done later, independently, by M. S. Raghunathan of Tata Institute in Bombay after a comparably long analysis.

At this stage however, the case of cocompact lattices still loomed before the international community of researchers as a blank wall. In 1974, in a brilliant stroke, Margoulis realized how to scale that wall. In 1965, I had introduced a strategy for solving a related problem: if one starts with a particular lattice, is there only one group that could contain that lattice and is there only one possible location for that lattice in the group? In a project extending until 1973, I succeeded in proving that the answer is "Yes." According to Margoulis, my proof was of special interest because of its new conceptualization of the problem and because it introduced, for the first time, the use of ergodic theory in its analysis.

Then Margoulis took a bold step in his analysis of cocompact lattices. He took a lattice in one setting and considered its (possibly degenerate) image in another setting. He then used a mixture of algebra, analysis, and number theory to finally solve the problem of the structure of these lattices. Actually, his results apply not only to lattices in continuous group but more generally to lattices in other sorts of groups as well, specifically to lattices in algebraic groups over either **R** or $Q_p$ for any prime number $p$.

When I lectured on Margoulis's results at Harvard in 1974, David Mumford, a Fields medalist in that year, entitled the talk "Recent breathtaking results of G. A. Margoulis." This unwonted adjective for a mathematical topic perhaps helps convey the electrifying excitement generated by Margoulis's result among the mathematicians of the world.

Unfortunately, Margoulis was not granted permission to travel to Helsinki to accept his Fields medal. In homage to his achievements, which were described by Jacques Tits of College de France at the award ceremony in Helsinki on 14 August 1978, the entire audience in Finlandia Hall rose to its feet, in a spontaneous gesture of admiration for the medalist who was so conspicuously absent.

GEORGE D. MOSTOW
*Department of Mathematics,*
*Yale University,*
*New Haven, Connecticut 06520*

# Antibodies (I): New Information About Gene Structure

Until recently, techniques for the direct examination of genes were not available. That situation has changed, however, as a result of the revolution in molecular biology that began early in this decade. It is now possible, for example, to pick out an individual gene from among the tens of thousands in the mammalian genome, then to use recombinant DNA techniques to manufacture enough copies of the gene to study and finally to determine the sequence of the nucleotides in that gene, and all in less than a year's time.

One of the more rewarding applications of the techniques has been to the study of antibody genes. Investigators are acquiring much new information about the numbers and arrangements of the genes in the mammalian genome. They have also determined the nucleotide sequences of at least four of these genes. The research has provided both a direct confirmation of a long-held hypothesis about the organization of antibody genes and some surprising new revelations about that organization. In addition, the gene studies are providing important clues to the solution of one of the long-standing problems of immunology—that is, how to account for the ability of a single animal to make as many as a million different antibodies.

What the DNA studies have confirmed about antibody gene arrangement is that there are separate genes coding for the variable and constant regions of antibody chains. An individual antibody molecule consists of four polypeptide chains—two identical light chains and two identical heavy chains (Fig. 1). Each of these polypeptide chains in turn consists of a variable region and a constant region. The variable regions of the light and heavy chains form the part of the antibody that combines with the appropriate antigen. Thus the variable regions must differ from antibody to antibody. But the amino acid sequence of the constant region is the same for all chains of the same type.

In 1965, William Dreyer of the California Institute of Technology and J. C. Bennett of the University of Alabama School of Medicine suggested "the two gene-one polypeptide theory" for the synthesis of antibody proteins. They proposed that two genes, one for the variable and one for the constant region, were needed for the production of a single antibody chain. Since then the evidence has been consistent with this hypothesis, but direct demonstration of the separate genes was not achieved until 1976, when Susumu Tonegawa and his colleagues at the Basel Institute for Immunology showed that the DNA segments coding for the constant and variable regions of mouse light chains are separate from one another in embryonic cells.

The surest piece of evidence for the separation came when Tonegawa and his colleagues, in collaboration with Walter Gilbert and Allan Maxam at Harvard University, determined the complete nucleotide sequence of a fragment of embryonic DNA carrying the gene for the variable region of a light chain. They found that beyond the codon (a sequence of three nucleotides specifying a particular amino acid) for amino acid 98 of the variable region, there was no agreement between the nucleotide sequence of the DNA and the amino acid sequence of any light chain. Thus, they concluded that the DNA beyond codon 98 did not specify any light chain structure and that the gene for the constant region could not be attached to this codon.

Although the genes for the variable and constant regions of an antibody protein might be separated from one another in embryonic cells, which do not make antibodies, immunologists assumed that the two genes would somehow get together in mature antibody-producing cells. And when the Basel workers looked at antibody gene patterns in a line of mature cells, they found that the genes coding for the variable and constant regions of the light chain produced by the cells appeared to be joined.

But on closer examination of the structure of the DNA encompassing both genes—genes Tonegawa and his colleagues thought would be connected directly to one another—the unexpected happened. They found that between the genes for the constant and variable regions of the light chains, the DNA contained a segment of about 1250 bases that was missing from the messenger RNA (mRNA) for the light chain in question. Thus, the DNA contained a nucleotide segment that could not be translated into protein structure since the sequence was


0036-8075/78/1020-298$00.50/0 

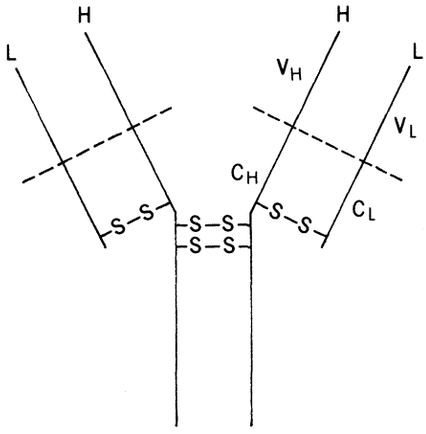


Fig. 1. Diagram of the structure of an antibody molecule. The letters *H* and *L* designate the heavy and light chains, respectively; *C* and *V* refer to the constant and variable regions of the appropriate chains.

not present in the mRNA that directs light chain synthesis. This intervening DNA sequence is one of the spacer sequences that have recently been discovered in a variety of genes from nucleated cells (*Science*, 3 February 1978, p. 517).

In addition to finding this long intervening sequence, the Basel and Harvard investigators also found a shorter segment of 93 untranslated nucleotides in the embryonic gene they sequenced. This intervening sequence was an insert within the DNA coding for the "leader" portion of the light chain. The insert is located near the junction between the leader DNA and the gene for the variable region. (Leader sequences of proteins are thought to be necessary for the transport of newly synthesized proteins out of the cell. They are eventually removed from the protein.)

The discovery that the gene for the variable region of the mouse light chain terminated after codon 98 was also unexpected, because the variable regions of light chains were generally thought to extend to amino acid 112. Of course, there was always the possibility that the variable region coded for by the analyzed gene was shorter than usual, but recent evidence suggests a much more intriguing explanation for the finding.

Termination of genes for light chain variable regions in the vicinity of codon 98 may be a general phenomenon. The nucleotide sequences of three additional genes have now been determined—two by Philip Leder, Jonathan Seidman, and their colleagues at the National Institute of Child Health and Human Development (NICHD) and the third by Tonegawa and his colleagues. They all terminate after codon 97.

The mouse produces only two main types of light chain, which are designated λ and κ. The genes studied by the Basel workers code for the variable regions of mouse λ chains and those sequenced by the NICHD group specify the variable regions of κ chains. Thus, the variable region genes for both types of chains turned out to be shorter than expected and the DNA coding for the 13 or so amino acids between amino acid 97 and the beginning of the constant region appeared to be missing. (This stretch of 13 amino acids is now called the J region because it joins the variable and constant regions of antibody chains.)

Tonegawa and his colleagues have recently found the missing DNA segment, however. In both gene-mapping and gene-sequencing experiments they have shown that, in embryonic cells, the DNA coding for the J region is not directly connected either to the variable or constant region genes. It is an undetermined number of nucleotides away from the gene for the variable region and 1250 nucleotides from the constant region gene. But in DNA from cells secreting light chains, the variable and J region genes become contiguous, although the constant region gene remains that same 1250 nucleotides away from the J sequence (Fig. 2).

Tonegawa has hypothesized that rearrangement of embryonic DNA to bring together the variable and constant region genes that are going to be expressed is part of the maturation process of cells that make antibodies, a suggestion supported by the recent findings on the arrangement of light chain genes in embryonic and mature cells. The rearrangement may also be a part of the process that activates a particular variable-constant region gene combination.

A given cell synthesizes only one species of antibody. In their examination of different cells, Tonegawa and his colleagues have observed that the only variable region gene expressed in a given cell line is the one joined to the gene for the constant region. It now appears that the joining event brings the variable region gene together with the J sequence. Moreover, depending on your definition of gene (which is not as clear as it used to be because of the discovery of intervening sequences in the genes of nucleated cells), it may now be necessary to invoke the "*three* gene-one polypeptide hypothesis" to explain the synthesis of antibody chains.

The concept of independently coded variable and J region sequences is supported by the results of protein structure analysis. Several investigators, including Michael Potter of the National Cancer

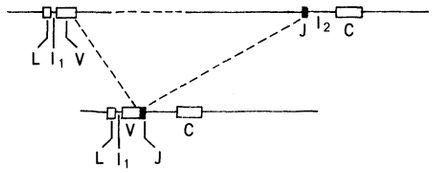


Fig. 2. Diagram of the arrangement of the gene segments needed to make up a single light chain. The upper portion of the diagram shows their arrangement in DNA from mouse embryo cells, and the lower portion indicates that in DNA from a line of mature antibody-producing cells. In the mature cells the *V* (for variable) and *J* (joining) region genes, which were far apart in embryonic DNA, are now adjacent, whereas the *J* and *C* (constant genes remain about 1250 nucleotides apart. *L* refers to the DNA coding for the leader sequence of the light chain, and $I_1$ and $I_2$ designate the two intervening sequences of DNA that are not translated into protein structure. [Adapted from a diagram by Susumu Tonegawa of the Basel Institute for Immunology]

Institute, David McKean of the Mayo Medical School, Martin Weigert of the Institute for Cancer Research, and Lee Hood of the California Institute of Technology, have been determining the amino acid sequences of the members of a group of light chains called the $V_κ21$ chains. These polypeptides all have very similar amino acid sequences from their amino terminal (amino acid 1) through amino acid 23, but beyond that their sequences begin to vary considerably. Weigert and Hood have now sequenced 22 of these chains. One of their findings is that the proteins appear to consist of two distinct segments, one ranging from amino acid 1 to 98 and the other from amino acid 99 to 112. In other words, arrangement of the protein segments corresponds to that of the genes.

The function of the J segment is unknown. But two possible roles can be visualized. One involves the provision of signals for the various nucleic acid rearrangements needed for the expression of antibody genes. Tonegawa hypothesizes that the left half of the J gene may carry a signal for the joining of the V and J region genes. In addition, work by several investigators indicates that the intervening sequence between the J and constant region genes is transcribed into RNA and then subsequently excised in some way. The right half of the J region gene might provide the signal for the excision.

The second function of the J region could be the generation of antibody diversity. How the new results on antibody and antibody gene structures are contributing to the clarification of the diversity problem will be considered in the second article of this series.—JEAN L. MARX