Underlying the development of antibody diversity is a unique pattern of gene organization and molecular events. THREE FAMILIES OF IMMUNOGLOBULIN GENES exist in mammals, one encoding HEAVY chains, another KAPPA chains, and the third LAMBDA chains. Each of these clusters contains one or more constant region genes and a number of variable region gene segments. The formation of a complete variable region of a light or heavy chain requires the joining of two or three separate genetic elements by a process of GENE REARRANGEMENT; a separate DNA rearrangement in the heavy-chain complex is required for subsequent CLASS-SWITCHING. Both germ-line and somatic events contribute to antibody diversity, including COMBINATORIAL JOINING, SOMATIC MUTATION and COMBINATORIAL ASSOCIATION. This gene organization and the requirement for gene rearrangement is unique to immunoglobulins, with the single exception of the related family of genes encoding the T-CELL RECEPTORS, whose functions will be discussed later.

The problem of the genetic basis of antibody diversity is one which was hotly debated through the 1960's and 70's, and only through the application of recombinant DNA technology have its essential features become well understood.

The basic problem can be set out as follows: if there are one million different antibodies which the immune system can produce, and if each of them has a unique primary structure (i.e., amino acid sequence), are there then a million genes required for their production, or can a smaller number of genes be modified in some systematic way to account for the total diversity? These two possibilities define the essential features of two competing hypotheses, GERM LINE generation of diversity on the one hand versus SOMATIC generation of diversity on the other.

We can simplify the problem by recognizing that the antibody combining site (which is, of course, responsible for specificity) is made up of two elements, the $V_H$ and $V_L$. If as few as one thousand $V_L$ domains combine randomly with the same number of $V_H$ domains, we could account for one million different combining sites ($10^6$).

If we use the kappa light chain system as an example (recognizing that the same arguments can be applied to the other V-region families), we can define the extreme forms of GERM LINE and SOMATIC theories as follows:

GERM-LINE THEORY -- For every kappa-chain V-region there exists one unique germ-line gene. A particular antibody-forming cell selects one of these and expresses it in unmodified form.

SOMATIC THEORY -- Only a single germ-line gene exists for all kappa-chain V-regions. A particular antibody-forming cell expresses this gene following a process of somatic mutation, which results in each cell expressing a different version of this gene.
Decades of genetic studies, culminating in the cloning and sequencing of immunoglobulin genes, have shown that significant features of both theories are correct. Many genes exist for V-regions of immunoglobulins, and these genes are somatically modified in a variety of ways in the course of their expression in antibody-forming cells.

STRUCTURE AND EXPRESSION OF IMMUNOGLOBULIN GENES

Three families of immunoglobulin genes exist, each on a separate chromosome. One includes kappa genes, another lambda genes, and the third includes all the heavy chain genes. Each family consists of a series of V-regions genetically linked (but still separated by long stretches of DNA) to one or more C-regions.

We can diagram the three human Ig clusters as follows:

<table>
<thead>
<tr>
<th>kappa:</th>
<th>lambda:</th>
<th>heavy:</th>
</tr>
</thead>
<tbody>
<tr>
<td>VK1</td>
<td>V(\lambda)1</td>
<td>Vh1</td>
</tr>
<tr>
<td>VK2</td>
<td>V(\lambda)2</td>
<td>Vh2</td>
</tr>
<tr>
<td>VK(n)</td>
<td>V(\lambda)(n)</td>
<td>Vh(n)</td>
</tr>
<tr>
<td>J(\kappa)</td>
<td>J(\lambda)</td>
<td>D(h)</td>
</tr>
<tr>
<td>C(\kappa)</td>
<td>C(\lambda)</td>
<td>Jh</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C(\mu)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C(\delta)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C(\gamma)3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C(\gamma)1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C(\alpha)2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C(\alpha)1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C(\gamma)2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C(\gamma)4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C(\epsilon)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C(\alpha)2</td>
</tr>
</tbody>
</table>

Figure 8-1

Let’s examine in some detail the structure of the kappa chain complex and the processes involved in its expression, referring to Fig. 8-2 below. This complex consists of a large number of V-region genes (about 55) genetically linked (by a long stretch of DNA) to a single copy of the constant region gene. An additional cluster of five short gene segments called J-segments is located a few thousand base pairs upstream (5’ direction) of the C-region gene, and each of these codes for the last 13 amino acids of the variable region. (NOTE: Don’t confuse these "J-segments" with the "J-chain," the polypeptide attached to IgM and polymeric IgA, which we will discuss in Chapter 9.) This is the "germ-line" configuration, present in germ cells (sperm and eggs) as well as all somatic cells other than Ig-producing cells, and is shown at the top of Figure 8-2.

The first step in expression of this gene is DNA REARRANGEMENT, involving the joining of one V-region and one J-segment, each chosen at random in any given B-cell. The result is the structure in line 2 of the figure, and the DNA which was originally between the selected V and J genes is cut out and lost in the form of a closed circular molecule; other V-region
MOLECULAR BASIS OF KAPPA GENE EXPRESSION

Germ-line configuration:

1) Germ-line DNA

Following DNA rearrangement:

2) Rearranged DNA

Following transcription:

3) Precursor mRNA

Following RNA splicing:

4) Mature mRNA

Following translation:

5) Precursor PEPTIDE

Following peptide processing:

6) Mature PEPTIDE

Figure 8-2
gene segments which happen to reside outside this excised segment (to the left, or 5’, of the V-region which is used (for example, V1 in line 2 of Figure 8-2) are retained, although they are no longer relevant to expression, described below. This process is unique to immunoglobulin (and T-cell receptor) genes.

The process of TRANSCRIPTION starts at the beginning of the rearranged V-region and continues past the end of the C-region, resulting in an immature mRNA whose structure is shown in line 3. The large intervening sequence (“intron”) between the J-segments and the C-region is removed by the process of RNA SPlicing, resulting in the mature mRNA shown in line 4. It includes a 200 nucleotide region at its 3’ end which is not translated, and a 3’ poly-A “tail”. The structure of this transcript and the processes by which it was produced (transcription and RNA splicing) are characteristic of most eukaryotic genes.

TRANSLATION of this message occurs on ribosomes associated with the rough endoplasmic reticulum (RER), and the resulting polypeptide has the structure shown in line 5. It is identical to the known structure of immunoglobulin kappa chains with a single exception -- it has an additional 13 or so amino acids at its amino terminal end, known as the LEADER or SIGNAL sequence. This sequence is required for the transport of the polypeptide across the endoplasmic reticulum membrane into the lumen of the ER, and is cleaved off when the polypeptide moves into the cisterna of the rough ER. The presence of the leader sequence and its proteolytic cleavage are general features of all secreted proteins in eukaryotes and prokaryotes.

The structure and mechanism of expression of lambda chains and heavy chains are similar to what we have just described for kappa chains--all have J-SEGMENTS, all show DNA REARRANGEMENT, TRANSCRIPTION, RNA SPlicing, TRANSLATION and proteolytic cleavage of the LEADER POLYPEPTIDE. Heavy chain gene structure is somewhat more complex, however, as there also exists an additional cluster of gene segments (known as “D”, for “diversity”, segments) which each encodes four amino acids between the V-region cluster and the J-segments. DNA rearrangement for H-chains thus involves two events, joining of a V with a D, and joining of the D with a J-segment. Transcription and the other processes discussed above take place as they do for kappa genes.

In each case, the end result is a polypeptide whose amino acid sequence has been determined by three or four separate genetic elements, and which is incorporated into the final immunoglobulin molecule by processes which will be discussed in Chapter 9.

ALTERNATE SPlicing IN B-CELLS

One unusual phenomenon still needs to be explained, i.e., the simultaneous synthesis of IgM and IgD by a single B-cell. This is the only example of a normal cell simultaneously producing two different kinds of immunoglobulin. The explanation derives from the fact that mu (µ) and delta (δ) constant region genes are adjacent to one another in the heavy chain gene complex, as illustrated in Figure 8-3. Using the same rearranged heavy chain V/D/J complex, a B-cell can make two kinds of mRNA--it can transcribe from the V-region through the end of the Cµ gene and make IgM, or it can transcribe all the way through the Cδ gene and make IgD by splicing out the Cµ region together with the intervening sequence during RNA splicing. Two different mRNAs can thus be made from a single gene complex. It should be emphasized that alternate splicing of RNA is a mechanism used by many other genes to generate diverse protein products.
SIMULTANEOUS SYNTHESIS OF IgM AND IgD IN B-CELLS
BY ALTERNATE RNA SPLICING

This process accounts for two important phenomena. First, IgM and IgD synthesis in B-cells can occur simultaneously and continuously (it is not a transient result of a switch from mu to delta). And second, it has been shown that the mu and delta chains are produced from the same chromosome, and not from the two different allelic copies. This is an extension of allelic exclusion which will be discussed in Chapter 9, and is known as haplotype exclusion.

Alternate splicing of mRNA also accounts for another important feature of immunoglobulin expression, namely the choice of whether a secreted versus a membrane-bound form of Ig is produced. All classes of Ig can be produced either as membrane-bound molecules expressed on the surface of B-cells, or as secreted molecules released into the serum or extracellular space. This difference is determined by which of two alternate exons is selected to be present at the 3' end of the heavy chain mRNA, which in turn depends on alternate mRNA splicing. Regulation of such mRNA splicing is therefore an important element in the differentiation of the B-cell lineage.

HEAVY CHAIN CLASS SWITCHING: A SECOND DNA REARRANGEMENT

While antibody-secreting cells begin their existence by secreting IgM, the progeny of an individual cell can go on to produce IgG, IgE or IgA. This phenomenon is known as CLASS SWITCHING and occurs by another DNA rearrangement event (shown in Figure 8-4, below), which relocates the already rearranged V/D/J complex from its original position near the Cµ gene to a position close to one of the other heavy-chain C-regions (γ3 in this example). This results in a new transcription unit and the synthesis of a heavy chain with same V-region but a new C-region.
MOLECULAR BASIS OF CLASS SWITCHING

Figure 8–4

This process accounts for two of the key features of class switching. First, as mentioned above, the \( V_\mu \)-region remains the same. Since switching does not change the light chains, this means that the specificity of the antibody (and its idiotype) cannot change. Second, class switching is unidirectional and irreversible. A cell can switch from one C-region to another situated to the right of the first (5'-->3' direction), but cannot go backwards (3'-->5' direction). The order of switching is therefore defined by the order of the heavy chain constant region genes on the chromosome. [NOTE: \( C_H \) genes are each made up of a number of exons and introns which are not shown in this schematic.]

SOURCES OF ANTIBODY DIVERSITY: SUMMARY

Let’s return to our original question of the Generation of Diversity and examine the genetic basis for the existence of a million or more antibody specificities. We’ll use what we now know of human kappa chain gene structure and expression as our primary example.

1) **About 50 \( V_\kappa \)-genes exist in the human genome.** This is a fairly large number, much higher than somatic theories predicted, although smaller than predicted by most germ-line theorists.

2) **Five different \( J_\kappa \)-segments exist in the human genome.** Thus, if every V-region can be used with any J-segment, a total of about 250 (50 x 5) different combinations can be made. (In the case of heavy chains, the existence of several D-segments increases this number even more.) This process is referred to as COMBINATORIAL JOINING.

3) **DNA rearrangement is imprecise.** The joining of V and J (and the joining of D segments to V and J) is a deliberately error-prone process, resulting in random nucleotide substitutions and deletion/insertions at the site of rearrangement. Therefore, the joining of a particular V-region and a particular J-segment can yield different results in different cells in which it occurs. If this increases the number of potential V-regions by about ten-fold, we now have about 2500 possible kappa chain V-regions (250 x 10).

4) **Somatic mutation occurs in rearranged V-regions.** Expressed kappa chain genes have been cloned and sequenced from many antibody secreting cells, and in most cases
they have been found to differ from any existing germ-line V-region in sequence positions other than the site of joining of V and J. These somatic mutations tend to be localized to the hypervariable regions, and occur by a local relaxation of the normal processes of error-correction in newly synthesized DNA. The presence of these mutations may increase the number of possible V-regions by another ten-fold or more, resulting in at least 25,000 different kappa V-region sequences (2500 x 10).

5) **Heavy and light chains associate in random combinations.** If a comparable number of different H-chain V-regions can be produced (which is an underestimate), then there are potentially some $6 \times 10^9$ different antibody combining sites ($25,000 \times 25,000$), a number considerably larger than the one million we initially set out to explain. This random association of H- and L-chains is referred to as COMBINATORIAL ASSOCIATION.

It has therefore become clear that the historical debate between proponents of GERM-LINE and SOMATIC theories of antibody diversity is no longer meaningful, and that elements of both theories are important in the generation of antibody diversity. Points 1, 2, and 5 above fall within the framework of the germ-line theory, while points 3 and 4 quite clearly represent somatic processes.

**T-CELL RECEPTORS: Ig-LIKE GENE ORGANIZATION AND GENE REARRANGEMENTS**

The organization of immunoglobulin V- and C-region segments, and the processes by which they are rearranged, are clearly well-adapted for the generation of the huge diversity which so important for the humoral immune system. Such gene rearrangements are, in fact, unique to the immunoglobulin genes, with the single exception of the T-cell receptor (TCR), whose roles in immune responses we will discuss later (see Chapters 14, 15 and 18). These molecules, although quite distinct from immunoglobulins, are evolutionarily related to them, and share many important properties with them. They are a highly diverse family of heterodimeric membrane receptors capable of specifically interacting with antigens, their polypeptide chains include variable and constant regions, and the genes encoding them are organized and rearranged in a manner very similar to those of immunoglobulins.

There are, however, three critical differences in the structure and expression of TCR's compared with Ig: 1) Each TCR molecule bears only a single combining site (it does not exhibit the bivalency characteristic of all antibody molecules); 2) Somatic mutation, an important component in the generation of antibody diversity, does not occur during TCR expression. 3) Unlike immunoglobulins which can exist as either soluble or membrane-bound forms, TCR's function only as membrane-bound molecules. In later chapters we’ll discuss the unique biology and antigen-recognition functions of the TCR, and it will become clear why these differences between Ig and TCR are centrally important.
Despite the important differences between immunoglobulin and T-cell antigen receptors, they both express unique antigen-combining sites which, as described above, can be assembled only as a result of V(D)J recombination of germ-line DNA. In both cases this process requires the participation of two proteins known as RAG-1 and RAG-2 (for "Recombination Activating Genes"). These recombinase enzymes first appeared during the evolution of cartilagenous fish, and are present only in those organisms capable of carrying out adaptive immune responses, namely the "higher" vertebrates, including fish, reptiles, birds, amphibians and mammals. Genetic knock-out of the RAG proteins in mice results in the complete inability to produce either immunoglobulins or T-cell receptors, and consequently the absence of both mature B-cells and T-cells. Such knock-outs have been widely used to manipulate and study the vertebrate immune system.

CHAPTER 8, STUDY QUESTIONS:

1. Describe the overall architecture of the three mammalian IMMUNOGLOBULIN GENE FAMILIES.
2. Describe the molecular processes involved in expression of a kappa light chain gene. What processes are unique to immunoglobulins? In what ways does expression of a heavy chain differ from that of light chains?
3. What are the various processes which contribute to antibody diversity?
4. In what ways can DNA rearrangements be used to help define and diagnose lymphocyte tumors?
5. What would be the expected phenotypic consequence of a null mutation in the human RAG-1 or RAG-2 gene?