Antibody Diversity: How Many Antibody Genes?

An individual animal may have the capacity to synthesize as many as a million different antibodies. Since mammalian cells are estimated to contain only about 30,000 genes that code for proteins, it has been difficult to reconcile this diversity of antibodies with the theory that information for the structure of different proteins is encoded in different genes.

About 15 years ago, three models were advanced to explain how antibody genes may be organized so as to encode information for the synthesis of antibodies in relatively small segments of DNA. The models lead to predictions of different numbers of antibody genes; but, until recently, it was impossible to test those predictions. Now investigators are using new techniques of molecular biology to count the number of antibody genes in the cells of a single animal. As a result of these gene counts, one of the three models—in which it is proposed that antibody diversity arises from mutations in a small set of antibody genes—is gaining increased acceptance.

The problem of explaining the genetic nature of antibody diversity can be simplified when the structure of antibody molecules is taken into account. Each antibody is composed of four polypeptide chains: two identical light chains, each of which consists of about 214 amino acids, and two identical heavy chains, each of which consists of about 440 amino acids. Light and heavy chains are assumed to combine at random to form antibody molecules. Thus a model of antibody diversity need only explain how information for the synthesis of 1000 different light chains and 1000 different heavy chains is encoded in DNA. This would allow for production of 1 million different antibodies.

Mouse light chains, which resemble those of humans, can be divided into two groups: κ chains and λ chains. A light chain from either group is identical to all others in the same group for about half the molecule (the constant, or C, region) and differs from others in the group for the remainder of the molecule (the variable, or V, region) (Fig. 1). Mouse light chains, then, have only two kinds of constant regions. Constant regions do not affect the specificity of an antibody for an antigen. Instead, it is the variable regions that are responsible for recognizing and binding to antigens.

All three models of the organization of antibody genes are based on the hypothesis that each light chain (or heavy chain) is coded by a gene for the variable region (V gene) and a gene for the constant region (C gene). In the first model, it is assumed that each cell of an animal inherits all of the genes necessary for the synthesis of at least 2000 different light and heavy chains and so at least 1 million different antibodies. However, none of those genes would be expressed in cells that do not produce antibodies. Each cell that produces an antibody usually expresses genes that code for the synthesis of only one kind of light chain and only one kind of heavy chain, and thus one kind of antibody. Proponents of this first model assume, moreover, that each V gene is adjacent to a C gene on the DNA (Fig. 2). This assumption is consistent with the theory that a messenger RNA (mRNA) for a light or heavy chain contains, in one molecule, a segment copied from a V gene adjacent to a segment copied from a C gene. Since all mouse light chains have one of two constant regions, the same two C genes would be repeated many times on mouse DNA.

The first model of the organization of antibody genes was previously questioned as a result of strong genetic evidence against it. Recently, it was conclusively disproved by experiments in which C genes were counted. A specific C gene was found to occur at most five times on mouse DNA. This result limits the types of models of the organization of antibody genes that may be advanced. The experiments were performed independently by several investigative groups, including those at the laboratories of Cesar Milstein of the Medical Research Council in Cambridge, England, Bernard Mach of the University of Geneva in Switzerland, and Philip Leder of the National Institute of Child Health and Human Development in Bethesda, Maryland, and by Janet Stavnezer of the University of California in San Francisco.

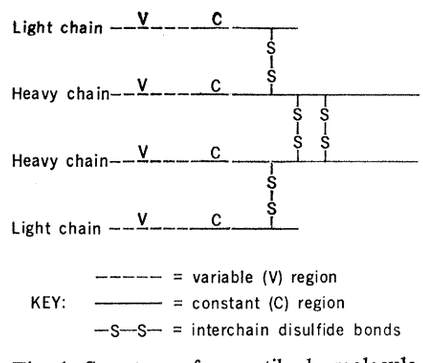

Fig. 1. Structure of an antibody molecule.

In order to count the number of times that C genes are repeated in a mouse cell, investigators at the four laboratories used molecular probes that are specific for the C genes for mouse κ chains. The probes were complementary DNA copies of part of an mRNA for a κ chain. The mRNA was obtained from a myeloma (a clone of cancerous antibody producing cells). Cells from a myeloma are all derived from one antibody producing cell and so they all synthesize identical copies of an antibody molecule. The mRNA for that antibody is present in the myeloma cells in large enough quantities to be isolated and purified.

Complementary DNA was copied from the mRNA's from a myeloma by means of a reverse transcriptase. The reverse transcriptase only yields a copy of 500 or so nucleotides of the 1100 nucleotides that compose the κ chain mRNA. However, this group of nucleotides includes part of the mRNA that contains the sequence of the C gene. Thus the complementary DNA copy of 500 nucleotides of κ chain mRNA can be used as a molecular probe that is specific for C genes. The probe will bind to C genes on mouse DNA when it is placed in solution with that DNA. The rate at which it can find and bind to its complementary C gene is a function of the number of times that the C gene is repeated on the DNA. The molecular probe for the C genes bound to DNA so slowly that Milstein, Mach, Leder, and Stavnezer concluded that mouse DNA contains very few copies of the C gene for κ chains and thus ruled out the first model of the organization of antibody genes.

The second and third models of the organization of antibody genes have proved more difficult to evaluate than the first. The second model resembles the first in that it is based on the hypothesis that all of the genes necessary for the synthesis of 1 million antibodies are inherited by each cell of an organism. However, the second model differs from the first in that, according to this model, C genes are represented only a few times per cell (Fig. 2). The question of how a copy of a V gene

and a copy of a nonadjacent C gene could compose one mRNA is left unanswered by this model, although various investigators have speculated as to how such an mRNA could occur. The second model leads to the prediction that an arbitrary mouse cell will have only a few C genes but will have as many V genes as variable regions of antibodies produced by the animal.

The third model of the organization of antibody genes differs from the first two in that it is based on the proposition that each cell of an animal inherits only a small set of antibody genes. Antibody diversity would result from random mutations of those genes in cells that produce antibodies. This third model leads to the prediction that an arbitrary mouse cell would have far fewer V genes than variable regions of antibodies produced by that animal.

The second and third models could be tested if the V genes in a mouse cell were counted and the number of V genes compared to the number of variable regions of the mouse's antibodies. Initial attempts to count V genes led to controversies. Terry Delovitch and Corrado Baglioni, who were then at the Massachusetts Institute of Technology in Cambridge, reported that there are at most a few hundred V genes per cell, whereas Alan Williamson of the University of Glasgow in Scotland reported that there may be several thousand V genes per cell. Recently, results reported by Leder and his associate Tasuku Honjo and by Susumu Tonegawa of the Basel Institute for Immunology in Switzerland and his associates have convinced many researchers that the number of V genes for light chains in mouse cells is less than the number of variable regions of mouse antibodies.

Leder and Honjo counted the number of V genes for λ chains in mouse cells. They chose to study λ chains because the amino acid sequences of these molecules had been determined by Martin Weigert and Italo Cesari while associated with Melvin Cohn of the Salk Institute in San Diego, California. Leder used results of Weigert and his associates as an essential part of his experimental design.

Weigert and his associates studied 18 different myelomas from an inbred strain of mouse. All of these myelomas consisted of cells that produced antibodies with λ chains. Although the myelomas arose spontaneously and independently, 12 of the 18 λ chains isolated from them had identical amino

(A) $V_1CV_2CV_3C \cdots V_nC$

(B) $V_1V_2V_3 \cdots V_nC$

Fig. 2. Models of the organization of antibody genes. (A) Each V gene is adjacent to a C gene. (B) C genes are represented only a few times per cell, whereas there are many V genes.

acid sequences. The remainder, which consisted of six different λ chains, varied by at most three amino acids. Each change took place in the variable region and all but one involved a single base change in the nucleotide sequence of a V gene; Weigert and Cohn interpret this as best explained by a mutation theory of antibody diversity.

Leder and Honjo used the result that seven variants of the mouse λ chain were isolated and sequenced to see if an arbitrary cell from the same inbred strain of mouse contained one or seven or more genes for λ chains. If each mouse cell inherits all the genes necessary for the production of 1 million different antibodies, each cell should have at least seven V genes for the λ chain.

The differences in the variable region of the seven different λ chains are so small that a molecular probe for one V gene should bind equally well to all the others. The experiment of Leder and Honjo was based on a determination of how quickly a molecular probe that consisted of an mRNA fragment that was transcribed from, and thus was complementary to, a V gene for a λ chain could find and bind a V gene on mouse DNA. If there were seven V genes for λ chains per cell, the probe would recognize them as seven complementary genes and would bind more rapidly than if there were only one V gene per cell. By analyzing binding rates, Leder and Honjo showed that the V gene molecular probe reassociated at about the same rate as a molecular probe for a hemoglobin gene. Since the hemoglobin gene is thought not to be repeated, Leder and Honjo believe that there are only one or a very few V genes for λ chains per cell.

Tonegawa and his colleagues designed an experiment that allowed them to count the number of V genes for κ chains of mouse cells. Mice produce many more κ than λ chains, and many more differences have been observed in the amino acid sequences of the variable regions of κ chains. However, those differences apparently do not occur at random. The variable regions of κ chains can be classified into subgroups. All members of a subgroup have the same variable region except for a few amino acids in a segment called the hypervariable region. The hypervariable region is the site where the antibody binds to an antigen. Thus changes in a hypervariable region will allow an antibody to recognize different antigens.

Tonegawa and his colleagues studied the rates at which various molecular probes that consist of mRNA's for κ chains bind to mouse DNA. They interpret their results as indicating that there are only as many V genes for κ chains as there are subgroups of the variable regions of κ chains. Each subgroup contains many variants and these variants, according to Tonegawa, arise by mutations of a basic subgroup gene in the region of the gene that codes for a hypervariable region.

As increasing numbers of immunologists are expressing a belief that antibody diversity arises from mutations, an old question has arisen: Are ad hoc mutation mechanisms necessary to account for observed variations in antibody genes? Such mechanisms have been, and are being, proposed, but there are no reasons to believe that they operate in antibody producing cells.

Weigert, Cohn, and their associates believe that randomly occurring mutations are sufficient to account for antibody diversity. They postulate that each cell in an animal contains genes that code for about 100 different light chains and 100 different heavy chains. When a cell produces an antibody that recognizes and binds to an antigen, that cell will be stimulated to divide rapidly and will produce many more copies of that antibody. As cells replicate, mutations will occur at random in their DNA sequences. When such a mutation occurs in the hypervariable region of a V gene in an antibody producing cell, that cell will produce an antibody with altered specificity for antigens.

Cohn and Weigert calculate that antibody cells replicate and die so often that randomly occurring mutations can account for antibody diversity. Whether or not such a process takes place, the fact that it is feasible adds considerable weight to the arguments in favor of a mutation theory of antibody diversity. Although there is no conclusive evidence for the mutation theory, it is becoming the theory of choice by default as predictions of alternative theories have been shown not to be fulfilled.

—GINA BARI KOLATA