative RNA processing act at different stages in the life of the B cell, and on different primary transcripts. B cells make a long heavy-chain transcript that can be processed to give either transmembrane IgM or IgD before they are stimulated by antigen. A B cell that is activated ceases to coexpress IgD with IgM, either because μ and δ sequences have been removed as a consequence of an isotype switch or, in IgM-secreting plasma cells, because transcription from the V_H promoter no longer extends through the C_δ exons. In activated B cells that differentiate to become antibody-secreting plasma cells, much of the transcript is spliced to the secreted rather than transmembrane form of whichever isotype the B cell happens to express.

4-18. Antibody C regions confer functional specialization.

The secreted antibodies protect the body in a variety of ways, as we briefly outline here and discuss further in Chapter 9. In some cases it is enough for the antibody simply to bind antigen. For instance, by binding tightly to a toxin or virus, an antibody can prevent it from recognizing its receptor on a host cell. The V regions on their own are sufficient for this. The C region is essential, however, for recruiting the help of other cells and molecules to destroy and dispose of pathogens to which the antibody has bound, and it confers functionally distinct properties on each of the various isotypes.

The C regions of antibodies have three main effector functions. First, the Fc portions of different isotypes are recognized by specialized receptors expressed by immune effector cells. The Fc portions of IgG1 and IgG3 antibodies are recognized by Fc receptors present on the surface of phagocytic cells such as macrophages and neutrophils, which can thereby bind and engulf pathogens coated with antibodies of these isotypes. The Fc portion of IgE binds to a high-affinity Fcε receptor on mast cells, basophils, and activated eosinophils, enabling these cells to respond to the binding of specific antigen by releasing inflammatory mediators. Second, the Fc portions of antigen: antibody complexes can bind to complement (see Fig. 1.24) and initiate the complement cascade, which helps to recruit and activate phagocytes, can aid the engulfment of microbes by phagocytes, and can also directly destroy pathogens. Third, the Fc portion can deliver antibodies to places they would not reach without active transport. These include the mucus secretions, tears, and milk (IgA), and the fetal blood circulation by transfer from the pregnant mother (IgG). In both cases, the Fc portion engages a specific receptor that leads to the active transport of the immunoglobulin through cells to reach different body compartments.

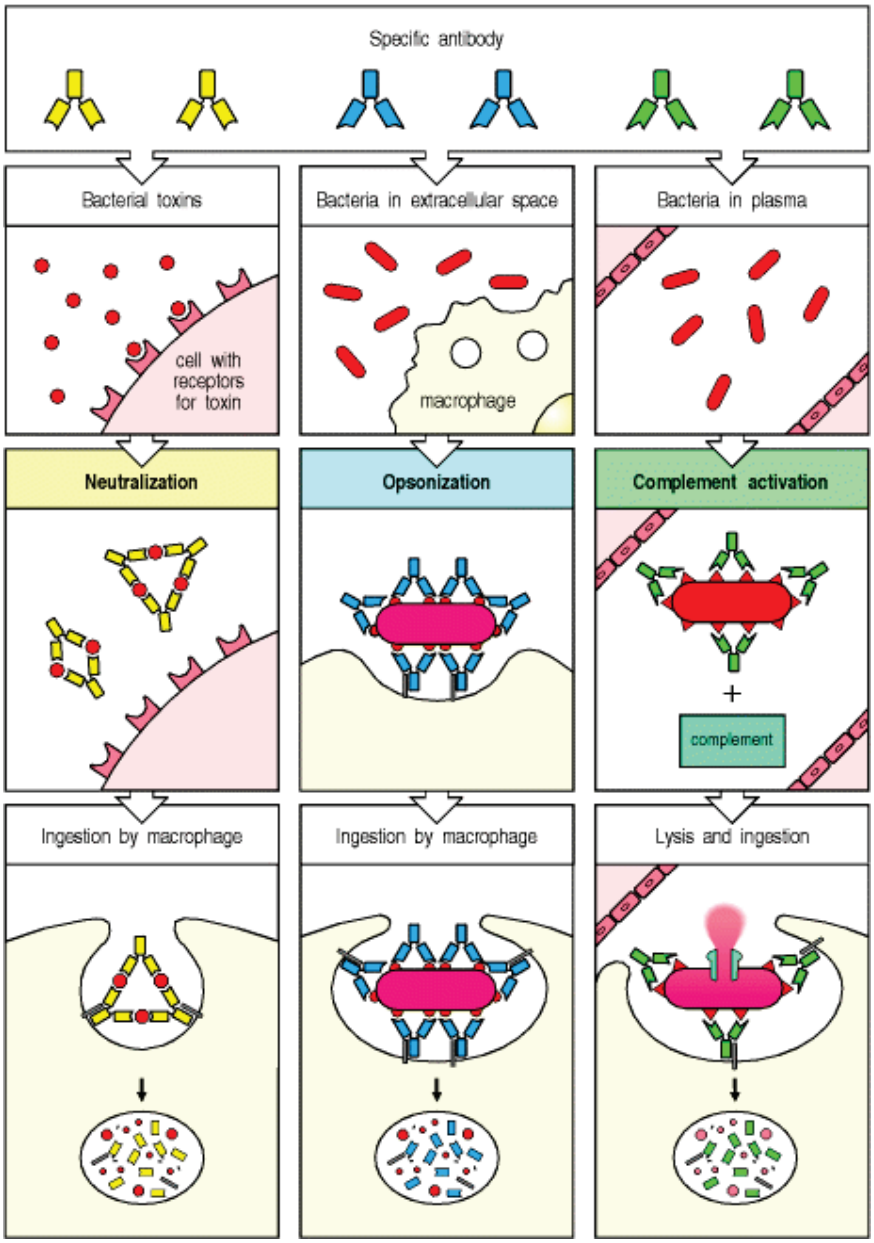


Figure 1.24. Antibodies can participate in host defense in three main ways. The left panels show antibodies binding to and neutralizing a bacterial toxin, thus preventing it from interacting with host cells and causing pathology. Unbound toxin can react with receptors on the host cell, whereas the toxin:antibody complex cannot. Antibodies also neutralize complete virus particles and bacterial cells by binding to them and inactivating them. The antigen:antibody complex is eventually scavenged and degraded by macrophages. Antibodies coating an antigen render it recognizable as foreign by phagocytes (macrophages and neutrophils), which then ingest and destroy it; this is called opsonization. The middle panels show opsonization and phagocytosis of a bacterial cell. The right panels show activation of the complement system by antibodies coating a bacterial cell. Bound antibodies form a receptor for the first protein of the complement system, which eventually forms a protein complex on the surface of the bacterium that, in some cases, can kill the bacterium directly. More generally, complement coating favors the uptake and destruction of the bacterium by phagocytes. Thus, antibodies target pathogens and their toxic products for disposal by phagocytes.

The role of the Fc portion in these effector functions can be demonstrated by studying enzymatically treated immunoglobulins that have had one or other domain of the Fc cleaved off (see Section 3-3) or, more recently, by genetic engineering, which permits detailed mapping of the exact amino acid residues within the Fc that are needed for particular functions. Many kinds of microorganism seem to have responded to the destructive potential of the Fc portion by manufacturing proteins that either bind to it or proteolytically cleave it, and so prevent the Fc region from working. Examples of these are Protein A and Protein G made by *Staphylococcus* species (Fig. 4.22), and Protein D of *Haemophilus* species. Researchers can exploit these proteins to help to map the Fc and as immunological reagents (see Appendix I, Section A-10). Not all immunoglobulin isotypes or subtypes have the same capacity to engage each

of the effector functions. The differential capabilities of each isotype are summarized in Fig. 4.16. For example, IgG1 and IgG3 have higher affinity for the most common type of Fc receptor.

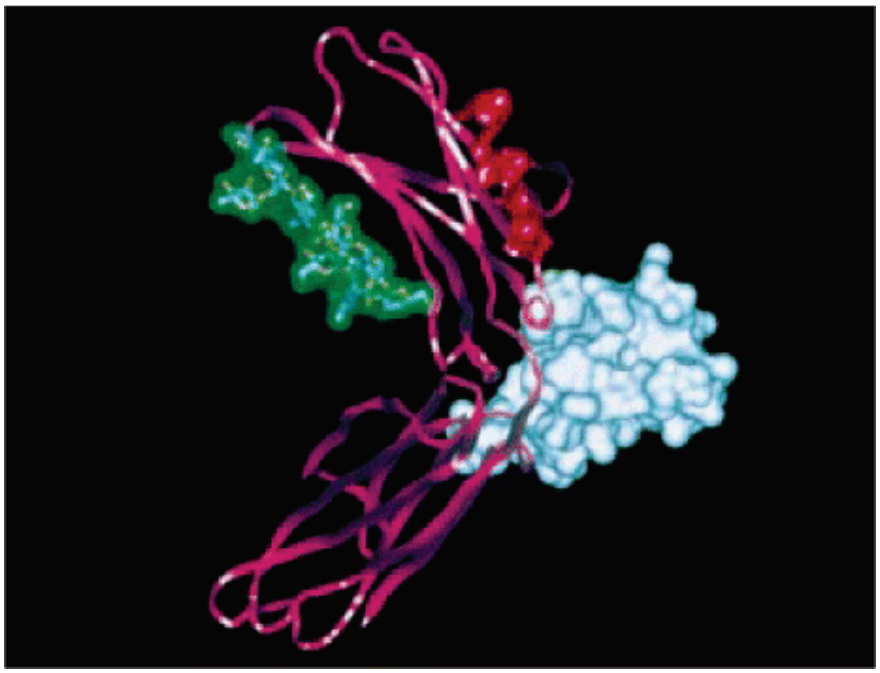


Figure 4.22. Protein A of *Staphylococcus aureus* bound to a fragment of the Fc region of IgG. A fragment of the Fc portion of a single IgG heavy chain is complexed with a fragment of the immunoglobulin-binding Protein A from *Staphylococcus aureus*. The Fc fragment has two domains, C_H2 and C_H3, shown in magenta. A carbohydrate chain is attached to an asparagine residue in the C_H2 domain: all the atoms are shown and the surface is outlined in green. The fragment of Protein A (white) is bound between the two domains of the Fc fragment. The amino acids that bind to the complement component C1q (red) lie in the C_H2 domain. Photograph courtesy of C. Thorpe.

4-19. IgM and IgA can form polymers.

Although all immunoglobulin molecules are constructed from a basic unit of two heavy and two light chains, both IgM and IgA can form multimers (Fig. 4.23). IgM and IgA C regions contain a 'tailpiece' of 18 amino acids that contains a cysteine residue essential for polymerization. An additional separate 15 kDa polypeptide chain called the J chain promotes polymerization by linking to the cysteines of the tailpiece, which is found only in the secreted forms of the μ and α chains. (This J chain should not be confused with the J gene segment; see Section 4-2.) In the case of IgA, polymerization is required for transport through epithelia, as we discuss in Chapter 9. IgM molecules are found as pentamers, and occasionally hexamers (without J chain), in plasma, whereas IgA in mucous secretions, but not in plasma, is mainly found as a dimer (see Fig. 4.23).

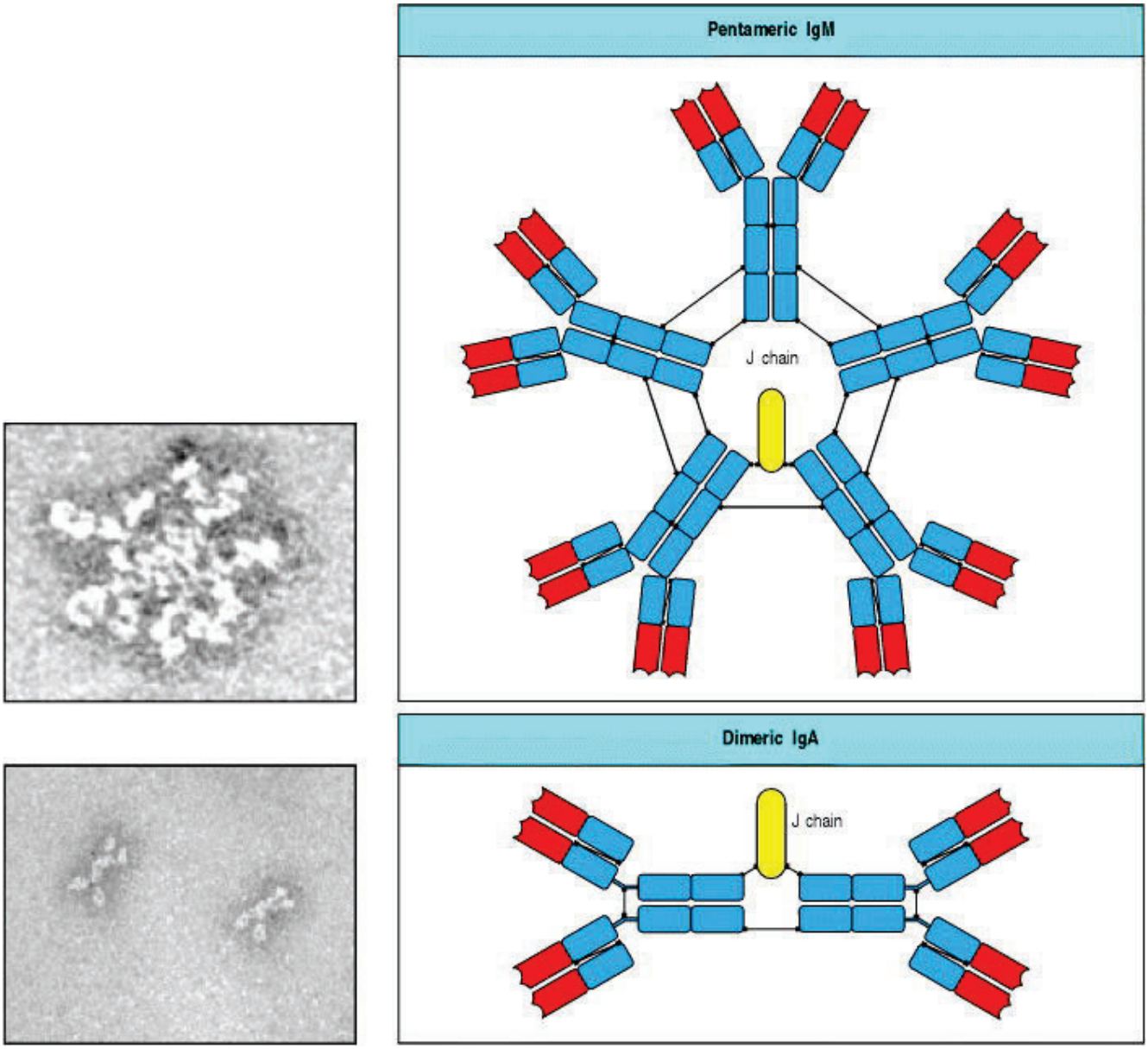


Figure 4.23. The IgM and IgA molecules can form multimers. IgM and IgA are usually synthesized as multimers in association with an additional polypeptide chain, the J chain. In pentameric IgM, the monomers are cross-linked by disulfide bonds to each other and to the J chain. The top left panel shows an electron micrograph of an IgM pentamer, showing the arrangement of the monomers in a flat disc. IgM can also form hexamers that lack a J chain but are more efficient in complement activation. In dimeric IgA, the monomers have disulfide bonds to the J chain as well as to each other. The bottom left panel shows an electron micrograph of dimeric IgA. Photographs (× 900,000) courtesy of K.H. Roux and J.M. Schiff.

The polymerization of immunoglobulin molecules is thought to be important in the binding of antibody to repetitive epitopes. The dissociation rate of an individual epitope from an individual antibody binding site influences the strength of binding, or affinity, of that site: the lower the dissociation rate, the higher the affinity (see Appendix I, Section A-9). An antibody molecule has two or more identical antigen-binding sites, and if it attaches to two or more repeating epitopes on a single target antigen, it will only dissociate when all sites dissociate. The dissociation rate of the whole antibody from the whole antigen will therefore be much slower than the rate for the individual binding sites, giving a greater effective total binding strength, or avidity. This consideration is particularly relevant for pentameric IgM, which has ten antigen-binding sites. IgM antibodies frequently recognize repetitive epitopes such as those expressed by bacterial cell-wall polysaccharides, but the binding of individual sites is often of low affinity because IgM is made early in immune responses, before somatic hypermutation and affinity maturation. Multisite binding makes up for this, dramatically improving the overall functional binding strength.

4-20. Various differences between immunoglobulins can be detected by antibodies.

When an immunoglobulin is used as an antigen, it will be treated like any other foreign protein and will elicit an antibody response. Anti-immunoglobulin antibodies can be made that recognize the amino acids that characterize the isotype of the injected antibody. Such anti-isotypic antibodies recognize all immunoglobulins of the same isotype in all members of the species from which the injected antibody came.

It is also possible to raise antibodies that recognize differences in immuno-globulins from members of the same species that are due to the presence of multiple alleles of the individual C genes in the population (genetic polymorphism). Such allelic variants are called allotypes. In contrast to anti-isotypic antibodies, anti-allotypic antibodies will recognize immuno-globulin of a particular isotype only in some members of a species. Finally, as individual antibodies differ in their V regions, one can raise antibodies against unique sequence variants, which are called idiotypes.

A schematic picture of the differences between idiotypes, allotypes, and isotypes is given in Fig. 4.24. Historically, the main features of immunoglobulins were defined by using isotypic and allotypic genetic markers identified by antisera raised in different species or in genetically distinct members of the same species (see Appendix I, Section A-10). The independent segregation of allotypic and isotypic markers revealed the existence of separate heavy-chain, κ, and λ genes.

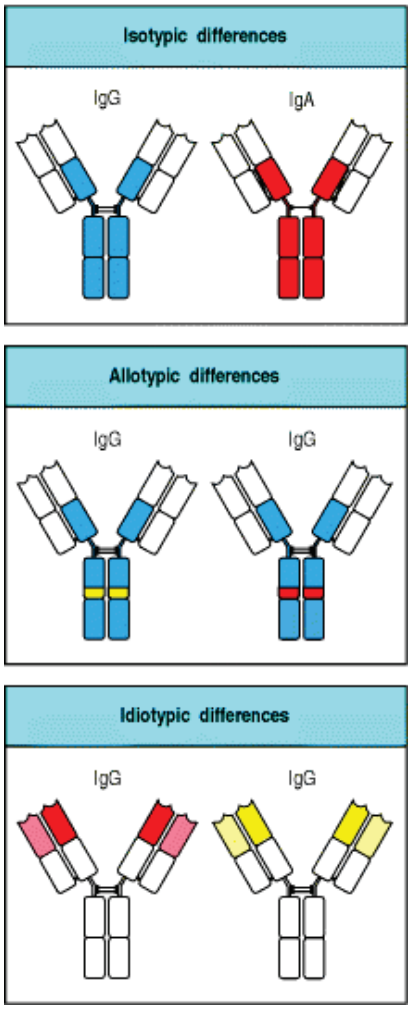


Figure 4.24. Different types of variation between immunoglobulins. Differences between constant regions due to usage of different C-region genes are called isotypes; differences due to different alleles of the same C gene are called allotypes; differences due to particular rearranged V_H and V_L genes are called idiotypes.

Summary.

The isotypes of immunoglobulins are defined by their heavy-chain C regions, each isotype being encoded by a separate C-region gene. The heavy-chain C-region genes lie in a cluster 3′ to the V-region gene segments. A productively rearranged V-region exon is initially expressed in association with μ and δ C_H regions, but the same V-region exon can subsequently be associated with any one of the other isotypes by the process of isotype switching, in

which the DNA is rearranged to place the V region 5′ to a different C-region gene. Unlike V(D)J recombination, isotype switching is always productive and occurs only in B cells activated by antigen. The immunological functions of the various isotypes differ; thus, isotype switching varies the response to the same antigen at different times or under different conditions. Immunoglobulin RNA can be processed in two different ways to produce either membrane-bound immunoglobulin, which acts as the B-cell receptor for antigen, or secreted antibody. In this way, the B-cell antigen receptor has the same specificity as the antibody that the B cell secretes upon activation.

Summary to Chapter 4.

Lymphocyte receptors are remarkably diverse, and developing B cells and T cells use the same basic mechanism to achieve this diversity. In each cell, functional genes for the immunoglobulin and T-cell receptor chains are assembled by somatic recombination from sets of separate gene segments that together encode the V region. The substrates for the joining process, arrays of V, D, and J gene segments, are similar among all the receptor loci, though there are some important differences in the details of their arrangement. The lymphoid-specific proteins, RAG-1 and RAG-2, direct the V(D)J recombination process in both T and B cells. These genes function in concert with other ubiquitous DNA-modifying enzymes and at least one other lymphoid-specific enzyme, TdT, to complete the joining process. As each type of gene segment is present in multiple, slightly different, versions, the random selection of gene segments for assembly is the source of enormous potential diversity. During the process of assembly, further diversity is introduced at the gene segment junctions through imprecise joining mechanisms. This diversity is concentrated in the DNA encoding the CDR3 loops of the receptors, which lie at the center of the antigen-binding site. The independent association of the two chains of immunoglobulins or T-cell receptors multiplies the overall diversity of the complete antigen receptor. In addition, mature B cells that are activated by antigen initiate a process of somatic point mutation of the V-region DNA, which creates numerous variants of the original assembled V region.

An important difference between immunoglobulins and T-cell receptors is that immunoglobulins exist in both membrane-bound (B-cell receptors) and secreted forms (antibodies). The ability to express both a secreted and membrane-bound form of the same molecule is due to differential splicing of mRNA to include exons that encode either the hydrophobic membrane anchor or the secreted tailpiece. Heavy-chain C regions contain three or four immunoglobulin domains, whereas the T-cell receptor chains have only one. Antibodies also have a variety of effector functions that are mediated by their C regions. Moreover, there are several alternative heavy-chain C regions for immunoglobulins, each with different effector functions. The same V region can be expressed along with different C regions through a process known as isotype switching. In this way, the progeny of a single B cell can express multiple different isotypes, thus maximizing the possible effector functions of any given antigen-specific antibody. The changes in immunoglobulin and T-cell receptor genes that occur during B-cell and T-cell development are summarized in Fig. 4.25.

			Process occurs in:	
Event	**Process**	**Nature of change**	**B cells**	**T cells**
V-region assembly	Somatic recombination of DNA	Irreversible	Yes	Yes
Junctional diversity	Imprecise joining, N-sequence insertion in DNA	Irreversible	Yes	Yes
Transcriptional activation	Activation of promoter by proximity to the enhancer	Irreversible but regulated	Yes	Yes
Switch recombination	Somatic recombination of DNA	Irreversible	Yes	No
Somatic hypermutation	DNA point mutation	Irreversible	Yes	No
IgM, IgD expression on surface	Differential splicing of RNA	Reversible, regulated	Yes	No
Membrane vs secreted form	Differential splicing of RNA	Reversible, regulated	Yes	No

Figure 4.25. Changes in immunoglobulin and T-cell receptor genes that occur during B-cell and T-cell development and differentiation. Those changes that establish immunological diversity are all irreversible, as they involve changes in B-cell or T-cell DNA. Certain changes in the organization of DNA or in its transcription are unique to B cells. Somatic hypermutation has not been observed in functional T-cell receptors. The B-cell-specific

processes, such as switch recombination, allow the same variable (V) region to be attached to several functionally distinct heavy-chain C regions, and thereby create functional diversity in an irreversible manner. By contrast, the expression of IgM versus IgD, and of membrane-bound versus secreted forms of all immunoglobulin types, can in principle be reversibly regulated.

References to Chapter 4

General references.

Casali, P., and Silberstein, L.E.S. (eds): Immunoglobulin gene expression in development and disease. *Ann. N.Y. Acad. Sci.* 1995, 76.

S.D. Fugmann, A.I. Lee, P.E. Shockett, I.J. Villey, and D.G. Schatz. The RAG proteins and V(D)J recombination: complexes, ends, and transposition *Annu. Rev. Immunol.* 2000. 18: 495-527. (PubMed)

S.D. Wagner and M.S. Neuberger. Somatic hypermutation of immunoglobulin genes *Annu. Rev. Immunol.* 1996. 14: 441-457. (PubMed)

Section references.

4-1 Immunoglobulin genes are rearranged in antibody-producing cells.

N. Hozumi and S. Tonegawa. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions *Proc. Natl. Acad. Sci. USA* 1976. 73: 3628-3632. (PubMed)

S. Tonegawa, C. Brack, N. Hozumi, and V. Pirrotta. Organization of immunoglobulin genes *Cold Spring Harbor Symp. Quant. Biol.* 1978. 42: 921-931. (PubMed)

T.A. Waldmann. The arrangement of immunoglobulin and T-cell receptor genes in human lymphoproliferative disorders *Adv. Immunol.* 1987. 40: 247-321. (PubMed)

4-2 The DNA sequence encoding a complete V region is generated by the somatic recombination of separate gene segments.

P. Early, H. Huang, M. Davis, K. Calame, and L. Hood. An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH *Cell* 1980. 19: 981-992. (PubMed)

S. Tonegawa, A.M. Maxam, R. Tizard, O. Bernard, and W. Gilbert. Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain *Proc. Natl. Acad. Sci. USA* 1978. 75: 1485-1489. (PubMed)

4-3 There are multiple different V-region gene segments.

G.P. Cook and I.M. Tomlinson. The human immunoglobulin V-H repertoire *Immunol. Today.* 1995. 16: 237-242. (PubMed)

R. Joho, I.L. Weissman, P. Early, J. Cole, and L. Hood. Organization of kappa light chain genes in germ-line and somatic tissue *Proc. Natl. Acad. Sci. USA* 1980. 77: 1106-1110. (PubMed)

R. Kofler, S. Geley, H. Kofler, and A. Helmberg. Mouse variable-region gene families—complexity, polymorphism, and use in nonautoimmune responses *Immunol. Rev.* 1992. 128: 5-21. (PubMed)

R. Maki, A. Traunecker, H. Sakano, W. Roeder, and S. Tonegawa. Exon shuffling generates an immunoglobulin heavy chain gene *Proc. Natl. Acad. Sci. USA* 1980. 77: 2138-2142. (PubMed)

F. Matsuda and T. Honjo. Organization of the human immunoglobulin heavy-chain locus *Adv. Immunol.* 1996. 62: 1-29. (PubMed)

4-4 Rearrangement of V, D, and J gene segments is guided by flanking DNA sequences.

U. Grawunder, R.B. West, and M.R. Lieber. Antigen receptor gene rearrangement *Curr. Opin. Immunol.* 1998. 10: 172-180. (PubMed)

E.E. Max, J.G. Seidman, and P. Leder. Sequences of five potential recombination sites encoded close to an immunoglobulin kappa constant region gene *Proc. Natl. Acad. Sci. USA* 1979. 76: 3450-3454. (PubMed)

M. Muramatsu, K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai, and T. Honjo. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme *Cell* 2000. 102: 553-563. (PubMed)

H. Sakano, K. Huppi, G. Heinrich, and S. Tonegawa. Sequences at the somatic recombination sites of immunoglobulin light-chain genes *Nature* 1979. 280: 288-294. (PubMed)

4-5 The reaction that recombines V, D, and J gene segments involves both lymphocyte-specific and ubiquitous DNA-modifying enzymes.

A. Agrawal and D.G. Schatz. RAG1 and RAG2 form a stable postcleavage synaptic complex with DNA containing signal ends in V(D)J recombination *Cell* 1997. 89: 43-53. (PubMed)

T. Blunt, N.J. Finnie, G.E. Taccioli, G.C.M. Smith, J. Demengeot, T.M. Gottlieb, R. Mizuta, A.J. Varghese, F.W. Alt, P.A. Jeggo, and S.P. Jackson. Defective DNA-dependent protein kinase activity is linked to V(D)J recombination and DNA-repair defects associated with the murine-scid mutation *Cell* 1995. 80: 813-823. (PubMed)

Z. Gu, S. Jin, Y. Gao, D.T. Weaver, and F.W. Alt. Ku70-deficient embryonic stem cells have increased ionizing radiosensitivity, defective DNA end-binding activity, and inability to support V(D)J recombination *Proc. Natl. Acad. Sci. USA* 1997. 94: 8076-8081. (PubMed) (Full Text in PMC)

Z.Y. Li, T. Otevrel, Y.J. Gao, H.L. Cheng, B. Seed, T.D. Stamato, G.E. Taccioli, and F.W. Alt. The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination *Cell* 1995. 83: 1079-1089. (PubMed)

M.A. Oettinger, D.G. Schatz, C. Gorka, and D. Baltimore. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination *Science* 1990. 248: 1517-1523. (PubMed)

P.E. Shockett and D.G. Schatz. DNA hairpin opening mediated by the RAG1 and RAG2 proteins *Mol. Cell. Biol.* 1999. 19: 4159-4166. (PubMed) (Full Text in PMC)

4-6 The diversity of the immunoglobulin repertoire is generated by four main processes.

L.J. Fanning, A.M. Connor, and G.E. Wu. Development of the immunoglobulin repertoire *Clin. Immunol. Immunopathol.* 1996. 79: 1-14. (PubMed)

M. Weigert, R. Perry, D. Kelley, T. Hunkapiller, J. Schilling, and L. Hood. The joining of V and J gene segments creates antibody diversity *Nature* 1980. 283: 497-499. (PubMed)

4-7 The multiple inherited gene segments are used in different combinations.

A. Lee, S. Desravines, and E. Hsu. IgH diversity in an individual with only one million B lymphocytes *Dev. Immunol.* 1993. 3: 211-222. (PubMed)

4-8 Variable addition and subtraction of nucleotides at the junctions between gene segments contributes to diversity in the third hypervariable region.

G.H. Gauss and M.R. Lieber. Mechanistic constraints on diversity in human V(D)J recombination *Mol. Cell. Biol.* 1996. 16: 258-269. (PubMed)

T. Komori, A. Okada, V. Stewart, and F.W. Alt. Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes [published erratum appears in *Science* 1993, 262:1957] *Science* 1993. 261: 1171-1175.

(PubMed)

M. Weigert, L. Gatmaitan, E. Loh, J. Schilling, and L. Hood. Rearrangement of genetic information may produce immunoglobulin diversity *Nature* 1978. 276: 785-790. (PubMed)

4-9 Rearranged V genes are further diversified by somatic hypermutation.

A.G. Betz, C. Rada, R. Pannell, C. Milstein, and M.S. Neuberger. Passenger transgenes reveal intrinsic specificity of the antibody hypermutation mechanism: clustering, polarity, and specific hot spots *Proc. Natl. Acad. Sci. USA* 1993. 90: 2385-2388. (PubMed) (Full Text in PMC)

D. McKean, K. Huppi, M. Bell, L. Straudt, W. Gerhard, and M. Weigert. Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin *Proc. Natl. Acad. Sci. USA* 1984. 81: 3180-3184. (PubMed)

M.S. Neuberger, M.R. Ehrenstein, N. Klix, C.J. Jolly, J. Yelamos, C. Rada, and C. Milstein. Monitoring and interpreting the intrinsic features of somatic hypermutation *Immunol. Rev.* 1998. 162: 107-116. (PubMed)

Papavasiliou, F.N., and Schatz, D.G.: Cell cycle regulated DNA double strand breaks in somatic hypermutation of immunoglobulin genes. *Nature* , in press.

J.E. Sale and M.S. Neuberger. TdT-accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line *Immunity* 1998. 9: 859-869. (PubMed)

I.M. Tomlinson, G. Walter, P.T. Jones, P.H. Dear, E.L.L. Sonnhammer, and G. Winter. The imprint of somatic hypermutation on the repertoire of human germline V genes *J. Mol. Biol.* 1996. 256: 813-817. (PubMed)

M.G. Weigert, I.M. Cesari, S.J. Yonkovich, and M. Cohn. Variability in the lambda light chain sequences of mouse antibody *Nature* 1970. 228: 1045-1047. (PubMed)

4-10 In some species most immunoglobulin gene diversification occurs after gene rearrangement.

K.L. Knight and M.A. Crane. Generating the antibody repertoire in rabbit *Adv. Immunol.* 1994. 56: 179-218. (PubMed)

C.A. Reynaud, B. Bertocci, A. Dahan, and J.C. Weill. Formation of the chicken B-cell repertoire—ontogeny, regulation of Ig gene rearrangement, and diversification by gene conversion *Adv. Immunol.* 1994. 57: 353-378. (PubMed)

C.A. Reynaud, C. Garcia, W.R. Hein, and J.C. Weill. Hypermutation generating the sheep immunoglobulin repertoire is an antigen independent process *Cell* 1995. 80: 115-125. (PubMed)

M. Vajdy, P. Sethupathi, and K.L. Knight. Dependence of antibody somatic diversification on gut-associated lymphoid tissue in rabbits *J. Immunol.* 1998. 160: 2725-2729. (PubMed)

4-11 The T-cell receptor loci comprise sets of gene segments and are rearranged by the same enzymes as the immunoglobulin loci.

L. Rowen, B.F. Koop, and L. Hood. The complete 685-kilobase DNA sequence of the human β T cell receptor locus *Science* 1996. 272: 1755-1762. (PubMed)

Y. Shinkai, G. Rathbun, K.P. Lam, E.M. Oltz, V. Stewart, M. Mendelsohn, J. Charron, M. Datta, F. Young, A.M. Stall, and F.W. Alt. RAG-2 deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement *Cell* 1992. 68: 855-867. (PubMed)

4-12 T-cell receptors concentrate diversity in the third hypervariable region.

M.M. Davis and P.J. Bjorkman. T-cell antigen receptor genes and T-cell recognition [published erratum appears in *Nature* 1988, 335:744] *Nature* 1988. 334: 395-402. (PubMed)

J.L. Jorgensen, U. Esser, B. Fazekas de St. Groth, P.A. Reay, and M.M. Davis. Mapping T-cell receptor-peptide contacts by variant peptide immunization of single-chain transgenics *Nature* 1992. 355: 224-230. (PubMed)

4-13 γ:δ T-cell receptors are also generated by gene rearrangement.

Y.H. Chien, M. washima, K.B. Kaplan, J.F. Elliott, and M.M. Davis. A new T-cell receptor gene located within the alpha locus and expressed early in T-cell differentiation *Nature* 1987. 327: 677-682. (PubMed)

A.C. Hayday, H. Saito, S.D. Gillies, D.M. Kranz, G. Tanigawa, H.N. Eisen, and S. Tonegawa. Structure, organization, and somatic rearrangement of T cell gamma genes *Cell* 1985. 40: 259-269. (PubMed)

J.J. Lafaille, A. DeCloux, M. Bonneville, Y. Takagaki, and S. Tonegawa. Junctional sequences of T cell receptor gamma delta genes: implications for gamma delta T cell lineages and for a novel intermediate of V-(D)-J joining *Cell* 1989. 59: 859-870. (PubMed)

S. Tonegawa, A. Berns, M. Bonneville, A.G. Farr, I. Ishida, K. Ito, S. Itohara, C.A. Janeway Jr, O. Kanagawa, and R. Kubo, *et al.* Diversity, development, ligands, and probable functions of gamma delta T cells *Adv. Exp. Med. Biol.* 1991. 292: 53-61. (PubMed)

4-14 Somatic hypermutation does not generate diversity in T-cell receptors.

B. Zheng, W. Xue, and G. Kelsoe. Locus-specific somatic hypermutation in germinal centre T cells *Nature* 1994. 372: 556-559. (PubMed)

4-15 The immunoglobulin heavy-chain isotypes are distinguished by the structure of their constant regions.

D.R. Davies and H. Metzger. Structural basis of antibody function *Annu. Rev. Immunol.* 1983. 1: 87-117. (PubMed)

4-16 The same V_H exon can associate with different C_H genes in the course of an immune response.

P. Revy, T. Muto, Y. Levy, F. Geissmann, A. Plebani, O. Sanal, N. Catalan, M. Forveille, R. Dufourcq-Lagelouse, A. Gennery, I. Tezcan, F. Ersoy, H. Kayserili, A.G. Ugazio, N. Brousse, M. Muramatsu, L.D. Notarangelo, K. Kinoshita, T. Honjo, A. Fischer, and A. Durandy. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the hyper-IgM syndrome (HIGM2) *Cell* 2000. 102: 565-575. (PubMed)

H. Sakano, R. Maki, Y. Kurosawa, W. Roeder, and S. Tonegawa. Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes *Nature* 1980. 286: 676-683. (PubMed)

J. Stavnezer. Immunoglobulin class switching *Curr. Opin. Immunol.* 1996. 8: 199-205. (PubMed)

4-17 Transmembrane and secreted forms of immunoglobulin are generated from alternative heavy-chain transcripts.

P. Early, J. Rogers, M. Davis, K. Calame, M. Bond, R. Wall, and L. Hood. Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways *Cell* 1980. 20: 313-319. (PubMed)

M.L. Peterson, E.R. Gimmi, and R.P. Perry. The developmentally regulated shift from membrane to secreted mu mRNA production is accompanied by an increase in cleavage-polyadenylation efficiency but no measurable change in splicing efficiency *Mol. Cell. Biol.* 1991. 11: 2324-2327. (PubMed)

J. Rogers, P. Early, C. Carter, K. Calame, M. Bond, L. Hood, and R. Wall. Two mRNAs with different 3′ ends encode membrane-bound and secreted forms of immunoglobulin mu chain *Cell* 1980. 20: 303-312. (PubMed)

4-18 Antibody C regions confer functional specialization.

B.A. Helm, I. Sayers, A. Higginbottom, D.C. Machado, Y. Ling, K. Ahmad, E.A. Padlan, and A.P.M. Wilson. Identification of the high affinity receptor binding region in human IgE *J. Biol. Chem.* 1996. 271: 7494-7500. (PubMed)

R. Jefferis, J. Lund, and M. Goodall. Recognition sites on human IgG for Fcg receptors—the role of glycosylation

Immunol. Lett. 1995. 44: 111-117. (PubMed)

Sensel, M.G., Kane, L.M., and Morrison, S.L.: Amino acid differences in the N-terminus of CH2 influence the relative abilities of IgG2 and IgG3 to activate complement. *Mol. Immunol.* 3:1019-1029.

4-19 IgM and IgA can form polymers.

B.A. Hendrickson, D.A. Conner, D.J. Ladd, D. Kendall, J.E. Casanova, B. Corthesy, E.E. Max, M.R. Neutra, C.E. Seidman, and J.G. Seidman. Altered hepatic transport of IgA in mice lacking the J chain *J. Exp. Med.* 1995. 182: 1905-1911. (PubMed)

M.J. Niles, L. Matsuuchi, and M.E. Koshland. Polymer IgM assembly and secretion in lymphoid and nonlymphoid cell-lines—evidence that J chain is required for pentamer IgM synthesis *Proc. Nat. Acad. Sci. USA* 1995. 92: 2884-2888. (PubMed) (Full Text in PMC)

4-20 Various differences between immunoglobulins can be detected by antibodies.

J.D. Capra and J.M. Kehoe. Structure of antibodies with shared idiotypy: the complete sequence of the heavy chain variable regions of two immunoglobulin M anti-gamma globulins *Proc. Natl. Acad. Sci. USA* 1974. 71: 4032-4036. (PubMed)

B.A. Fields, F.A. Goldbaum, X. Ysern, R.J. Poljak, and R.A. Mariuzza. Molecular basis of antigen mimicry by an anti idiotope *Nature* 1995. 374: 739-742. (PubMed)

T.T. Wu, E.A. Kabat, and H. Bilofsky. Similarities among hypervariable segments of immunoglobulin chains *Proc. Natl. Acad. Sci. USA* 1975. 72: 5107-5110. (PubMed)