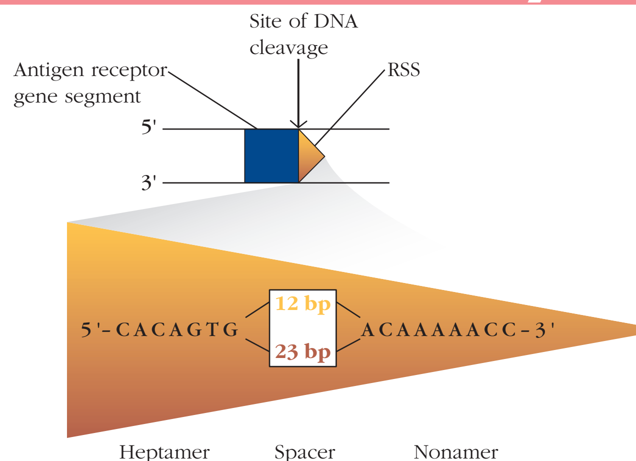
7

The Organization and Expression of Lymphocyte Receptor Genes



The Recombination Signal Sequence (RSS) serves as the site of DNA cleavage. The RSS is composed of conserved heptamer and nonamer sequences, separated by a spacer region of 12 or 23 bp, which is conserved in length, but not in sequence. Cleavage occurs at the junction of the heptamer and the variable region coding segment. *[Adapted from D. G. Schatz and Y. Ji, 2011, Recombination centres and the orchestration of V(D)J recombination,* Nature Reviews. Immunology **11:***251–263.]*

- The Puzzle of Immunoglobulin Gene Structure
- Multigene Organization of Ig Genes
- The Mechanism of V(D)J Recombination
- B-Cell Receptor Expression
- T-Cell Receptor Genes and Expression

To protect its host, the immune system must recognize a vast array of rapidly evolving microorganisms. To accomplish this, it must generate a diverse and flexible repertoire of receptor molecules, while minimizing the expression of receptors that recognize self antigens. As described in Chapter 3, each B or T lymphocyte expresses a unique antigen-specific receptor. When these receptors bind to their corresponding antigens under the appropriate conditions (described in Chapters 11 and 12), T and B lymphocytes proliferate and differentiate into effector cells that eliminate the microbial threat (Chapter 13).

In Chapter 3, we described the biochemistry of the T and B lymphocyte receptors and the secreted antibodies formed by B lymphocytes following antigen stimulation. We also outlined the experiments which demonstrated that secreted antibodies are identical in antigen-binding specificity to the B-cell receptors of the secreting cell. In this chapter, we address the question of how an organism can encode and express receptors capable of recognizing a constantly evolving universe of microbial threats using a finite amount of genetic information.

The production of specific lymphocyte receptors employs a number of genetic mechanisms that are unique to the immune system. In 1987, Susumu Tonegawa won the Nobel Prize for Physiology or Medicine "for his discovery of the genetic principle for generation of antibody diversity," a discovery that challenged the fundamental concept that one gene encoded one polypeptide chain. Tonegawa and his colleagues showed that *the antibody light chain was encoded in the germ line by not one but three families of gene segments separated by kilobases of DNA* (Figure 7-1). (The germ-line DNA is the genetic information encoded in the sperm and egg, which can be passed on to future generations.) Their work demonstrated that two DNA segments, one from each family, are conjoined, only in B lymphocytes, to create the mature form of the light-chain *variable* region of the immunoglobulin (Ig) gene. A third segment encodes the *constant* region of the gene. Different combinations of segments are used in each B cell, to create the diverse repertoire of light-chain receptor genes. Subsequent experiments have shown that all of the B- and T-cell receptor genes are assembled from multiple gene segments by similar rearrangements.

We describe below the unique genetic arrangements of T- and B-cell receptor gene segments, and the mechanisms by which they are rearranged and expressed. We will address here only those mechanisms that shape the receptor repertoires of mouse and human **naïve** B and T cells, which have not yet been exposed to antigen.

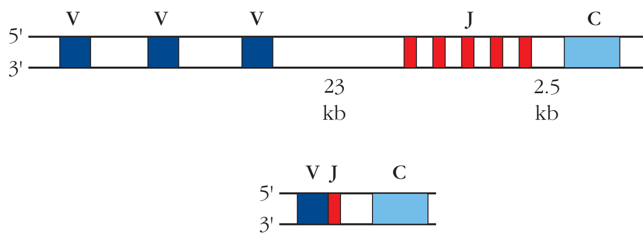


FIGURE 7-1 The antibody light-chain gene encodes three families of DNA segments. During B-cell development, one V segment and one J segment (which encode contiguous parts of the light-chain variable region) join together with the C (constant) region to form the gene for the antibody light chain. This gene rearrangement occurs in the DNA, prior to gene transcription into mRNA.

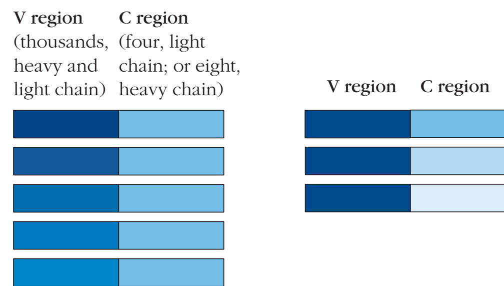


FIGURE 7-2 Early sequencing studies indicated that the light chain may be encoded in more than one segment. (a) Many Ig variable regions in both heavy and light chains could be found in association with few constant regions. (b) The same variable region can be found in contiguity with several different heavy-chain constant regions.

Additional layers of diversity are generated in B cells following antigen exposure and those will be addressed in Chapter 12. These powerful mechanisms include the generation of antibodies of different heavy-chain classes, each capable of mediating a discrete set of effector functions, as well as the creation of modified antigen-binding regions by somatic hypermutation, followed by antigen-driven selection. Both of these processes are triggered in human and mouse B cells only after antigen contact.

The Puzzle of Immunoglobulin Gene Structure

The immune system relies on a vast array of ***B*-cell receptors (BCRs)** that possess the ability to bind specifically to a correspondingly large number of potential pathogens. The first indication of the immense size of the antibody repertoire was provided by immunologists using synthetic molecules to stimulate antibody production. They discovered that antibodies can discriminate between small synthetic molecules differing in as little as the position of an amino or hydroxyl group on a phenyl ring. If the immune system can discriminate between small molecules that it had presumably never encountered during evolutionary selection and that differ in such subtle ways, then, it was reasoned, the number of potential antibodies must be very large indeed. A series of experiments conducted in the late 1970s and early 1980s estimated the number of different BCRs generated in a normal mouse immune system to be at least 10^7, but we now know that estimate was many orders of magnitude too small.

Investigators trying to make sense of Ig genetics were also faced with an additional puzzle: protein sequencing of mouse and human antibody heavy and light chains revealed that the first (amino terminal) 110 amino acids of antibody heavy and light chains are extremely variable among different antibody molecules. This region was therefore defined as the ***variable* (V) region** of the antibody molecule. In contrast, the remainder of the light and heavy chains could be classified into one of only four sequences (light chain) or eight sequences (heavy chain) (Figure 7-2a) and was, therefore, named the ***constant* (C) region**. This raised an intriguing question: If each of 10^5 to 10^7 antibodies is encoded by a separate gene, how could the constant region part of the gene remain constant in sequence in the face of evolutionary drift?

Furthermore, antibodies could be found in which the same antibody variable (V) region was associated with more than one heavy-chain constant (C) region (Figure 7-2b), lending further support to the possibility that the expression of the variable and constant regions of each antibody chain were independently controlled. (These additional constant regions are generated by the antigen-induced process of class switch recombination, discussed in Chapter 12. In the current context, the only point of relevance is the independence with which the variable and constant regions appear to be expressed.) It was rapidly becoming clear that the solution to the antibody gene puzzle was more complex than had previously been imagined.

Investigators Proposed Two Early Theoretical Models of Antibody Genetics

Classical **germ-line theories** of antibody diversity suggested that the genetic information for each antibody is encoded, in its entirety, within the germ-line genome. This means that the genes encoding the entire sequence of every antibody heavy or light chain that the animal could ever make must be present in every cell. However, a quick calculation is sufficient to demonstrate that if there are 10^7 or more antibodies, each of which requires approximately 2,000 nucleotides, this would require massive expenditure of genetic information—indeed, more DNA would be required to encode the receptors of the immune system than is available to the organism. Although arguments were initially made that the dedication of a considerable fraction of the genome to the immune system may represent a reasonable evolutionary strategy, it

became clear, as estimates of the size of the antibody repertoire were revised upward, that there simply was not enough DNA to go around and new ideas must be found.

In 1965, William Dreyer and J. Claude Bennett proposed that antibody heavy and light chains are each encoded in two separate segments in the germ-line genome and that one of each of the V region- and C region-encoding segments are brought together in B-cell DNA to form complete antibody heavy and light-chain genes. The idea that DNA in somatic cells might engage in recombinatorial activity was revolutionary. However, germ-line theorists began to modify their ideas to embrace this possibility.

Others suggested the equally innovative idea that the number of variable region genes in the genome might be extremely limited, and proposed the **somatic hypermutation theory**. According to this hypothesis, a limited number of antibody genes is acted upon by unknown mutational mechanisms in somatic cells to generate a diverse receptor repertoire in mature B lymphocytes. This latter idea had the advantage of explaining how a large repertoire of antibodies could be generated from a relatively small number of genes, but the disadvantage that such a process, like the somatic cell gene recombination suggested by others, had never been observed. There was additional argument over whether mutation would occur prior to, or only after, antigen contact and, therefore, whether mutation was responsible for generation of the so-called "primary repertoire" that exists prior to antigen binding by the B cell.

Heated debate continued between the proponents of modified germ-line versus somatic mutation theories throughout the early 1970s, until a seminal set of experiments revealed that both sides were correct. We now know that each antibody molecule is encoded by multiple, germ-line, variable-region gene segments, which are rearranged differently in each naïve immune cell to produce a diverse primary receptor repertoire. These rearranged genes are then further acted upon after antigen encounter by somatic hypermutation and antigenic selection, resulting in an expanded and exquisitely honed repertoire of antigen-specific B cells (see Chapter 12).

Breakthrough Experiments Revealed That Multiple Gene Segments Encode the Light Chain

From the mid 1970s until the mid 1980s, a small group of brilliant immunologists completed a series of experiments that fundamentally altered the way in which scientists think about the genetics of immune receptor molecules. The first breakthrough occurred when Susumu Tonegawa showed, as Dreyer and Bennett had predicted, that multiple gene segments encode the antibody light chain. His achievement is all the more impressive because the modern tools of molecular biology were not yet available. The experiment he performed with Nobumichi Hozumi is described in Classic Experiment Box 7-1.

CLASSIC EXPERIMENT BOX 7-1

Hozumi and Tonegawa's Experiment: DNA Recombination Occurs in immunoglobulin Genes in Somatic Cells

The paradigm-shifting experiment of Hozumi and Tonegawa was designed to determine whether the DNA encoding Ig light-chain constant and variable regions existed in separate segments in non antibody-producing cells, such that a single constant region gene could associate with different variable region genes in different B cells. They hypothesized that this might be the case, because of Ig amino acid sequencing data showing that the sequences of the constant regions of many Ig light chains were identical. They reasoned that if multiple copies of a constant region gene existed, then each copy would be expected to accumulate silent or neutral mutations over time. The most likely explanation for the finding of a single constant region amino acid sequence arranged in tandem with many different variable regions was, therefore, that a single constant region gene segment was cooperating with multiple variable region gene segments in different cells or situations to generate the light-chain gene. However, this notion was heretical to a generation of scientists brought up on the concept that one gene encodes one polypeptide chain.

Scientists are accustomed to thinking about the concept of alternative RNA splicing, wherein different proteins may be encoded by the same piece of DNA by differentially using particular RNA segments, cut and spliced following transcription of a long precursor RNA transcript. However, this experiment asked an entirely new question: *Can a piece of DNA change its place on a chromosome in a somatic cell?*

To serve as their source of germ-line DNA, Hozumi and Tonegawa used DNA from an organ in which Ig genes would not be expressed, *embryonic liver* in this case. (Sperm and egg DNA would have been much more difficult to obtain.) For B-cell DNA, they used DNA from an antibody-producing plasma cell tumor line, MOPC 321, which secretes fully functional κ light chains. They separately cut both sets of DNA with the same restriction enzyme and used radioactive probes to determine *the sizes of the DNA fragments* on which the variable and constant regions of the light-chain gene were found in the two sets of cells.

(continued)

CLASSIC EXPERIMENT *(continued)*

Below, we describe how Hozumi and Tonegawa conducted their experiment, and indicate how the experimental protocol would be modified using the reagents and techniques available to a modern molecular biologist. We then display their data as it would appear both in the modern gel format as well as in its original form.

Hozumi and Tonegawa:

a. Purified genomic DNA from embryonic liver cells and from the MOPC 321 tumor cell line.

b. Cut the two sets of genomic DNA with a restriction endonuclease (BamHI) and separated the DNA fragments using electrophoresis. (Today we would use a polyacrylamide gel and electrophorese submicrogram quantities of samples over the course of a few hours; Hozumi and Tonegawa used a foot-long agarose gel that needed 2 liters of agarose. They loaded 5 mg of DNA and ran the gel for 3 days.)

c. Made ^{125}I-labeled nucleic acid probes specific for the two regions of mouse κ light chain. One of Tonegawa's probes was a full-length radiolabeled piece of mRNA encoding the entire κ-chain sequence. The other probe was the 3′ half of the sequence, which would hybridize to the constant, but not to the variable region of the κ-chain gene. Today, making enzymatically labeled, stable DNA probes for any particular sequence is a safe and relatively straightforward task. In Hozumi and Tonegawa's day, this task was significantly more challenging.

d. Probed the nuclease-generated DNA fragments to determine the size of the fragments carrying the variable and constant region sequences. We would normally use a Southern blot procedure, electrophoresing the fragments, blotting the gel with nitrocellulose paper, and then probing the paper with enzyme-labeled fragments complementary to the sequences of interest, developing the blots with luminescent or fluorescent substrates. Hozumi and Tonegawa's approach was much more time-consuming. They cut the gel into about 30 slices, melted the agarose, and separately eluted the DNA from each slice. To each DNA sample, they added radiolabeled RNA, allowed it to anneal, and removed the un-annealed RNA using RNase. The radioactivity remaining in each fraction was then plotted against the size of the DNA in the slice.

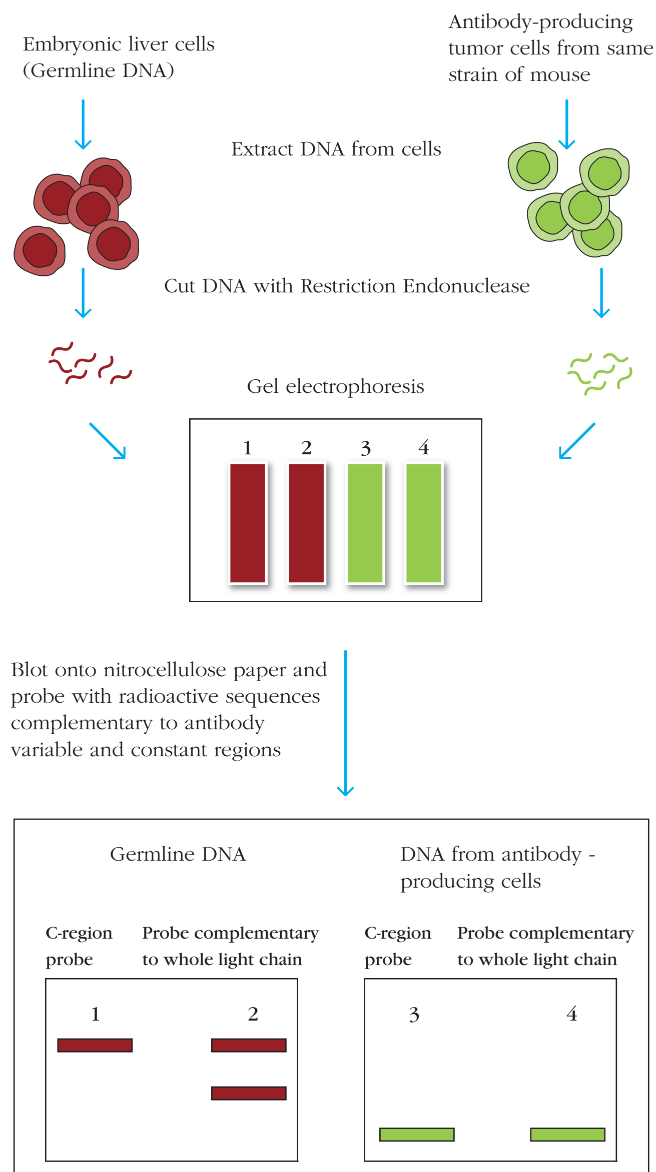


FIGURE 1
The κ light-chain gene is formed by DNA recombination between variable and constant region gene segments. DNA from embryonic liver cells was used as a source of germ-line DNA, and DNA from a B-cell tumor cell line was used as an example of DNA from antibody-producing cells. In embryonic liver, the DNA sequences encoding the variable and constant regions, respectively, were located on different restriction endonuclease fragments. However, these two sequences were co-located on a single restriction fragment in the myeloma DNA. (See text for details of the experiment.)

Figure 1 shows the results that would be obtained from a modern-day Southern

BOX 7-1

blot of Tonegawa's fragments. The germ-line DNA blot probed with the constant region mRNA sequence probe (lane 1) shows only a single band. This indicates that the restriction endonuclease is not cutting in the middle of the light-chain constant region, but rather that the whole constant region sequence is encoded within a single fragment.

In contrast, probing the blot with the whole light-chain sequence (V and C, lane 2) yields two bands. Since the difference between lanes 1 and 2 is the presence of the variable region sequence in the probe used for lane 2, the presence of two bands in lane 2 indicates that the restriction endonuclease used to generate the DNA fragments is cutting somewhere in between the variable and constant regions in the germ-line DNA, such that the variable region lies on a DNA fragment distinct from that bearing the constant region.

Analyzing the blots from the plasma cell tumor, we first note that the large fragment containing the constant region DNA and the midsized fragment bearing the variable region DNA have disappeared. Both the variable and constant region gene segments are now located on smaller fragments. This implies that a new pair of restriction endonuclease sites now forms the boundaries of each of the constant and variable region fragments. *Right away, we can tell that the DNA environment around the light-chain genes changes as the B cell differentiates*.

Next, we note that the sizes of the DNA fragments on which the constant and variable region gene segments are located are apparently the same. This implies, although it does not yet prove, that the movement of the constant and/or variable region gene segments has brought them into close proximity with one another, such that the constant and variable region gene segments co-locate on the same fragment. An alternative explanation is that they have both altered their locations and the similarity in the size of the fragments is coincidental. DNA sequencing supported the former interpretation: *as the B cell differentiates, the variable and constant region gene segments are moved from distant regions of the chromosome into close apposition with one another*.

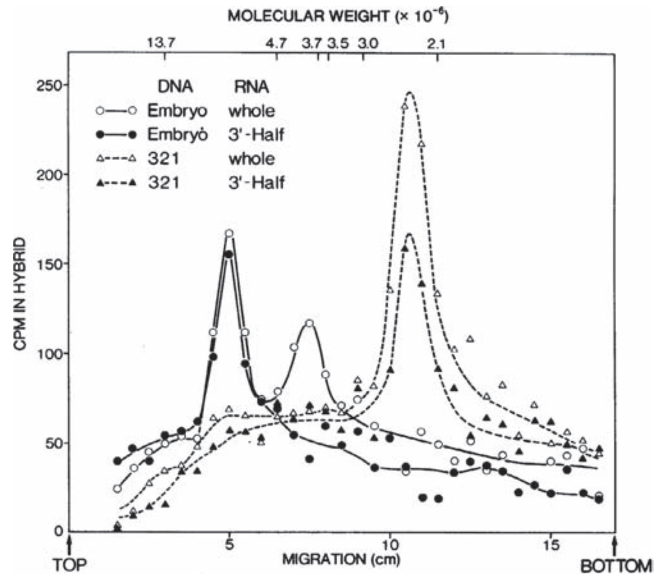


FIGURE 2

The original data from Hozumi and Tonegawa's classic experiment proving immunoglobulin gene recombination occurs in B cells. DNA from the sources shown was run out on an agarose gel for 3 days at 4°C. The gel was sliced, the DNA eluted, and then each sample was hybridized with radiolabeled probes for either the constant region of the κ light chain, or the whole chain. The plot shows the amount of radioactivity in each fraction as a function of the migration distance, which reflects the molecular weight of the DNA fragment. (See text for details.) *[From N. Hozumi and S. Tonegawa, 1976,* Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions, *Proceedings of the National Academy of Sciences USA* **73:***3628–3632.]*

It is all too easy, with modern molecular biology technologies, to forget what a tour de force this experiment actually represented. Few of the reagents or pieces of apparatus we commonly encounter in the molecular biology laboratory today were available to Hozumi and Tonegawa—they had to make their own.

The original paper speaks of purifying their own restriction endonuclease (BamHI) from a bacterial sample obtained from a colleague, and of performing the electrophoresis at 4°C for 3 days. Furthermore, technical difficulties precluded their being able to make a good 5′ (V-region) probe. They therefore probed each sample with a 3′ (C-region) probe, as well as one that bound to the whole light chain, and then inferred 5′ (V-region) binding by subtraction. It was an extraordinary piece of work. Their original data are shown in Figure 2.

What happened to the DNA on the alternative allele that did not encode the tumor cell secreted light chain? Subsequent analysis of the DNA from this tumor showed that the DNA from both chromosomes had undergone rearrangement. Hozumi and Tonegawa were fortunate that the variable regions used by both rearrangements were close to one another and so the fragment patterns overlapped.

Hozumi, N., and S. Tonegawa. 1976. Evidence for somatic rearrangement of Ig genes coding for variable and constant regions. *Proceedings of the National Academy of Sciences USA* **73**:3628–3632.

Tonegawa's experiment showed that, in the DNA from non-antibody-producing, embryonic liver cells, there is a BamHI endonuclease site between the variable and constant regions. We know this because probes for the variable and constant regions each recognized a different DNA fragment in BamHI-digested germ-line DNA. However, in the antibody-producing tumor cell, the DNA encoding the variable and constant regions appeared to be combined in just one fragment, thus demonstrating that DNA rearrangement must have occurred during the formation of an antibody light-chain gene (see Classic Experiment Box 7-1 for further details).

Although this experiment demonstrated that the V and C regions of antibody genes were located in different contexts in the DNA of non-antibody-producing cells, it did not speak to their relative locations; indeed, the initial experiment did not rule out the possibility that the V and C fragments could be encoded on different chromosomes in the embryonic cells. However, sequencing experiments subsequently showed that the segments encoding the V and C genes of the κ light chains are on the same chromosome and that, in non B cells, the two segments are separated by a long non-coding DNA sequence.

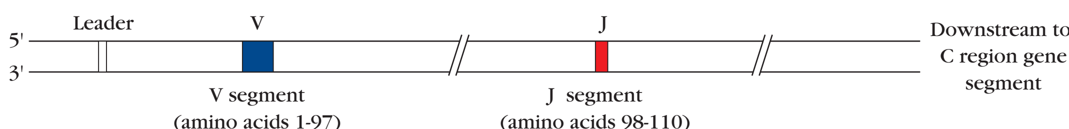
The impact of this result on the biological community was profound. For the first time, DNA was shown to be cut and rejoined during the process of cell differentiation. This experiment not only provided the experimental proof of Dreyer and Bennett's prediction; it also paved the way for the next surprising finding.

Tonegawa's experiment had identified the V and C segments encoding the kappa light chain. However, when scientists in his group sequenced the antibody light-chain DNA, they encountered another unexpected result. As expected, the embryonic (unrearranged) V region segment had, at its 5′ terminus, a short hydrophobic leader sequence, a common feature of membrane proteins necessary to guide the nascent protein chain into the membrane (Figure 7-3a). A 93 bp sequence of non-coding DNA separated the leader sequence from a long stretch of DNA that encoded the first 97 amino acids of the V region. But the light-chain V region domain is approximately 110 amino acids long. Where was the coding information for the remaining 13 amino acids?

Sequencing of the light-chain constant region fragment from embryonic DNA provided the answer. Upstream from the constant region coding sequence, and separated from it by a non-coding DNA segment of 1250 bp, were the 39 bp encoding the remaining 13 amino acids of the V region. This additional light-chain coding segment was named the ***joining* (J) gene segment** (shown in red in Figure 7-3a). Further sequencing of mouse and human light-chain variable and constant region genes confirmed Tonegawa's second, astonishing finding. Not only are the variable and constant regions encoded in two separate segments, but the light-chain variable regions themselves are encoded in two separate gene segments, the V and J segments, that are made contiguous only in antibody-producing cells.

The heavy-chain variable region gene was then shown to require yet a third genetic segment. Adopting a similar strategy of cloning and sequencing Ig heavy-chain genes, Lee Hood's group identified a germ-line ***H*eavy-chain *V*ariable Region (V_H)** gene fragment that encoded amino acid residues 1–101 of the antibody heavy chain and a second fragment that included a ***H*eavy-chain *J*oining gene segment (J_H)** that determined the sequence of amino acid residues 107–123. Neither of these segments contained the DNA sequence necessary to encode residues 102–106 of the heavy chain. Significantly, these missing residues were included in the third complementarity-determining region of the antibody heavy chain, **CDR3**, which provides contact residues in the binding of most antigens (see Chapter 3). Gene segments encoding this part of the antibody heavy chain were eventually located 5′ of the J region in mouse embryonic DNA by Hood and colleagues (Figure 7-3b). The importance of the contribution made by this gene segment to the diversity of antibody specificities is denoted by its name; the ***Diversity* (D) region**. (Because there is no D region in the light chain, immunologists usually drop the subscript denoting the heavy chain.)

Thus, the variable region of the heavy chain of the antibody molecule is encoded by three discrete gene segments,



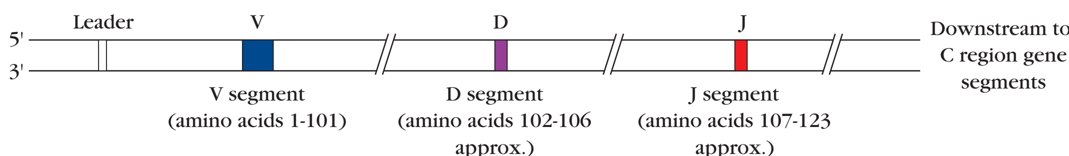


FIGURE 7-3 Antibody light chains are encoded in two segments—V and J, whereas antibody heavy chains are encoded in three segments—V, D, and J.

TABLE 7-1 Chromosomal locations of immunoglobulin genes in humans and mice

Gene	Human chromosome	Mouse chromosome
λ light chain	22	16
κ light chain	2	6
Heavy chain	14	12

and the variable region of the light chain by two segments, in the germ-line genome. These segments are brought together by a process of DNA recombination that occurs only in the B lymphocyte lineage to create the complete variable region gene. Furthermore, the DNA at the junction between the V and J segments for light chains, and at the VD and DJ junctions in heavy chains, accounts for the extraordinary diversity that was first observed by Kabat and Wu in the CDR3 regions of both heavy and light chains (see Chapter 3). Below, we will describe how further genetic diversity is generated at these junctions by additional processes unique to immune system genetics.

Multigene Organization of Ig Genes

Recall that Ig proteins consist of two identical heavy chains and two identical light chains (see Chapter 3). The light chains can be either **kappa (κ) light chains** or **lambda (λ) light chains**. The heavy-chain, kappa, and lambda gene families are each encoded on separate chromosomes (Table 7-1).

Kappa Light-Chain Genes Include V, J, and C Segments

The mouse Ig_κ locus spans 3.5 Mb and includes 91 potentially functional V_κ genes, which have been grouped into 18 V gene families based on sequence homology (Figure 7-4a). (If two sequences are greater than 80% identical, they are classified as belonging to the same V gene family.) These V gene families can be further grouped into V gene clans, based again on sequence homology. Each V_κ segment includes the leader exon encoding the signal peptide. Individual V_κ segments are separated by non-coding gaps of 5 to 100 kb. The transcriptional orientation of particular V_κ segments may be in the same or in the opposite direction as the constant region segment. The relative orientations of the variable and constant region segments do not affect the use of the segments, but do alter some details of the recombinational mechanism that creates the complete light-chain gene, as will be discussed later.

Downstream of the V_κ region cluster are four functional J_κ segments and one pseudogene, or other nonfunctional open reading frame (Table 7-2). A similar arrangement is found in the human V_κ locus, although the numbers of V and J gene segments vary. A single C_κ segment is found downstream of the J region, and all kappa light-chain constant regions are encoded by this segment.

Lambda Light-Chain Genes Pair Each J Segment with a Particular C Segment

The mouse Ig_λ locus spans a region of approximately 190 kb. Lambda light chains are only found in 5% of mouse antibodies because of a deletional event in the mouse genome that has eliminated most of the lambda light-chain variable region segments. It was therefore not surprising to discover that

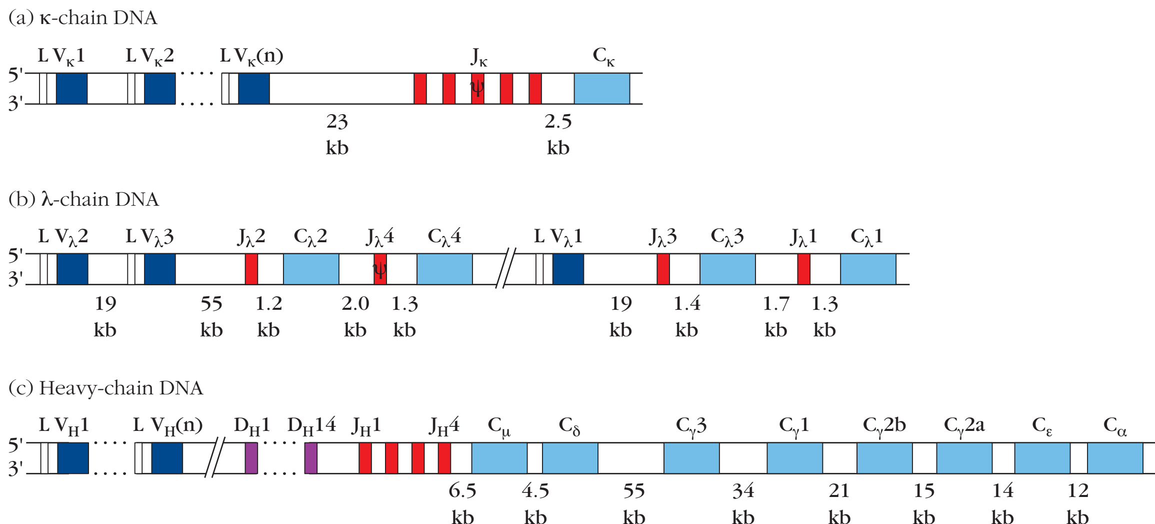


FIGURE 7-4 Organization of immunoglobulin germ-line gene segments in the mouse. The (a) κ light chain and (b) λ light chains are encoded by V, J, and C gene segments. The (c) heavy chain is encoded by V, D, J, and C gene segments. The distances in kilobases (kb) separating the various gene segments in mouse germ-line DNA are shown below each diagram.

TABLE 7-2 Immunoglobulin variable region gene numbers in humans and mice

	Human	Mouse
V_κ	34–48 functional; 8 ORFs; 30 pseudogenes.	91 functional; 9 ORFs; 60 pseudogenes
J_κ	5 functional; multiple alleles	4 functional; 1 ORF
V_λ	33 functional; 6 ORFs; 36 pseudogenes	3–8 functional
J_λ	5 functional; 2 ORFs	4 functional; 1 pseudogene
V_H	38–44 functional; 4 ORFs; 79 pseudogenes	101 functional; 69 pseudogenes*
D	23 functional; 4 ORFs	19 functional; 7 ORFs; 6 pseudogenes
J_H	6 functional; 3 pseudogenes; several different alleles found	4 functional; again, multiple haplotypes

* Gene numbers of the mouse V_H locus refer to the sequenced chromosome 14 of the C57 BL/6 mouse.

A germ-line gene is considered to be functional if the coding region has an open reading frame (ORF) without a stop codon, and if there is no described defect in the splicing sites, RSS, and/or regulatory elements.

A germ-line entity is considered to be an ORF if the coding region has an open reading frame but alterations have been described in the splicing sites, RSSs, and/or regulatory elements, and/or changes of conserved amino acids have been suggested by the investigators to lead to incorrect folding and/or the entity is an orphon, a nonfunctional gene located outside the main chromosomal locus.

A germ-line entity is considered to be a pseudogene if the coding region has a stop codon(s) and/or a frameshift mutation. In the case of V gene segments, these mutations may be either in the V gene coding sequence, or in the leader sequence.

SOURCE: Gene numbers and definitions summarized from the International Immunogenetics Information System Web site: http://imgt.org. Accessed November 16, 2011.

there are only three fully functional V_λ gene segments, although this number varies somewhat by strain (see Table 7-2). Interestingly, there are also three fully functional λ chain constant regions, each one associated with its own J region segment (Figure 7-4b). The J-C_λ4 segments are not expressed because of a splice site defect.

Since recombination of *Ig* gene segments always occurs in the downstream direction (V to J), the location of the V_λ1 variable region sequence upstream of the JC_λ3 and JC_λ1 segments but downstream from JC_λ2 means that V_λ1 is always expressed with either JC_λ3 or JC_λ1 but never with JC_λ2. For the same reason, V_λ2 is usually associated with the JC_λ2 pairing, although occasional pairings of V_λ2 with the 190 kb distant JC_λ1 segments have been observed.

In humans, 40% of light chains are of the λ type and about 30 V_λ-chain gene segments are used in mature antibody light chains (see Table 7-2). Downstream from the human V_λ locus is a series of seven JC_λ pairs, of which four pairs are fully functional.

Heavy-Chain Gene Organization Includes V_H, D, J_H, and C_H Segments

Multiple V_H gene segments lie across a region of approximately 3.0 Mb in both mice and humans (Figure 7-4c). These segments can be classified, like V_κ segments, into families of homologous sequences. Humans express at least 38 functional V_H segments and mice approximately 101 (see Table 7-2). Downstream from the cluster of mouse V_H region segments is an 80 kb region containing approximately 14 D regions (the actual number varies among different mouse strains). Just 0.7 kb downstream of the most 3′ D segment is the J_H region cluster, which contains four functional J_H regions. A further gap separates the last J_H segment from the first constant region exon, C_μ1.

The human V_H locus has a similar arrangement with approximately 30 functional D segments, and 6 functional J_H segments. Human D regions can be read in all three reading frames, whereas mouse D regions are mainly read in reading frame 1, because of the presence of stop codons in reading frames 2 and 3 in most mouse D regions.

The eight constant regions of antibody heavy chains are encoded in a span of 200 kb of DNA downstream from the J_H locus, as illustrated in Figure 7-4c. Recall that the constant regions of antibodies determine their heavy-chain class and, ultimately, their effector functions (see Chapter 13).

The Mechanism of V(D)J Recombination

In V(D)J recombination, the DNA encoding a complete antibody V-region is assembled from V, D, and J (heavy chain) or V and J (light chain) segments that are initially separated by many kilobases of DNA. Each developing B cell generates a novel pair of variable region genes by recombination at the level of genomic DNA. Recombination is catalyzed by a set of enzymes, many of which are also involved in DNA repair functions (Table 7-3), and is directed to the appropriate sites on the Ig gene by recognition of specific DNA sequence motifs called ***Recombination Signal Sequences (RSSs)***. These sequences ensure that one of each

TABLE 7-3 Proteins involved in V(D)J recombination

Protein	Function
RAG-1/2	Lymphoid-specific complex of two proteins that catalyze DNA strand breakage and rejoin to form signal and coding joints
TdT	Lymphoid-specific protein that adds N region nucleotides to the joints between gene segments in the Ig heavy chain and at all joints between TCR gene segments
HMG1/2 proteins	Stabilize binding of RAG1/2 to Recombination Signal Sequences (RSSs), particularly to the 23-bp RSS; stabilize bend introduced into the 23-bp spacer DNA by the RAG1/2 proteins
Ku70 and Ku80 heterodimers	Binds DNA coding and signal ends and holds them in protein-DNA complex
DNA PKcs	In complex with Ku proteins, recruits and phosphorylates Artemis
Artemis	Opens the coding end hairpins
XRCC4	Stabilizes and activates DNA ligase IV
DNA ligase IV	In complex with XRCC4, and Cernunnos ligates DNA ends
Cernunnos	With XRCC4, activates DNA ligase IV

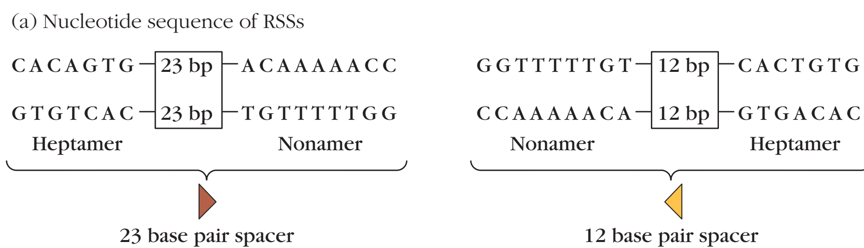
type of segment (V and J for the light chain, or V, D, and J for the heavy chain) is included in the recombined variable region gene. During cleavage and ligation of the segments, the DNA is edited in various ways, adding further variability to the recombined gene.

Recombination Is Directed by Signal Sequences

If recombination is to occur in the DNA of every lymphocyte, then a mechanism must exist to ensure that it only occurs in antibody and T-cell receptor genes, and that the moving DNA segments end up exactly where they should be in the genome. Otherwise, dire consequences, including malignancy, can ensue.

In the late 1970s, investigators working with light-chain genes, described two blocks of conserved sequences—a **nonamer** (a set of nine base pairs) and a **heptamer** (a set of seven base pairs)—that are highly conserved and occur in the noncoding regions upstream of each J segment. The heptamer appeared to end exactly at the J region coding sequence. Further sequencing showed that the same motif was repeated in an inverted manner on the downstream side of the V region coding sequences, again with the heptamer sequence ending flush with the V-region gene segment (Figure 7-5a).

A further noteworthy feature of these sequences was the presence of a spacer sequence of either 12 or 23 bp between the heptamer and the nonamer. The significance of the spacer lengths was clear; they represented one, or two, turns



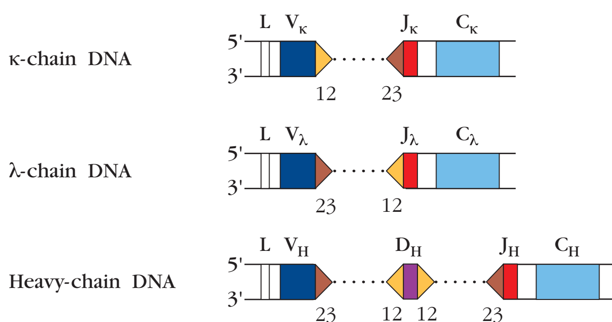


FIGURE 7-5 Two conserved sequences in light-chain and heavy-chain DNA function as recombination signal sequences (RSSs). (a) Both signal sequences consist of a conserved heptamer and conserved AT-rich nonamer; these are separated by nonconserved spacers of 12 or 23 bp. (b) The two types of RSS have characteristic locations within λ-chain, κ-chain, and heavy-chain germ-line DNA. During DNA rearrangement, gene segments adjacent to the 12-bp RSS can join only with segments adjacent to the 23-bp RSS.

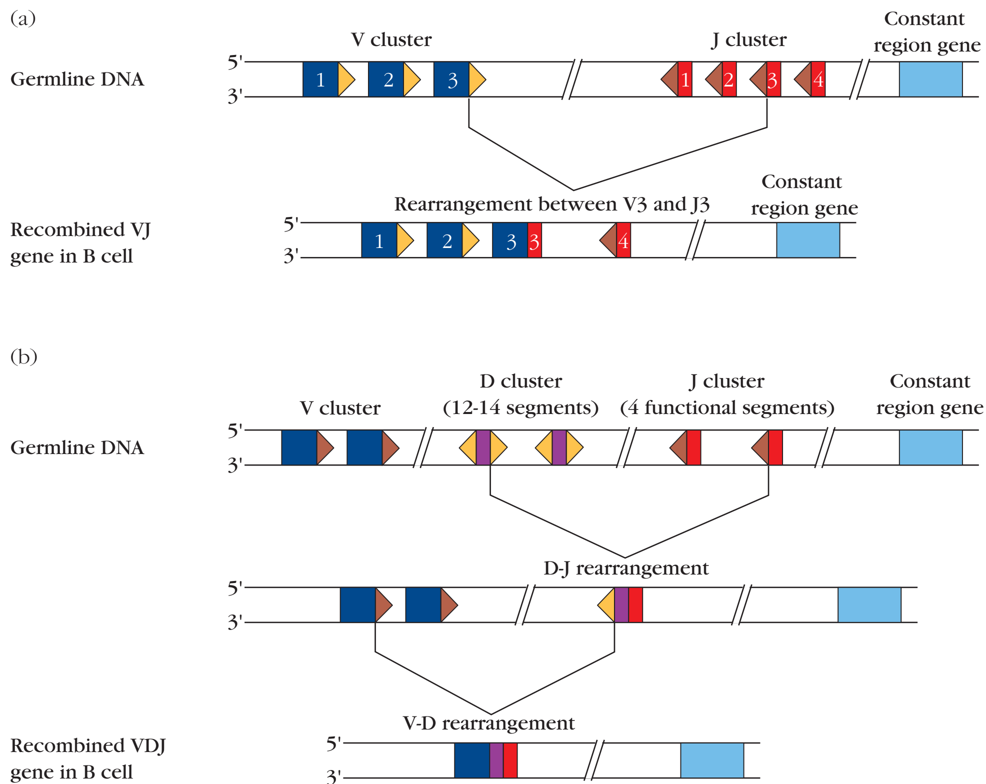


FIGURE 7-6 Recombination between gene segments is required to generate a complete light chain gene. (a) Recombination between a V region (in this case, V3) and a J region (in this case J3) generates a single VC light-chain gene in each B cell. The recombinase enzymes recognize the RSS downstream of the V region (orange triangle) and upstream of the J region (brown triangle). In every case, an RSS with a 12-bp (one-turn) spacer is paired with an RSS with a 23-bp (two-turn) spacer. This ensures that there is no inadvertent V-V or J-J joining. (b) Recombination of V (blue), D (purple), and J (red) segments creates a complete heavy-chain variable region gene. Again, the recombinase enzyme recognizes the RSS sequences downstream of the V region, up- and downstream of the D region, and upstream of the J region, pairing 23-bp spacers with 12-bp spacers.

of the double helix. Thus, the spacer sequence ensures that the ends of the nonamer and heptamer closest to the spacers would be on the same side of the double helix and, thus, accessible to binding by the same enzyme. The investigators correctly concluded that they had discovered the signal that directs recombination between the V and J gene segments and termed this heptamer 12/23 nonamer motif, the recombination signal sequence (RSS).

To summarize, the RSS consists of three elements:

- An absolutely conserved, 7-bp (heptamer) consensus sequence 5′-CACAGTG –3′
- A less conserved spacer of either 12 or 23 bp
- A second conserved, 9-bp (nonamer) consensus sequence 5′-ACAAAAACC-3′

In the heavy-chain gene segments, a similar pattern was noted. The spacer regions separating the heptamer and nonamer pairs were 23 bp in length following the V and preceding the J segments and 12 bp in length before and after the D segments. The relative locations of the 12 and 23 base pair spacers (Figure 7-5b) suggested that the VDJ recombinase enzyme is designed to bring together one sequence with a 12-bp spacer with one sequence with a 23-bp spacer, something we now know to be the case.

Figures 7-6a and 7-6b illustrate the generation of complete light-chain and heavy-chain genes from individual V and J and V, D, and J segments, respectively.

Gene Segments Are Joined by the RAG1/2 Recombinase

Two proteins, encoded by closely linked genes, ***RAG1*** **(Recombination Activating Gene 1)** and ***RAG2*** **(Recombination Activating Gene 2)**, were shown to be necessary for recombining antibody genes. The *RAG1* and *RAG2* genes are encoded just 8 kilobases apart and are transcribed together. The expression of RAG1 and RAG2 is developmentally regulated in both T and B cells (see Chapters 9 and 10) and, although RAG1 is expressed at all phases of the cell cycle, RAG2 is stable only in G_0- or G_1-phase cells. RAG1 is the predominant recombinase; it forms a complex with the RSS that is stabilized by the binding of RAG2. RAG2 by itself does not exhibit detectable RSS binding activity.

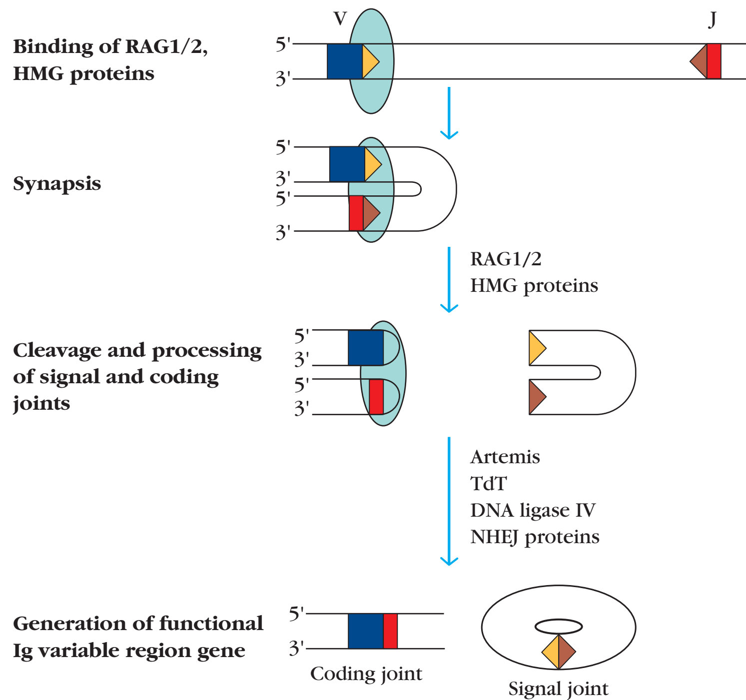


FIGURE 7-7 Overview of recombination of immunoglobulin variable region genes. The RAG1/2 complex (represented together by the green oval) binds the RSSs and catalyzes recombination. Other enzymes fill in or cleave nucleotides at the coding end, and ligase completes the process. See text for details. *[Adapted from Krangel, M. Nature Immunology 4, p. 625. 2003]*

Only three of the proteins implicated in V(D)J recombination are unique to lymphocytes: RAG1, RAG2, and ***T*erminal *d*eoxynucleotidyl *T*ransferase (TdT)**, which is responsible for the generation of additional diversity in the CDR3 region of the antibody heavy chain (as we will see below). TdT is expressed only in developing lymphocytes and adds untemplated "N" nucleotides to the free 3′ termini of coding ends following their cleavage by RAG1/2 recombinases.

Other enzymes participating in the recombination process are not lymphoid specific. Whereas binding of the RSS by RAG1/2 can occur in the absence of any other proteins, other cellular factors, most of which are part of the *N*on-*H*omologous *E*nd *J*oining (NHEJ) pathway of DNA repair are necessary for completion of V(D)J recombination. These other non-lymphocyte-specific proteins known to participate in V(D)J joining are described in Table 7-3.

V(D)J Recombination Results in a Functional Ig Variable Region Gene

The process of V(D)J recombination occurs in several phases (Figure 7-7). The end product of each successful rearrangement is an intact Ig gene, in which V and J (light chains) or V, D, and J (heavy chains) segments are brought together to create a complete heavy or light chain gene. The new joints in the antibody V region gene, created by this recombination process are referred to as **coding joints**. The joints between the heptamers from the RSSs are referred to as **signal joints**.

The first phase of this process, DNA recognition and cleavage, is catalyzed by the RAG1/2 proteins. The second phase, end processing and joining, requires a more complex set of enzymatic activities in addition to RAG1/2, including Artemis, TdT, DNA ligase IV, and other NHEJ proteins. The individual steps involved in the process of recombination between V_κ and J_κ segments are shown sequentially in Figure 7-8.

Step 1 *Recognition of the heptamer-nonamer Recombination Signal Sequence (RSS) by the RAG1/RAG2 enzyme complex.* The RAG1/2 recombinase forms a complex with the heptamer-nonamer RSSs contiguous with the two gene segments to be joined. Complex formation is initiated by recognition of the nonamer RSS sequences by RAG1 and the 12-23 rule is followed during this binding.

Step 2 *One-strand cleavage at the junction of the coding and signal sequences.* The RAG1/2 proteins then perform one of their unique functions: the creation of a single-strand nick, 5′ of the heptameric signal sequence on the coding strand of each V segment and a similar nick on the non-coding strand exactly at the heptamer-J region junction. (Figure 7-8 shows this process for the V segment only.)

Step 3 *Formation of V and J region hairpins and blunt signal ends.* The free 3′ hydroxyl group at the end of the coding strand of the V_K segment now attacks the phosphate group on the opposite, non-coding V_K strand, forming a new covalent bond *across* the double helix and yielding a DNA hairpin structure on the V-segment side of the break (coding end). Simultaneously, a blunt DNA end is formed at

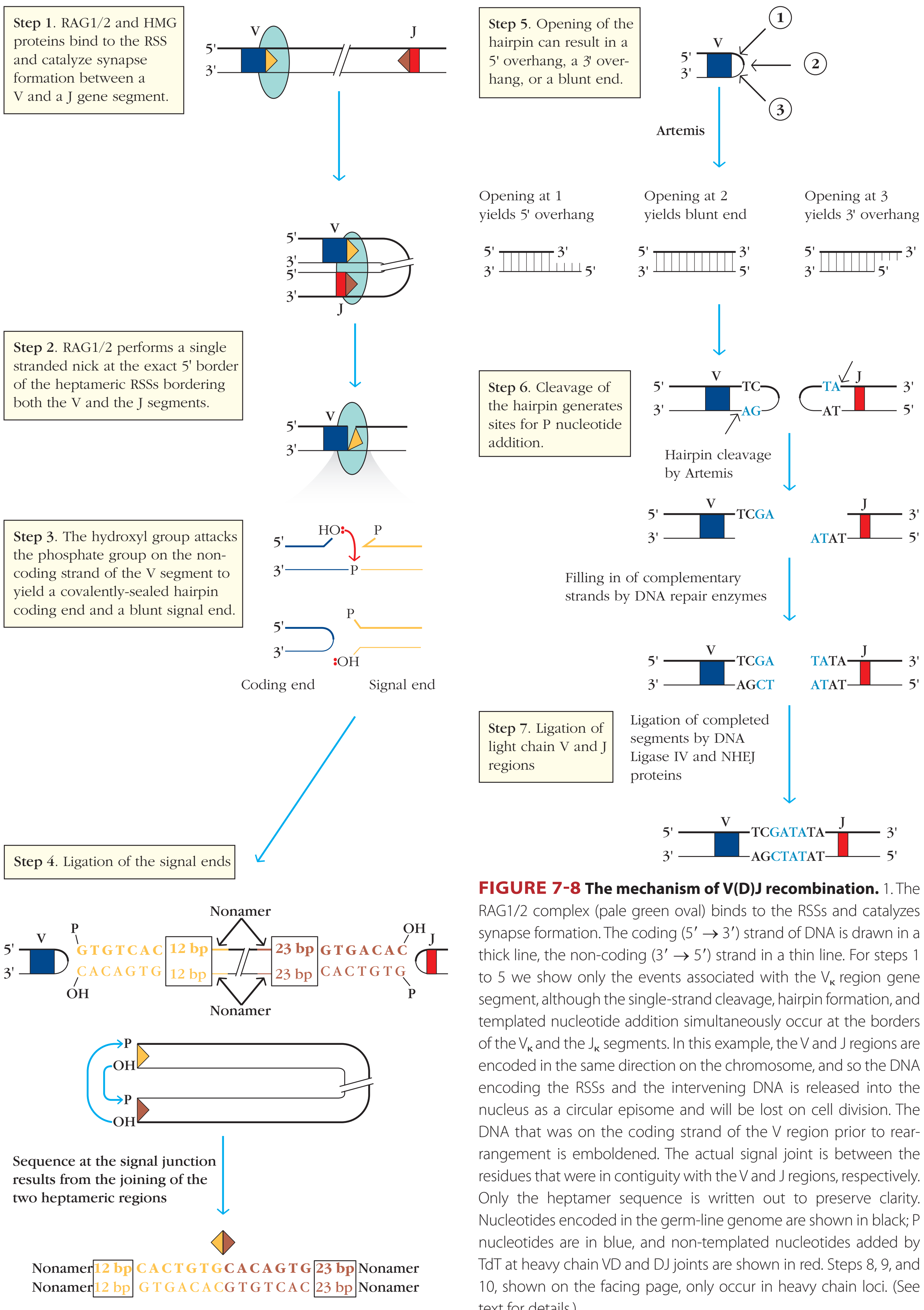
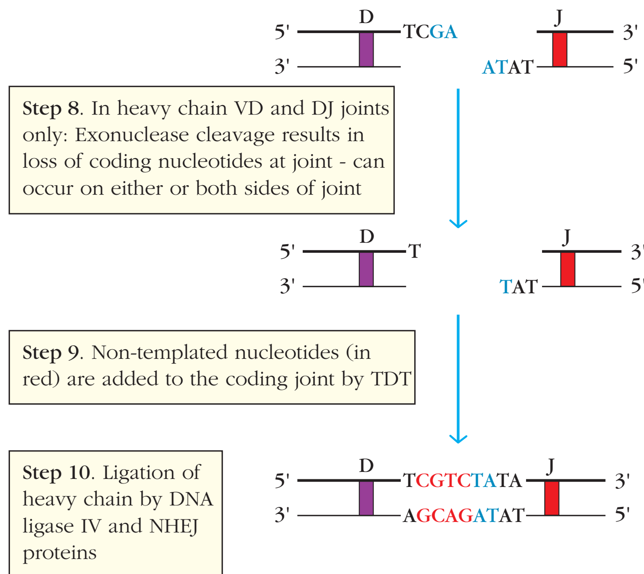


FIGURE 7-8 The mechanism of V(D)J recombination. 1. The RAG1/2 complex (pale green oval) binds to the RSSs and catalyzes synapse formation. The coding (5′ → 3′) strand of DNA is drawn in a thick line, the non-coding (3′ → 5′) strand in a thin line. For steps 1 to 5 we show only the events associated with the V_κ region gene segment, although the single-strand cleavage, hairpin formation, and templated nucleotide addition simultaneously occur at the borders of the V_κ and the J_κ segments. In this example, the V and J regions are encoded in the same direction on the chromosome, and so the DNA encoding the RSSs and the intervening DNA is released into the nucleus as a circular episome and will be lost on cell division. The DNA that was on the coding strand of the V region prior to rearrangement is emboldened. The actual signal joint is between the residues that were in contiguity with the V and J regions, respectively. Only the heptamer sequence is written out to preserve clarity. Nucleotides encoded in the germ-line genome are shown in black; P nucleotides are in blue, and non-templated nucleotides added by TdT at heavy chain VD and DJ joints are shown in red. Steps 8, 9, and 10, shown on the facing page, only occur in heavy chain loci. (See text for details.)



the edge of the heptameric signal sequence. The same process occurs simultaneously on the J_κ side of the incipient joint. At this stage, the RAG1/2 proteins and HMG proteins are still associated with the coding and signal ends of both the V and J segments in a postcleavage complex.

Step 4 *Ligation of the signal ends.* DNA ligase IV then ligates the free blunt ends to form the signal joint. The involvement of particular enzymes in this process was deduced from observations of V(D)J recombination in natural and artificially generated systems lacking one or more enzymes (see Table 7-3).

Step 5 *Hairpin cleavage.* The hairpins at the ends of the V and J regions are now opened in one of three ways. The identical bond that was formed by the reaction described in step 3 above, may be reopened to create a blunt end at the coding joint. Alternatively, the hairpin may be opened asymmetrically on the "top" or on the "bottom" strand, to yield a 5′ or a 3′ overhang, respectively. A 3′ overhang is more common in in vivo experiments. Hairpin opening is catalyzed by **Artemis**, a member of the NHEJ pathway.

Step 6 *Overhang extension, leading to palindromic nucleotides.* In Ig light-chain rearrangements, the resulting overhangs can act as substrates for extension DNA repair enzymes, leading to double stranded ***p*alindromic (P) nucleotides** at the coding joint. For example, the top row of bases in the V region in the 5′ to 3′ direction reads TCGA. Reading backward on the bottom strand from the point of ligation also yields TCGA. The palindromic nature of the bases at this joint is a direct function of an asymmetric hairpin opening reaction. P nucleotide addition can also occur at both the VD and DJ joints of the heavy-chain gene segments, but, as described below, other processes can intervene to add further diversity at the V_H-D and D-J_H junctions.

Step 7 *Ligation of light-chain V and J Segments.* Members of the NHEJ pathway repair both the signal and the coding joints, but the precise roles of each, and potentially other enzymes in this process, have yet to be fully characterized.

During B-cell development, Ig heavy-chain genes are rearranged first, followed by the light-chain genes. This temporal dissociation of the two processes enables two additional diversifying mechanisms to act on heavy-chain V region segments. The enzymes responsible for these mechanisms are usually turned off before light-chain rearrangements begin.

Comparative sequence analysis of germ-line and mature B-cell Ig genes demonstrated that both loss of templated nucleotides (nucleotides found in the germ-line DNA) and addition of untemplated nucleotides (nucleotides not found in germ-line DNA) could be identified in heavy-chain sequences. Two distinct enzyme-catalyzed activities are responsible for these findings.

Step 8 *Exonuclease trimming.* An exonuclease activity, which has yet to be identified, trims back the edges of the V region DNA joints. Since the RAG proteins themselves can trim DNA near a 3′ flap, it is possible that the RAG proteins may cut off some of the lost nucleotides. Alternatively, Artemis has also been shown to have exonuclease, as well as endonuclease activity, and could be the enzyme responsible for the V(D)J-associated exonuclease function. Exonuclease trimming does not necessarily occur in sets of three nucleotides, and so can lead to out-of-phase joining. V segment sequences in which trimming has caused the loss of the correct reading frame for the chain cannot encode antibody molecules, and such rearrangements are said to be **unproductive**. As mentioned above, such exonuclease trimming is more common at the two heavy-chain V gene joints (V-D and D-J) than at the light-chain V-J joint. In cases where trimming is extensive, it can lead to the loss of the entire D region as well as the elimination of any P nucleotides formed as a result of asymmetric hairpin cleavage.

Step 9 *N nucleotide addition.* **Non-templated (N) nucleotides** are added by TdT to the coding joints of heavy-chain genes after hairpin cleavage. This enzyme can add up to 20 nucleotides to each side of the joint. The two ends are held together throughout this process by the enzyme complex, and again, loss of the correct phase may occur if nucleotides are not added in the correct multiples of three required to preserve the reading frame.

Step 10 *Ligation and repair of the heavy-chain gene.* This occurs as for the light-chain genes.

In describing V(D)J recombination, investigators must explain not only the mechanism of RSS recognition, cleavage, and ligation but also address the question of how two RSSs, located many kilobases distant from one another in the linear DNA sequence, are brought into close apposition. Furthermore, once successful recombination has occurred on one heavy-chain and one light-chain allele, this information must be communicated to the homologous chromosome, so that the other alleles can be silenced. How does this occur?

Recent research indicates that both the structure and the location of recombinationally active, Ig V region DNA

change significantly as B-cell development proceeds, and that some of these changes are signaled by epigenetic alterations in the chromatin structure, mediated by specific methylation reactions on associated histone residues. DNA close to the nuclear membrane (pericentric DNA) does not include recombinationally active V regions; rather, DNA undergoing recombination moves away from the nuclear membrane, toward the center of the nucleus. The structure of the chromatin undergoing recombination also alters, such that long stretches of DNA are condensed into loops that allow the recombining sequences to come into closer contact with one another. Once successful recombination has occurred, the inactive allele has been shown to migrate to the pericentric regions. Thus, the recombination process is occurring within an actively modulating nucleoplasmic context, and is controlled by enzymes that alter chromatin structure, in addition to those that cleave and recombine the DNA.

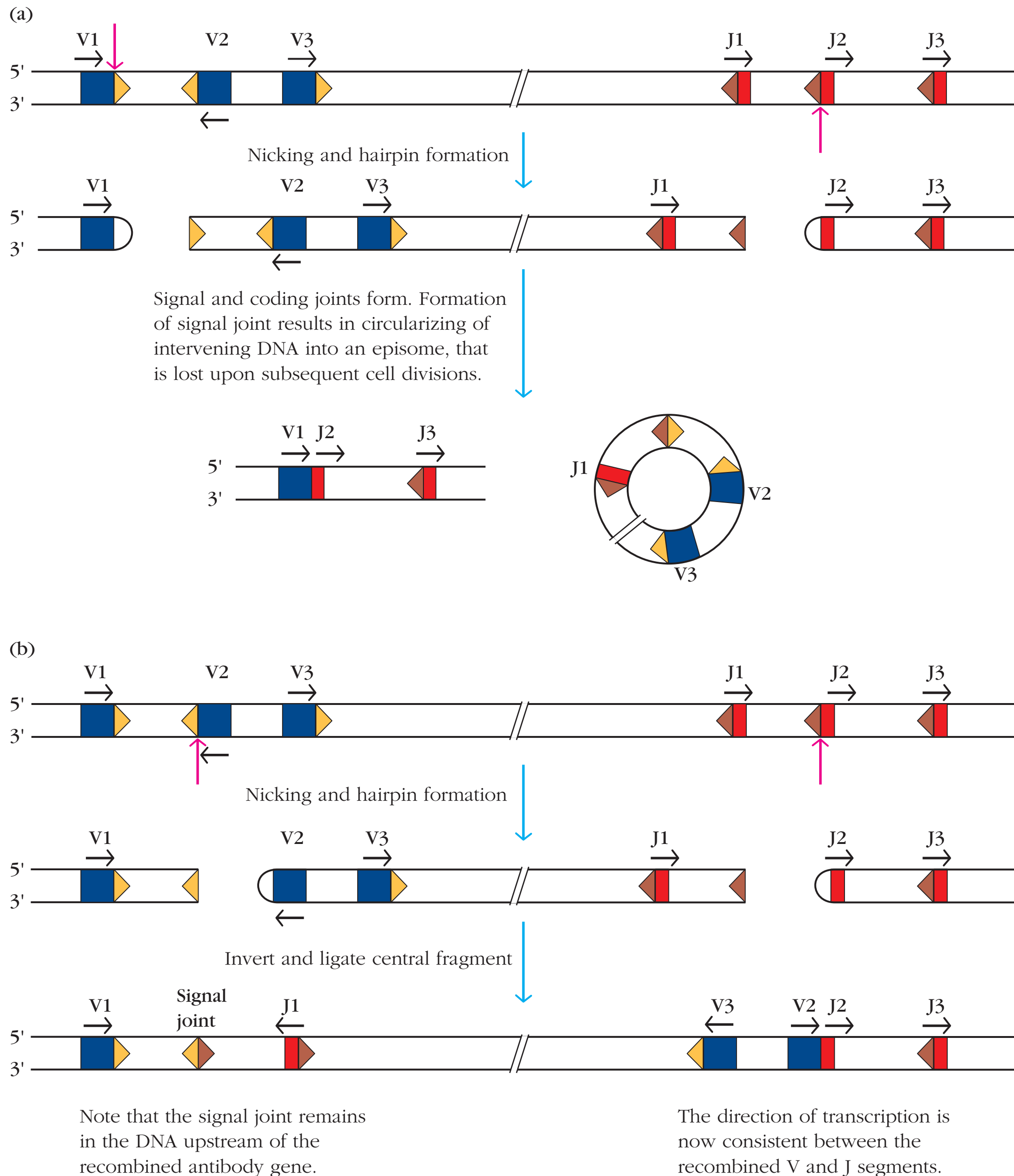


FIGURE 7-9 Recombination can occur between DNA segments aligned in the same, or opposite, transcriptional direction on the chromosome. (a) Recombination is shown occurring between V1 and J2, which are both encoded in the same transcriptional orientation, from left to right. The intervening DNA is excised as a circular episome. (b) Recombination is shown occurring between V2 and J2, which are encoded in opposite transcriptional orientations. In this case, the DNA containing the signal joint remains inverted in the DNA upstream of the recombined pair.

TABLE 7-4 Combinatorial antibody diversity in humans

Nature of segment	Number of heavy-chain segments (estimated)	Number of κ-chain segments (estimated)	Number of λ-chain segments (estimated)
V	41	41	33
D	23		
J	6	5	5
Possible number of combinations	$41 \times 23 \times 6 = 5658$	$41 \times 5 = 205$	$30 \times 5 = 165$
Possible number of heavy-light chain combinations in the human $= 5658 \times (205 + 165) = 2.09 \times 10^6$			

V(D)J Recombination Can Occur between Segments Transcribed in Either the Same or Opposite Directions

Sequencing of chromosomal fragments containing multiple antibody V region genes showed that transcription of some V region genes occurs in a direction opposite from that of downstream D and J segments. However, work with artificial recombinase substrates has demonstrated that the same general mechanism of recombination is used, regardless of the transcriptional direction of the gene segments involved. When recombination occurs between two gene segments that are transcribed in the same direction, the intervening DNA is deleted and lost (Figure 7-9a). When they are transcribed in opposite directions, the DNA that was located between the V and J segments is retained, in an inverted orientation, in the DNA upstream of the rearranged VJ region (Figure 7-9b).

Five Mechanisms Generate Antibody Diversity in Naïve B Cells

This description of the complex and sophisticated apparatus by which Ig genes are created allows us to understand how such an immensely diversified antibody repertoire can be generated from a finite amount of genetic material. To summarize, the diversity of the naïve BCR repertoire is shaped by the following mechanisms:

1. *Multiple gene segments* exist at heavy (V, D, and J) and light-chain (V and J) loci. These can be combined with one another to provide extensive combinatorial diversity (Table 7-4).
2. *P nucleotide addition* results when the DNA hairpin at the coding joint is cleaved asymmetrically. Filling in the single-stranded DNA piece resulting from this asymmetric cleavage generates a short palindromic sequence.
3. *Exonuclease trimming* sometimes occurs at the VDJ and VJ junctions, causing loss of nucleotides.
4. *Non-templated N nucleotide addition* in heavy chains results from TdT activity. Mechanisms 2, 3, and 4 give rise to extra diversity at the junctions between gene segments, which contribute to CDR3.

   In addition to these four mechanisms for generating antibody diversity that operate on individual heavy- or light-chain variable segments, the combination of different heavy and light-chain pairs to form a complete antibody molecule provides further opportunities for increasing the number of available antibody combining sites (Table 7-4).
5. *Combinatorial diversity:* The same heavy chain can combine with different light chains, and vice versa.

These five mechanisms are responsible for the creation of the diverse repertoire of BCRs and antibodies that is available to the organism before any contact with pathogens or antigen has occurred. Following antigenic stimulation, B cells are able to use yet another mechanism, unique to the immune system, to further diversify and refine the antigen-specific receptors and antibodies: *somatic hypermutation*. As described in Chapter 12, a specialized enzyme complex targets the genes encoding the variable regions of Ig genes *only in those B cells that have undergone antigen-specific activation in the presence of T-cell help*. B cells are exposed to successive cycles of mutation at the BCR loci, followed by antigen-mediated selection in specialized regions of the lymph node and spleen. The end result of this process is that the average affinity of antigen-specific BCRs and antibodies formed at the end of an immune response is considerably higher than that at its instigation. This process is referred to as **affinity maturation**.

This section has described the process of the generation of the primary Ig variable region repertoire as it occurs in humans and rodents. Although the same principles apply to most vertebrate species, different species have evolved their own variations. For example, the process of gene conversion is used in chickens, and some species, such as sheep and cows, use somatic hypermutation in the generation of the primary as well as the antigen-experienced repertoire.

In Evolution Box 7-2, we describe the evolution of this system of recombined lymphocyte receptors, addressing the current hypothesis that the key event was the introduction of the *RAG1/2* gene segment into the early vertebrate genome as a transposon.

EVOLUTION

Evolution of Recombined Lymphocyte Receptors

Scientists are slowly beginning to understand how the process of V(D)J recombination may have evolved. With that understanding has come an appreciation for the fact that the BCRs and TCRs of the vertebrate adaptive immune system may not be the only immune receptor molecules generated by recombinatorial genetics.

Ig-like receptors have been identified in species as ancient as the earliest jawed vertebrates. However, extensive analyses of the only surviving jawless vertebrates (Agnathans), the hagfish and lamprey, show no evidence of TCR or BCR V(D)J segments, RAG1/2 genes, or genes encoding elements of a primitive *M*ajor *H*isto-*C*ompatibility *C*omplex (MHC) system. This suggests that the RSS-based, recombinatorially generated set of adaptive immune receptors first emerged in a common ancestor of jawed vertebrates, most probably more than 500 million years ago (Figure 1).

The driving force in the development of the adaptive immune system was most probably the incorporation of an ancestor of the RAG1 genes into the ancestral genome in the form of a transposable element. This hypothesis is supported by several observations and experiments:

- The DNA-binding region of RAG1 is strikingly homologous to that of known transposable elements and is evolutionarily related to members of the *Transib* transposase family, currently expressed in species such as fruit flies, mosquitoes, silkworm, and the red flour beetle.
- The inverted repeats in the RSSs of the Ig and TCR gene segments are structurally similar to those found in other transposons.
- The mechanism of action of the RAG proteins involves formation of a DNA hairpin intermediate reminiscent of the action of certain transposases. For example, another phylogenetically conserved transposon in flies, the HERMES transposon, has been shown to induce a double-strand break via the hairpin-formation mechanism described above for V(D)J recombination.
- RAG proteins have the ability to transpose an RSS-containing segment of DNA to an unrelated target DNA in vitro.

Current thinking supports the hypothesis that the appearance of the transposon in a primordial antigen receptor gene might have separated the receptor gene into two or more pieces. This hypothesis is supported by the presence, in lower chordates, of genes that are related to BCR and TCR V regions, and that therefore could

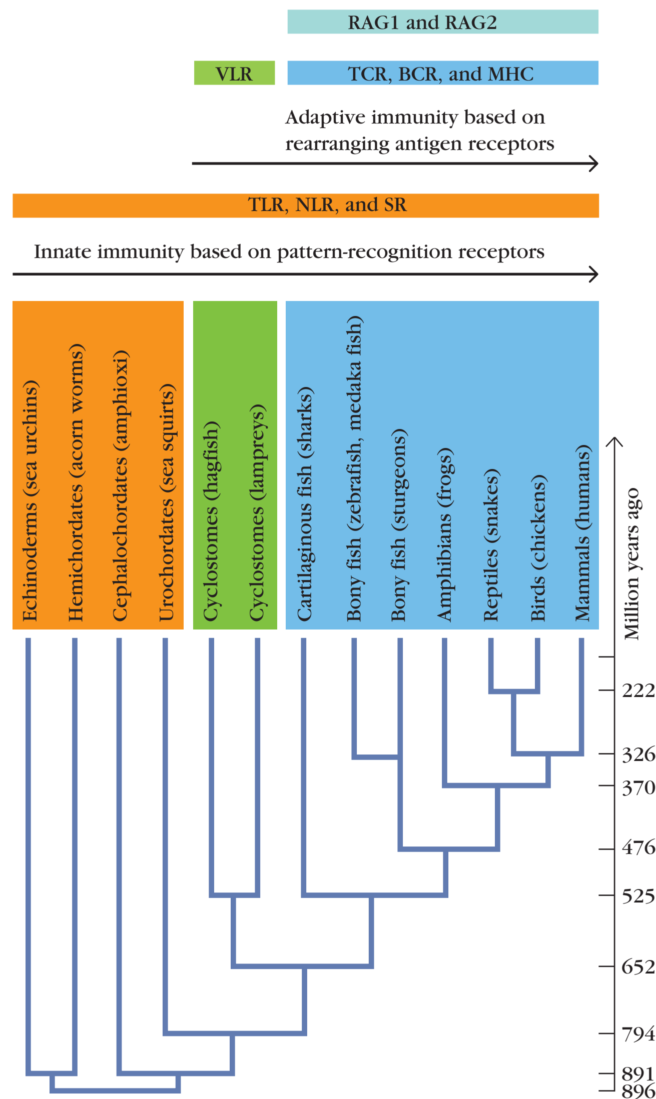


FIGURE 1

The emergence of the RSS-dependent recombination-based adaptive immune system coincides with the first appearance of the jawed vertebrates. SR = scavenger receptor. *[Adapted from Figure 1, M. F. Flajnik and M. Kasahara, 2010, Origin and evolution of the adaptive immune system: Genetic events and selective pressures,* Nature Reviews Genetics **11:**47–59.]

BOX 7-2

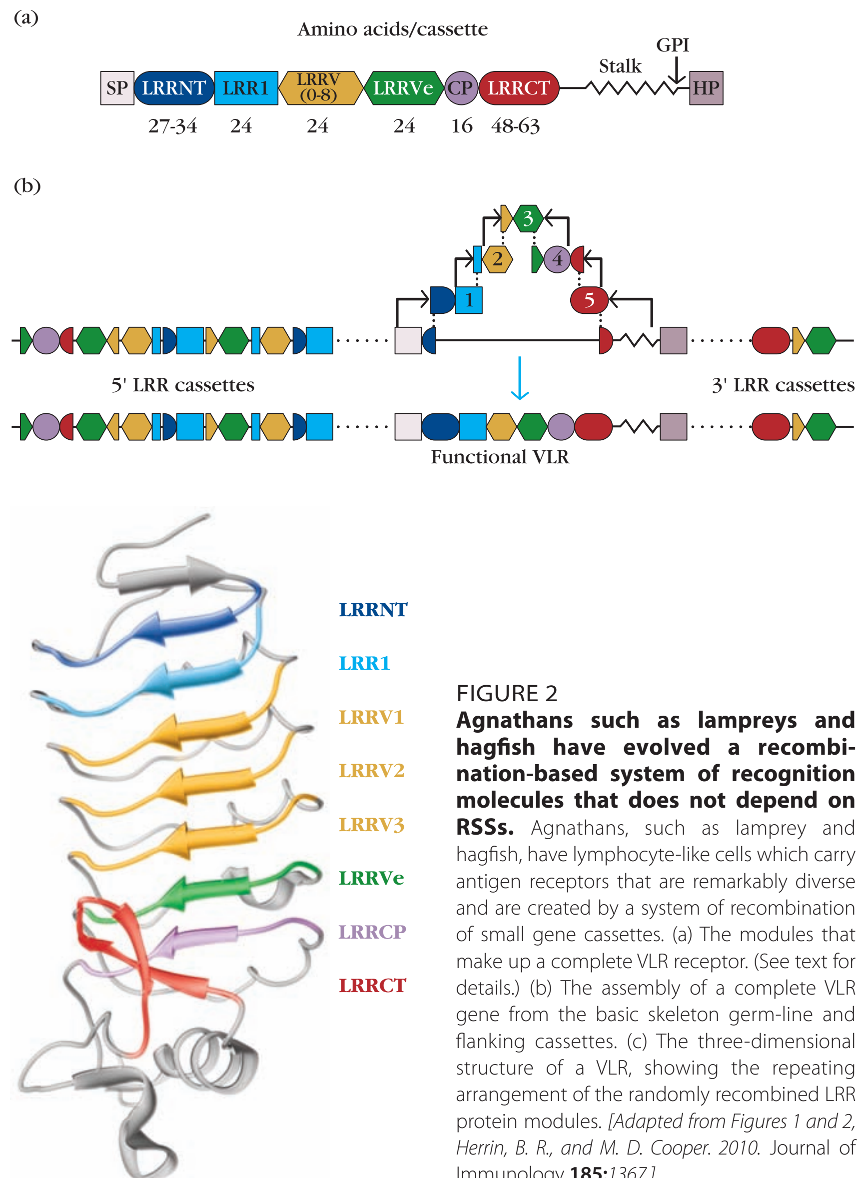


FIGURE 2
Agnathans such as lampreys and hagfish have evolved a recombination-based system of recognition molecules that does not depend on RSSs. Agnathans, such as lamprey and hagfish, have lymphocyte-like cells which carry antigen receptors that are remarkably diverse and are created by a system of recombination of small gene cassettes. (a) The modules that make up a complete VLR receptor. (See text for details.) (b) The assembly of a complete VLR gene from the basic skeleton germ-line and flanking cassettes. (c) The three-dimensional structure of a VLR, showing the repeating arrangement of the randomly recombined LRR protein modules. *[Adapted from Figures 1 and 2, Herrin, B. R., and M. D. Cooper. 2010.* Journal of Immunology **185:***1367.]*

have provided the substrate for the transposon invasion. When the transposon subsequently jumped out of the gene, the RSSs would have been left behind. Multiple such rounds of transposition followed by gene duplication would have given rise to the TCR and BCR loci we know today. Interestingly, our knowledge of phylogeny suggests that two waves of gene duplications occurred at the time of vertebrate origin, right around the time of the proposed transposon entry.

Much still remains unknown about the evolution of the adaptive immune receptor molecules. What was the nature of the primordial antigen receptor that was the target for the first transposition? At what stage did the RAG2 gene become associated with the RAG1 component? And how do we account for the evolution of the Ig receptors of species such as the shark, in which fully rearranged V(D)J genes exist in the germ-line genome?

Recent evidence has suggested that the recombinatorial strategy for generating antigen receptors is not limited to the RSS-based Ig and TCR systems of vertebrates. Jawless fish, such as lampreys and hagfish, possess lymphocyte-like cells that can be stimulated to divide and differentiate. Moreover, these cells release specific agglutinins after immunization with antigens, and higher serum levels of these agglutinins are secreted after a secondary, than after a primary, immunization, suggesting that immunological memory exists in these fish. These lymphocyte-like cells therefore appear to express all the hallmarks of adaptive immunity, and yet the agglutinins they secrete do not appear to have an Ig-like structure.

Analysis of the agglutinins has revealed that activated lamprey "lymphocytes" express abundant quantities of *l*eucine-*r*ich-*r*epeat (LRR)-containing proteins. LRRs are protein motifs that are frequently associated with protein-protein recognition, and such motifs have already been encountered in the context of the Toll-like receptors described in Chapter 5. The lamprey agglutinin receptors are generated by recombination of gene segments during lymphocyte development, and the LRR receptor repertoires of these fish are remarkably diverse. These receptors have therefore been named *v*ariable *l*ymphocyte *r*eceptor (VLR) molecules.

Figure 2a shows an example of the arrangement of the protein modules in one of these VLR molecules. At the amino terminal of the protein is an invariant signal peptide, followed by a 27–34 residue *N*-*T*erminal *L*eucine-*R*ich *R*epeat (LRRNT). This is followed by the first of several 24-residue *L*eucine-*R*ich *R*epeat (LRR) modules, LRR1, connected to a series of up to eight 24-residue LRR modules with *vari*able sequences (LRRVs). At the C-terminal of the molecule, a 24-residue *end LRRV* segment (LRRVe) is attached to a short 16-residue *c*onnecting *p*eptide (CP), which culminates in a 48–63 residue *C*-*t*erminal LRR (LRRCT). The VLR molecules are attached to the lymphocyte membrane by an invariant stalk, rich in threonine and proline residues, connecting to a *g*lycosyl-*p*hosphatidyl-*i*nositol

(continued)

EVOLUTION *(continued)* BOX 7-2

(GPI) anchor and a hydrophobic peptide. Lymphocyte activation leads to phospholipase cleavage at the GPI anchor, enabling soluble forms of the VLRs to be released from the lymphocyte following antigen stimulation.

The assembly process that generates the completed protein occurs only in lymphocytes. In the germ-line DNA, LRR cassettes flank the skeleton *LRR* genes, which consist initially only of sequences coding for parts of the LRRNT and LRRCT, separated by non-coding DNA (Figure 2b). During lymphocyte development, the non-coding sequence is replaced with variable LRRs. Gene segments are copied from one part of the genome into another in a one-way process, similar to gene conversion. Each lamprey lymphocyte expresses a unique VLR gene from a single allele, and the diversity of the repertoire is limited only by the number of lymphocytes.

Figure 2c shows a cartoon of the structure of one of these receptors, generated from an x-ray crystallographic model. Comparison of the sequences of several of these primarily β-stranded structures has shown that their sequence diversity is concentrated on the concave (left) side of the molecule and can be attributed to the inherent diversity of the LRR cassettes.

Did the two LRR- and V(D)J-based, recombinatorially derived sets of immune receptors exist side by side in an ancestral species? This we do not know. Agnathans other than lampreys and hagfish were extinguished 400 million years ago, and we have access only to the fossil remains of ostracoderms (agnathans with dermal skeletons), which are thought to be the ancestors of the gnathostomes. One theory would hold that the VLR recombinatorial system evolved in an ancestor common only to hagfish and lampreys, and could have been causal in their ability to survive when other agnathan species succumbed to environmental insult. Alternatively, the VLR and Ig-based systems may have co-existed for a time, with lymphocytes bearing both types of receptors, until one of the systems was lost.

Combinatorial genetics in the generation of the vertebrate antigen receptors may not be as unique a mechanism as immunologists had previously thought.

B-Cell Receptor Expression

The expression of a receptor on the surface of a B cell is the end result of a complex and tightly regulated series of events. First, the cell must ensure that the various gene recombination events culminate in productive rearrangements of both the heavy- and light-chain loci. Second, only one heavy-chain and one light-chain allele must be expressed in each B cell. Finally, the receptor must be tested to ensure it does not bind self antigens, in order to protect the host against the generation of an autoimmune response.

Allelic Exclusion Ensures That Each B Cell Synthesizes Only One Heavy Chain and One Light Chain

The random nature of Ig heavy- and light-chain gene rearrangement means that more than one heavy-light-chain pair could potentially be expressed on the surface of individual B cells. Furthermore, since each heavy chain could potentially combine with both light chains and vice versa, this could result in the creation of B cells bearing a variety of different antigen-binding sites. Whereas this opportunity to increase the number of available receptors may initially sound advantageous to the organism, in practice the presence of more than one receptor per B cell creates prohibitive difficulties for those mechanisms that protect against the generation of autoimmunity. The mechanism by which B cells ensure that only one heavy- and one light-chain allele are transcribed and translated is referred to as **allelic exclusion**.

The rearrangement of *Ig* genes occurs in an ordered way, and begins with recombination at one of the two homologous chromosomes carrying the heavy-chain loci. The production of a complete heavy chain and its expression on the B-cell surface in concert with a surrogate light chain, made up of two proteins, VpreB and λ5 (Figure 7-10), signals the end of heavy-chain gene rearrangement and, thus, only one antibody heavy chain is allowed to complete the rearrangement process. The mechanism by which this allelic exclusion occurs is still under investigation. However, we do know that following successful arrangement at one allele, the V_H gene locus on the other chromosomes is methylated and recruited to heterochromatin.

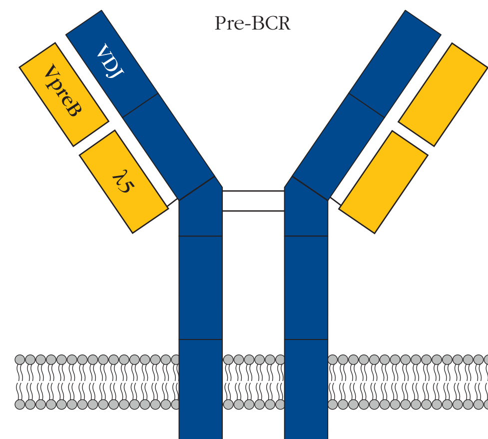


FIGURE 7-10 Ig heavy-chain expression on the cell surface in concert with VpreB and λ5 signals, via Igα/Igβ, the end of heavy-chain rearrangement.