IMMUNE DIVERSITY

5.25.1 Introduction

Key Terms

An **immune response** is an organism's reaction, mediated by components of the immune system, to an antigen.

An **antigen** is any foreign substance whose entry into an organism provokes an immune response by stimulating the synthesis of an antibody (an immunoglobulin protein that can bind to the antigen).

A **B cell** is a lymphocyte that produces antibodies. B cells development occurs primarily in bone marrow.

T cells are lymphocytes of the T (thymic) lineage; may be subdivided into several functional types. They carry TcR (T-cell receptor) and are involved in the cell-mediated immune response.

The **humoral response** is an immune response that is mediated primarily by antibodies. It is defined as immunity that can be transferred from one organism to another by serum antibody.

An **immunoglobulin (Antibody)** is a class of protein that is produced by B cells in response to antigen.

An **antibody** is a protein (immunoglobulin) produced by B lymphocyte cells that recognizes a particular 'foreign antigen', and thus triggers the immune response.

A **helper T cell** is a T lymphocyte that activates macrophages and stimulates B cell proliferation and antibody production. Helper T cells usually express cell surface CD4 but not CD8.

Two mutants are said to **complement** each other when a diploid that is heterozygous for each mutation produces the wild type phenotype.

The **cell-mediated response** is the immune response that is mediated primarily by T lymphocytes. It is defined based on immunity that cannot be transferred from one organism to another by serum antibody.

A **cytotoxic T cell** is a T lymphocyte (usually $CD8^+$) that can be stimulated to kill cells containing intracellular pathogens, such as viruses.

The **T cell receptor (TCR)** is the antigen receptor on T lymphocytes. It is clonally expressed and binds to a complex of MHC class I or class II protein and antigen-derived peptide.

The **major histocompatibility complex (MHC)** is a chromosomal region containing genes that are involved in the immune response. The genes encode proteins for antigen presentation, cytokines, and complement proteins. The MHC is highly polymorphic.

Tolerance is the lack of an immune response to an antigen (either self antigen or foreign antigen) due to clonal deletion.

An **autoimmune disease** is a pathological condition in which the immune response is directed to self antigen.

Clonal deletion describes the elimination of a clonal population of lymphocytes. At certain stages of lymphocyte development, clonal deletion can be induced when lymphocyte antigen receptors bind to their cognate antigen.

A **superfamily** is a set of genes all related by presumed descent from a common ancestor, but now showing considerable variation.

It is an axiom of genetics that the genetic constitution created in the zygote by the combination of sperm and egg is inherited by all somatic cells of the organism. We look to differential control of gene expression, rather than to changes in DNA content, to explain the different phenotypes of particular somatic cells.

Yet there are exceptional situations in which the reorganization of certain DNA sequences is used to regulate gene expression or to create new genes. The immune system provides a striking and extensive case in which the content of the genome changes, when recombination creates active genes in lymphocytes. Other cases are represented by the substitution of one sequence for another to change the mating type of yeast or to generate new surface antigens by trypanosomes (see *Molecular Biology 4.18 Rearrangement of DNA*).

The **immune response** of vertebrates provides a protective system that distinguishes foreign proteins from the proteins of the organism itself. Foreign material (or part of the foreign material) is recognized as comprising an **antigen**. Usually the antigen is a protein (or protein-attached moiety) that has entered the bloodstream of the animal – for example, the coat protein of an infecting virus. Exposure to an antigen initiates production of an immune response that *specifically recognizes the antigen and destroys it.*

Immune reactions are the responsibility of white blood cells – the B and T lymphocytes, and macrophages. The lymphocytes are named after the tissues that produce them. In mammals, **B cells** mature in the bone marrow, while **T cells** mature in the thymus. *Each class of lymphocyte uses the rearrangement of DNA as a mechanism for producing the proteins that enable it to participate in the immune response.*

The immune system has many ways to destroy an antigenic invader, but it is useful to consider them in two general classes. Which type of response the immune system mounts when it encounters a foreign structure depends partly on the nature of the antigen. The response is defined according to whether it is executed principally by B cells or T cells.

The **humoral response** depends on B cells. It is mediated by the secretion of antibodies, which are **immunoglobulin** proteins. *Production of an **antibody** specific for a foreign molecule is the primary event responsible for recognition of an antigen.* Recognition requires the antibody to bind to a small region or structure on the antigen.

The function of antibodies is represented in **Figure 25.1**. Foreign material circulating in the bloodstream, for example, a toxin or pathogenic bacterium, has a surface that presents antigens. The antigen(s) are recognized by the antibodies, which form an

antigen-antibody complex. This complex then attracts the attention of other components of the immune system.

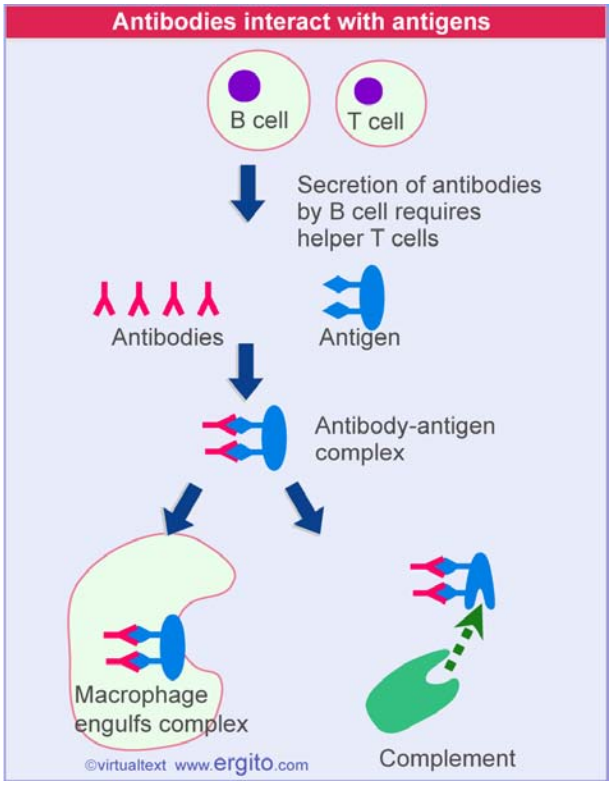

Figure 25.1 Humoral immunity is conferred by the binding of free antibodies to antigens to form antigen-antibody complexes that are removed from the bloodstream by macrophages or that are attacked directly by the complement proteins.

The humoral response depends on these other components in two ways. First, B cells need signals provided by T cells to enable them to secrete antibodies. These T cells are called **helper T cells**, because they assist the B cells. Second, antigen-antibody formation is a trigger for the antigen to be destroyed. The major pathway is provided by the action of **complement**, a component whose name reflects its ability to "complement" the action of the antibody itself. Complement consists of a set of ~20 proteins that function through a cascade of proteolytic actions. If the target antigen is part of a cell, for example, an infecting bacterium, the action of complement culminates in lysing the target cell. The action of complement also provides a means of attracting macrophages, which scavenge the target cells or their products. Alternatively, the antigen-antibody complex may be taken up directly by macrophages (scavenger cells) and destroyed.

The **cell-mediated response** is executed by a class of T lymphocytes called **cytotoxic T cells** (also called killer T cells). The basic function of the T cell in recognizing a target antigen is indicated in **Figure 25.2**. A cell-mediated response typically is elicited by an intracellular parasite, such as a virus that infects the body's own cells. As a result of the viral infection, fragments of foreign (viral) antigens are displayed on the surface of the cell. These fragments are recognized by the **T cell receptor** (**TCR**), which is the T cells' equivalent of the antibody produced by a B

cell.

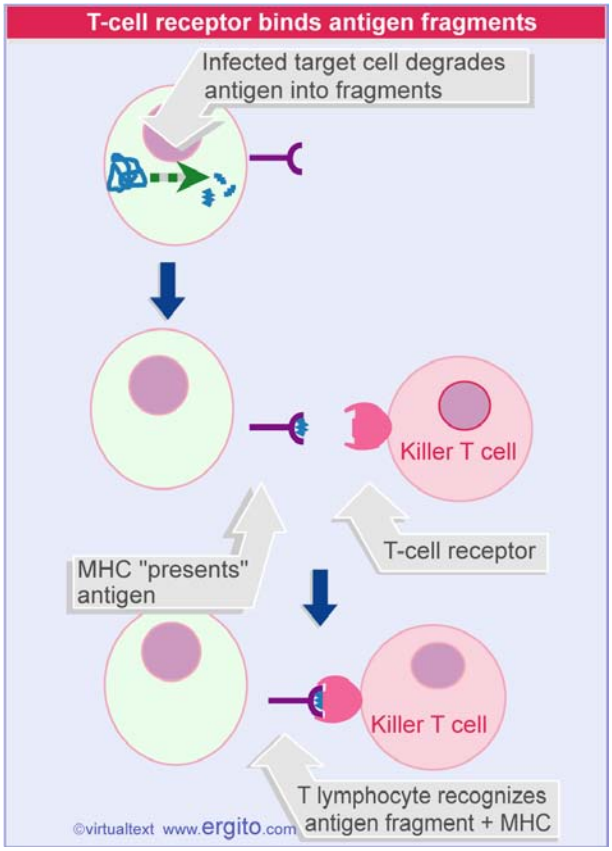

Figure 25.2 In cell-mediated immunity, killer T cells use the T-cell receptor to recognize a fragment of the foreign antigen which is presented on the surface of the target cell by the MHC protein.

A crucial feature of this recognition reaction is that *the antigen must be* presented *by a cellular protein that is a member of the* **MHC** (**major histocompatibility complex**). The MHC protein has a groove on its surface that binds a peptide fragment derived from the foreign antigen. The combination of peptide fragment and MHC protein is recognized by the T cell receptor. Every individual has a characteristic set of MHC proteins. They are important in graft reactions; a graft of tissue from one individual to another is rejected because of the difference in MHC proteins between the donor and recipient, an issue of major medical importance. The demand that the T lymphocytes recognize both foreign antigen and MHC protein ensures that the cell-mediated response acts only on host cells that have been infected with a foreign antigen. (We discuss the division of MHC proteins into the general types of class I and class II later in *Molecular Biology 5.25.20 The major histocompatibility locus codes for many genes of the immune system*.)

The purpose of each type of immune response is to attack a foreign target. Target recognition is the prerogative of B-cell immunoglobulins and T cell receptors. A crucial aspect of their function lies in the ability to distinguish "self" from "nonself." Proteins and cells of the body itself must *never* be attacked. Foreign targets must be *destroyed entirely*. The property of failing to attack "self" is called **tolerance**. Loss of

this ability results in an **autoimmune disease**, in which the immune system attacks its own body, often with disastrous consequences.

What prevents the lymphocyte pool from responding to "self" proteins? Tolerance probably arises early in lymphocyte cell development when B and T cells that recognize "self" antigens are destroyed. This is called **clonal deletion**. In addition to this negative selection, there is also positive selection for T cells carrying certain sets of T cell receptors.

A corollary of tolerance is that it can be difficult to obtain antibodies against proteins that are closely related to those of the organism itself. As a practical matter, therefore, it may be difficult to use (for example) mice or rabbits to obtain antibodies against human proteins that have been highly conserved in mammalian evolution. The tolerance of the mouse or rabbit for its own protein may extend to the human protein in such cases.

Each of the three groups of proteins required for the immune response – immunoglobulins, T cell receptors, MHC proteins – is diverse. Examining a large number of individuals, we find many variants of each protein. Each protein is coded by a large family of genes; and in the case of antibodies and the T cell receptors, the diversity of the population is increased by DNA rearrangements that occur in the relevant lymphocytes.

Immunoglobulins and T cell receptors are direct counterparts, each produced by its own type of lymphocyte. The proteins are related in structure, and their genes are related in organization. The sources of variability are similar. The MHC proteins also share some common features with the antibodies, as do other lymphocyte-specific proteins. In dealing with the genetic organization of the immune system, we are therefore concerned with a series of related gene families, indeed a **superfamily** that may have evolved from some common ancestor representing a primitive immune response.

IMMUNE DIVERSITY

5.25.2 Clonal selection amplifies lymphocytes that respond to individual antigens

Key Terms

The **clonal selection** theory proposed that each lymphocyte expresses a single antigen receptor specificity and that only those lymphocytes that bind to a given antigen are stimulated to proliferate and to function in eliminating that antigen. Thus, the antigen "selects" the lymphocytes to be activated. Clonal selection is now an established principle in immunology.

The **primary immune response** is an organism's immune response upon first exposure to a given antigen. It is characterized by a relatively shorter duration and lower affinity antibodies than in the secondary immune response.

A **memory cell** is a lymphocyte that has been stimulated during the primary immune response to antigen and that is rapidly activated upon subsequent exposure to that antigen. Memory cells respond more rapidly to antigen than naive cells.

The **secondary immune response** is an organism's immune response upon a second exposure to a given antigen. This second exposure is also referred to as a "booster". The secondary immune response is characterized by a more rapid induction, greater magnitude, and higher affinity antibodies than the primary immune response.

A **hapten** is a small molecule that acts as an antigen when conjugated to a protein.

An **antigenic determinant** is the portion of an antigen that is recognized by the antigen receptor on lymphocytes. It is also called an epitope.

An **epitope** is the portion of an antigen that is recognized by the antigen receptor on lymphocytes. It is also called an antigenic determinant.

Key Concepts

- Each B lymphocyte expresses a single immunoglobulin and each T lymphocyte expresses a single T cell receptor.
- There is a very large variety of immunoglobulins and T cell receptors.
- Antigen binding to an immunoglobulin or T cell receptor triggers clonal multiplication of the cell.

The name of the immune response describes one of its central features. After an organism has been exposed to an antigen, it becomes *immune* to the effects of a new infection. Before exposure to a particular antigen, the organism lacks adequate capacity to deal with any toxic effects. This ability is acquired during the immune response. After the infection has been defeated, the organism retains the ability to respond rapidly in the event of a re-infection.

These features are accommodated by the **clonal selection** theory illustrated in **Figure**

25.3. The pool of lymphocytes contains B cells and T cells carrying a large variety of immunoglobulins or T cell receptors. *But any individual B lymphocyte produces one immunoglobulin, which is capable of recognizing only a single antigen; similarly any individual T lymphocyte produces only one particular T cell receptor*.

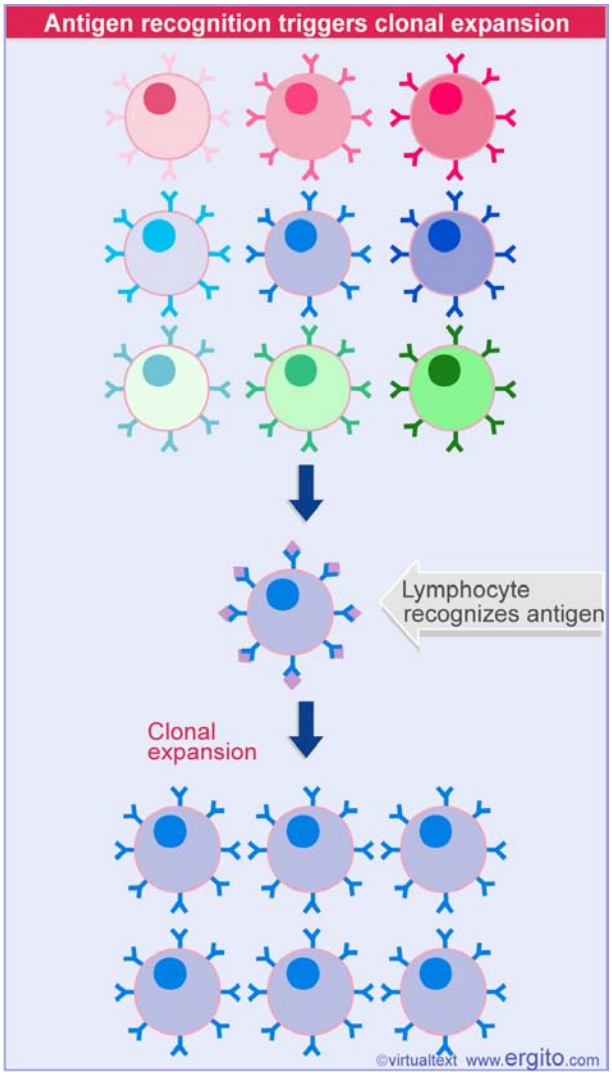

Figure 25.3 The pool of immature lymphocytes contains B cells and T cells making antibodies and receptors with a variety of specificities. Reaction with an antigen leads to clonal expansion of the lymphocyte with the antibody (B cell) or receptor (T cell) that can recognize the antigen.

In the pool of immature lymphocytes, the unstimulated B cells and T cells are morphologically indistinguishable. But on exposure to antigen, a B cell whose antibody is able to bind the antigen, or a T cell whose receptor can recognize it, is stimulated to divide, probably by some feedback from the surface of the cell, where the antibody/receptor-antigen reaction occurs. The stimulated cells then develop into mature B or T lymphocytes, which includes morphological changes involving (for example) an increase in cell size (especially pronounced for B cells).

The initial expansion of a specific B- or T cell population upon first exposure to an antigen is called the **primary immune response**. Large numbers of B or T lymphocytes with specificity for the target antigen are produced. Each population represents a clone of the original responding cell. Antibody is secreted from the B cells in large quantities, and it may even come to dominate the antibody population.

After a successful primary immune response has been mounted, the organism retains B cells and T cells carrying the corresponding antibody or receptor. These **memory cells** represent an intermediate state between the immature cell and the mature cell. They have not acquired all of the features of the mature cell, but they are long-lived, and can rapidly be converted to mature cells. Their presence allows a **secondary immune response** to be mounted rapidly if the animal is exposed to the same antigen again.

The pool of immature lymphocytes in a mammal contains ~10^{12} cells. This pool contains some lymphocytes that have unique specificities (because a corresponding antigen has never been encountered), while others are represented by up to 10^6 cells (because clonal selection has expanded the pool to respond to an antigen).

What features are recognized in an antigen? Antigens are usually macromolecular. Although small molecules may have antigenic determinants and can be recognized by antibodies, usually they are not effective in provoking an immune response (because of their small size). But they do provoke a response when conjugated with a larger carrier molecule (usually a protein). A small molecule that is used to provoke a response by such means is called a **hapten**.

Only a small part of the surface of a macromolecular antigen is actually recognized by any one antibody. The binding site consists of only 5-6 amino acids. Of course, any particular protein may have more than one such binding site, in which case it provokes antibodies with specificities for different regions. The region provoking a response is called an **antigenic determinant** or **epitope**. When an antigen contains several epitopes, some may be more effective than others in provoking the immune response; in fact, they may be so effective that they entirely dominate the response.

How do lymphocytes find target antigens and where does their maturation take place? Lymphocytes are peripatetic cells. They develop from immature stem cells that are located in the adult bone marrow. They migrate to the peripheral lymphoid tissues (spleen, lymph nodes) either directly via the bloodstream (if they are B cells) or via the thymus (where they become T cells). The lymphocytes recirculate between blood and lymph; the process of dispersion ensures that an antigen will be exposed to lymphocytes of all possible specificities. When a lymphocyte encounters an antigen that binds its antibody or receptor, clonal expansion begins the immune response.

IMMUNE DIVERSITY

5.25.3 Immunoglobulin genes are assembled from their parts in lymphocytes

Key Terms

The immunoglobulin **light chain** is one of two types of subunits in an antibody tetramer. Each antibody contains two light chains. The N-terminus of the light chain forms part of the antigen recognition site.

The immunoglobulin **heavy chain** is one of two types of subunits in an antibody tetramer. Each antibody contains two heavy chains. The N-terminus of the heavy chain forms part of the antigen recognition site, whereas the C-terminus determines the subclass (isotype).

The **variable region (V region)** of an immunoglobulin chain is coded by the V gene and varies extensively when different chains are compared, as the result of multiple (different) genomic copies and changes introduced during construction of an active immunoglobulin.

Constant regions (C region) of immunoglobulins are coded by C genes and are the parts of the chain that vary least. Those of heavy chains identify the type of immunoglobulin.

A **V gene** is sequence coding for the major part of the variable (N-terminal) region of an immunoglobulin chain.

C genes code for the constant regions of immunoglobulin protein chains.

Somatic recombination describes the process of joining a C gene to a C gene in a lymphocyte to generate an immunoglobulin or T cell receptor.

Key Concepts

- An immunoglobulin is a tetramer of two light chains and two heavy chains.
- Light chains fall into the lambda and kappa families; heavy chains form a single family.
- Each chain has an N-terminal variable region (V) and a C-terminal constant region (C).
- The V domain recognizes antigen and the C domain provides the effector response.
- V domains and C domains are separately coded by V gene segments and C gene segments.
- A gene coding for an intact immunoglobulin chain is generated by somatic recombination to join a V gene segment with a C gene segment.

A remarkable feature of the immune response is an animal's ability to produce an appropriate antibody whenever it is exposed to a new antigen. How can the organism be prepared to produce antibody proteins each designed specifically to recognize an

antigen whose structure cannot be anticipated?

For practical purposes, we usually reckon that a mammal has the ability to produce 10^6-10^8 different antibodies. Each antibody is an immunoglobulin tetramer consisting of two identical **light chains** (L) and two identical **heavy chains** (H). If any light chain can associate with any heavy chain, to produce 10^6–10^8 potential antibodies requires 10^3–10^4 different light chains and 10^3–10^4 different heavy chains.

There are 2 types of light chain and ~10 types of heavy chain. Different classes of immunoglobulins have different effector functions. The class is determined by the heavy chain constant region, which exercises the effector function (see **Figure 25.17**).

The structure of the immunoglobulin tetramer is illustrated in **Figure 25.4**. Light chains and heavy chains share the same general type of organization in which each protein chain consists of two principal regions: the N-terminal **variable region** (**V region**); and the C-terminal **constant region** (**C region**). They were defined originally by comparing the amino acid sequences of different immunoglobulin chains. As the names suggest, the variable regions show considerable changes in sequence from one protein to the next, while the constant regions show substantial homology.

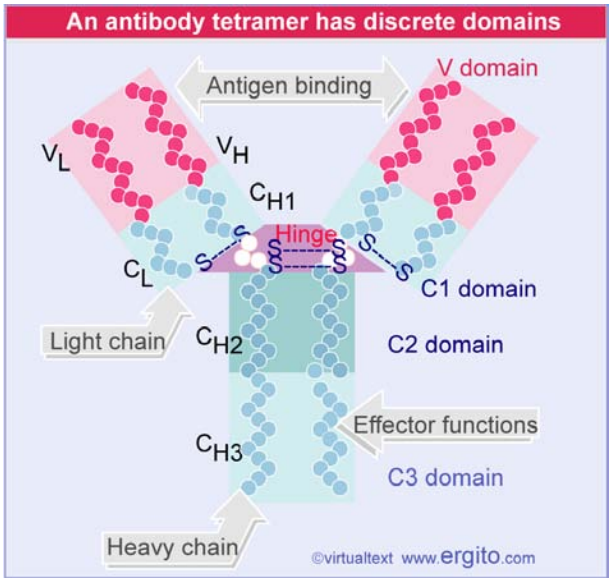

Figure 25.4 Heavy and light chains combine to generate an immunoglobulin with several discrete domains.

Corresponding regions of the light and heavy chains associate to generate distinct domains in the immunoglobulin protein.

The variable (V) domain is generated by association between the variable regions of the light chain and heavy chain. *The V domain is responsible for recognizing the antigen.* An immunoglobulin has a Y-shaped structure in which the arms of the Y are identical, and each arm has a copy of the V domain. Production of V domains of different specificities creates the ability to respond to diverse antigens. The total

number of variable regions for either light- or heavy-chain proteins is measured in hundreds. *So the protein displays the maximum versatility in the region responsible for binding the antigen.*

The number of constant regions is vastly smaller than the number of variable regions – typically there are only 1-10 C regions for any particular type of chain. The constant regions in the subunits of the immunoglobulin tetramer associate to generate several individual C domains. The first domain results from association of the single constant region of the light chain (C_L) with the C_{H1} part of the heavy-chain constant region. The two copies of this domain complete the arms of the Y-shaped molecule. Association between the C regions of the heavy chains generates the remaining C domains, which vary in number depending on the type of heavy chain.

Comparing the characteristics of the variable and constant regions, we see the central dilemma in immunoglobulin gene structure. How does the genome code for a set of proteins in which any individual polypeptide chain must have one of <10 possible C regions, but can have any one of several hundred possible V regions? It turns out that the number of coding sequences for each type of region reflects its variability. There are many genes coding for V regions, but only a few genes coding for C regions.

In this context, *"gene" means a sequence of DNA coding for a discrete part of the final immunoglobulin polypeptide* (heavy or light chain). So **V genes** code for variable regions and **C genes** code for constant regions, although *neither type of gene is expressed as an independent unit.* To construct a unit that can be expressed in the form of an authentic light or heavy chain, a V gene must be joined physically to a C gene. In this system, two "genes" code for one polypeptide. To avoid confusion, we will refer to these units as "gene segments" rather than "genes."

The sequences coding for light chains and heavy chains are assembled in the same way: *any one of many V gene segments may be joined to any one of a few C gene segments.* This **somatic recombination** occurs *in the B lymphocyte in which the antibody is expressed.* The large number of available V gene segments is responsible for a major part of the diversity of immunoglobulins. However, not all diversity is coded in the genome; some is generated by changes that occur during the process of constructing a functional gene (for review see 261; 263).

Essentially the same description applies to the formation of functional genes coding for the protein chains of the T cell receptor. Two types of receptor are found on T cells, one consisting of two types of chain called α and β, the other consisting of γ and δ chains. Like the genes coding for immunoglobulins, the genes coding for the individual chains in T cell receptors consist of separate parts, including V and C regions, that are brought together in an active T cell (see *Molecular Biology 5.25.18 T cell receptors are related to immunoglobulins*; for review see 770).

The crucial fact about the synthesis of immunoglobulins, therefore, is that *the arrangement of V gene segments and C gene segments is different in the cells producing the immunoglobulins (or T cell receptors) from all other somatic cells or germ cells* (768).

The construction of a functional immunoglobulin or T cell receptor gene might seem

to be a Lamarckian process, representing a change in the genome that responds to a particular feature of the phenotype (the antigen). At birth, the organism does not possess the functional gene for producing a particular antibody or T cell receptor. It possesses a large number of V gene segments and a smaller number of C gene segments. The subsequent construction of an active gene from these parts allows the antibody/receptor to be synthesized so that it is available to react with the antigen. The clonal selection theory requires that this rearrangement of DNA occurs *before the exposure to antigen,* which then results in *selection* for those cells carrying a protein able to bind the antigen. The entire process occurs in somatic cells and does not affect the germline; so the response to an antigen is not inherited by progeny of the organism.

There are two families of immunoglobulin light chains, κ and λ, and one family containing all the types of heavy chain (H). Each family resides on a different chromosome, and consists of its own set of both V gene segments and C gene segments. This is called the *germline pattern*, and is found in the germline and in somatic cells of all lineages other than the immune system.

But in a cell expressing an antibody, each of its chains – one light type (either κ or λ) and one heavy type – is coded by a single intact gene. The recombination event that brings a V gene segment to partner a C gene segment creates an active gene consisting of exons that correspond precisely with the functional domains of the protein. The introns are removed in the usual way by RNA splicing.

Recombination between V and C gene segments to give functional loci occurs in a population of immature lymphocytes. A B lymphocyte usually has only one productive rearrangement of light-chain gene segments (either κ or λ) and one of heavy-chain gene segments. Similarly, a T lymphocyte productively rearranges an α gene and a β gene, or one δ gene and one γ gene. The antibody or T cell receptor produced by any one cell is determined by the particular configuration of V gene segments and C gene segments that has been joined.

The principles by which functional genes are assembled are the same in each family, but there are differences in the details of the organization of the V and C gene segments, and correspondingly of the recombination reaction between them. In addition to the V and C gene segments, other short DNA sequences (including J segments and D segments) are included in the functional somatic loci (for review see 267; 264).

Reviews

261. Tonegawa, S. (1983). *Somatic generation of antibody diversity*. Nature 302, 575-581.

263. Alt, F. W., Blackwell, T. K., and Yancopoulos, G. D. (1987). *Development of the primary antibody repertoire*. Science 238, 1079-1087.

264. Blackwell, T. K. and Alt, F. W. (1989). *Mechanism and developmental program of immunoglobulin gene rearrangement in mammals*. Annu. Rev. Genet. 23, 605-636.

267. Yancopoulos, G. D. and Alt, F. W. (1986). *Regulation of the assembly and expression of variable-region genes*. Annu. Rev. Immunol. 4, 339-68.

770. Hood, L., Kronenberg, M., and Hunkapiller, T. (1985). *T cell antigen receptors and the immunoglobulin supergene family*. Cell 40, 225-229.

References

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

IMMUNE DIVERSITY

5.25.4 Light chains are assembled by a single recombination

Key Terms

J segments (joining segments) are coding sequences in the immunoglobulin and T cell receptor loci. The J segments are between the variable (V) and constant (C) gene segments.

Key Concepts

- A lambda light chain is assembled by a single recombination between a V gene and a J-C gene segment.
- The V gene segment has a leader exon, intron, and variable-coding region.
- The J-C gene segment has a short J-coding exon, intron, and C-coding region.
- A kappa light chain is assembled by a single recombination between a V gene segment and one of five J segments preceding the C gene.

A λ light chain is assembled from two parts, as illustrated in **Figure 25.5**. The V gene segment consists of the leader exon (L) separated by a single intron from the variable (V) segment. The C gene segment consists of the J segment separated by a single intron from the constant (C) exon.

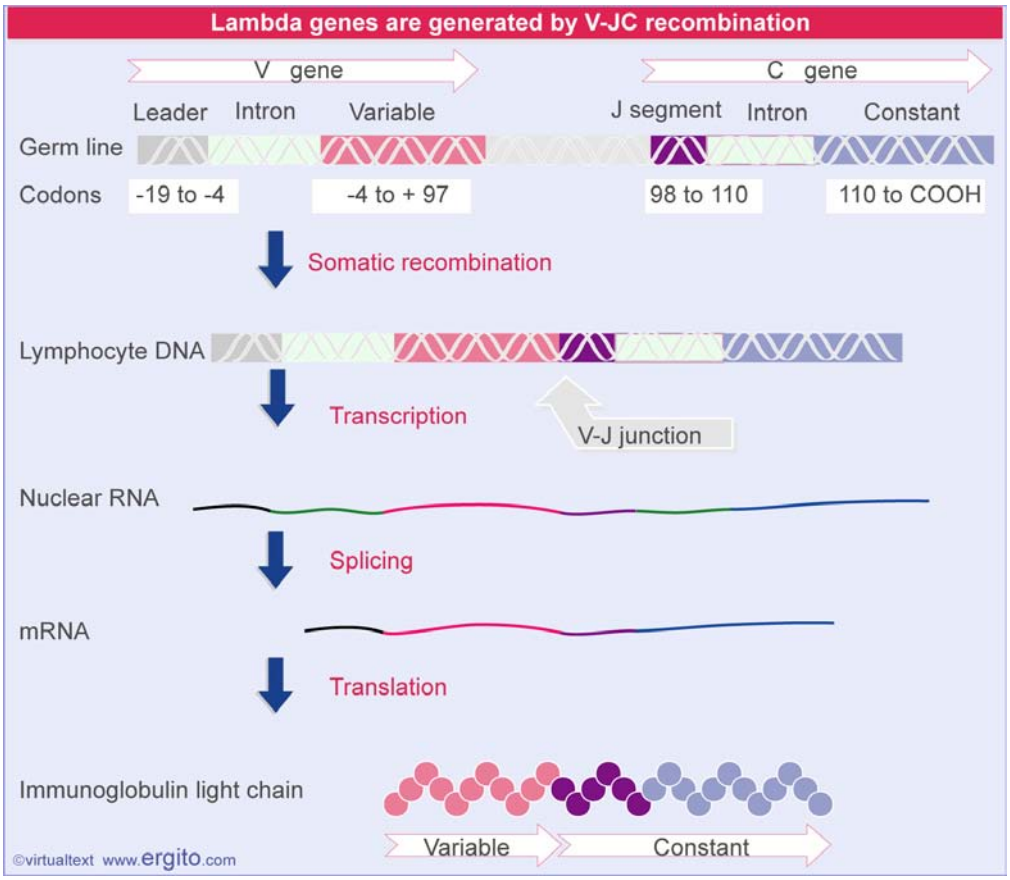

Figure 25.5 The lambda C gene segment is preceded by a J segment, so that V-J recombination generates a functional lambda light-chain gene.

The name of the **J segment** is an abbreviation for joining, since it identifies the region to which the V segment becomes connected. So the joining reaction does not directly involve V and C gene segments, but occurs via the J segment; when we discuss the joining of "V and C gene segments" for light chains, we really mean V-JC joining.

The J segment is short and codes for the last few (13) amino acids of the variable region, as defined by amino acid sequences. In the intact gene generated by recombination, the V-J segment constitutes a single exon coding for the entire variable region.

The consequences of the κ joining reaction are illustrated in **Figure 25.6**. A κ light chain also is assembled from two parts, but there is a difference in the organization of the C gene segment. A group of five J segments is spread over a region of 500-700 bp, separated by an intron of 2-3 kb from the C_κ exon. In the mouse, the central J segment is nonfunctional (ψJ3). A V_κ segment may be joined to any one of the J segments (769).

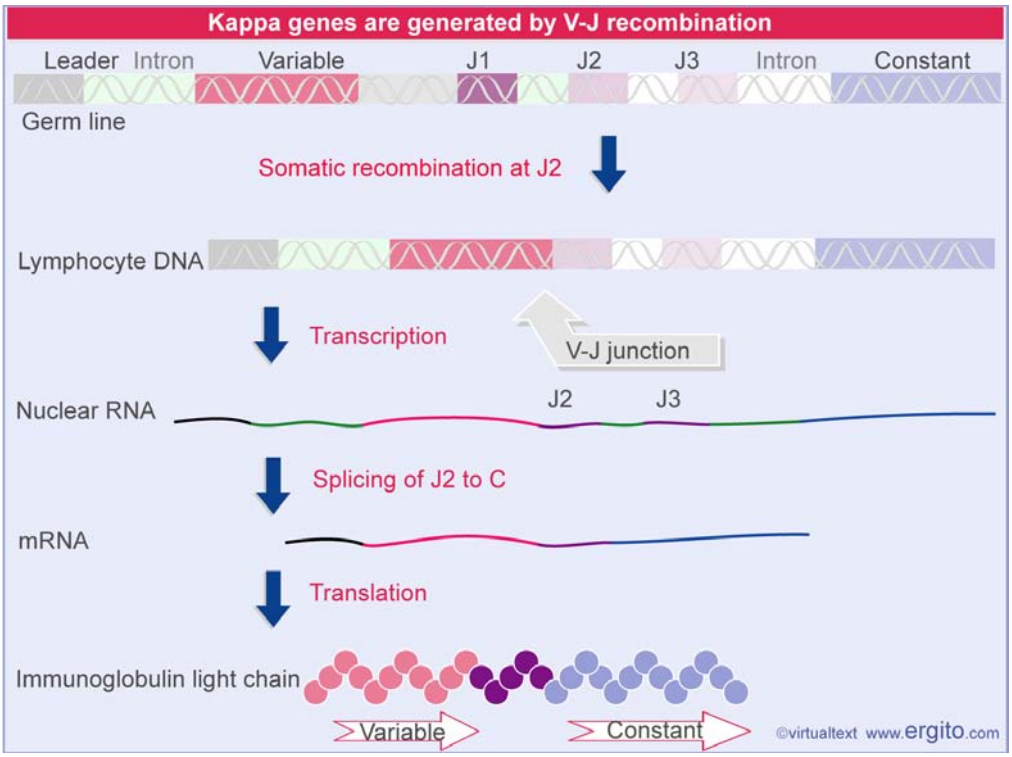

Figure 25.6 The kappa C gene segment is preceded by multiple J segments in the germ line. V-J joining may recognize any one of the J segments, which is then spliced to the C gene segment during RNA processing.

Whichever J segment is used becomes the terminal part of the intact variable exon. Any J segments on the left of the recombining J segment are lost (J1 has been lost in the figure). Any J segment on the right of the recombining J segment is treated as part of the intron between the variable and constant exons (J3 is included in the intron that is spliced out in the figure).

All functional J segments possess a signal at the left boundary that makes it possible to recombine with the V segment; and they possess a signal at the right boundary that can be used for splicing to the C exon. Whichever J segment is recognized in DNA joining uses its splicing signal in RNA processing.

References

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

IMMUNE DIVERSITY

5.25.5 Heavy chains are assembled by two recombinations

Key Terms

The **D segment** is an additional sequence that is found between the V and J regions of an immunoglobulin heavy chain.

Key Concepts

- The units for heavy chain recombination are a V gene, D segment, and J-C gene segment.
- The first recombination joins D to J-C.
- The second recombination joins V to D-J-C.
- The C segment consists of several exons.

Heavy chain construction involves an additional segment. The **D segment** (for diversity) was discovered by the presence in the protein of an extra 2-13 amino acids between the sequences coded by the V segment and the J segment. An array of >10 D segments lies on the chromosome between the V_H segments and the 4 J_H segments.

V-D-J joining takes place in two stages, as illustrated in **Figure 25.7**. First one of the D segments recombines with a J_H segment; then a V_H segment recombines with the DJ_H combined segment. The reconstruction leads to expression of the adjacent C_H segment (which consists of several exons). (We discuss the use of different C_H gene segments in *Molecular Biology 5.25.12 Class switching is caused by DNA recombination*; now we will just consider the reaction in terms of the connection to one of several J segments that precede a C_H gene segment.)

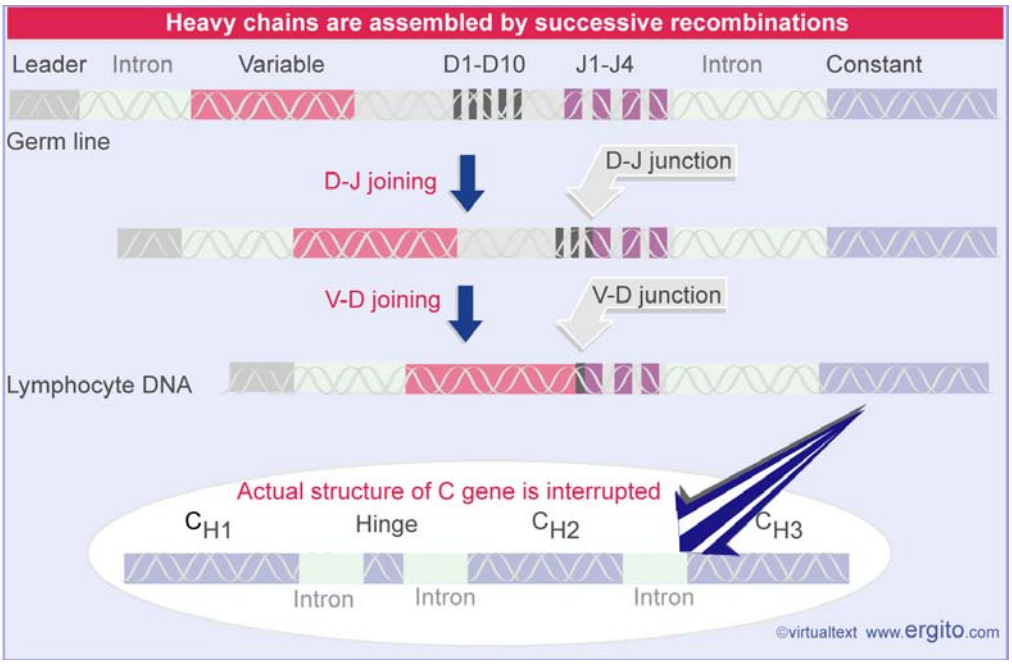

Figure 25.7 Heavy genes are assembled by sequential joining reactions. First a D segment is joined to a J segment; then a V gene segment is joined to the D segment.

The D segments are organized in a tandem array. The mouse heavy-chain locus contains 12 D segments of variable length; the human locus has ~30 D segments (not all necessarily active). Some unknown mechanism must ensure that the *same* D segment is involved in the D-J joining and V-D joining reactions. (When we discuss joining of V and C gene segments for heavy chains, we assume the process has been completed by V-D and D-J joining reactions.)

The V gene segments of all three immunoglobulin families are similar in organization. The first exon codes for the signal sequence (involved in membrane attachment), and the second exon codes for the major part of the variable region itself (<100 codons long). The remainder of the variable region is provided by the D segment (in the H family only) and by a J segment (in all three families).

The structure of the constant region depends on the type of chain. For both κ and λ light chains, the constant region is coded by a single exon (which becomes the third exon of the reconstructed, active gene). For H chains, the constant region is coded by several exons; corresponding with the protein chain shown in **Figure 25.4**, separate exons code for the regions C_{H1}, hinge, C_{H2}, and C_{H3}. Each C_H exon is ~100 codons long; the hinge is shorter. The introns usually are relatively small (~300 bp).

This content is available online at http://www.ergito.com/main.jsp?bcs=MBIO.5.25.5

IMMUNE DIVERSITY

5.25.6 Recombination generates extensive diversity

Key Concepts

- A light chain locus can produce >1000 chains by combining 300 V genes with 4-5 C genes.
- An H locus can produce >4000 chains by combining 300 V genes, 20 D segments, and 4 J segments.

Now we must examine the different types of V and C gene segments to see how much diversity can be accommodated by the variety of the coding regions carried in the germline. In each light Ig gene family, many V gene segments are linked to a much smaller number of C gene segments.

Figure 25.8 shows that the λ locus has ~6 C gene segments, each preceded by its own J segment. The λ locus in mouse is much less diverse than the human locus. The main difference is that in mouse there are only two Vλ gene segments; each is linked to two J-C regions. Of the 4 Cλ gene segments, one is inactive. At some time in the past, the mouse suffered a catastrophic deletion of most of its germline Vλ gene segments.

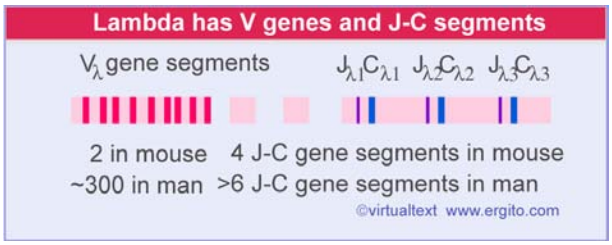

Figure 25.8 The lambda family consists of V gene segments linked to a small number of J-C gene segments.

Figure 25.9 shows that the κ locus has only one C gene segment, although it is preceded by 5 J segments (one of them inactive). The Vκ gene segments occupy a large cluster on the chromosome, upstream of the constant region. The human cluster has two regions. Just preceding the Cκ gene segment, a region of 600 kb contains the 5 Jκ segments and 40 Vκ gene segments. A gap of 800 kb separates this region from another group of 36 Vκ gene segments.

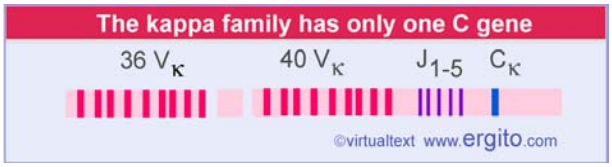

Figure 25.9 The human and mouse kappa families consist of V gene segments linked to 5 J segments connected to a single C gene segment.

The V κ gene segments can be subdivided into families, defined by the criterion that members of a family have >80% amino acid identity. The mouse family is unusually large, ~1000 genes, and there are ~18 V κ families, varying in size from 2-100 members. Like other families of related genes, therefore, related V gene segments form subclusters, generated by duplication and divergence of individual ancestral members. However, many of the V segments are inactive pseudogenes, and <50 of the ~300 total are likely to be used to generate immunoglobulins.

A given lymphocyte generates *either* a κ*or* a λ light chain to associate with the heavy chain. In man, ~60% of the light chains are κ and ~40% are λ. In mouse, 95% of B cells express the κ type of light chain, presumably because of the reduced number of λ gene segments.

The single locus for heavy chain production in Man consists of several discrete sections, as summarized in **Figure 25.10**. It is similar in the mouse, where there are more V_H gene segments, fewer D and J segments, and a slight difference in the number and organization of C gene segments. The 3′ member of the V_H cluster is separated by only 20 kb from the first D segment. The D segments are spread over ~50 kb, and then comes the cluster of J segments. Over the next 220 kb lie all the C_H gene segments. There are 9 functional C_H gene segments and 2 pseudogenes. The organization suggests that a γ gene segment must have been duplicated to give the subcluster of γ- γ- ε - α, after which the entire group was then duplicated.

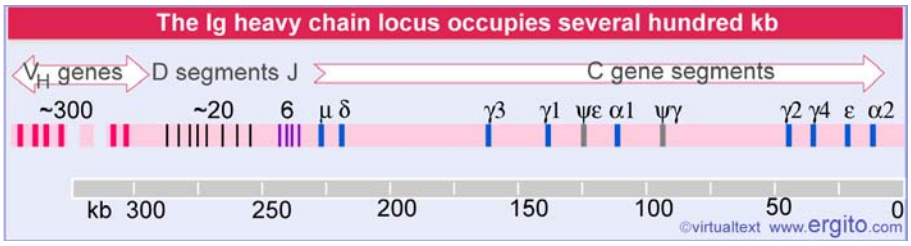

Figure 25.10 A single gene cluster in man contains all the information for heavy-chain gene assembly.

How far is the diversity of germline information responsible for V region diversity in immunoglobulin proteins? By combining any one of ~50 V gene segments with any one of 4-5 J segments, a typical light chain locus has the potential to produce some 250 chains. There is even greater diversity in the H chain locus; by combining any one of ~50 V_H gene segments, 20 D segments, and 4 J segments, the genome potentially can produce 4000 variable regions to accompany any C_H gene segment. In mammals, this is the starting point for diversity, but additional mechanisms introduce further changes. *When closely related variants of immunoglobulins are examined, there often are more proteins than can be accounted for by the number of*

corresponding V gene segments. The new members are created by somatic changes in individual genes during or after the recombination process (see *Molecular Biology 5.25.14 Somatic mutation generates additional diversity in mouse and man*).

IMMUNE DIVERSITY

5.25.7 Immune recombination uses two types of consensus sequence

Key Concepts

- The consensus sequence used for recombination is a heptamer separated by either 12 or 23 base pairs from a nonamer.
- Recombination occurs between two consensus sequences that have different spacings.

Assembly of light- and heavy-chain genes involves the same mechanism (although the number of parts is different). The same consensus sequences are found at the boundaries of all germline segments that participate in joining reactions. Each consensus sequence consists of a heptamer separated by either 12 or 23 bp from a nonamer.

Figure 25.11 illustrates the relationship between the consensus sequences at the mouse Ig loci. At the κ locus, each V κ gene segment is followed by a consensus sequence with a 12 bp spacing. Each J κ segment is preceded by a consensus sequence with a 23 bp spacing. The V and J consensus sequences are inverted in orientation. At the λ locus, each V λ gene segment is followed by a consensus sequence with 23 bp spacing, while each J λ gene segment is preceded by a consensus of the 12 bp spacer type (771).

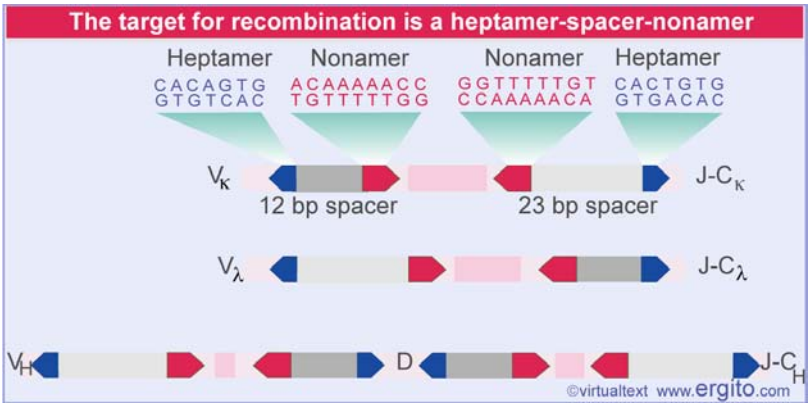

Figure 25.11 Consensus sequences are present in inverted orientation at each pair of recombining sites. One member of each pair has a spacing of 12 bp between its components; the other has 23 bp spacing.

The rule that governs the joining reaction is that *a consensus sequence with one type of spacing can be joined only to a consensus sequence with the other type of spacing.* Since the consensus sequences at V and J segments can lie in either order, the different spacings do not impart any directional information, but serve to prevent one V gene segment from recombining with another, or one J segment from recombining

with another.

This concept is borne out by the structure of the components of the heavy gene segments. Each V_H gene segment is followed by a consensus sequence of the 23 bp spacer type. The D segments are flanked on either side by consensus sequences of the 12 bp spacer type. The J_H segments are preceded by consensus sequences of the 23 bp spacer type. So the V gene segment must be joined to a D segment; and the D segment must be joined to a J segment. A V gene segment cannot be joined directly to a J segment, because both possess the same type of consensus sequence.

The spacing between the components of the consensus sequences corresponds almost to one or two turns of the double helix. This may reflect a geometric relationship in the recombination reaction. For example, the recombination protein(s) may approach the DNA from one side, in the same way that RNA polymerase and repressors approach recognition elements such as promoters and operators.

References

771. Lewis, S., Gifford, A., and Baltimore, D. (1985). *DNA elements are asymmetrically joined during the site-specific recombination of kappa immunoglobulin genes*. Science 228, 677-685.

IMMUNE DIVERSITY

5.25.8 Recombination generates deletions or inversions

Key Terms

A **signal end** is produced during recombination of immunoglobulin and T cell receptor genes. The signal ends are at the termini of the cleaved fragment containing the recombination signal sequences. The subsequent joining of the signal ends yields a signal joint.

A **coding end** is produced during recombination of immunoglobulin and T cell receptor genes. Coding ends are at the termini of the cleaved V and (D)J coding regions. The subsequent joining of the coding ends yields a coding joint.

Key Concepts

- Recombination occurs by double-strand breaks at the heptamers of two consensus sequences.
- The signal ends of the fragment between the breaks usually join to generate an excised circular fragment.
- The coding ends are covalently linked to join V to J-C (L chain) or D to J-C and V to D-J-C (H chain).
- If the recombining genes are in inverted instead of direct orientation, there is an inversion instead of deletion of an excised circle.

Recombination of the components of immunoglobulin genes is accomplished by a physical rearrangement of sequences, involving breakage and reunion, but the mechanism is different from homologous recombination. The general nature of the reaction is illustrated in **Figure 25.12** for the example of a κ light chain. (The reaction is similar at a heavy chain locus, except that there are two recombination events: first D-J, then V-DJ.)

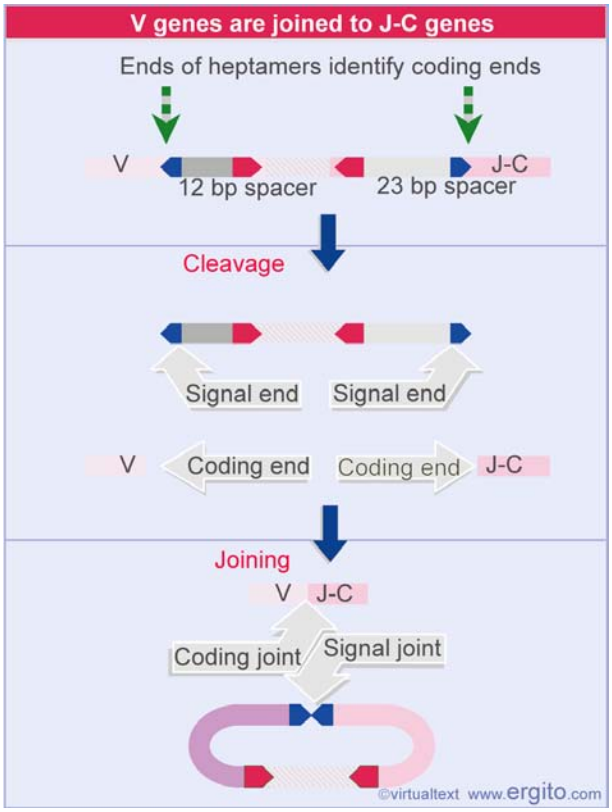

Figure 25.12 Breakage and reunion at consensus sequences generates immunoglobulin genes.

Breakage and reunion occur as separate reactions. A double-strand break is made at the heptamers that lie at the ends of the coding units. This releases the entire fragment between the V gene segment and J-C gene segment; the cleaved termini of this fragment are called **signal ends**. The cleaved termini of the V and J-C loci are called **coding ends**. The two coding ends are covalently linked to form a coding joint; this is the connection that links the V and J segments. If the two signal ends are also connected, the excised fragment would form a circular molecule.

We have shown the V and J-C loci as organized in the same orientation. As a result, the cleavage at each consensus sequence releases the region between them as a linear fragment. If the signal ends are joined, it is converted into a circular molecule, as indicated in **Figure 25.12**. Deletion to release an excised circle is the predominant mode of recombination at the immunoglobulin and TCR loci.

In some exceptional cases the V gene segment is inverted in orientation on the chromosome relative to the J-C loci. In such a case, breakage and reunion inverts the intervening material instead of deleting it. The outcomes of deletion versus inversion are the same as shown previously for homologous recombination between direct or inverted repeats in **Figure 16.9** and **Figure 16.10**. There is one further proviso, however; recombination with an inverted V gene segment makes it *necessary* for the signal ends to be joined, because otherwise there is a break in the locus. Inversion occurs in TCR recombination, and also sometimes in the κ light chain locus.

IMMUNE DIVERSITY

5.25.9 The RAG proteins catalyze breakage and reunion

Key Terms

A **P nucleotide** sequence is a short palindromic (inverted repeat) sequence that is generated during rearrangement of immunoglobulin and T cell receptor V, (D), J gene segments. P nucleotides are generated at coding joints when RAG proteins cleave the hairpin ends generated during rearrangement.

An **N nucleotide** sequence is a short non-templated sequence that is added randomly by the enzyme at coding joints during rearrangement of immunoglobulin and T cell receptor genes. N nucleotides augment the diversity of antigen receptors.

Key Concepts

- The RAG proteins are necessary and sufficient for the cleavage reaction.
- RAG1 recognizes the nonamer consensus sequences for recombination. RAG2 binds to RAG1 and cleaves at the heptamer.
- The reaction resembles the topoisomerase-like resolution reaction that occurs in transposition.
- It proceeds through a hairpin intermediate at the coding end; opening of the hairpin is responsible for insertion of extra bases (P nucleotides) in the recombined gene.
- Deoxynucleoside transferase inserts additional N nucleotides at the coding end.
- The codon at the site of the V-(D)J joining reaction has an extremely variable sequence and codes for amino acid 96 in the antigen-binding site.
- The double-strand breaks at the coding joints are repaired by the same system involved in nonhomologous end-joining of damaged DNA.
- An enhancer in the C gene activates the promoter of the V gene after recombination has generated the intact immunoglobulin gene.

The proteins RAG1 and RAG2 are necessary and sufficient to cleave DNA for V(D)J recombination. They are coded by two genes, separated by <10 kb on the chromosome, whose transfection into fibroblasts causes a suitable substrate DNA to undergo the V(D)J joining reaction (1260; 1261) (see *Great Experiments 12.5 Identification of the V(D)J recombination activating genes, RAG1 and RAG2*). Mice that lack either *RAG1* or *RAG2* are unable to recombine their immunoglobulins or T cell receptors, and as a result have immature B and T lymphocytes. The RAG proteins together undertake the catalytic reactions of cleaving and rejoining DNA, and also provide a structural framework within which the reactions occur (3022; 3023; 3025).

RAG1 recognizes the heptamer/nonamer signals with the appropriate 12/23 spacing and recruits RAG2 to the complex. The nonamer provides the site for initial recognition, and the heptamer directs the site of cleavage.

The reactions involved in recombination are shown in **Figure 25.13**. The complex nicks one strand at each junction. The nick has 3′–OH and 5′–P ends. The free 3′–OH end then attacks the phosphate bond at the corresponding position *in the other strand of the duplex.* This creates a hairpin at the coding end, in which the 3′ end of one strand is covalently linked to the 5′ end of the other strand; it leaves a blunt double-strand break at the signal end.

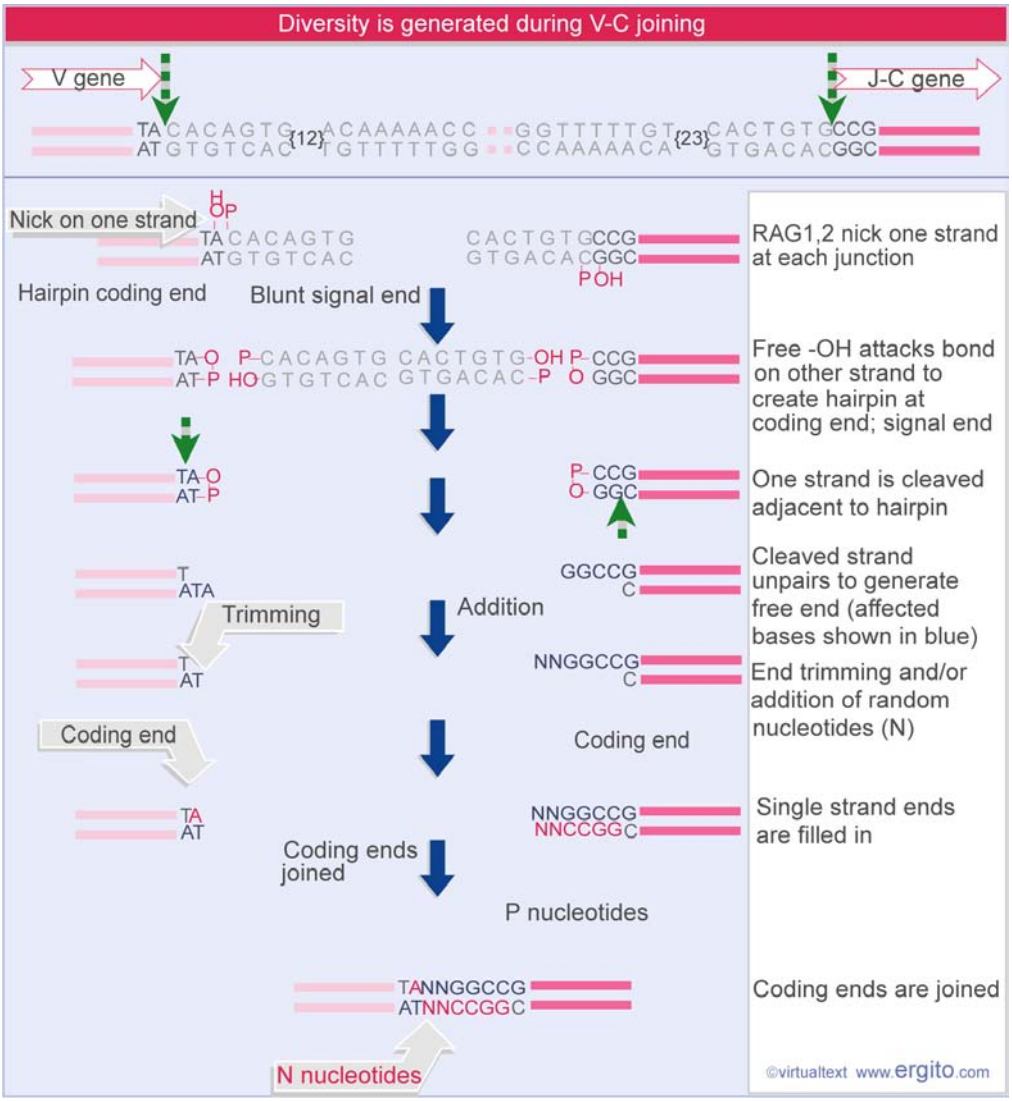

Figure 25.13 Processing of coding ends introduces variability at the junction.

This second cleavage is a transesterification reaction in which bond energies are conserved. It resembles the topoisomerase-like reactions catalyzed by the resolvase proteins of bacterial transposons (see *Molecular Biology 4.16.9 TnA transposition requires transposase and resolvase*). The parallel with these reactions is supported

further by a homology between RAG1 and bacterial invertase proteins (which invert specific segments of DNA by similar recombination reactions). In fact, the RAG proteins can insert a donor DNA whose free ends consist of the appropriate signal sequences (heptamer–12/23–spacer nonamer) into an unrelated target DNA in an *in vitro* transposition reaction (1196; 1197; 1198). This suggests that somatic recombination of immune genes evolved from an ancestral transposon. It also suggests that the RAG proteins are responsible for chromosomal translocations in which Ig or TCR loci are connected to other loci (see *Molecular Biology 6.30.12 Proto-oncogenes can be activated by translocation*).

The hairpins at the coding ends provide the substrate for the next stage of reaction. If a single-strand break is introduced into one strand close to the hairpin, an unpairing reaction at the end generates a single-stranded protrusion. Synthesis of a complement to the exposed single strand then converts the coding end to an extended duplex. This reaction explains the introduction of **P nucleotides** at coding ends; they consist of a few extra base pairs, related to, but reversed in orientation from, the original coding end.

Some extra bases also may be inserted, apparently with random sequences, between the coding ends. They are called **N nucleotides**. Their insertion occurs via the activity of the enzyme deoxynucleoside transferase (known to be an active component of lymphocytes) at a free 3′ coding end generated during the joining process.

Changes in sequence during recombination are therefore a consequence of the enzymatic mechanisms involved in breaking and rejoining the DNA. In heavy chain recombination, base pairs are lost or inserted at the V_H-D or D-J or both junctions. Deletion also occurs in V λ-J λ joining, but insertion at these joints is unusual (773; for review see 274; 275). The changes in sequence affect the amino acid coded at V-D and D-J junctions in heavy chains or at the V-J junction in light chains.

These various mechanisms together ensure that a coding joint may have a sequence that is different from what would be predicted by a direct joining of the coding ends of the V, D, and J regions.

Changes in the sequence at the junction make it possible for a great variety of amino acids to be coded at this site. It is interesting that the amino acid at position 96 is created by the V-J joining reaction. It forms part of the antigen-binding site and also is involved in making contacts between the light and heavy chains. So the maximum diversity is generated at the site that contacts the target antigen.

Changes in the number of base pairs at the coding joint affect the reading frame. The joining process appears to be random with regard to reading frame, so that probably only one third of the joined sequences retain the proper frame of reading through the junctions. If the V-J region is joined so that the J segment is out of phase, translation is terminated prematurely by a nonsense codon in the incorrect frame. We may think of the formation of aberrant genes as comprising the price the cell must pay for the increased diversity that it gains by being able to adjust the sequence at the joining site.

Similar although even greater diversity is generated in the joining reactions that

involve the D segment of the heavy chain. The same result is seen with regard to reading frame; nonproductive genes are generated by joining events that place J and C out of phase with the preceding V gene segment.

The joining reaction that works on the coding end uses the same pathway of nonhomologous end-joining (NHEJ) that repairs double-strand breaks in cells (for review see 2514) (See *Molecular Biology 4.15.28 Eukaryotic cells have conserved repair systems*). The initial stages of the reaction were identified by isolating intermediates from lymphocytes of mice with the *SCID* mutation, which results in a much reduced activity in immunoglobulin and TCR recombination. *SCID* mice accumulate broken molecules that terminate in double-strand breaks at the coding ends, and are thus deficient in completing some aspect of the joining reaction. The *SCID* mutation inactivates a DNA-dependent protein kinase (DNA-PK). The kinase is recruited to DNA by the Ku70 and Ku80 proteins, which bind to the DNA ends. DNA-PK phosphorylates and thereby activates the protein Artemis, which nicks the hairpin ends (2397) (it also has exonuclease and endonuclease activities that function in the NHEJ pathway). The actual ligation is undertaken by DNA ligase IV and also requires the protein XRCC4. As a result, mutations in the Ku proteins or in XRCC4 or DNA ligase IV are found among human patients who have diseases caused by deficiencies in DNA repair that result in increased sensitivity to radiation.

What is the connection between joining of V and C gene segments and their activation? Unrearranged V gene segments are not actively represented in RNA. But when a V gene segment is joined productively to a C_κ gene segment, the resulting unit is transcribed. However, since the sequence upstream of a V gene segment is not altered by the joining reaction, *the promoter must be the same in unrearranged, nonproductively rearranged, and productively rearranged genes.*

A promoter lies upstream of every V gene segment, but is inactive. It is activated by its relocation to the C region. The effect must depend on sequences downstream. What role might they play? An enhancer located within or downstream of the C gene segment activates the promoter at the V gene segment. The enhancer is tissue specific; it is active only in B cells. Its existence suggests the model illustrated in **Figure 25.14**, in which the V gene segment promoter is activated when it is brought within the range of the enhancer.

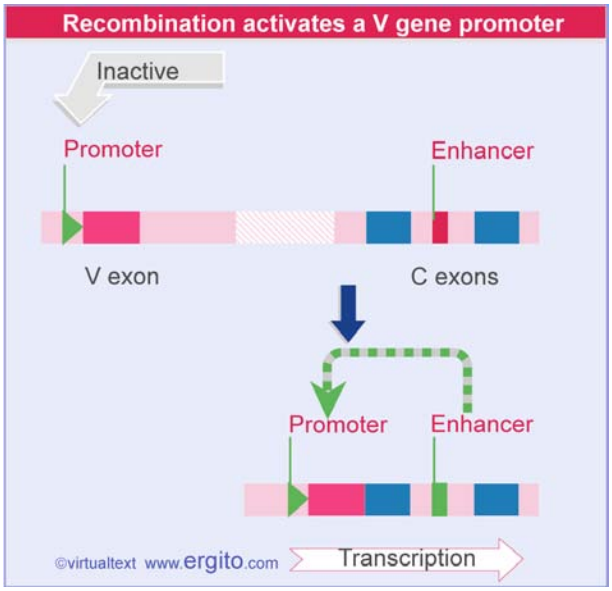

Figure 25.14 A V gene promoter is inactive until recombination brings it into the proximity of an enhancer in the C gene segment. The enhancer is active only in B lymphocytes.

Last updated on 10-16-2002

Reviews

274. Gellert, M. (1992). *Molecular analysis of VDJ recombination*. Annu. Rev. Genet. 26, 425-446.

275. Schatz, D. G., Oettinger, M. A., and Schlissel, M. S. (1992). *VDJ recombination: molecular biology and regulation*. Annu. Rev. Immunol. 10, 359-383.

2514. Jeggo, P. A. (1998). *DNA breakage and repair.* Adv. Genet. 38, 185-218.

References

773. Roth, D. B., Menetski, J. P., Nakajima, P. B., Bosma, M. J., and Gellert, M. (1992). *V D J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in SCID mouse thymocytes*. Cell 70, 983-991.

1196. Agrawal, A., Eastman, Q. M., and Schatz, D. G. (1998). *Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system*. Nature 394, 744-751.

1197. Hiom, K., Melek, M., and Gellert, M. (1998). *DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations*. Cell 94, 463-470.

1198. Melek, M. and Gellert, M. (2000). *RAG1/2-mediated resolution of transposition intermediates: two pathways and possible consequences.* Cell 101, 625-633.

1260. Schatz, D. G. and Baltimore, D. (1988). *Stable expression of immunoglobulin gene V(D)J recombinase activity by gene transfer into 3T3 fibroblasts.* Cell 53, 107-115.

1261. Schatz, D. G., Oettinger, M. A., and Baltimore, D. (1989). *The V(D)J recombination activating gene, RAG-1*. Cell 59, 1035-1048.

2397. Ma, Y., Pannicke, U., Schwarz, K., and Lieber, M. R. (2002). *Hairpin Opening and Overhang Processing by an Artemis/DNA-Dependent Protein Kinase Complex in Nonhomologous End Joining and V(D)J Recombination.* Cell 108, 781-794.

3022. Yarnell Schultz, H., Landree, M. A., Qiu, J. X., Kale, S. B., and Roth, D. B. (2001). *Joining-deficient RAG1 mutants block V(D)J recombination in vitro and hairpin opening in vitro.* Mol. Cell 7, 65-75.

3023. Qiu, J. X., Kale, S. B., Yarnell Schultz, H., and Roth, D. B. (2001). *Separation-of-function mutants reveal critical roles for RAG2 in both the cleavage and joining steps of V(D)J recombination.* Mol. Cell 7, 77-87.

3025. Tsai, C. L., Drejer, A. H., and Drejer, A. H. (2002). *Evidence of a critical architectural function for the RAG proteins in end processing, protection, and joining in V(D)J recombination.* Genes Dev. 16, 1934-1949.

IMMUNE DIVERSITY

5.25.10 Allelic exclusion is triggered by productive rearrangement

Key Terms

Allelic exclusion describes the expression in any particular lymphocyte of only one allele coding for the expressed immunoglobulin. This is caused by feedback from the first immunoglobulin allele to be expressed that prevents activation of a copy on the other chromosome.

The recombination of V, (D), J gene segments results in a **productive rearrangement** if all the rearranged gene segments are in the correct reading frame.

The recombination of V, (D), J gene segments results in a **nonproductive rearrangement** if the rearranged gene segments are not in the correct reading frame. A nonproductive rearrangement occurs when nucleotide addition or subtraction disrupts the reading frame or when .

Key Concepts

- Recombination to generate an intact immunoglobulin gene is productive if it leads to expression of an active protein.
- A productive rearrangement prevents any further rearrangement from occurring, but a nonproductive rearrangement does not.
- Allelic exclusion applies separately to light chains (only one kappa *or* lambda may be productively rearranged) and to heavy chains (one heavy chain is productively rearranged).

Each B cell expresses a single type of light chain and a single type of heavy chain, because only a single productive rearrangement of each type occurs in a given lymphocyte, to produce one light and one heavy chain gene. Because each event involves the genes of only *one* of the homologous chromosomes, *the alleles on the other chromosome are not expressed in the same cell.* This phenomenon is called **allelic exclusion**.

The occurrence of allelic exclusion complicates the analysis of somatic recombination. A probe reacting with a region that has rearranged on one homologue will also detect the allelic sequences on the other homologue. We are therefore compelled to analyze the different fates of the two chromosomes together.

The usual pattern displayed by a rearranged active gene can be interpreted in terms of a deletion of the material between the recombining V and C loci.

Two types of gene organization are seen in active cells:

- Probes to the active gene may reveal one rearranged copy and one germline copy. We assume then that joining has occurred on one chromosome, while the other chromosome has remained unaltered.

- Two different rearranged patterns may be found, indicating that the chromosomes have suffered independent rearrangements. In some of these instances, material between the recombining V and C gene segments is entirely absent from the cell line. This is most easily explained by the occurrence of independent deletions (resulting from recombination) on each chromosome.

When two chromosomes both lack the germline pattern, usually only one of them has passed through a **productive rearrangement** to generate a functional gene. The other has suffered a **nonproductive rearrangement**; this may take several forms, but in each case the gene sequence cannot be expressed as an immunoglobulin chain. (It may be incomplete, for example because D-J joining has occurred but V-D joining has not followed; or it may be aberrant, with the process completed, but failing to generate a gene that codes for a functional protein.)

The coexistence of productive and nonproductive rearrangements suggests the existence of a feedback loop to control the recombination process. A model is outlined in **Figure 25.15**. Suppose that each cell starts with two loci in the unrearranged germline configuration Ig^0. Either of these loci may be rearranged to generate a productive gene Ig^+ or a nonproductive gene Ig^-.

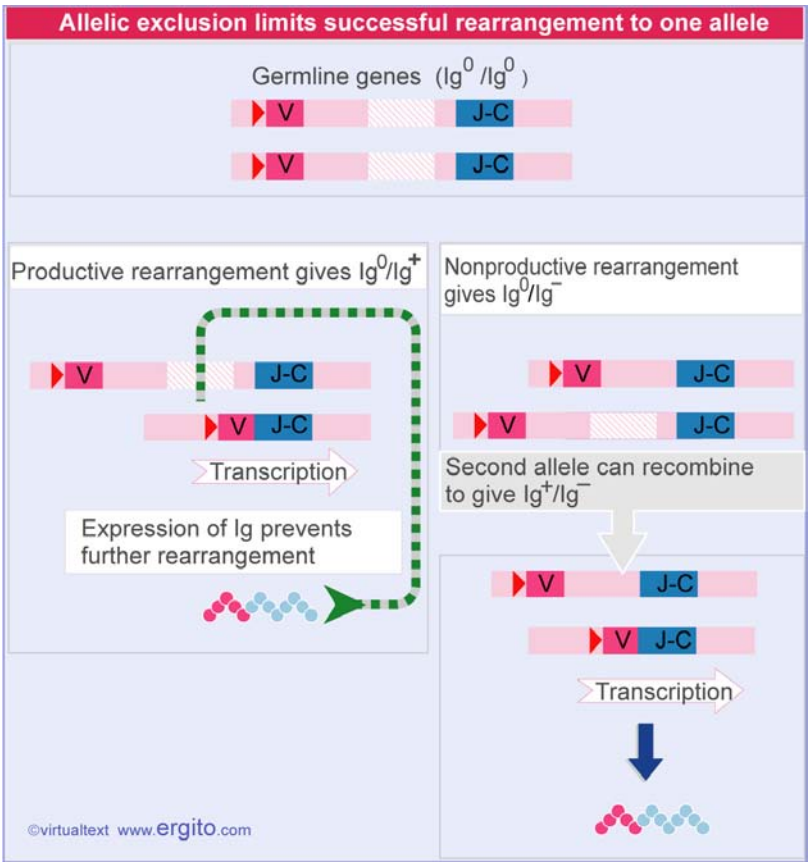

Figure 25.15 A successful rearrangement to produce an active light or heavy chain suppresses further rearrangements of the same type, and results in allelic exclusion.

If the rearrangement is productive, the synthesis of an active chain provides a trigger to prevent rearrangement of the other allele. The active cell has the configuration Ig^0/Ig^+.

If the rearrangement is nonproductive, it creates a cell with the configuration Ig^0/Ig^-. There is no impediment to rearrangement of the remaining germline allele. If this rearrangement is productive, the expressing cell has the configuration Ig^+/Ig^-. Again, the presence of an active chain suppresses the possibility of further rearrangements.

Two successive nonproductive rearrangements produce the cell Ig^-/Ig^-. In some cases an Ig^-/Ig^- cell can try yet again. Sometimes the observed patterns of DNA can only have been generated by successive rearrangements.

The crux of the model is that the cell keeps trying to recombine V and C gene segments until a productive rearrangement is achieved. Allelic exclusion is caused by the suppression of further rearrangement as soon as an active chain is produced. The use of this mechanism *in vivo* is demonstrated by the creation of transgenic mice whose germline has a rearranged immunoglobulin gene. Expression of the transgene in B cells suppresses the rearrangement of endogenous genes (for review see 269).

Allelic exclusion is independent for the heavy- and light-chain loci. Heavy chain

genes usually rearrange first. Allelic exclusion for light chains must apply equally to both families (cells may have *either* active κ or λ light chains). It is likely that the cell rearranges its κ genes first, and tries to rearrange λ only if both κ attempts are unsuccessful.

There is an interesting paradox in this series of events. The same consensus sequences and the same V(D)J recombinase are involved in the recombination reactions at H, κ, and λ loci. Yet the three loci rearrange in a set order. What ensures that heavy rearrangement precedes light rearrangement, and that κ precedes λ? The loci may become accessible to the enzyme at different times, possibly as the result of transcription. Transcription occurs even before rearrangement, although of course the products have no coding function. The transcriptional event may change the structure of chromatin, making the consensus sequences for recombination available to the enzyme.

Reviews

269. Storb, U. (1987). *Transgenic mice with immunoglobulin genes*. Annu. Rev. Immunol. 5, 151-174.

IMMUNE DIVERSITY

5.25.11 Early heavy chain expression can be changed by RNA processing

Key Concepts

- All lymphocytes start by synthesizing the membrane-bound form of IgM.
- A change in RNA splicing causes this to be replaced by the secreted form when the B cell differentiates.

The period of IgM synthesis that begins lymphocyte development falls into two parts, during which different versions of the μ constant region are synthesized:

- As a stem cell differentiates to a pre-B lymphocyte, an accompanying light chain is synthesized, and the IgM molecule ($L_2\mu_2$) appears at the surface of the cell. This form of IgM contains the μ_m version of the constant region (*m* indicates that IgM is located in the membrane). The membrane location may be related to the need to initiate cell proliferation in response to the initial recognition of an antigen.
- When the B lymphocyte differentiates further into a plasma cell, the μ_s version of the constant region is expressed. The IgM actually is secreted as a pentamer IgM_5J, in which J is a joining polypeptide (no connection with the J region) that forms disulfide linkages with μ chains. Secretion of the protein is followed by the humoral response depicted in **Figure 25.1**.

The μ_m and μ_s versions of the μ heavy chain differ only at the C-terminal end. The μ_m chain ends in a hydrophobic sequence that probably secures it in the membrane. This sequence is replaced by a shorter hydrophilic sequence in μ_s; the substitution allows the μ heavy chain to pass through the membrane. The change of C-terminus is accomplished by an alternative splicing event, which is controlled by the 3′ end of the nuclear RNA, as illustrated in **Figure 25.16**.

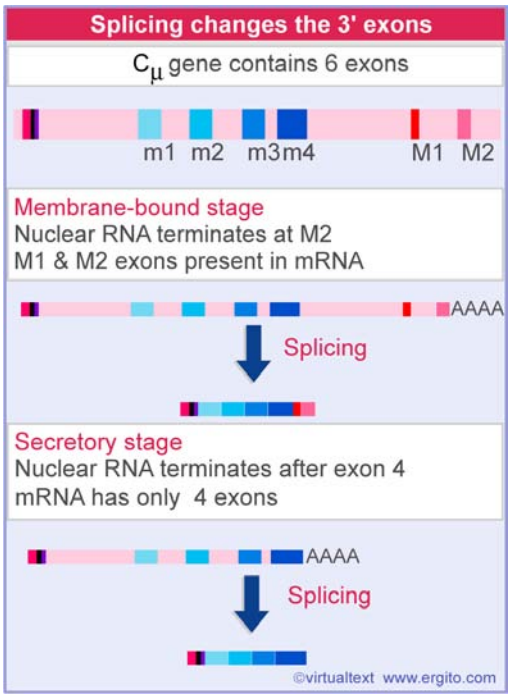

Figure 25.16 The 3′ end controls the use of splicing junctions so that alternative forms of the heavy gene are expressed.

At the membrane-bound stage, the RNA terminates after exon M2, and the constant region is produced by splicing together six exons. The first four exons code for the four domains of the constant region. The last two exons, M1 and M2, code for the 41-residue hydrophobic C-terminal region and its nontranslated trailer. The 5′ splice junction within exon 4 is connected to the 3′ splice junction at the beginning of M1.

At the secreted stage, the nuclear RNA terminates after exon 4. The 5′ splice junction within this exon that had been linked to M1 in the membrane form is ignored. This allows the exon to extend for an additional 20 codons.

A similar transition from membrane to secreted forms is found with other constant regions. The conservation of exon structures suggests that the mechanism is the same.

This content is available online at http://www.ergito.com/main.jsp?bcs=MBIO.5.25.11

IMMUNE DIVERSITY

5.25.12 Class switching is caused by DNA recombination

Key Terms

Class switching describes a change in Ig gene organization in which the C region of the heavy chain is changed but the V region remains the same.

An **S region** is an intron sequence involved in immunoglobulin class switching. S regions consist of repetitive sequences at the 5' ends of gene segments encoding the heavy chain constant regions.

Key Concepts

- Immunoglobulins are divided into five classes according to the type of constant region in the heavy chain.
- Class switching to change the C_H region occurs by a recombination between S regions that deletes the region between the old C_H region and the new C_H region.
- Multiple successive switch recombinations can occur.

The *class* of immunoglobulin is defined by the type of C_H region. **Figure 25.17** summarizes the five Ig classes. IgM (the first immunoglobulin to be produced by any B cell) and IgG (the most common immunoglobulin) possess the central ability to activate complement, which leads to destruction of invading cells. IgA is found in secretions (such as saliva), and IgE is associated with the allergic response and defense against parasites.

There are five types of heavy chain					
Type	IgM	IgD	IgG	IgA	IgE
Heavy chain	μ	δ	γ	α	ε
Structure	$(\mu_2L_2)_5J$	δ_2L_2	γ_2L_2	$(\alpha_2L_2)_2J$	ε_2L_2
Proportion	5%	1%	80%	14%	<1%
Effector function	Activates complement	Development of tolerance (?)	Activates complement	Found in secretions	Allergic response

Figure 25.17 Immunoglobulin type and function is determined by the heavy chain. J is a joining protein in IgM; all other Ig types exist as tetramers.

All lymphocytes start productive life as immature cells engaged in synthesis of IgM. Cells expressing IgM have the germline arrangement of the C_H gene segment cluster shown in **Figure 25.10**. The V-D-J joining reaction triggers expression of the C_μ gene segment. A lymphocyte generally produces only a single class of immunoglobulin at any one time, but the class may change during the cell lineage. A change in expression is called **class switching**. It is accomplished by a substitution in the type of C_H region that is expressed. Switching can be stimulated by

environmental effects; for example, the growth factor TGF β causes switching from C_μ to C_α.

Switching involves only the C_H gene segment; the same V_H gene segment continues to be expressed. So a given V_H gene segment may be expressed successively in combination with more than one C_H gene segment. The same light chain continues to be expressed throughout the lineage of the cell. Class switching therefore allows the type of effector response (mediated by the C_H region) to change, while maintaining the same capacity to recognize antigen (mediated by the V regions).

Changes in the expression of C_H gene segments are made in two ways. The majority occur via further DNA recombination events, involving a system different from that concerned with V-D-J joining (and able to operate only later during B cell development). Another type of change occurs at the level of RNA processing, but generally this is involved with changing the C-terminal sequence of the C_H region rather than its class (see *Molecular Biology 5.25.11 Early heavy chain expression can be changed by RNA processing*).

Cells expressing downstream C_H gene segments have deletions of C_μ and the other gene segments preceding the expressed C_H gene segment. Class switching is accomplished by a recombination to bring a new C_H gene segment into juxtaposition with the expressed V-D-J unit. The sequences of switched V-D-J-C_H units show that the sites of switching lie upstream of the C_H gene segments themselves. The switching sites are called **S regions**. **Figure 25.18** depicts two successive switches.

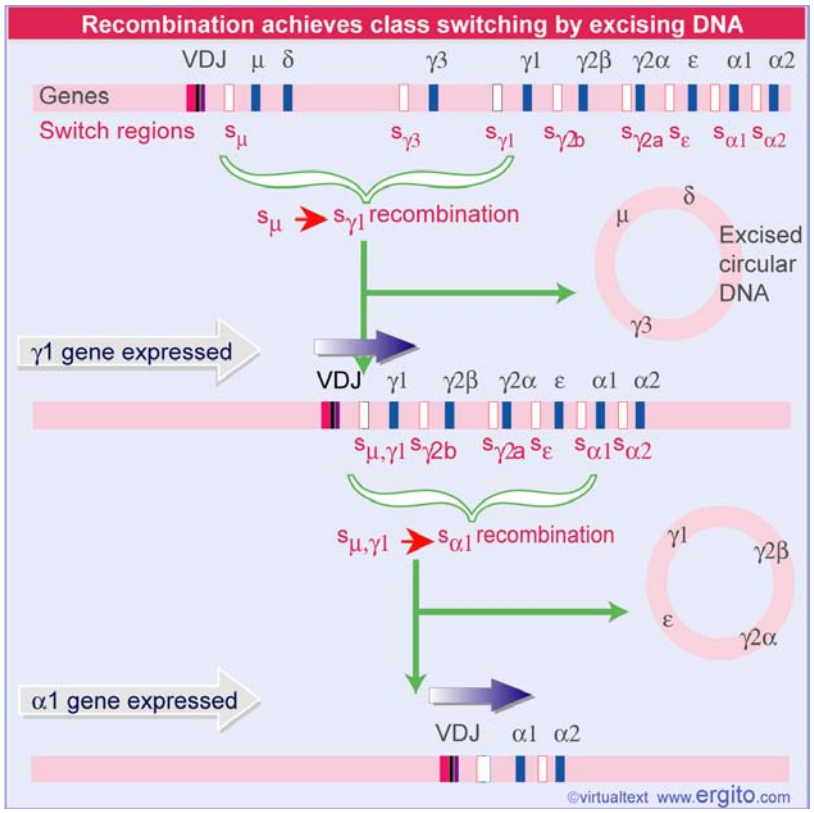

Figure 25.18 Class switching of heavy genes may occur by recombination between switch regions (S), deleting the material between the recombining S sites. Successive switches may occur.

In the first switch, expression of C_{μ} is succeeded by expression of C_{γ_1}. The C_{γ_1} gene segment is brought into the expressed position by recombination between the sites S_{μ} and S_{γ_1}. The S_{μ} site lies between V-D-J and the C_{μ} gene segment. The S_{γ_1} site lies upstream of the C_{γ_1} gene segment. The DNA sequence between the two switch sites is excised as a circular molecule.

The linear deletion model imposes a restriction on the heavy-gene locus: *once a class switch has been made, it becomes impossible to express any C_H gene segment that used to reside between C_{μ} and the new C_H gene segment.* In the example of **Figure 25.18**, cells expressing C_{γ_1} should be unable to give rise to cells expressing C_{γ_3}, which has been deleted.

However, it should be possible to undertake another switch to any C_H gene segment *downstream* of the expressed gene. **Figure 25.18** shows a second switch to C_{α} expression, accomplished by recombination between S_{α_1} and the switch region S_{μ,γ_1} that was generated by the original switch.

We assume that all of the C_H gene segments have S regions upstream of the coding sequences. We do not know whether there are any restrictions on the use of S regions. Sequential switches do occur, but we do not know whether they are optional or an obligatory means to proceed to later C_H gene segments. We should like to know whether IgM can switch directly to *any* other class. Because the S regions lie within

the introns that precede the C_H coding regions, switching does not alter the translational reading frame.

Last updated on 5-2-2002

IMMUNE DIVERSITY

5.25.13 Switching occurs by a novel recombination reaction

Key Concepts

- Switching occurs by a double-strand break followed by the non-homologous end joining reaction.
- The important feature of a switch region is the presence of inverted repeats.
- Switching requires activation of promoters that are upstream of the switch sites.

We know that switch sites are not uniquely defined, because different cells expressing the same C_H gene segment prove to have recombined at different points. Switch regions vary in length (as defined by the limits of the sites involved in recombination) from 1-10 kb. They contain groups of short inverted repeats, with repeating units that vary from 20-80 nucleotides in length. The primary sequence of the switch region does not seem to be important; what matters is the presence of the inverted repeats (2433).

An S region typically is located ~2 kb upstream of a C_H gene segment. The switching reaction releases the excised material between the switch sites as a circular DNA molecule (1208; 1209; 1210). Two of the proteins required for the joining phase of VDJ recombination (and also for the general nonhomologous end-joining pathway, NHEJ), Ku and DNA-PKcs, are required, suggesting that the joining reaction may use the NHEJ pathway (1213; 1214). Basically this implies that the reaction occurs by a double-strand break followed by rejoining of the cleaved ends.

We can put together the features of the reaction to propose a model for the generation of the double-strand break (for review see 2434). The critical points are:

- transcription through the S region is required;
- the inverted repeats are crucial;
- and the break can occur at many different places within the S region.

Figure 25.19 shows the stages of the class switching reaction. A promoter (I) lies immediately upstream of each switch region. Switching requires transcription from this promoter (1211; 1212). The promoter may respond to activators that respond to environmental conditions, such as stimulation by cytokines, thus creating a mechanism to regulate switching. The first stage in switching is therefore to activate the I promoters that are upstream of each of the switch regions that will be involved. When these promoters are activated, they generate sterile transcripts that are spliced to join the I region with the corresponding heavy constant region.

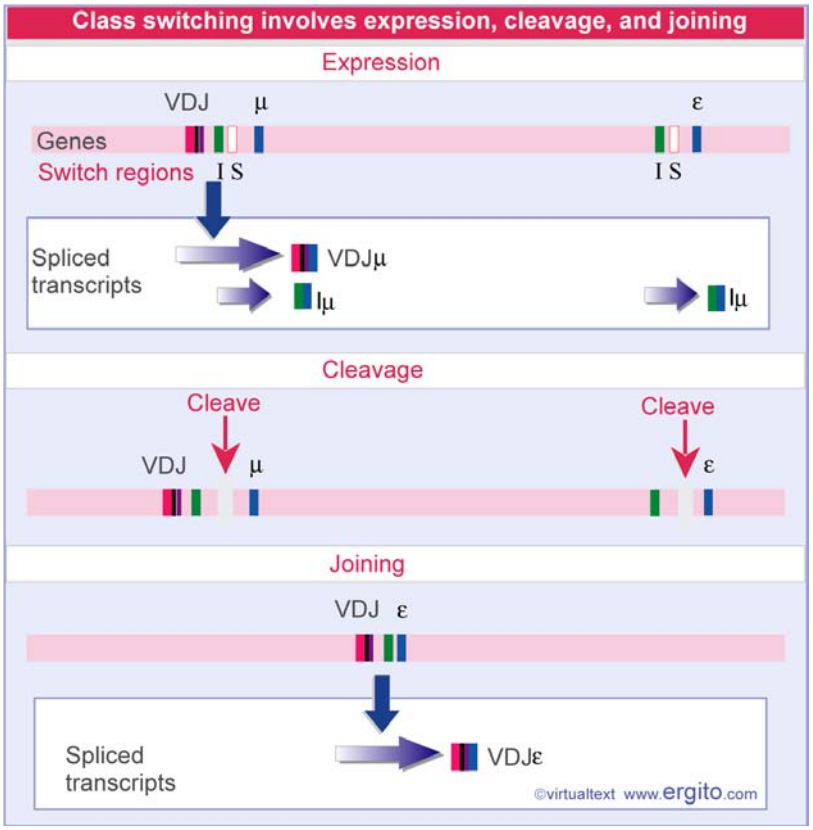

Figure 25.19 Class switching passes through discrete stages. The I promoters initiate transcription of sterile transcripts. The switch regions are cleaved. Joining occurs at the cleaved regions.

The key insight into the mechanism of switching was the discovery of the requirement for the enzyme AID (activation-induced cytidine deaminase) (for review see 4845). In the absence of AID, class switching is blocked before the nicking stage (1206, 1207). Somatic mutation is also blocked, showing an interesting connection between two important processes in immune diversification (see *Molecular Biology 5.25.15 Somatic mutation is induced by cytidine deaminase and uracil glycosylase*).

AID is expressed only at a specific stage during the differentiation of B lymphocytes, restricting the processes of class switching and somatic mutation to this stage. AID is a member of a class of enzymes that act on RNA to change a cytidine to a uridine (see *Molecular Biology 5.26.10 RNA editing occurs at individual bases*). However, AID has a different specificity, and acts on single-stranded DNA (4846).

Another enzyme is also required for both class switching and somatic mutation. This is UNG, a uracil DNA glycosylase, which removes the uracil that AID generates by deaminating cytidine. Mice that are deficient in UNG have a 10-fold reduction in class switching. This suggests a model in which the successive actions of AID and UNG create sites from which a base has been removed in DNA. Different consequences follow in the class switching and somatic mutation systems.

The source of the single-stranded DNA target for AID is generated by the process of

sterile transcription, most probably by exposing the nontemplate strand of DNA that is displaced when the other strand is used as template for RNA synthesis. This is supported by the observation that AID preferentially targets cytidines in the nontemplate strand (4847).

To cause class switching, these sites are converted into breaks in the nucleotide chain that provide the cleavage events shown in **Figure 25.19**. The broken ends are joined by the NHEJ pathway, which is a repair system that acts on double-strand breaks in DNA (see *Molecular Biology 4.15.29 A common system repairs double-strand breaks*). We do not know yet how the abasic sites are converted into double-strand breaks. The MSH system that is involved in repairing mismatches in DNA may be required, because mutations in the gene *MSH2* reduce class switching.

One unexplained feature is the involvement of inverted repeats. One possibility is that hairpins are formed by an interaction between the inverted repeats on the displaced nontemplate strand, as shown in **Figure 25.20**. In conjunction with the generation of abasic sites on this strand, this might lead to breakage.

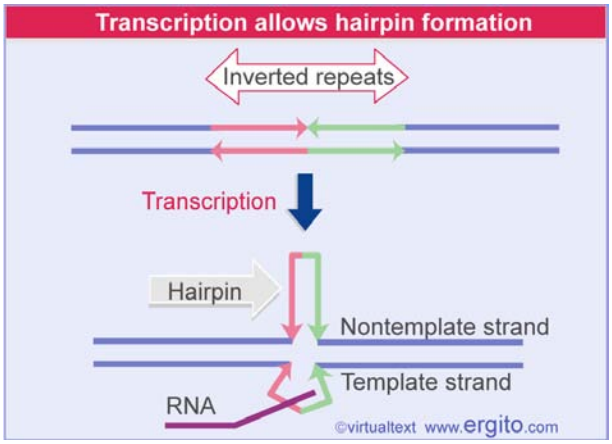

Figure 25.20 When transcription separates the strands of DNA, one strand may form an inverted repeat if the sequence is palindromic.

Another critical question that remains to be answered is how the system is targeted to the appropriate regions in the heavy chain locus, and what controls the use of switching sites.

Last updated on March 16, 2004

Reviews

2434. Honjo, T., Kinoshita, K., and Muramatsu, M. (2002). *Molecular mechanism of class switch recombination: linkage with somatic hypermutation.* Annu. Rev. Immunol. 20, 165-196.

4845. Li, Z., Woo, C. J., Iglesias-Ussel, M. D., Ronai, D., and Scharff, M. D. (2004). *The generation of antibody diversity through somatic hypermutation and class switch recombination.* Genes Dev. 18, 1-11.

References

1206. Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). *Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme*. Cell 102, 553-563.

1207. Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A., Tezcan, I., Ersoy, F., Kayserili, H., Ugazio, A. G., Brousse, N., Muramatsu, M., Notarangelo, L. D., Kinoshita, K., and Honjo, T. (2000). *Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2)*. Cell 102, 565-575.

1208. Iwasato, T., Shimizu, A., Honjo, T., and Yamagishi, H. (1990). *Circular DNA is excised by immunoglobulin class switch recombination.* Cell 62, 143-149.

1209. Matsuoka, M., Yoshida, K., Maeda, T., Usuda, S., and Sakano, H. (1990). *Switch circular DNA formed in cytokine-treated mouse splenocytes: evidence for intramolecular DNA deletion in immunoglobulin class switching.* Cell 62, 135-142.

1210. von Schwedler, U., Jack, H. M., and Wabl, M. (1990). *Circular DNA is a product of the immunoglobulin class switch rearrangement.* Nature 345, 452-456.

1211. Gu, H., Zou, Y. R., and Rajewsky, K. (1993). *Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting.* Cell 73, 1155-1164.

1212. Xu, L., Gorham, B., Li, S. C., Bottaro, A., Alt, F. W., and Rothman, P. (1993). *Replacement of germ-line epsilon promoter by gene targeting alters control of immunoglobulin heavy chain class switching.* Proc. Natl. Acad. Sci. USA 90, 3705-3709.

1213. Rolink, A., Melchers, F., and Andersson, J. (1996). *The SCID but not the RAG-2 gene product is required for S mu-S epsilon heavy chain class switching.* Immunity 5, 319-330.

1214. Manis, J. P., Gu, Y., Lansford, R., Sonoda, E., Ferrini, R., Davidson, L., Rajewsky, K., and Alt, F. W. (1998). *Ku70 is required for late B cell development and immunoglobulin heavy chain class switching.* J. Exp. Med. 187, 2081-2089.

2433. Kinoshita, K., Tashiro, J., Tomita, S., Lee, C. G., and Honjo, T. (1998). *Target specificity of immunoglobulin class switch recombination is not determined by nucleotide sequences of S regions.* Immunity 9, 849-858.

4846. Bransteitter, R., Pham, P., Scharff, M. D., and Goodman, M. F. (2003). *Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase.* Proc. Natl. Acad. Sci. USA 100, 4102-4107.

4847. Pham, P., Bransteitter, R., Petruska, J., and Goodman, M. F. (2003). *Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation.* Nature 424, 103-107.

IMMUNE DIVERSITY

5.25.14 Somatic mutation generates additional diversity in mouse and man

Key Terms

A **somatic mutation** is a mutation occurring in a somatic cell, and therefore affecting only its daughter cells; it is not inherited by descendants of the organism.

Hypermutation describes the introduction of somatic mutations in a rearranged immunoglobulin gene. The mutations can change the sequence of the corresponding antibody, especially in its antigen-binding site.

Hybridoma is a cell line produced by fusing a myeloma with a lymphocyte; it continues indefinitely to express the immunoglobulins of both parents.

Key Concepts

- Active immunoglobulin genes have V regions with sequences that are changed from the germline because of somatic mutation.
- The mutations occur as substitutions of individual bases.
- The sites of mutation are concentrated in the antigen-binding site.
- The process depends on the enhancer that activates transcription at the Ig locus.

Comparisons between the sequences of expressed immunoglobulin genes and the corresponding V gene segments of the germline show that new sequences appear in the expressed population. Some of this additional diversity results from sequence changes at the V-J or V-D-J junctions that occur during the recombination process. However, other changes occur upstream at locations within the variable domain.

Two types of mechanism can generate changes in V gene sequences after rearrangement has generated a functional immunoglobulin gene. In mouse and Man, the mechanism is the induction of **somatic mutations** at individual locations within the gene specifically in the active lymphocyte. The process is sometimes called **hypermutation**. In chicken, rabbit, and pig, a different mechanism uses gene conversion to change a segment of the expressed V gene into the corresponding sequence from a different V gene (see *Molecular Biology 5.25.16 Avian immunoglobulins are assembled from pseudogenes*).

A probe representing an expressed V gene segment can be used to identify all the corresponding fragments in the germline. Their sequences should identify the complete repertoire available to the organism. Any expressed gene whose sequence is different must have been generated by somatic changes.

One difficulty is to ensure that every potential contributor in the germline V gene segments actually has been identified. This problem is overcome by the simplicity of

the mouse λ chain system. A survey of several myelomas producing λ_1 chains showed that many have the sequence of the single germline gene segment. *But others have new sequences that must have been generated by mutation of the germline gene segment.*

To determine the frequency of somatic mutation in other cases, we need to examine a large number of cells in which the same V gene segment is expressed. A practical procedure for identifying such a group is to characterize the immunoglobulins of a series of cells, all of which express an immune response to a particular antigen.

(Epitopes used for this purpose are small molecules – haptens – whose discrete structure is likely to provoke a consistent response, unlike a large protein, different parts of which provoke different antibodies. A hapten is conjugated with a nonreactive protein to form the antigen. The cells are obtained by immunizing mice with the antigen, obtaining the reactive lymphocytes, and sometimes fusing these lymphocytes with a myeloma [immortal tumor] cell to generate a **hybridoma** that continues to express the desired antibody indefinitely.)

In one example, 10 out of 19 different cell lines producing antibodies directed against the hapten phosphorylcholine had the same V_H sequence. This sequence was the germline V gene segment T15, one of four related V_H genes. The other 9 expressed gene segments differed from each other and from all 4 germline members of the family. They were more closely related to the T15 germline sequence than to any of the others, and their flanking sequences were the same as those around T15. This suggested that they arose from the T15 member by somatic mutation (1022).

Figure 25.21 shows that sequence changes are localized around the V gene segment, extending in a region from ~150 bp downstream of the V gene promoter for ~1.5 kb. They take the form of substitutions of individual nucleotide pairs. Usually there are ~3-15 substitutions, corresponding to <10 amino acid changes in the protein. They are concentrated in the antigen-binding site (thus generating the maximum diversity for recognizing new antigens). Only some of the mutations affect the amino acid sequence, since others lie in third-base coding positions as well as in nontranslated regions (for review see 270; 271).

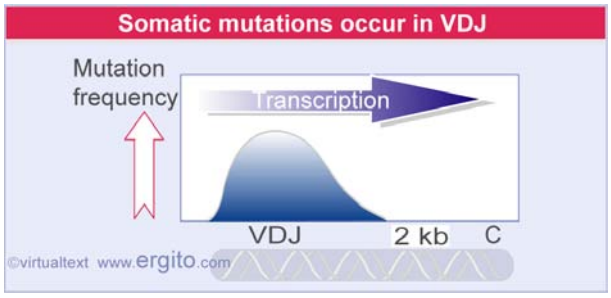

Figure 25.21 Somatic mutation occurs in the region surrounding the V segment and extends over the joined VDJ segments.

The large proportion of ineffectual mutations suggests that somatic mutation occurs more or less at random in a region including the V gene segment and extending beyond it. There is a tendency for some mutations to recur on multiple occasions.

These may represent hotspots as a result of some intrinsic preference in the system.

Somatic mutation occurs during clonal proliferation, apparently at a rate ~10^{-3} per bp per cell generation. Approximately half of the progeny cells gain a mutation; as a result, cells expressing mutated antibodies become a high fraction of the clone.

In many cases, a single family of V gene segments is used consistently to respond to a particular antigen. Upon exposure to an antigen, presumably the V region with highest intrinsic affinity provides a starting point. Then somatic mutation increases the repertoire. Random mutations have unpredictable effects on protein function; some inactivate the protein, others confer high specificity for a particular antigen. The proportion and effectiveness of the lymphocytes that respond is increased by selection among the lymphocyte population for those cells bearing antibodies in which mutation has increased the affinity for the antigen.

Last updated on March 11, 2004

Reviews

270. French, D. L., Laskov, R., and Scharff, M. D. (1989). *The role of somatic hypermutation in the generation of antibody diversity*. Science 244, 1152-1157.

271. Kocks, C. and Rajewsky, K. (1989). *Stable expression and somatic hypermutation of antibody V regions in B-cell developmental pathways*. Annu. Rev. Immunol. 7, 537-559.

References

1022. Kim, S., Davis, M., Sinn, E., Patten, P., and Hood, L. (1981). *Antibody diversity: somatic hypermutation of rearranged VH genes.* Cell 27, 573-581.

IMMUNE DIVERSITY

5.25.15 Somatic mutation is induced by cytidine deaminase and uracil glycosylase

Key Concepts

- A cytidine deaminase is required for somatic mutation as well as for class switching.
- Uracil-DNA glycosylase activity influences the pattern of somatic mutations.
- Hypermutation may be initiated by the sequential action of these enzymes.

Somatic mutation has many of the same requirements as class switching (see *Molecular Biology 5.25.13 Switching occurs by a novel recombination reaction* (1206; 1207; 3433; for review see 4845):

- transcription must occur in the target region (as shown in this case by the demand for the enhancer that activates transcription at each Ig locus (1215);
- it requires the enzymes AID (activation induced cytidine deaminase) and UNG (uracil-DNA-glycosylase);
- the MSH mismatch-repair system is involved.

The way in which the removal of the deaminated base leads to somatic mutation is suggested by the experiment summarized in **Figure 25.22**. When AID deaminates cytosine, it generates uracil, which then is removed from the DNA by UNG. Normally all possible substitutions occur at the abasic site. But if the action of uracil-DNA glycosylase is blocked, we see a different result. If uracil is not removed from DNA, it should pair with adenine during replication. The ultimate result is to replace the original C-G pair with a T-A pair. Uracil-DNA glycosylase can be blocked by introducing into cells the gene coding for a protein that inhibits the enzyme. (The gene is a component of the bacteriophage PSB-2, whose genome is unusual in containing uracil, so that the enzyme needs to be blocked during a phage infection.) When the gene is introduced into a lymphocyte cell line, there is a dramatic change in the pattern of mutations, with almost all comprising the predicted transition from C-G to A-T (2936).

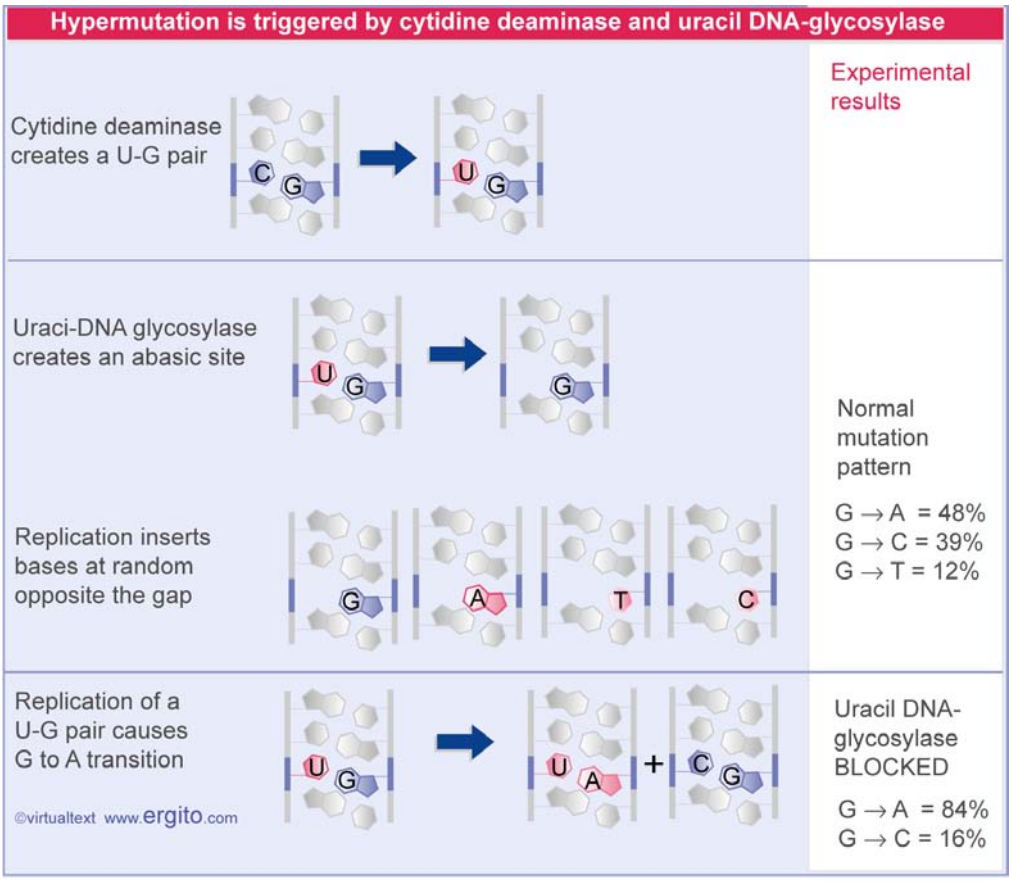

Figure 25.22 When the action of cytidine deaminase (top) is followed by that of uracil DNA-glycosylase, an abasic site is created. Replication past this site should insert all four bases at random into the daughter strand (center). If the uracil is not removed from the DNA, its replication causes a C-G to T-A transition.

The key event in generating a random spectrum of mutations is therefore to create the abasic site. Then the MSH repair system is recruited to excise and replace the stretch of DNA containing the damage. The simplest possibility is that the replacement is performed by an error-prone DNA polymerase, mutations may be introduced. Another possibility is that so many abasic sites are created that the repair systems are overwhelmed. When replication occurs, this could lead to the random insertion of bases opposite the abasic sites. We don't know yet what restricts the action of this system to the target region for hypermutation.

The difference in the systems is at the end of the process, when double-strand breaks are introduced in class switching, but individual point mutations are created during somatic mutation. We do not yet know exactly where the systems diverge. One possibility is that breaks are introduced at abasic sites in class switching, but the sites are erratically repaired in somatic mutation. Another possibility is that breaks are introduced in both cases, but are repaired in an error-prone manner in somatic mutation (for review see 2434).

Last updated on March 16, 2004

Reviews

2434. Honjo, T., Kinoshita, K., and Muramatsu, M. (2002). *Molecular mechanism of class switch recombination: linkage with somatic hypermutation.* Annu. Rev. Immunol. 20, 165-196.

3433. Kinoshita, K. and Honjo, T. (2001). *Linking class-switch recombination with somatic hypermutation.* Nat. Rev. Mol. Cell Biol. 2, 493-503.

4845. Li, Z., Woo, C. J., Iglesias-Ussel, M. D., Ronai, D., and Scharff, M. D. (2004). *The generation of antibody diversity through somatic hypermutation and class switch recombination.* Genes Dev. 18, 1-11.

References

1206. Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). *Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme*. Cell 102, 553-563.

1207. Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A., Tezcan, I., Ersoy, F., Kayserili, H., Ugazio, A. G., Brousse, N., Muramatsu, M., Notarangelo, L. D., Kinoshita, K., and Honjo, T. (2000). *Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2)*. Cell 102, 565-575.

1215. Peters, A. and Storb, U. (1996). *Somatic hypermutation of immunoglobulin genes is linked to transcription initiation.* Immunity 4, 57-65.

2936. Di Noia, J. and Neuberger, J. (2002). *Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase.* Nature 419, 43-48.

IMMUNE DIVERSITY

5.25.16 Avian immunoglobulins are assembled from pseudogenes

Key Concepts

- An immunoglobulin gene in chicken is generated by copying a sequence from one of 25 pseudogenes into the V gene at a single active locus.

The chick immune system is the paradigm for rabbits, cows, and pigs, which rely upon using the diversity that is coded in the genome. A similar mechanism is used by both the single light chain locus (of the λ type) and the H chain locus. The organization of the λ locus is drawn in **Figure 25.23**. It has only one functional V gene segment, J segment, and C gene segment. Upstream of the functional V λ_1 gene segment lie 25 V λ pseudogenes, organized in either orientation. They are classified as pseudogenes because either the coding segment is deleted at one or both ends, or proper signals for recombination are missing (or both). This assignment is confirmed by the fact that only the V λ_1 gene segment recombines with the J-C λ gene segment (772).

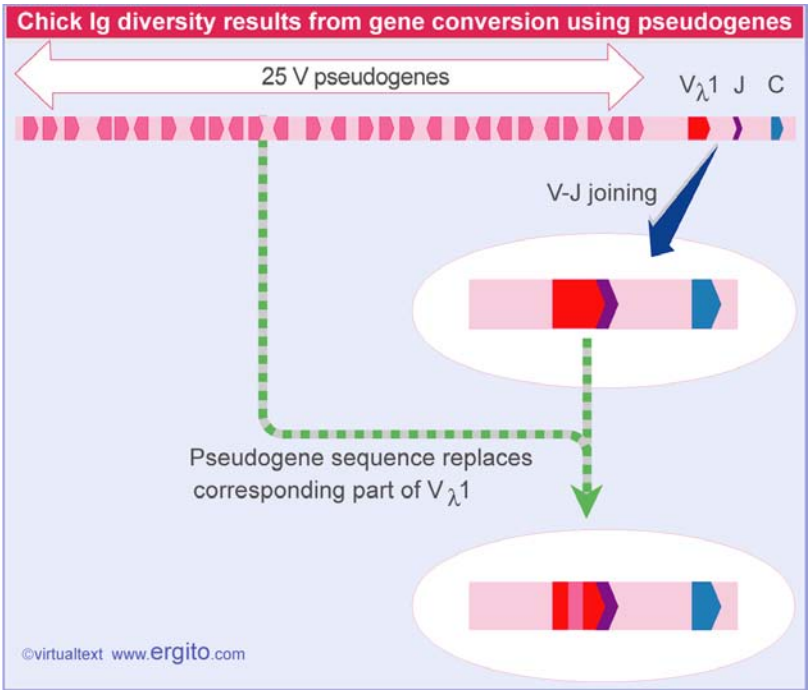

Figure 25.23 The chicken lambda light locus has 25 V pseudogenes upstream of the single functional V-J-C region. But sequences derived from the pseudogenes are found in active rearranged V-J-C genes.

But sequences of active rearranged V λ-J-C λ gene segments show considerable diversity! A rearranged gene has one or more positions at which a cluster of changes

has occurred in the sequence. A sequence identical to the new sequence can almost always be found in one of the pseudogenes (which themselves remain unchanged). The exceptional sequences that are not found in a pseudogene always represent changes at the junction between the original sequence and the altered sequence.

So a novel mechanism is employed to generate diversity. Sequences from the pseudogenes, between 10 and 120 bp in length, are substituted into the active $V_{\lambda 1}$ region by gene conversion. The unmodified $V_{\lambda 1}$ sequence is not expressed, even at early times during the immune response. A successful conversion event probably occurs every 10-20 cell divisions to every rearranged $V_{\lambda 1}$ sequence. At the end of the immune maturation period, a rearranged $V_{\lambda 1}$ sequence has 4-6 converted segments spanning its entire length, derived from different donor pseudogenes. If all pseudogenes participate, this allows 2.5×10^8 possible combinations!

The enzymatic basis for copying pseudogene sequences into the expressed locus depends on enzymes involved in recombination and is related to the mechanism for somatic hypermutation that introduces diversity in mouse and man (2154). Some of the genes involved in recombination are required for the gene conversion process; for example, it is prevented by deletion of *RAD54*. Deletion of other recombination genes (*XRCC2*, *XRCC3*, and *RAD51B*) has another, very interesting effect: somatic mutation occurs at the V gene in the expressed locus. The frequency of the somatic mutation is ~10× greater than the usual rate of gene conversion.

These results show that the absence of somatic mutation in chick is not due to a deficiency in the enzymatic systems that are responsible in mouse and man. The most likely explanation for a connection between (lack of) recombination and somatic mutation is that unrepaired breaks at the locus trigger the induction of mutations. The reason why somatic mutation occurs in mouse and man but not in chick may therefore lie with the details of the operation of the repair system that operates on breaks at the locus. It is more efficient in chick, so that the gene is repaired by gene conversion before mutations can be induced.

Last updated on 10-23-2001

References

772. Reynaud, C. A., Anquez, V., Grimal, H., and Weill, J. C. (1987). *A hyperconversion mechanism generates the chicken light chain preimmune repertoire*. Cell 48, 379-388.

2154. Sale, J. E., Calandrini, D. M., Takata, M., Takeda, S., and Neuberger, M. S. (2001). *Ablation of XRCC2/3 transforms immunoglobulin V gene conversion into somatic hypermutation.* Nature 412, 921-926.

IMMUNE DIVERSITY

5.25.17 B cell memory allows a rapid secondary response

Key Terms

The **primary immune response** is an organism's immune response upon first exposure to a given antigen. It is characterized by a relatively shorter duration and lower affinity antibodies than in the secondary immune response.

The **secondary immune response** is an organism's immune response upon a second exposure to a given antigen. This second exposure is also referred to as a "booster". The secondary immune response is characterized by a more rapid induction, greater magnitude, and higher affinity antibodies than the primary immune response.

B cell memory is responsible for rapid antibody production during a secondary immune response and subsequent responses. Memory B cells produce antibodies of higher affinity than naive B cells.

The **B cell receptor (BCR)** is the antigen receptor complex on the cell surface of B lymphocytes. It consists of membrane-bound immunoglobulin bound noncovalently to Ig α and Ig β chains.

Key Concepts

- The primary response to an antigen is mounted by B cells that do not survive beyond the response period.
- Memory B cells are produced that have specificity for the same antigen but that are inactive.
- A reexposure to antigen triggers the secondary response in which the memory cells are rapidly activated.

We are now in a position to summarize the relationship between the generation of high-affinity antibodies and the differentiation of the B cell. **Figure 25.24** shows that B cells are derived from a self-renewing population of stem cells in the bone marrow. Maturation to give B cells depends upon Ig gene rearrangement, which requires the functions of the *SCID* and *RAG1,2* (and other) genes. If gene rearrangement is blocked, mature B cells are not produced. The antibodies carried by the B cells have specificities determined by the particular combinations of V(D)J regions, and any additional nucleotides incorporated during the joining process.

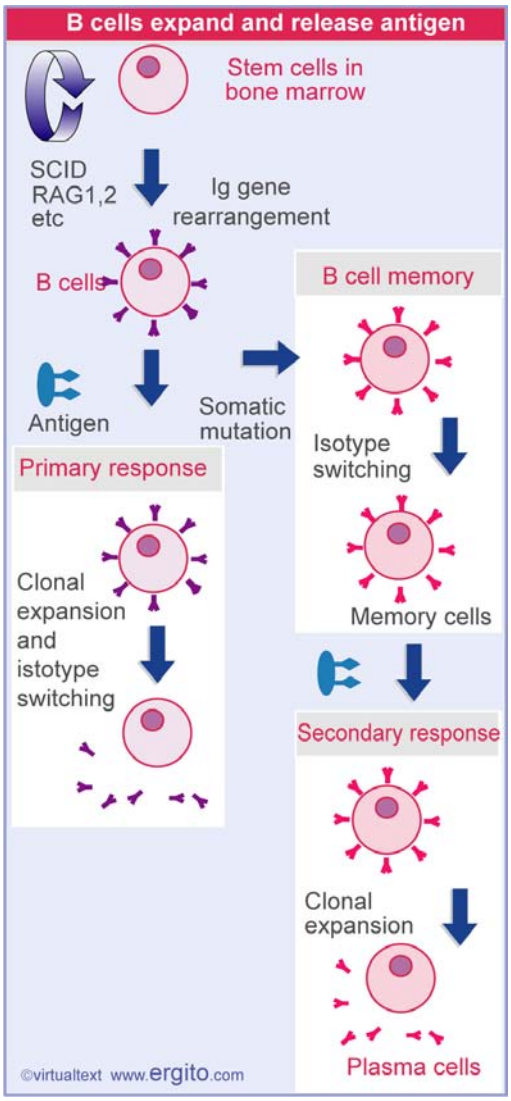

Figure 25.24 B cell differentiation is responsible for acquired immunity. Pre-B cells are converted to B cells by Ig gene rearrangement. Initial exposure to antigen provokes both the primary response and storage of memory cells. Subsequent exposure to antigen provokes the secondary response of the memory cells.

Exposure to antigen triggers two aspects of the immune response. The **primary immune response** occurs by clonal expansion of B cells responding to the antigen. This generates a large number of plasma cells that are specific for the antigen; isotype switching occurs to generate the appropriate type of effector response. The population of cells concerned with the primary response is a dead end; these cells do not live beyond the primary response itself.

Provision for a **secondary immune response** is made through the phenomenon of **B cell memory**. Somatic mutation generates B cells that have increased affinity for the antigen. These cells do not trigger an immune response at this time, although they may undergo isotype switching to select other forms of C_H region. They are stored as

memory cells, with appropriate specificity and effector response type, but are inactive. They are activated if there is a new exposure to the same antigen. Because they are pre-selected for the antigen, they enable a secondary response to be mounted very rapidly, simply by clonal expansion; no further somatic mutation or isotype switching occurs during the secondary response (for review see 276).

The pathways summarized in **Figure 25.24** show the development of acquired immunity, that is, the response to an antigen. In addition to these cells, there is a separate set of B cells, named the Ly-1 cells. These cells have gone through the process of V gene rearrangement, and apparently are selected for expression of a particular repertoire of antibody specificities. They do not undergo somatic mutation or the memory response. They may be involved in natural immunity, that is, an intrinsic ability to respond to certain antigens.

A more detailed view of B cell development is shown in **Figure 25.25**. The first step is recombination between the D and J segments of the μ heavy chain. This is succeeded by V-D recombination, generating a μ heavy chain. Several recombination events, involving a succession of nonproductive and productive rearrangements, may occur, as shown previously in **Figure 25.15**. These cells express a protein resembling a λ chain, called the surrogate light (SL) chain, which is expressed on the surface and associates with the μ heavy chain to form the pre-B-receptor. It resembles an immunoglobulin complex, but does not function as one.

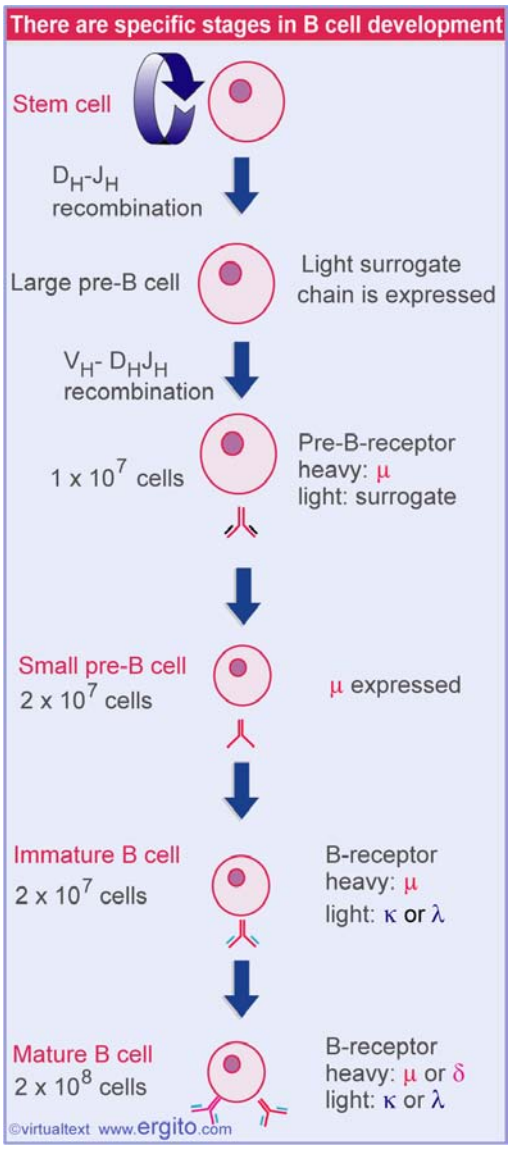

Figure 25.25 B cell development proceeds through sequential stages.

The production of μ chain represses synthesis of SL chain, and the cells divide to become small pre-B cells. Then light chain is expressed and functional immunoglobulin appears on the surface of the immature B cells. Further cell divisions occur, and the expression of δ heavy chain is added to that of μ chain, as the cells mature into B cells.

Immunoglobulins function both by secretion from B cells and by surface expression. **Figure 25.26** shows that the active complex on the cell surface is called the **B cell receptor** (BCR), and consists of an immunoglobulin associated with transmembrane proteins called Ig α and Ig β. They provide the signaling components that trigger intracellular pathways in response to antigen-antibody binding. The activation of the BCR is also influenced by interactions with other receptors, for example, to mediate the interaction of antigen-activated B cells with helper T cells.

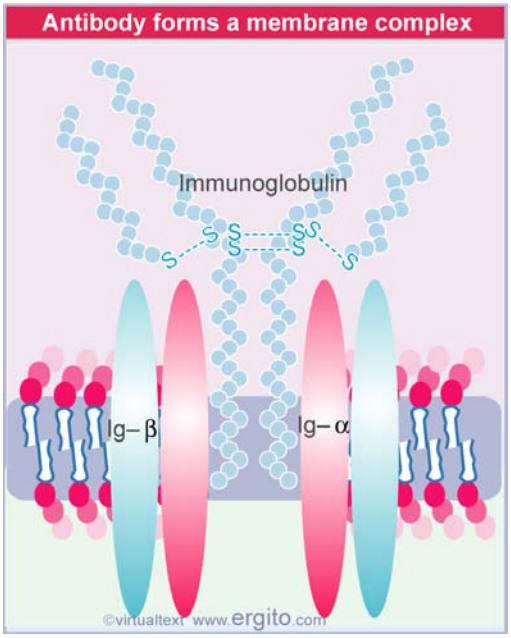

Figure 25.26 The B cell antigen receptor consists of an immunoglobulin tetramer (H_2L_2) linked to two copies of the signal-transducing heterodimer (Ig α β).

Reviews

276. Rajewsky, K. (1996). *Clonal selection and learning in the antibody system*. Nature 381, 751-758.

IMMUNE DIVERSITY

5.25.18 T cell receptors are related to immunoglobulins

Key Concepts

- T cells use a similar mechanism of V(D)J-C joining to B cells to produce either of two types of T cell receptor.
- TCR α β is found on >95% of T lymphocytes, and TCR γ δ is found on <5%.

The lymphocyte lineage presents an example of evolutionary opportunism: a similar procedure is used in both B cells and T cells to generate proteins that have a variable region able to provide significant diversity, while constant regions are more limited and account for a small range of effector functions. T cells produce either of two types of T cell receptor. The different T cell receptors are synthesized at different times during T cell development, as summarized in **Figure 25.27** (for review see 266; 268).

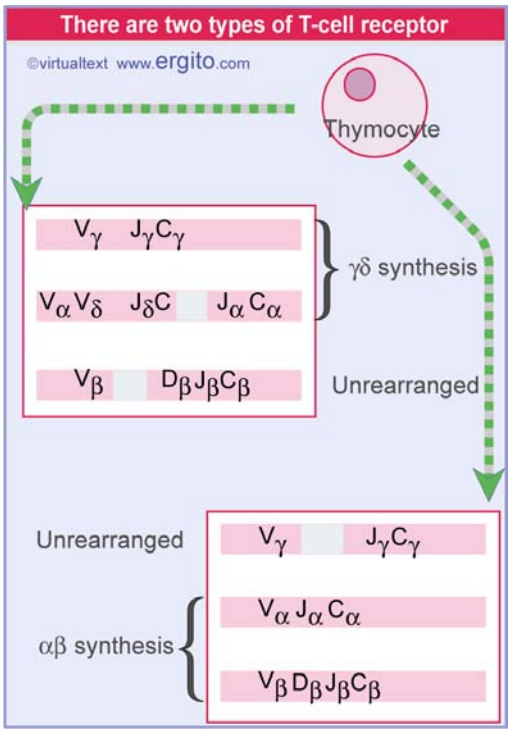

Figure 25.27 The γ δ receptor is synthesized early in T-cell development. TCR α β is synthesized later and is responsible for "classical" cell-mediated immunity, in which target antigen and host histocompatibility antigen are recognized together.

The γ δ receptor is found on <5% of T lymphocytes. It is synthesized only at an early

stage of T cell development. In mice, it is the only receptor detectable at <15 days of gestation, but has virtually been lost by birth at day 20.

TCR α β is found on >95% of lymphocytes. It is synthesized later in T cell development than γ δ. In mice, it first becomes apparent at 15–17 days after gestation. By birth it is the predominant receptor. It is synthesized by a separate lineage of cells from those involved in TCR γ δ synthesis, and involves independent rearrangement events.

Like immunoglobulins, a TCR must recognize a foreign antigen of unpredictable structure. The problem of antigen recognition by B cells and T cells is resolved in the same way, and the organization of the T cell receptor genes resembles the immunoglobulin genes in the use of variable and constant regions. *Each locus is organized in the same way as the immunoglobulin genes, with separate segments that are brought together by a recombination reaction specific to the lymphocyte.* The components are the same as those found in the three Ig families.

The organization of the TCR proteins resembles that of the immunoglobulins. The V sequences have the same general internal organization in both Ig and TCR proteins. The TCR C region is related to the constant Ig regions and has a single constant domain followed by transmembrane and cytoplasmic portions. Exon-intron structure is related to protein function.

The resemblance of the organization of TCR genes with the Ig genes is striking. As summarized in **Figure 25.28**, the organization of TCR α resembles that of Ig κ, with V gene segments separated from a cluster of J segments that precedes a single C gene segment. The organization of the locus is similar in both Man and mouse, with some differences only in the number of V_α gene segments and J_α segments. (In addition to the α segments, this locus also contains δ segments, which we discuss shortly (for review see 273).

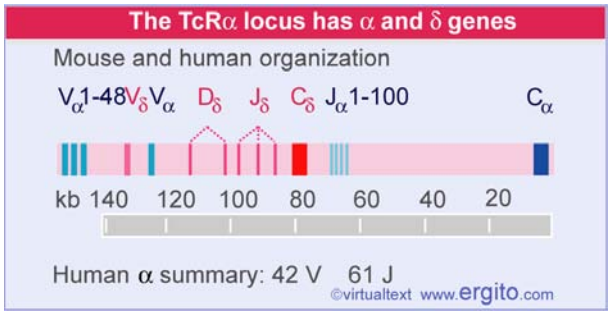

Figure 25.28 The human TCR α locus has interspersed α and δ segments. A V_δ segment is located within the V_α cluster. The D-J-C_δ segments lie between the V gene segments and the J-C_α segments. The mouse locus is similar, but has more V_δ segments.

The components of TCR β resemble those of IgH. **Figure 25.29** shows that the organization is different, with V gene segments separated from two clusters each containing a D segment, several J segments, and a C gene segment. Again the only differences between Man and mouse are in the numbers of the V_β and J_β units.

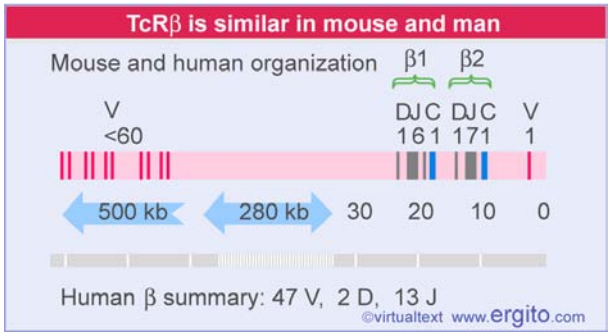

Figure 25.29 The TCR β locus contains many V gene segments spread over ~500 kb, and lying ~280 kb upstream of the two D-J-C clusters.

Diversity is generated by the same mechanisms as in immunoglobulins. Intrinsic diversity results from the combination of a variety of V, D, J, and C segments; some additional diversity results from the introduction of new sequences at the junctions between these components (in the form of P and N nucleotides; see **Figure 25.13**). Some TCR β chains incorporate two D segments, generated by D-D joins (directed by an appropriate organization of the nonamer and heptamer sequences). A difference between TCR and Ig is that somatic mutation does not occur at the TCR loci. Measurements of the extent of diversity show that the 10^{12} T cells in Man contain 2.5×10^7 different α chains associated with 10^6 different β chains.

The same mechanisms are likely to be involved in the reactions that recombine Ig genes in B cells and TCR genes in T cells. The recombining TCR segments are surrounded by nonamer and heptamer consensus sequences identical to those used by the Ig genes. This argues strongly that the same enzymes are involved. Most rearrangements probably occur by the deletion model (see **Figure 25.12**). We do not know how the process is controlled so that Ig loci are rearranged in B cells, while T cell receptors are rearranged in T cells.

The organization of the γ locus resembles that of Ig λ, with V gene segments separated from a series of J-C segments. **Figure 25.30** shows that this locus has relatively little diversity, with ~8 functional V segments. The organization is different in Man and mouse. Mouse has 3 functional J-C loci, but some segments are inverted in orientation. Man has multiple J segments for each C gene segment.

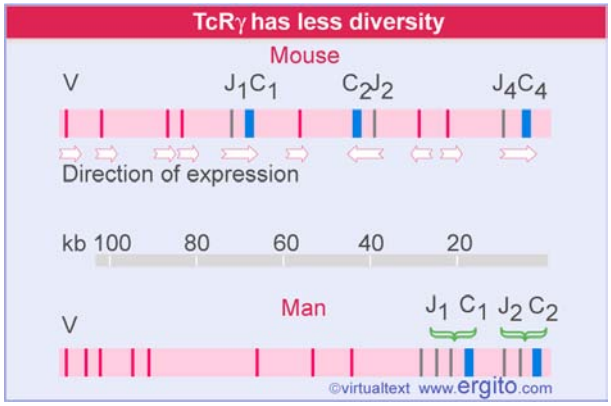

Figure 25.30 The TCR γ locus contains a small number of functional V gene segments (and also some pseudogenes; not shown), lying upstream of the J-C loci.

The δ subunit is coded by segments that lie at the TCR α locus, as illustrated previously in **Figure 25.28**. The segments D δ-D δ-J δ-C δ lie between the V gene segments and the J α-C α segments. Both of the D segments may be incorporated into the δ chain to give the structure VDDJ. The nature of the V gene segments used in the δ rearrangement is an interesting question. Very few V sequences are found in active TCR δ chains. In man, only one V gene segment is in general use for δ rearrangement. In mouse, several V δ segments are found; some are unique for δ rearrangement, but some are also found in α rearrangements. The basis for specificity in choosing V segments in α and δ rearrangement is not known. One possibility is that many of the V α gene segments can be joined to the DDJ δ segment, but that only some (therefore defined as V δ) can give active proteins (for review see 272).

While for the present we have labeled the V segments that are found in δ chains as V δ gene segments, we must reserve judgment on whether they are really unique to δ rearrangement. The interspersed arrangement of genes implies that synthesis of the TCR α β receptor and the γ δ receptor is mutually exclusive at any one allele, because the δ locus is lost entirely when the V α-J α rearrangement occurs.

Rearrangements at the TCR loci, like those of immunoglobulin genes, may be productive or nonproductive. The β locus shows allelic exclusion in much the same way as immunoglobulin loci; rearrangement is suppressed once a productive allele has been generated. The α locus may be different; several cases of continued rearrangement suggest the possibility that substitution of V α sequences may continue after a productive allele has been generated.

Reviews

266. Kronenberg, M., Siu, G., Hood, L. E., and Shastri, N. (1986). *The molecular genetics of the T-cell antigen receptor and T-cell antigen recognition*. Annu. Rev. Immunol. 4, 529-591.

268. Marrack, P. and Kappler, J. (1987). *The T-cell receptor*. Science 238, 1073-1079.

272. Raulet, D. H. (1989). *The structure, function, and molecular genetics of the gamma/delta T-cell receptor*. Annu. Rev. Immunol. 7, 175-207.

273. Davis, M. M. (1990). *T-cell receptor gene diversity and selection*. Annu. Rev. Biochem. 59, 475-496.

IMMUNE DIVERSITY

5.25.19 The T cell receptor functions in conjunction with the MHC

Key Terms

A **cytotoxic T cell** is a T lymphocyte (usually $CD8^+$) that can be stimulated to kill cells containing intracellular pathogens, such as viruses.

A **helper T cell** is a T lymphocyte that activates macrophages and stimulates B cell proliferation and antibody production. Helper T cells usually express cell surface CD4 but not CD8.

A **DP thymocyte** is a double positive thymocyte. It is an immature T cell that expresses cell surface CD4 and CD8. Selection of DP thymocytes in the thymus yields mature T cells expressing either CD4 or CD8.

CD3 is a complex of proteins that associates with the T cell antigen receptor's α and β chains. Each complex consists of one each of the δ, ε, γ chains and two ζ chains.

Key Concepts

- The TCR recognizes a short peptide that is bound in a groove of an MHC protein on the surface of the presenting cell.

T cells with α β receptors are divided into several subtypes that have a variety of functions connected with interactions between cells involved in the immune response. **Cytotoxic T cells** (also known as killer T cells) possess the capacity to lyse an infected target cell. **Helper T cells** assist T cell-mediated target killing or B cell-mediated antibody-antigen interaction.

A major difference between the antibodies of B cells and the receptors of T cells is the way that they handle the antigen. An antibody recognizes a short region (an epitope) within the antigen. The T cell receptor binds to a small peptide (4-5 amino acids long) that has been processed from the antigen by cleavage reactions. The peptide fragment is "presented" to the T cell by an MHC protein. So the T cell simultaneously recognizes the foreign antigen and an MHC protein carried by the presenting cell, as illustrated previously in **Figure 25.2**. Both helper T cells and killer T cells work in this way, but they have different requirements for the presentation of antigen; different types of MHC protein are used in each case (see *Molecular Biology 5.25.20 The major histocompatibility locus codes for many genes of the immune system*). Helper T cells require antigen to be presented by an MHC class II protein, while killer T cells require antigen to be presented by an MHC class I protein.

The TCR α β receptor is responsible for helper T cell function in humoral immunity, and for killer T cell function in cell-mediated immunity. This places upon it the responsibility of recognizing both the foreign antigen and the host MHC protein. The MHC protein binds a short peptide derived from the foreign antigen, and the TCR

then recognizes the peptide in a groove on the surface of the MHC. The MHC is said to *present* the peptide to the TCR. (The peptide is generated when the proteasome degrades the foreign protein to generate fragments of 8-10 residues long, as described in *Molecular Biology Supplement 8.32 The proteasome is a large machine that degrades ubiquitinated proteins*.) A given TCR has specificity for a particular MHC as well as for the foreign antigen. The basis for this dual capacity is one of the most interesting issues to be defined about the α β TCR.

Recombination to generate functional TCR chains is linked to the development of the T lymphocyte, as summarized in **Figure 25.31**. The first stage is rearrangement to form an active TCR β chain. This binds a nonrearranging surrogate α chain, called preTCR α. At this stage, the lymphocyte has not expressed either of the surface proteins CD4 or CD8. The preTCR heterodimer then binds to the CD3 signaling complex (see below). Signaling from the complex triggers several rounds of cell division, during which α chains are rearranged, and the CD4 and CD8 genes are turned on, so that the lymphocyte is converted from $CD4^-CD8^-$ to $CD4^+CD8^+$. This point in development is called β *selection*. It generates **DP thymocytes**.

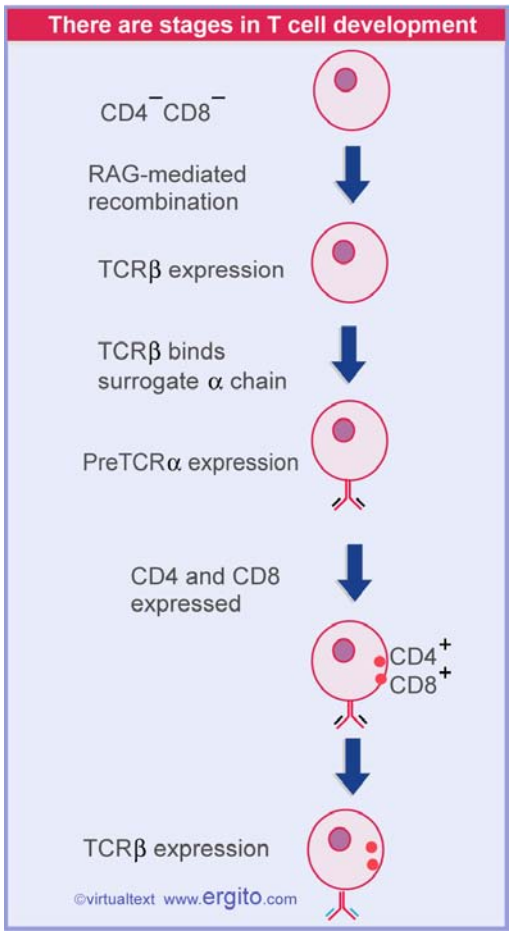

Figure 25.31 T cell development proceeds through sequential stages.

α chain rearrangement continues in the DP thymocytes. The maturation process continues by both positive selection (for mature TCR complexes able to bind a

ligand) and by negative selection (against complexes that interact with inappropriate – self – ligands). Both types of selection involve interaction with MHC proteins. The DP thymocytes either die after ~3 days or become mature lymphocytes as the result of the selective processes. The surface TCR α β heterodimer becomes crosslinked on the surface during positive selection (which rescues the thymocytes from cell death), and then, if they survive the subsequent negative selection, allows them to give rise to the separate T lymphocyte classes which are $CD4^+CD8^-$ and $CD4^-CD8^+$ (907).

The T cell receptor is associated with a complex of proteins called **CD3**, which is involved in transmitting a signal from the surface of the cell to the interior when its associated receptor is activated by binding antigen. Our present picture of the components of the receptor complex on a T cell is illustrated in **Figure 25.32**. The important point is that the interaction of the TCR variable regions with antigen causes the ζ subunits of the CD3 complex to activate the T cell response. The activation of CD3 provides the means by which either α β or γ δ TCR signals that it has recognized an antigen. This is comparable to the constitution of the B cell receptor, in which immunoglobulin associates with the Ig α β signaling chains (see **Figure 25.26**).

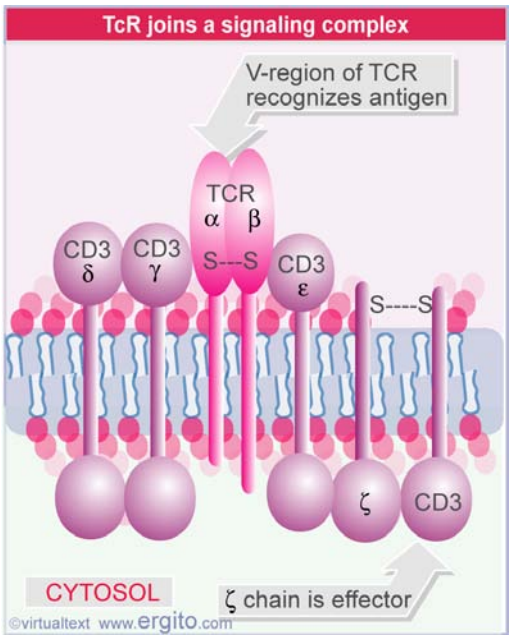

Figure 25.32 The two chains of the T-cell receptor associate with the polypeptides of the CD3 complex. The variable regions of the TCR are exposed on the cell surface. The cytoplasmic domains of the ζ chains of CD3 provide the effector function.

A central dilemma about T cell function remains to be resolved. Cell-mediated immunity requires two recognition processes. Recognition of the foreign antigen requires the ability to respond to novel structures. Recognition of the MHC protein is of course restricted to one of those coded by the genome, but, even so, there are many different MHC proteins. So considerable diversity is required in both recognition reactions. Although helper and killer T cells rely upon different classes

of MHC proteins, they use the same pool of α and β gene segments to assemble their receptors. Even allowing for the introduction of additional variation during the TCR recombination process, it is not clear how enough different versions of the T cell receptor are made available to accommodate all these demands.

Last updated on 1-11-2000

Reviews

907. Goldrath, A. W. and Bevan, M. J. (1999). *Selecting and maintaining a diverse T cell repertoire.* Nature 402, 255-262.

IMMUNE DIVERSITY

5.25.20 The major histocompatibility locus codes for many genes of the immune system

Key Terms

The **H2 locus** is the mouse major histocompatibility complex, a cluster of genes on chromosome 17. The genes encode proteins for antigen presentation, cytokines, and complement proteins.

The **HLA** locus is the human major histocompatibility complex, a cluster of genes on chromosome 6. The genes encode proteins for antigen presentation, cytokines, and complement proteins.

An **MHC class I protein** mostly presents, to $CD8^+$ T cells, peptides that are produced by proteolytic degradation in the cytosol.

Transplantation antigen is protein coded by a major histocompatibility locus, present on all mammalian cells, involved in interactions between lymphocytes.

An **MHC class II protein** mostly presents, to $CD4^+$ T cells, peptides that are produced by proteolytic degradation in the endocytic pathway.

Two mutants are said to **complement** each other when a diploid that is heterozygous for each mutation produces the wild type phenotype.

The **external domain** is the part of a plasma membrane protein that extends outside of the cell. Upon internalization, the protein's external domain extends into the lumen (the topological equivalent of the outside of the cell) of an organelle.

The **transmembrane region (transmembrane domain)** is the part of a protein that spans the membrane bilayer. It is hydrophobic and in many cases contains approximately 20 amino acids that form an α -helix. It is also called the transmembrane domain.

The **cytoplasmic domain** is the part of a transmembrane protein that is exposed to the cytosol.

Key Concepts

- The MHC locus codes for the class I and class II proteins as well as for other proteins of the immune system.
- Class I proteins are the transplantation antigens that are responsible for distinguishing "self" from "nonself" tissue.
- An MHC class I protein is active as a heterodimer with β_2 microglobulin.
- Class II proteins are involved in interactions between T cells.
- An MHC class II protein is a heterodimer of α and β chains.

The major histocompatibility locus occupies a small segment of a single chromosome in the mouse (where it is called the **H2 locus**) and in man (called the

HLA locus). Within this segment are many genes coding for functions concerned with the immune response. At those individual gene loci whose products have been identified, many alleles have been found in the population; the locus is described as highly *polymorphic*, meaning that individual genomes are likely to be different from one another. Genes coding for certain other functions also are located in this region.

Histocompatibility antigens are classified into three types by their immunological properties. In addition, other proteins found on lymphocytes and macrophages have a related structure and are important in the function of cells of the immune system:

MHC class I proteins are the **transplantation antigens**. They are present on every cell of the mammal. As their name suggests, these proteins are responsible for the rejection of foreign tissue, which is recognized as such by virtue of its particular array of transplantation antigens. In the immune system, their presence on target cells is required for the cell-mediated response. The types of class I proteins are defined serologically (by their antigenic properties). The murine class I genes code for the H2-K and H2-D/L proteins. Each mouse strain has one of several possible alleles for each of these functions. The human class I functions include the classical transplantation antigens, HLA-A, B, C.

MHC class II proteins are found on the surfaces of both B and T lymphocytes as well as macrophages. These proteins are involved in communications between cells that are necessary to execute the immune response; in particular, they are required for helper T cell function. The murine class II functions are defined genetically as I-A and I-E. The human class II region (also called HLA-D) is arranged into four subregions, DR, DQ, DZ/DO, DP.

The **complement** proteins are coded by a genetic locus that is also known as the S region; S stands for serum, indicating that the proteins are components of the serum. Their role is to interact with antibody-antigen complexes to cause the lysis of cells in the classical pathway of the humoral response.

The *Qa* and *Tla* loci proteins are found on murine hematopoietic cells. They are known as differentiation antigens, because each is found only on a particular subset of the blood cells, presumably related to their function. They are structurally related to the class I H2 proteins, and like them are polymorphic.

We can now relate the types of proteins to the organization of the genes that code for them. The MHC region was originally defined by genetics in the mouse, where the classical H2 region occupies 0.3 map units. Together with the adjacent region where mutations affecting immune function are also found, this corresponds to a region of ~2000 kb of DNA (for review see 262; 265). The MHC region has been completely sequenced in several mammals, and also some birds and fish. By comparing these sequences, we find that the organization has been generally conserved (for review see 3772).

The gene organization in mouse and Man is summarized in **Figure 25.33**. The genomic regions where the class I and class II genes are located mark the original boundaries of the locus (going in the direction from telomere to centromere, right to left as shown in the figure). The genes in the region that separates the class I and class II genes code for a variety of functions; this is called the class III region.

Defining the ends of the locus varies with the species, and the region beyond the class I genes on the telomeric side is called the extended class I region. Similarly, the region beyond the class II genes on the centromeric side is called the extended class II region. The major difference between mouse and Man is that the extended class II region contains some class I (H2-K) genes in the mouse.

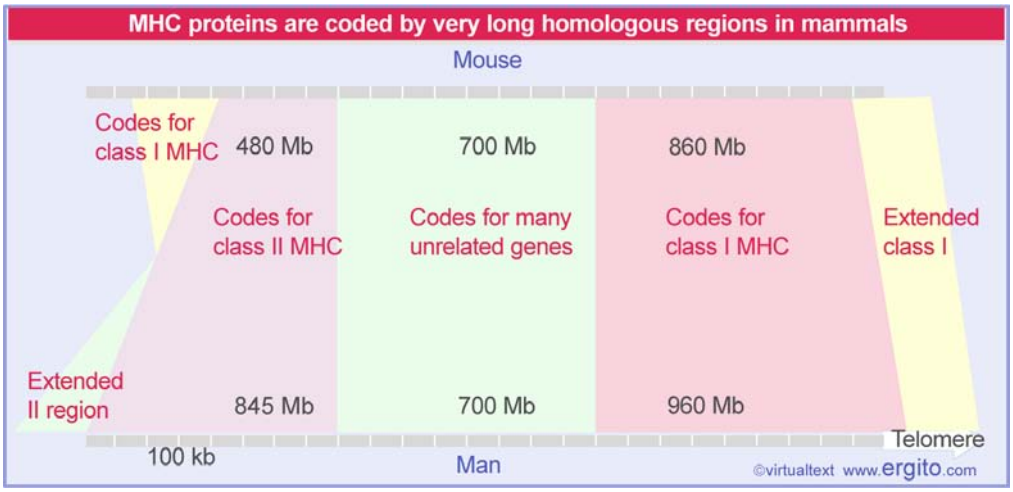

Figure 25.33 The MHC region extends for >2 Mb. MHC proteins of classes I and II are coded by two separate regions. The class III region is defined as the segment between them. The extended regions describe segments that are syntenic on either end of the cluster. The major difference between mouse and Man is the presence of H2 class I genes in the extended region on the left. The murine locus is located on chromosome 17, and the human locus on chromosome 6.

There are several hundred genes in the MHC regions of mammals, but it is possible for MHC functions to be provided by far fewer genes, as in the case of the chicken, where the MHC region is 92 Kb and has only 9 genes (910).

As in comparisons of other gene families, we find differences in the exact numbers of genes devoted to each function. Because the MHC locus shows extensive variation between individuals, the number of genes may be different in different individuals. As a general rule, however, a mouse genome has fewer active H2 genes than a human genome. The class II genes are unique to mammals (except for one subgroup), and as a rule, birds and fish have different genes in their place. There are ~8 functional class I genes in Man and ~30 in the mouse. The class I region also includes many other genes. The class III regions are closely similar in Man and mouse.

All MHC proteins are dimers located in the plasma membrane, with a major part of the protein protruding on the extracellular side. The structures of class I and class II MHC proteins are related, although they have different components, as summarized in **Figure 25.34**.

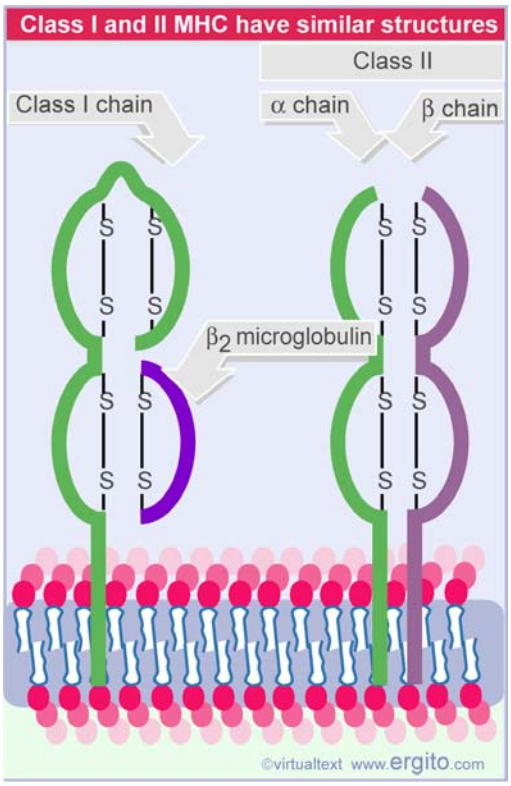

Figure 25.34 Class I and class II histocompatibility antigens have a related structure. Class I antigens consist of a single (α) polypeptide, with three external domains (α 1, α 2, α 3), that interacts with β 2 microglobulin (β 2 m). Class II antigens consist of two (α and β) polypeptides, each with two domains (α 1 & α 2, β 1 & β 2) with a similar overall structure.

Class II antigens consist of two chains, α and β, whose combination generates an overall structure in which there are two extracellular domains.

All class I MHC proteins consist of a dimer between the class I chain itself and the β2-microglobulin protein. The class I chain is a 45 kD transmembrane component that has three **external domains** (each ~90 amino acids long, one of which interacts with β2 microglobulin), a **transmembrane region** of ~40 residues, and a short **cytoplasmic domain** of ~30 residues that resides within the cell.

The β2 microglobulin is a secreted protein of 12 kD. It is needed for the class I chain to be transported to the cell surface. Mice that lack the β2 microglobulin gene have no MHC class I antigen on the cell surface.

The organization of class I genes summarized in **Figure 25.35** coincides with the protein structure. The first exon codes for a signal sequence (cleaved from the protein during membrane passage). The next three exons code for each of the external domains. The fifth exon codes for the transmembrane domain. And the last three rather small exons together code for the cytoplasmic domain. The only difference in the genes for human transplantation antigens is that their cytoplasmic

domain is coded by only two exons.

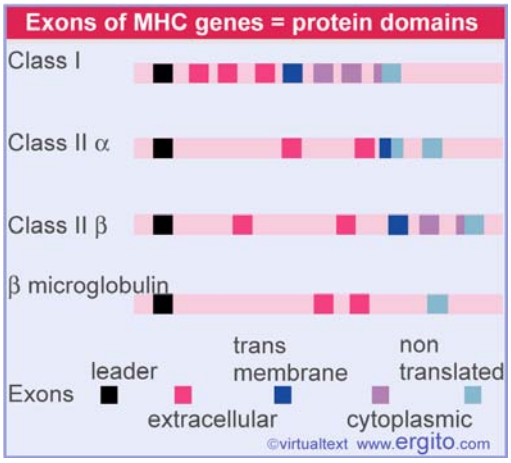

Figure 25.35 Each class of MHC genes has a characteristic organization, in which exons represent individual protein domains

The exon coding for the third external domain of the class I genes is highly conserved relative to the other exons. The conserved domain probably represents the region that interacts with β2 microglobulin, which explains the need for constancy of structure. This domain also exhibits homologies with the constant region domains of immunoglobulins.

What is responsible for generating the high degree of polymorphism in these genes? Most of the sequence variation between alleles occurs in the first and second external domains, sometimes taking the form of a cluster of base substitutions in a small region. One mechanism involved in their generation is gene conversion between class I genes. Pseudogenes are present as well as functional genes.

The gene for β2 microglobulin is located on a separate chromosome. It has four exons, the first coding for a signal sequence, the second for the bulk of the protein (from amino acids 3 to 95), the third for the last four amino acids and some of the nontranslated trailer, and the last for the rest of the trailer.

The length of β2 microglobulin is similar to that of an immunoglobulin V gene; there are certain similarities in amino acid constitution; and there are some (limited) homologies of nucleotide sequence between β2 microglobulin and Ig constant domains or type I gene third external domains. All the groups of genes that we have discussed in this chapter may have descended from a common ancestor that coded for a primitive domain.

Last updated on 4-15-2003

Reviews

262. Steinmetz, M. and Hood, L. (1983). *Genes of the MHC complex in mouse and man*. Science 222, 727-732.

265. Flavell, R. A., Allen, H., Burkly, L. C., Sherman, D. H., Waneck, G. L., and Widera, G. (1986). *Molecular biology of the H-2 histocompatibility complex*. Science 233, 437-443.

3772. Kumnovics, A., Takada, T., and Lindahl, K. F. (2003). *Genomic organization of the Mammalian MHC.* Annu. Rev. Immunol. 21, 629-657.

References

910. Kaufman, J. et al. (1999). *The chicken B locus is a minimal essential major histocompatibility complex.* Nature 401, 923-925.

IMMUNE DIVERSITY

5.25.21 Innate immunity utilizes conserved signaling pathways

Key Terms

Adaptive immunity is the response mediated by lymphocytes that are activated by their specific interaction with antigen. The adaptive immune response develops over several days as lymphocytes with antigen-specific receptors are stimulated to proliferate and become effector cells. It is responsible for immunological memory.

Acquired immunity is another term for adaptive immunity.

Innate immunity is the rapid response mediated by cells with non-varying (germline-encoded) receptors that recognize pathogen. The cells of the innate immune response act to eliminate the pathogen and initiate the adaptive immune response.

A **pathogen-associated molecular pattern is (PAMP)** a molecular structure on the surface of a pathogen. A given PAMP may be conserved across a large number of pathogens. During an immune response, PAMPs may be recognized by receptors on cells that mediate innate immunity.

A **lipopolysaccharide is (LPS)** a molecule containing both lipid and sugar components. It is present in the outer membrane of Gram-negative bacteria. It is also an endotoxin responsible for inducing septic shock during an infection.

An **endotoxin** is a toxin that is present on the surface of Gram-negative bacteria (as opposed to exotoxins, which are secreted). LPS is an example of an endotoxin.

A **Toll-like receptor is (TLR)** a plasma membrane receptor that is expressed on phagocytes and other cells and is involved in signaling during the innate immune response. TLRs are related to IL-1 receptors.

The **leucine-rich region is (LLR)** a motif found in the extracellular domains of some surface receptor proteins in animal and plant cells.

Key Concepts

- Innate immunity is triggered by receptors that recognize motifs (PAMPs) that are highly conserved in bacteria or other infective agents.
- Toll-like receptors are commonly used to activate the response pathway.
- The pathways are highly conserved from invertebrates to vertebrates, and an analogous pathway is found in plants.

The immune response described in this chapter comprises a set of reactions that respond to a pathogen by selecting lymphocytes (B cells or T cells) whose receptors (antibodies or TCR) have a high affinity for the pathogen. The basis for this selective process is the generation of a very large number of receptors so as to create a high possibility of recognizing a foreign molecule. Receptors that recognize the body's

own proteins are screened out early in the process. Activation of the receptors on B cells triggers the pathways of the humoral response; activation of the receptors on T cells triggers the pathways of the cell-mediated response. The overall response to an antigen via selection of receptors on the lymphocytes is called **adaptive immunity** (or **acquired immunity**). The response typically is mounted over several days, following the initial activation of B cells or T cells that recognize the foreign pathogen. The organism retains a memory of the response, which enables it to respond more rapidly if it is exposed again to the same pathogen. The principles of adaptive immunity are similar through the vertebrate kingdoms, although details vary.

Another sort of immune response occurs more quickly and is found in a greater range of animals (including those that do not have adaptive immunity). **Innate immunity** depends on the recognition of certain, pre-defined patterns in foreign pathogens. These patterns are motifs that are conserved in the pathogens because they have an essential role in their function, but they are not found in higher eukaryotes. The motif is typically recognized by a receptor dedicated to the purpose of triggering the innate response upon an infection. Receptors that trigger the innate response are found on cells such as neutrophils and macrophages, and cause the pathogen to be phagocytosed and killed. The response is rapid, because the set of receptors is already present on the cells and does not have to be amplified by selection. It is widely conserved and is found in organisms ranging from flies to Man. When the innate response is able to deal effectively with an infection, the adaptive response will not be triggered. There is some overlap between the responses in that they activate some of the same pathways, so cells activated by the innate response may subsequently participate in the adaptive response.

The motifs that trigger innate immunity are sometimes called pathogen-associated molecular patterns (**PAMPs**). **Figure 25.36** shows that they are widely distributed across broad ranges of organisms. Formyl-methionine is used to initiate most bacterial proteins, but is not found in eukaryotes. The peptidoglycan of the cell wall is unique to bacteria. **Lipopolysaccharide** (**LPS**) is a component of the outer membrane of most gram-negative bacteria; also known as **endotoxin**, it is responsible for septic shock syndrome.

PAMPs are ubiquitous

Organism	Pathogen	Location
All bacteria	formyl-Methionine	Most proteins
Most bacteria	peptidoglycan	Cell wall
Gram-negative bacteria	lipopolysaccharide	Cell wall
Yeast	zymosan	Cell wall

Figure 25.36 Pathogen-associated molecular patterns (PAMPs) are compounds that are common to large ranges of bacteria or yeasts and are exposed when an infection occurs.

A key insight into the nature of innate immunity was the discovery of the involvement of **Toll-like receptors** (**TLR**) (for review see 2382). The receptor Toll, which is related to mammalian IL1 receptor, triggers the pathway in *Drosophila* that controls dorsal-ventral development (see *Molecular Biology 6.31.9 Dorsal-ventral development uses localized receptor-ligand interactions*). This leads to activation of the transcription factor dorsal, a member of the Rel family, which is related to the mammalian factor NF- κB. The pathway of innate immunity is parallel to the Toll pathway, with similar components. In fact, one of the first indications of the nature of innate immunity in flies was the discovery of the transcription factor Dif (dorsal-related immunity factor), which is activated by one of the pathways (2378).

Flies have no system of adaptive immunity, but are resistant to microbial infections (for review see 2385). This is because their innate immune systems trigger synthesis of potent antimicrobial peptides. Seven distinct peptides have been identified in *Drosophila*, where they are synthesized in the fat body (the equivalent organ to the liver). Two of the peptides are antifungal, five act largely on bacteria. The general mode of action is to kill the target organism by permeabilizing its membrane. All of these peptides are coded by genes whose promoters respond to transcription factors of the Rel family. **Figure 25.37** summarizes the components of the innate pathways in *Drosophila*.

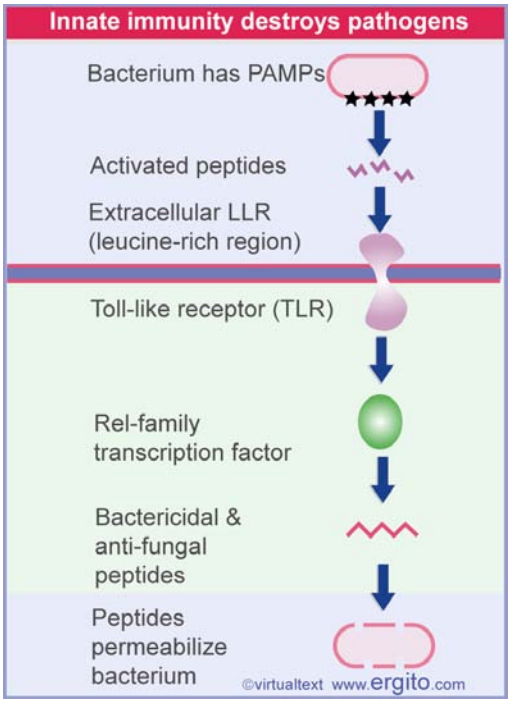

Figure 25.37 Innate immunity is triggered by PAMPs. In flies, they cause the production of peptides that activate Toll-like receptors. The receptors lead to a pathway that activates a transcription factor of the Rel family. Target genes for this factor include bactericidal and anti-fungal peptides. The peptides act by permeabilizing the membrane of the pathogenic organism.

Two innate response pathways function in *Drosophila*, one responding principally to

fungi, the other principally to gram-negative bacteria. Gram-positive bacteria may be able to trigger both pathways. **Figure 25.38** outlines the steps in each pathway. Fungi and gram-positive bacteria activate a proteolytic cascade that generates peptides that activate a Toll-like receptor. This is the NF- κB-like pathway. The dToll receptor activates the transcription factor Dif (a relative of NF- κB), leading ultimately to activation of the anti-fungal peptide drosomycin (2379; 2381). Gram-negative bacteria trigger a pathway via a different receptor that activates the transcription factor Relish, leading to production of the bactericidal peptide attacin (2380). This pathway is called the Imd pathway after one of its components, a protein that has a "death domain" related to those found in the pathways for apoptosis.

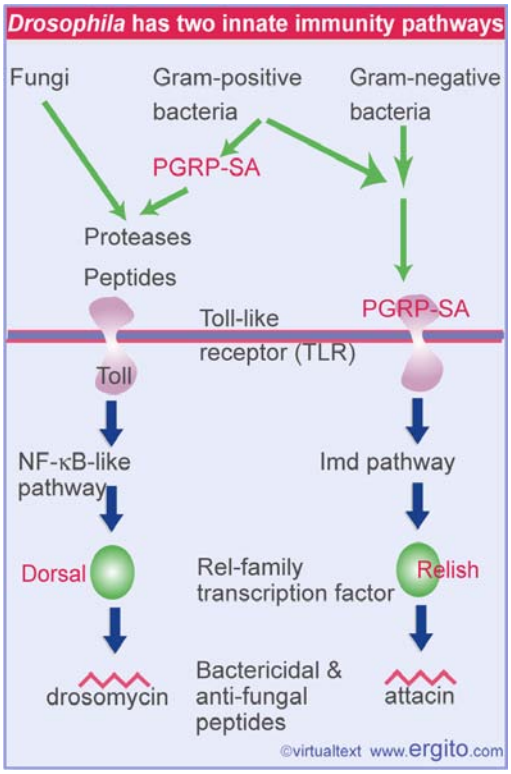

Figure 25.38 One of the *Drosophila* innate immunity pathways is closely related to the mammalian pathway for activating NF-kB; the other has components related to those of apoptosis pathways.

The key agents in responding to the bacteria are proteins called PGRPs because of their high affinities for bacterial peptidoglycans. There are two types of these proteins. PGRP-SAs are short extracellular proteins. They probably function by activating the proteases that trigger the Toll pathway. PGRP-LCs are transmembrane proteins with an extracellular PGRP domain. Their exact role has to be determined.

The innate immune response is highly conserved (for review see 2459, 4381). Mice that are resistant to septic shock when they are treated with LPS have mutations in the Toll-like receptor TLR4 (2383). A human homologue of the Toll receptor can activate some immune-response genes, suggesting that the pathway of innate immunity may also function in Man (2384). The pathway downstream of the

Toll-like receptors is generally similar in all cases, typically leading to activation of the transcription factor NF- κB. We do not yet know whether the upstream pathway is conserved, in particular whether the PAMPs function by generating ligands that in turn activate the Toll-like receptors or whether they might interact directly with them. The pathway upstream of the Toll-like receptors is different in mammals and flies, because the pathogens directly activate mammalian Toll-like receptors. In the case of LPS, the pathogen binds to the surface protein CD14; this enables CD14 to activate TLR4, triggering the innate response pathway (2383). There are ~20 receptors in the TLR (Toll-like receptor) class in the human genome, which gives some indication of how many pathogens can trigger the innate response.

Plants have extensive defense mechanisms, among which are pathways analogous to the innate response in animals (for review see 2387). The same principle applies that PAMPs are the motifs that identify the infecting agent as a pathogen. The proteins that respond to the pathogens are coded by a class of genes called the disease resistance genes. Many of these genes code for receptors that share a property with the TLR class of animal receptors: the extracellular domain has a motif called the **leucine-rich region** (**LLR**). The response mechanism is different from animal cells, and is directed to activating a MAPK cascade. Many different pathogens activate the same cascade, which suggests that a variety of pathogen-receptor interactions converge at or before the activation of the first MAPK (2386).

Last updated on November 10, 2003

Reviews

2382. Aderem, A. and Ulevitch, R. J. (2000). *Toll-like receptors in the induction of the innate immune response.* Nature 406, 782-787.

2385. Hoffmann, J. A., Kafatos, F. C., Janeway, C. A., and Ezekowitz, R. A. (1999). *Phylogenetic perspectives in innate immunity.* Science 284, 1313-1318.

2387. Dangl, J. L. and Jones, J. D. (2001). *Plant pathogens and integrated defence responses to infection.* Nature 411, 826-833.

2459. Janeway, C. A. and Medzhitov, R. (2002). *Innate immune recognition.* Annu. Rev. Immunol. 20, 197-216.

References

2378. Ip, Y. T., Reach, M., Engstrom, Y., Kadalayil, L., Cai, H., GonzÃ¡lez-Crespo, S., Tatei, K., and Levine, M. (1993). *Dif, a dorsal-related gene that mediates an immune response in Drosophila.* Cell 75, 753-763.

2379. Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J. M., and Hoffmann, J. A. (1996). *The dorsoventral regulatory gene cassette spÃtzle/Toll/cactus controls the potent antifungal response in Drosophila adults.* Cell 86, 973-983.

2380. Williams, M. J., Rodriguez, A., Kimbrell, D. A., and Eldon, E. D. (1997). *The 18-wheeler mutation reveals complex antibacterial gene regulation in Drosophila host defense.* EMBO J. 16, 6120-6130.

2381. Rutschmann, S., Jung, A. C., Hetru, C., Reichhart, J. M., Hoffmann, J. A., and Ferrandon, D. (2000). *The Rel protein DIF mediates the antifungal but not the antibacterial host defense in Drosophila.* Immunity 12, 569-580.

2383. Poltorak, A., He, X., Smirnova, I., Liu, M. Y., Huffel, C. V., Du, X., Birdwell, D., Alejos, E., Silva, M., Galanos, C., Freudenberg, M., Ricciardi-Castagnoli, P., Layton, B., and Beutler, B. (1998). *Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene.* Science 282, 2085-2088.

2384. Medzhitov, R., Preston-Hurlburt, P., and Janeway, C. A. (1997). *A human homologue of the Drosophila Toll protein signals activation of adaptive immunity.* Nature 388, 394-397.

2386. Asai, T., Tena, G., Plotnikova, J., Willmann, M. R., Chiu, W. L., Gomez-Gomez, L., Boller, T., Ausubel, F. M., and Sheen, J. (2002). *MAP kinase signalling cascade in Arabidopsis innate immunity.* Nature 415, 977-983.

4381. Hoffmann, J. A. (2003). *The immune response of Drosophila.* Nature 426, 33-38.

IMMUNE DIVERSITY

5.25.22 Summary

Immunoglobulins and T cell receptors are proteins that play analogous functions in the roles of B cells and T cells in the immune system. An Ig or TCR protein is generated by rearrangement of DNA in a single lymphocyte; exposure to an antigen recognized by the Ig or TCR leads to clonal expansion to generate many cells which have the same specificity as the original cell. Many different rearrangements occur early in the development of the immune system, creating a large repertoire of cells of different specificities.

Each immunoglobulin protein is a tetramer containing two identical light chains and two identical heavy chains. A TCR is a dimer containing two different chains. Each polypeptide chain is expressed from a gene created by linking one of many V segments via D and J segments to one of a few C segments. Ig L chains (either κ or λ) have the general structure V-J-C, Ig H chains have the structure V-D-J-C, TCR α and γ have components like Ig L chains, and TCR δ and β are like Ig H chains.

Each type of chain is coded by a large cluster of V genes separated from the cluster of D, J, and C segments. The numbers of each type of segment, and their organization, are different for each type of chain, but the principle and mechanism of recombination appear to be the same. The same nonamer and heptamer consensus sequences are involved in each recombination; the reaction always involves joining of a consensus with 23 bp spacing to a consensus with 12 bp spacing. The cleavage reaction is catalyzed by the RAG1 and RAG2 proteins, and the joining reaction is catalyzed by the same NHEJ pathway that repairs double-strand breaks in cells. The mechanism of action of the RAG proteins is related to the action of site-specific recombination catalyzed by resolvases.

Although considerable diversity is generated by joining different V, D, J segments to a C segment, additional variations are introduced in the form of changes at the junctions between segments during the recombination process. Changes are also induced in immunoglobulin genes by somatic mutation, which requires the actions of cytidine deaminase and uracil glycosylase. Mutations induced by cytidine deaminase probably lead to removal of uracil by uracil glycosylase, followed by the induction of mutations at the sites where bases are missing.

Allelic exclusion ensures that a given lymphocyte synthesizes only a single Ig or TCR. A productive rearrangement inhibits the occurrence of further rearrangements. Although the use of the V region is fixed by the first productive rearrangement, B cells switch use of C_H genes from the initial μ chain to one of the H chains coded farther downstream. This process involves a different type of recombination in which the sequences between the VDJ region and the new C_H gene are deleted. More than one switch occurs in C_H gene usage. Class switching requires the same cytidine deaminase that is required for somatic mutation, but its role is not known. At an earlier stage of Ig production, switches occur from synthesis of a membrane-bound version of the protein to a secreted version. These switches are accomplished by alternative splicing of the transcript.

Innate immunity is a response triggered by receptors whose specificity is predefined for certain common motifs found in bacteria and other infective agents. The receptor that triggers the pathway is typically a member of the Toll-like class, and the pathway resembles the pathway triggered by Toll receptors during embryonic development. The pathway culminates in activation of transcription factors that cause genes to be expressed whose products inactivate the infective agent, typically by permeabilizing its membrane.