MOLECULAR MEDICINE

JUMPING GENES AND THE IMMUNOGLOBULIN V GENE SYSTEM

ROBERT S. SCHWARTZ, M.D.

MULTIPLE myeloma begins when one B cell of approximately 10 billion in the body undergoes mutations that enable it to outgrow most other cells in the bone marrow. When the neoplastic clone consists of 10^{11} B cells, serum electrophoresis barely detects the clone's signature molecule, a monoclonal immunoglobulin; with 10^{12} members, the gammopathy is unmistakable; and with 10^{13} myeloma cells, symptoms begin to appear. Not the least of the clinical problems in patients with myeloma is susceptibility to bacterial infection, a paradox given the huge amounts of gamma globulin secreted by the tumor. But most of the antibodies in the serum are worthless securities. They are a population of monoclonal immunoglobulins, and thus all of them have identical antigen-binding sites. The range of protection they provide, if any, is extremely narrow. To be protective, the immune system needs a way of coping with the diversity of the microbial environment. In other words, a proper antimicrobial defense requires billions of different kinds of antibody molecules and T cells.

Few problems in biology have matched the enigma of immunologic diversity. For decades, enormous quantities of paper and ink were consumed in addressing this question. Some immunologists believed there was an all-purpose antibody, which the antigen molded to a custom fit; others thought that each antibody corresponded to one gene, suggesting that there were tens of thousands of immunoglobulin genes. The solution came in 1976, when Hozumi and Tonegawa showed that myeloma cells constructed their monoclonal immunoglobulins from different gene segments, which the myeloma precursor cells had fused into a single coding sequence. This work was the foundation of our present understanding of the structure and function of antibodies and the antigen receptors of B cells and T cells.

Lymphocytes maximize their diversity by compiling the genetic code for their antigen receptors from an assortment of physically discrete segments of DNA. In part, the strategy resembles a game in which each player is dealt a hand from a freshly shuffled deck of 52 cards. If each player receives six cards, about 20 million different hands are possible. However, lymphocytes have the potential to generate more than 10^{14} different antigen receptors. How do they do this? In answering this question, I shall focus on immunoglobulin genes. Similar principles apply to genes for T-cell receptors but differ in important details.

Immunoglobulins embedded in the surface membrane of the B lymphocyte serve as the cell's antigen receptors. All the receptors in each B cell are the same, but each new clone of B cells expresses a different receptor. Thus, within the whole population of billions of B cells, there is an enormous variety of immunoglobulin receptors. The dynamic turnover of B cells adds another dimension to the diversity of antibodies. Millions of fresh recruits — each with a novel receptor — leave the marrow daily to replace the millions of B cells that die every day in the lymph nodes and spleen. The continuous seeding of lymphoid tissue with new repertoires of B cells gives the immune system a remarkable capacity to respond to a changing environment.

All this diversity rests on about 200 genes. These variable-region genes (V genes) are the physical substrate for antibody variation. They provide the DNA from which the immature B cell fashions its receptor's coding region by cutting, rearranging, and revising individual gene segments. This process culminates in an antibody whose Y shape delineates the two functionally distinct sections of the molecule (Fig. 1).

The diversity of the antibody is in the arms of the Y. These arms, the variable regions, consist of matched pairs of polypeptides, the light and heavy chains. Each pair forms a surface or groove that serves as a binding site for antigen; an antibody with two variable regions has two identical binding sites. Each light and heavy chain of the pair contains three subdivisions called "hypervariable regions," which account for most of the variation in immunoglobulins (Fig. 1). The six hypervariable regions, three on the light chain and three on the heavy chain, are the principal contributors to the antibody's specificity and individuality.

The stem of the Y, the constant region, specifies the antibody's effector properties, such as complement fixation or passage through the placenta. The variability of this region is limited to the five immunoglobulin isotypes: IgM, IgG, IgA, IgD, and IgE; each corresponds to a particular constant-region gene (C gene): Cμ, Cγ, Cα, Cδ, and Cϵ, respectively. Light chains do not contribute to the constant region, but the 3′ end of the heavy-chain gene does.

Organization of the coding region for the antigen receptor begins early in the maturation of lymphocytes. The cell links physically discrete genes, some separated by thousands of intervening bases, to form a continuous coding unit for the receptor. The κ and λ loci contain the variable-region genes for the two families of light chains, kappa and lambda (Fig. 1). These genes are of two types: variable (Vκ or Vλ) and joining (Jκ or Jλ). There are three kinds of heavy-chain variable-region genes: variable (VH), diversity (DH), and joining (JH). Strictly speaking, these DNA sequences are not genes but exons, because the B cell translates them into messenger RNA only when they have been joined into a coding unit. Nevertheless, the convention is to call the individual VH, DH, and JH units "genes," "gene segments," or simply "segments."

The B cell has available approximately 50 VH genes, about 30 DH genes, and 6 JH genes, but it uses only one of each kind to assemble the VH–DH–JH coding sequence of its heavy-chain variable region.

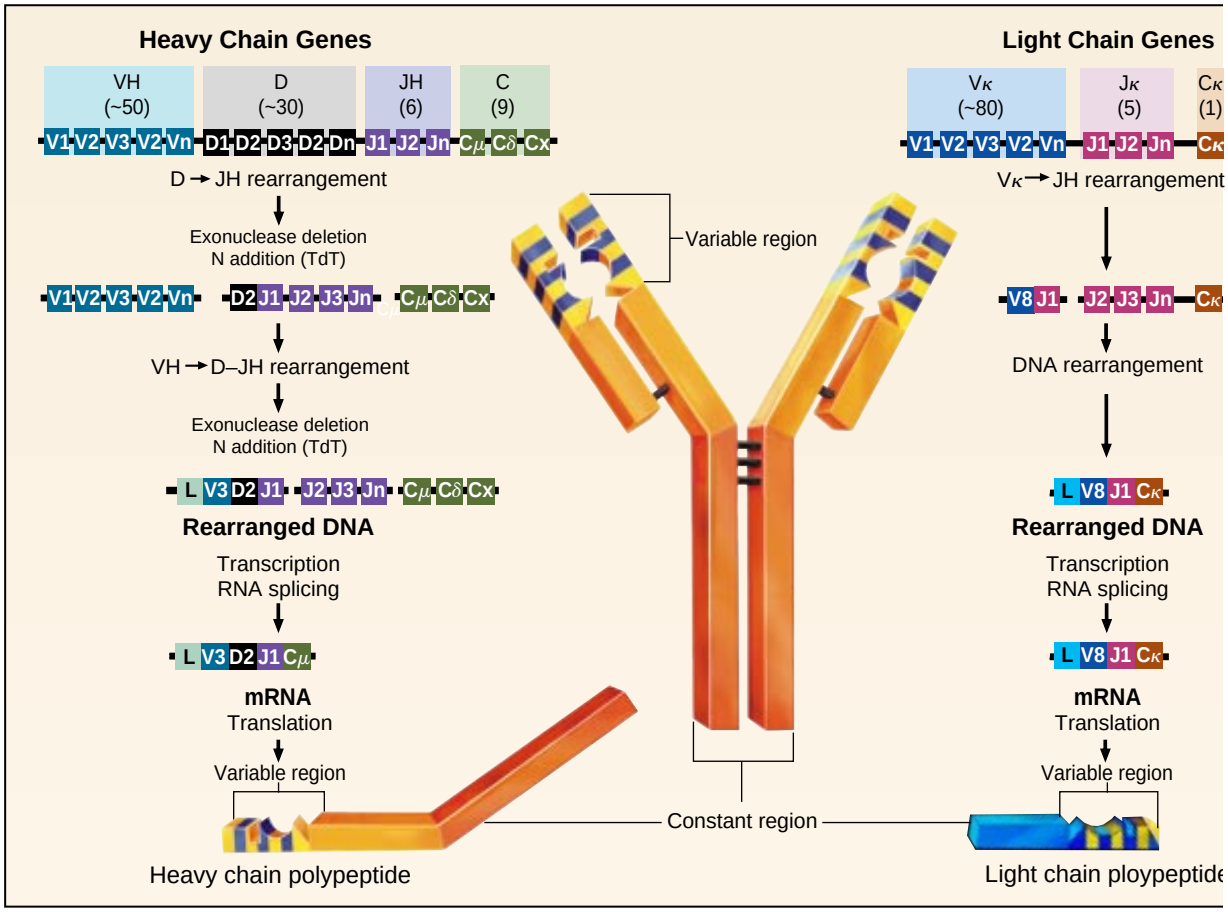

Figure 1. The Molecular Basis of Antibody Diversity.

The immunoglobulin molecule (center) consists of identical pairs of heavy and light chains. Each chain has two main divisions, the variable region and the constant region. Three subdivisions within the variable region, called the hypervariable regions, account for most of the variation among antibodies (blue bars). The six hypervariable regions, three on the heavy chain and three on the light chain, form a fold, pocket, or surface that binds the antigen. Between these hypervariable regions, the four framework regions (yellow bars) support the three-dimensional structure of the variable region's immunoglobulin fold. The coding region for the heavy chain (left) arises by random rearrangement of four kinds of gene segments: VH, DH, JH, and CH. Precise enzymatic cuts in the B cell's double-stranded DNA, followed by ligation of the cut ends of the DNA, are key steps in establishing the immunoglobulin gene-coding unit. DNA between the cut segments forms loops, which nucleases excise and degrade. In the example shown here, an early step is the juxtaposition of 1 of 30 D genes (D2) next to 1 of 6 JH genes (J1). Next, a VH segment (V3 in the example), which may be located thousands of kilobases away from the D2 gene, links up with the D2–J1 unit. Before the D2–J1 and V3–D2–J1 genes fuse, exonucleases and terminal deoxynucleotidyl transferase (TdT) revise the cut ends of the segments by randomly removing or adding nucleotides. These somatic modifications have a profound influence on the diversity of the hypervariable region nearest the constant region of the heavy chain. This so-called third hypervariable region corresponds to the coding region from the 3′ end of the VH segment to the 5′ end of the JH gene, between the junctions of VH to DH and DH to JH. Formation of the light-chain coding region (right) is similar in principle, except that D segments do not participate in the process. Exonucleases and TdT may modify the cut ends of light-chain genes, but to a much smaller extent than in the case of heavy-chain genes. TdT is also used infrequently in forming the heavy chain in the fetus. The number of VH, DH, and Vκ genes is approximate; there are many nonfunctional variable-region genes (pseudogenes), which may represent evolutionary junk. L denotes leader region, and mRNA messenger RNA. The black bars represent disulfide bonds that hold the chains of the antibody molecule together.

We do not know how a particular gene segment enters into the process of VH–DH–JH gene rearrangement. The focus has been on two recombination-activating genes: *RAG-1* and *RAG-2,* which are essential for the VH–DH–JH gene recombination. Studies in mice have shown that when either gene is disabled by the gene-targeting method, the V-gene segments cannot be recombined.

To recombine scattered VH genes and organize them into a unified and translatable coding region, DNA must be cut, the intervening region eliminated, and the severed ends of the two strands of DNA repaired. Enzymes mediate all these steps. Signals flanking the DNA on each DH and JH gene segment and on one side of the V gene guide the cutting enzyme. These recombination-signal sequences ensure the correct linkage of VH, DH, and JH (or VL and JL) segments and prevent nonsense rearrangements between like segments, such

as two VH or two JH segments. The cut segments — DH11 and JH2, for example — fuse, forming a continuous DH11–JH2 sequence from which the intervening DNA has been eliminated (Fig. 1).

The potential to generate unique combinations by a process like the card game described above or VH-gene recombination is large but not unlimited. Other mechanisms, which can increase the diversity of antigen receptors by orders of magnitude, come into play after the recombination is complete. Special enzymes introduce random somatic alterations in the DNA code where the V segments join. In the example of the DH11–JH2 sequence, exonucleases "nibble" several bases from the cut ends of the DH11 and JH2 segments. Another enzyme, terminal deoxynucleotidyl transferase (TdT), randomly adds up to 20 nucleotides, called N additions, to the joint between the two cut ends. These random losses and gains of nucleotides in the recombined VH-gene segments make the newly assembled coding regions for the heavy- and light-chain polypeptides virtually unique. Indeed, the variable region of the antigen receptor provides a molecular signature distinctive enough to serve as the target in searching for residual malignant cells after the administration of chemotherapy in patients with lymphoma or leukemia.

The immature B cell assembles its V genes in a specific order. First, it joins heavy-chain segments in the following sequence: DH→JH; VH→DH–JH. After RNA-processing enzymes excise introns and any other intervening segments, the messenger RNA corresponding to the linked VH–DH–JH–Cμ segments is the transcription unit for heavy chains (Fig. 1). The appearance of cytoplasmic μ chains, a landmark in B-cell differentiation, is clinically useful in classifying B-cell neoplasms. The presence of the heavy chain on the B-cell membrane has considerable functional importance: it inhibits further rearrangement of VH-gene segments and signals the rearrangement of Vκ segments. If errors cause that process to fail, the B cell tries to make a Vλ coding unit. These successive efforts — Vκ first, then Vλ — account for the 60:40 ratio of Vκ to Vλ light chains in populations of immunoglobulins. Like any other nucleated cell, the B cell has two sets of V genes, one from each of the person's parents. If the B cell were to express more than one kind of antigen receptor by rearranging V genes from both parental loci, it would lose specificity and efficiency — contact with antigen "A" would trigger not only anti-A antibodies but also anti-B antibodies. A process called allelic exclusion restricts the rearrangement to one of the two parental loci. This mechanism explains why all the cells in a particular B-cell clone, B-cell lymphoma, or monoclonal gammopathy have only one kind of light chain and one kind of heavy chain.

Unlike genes for T-cell receptors, the rearranged V genes in a mature B cell can mutate. Somatic mutation of V genes can increase the ability of the surface immunoglobulin receptors to bind the antigen, thereby giving the B-cell clone a growth advantage over other B cells that cannot respond to the antigen. Such mutated clones arise in germinal centers of lymph nodes and spleen under the influence of signals from T cells. The mutations typically occur in the hypervariable segments of heavy- or light-chain genes (often both). The functional influence of these mutations is considerable, because the antigen-binding site depends on just 5 to 10 amino acids. A mutation that changes only one amino acid can increase or decrease the antibody's affinity 10-fold or even change the antibody's specificity. The antibody response thus proceeds by a Darwinian mechanism in which B-cell clones with receptors having the highest affinity for the antigen ultimately dominate the population. Indeed, studies of B cells that were plucked one by one from germinal centers of human spleens have shown that a few related clones with mutated VH and VL genes predominate in individual germinal centers.

Clinical applications have been found for virtually all the molecular elements of lymphocyte diversification. A previous article in this series explained the molecular diagnosis of malignant clones of B cells or T cells; we have seen how investigators can exploit a clone's specific molecular signature to seek residual leukemia or lymphoma; the use of TdT to classify leukemias has become a standard procedure; and V-gene mutations can be used to trace the origins of lymphomas and leukemias, as well as autoantibodies. Monoclonal antibodies constructed from mouse V genes and human C genes by molecular engineering are already in clinical use. In the not-too-distant future, "phage farms," exploiting the principle of V-gene recombination, will produce antibodies made to order entirely in vitro for any clinically useful immunochemical specificity. The molecular technology of immunoglobulin genes is now so advanced that the only impediment to new clinical applications is the investigator's imagination.

Recommended Reading

Cook GP, Tomlinson IM. The human immunoglobulin VH repertoire. Immunol Today 1995;16:237-42.

Hozumi N, Tonegawa S. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. Proc Natl Acad Sci U S A 1976;73:3628-32.

Schatz DG, Oettinger MA, Schlissel MS. V(D)J recombination: molecular biology and regulation. Annu Rev Immunol 1992;10:359-83.

Schwartz RS. Jumping genes. N Engl J Med 1995;332:941-4.

Stewart AK, Schwartz RS. Immunoglobulin V regions and the B cell. Blood 1994;83:1717-30.