75

an Article from SCIENTIFIC AMERICAN

OCTOBER, 1985 VOL. 253, NO. 4

The Molecules of the Immune System

The proteins that recognize foreign invaders are the most diverse proteins known. They are encoded by hundreds of scattered gene fragments, which can be combined in millions or billions of ways

by Susumu Tonegawa

The immune system is clearly essential to survival; without it death from infection is all but inevitable. Even apart from its vital function the immune system is a fascinating example of biological ingenuity. The cells and molecules of this defensive network maintain constant surveillance for infecting organisms. They recognize an almost limitless variety of foreign cells and substances, distinguishing them from those native to the body itself. When a pathogen enters the body, they detect it and mobilize to eliminate it. They "remember" each infection, so that a second exposure to the same organism is dealt with more efficiently. Furthermore, they do all this on a quite small defense budget, demanding only a moderate share of the genome and of the body's resources.

The critical event in mounting an immune response is the recognition of chemical markers that distinguish self from nonself. The molecules entrusted with this task are proteins whose most intriguing property is their variability of structure. In general all the molecules of a given protein made by an individual are absolutely identical: they have the same sequence of amino acids. At most there may be two versions of a protein, specified by maternal and paternal genes. The recognition proteins of the immune system, in contrast, come in millions or perhaps billions of slightly different forms. The differences enable each molecule to recognize a specific target pattern.

The most familiar of the recognition proteins are the antibodies, or immunoglobulins. Much has been learned of their structure and, equally important, of the genetic mechanisms responsible for their diversity. It turns out that vast numbers of antibodies are made by reshuffling a much smaller set of gene fragments. Thus antibody genes offer dramatic evidence that DNA is not an inert archive but can be altered during the life span of an individual. In the synthesis of antibodies the cutting and joining of gene sequences is not a mere incidental feature of the genetic process; it is essential to the functioning of the immune system.

Another class of recognition molecules consists of the proteins called *T*-cell receptors. Because they are more difficult to isolate, their properties are not yet as well known as those of the antibodies. In structure and evolutionary origins they are clearly related to the immunoglobulins, and a similar genetic mechanism accounts for their diversity, but their mode of operation is subtly different. A *T*-cell receptor recognizes only those cells that bear both self and nonself markers. By this curious means *T* cells are given the ability both to act directly against viral infections and to regulate other components of the immune system.

The primary cells of the immune system are the small white blood cells called lymphocytes. Like other blood cells, they are derived from stem cells in the bone marrow. In mammals one class of lymphocytes, the *B* cells, complete their maturation in the bone marrow. A second class, the *T* cells, undergo further differentiation in the thymus gland. Cells of the two classes are similar in size and appearance, but they take part in different forms of immune response.

B lymphocytes are the cells that manufacture antibodies. Their basic mode of action can be understood in terms of the clonal selection theory proposed 30 years ago by Sir Macfarlane Burnet. As each *B* cell matures in the bone marrow, it becomes committed to the synthesis of antibodies that recognize a specific antigen, or molecular pattern. In the simplest case all the descendants of each such cell retain the same specificity, and so they form a clone of immunologically identical cells. (Actually some variation is introduced as the cells proliferate.)

The antibodies made by a *B* cell remain bound to the cell membrane, where they are displayed on the surface as receptor molecules. When an antigen binds to an antibody in the membrane, the cell is stimulated to proliferate; this is the clonal-selection process. In general many clones respond to a single infection. The antigenic markers recognized by antibodies are comparatively small patterns of molecular structure, and a single virus

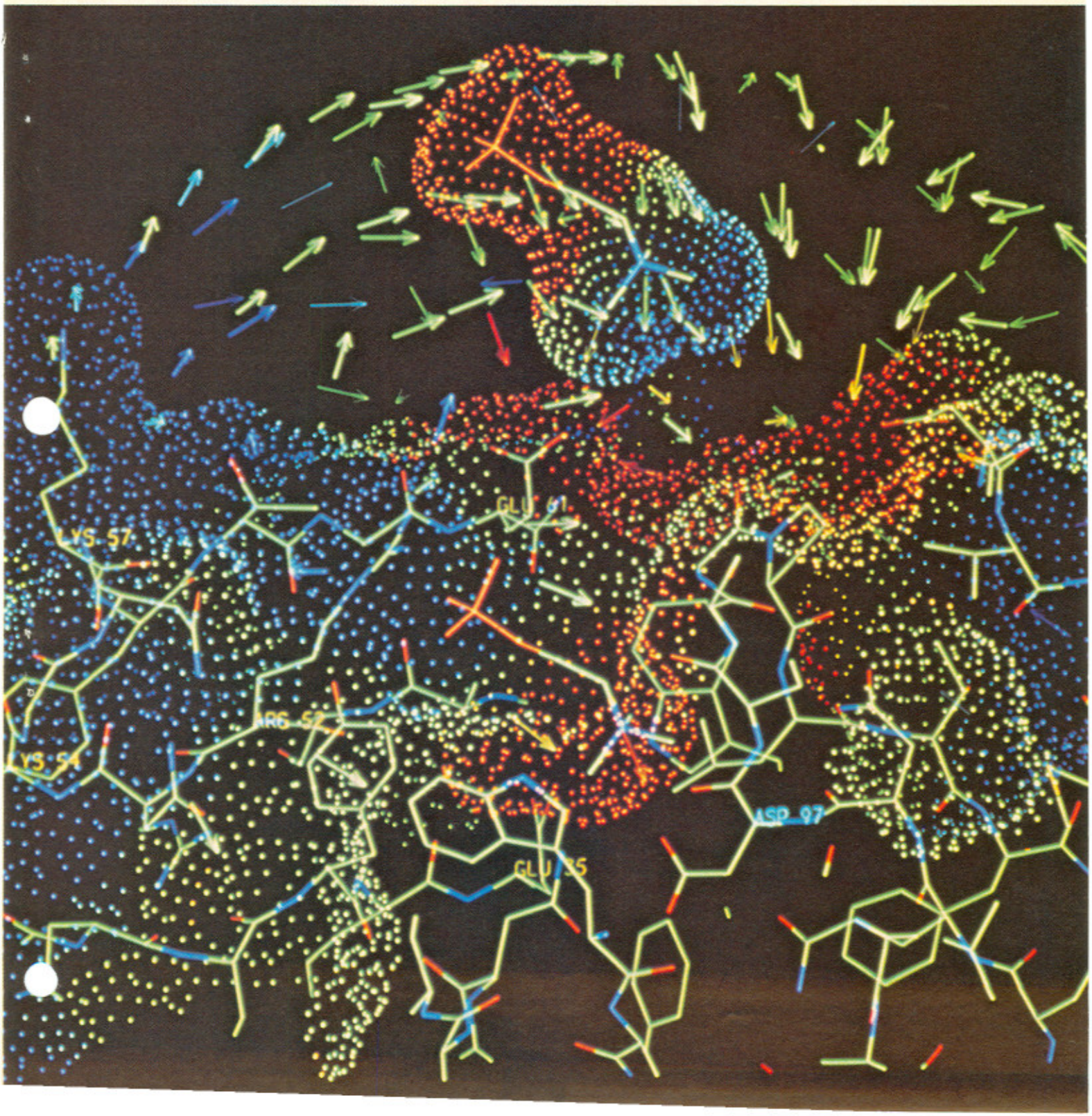
BINDING OF AN ANTIGEN to an antibody is a central event in the recognition of foreign organisms in the body. In the computer-generated image on the opposite page the bound substance is not an actual antigen but a hapten, a small molecule with an affinity for a particular antibody. The hapten shown is phosphocholine. It is guided to the antigen-binding site by electrostatic interactions and fits into a cleft on the antibody surface. Its orientation on approaching the binding site, as depicted in the top center of the image, is suggested by a calculation done by Elizabeth D. Getzoff, John A. Tainer and Arthur J. Olson of the Research Institute of Scripps Clinic; the calculation is based on the atomic structure of the antibody-hapten complex, which was determined by Eduardo A. Padlan, Gerson H. Cohen and David R. Davies of the National Institutes of Health. The skeleton of the protein and that of the approaching hapten molecule are shown enveloped in dots that represent the surfaces accessible to water molecules; another hapten, installed in the antigen-binding site directly below the first hapten, is shown as a skeleton only. The colors of the dots indicate the calculated electrostatic potential of various regions of the molecular surface; blue is the most positive and red the most negative. Arrows show the direction of the electrostatic field, and their colors indicate the electrostatic potential at their points of origin. The image was made with the programs GRAMPS, developed by Olson and T. J. O'Donnell of Abbott Laboratories, and GRANNY, written by Olson and Michael L. Connolly of Scripps Clinic.

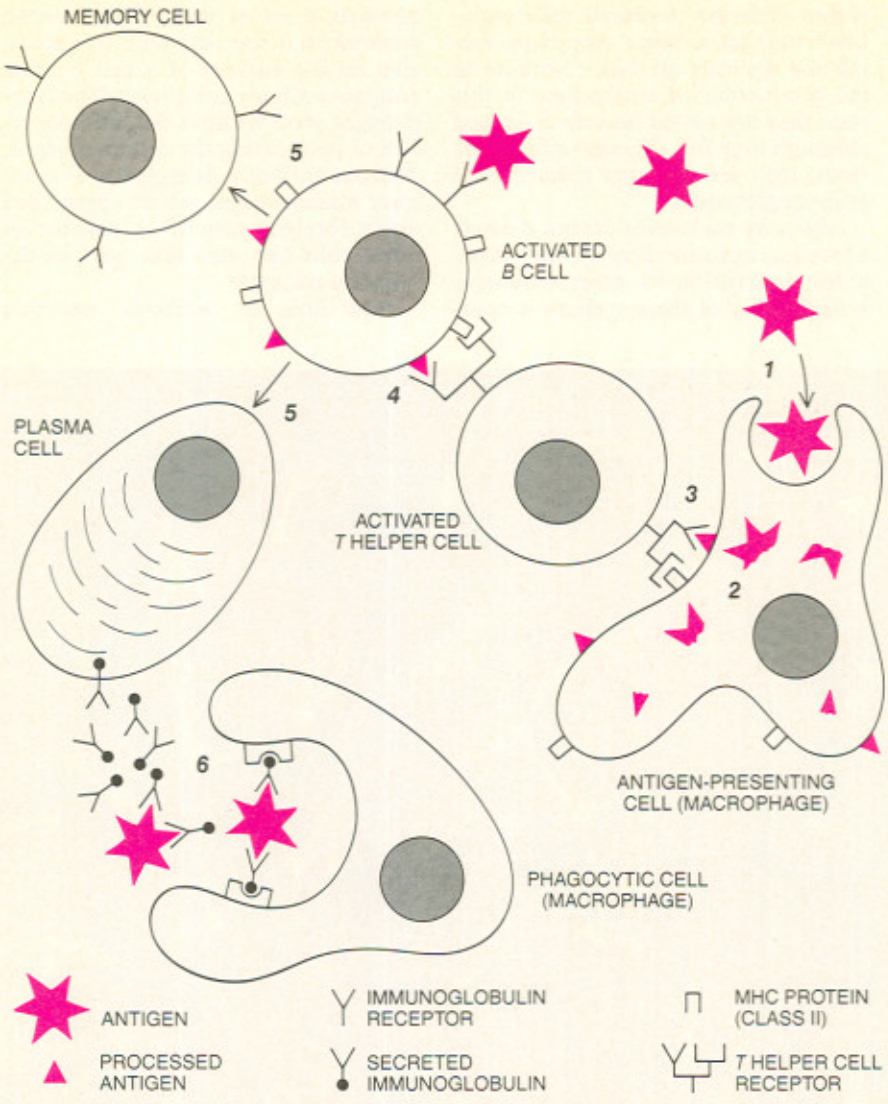
or bacterium carries many recognizable patterns.

Some of the progeny of the selected clones remain as circulating *B* lymphocytes. They serve as the immune system's memory, providing a faster response to any subsequent exposure to the same antigens. The memory cells are responsible for the immunity that develops following many infections or as a result of vaccination. Other members of the selected *B*-cell clones undergo "terminal differentiation": they grow larger, stop reproducing and dedicate all their resources to the production of antibodies. In this state they are called plasma cells, and although they live for only a few days more, they secrete large quantities of immunoglobulins.

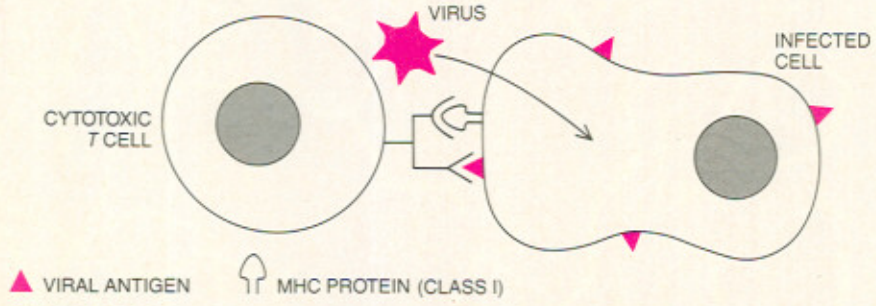
Antibody molecules cannot destroy a foreign organism directly; they mark it for destruction by other defensive systems. One of these systems is complement, a set of more than a dozen proteins that are activated in succession on the surface of a cell bearing antigen-antibody complexes. The complement proteins have the ultimate effect of perforating the cell membrane. Antigen-antibody complexes also attract macrophages, which engulf and digest foreign particles. A number of other cells can also take part in the immune response.

How does an antibody molecule





IMMUNE RESPONSE TO INFECTION mobilizes several cooperating populations of cells. *B* cells carry immunoglobulins as surface receptors that recognize and bind to circulating antigens; in general, however, the *B* cells are not activated by this process alone. First the antigen must be taken up by an antigen-presenting cell (*1*); a macrophage can serve in this role. The antigen is processed by the macrophage (*2*) and then displayed on its surface. There it is recognized by a *T* helper cell, which is thereby activated (*3*). The *T* helper cell then activates *B* cells carrying the same processed antigen (*4*). The activated *B* cells proliferate and undergo terminal differentiation (*5*). Some of the progeny become memory cells, which provide a quicker response to future infections, whereas others develop into antibody-secreting plasma cells. The secreted antibodies bind to the antigen, thereby marking it for destruction by various other components of the immune system, including macrophages (*6*).



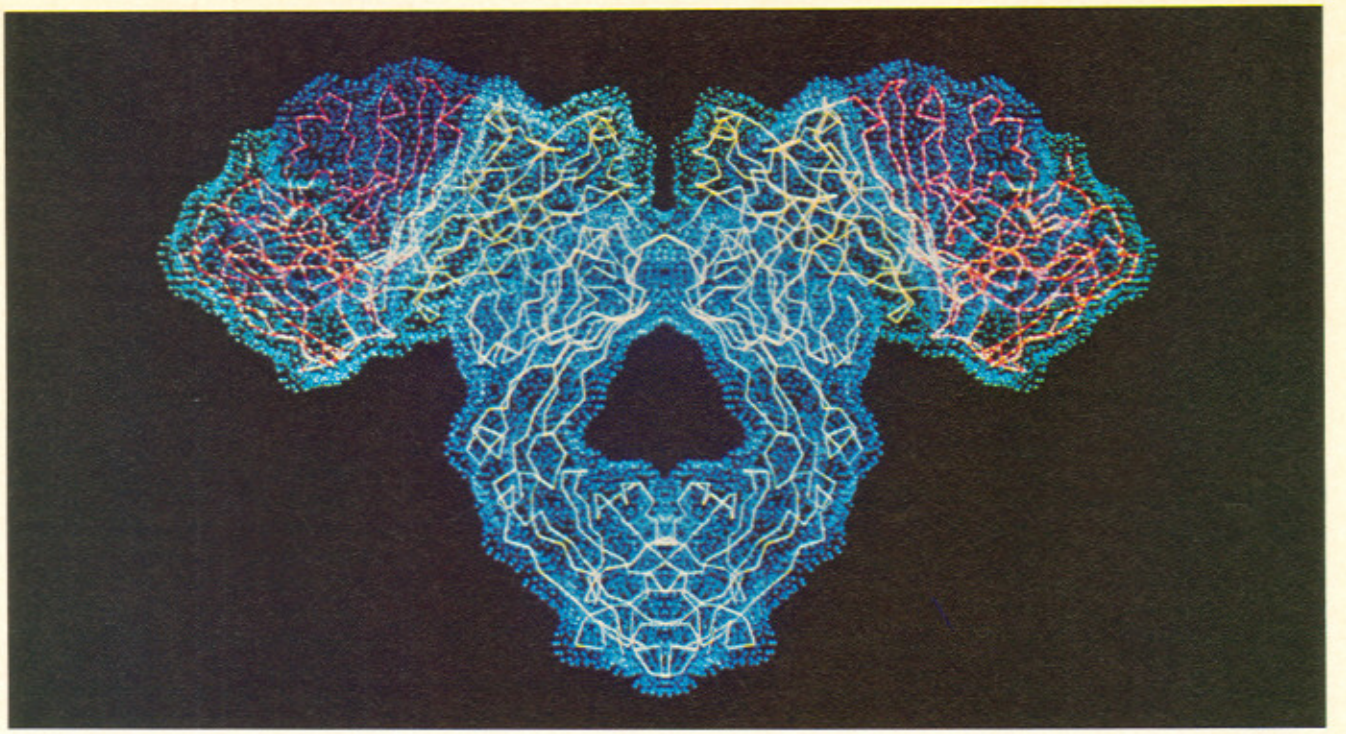
RESPONSE TO VIRAL INFECTION calls on other elements of the immune system. When a virus enters a cell, viral proteins are left behind embedded in the cell membrane. Cytotoxic *T* cells specifically recognize such foreign molecules displayed in combination with proteins that identify the host. They are the Class I proteins of the major histocompatibility complex (MHC). The cell infected by the virus is killed by the cytotoxic *T* cell.

recognize an antigen? The answer was found by analyzing the amino acid sequence and the three-dimensional structure of antibodies.

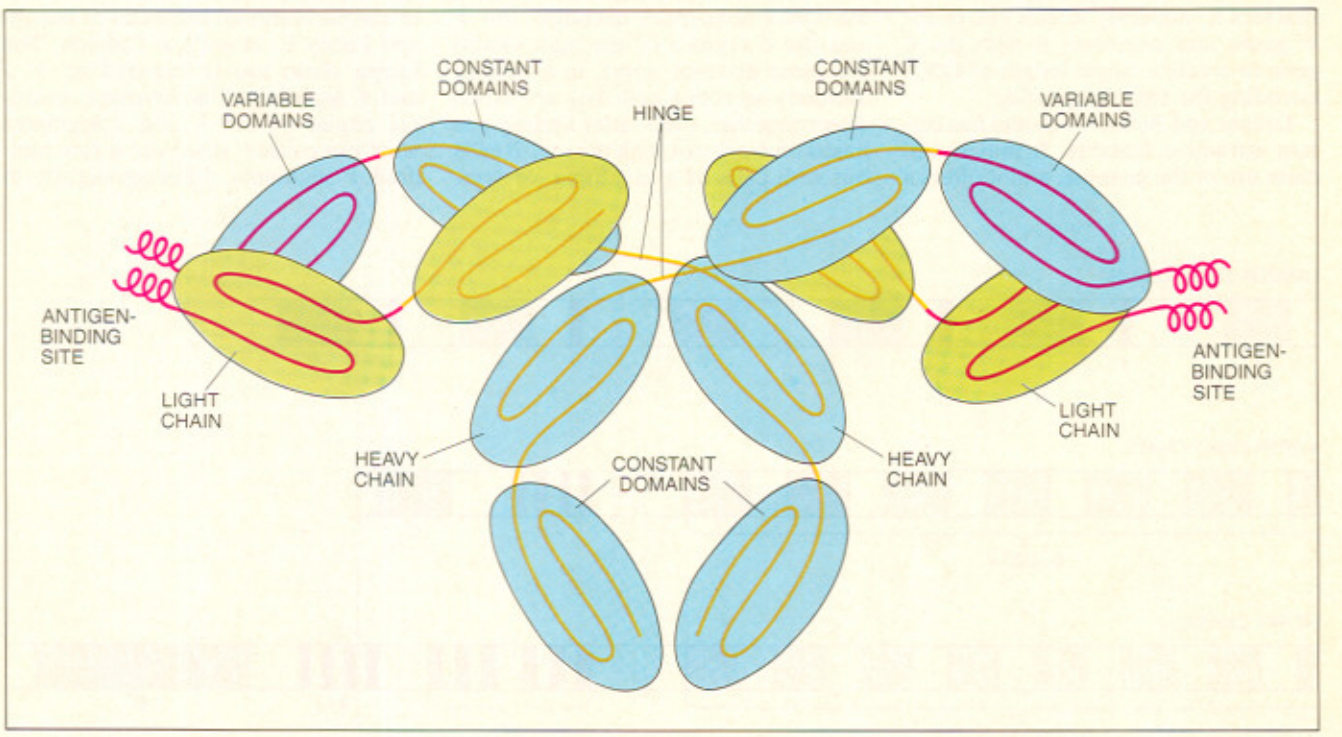
A basic antibody molecule consists of four polypeptide strands: there are two identical light chains of about 220 amino acids and two identical heavy chains of either 330 or 440 amino acids. The four chains are held together by disulfide bonds and noncovalent bonds to form a Y-shaped molecule. Both the heavy and the light chains are built up from a common domain, or structural subunit, of about 110 amino acids. It would appear that the gene for some prototypical protein of about this size has been repeatedly duplicated and altered to give rise to the genes for both immunoglobulin chains. A light chain has two somewhat different copies of the domain and a heavy chain has either three or four copies. All the copies fold up into broadly similar three-dimensional structures.

In both heavy and light chains the domain at the amino end of the polypeptide—the end synthesized first—differs in an important way from the other domains. The amino-terminal domain is where most of the variation in amino acid sequence is found. In the folded antibody the variable regions make up the terminal half of the arms of the Y; each arm incorporates the variable region of one heavy chain and one light chain. Within the variable region of each chain there are three small segments where the amino acid sequences are found to be particularly diverse. These "hypervariable" segments come together at the end of each arm to form a cleft that acts as the antigen-binding site. The specificity of the molecule depends on the shape of the cleft and on the properties of the chemical groups that line its walls; thus the nature of the antigen recognized by an antibody is determined primarily by the sequence of amino acids in the hypervariable regions.

One further aspect of the structure of antibodies must be mentioned. Even in the constant regions not all molecules are identical. In mammals the light-chain constant regions are of two types, designated kappa and lambda. There are five classes of heavy-chain constant regions: mu, delta, gamma, epsilon and alpha. Antibodies with the same variable regions but different heavy-chain classes recognize the same antigens but have different roles in the immune response. For example, the membrane-bound antibodies that serve as *B*-cell receptors incorporate mu or delta chains, and most of the antibodies secreted in response to an an-



ANTIBODY MOLECULE is a Y-shaped protein made up of four polypeptide chains. Two heavy chains (*blue surfaces*) extend from the stem of the Y into the arms; two light chains (*green surfaces*) are confined to the arms. Each polypeptide has both constant regions (*white and yellow skeleton*) and variable regions (*red skeleton*). All antibodies of a given type have the same constant regions, but the variable regions differ from one clone of *B* cells to another. At the end of each arm the heavy- and light-chain variable regions fold to create an antigen-binding site. The image was made by Olson with the computer programs used in making the one on page 123.



STRUCTURE OF AN ANTIBODY can be analyzed in terms of repeated domains, or independent folding units. A light chain consists of two such domains; the heavy chains shown here have four domains, although some heavy chains have three. Within a domain the polypeptide chain assumes a characteristic pattern of folding that includes several strands of the substructure called a beta sheet. The variable region of each polypeptide is confined to a single domain at the amino end of the chain. Three loops (called hypervariable regions) within the variable domain contribute to the antigen-binding site. The domain structure shown is schematic; the actual folding pattern is more complex. Similar domains appear in the *T*-cell-receptor protein and in the proteins of the major histocompatibility complex, which identify an individual's cells. All three families of molecules have probably evolved from a common ancestor.

tigen include gamma or alpha chains.

For many years there were two competing theories of the genetic origin of antibody diversity. One school of thought held that the germ line (the complement of genes passed from one generation to the next) must include a separate gene for every polypeptide that ultimately appears in an antibody. In this germ-line theory immunoglobulin genes are expressed in exactly the same way as those for any other proteins, and no special gene-processing mechanisms are needed. On the other hand, the model requires an enormous number of immunoglobulin genes.

The second theory supposes there are only a limited number of antibody genes in the germ line, and they somehow diversify as *B* lymphocytes emerge from their stem cells. In other words, the diversification takes place in the somatic, or body, cells rather than in the germ cells.

An interesting variation of the germ-line theory was introduced in 1965 by William J. Dreyer and J. Claude Bennett of the California Institute of Technology. They suggested that for each type of antibody polypeptide chain the germ line includes many *V* genes (one to encode every possible variable region) and a single *C* gene for the constant region. As the cell matures it randomly selects one of the *V* genes and combines it with the *C* gene to create a single length of DNA encoding the full polypeptide.

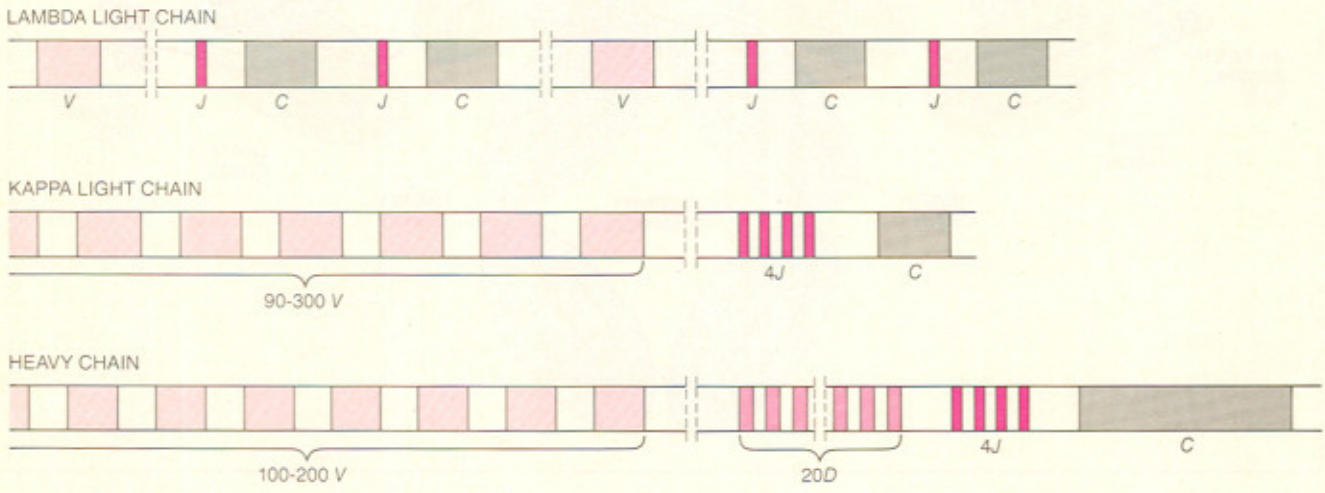
Dreyer and Bennett's model has certain attractive features. It makes efficient use of the genome, and it offers a natural explanation of how antibody molecules can vary greatly in one part of their structure and remain constant in other parts. Until the mid-1970's, however, there was a major impediment to acceptance of the theory: it required some means of rearranging genes in somatic cells. No such mechanisms were known, and many workers considered them unlikely to exist. The doctrines that one gene always encodes one polypeptide and that the genome remains constant throughout an organism's development were then considered universally established principles of biology.

In the past 10 years the application of recombinant-DNA technology to the study of immunoglobulin genes has shown that they do undergo somatic recombination, but in ways much more complicated than Dreyer and Bennett supposed. Through these complex rearrangements great diversity is generated in the *V* regions.

The first evidence of somatic recombination in immunoglobulin genes was found by Nobumichi Hozumi and me in 1976, when we were both working at the Institute for Immunology in Basel, Switzerland. Our experiments made use of restriction enzymes, which cut DNA at points marked by a particular sequence of nucleotides. The results showed that in embryonic mouse cells the DNA sequences encoding the *V* and the *C* regions of the light chains are some distance apart. In a mature antibody-secreting cell they are much closer together. (The latter finding was based on work not with normal *B* cells but with cells of a myeloma, or lymphocyte cancer. Such malignant cells are much easier to grow in culture.)

The mechanisms responsible for the shuffling of immunoglobulin DNA sequences became clearer when fragments of the DNA were cloned in bacteria and analyzed. This was first done by Ora Bernard and me in Basel, in collaboration with Allan Maxam and Walter Gilbert of Harvard University. Working with a DNA clone derived from embryonic mouse cells, we determined the nucleotide sequence of a segment of DNA encompassing the *V* gene of a particular lambda light chain. To our surprise we found that the nucleotides corresponding to the last 13 amino acids of the variable region were missing. They were discovered by Christine Brack of my laboratory. The missing fragment is thousands of base pairs away from the DNA encoding the rest of the *V* region and lies about 1,300 base pairs "upstream" from the start of the *C* region. This small interposed segment has been named *J*, for joining. Each lambda light chain is assembled by combining the scattered *V*, *J* and *C* segments.

A similar analysis was soon carried out for the kappa light chain and for the heavy-chain variable region. The work was done in several laboratories, notably my own in Basel, Philip Leder's at the National Institutes of Health and Leroy E. Hood's at Caltech. The kappa chain too is encoded by *V*, *J* and *C* segments. Furthermore, multiple copies of the *V* and *J* segments were discovered: there are a few hundred *V* segments, differing slightly in



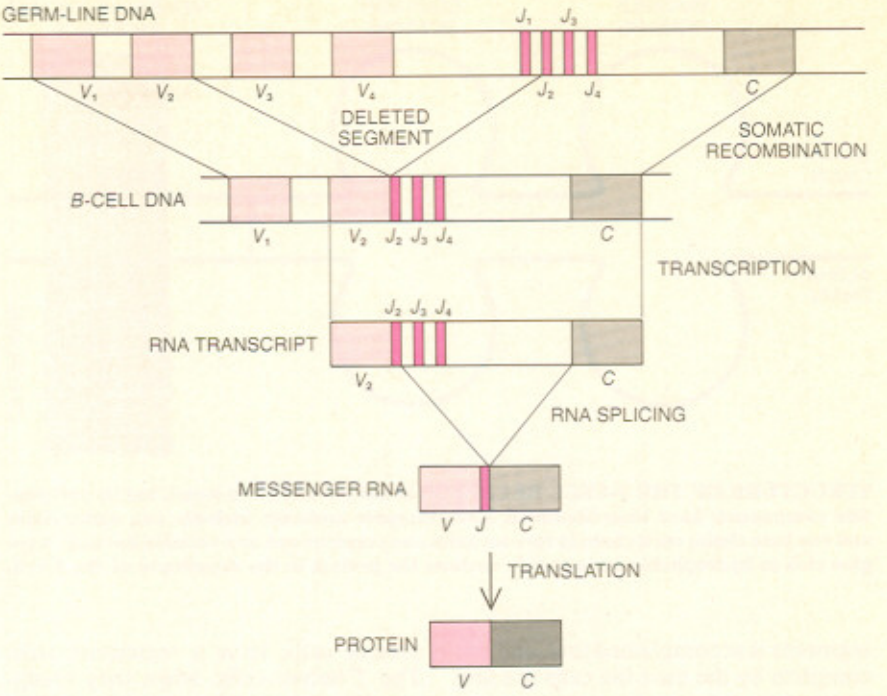
GENES FOR ANTIBODIES are broken up into small segments scattered widely throughout the genome. Two kinds of light chain appear in mammalian antibody molecules. For the lambda light chain of the mouse there are two *V* genes that encode most of the variable region and four *C* genes for the constant region. Upstream of each *C* gene is a short segment of DNA designated *J*, for joining, which specifies the remainder of the variable region. Either *V* gene can be combined with any pair of *J* and *C* genes. For the kappa light chain there are a few hundred *V* segments, four *J* segments and a single *C* gene. The heavy-chain genes are similar, except that the DNA for the variable region is further subdivided: in addition to the *V* and *J* segments there are about 20 *D* (for diversity) segments. Each set of genes is on a different chromosome. The *T*-cell-receptor genes are organized much as the heavy-chain genes are.

amino acid sequence, and four distinct *J* segments. The number of possible kappa-chain variable regions is the product of these numbers.

The potential diversity of the heavy chains is even greater. In addition to *V* and *J* segments the genes for the heavy-chain variable region include a third fragment designated *D* (for diversity). Mouse germ-line cells have a few hundred *V* segments, about 20 *D* segments and four *J* segments. In principle they can be brought together in well over 10,000 combinations. Combining a light chain with a heavy chain can probably yield more than 10 million distinct antigen-binding sites.

The assembly of a functioning immunoglobulin gene takes place in two stages. First the *V* and *J* segments in a light chain or the *V*, *D* and *J* segments in a heavy chain are brought together within the DNA. An RNA transcript is then made of the entire length of DNA, including the *V*-*J* or *V*-*D*-*J* complex, the *C* gene and the intron, or noncoding intervening sequence, that separates them. Finally the intron is excised and the messenger RNA is exported from the nucleus to be translated into protein.

The second stage in this process relies on mechanisms of RNA splicing that are common to many families of genes in eukaryotic cells. The first stage, in which the DNA itself rather than the RNA transcript is altered in a highly specific manner, is more unusual and may even be unique to the immune system. It evidently employs a set of enzymes that can bring together distant *V*, *D* and *J* segments, often deleting all the DNA that separates them. The enzymes themselves have not been isolated, but signal sequences that probably guide their action have been discovered. For example, just downstream of each *V* gene for the kappa chain there is a distinctive pattern composed of a heptamer, or seven-nucleotide unit, followed by a spacer and a nonamer, or nine-nucleotide unit. Just upstream of the *J* segment there is an approximately complementary nonamer-spacer-heptamer pattern. These units could provide a template for the enzymes that cut and rejoin the double helix. Similar signal sequences are found in heavy-chain genes, arranged so that a *D* segment will be included between the *V* and the *J* segments.



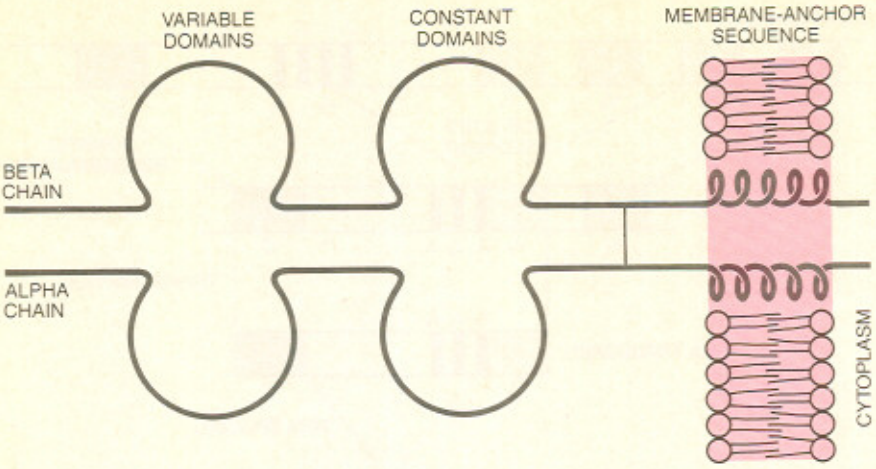
ASSEMBLY OF AN ANTIBODY GENE from scattered fragments is done in two stages, shown for a kappa light chain. First randomly selected *V* and *J* segments are fused by enzymes that delete all the DNA lying between them. Here the gene segments labeled V_3, V_4 and J_1 are deleted, bringing together V_2 and J_2. Next the entire length of DNA from the start of V_2 to the end of the *C* gene is transcribed into RNA. Standard RNA-splicing enzymes, which take part in the expression of many genes, excise all the RNA from the end of J_2 to the start of *C*. The resulting sequence of messenger RNA is translated into protein.

The many possible combinations that can be formed from several hundred gene segments are the key to antibody diversity, but there are at least two additional sources of variety. One of these is a lack of precision in the DNA-splicing machinery that fuses *V*, *D* and *J* segments. The site of the junction can vary by several base pairs. Furthermore, in some cases additional base pairs are inserted in the process of combining segments. Both kinds of change can obviously alter the amino acid sequence of the polypeptide. As a result, even if two antibodies are specified by the same collection of gene segments, they may still have slightly different antigen-binding sites.

Another major source of diversity is somatic mutation. In 1970 Martin Weigert, working in Melvin Cohn's laboratory at the Salk Institute for Biological Studies, determined the amino acid sequences of the lambda light chains derived from 18 mouse myelomas. All the mice were of the same inbred strain and so should have been genetically identical. Weigert found that 12 of the lambda chains were indeed identical but that the other six differed both from the majority sequence and from one another. Spontaneous genetic changes in the developing cells were a likely explanation, but cogent evidence of somatic mutation was not obtained until immunoglobulin genes were cloned and sequenced. In 1977 Brack and Bernard showed that the inbred mouse strain carries only one germ-line *V*-region gene for the lambda chain and that its nucleotide sequence corresponds to the amino acid sequence found in 12 of the myelomas. The logical conclusion is that the six variants arose by somatic mutation.

Since then amino acid sequences have been compared with germ-line nucleotide sequences for a number of kappa light chains and heavy chains. In every case the diversity observed in the proteins has been greater than that of the germ-line DNA. Mutations are seen in the variable domain and in the immediately adjacent regions but not in the constant domains. Estimates of the rate of mutation suggest there should be one change in the *V* region for every three to 30 cell divisions, a rate several orders of magnitude greater than the average mutation rate for eukaryotic genes. It therefore seems that *B* cells or their progenitors carry an enzymatic apparatus for inducing mutations in the variable region of immunoglobulin genes. As yet the nature of the enzymes is entirely unknown.

The presence of both combinatorial and mutational mechanisms for the diversification of antibody genes is intriguing. Why have two systems evolved to meet the same need? Recent studies suggest a plausible explanation. Both mechanisms seem to be under strict control during the development of *B* lymphocytes. The recombination of the immunoglobulin gene



STRUCTURE OF THE *T*-CELL RECEPTOR is not yet known in detail, but its polypeptide components have been identified. Each receptor molecule includes one alpha chain and one beta chain; each chain in turn includes one constant and one variable domain. A region rich in hydrophobic amino acids anchors the protein in the membrane of the *T* cell.

segments is accomplished first, and it is complete by the time the cells are first exposed to antigens. It creates a population of cells that vary widely in their specificity, from which a few cells are selected by any given antigen. The mutational mechanism is then called into action during the proliferation of the selected *B*-cell clones. By altering individual nucleotide bases the mutations fine-tune the immune response, creating immunoglobulin genes whose products better match the antigen.

The effects that DNA-joining inaccuracy, base-pair insertion and somatic mutation have on antibody diversity are hard to quantify, but they surely increase the number of distinct antigen-binding sites by a factor of 100, and the factor is probably still larger. Thus if the combinatorial mechanisms alone give rise to 10 million antibodies, the total number could well be greater than a billion.

Given the intricacy of the *B* cells and their antibody-producing machinery, it is somewhat daunting to realize they constitute only half of the immune system. The *T* cells are just as complex and are essential to immunological competence. An animal deprived of *T* cells cannot mount an effective immune response to most antigens even though its *B* cells are intact.

There are three known subpopulations of *T* cells, all of them identical in appearance but distinguished by function. The cytotoxic *T* cells kill their target cells directly. The method of destruction is not known; an activated cytotoxic *T* cell binds to its target but does not engulf it (as a macrophage would), causing a lesion that kills the target cell. The other populations of *T* cells, called *T* helper cells and *T* suppressor cells, have a regulatory role. The *T* helper cells, when they recognize an antigen, stimulate other components of the immune system, including *B* cells and other *T* cells specific for the same antigen. The suppressor cells have just the opposite effect, that is, they diminish the activity of the same groups of cells.

The name *T* helper cell suggests an auxiliary or subordinate role, as if the cells merely abet a response that would take place even in their absence. Actually the *T* helper cells may be the master switch of the immune system. *B* cells, for example, recognize antigens independently of *T*-cell stimulation, but their proliferation and terminal differentiation usually requires activation by *T* helper cells. The *T* suppressor cells would seem to be equally important: by providing negative feedback they make the immune response self-limiting. They may also have a part in eliminating *B* and *T* cells directed against the self.

Because the *T* cells are antigen-specific, they must have receptor molecules analogous to the membrane-bound immunoglobulins of *B* cells. This fact was recognized more than 20 years ago, but the *T*-cell receptors proved difficult to analyze or even identify because they are not secreted in large quantities the way antibodies are. The receptors were first glimpsed in experiments done by James P. Allison of the University of Texas at Austin, John W. Kappler of the National Jewish Hospital in Denver and Ellis L. Reinherz of the Harvard Medical School. They prepared antibodies that bind to a protein on the *T*-cell surface; the protein identified in this way was considered a good candidate for the role of a receptor because it varies in structure from one clone of cells to another. The mass of the protein is about two-thirds that of an immunoglobulin, and it consists of two subunits, designated alpha and beta.

In 1984 Tak W. Mak and his colleagues at the University of Toronto and Mark M. Davis and his colleagues at the Stanford University School of Medicine cloned and sequenced a gene that is expressed and rearranged in *T* cells but not in *B* cells. Mak worked with cells from a human *T*-cell leukemia and Davis with a hybridoma, a cell line created by fusing a mouse *T* helper cell with a malignant *T* cell. In spite of the disparate origins of the two genes they were found to encode the same protein.

The nucleotide sequences analyzed by Mak and Davis are homologous to those of immunoglobulin genes, and there are also larger-scale indications of a familial resemblance to immunoglobulins. The genes are divided into scattered segments that can be rearranged in the developing *T* cell, and the upstream segments (corresponding to the amino end of the polypeptide) are variable whereas the downstream segments have a constant sequence. As in membrane-bound immunoglobulins, a series of hydrophobic amino acids near the carboxyl end of the protein anchor the molecule in the membrane. A direct determination of amino acid sequences by Reinherz and his coworkers has since confirmed that the genes specify the beta subunit of the *T*-cell receptor.

Two more *T*-cell DNA clones were isolated by Haruo Saito, working in my laboratory at the Massachusetts Institute of Technology, and David M. Kranz in Herman N. Eisen's laboratory, also at M.I.T. In this case the genes came from mouse cytotoxic *T* cells; nevertheless, the downstream part of one clone is essentially identical with the constant-region gene for the helper-cell beta chain. The second DNA sequence has a number of properties in common with the beta-chain genes. It is homologous to immunoglobulin genes, is made up of segments that are rearranged and expressed only in *T* cells and has a hydrophobic anchor segment. The logical hypothesis was that the gene encodes the alpha chain of the receptor molecule.

Soon afterward, however, Saito isolated a third gene from the same clone of cytotoxic *T* cells. It too has all the properties expected of a *T*-cell receptor, and it has an additional factor in its favor. Chemical analysis of the protein, done in parallel with the gene-cloning studies, showed that the receptor protein has carbohydrate side

chains attached to it through the amino acid asparagine. The earlier alpha-chain candidate has no asparagine units in the appropriate positions, whereas the new one has several suitable asparagines. A partial determination of the amino acid sequence of the alpha subunit by Kappler and his co-workers has confirmed that the third of the DNA clones is the true alpha-chain gene. The same gene was also isolated from a helper-cell hybridoma by Y.-H. Chien and others in Davis' laboratory at Stanford.

In this account the second gene found by Saito and Kranz—the discarded alpha-chain candidate—would seem to be left without a function. It is so closely related to the other genes, however, that it almost certainly has some role in the recognition of antigens. The protein it is presumed to encode is now designated the gamma chain; I shall discuss below what part it might play in the *T* cell's action.

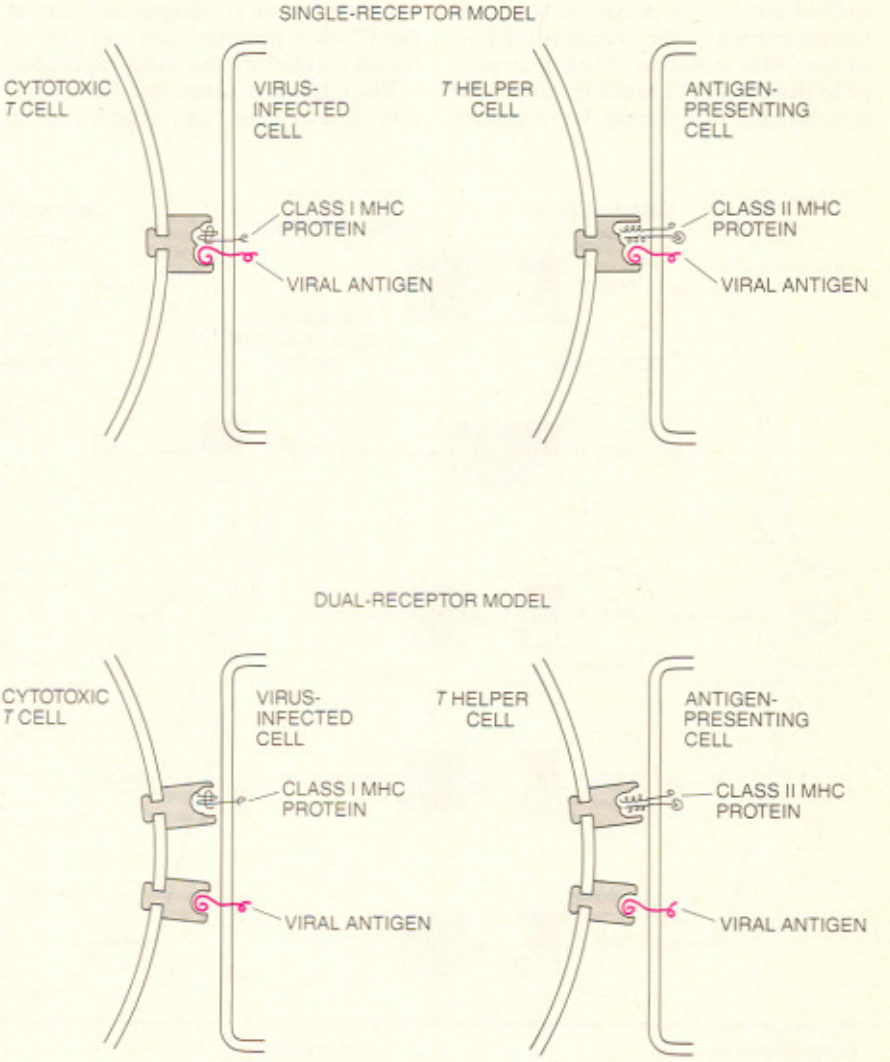
From the nucleotide sequences specifying the alpha and beta chains much of the structure of the *T*-cell receptor can be inferred. Each chain is composed of two domains, which are similar in overall structure to the repeated domain of the immunoglobulins. The degree of sequence homology with the immunoglobulins is between 25 and 35 percent. The two chains are linked by a disulfide bond between the constant region and the membrane-anchoring peptide. Molecules derived from helper cells and cytotoxic *T* cells have identical constant regions in both the alpha and the beta chains.

The molecular genetics of the *T*-cell-receptor molecules is also strikingly similar to that of immunoglobulins. The variable regions of both receptor chains are encoded by three gene segments, corresponding to the *V*, *D* and *J* segments, which are scattered along a chromosome in germ-line cells but are fused in mature *T* lymphocytes. The heptamer-nonamer signal sequences associated with immunoglobulin genes are also found near the *T*-cell-receptor segments, indicating that the same system of enzymes or a very similar one is employed to mediate the somatic recombination.

Considering all the similarities between genes for immunoglobulins and those for *T*-cell receptors, it seems reasonable to speculate that the two kinds of protein might recognize antigens in the same way. In other words, the *T*-cell receptors might have an antigen-binding site formed by clusters of highly variable amino acids in specific subregions within the variable regions of the alpha and beta chains. It is an appealing hypothesis in that it supplies a single explanation for the recognition abilities of both proteins. Even if the hypothesis turns out to be correct, however, it cannot be the whole story. The reason is that the two branches of the immune system recognize antigens in different circumstances. A *B* cell can respond to an antigen alone, but an individual's *T* cells are activated only if the antigen is displayed on the surface of a cell that also carries markers of the individual's own identity.

To describe this difference in antigen response it is necessary to introduce the molecules that serve as markers of individual identity. They are proteins encoded by a large cluster of genes called the major histocompatibility complex, or MHC, and they make up a third class of proteins with a vital role in immune recognition.

The MHC proteins were discovered in tissue-grafting experiments. Unless the donor and the recipient of a graft are genetically identical (as in the case of identical twins or mice of an inbred strain) the graft is generally rejected because the recipient mounts an immune response to the donor's MHC proteins. The prevalence of graft rejection implies that unrelated individuals almost always express different sets of MHC genes. Indeed, apart from immunoglobulins and *T*-cell receptors, the MHC proteins are the most diverse ones known. Whereas antibodies and *T*-cell receptors vary from one cell to the next, however, the MHC proteins



***T*-CELL-RECEPTOR SYSTEM, unlike an antibody molecule, does not respond to an antigen alone; the antigen must be presented on the surface of a cell that also displays one of the proteins of the major histocompatibility complex (MHC). Cytotoxic *T* cells recognize an antigen in combination with a Class I MHC protein, found on almost all body cells. *T* helper cells bind to an antigen associated with a Class II MHC protein; the two molecules are confined to cells of the immune system, such as macrophages and lymphocytes. It is not yet clear whether *T* cells have a receptor with two binding sites or two separate receptor molecules.**

differ from one individual to another.

Two classes of MHC proteins have been identified. A Class I molecule consists of a large polypeptide chain (about the size of an immunoglobulin heavy chain) linked to a much smaller subunit called beta-2 microglobulin. Class I MHC proteins are found on the surface of virtually all cells. The Class II proteins, in contrast, appear only on a few types of cells that have a part in the immune response, such as *B* lymphocytes, macrophages and specialized epithelial cells. A Class II molecule is also made up of two polypeptide chains, both about the size of an immunoglobulin light chain. All the MHC polypeptides exhibit some degree of homology with immunoglobulins, although the resemblance is not as strong as that between *T*-cell receptors and immunoglobulins.

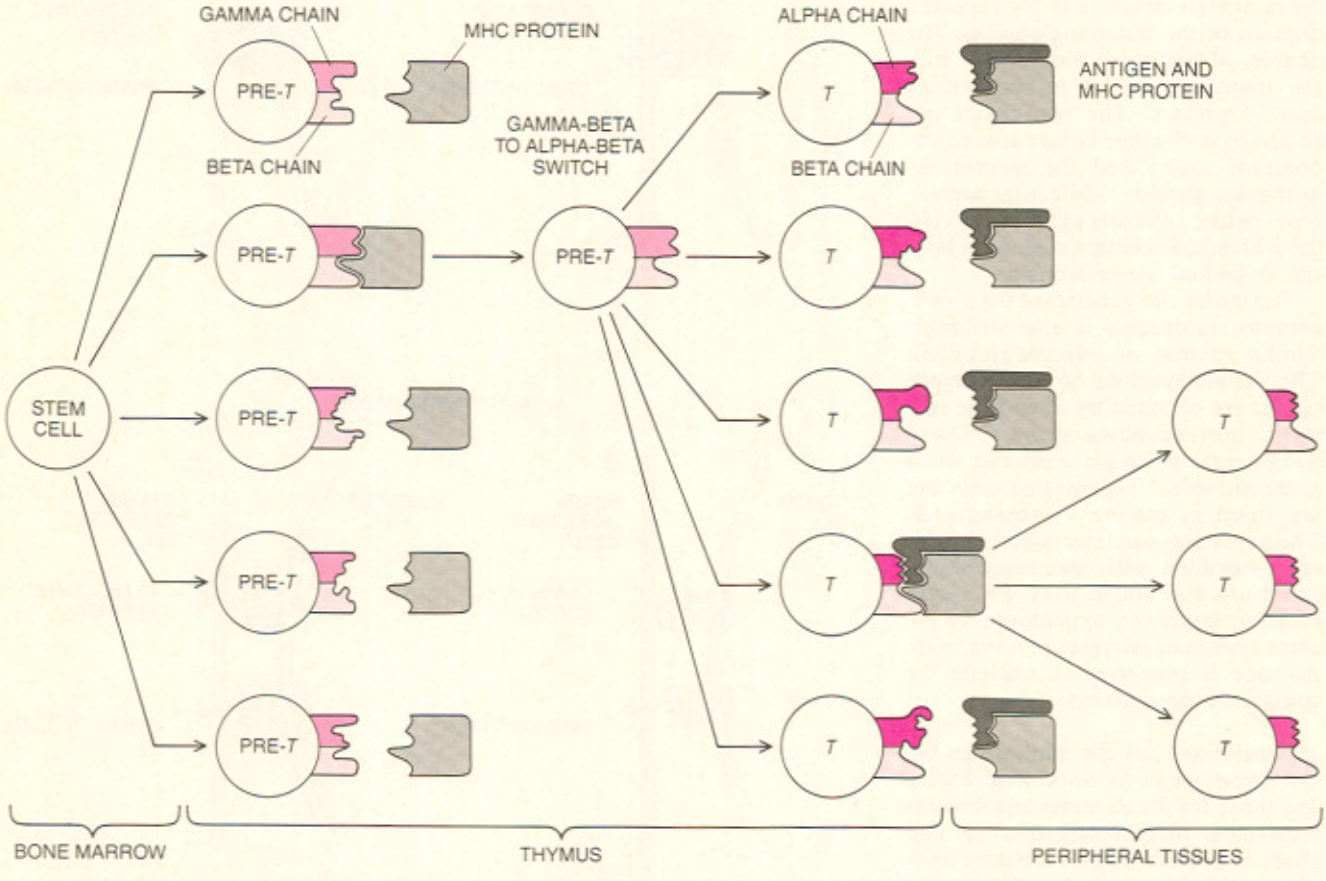
Since the transfer of tissue between individuals is rare in nature, graft rejection cannot be the primary function of the MHC proteins. Their true purpose lies elsewhere in the immune system, namely in directing the responses of *T* cells. A *T* cell recognizes both an antigen and a self MHC protein on the surface of a single cell. The requirement of dual stimuli is called MHC restriction. Cytotoxic *T* cells respond to antigen together with a Class I MHC protein; *T* helper cells require a Class II protein.

Of what benefit to the organism is MHC restriction? Its effect is to direct the activities of *T* cells to the surface of the animal's own cells rather than to bacteria or free foreign molecules. One plausible idea is that the cytotoxic *T* cell arose to provide protection against viral infection. When a virus enters a cell, coat proteins encoded by the viral genome are displayed on the cell membrane. Hence the infected cell has just the right pattern of surface markers for recognition by *T* cells: a foreign molecule in combination with native proteins. A cytotoxic *T* cell, recognizing the viral antigen and one of the Class I proteins, can kill the infected cell before the virus replicates.

The Class II proteins and the regulatory *T* cells may have evolved to increase the efficiency of the immune response. *T* helper cells can be triggered by antigen-presenting cells that take up circulating antigens and display them on their surface along with Class II MHC proteins. The presence of Class II molecules on *B* lymphocytes and macrophages may be the key to how helper cells communicate with those cells and thereby recruit their participation in an immune response.

If a *T* cell must recognize two surface markers, the question arises of whether it has two separate receptors or a single dual-function receptor. Some recent experiments seem to favor the single-receptor model, but the results are by no means conclusive. If a second receptor is eventually found, it may turn out to incorporate the "orphaned" gamma chain, which has all the properties expected of a receptor protein but has no place in the current scheme of *T*-cell operations.

There is another possible role for the gamma chain. *T* lymphocytes become mature and functional only after a pe-



THYMIC "EDUCATION" is a necessary stage in the development of functional *T* lymphocytes. A model proposed by David Raulet and the author offers one possible explanation of the developmental process. According to the model, immature *T* cells first make receptor proteins with polypeptide chains called gamma (*color*) and beta (*light color*). In the thymus the lymphocytes are exposed to MHC proteins, and only those cells with sufficient affinity for these markers of self-identity are allowed to propagate. If the selected cells were released from the thymus, however, they would attack the body's own tissues. The affinity of the receptor molecules for self-antigens must therefore be reduced. Each receptor retains the beta chain of the selected clone, but the gamma chain is replaced by one of a variety of alpha chains (*solid color*). The modified *T* cells respond to a self-MHC protein in combination with an antigen.

riod of residence in the thymus, and it is during this period of "thymic education" that *T* cells come to recognize antigens only in combination with the individual's own MHC proteins. No definitive explanation of thymic education has been given, but many immunologists agree that the crucial step must be selection of a subpopulation of immature *T* cells through their interaction with self MHC proteins displayed to them by thymic cells. According to one model, each immature *T* cell responds to just one MHC protein (or to a small group of them), but the total population includes cells responsive to all possible markers. In the thymus only those cells that have sufficient affinity for the native MHC proteins are allowed to propagate and continue their differentiation.

For this scheme to work immature *T* cells must be able to recognize and respond to MHC molecules alone, without an accompanying antigen. When the mature *T* cells are released from the thymus, they have obviously lost that capability; otherwise they would be aggressive against the body's own tissues. What is the biochemical basis of the change in reactivity? Recent studies indicate that in immature *T* cells the alpha gene is expressed at low levels, whereas the beta and gamma genes produce larger quantities of protein. On the basis of these findings David Raulet of M.I.T. and I have proposed a model of *T*-cell development that we call the gamma-beta-to-alpha-beta switch.

In the model immature cells have receptors made up of a gamma chain and a beta chain and are responsive to MHC proteins alone. In the course of differentiation the gamma gene is turned off and the alpha gene is turned on, so that mature cells have alpha-beta receptors. The change reduces the cell's affinity for self MHC proteins but, because the beta chain is still present, does not extinguish it entirely. An analogous mechanism operates in red blood cells when they switch from the fetal to the adult form of hemoglobin.

Our proposal for the gamma chain's function has not yet been tested, but the tools are now in hand to settle this question and many others about the nature of the *T*-cell receptor. Ideally, structural and genetic studies would yield an understanding of these molecules at the same level of detail that can now be given for the immunoglobulins. At that point one might hope to resolve some of the major remaining enigmas of immunology: how *T* lymphocytes develop in the thymus, how they recognize their target cells and how they control the rest of the immune system.