Somatic Generation of Immune Diversity

Susumu Tonegawa

KEY WORDS: immune diversity; lymphocyte; receptors; somatic mutation; T cell.

One day in the fall of 1970, I received an airmail letter from Renato Dulbecco who was traveling in Europe. At that time I was a postdoctoral fellow in his laboratory at the Salk Institute. The letter, written on stationery of the Hotel Hassler in Rome, said:

"Dear Susumu,

I don't know what arrangements you have made for after your departure from La Jolla at the end of the year but I would like to mention to you another possibility. The Institute of Immunology in Basel, Switzerland will start operating in a month. They already have an excellent collection of immunologists, but have not yet built an adequate background in molecular biology. I talked about you to Niels Jerne, the Director, and they are interested in having you there. There are many immunologically interesting phenomena obtained with crude RNA preparations but they are unreliable because RNA is not characterized. In general, it seems the best system for understanding development at a molecular level and you may like to get into such a field. If you are interested, write to Niels K. Jerne, Basel Institute for Immunology, 487 Grenzacherstrasse . . ."

Thanks partly to this remarkably prophetic letter and partly to the US immigration law that prevented me from remaining in the USA, in February 1971 I found myself in this cozy Swiss town almost completely surrounded by immunologists. For someone who had had no formal training in immunology whatsoever and had never even visited Switzerland, it was a rather drastic change. Indeed, the first twelve months at the Basel Institute were not easy.

After I arrived in Basel I initially attempted to continue the project of my days in Dulbecco's laboratory, namely, the transcriptional control of the simian virus 40 genes. However I soon realized that this was not a subject that aroused great interest in an Institute almost entirely staffed by immunologists nor one that allowed me to take full advantage of my many talented colleagues. I therefore decided to learn immunology by talking to them, reading papers, and asking

Center for Cancer Research and Department of Biology Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

questions. An immunologist, Ita Askonas, and a geneticist, Charlie Steinberg, became my tutors and were most helpful to me in getting into a new field. It was during this process that I was introduced to the problem of the origin of antibody diversity.

THE PROBLEM

Immunologists agreed that an individual vertebrate synthesizes many millions of structurally different forms of antibody molecules even before it encounters an antigen. Moreover, Gerald Edelman and Rodney Porter had shown that a typical antibody molecule is composed of two identical light chains and two identical heavy chains (1, 2). It had also been found that each of these two types of chain exhibits great sequence variability in the amino terminal region between one antibody molecule and the next and no sequence variability in the carboxyl terminal regions (3). These two regions were then referred to as the variable, or *V,* and the constant, or *C,* regions. However immunologists and geneticists were divided for many years into two schools of thought with respect to the issue of whether the genetic diversity required for the synthesis of these proteins is generated during evolution and is carried in the germline or during development in which case it would be present in somatic but not germline cells. One school of thought held that the germline must include a separate gene for every polypeptide that ultimately appears in an antibody molecule (4). In this germline theory, antibody or immunoglobulin genes are expressed in exactly the same way as those for any other protein, and no special gene-processing mechanisms are needed. On the other hand, the model requires an enormous number of immunoglobulin genes inherited from the parents. While the four chain structure of an immunoglobulin molecule allows diversity to be generated by chain paring, the number of genes required for both light and heavy chains is still very large. One major difficulty for germline theories of antibody diversity was the observation that all antibody polypeptide chains of a given type share a common genetic marker (allotype) that segregates as a single Mendelian gene. If there were many thousands of light and heavy chain genes, how could the same genetic marker in all of these genes have been maintained?

The second theory supposed that there are only a limited number of antibody genes in the germline, and that these genes somehow diversify as the antibody-forming B lymphocytes emerge from their stem cells. In other words, the diversification of antibody gene sequences takes place in specialized somatic, or body, cells rather than being carried from generation to generation by the germ cells (5–7). One attraction of this latter theory is that it relieves the host of the need to commit a disproportionately large fraction of the inherited genes to code for antibodies, but the theory demands an unprecedented mechanism for diversifying the inherited genes somatically.

Arguments for and against these contrasting ideas were made both vocally and in written form for many years. However, all of these arguments were based on the interpretation of amino acid sequences of immunoglobulin polypeptide

chains or on the generally accepted principles of evolution and genetics. No direct evidence for either view had been obtained. This was because no technique was available that would allow an analysis of the fine structure of specific genes from higher organisms.

GENE COUNTING

In the early seventies the technology for purifying a specific eukaryotic mRNA was just becoming available. Furthermore a method to determine the number of copies of a specific gene by kinetic analysis of nucleic acid hybridization had already been established (8, 9). These technical developments led some scientists, including myself, to think that one can experimentally determine the number of immunoglobulin genes contained in a germline genome and thereby decide which of the two major theories of antibody diversity is correct. The validity of this approach is based in part on the fact that the V region of a given chain type, while being different, exhibits a high degree of amino acid sequence homology. It was therefore thought a mRNA coding for a specific immunoglobulin polypeptide chain would hybridize not only with its own gene but also with many other immunoglobulin genes, it they existed in a germline genome.

I thus obtained mouse myeloma cells and put my effort to purifying immunoglobulin mRNA and carrying out the hybridization studies. However the initial studies focusing on the mouse κ light chain and heavy chain genes gave ambiguous results. The difficulty was primarily due to uncertainty about the purity of the mRNA used as the hybridization probe as well as a lack of knowledge on the extent to which a probe will hybridize with the related but not identical genes, and the precise effect of sequence differences on hybridization kinetics. Thus it turned out to be nearly impossible to make a convincing interpretation of the data obtained in these early studies in relation to the issue of the evolutionary versus somatic generation of antibody diversity.

One subsequent series of experiments which I carried out on genes coding for the mouse λ light chains, however, was very encouraging (10). Using a mRNA preparation that was more than 95% pure, I could show that the mouse λ light chain gene is reiterated no more than the β globin gene. The latter gene had been shown to be essentially unique. Fortunately, Weigert, Cohn and their coworkers had identified at least eight different V_λ region sequences among BALB/c-derived myelomas (11). Since these V regions were highly homologous, differing by only one, two or three amino acid residues, it was very likely that the corresponding genes would crosshybridize extensively if they existed separately in the germline genome. Furthermore statistical analysis of λ light chain-secreting myelomas strongly suggested that a BALB/c mouse has the capacity to synthesize many more than the eight different V_λ regions identified. Thus the number of the mouse λ genes determined experimentally (no more than a few) was far smaller than the number of different V_λ regions (at least eight, most probably many more)

detected in proteins. On the basis of these results I was convinced that a somatic diversification occurs in this gene system.

REARRANGEMENT

In the meantime I became aware that some immunologists had been speculating that immunoglobulin polypeptide chains may be encoded by two separate DNA segments, one each for the *V* and *C* regions. Drawing an analogy from the elegant Campbell model (12) on the integration and excision of a phage λ genome, Dreyer and Bennett had further suggested that one of many "*V* genes" may be excised out from the original chromosomal position and joined with the single "*C* gene" in an immunoglobulin-producing B cell (13). This model successfully explained the maintenance of the common genetic marker in all immunoglobulin polypeptide chains of a given type by postulating a single *C* gene for that chain type. Although a somatic recombination between the "*V* and *C* genes" is an inherent aspect of the model, it is clearly a version of the germline theory of antibody diversity because the model assumed that the germline genome carries many "*V* genes", one for every *V* region that an organism can synthesize.

When the Dreyer and Bennett model was published in 1965, it was not accepted widely by biologists. This is understandable because the model was built on two hypotheses, both of which violated the then current dogmas of biology. These are the principles of one gene encoding one polypeptide chain and of the constancy of the genome during ontogeny and cell differentiation. My personal reaction to the model when I learned of it in the early nineteen seventies was also that of scepticism. However, at the same time I thought that the model might be testable if one were to use restriction enzymes. While in Dulbecco's laboratory I had heard of Daniel Nathan's breakthrough in the analysis of the SV40 genome by an application of the then newly discovered restriction enzymes (14). As one who used to struggle to define the transcriptional units of this DNA virus I was keenly aware of the power of these enzymes for the analysis of DNA structure. However, an extension of the restriction enzyme analysis from a viral genome of 5×10^3 base pairs to the 2×10^9 base pair genome of an eukaryote as complex as a mouse, required the use of an additional trick for the detection of a specific DNA fragment in a vast array of irrelevant fragments. An obvious solution seemed to lie in the combination of an electrophoretic separation of enzyme-digested DNA and the sensitive technique of nucleic acid hybridization. I discussed with Charlie Steinberg the need for developing a method that allows an *in situ* detection of a specific DNA sequence among the electrophoretically fractionated DNA fragments but we really could not come up with a good idea worthy of exploring. As we all now know, a very simple and elegant method ideal for this purpose was later developed by Edward Southern (15).

A few weeks passed by before I accidentally saw in one of the Institute's cold rooms a huge plexiglass tray in which someone was fractionating serum proteins by starch gel electrophoresis. I thought one may be able to fractionate a sufficient

amount of digested DNA in a gel of such dimensions, so that the DNA eluted from gel slices could be used for liquid phase hybridization. A quick calculation seemed to indicate that the experiment was feasible. Nobumichi Hozumi, a postdoctoral fellow in my laboratory, and I therefore decided to give it a try although we were keenly aware of the intense labor required by this type of experiment. As hybridization probes we used purified κ or λ light chain mRNA ($V + C$-probe) and its 3'-half fragment (C-probe) that had been iodinated to a high specific activity. The rationale of the experiment was as follows. First, if an immunoglobulin polypeptide chain is encoded by two "genes" V and C in the germline genome, it is highly probable that treatment with a restriction enzyme will separate these DNA sequences into fragments of distinct size, thus allowing their electrophoretic separation. Second, if a somatic rearrangement joins the V and C "genes" it is also highly probable that the myeloma DNA digested with the same restriction enzyme will contain a DNA fragment carrying both V and C "genes".

The results obtained were clear cut. To our pleasant surprise the patterns of hybridization of the embryo (a substitute of germline) DNA and a κ-myeloma DNA were not only drastically different but also perfectly consistent with the occurrence of separate V and C "genes" and a joined V plus C gene, respectively (16). We were of course aware of the alternative interpretations of the results, such as a fortuitious modification of the enzyme cleavage sites in one of the two types of DNA. However, we considered these alternative explanations of the results unlikely because they all required multiple fortuitous events. Our confidence was fortified soon afterwards as the development of Southern blot techniques allowed us to carry out more extensive analysis using a variety of restriction enzymes and myeloma cells.

JOINING OF GENE SEGMENTS

While the experiments with restriction enzymes were informative, details of the rearrangement were difficult to come by with this approach. Fortunately, recombinant DNA technology was just becoming available and was the ideal means for this purpose. Debates on the possible hazards of this type of research were flaring initially in the United States and shortly afterwards in European countries. In order to make sure that our research would not become a target of controversy, Charlie and I got in touch with Werner Arber at the University of Basel who was coordinating recombinant DNA research activities in Switzerland. A small informal work group was set up by the local researchers interested in this technique. The consensus of the group which was supported by most of the other Swiss researchers was that we should all follow the practices and guidelines being adopted in the United States. We met about once a month and exchanged information regarding both ethical and practical aspects of the technology.

On the basis of the previous experiments attempting to count immunoglobulin genes, I thought that it would be wise to start with the mouse λ light chain system, the simplest of all chain types that had been studied. Our goal was

to clone the V_λ and C_λ "genes" in the germline state from embryonic cells as well as the rearranged V plus C "genes" from a λ myeloma, and to determine the relationship between these genomic DNA clones by electronmicroscopy and DNA sequencing. No precedent existed at that time for cloning "unique" eukaryotic genes. We therefore had to devise a few tricks as we attempted to clone the first immunoglobulin gene. For instance, our available probe at that time was again 95% pure mRNA rather than a cDNA clone. This situation made the screening of a large number of DNA clones difficult because of the high background. To avoid this problem we pre-enriched the λ gene-containing genomic DNA fragments as much as possible using preparative R-loop formation (17, 18), so that the DNA library constructed would have the clone of interest at a high frequency.

Starting with the embryonic DNA we could isolate a clone that clearly hybridized specifically with the λ mRNA (18). When an electronmicroscopist, Christine Brack, who had just joined us from the Biozentrum of the University of Basel, examined the mixture of this clone and λ mRNA that had been annealed under an appropriate condition, she found a beautiful R-loop from which about a half of the mRNA strand protruded. This and additional analysis convinced us that we had cloned a V_λ "gene" to which no C "gene" was contigously attached, thus confirming at the DNA clone level that the V and C "genes" are indeed separate in the germline genome. A subsequent DNA sequencing study carried out in collaboration with Allan Maxam and Walter Gilbert of Harvard University revealed that this DNA clone corresponded to the V "gene" for the λ_2 subtype (19).

In the meantime Minoru Hirama, another postdoctoral fellow, succeeded in preparing λ and κ cDNA clones. Once these probes became available isolation of the genomic clones became much easier. My assistant Rita Schuller and I isolated a number of genomic DNA clones from λ and κ chain-synthesizing myelomas as well as from embryos (20, 21). Analysis of these DNA clones by electronmicroscopy, by restriction enzyme mapping, and by DNA sequencing not only confirmed the somatic rearrangement of immunoglobulin genes but also revealed some striking features of their arrangement and rearrangement (Fig. 1). These can be summarized as follows:

1. Although the V and C "genes" are rearranged and are much closer to each other in myeloma cells than in embryo cells, they are not contiguous and are separated by a few kilobases of DNA sequence that does not participate in coding of the polypeptide chain. This untranslated DNA sequence present within the rearranged, complete immunoglobulin gene was unanticipated and was also among the first demonstrations of an intron in eukaryotic genes (22).

2. The V "gene" found in the germline genome is about 13 codons short when it is compared to the length of the conventionally defined V region. The missing codons were found in a short stretch of DNA referred to as a J (or joining) gene segment that is located many kilobases away from the incomplete V "gene" (referred to as a V gene segment) and a few kilobases upstream of the C "gene" (also referred to as a C gene segment). In
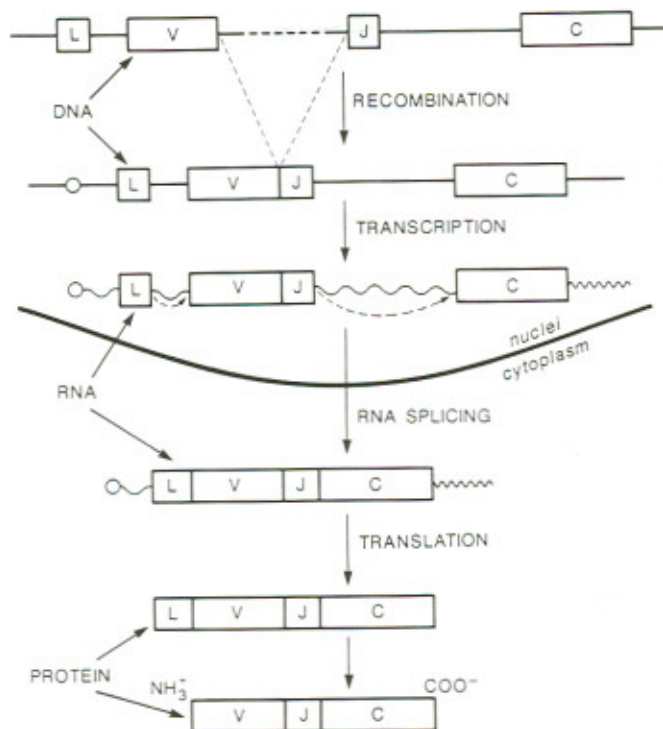
Fig. 1. The basic scheme for rearrangement and expression of an immunoglobulin light chain gene. At top is an arrangement of the gene segments on a germline genome. Somatic rearrangement links the *V* and *J* gene segment and generates a complete light chain gene shown just below the germline genome. The entire gene containing the leader exon (*L*), the *V* region exon (*V* and *J*), the *C* region exon (*C*), and the introns present between these exons are transcribed into a premRNA in the nuclei of the B cell. The premRNA is processed by RNA splicing as it is transported from the nuclei to the cytoplasm. The resulting mRNA devoid of introns is translated in the endoplasmic reticulum into a nascent polypeptide chain from which a mature λ light chain is generated after cleavage of the signal peptide.

myeloma cells the rearrangement event attaches the *J* gene segment to the *V* gene segment and thereby creates a complete *V* region "gene" (20, 23).

3. The signal peptide is encoded in yet another DNA segment referred to as the *L* (or leader) exon that is separated from the *V* gene segment by a short intron (19, 23).

Finding that the V_λ "gene" was split into two gene segments, V_λ and J_λ, in the germline genome was completely unexpected. But as soon as this discovery was made its implication for the somatic generation of antibody diversity was obvious. If the germline genome carries multiple copies of different *V* and *J* gene segments the number of complete *V* "genes" that can be generated by random joinings between these two types of gene segments would be much greater than the total number of the inherited gene segments. Thus, contrary to the Dreyer and Bennett original concept, DNA rearrangement can provide a major means for the somatic diversification of antibody molecules. The amino acid sequence

data of the κ light and heavy chains were consistent with this concept (24, 25). Indeed, the nucleotide sequence analysis of the mouse κ chain gene complex carried out both in my laboratory and in Phillip Leder's laboratory at The United States National Institutes of Health confirmed that a germline genome contains multiple V and J gene segments and that these gene segments are joined in different combinations in each myeloma cell (20, 26). Four different J_κ gene segments were found several kilobases upstream of the C_κ gene segment. The exact number of V_κ gene segments is unknown even today, but it is estimated to be two to three hundred (27).

HEAVY CHAIN GENES

Inasmuch as an immunoglobulin heavy chain is also composed of V and C regions, it was reasonable to expect that its gene also would undergo the type of DNA rearrangement described for the light chain genes. This supposition was confirmed by Leroy Hood and his coworkers at California Institute of Technology and by ourselves (Fig. 2) (28, 29). As in κ genes four J gene segments were found several kilobases upstream of the C gene segments coding for the C region of the μ class heavy chain. Multiple V gene segments were also identified.

While these features of the organization of heavy chain genes are essentially the same as those of the light chain genes, one observation made during these
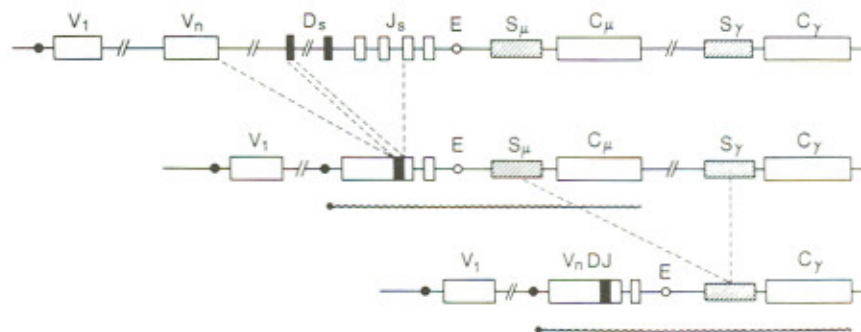
Fig. 2. Organization of the immunoglobulin heavy chain gene family. At top, middle, and bottom are organization in a germline genome, in a genome of B cells synthesizing a μ class heavy chain, and in a genome of a plasma cell synthesizing a γ class heavy chain, respectively. A mouse haploid genome carries several hundred different V gene segments, about a dozen D gene segments, four J gene segments, and one copy of C gene segment for each of the eight different classes or subclasses of immunoglobulin heavy chains. In a virgin B cell one copy each of the V, D, and J gene segment pools have been linked up and the joined VDJ DNA sequence is transcribed into a premRNA together with the $C\mu$ gene segment. In different B cells of the same organism a different set of V, D, and J gene segments are usually hooked up and expressed. As the virgin B cell differentiates either to a plasma cell or to a memory B cell (see Fig. 5) the second type of somatic recombination called "switch recombination" often occurs between a region (S_μ) located upstream of the $C\mu$ gene segment and another region (S_γ) located upstream of the C_γ gene segment. As shown at the bottom, the switch recombination replaces the C_μ gene segment with the C_γ gene segment without changing the VDJ exon. Filled circles designate transcriptional promotors present at the upstream of every V gene segment. The open circle designates the transcriptional enhancer (102, 103) which together with the promotor activate the rearranged heavy chain gene for a high level expression.

studies suggested that the somatic assembly of gene segments plays an even more prominent role in the diversification of heavy chains than of light chains. It was found that from one or two to a dozen amino acid codons that are present in the $V-J$ junction region of the assembled gene are not found in either of the apparently corresponding germline V or J gene segments (30, 31). This suggested that a third type of short gene segment referred to as D (or diversity) might participate in the somatic assembly of a heavy chain gene. Indeed, Hitoshi Sakano and Yoshi Kurosawa, two postdoctoral fellows in my laboratory, soon discovered about a dozen D gene segments (32, 33) which were subsequently mapped in a region upstream of the J cluster in the germline genome (34, 35). Thus the construction of a complete heavy chain V "gene" requires two DNA recombinational events, one joining a V with a D gene segment and the other the same D with a J gene segment.

RECOMBINATION RULE

The joining of $V-J$ or $V-D-J$ involves a site-specific recombination. It might therefore be expected that these gene segments would carry sequences in the vicinity of the joining ends that are recognized by a putative site-specific recombinase. Furthermore, such recognition sequences are likely to be common for all gene segments of a given type (e.g. V_κ's), because they all seem to be capable of joining with the common set of gene segments of the appropriate type (e.g. J_κ's). There indeed are a heptamer and a nonamer that are conserved in the region immediately downstream of each V_κ gene segment (Fig. 3) (36, 37). Sequences complementary to the V_κ heptamer and nonamer were also found in the region immediately upstream of each of the four J_κ gene segments. The same sets of sequences were also found in the corresponding regions of the V_λ and J_λ gene segments (36). When the heavy chain V and J gene segments were analyzed subsequently they too had the common conserved sequences (30, 31). Furthermore D gene segments carry the heptamer and nonamer sequences both upstream and downstream (32, 33). Another interesting feature of these putative recognition sequences is the fact that the length of the spacer between the heptamer and nonamer is either 12 or 23 base pairs (30, 31). Furthermore a gene segment carrying a recognition sequence with one type of spacer is able to join only with a gene segment with the spacer of the other type. This 12/23 base pair spacer rule seems to be adhered to strictly. Little is currently known about the recombinase but proteins with an affinity to the heptamer or nonamer have been identified in the extract of Abelson virus transformed pre B cell lines in which the rearrangement occurs *in vitro* at a relatively high frequency (38, 39).

DIVERSITY GENERATED AT THE JOINS

When the deduced amino acid sequence of a germline J_κ gene segment was compared with the determined amino acid sequences of those κ chains that are
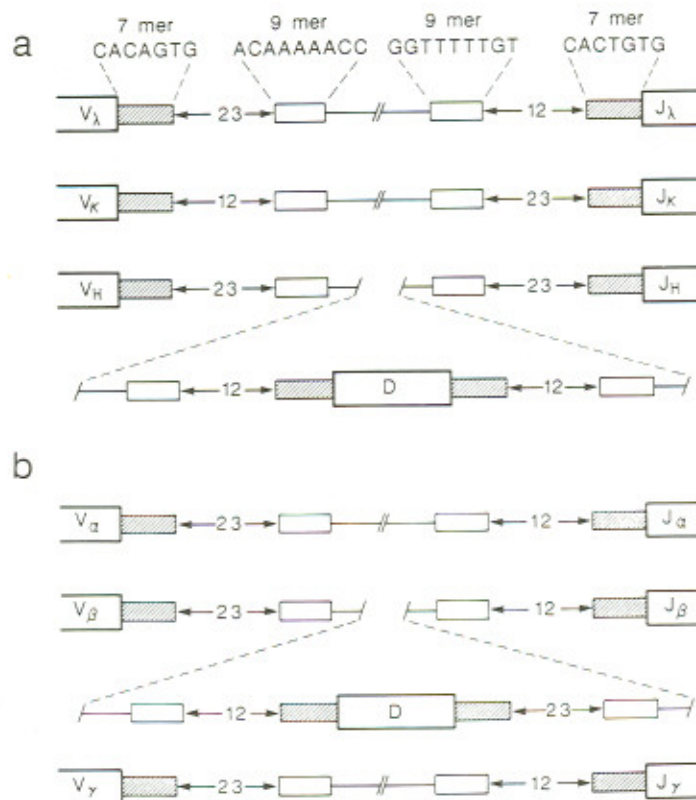
Fig. 3. Putative recognition sequences for the rearrangement of immunoglobulin and T cell receptor genes. The conserved heptamer and nonamer sequences and the length of the spacer between these sequences are schematically illustrated for immunoglobulin (Panel a) and for T cell receptor (Panel b) gene families. The sequences shown on top are consensus sequences. Individual sequences may deviate from these consensus sequences by a few nucleotides.

encoded in part by that J_κ gene segment, it was noticed that the 5' end of the J_κ gene segment is not prefixed but rather shifts toward upstream or downstream by several base pairs in different joining events (36, 37). This flexibility in the precise site of the joining was subsequently found to be a characteristic of the joining ends of other gene segments rather than of just J_κ-gene segments (31). It applies even when the same pair of gene segments were joined in different B cell percursors, such that the completed V "genes" are likely to have slightly different codons in the junction regions.

The $V-D$ and $D-J$ junctions exhibit diversity of yet another type. We found that up to a dozen base pairs of essentially random sequence are inserted in these junctions apparently without a template during the breakage and reunion of the recombining gene segments (32, 33). While the precise mechanism is yet unknown, the terminal deoxynucleotide transferase which is found in early B lymphatic nuclei or an enzyme with similar characteristics is thought to play a role in this phenomenon (40).

The part of the *V* region affected by the above two diversification mechanisms is limited. But this does not mean that they do not play a significant role in the determination of antibody specificity. On the contrary, the junctions encode the most variable two of the six loops of polypeptides that make up the antigen binding region of the antibody molecule (Fig. 4). Furthermore, specific cases are known where the affinity of an antibody to a defined antigen is drastically altered by a slight change in one junctional sequence (41). Thus the junctional variation also is a potent somatic generator of antibody diversity.
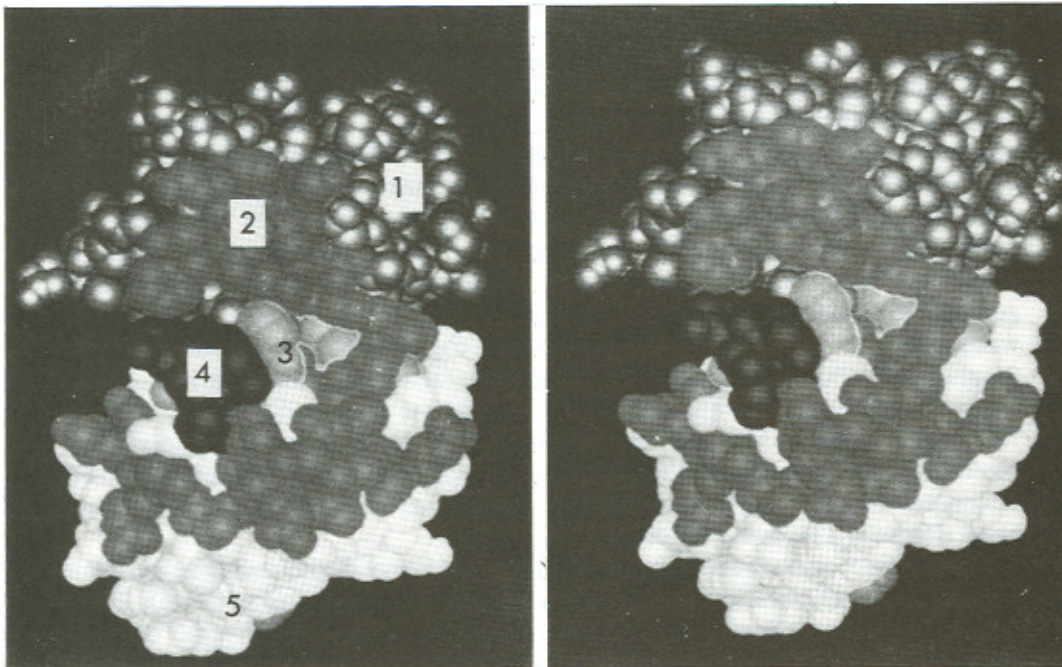
Fig. 4. Space-filling, stereo image of an antibody combining site. Atomic coordinates of mouse immunoglobulin MOPC 603 (104) were used to produce the picture. The heavy chain variable domain is color-coded dark grey, the light chain variable domain light grey. The hypervariable regions (except the VH third hypervariable region) are blue, the heavy chain segment coded for by the *D* gene is red, and the heavy and light chain segments coded for by the *J* genes are yellow. The *D* segment corresponds virtually exactly to the third heavy chain hypervariable region; hypervariable regions were defined as in Novotney *et al.* (105) except for the heavy chain second hypervariable region, which is marked as defined by Kabat *et al.* (25). The antigen of this particular immunoglobulin, phosphoryl choline, binds into the cavity in the middle of the picture in between the *VH* and *VL* domains, making contacts to amino acid residues belonging to the *VH* and *J* segments of the heavy chain and the *VL* segment of the light chain. Importance of the *D* segment is well illustrated in the two crystallographic structures of antibodies which bind the protein antigen lysozyme (106, 107). There, the contact area contributed by the *D* segment amount to 50% and 24%, respectively, of the total heavy chain contact area. This image was computer-generated by Jiri Novotny using the program SPHERE of Robert Bruccoleri. Because of reproduction difficulties, a black and white version of the original colour photograph is shown: 1 = dark grey (heavy chain variable domain); 2 = blue (hypervariable regions); 3 = orange (heavy and light chain *J* regions); 4 = red (heavy chain *D* region; 5 = light grey (light chain variable domain).

SOMATIC MUTATION

When F. Macfarlane Burnet proposed the clonal selection theory he recognized the need for some kind of random genetic process in order to generate antibodies able to bind specifically to the vast variety of antigens (42). He considered somatic mutations as the most plausible mechanism. Subsequently, this idea was adopted and forcefully presented by many including Joshua Lederberg, Niels Kaj Jerne, and Melvin Cohn (5, 6, 7)

The amino acid sequence data accumulated by Martin Weigert in Melvin Cohn's laboratory at the Salk Institute provided an excellent opportunity to examine directly the role of somatic mutations in antibody diversity (7, 11). They had analyzed the λ_1 light chains derived from eighteen myelomas. All the mice were of an inbred strain BALB/c and so should have been genetically identical. They found that twelve of the $V_{\lambda 1}$ regions were identical but that the other six differed both from the majority sequence and from one another by only one, two, or three amino acid residues. They proposed that BALB/c mice may carry only one germline $V_{\lambda 1}$ "gene" which codes for the majority sequence, and that all the other $V_{\lambda 1}$ regions observed are encoded by somatic mutants of this single $V_{\lambda 1}$ "gene" that arose in B cell development. As I already mentioned in an earlier section our gene-counting experiment by hybridization kinetics suggested that the germline BALB/c genome carries no more than a few $V_{\lambda 1}$ "genes". This number was reduced to one when we reevaluated the copy number by the more reliable Southern blotting method (20). The final proof of somatic mutation in $V_{\lambda 1}$ came when we cloned and sequenced the sole germline $V_{\lambda 1}$ gene segment and the rearranged λ_1 genes expressed in a myeloma (23). As Weigert and Cohn guessed the nucleotide sequence of the germline $V_{\lambda 1}$ gene segment corresponded to the major amino acid sequence, while the λ_1 gene expressed in the myeloma had been altered by single base changes.

Since this work several subsets of κ light and heavy chains and their germline V gene segments have been analyzed by cloning and sequencing (43–46). These results have all confirmed that somatic mutations further amplify the diversity encoded in the germline genome. Particularly revealing was the analysis carried out by Patricia J. Gearhart, Leroy Hood and their coworkers for the V_H regions associated with the binding of phosphorylcholine (PC). They demonstrated that single base changes can be extensive and yet are restricted to the joined VDJ sequences and the immediately adjacent regions (47, 48).

DEVELOPMENTAL CONTROL OF REARRANGEMENT AND HYPERMUTATION

Why have two extraordinary somatic genetic mechanisms, recombination and hypermutation, evolved in the immune system in order to carry out what appears to be one task, namely to diversify antibodies?

I believe that the answer may be the differential roles of these two genetic mechanisms. Thanks to the efforts of several independent groups of cellular and
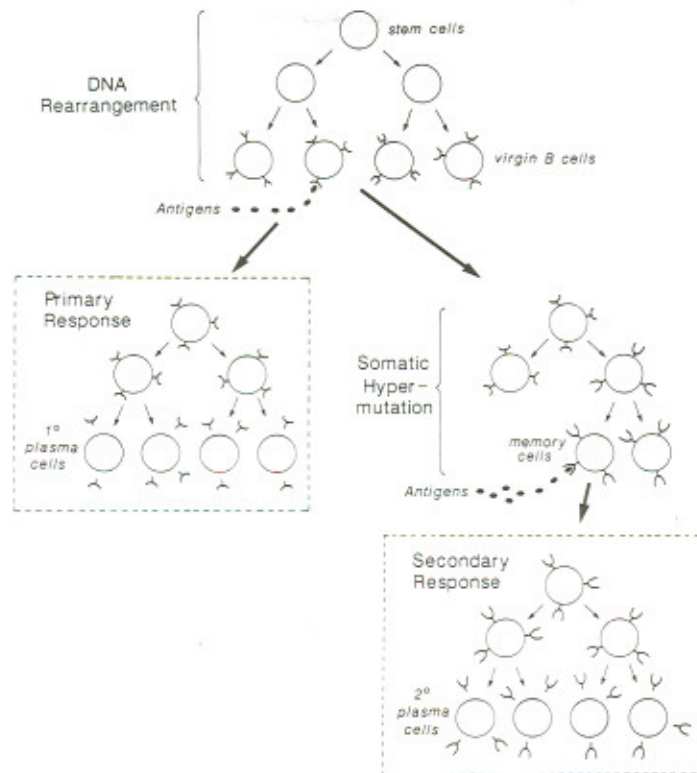
Fig. 5. Differentiation of B cells. Note that the receptors present on the memory cells and the antibody molecules secreted by the plasma cells of the secondary response have a tighter fit to the antigen than the receptors on the ancestral virgin B cells or the antibodies secreted by the plasma cells of the primary response. See text for the full explanation.

molecular immunologists a general picture is emerging that describes the relationship between the states of B cell development and the occurrence of somatic recombination or mutation (Fig. 5) (49–55). Somatic recombinations contributing to diversity are initiated first for the heavy chain and then for the light chain during the differentiation of progenitor cells, and the completion of somatic recombination is accompanied by the appearance of virgin B cells (56–58). These B cells form clones each of which is composed of cells bearing homogenous IgM molecules as surface receptors. Thus, somatic recombination is completed *prior to* any possible interaction of a B cell with antigens.

When an antigen enters the lymphatic system for the first time, it will be screened by these virgin B cells. The small fraction of these B cells that happen to have sufficient affinity for the antigenic determinants in question will respond and follow either of two pathways: they will produce the primary antibody response, or they will contribute to the generation of memory B cells. In the former pathway, the selected B cells will proliferate and differentiate into antibody-secreting plasma cells. During this process, the C region of the heavy chain can

switch from μ to another class, but mutation is rare in either the heavy or the light chain V region. Consequently, the antibodies secreted by plasma cells in the primary response would largely have the same V regions as the immunoglobulin receptors on the virgin B cells from which they derive.

By contrast, immunoglobulin remains in the cell surface receptor form during the other pathway taken by the antigen-activated virgin B cells, namely the generation of memory B cells. During this process the hypermutation apparatus appears to be most active and the rate of the mutation approaches 10^{-3} base substitution per cell per generation. Antigen selects, in a stepwise fashion, better and better fitting mutants so that the immunoglobulins on the surface of memory B cells achieve a substantially higher affinity than the immunoglobulins on the ancestral virgin B cells. Switch recombination also occurs frequently during this process too. When the same antigen as the one that elicited the primary response reenters the body, the memory B cells are selectively propagated and differentiate into plasma cells. This is the so-called secondary antibody response, which, therefore, consists of high affinity antibody of "mature" isotype and these antibodies show extensive somatic mutation in their V regions. Somatic mutations appear to cease after memory cells are generated, and little or no further mutation takes place during the secondary antibody response.

This scheme of B cell differentiation can be rephrased as follows. An organism is prepared for infection with pathogens bearing virtually any antigens with a large variety of resting B cells. These B cells bear unique immunoglobulin receptors encoded by one copy each of complete light and heavy chain genes that have been constructed by a random or quasirandom assembly of the inherited gene segments. Since the assembly occurs independent of antigens and since the inherited gene segments are not usually selected during evolution for precise fit to particular antigens, the antibody secreted by the plasma cells derived directly from the selected resting virgin B cells during a primary antibody response usually have a relatively low affinity. By contrast, the frequent single base changes that occur during the generation of memory B cells provide the organism with a great variety of finely altered immunoglobulin receptors from which only those with the best fit to the antigen in question will be selected. Since the plasma cells generated during the secondary antibody response are mostly direct descendants of these memory B cells having no further alterations in the antigen-combining sites, these antibodies usually exhibit a much higher affinity for antigen than do primary antibodies. This explains the long known phenomenon of affinity maturation of antibodies during the course of repeated immunizations (59).

Thus somatic creation of antibody genes can be viewed as a two step process. In the first step, blocks of gene segments are employed to build, in an antigen-independent fashion, a set of genes coding for antibodies of great diversity but with low affinity. In the second step, once the antigen is defined, a small selected set of B cells bearing low affinity antibodies as cell surface receptors undergo somatic mutations with the result that a fraction of them develop a higher affinity to that antigen and can be selected for further expansion. This process improves the ability of the immune system to detect a low

2

concentration of antigens. One wonders what happens to those cells in which mutation did not improve affinity. A recent study suggests that at least some of these cells may be set aside for selection by different antigens (54). Thus somatic mutation may also contribute to the repertoire of receptors specific for antigens not previously introduced into an immune system.

T CELL RECEPTORS

As the mystery of the genetic origins of antibody diversity was unravelled at least in its basic outlines, it seemed natural to extend our research to "the other half" of the lymphoid system, namely T cells. Although we often discussed the idea of research on the nature of antigen recognition by T cells in the laboratory in the late seventies while I was still in Basel, the real work did not start until the early eighties in my new laboratory at MIT. Although T cells were known to recognize and distinguish antigens as precisely as B cells, nothing was known about the biochemical nature of the molecules responsible for this task, namely T cell receptors (TCR). This lack of information was in stark contrast to the wealth of information about antibodies. Much debate took place among cellular immunologists on the nature of these molecules. Some argued that T cell receptors are just another class of immunoglobulins. Others thought T cell receptors would have to be quite different from immunoglobulins. Indeed studies carried out in the late seventies had shown that the way in which a T cell recognizes an antigen is quite different from the way a B cell does: the T cell reacts to antigens on a cell surface, and the T cell receptor simultaneously recognizes both an antigen and a determinant present on a glycoprotein encoded by a gene in the major histocompatibility complex (MHC) (60–62). This discovery raised another issue: does a T cell recognize two determinants with one receptor, or does it have two receptors, one for antigen and the other for a MHC product?

The receptor protein was first detected in 1983 in experiments carried out by three independent groups of scientists headed by James P. Allison, Ellis L. Reinherz, and Phillipa Marrack and John Kappler (63–65). They prepared antibodies that bind to a protein on the T cell surface: since these proteins were similar but exhibited clonally distributed structural diversity, they were thought to be a good candidate for the receptor. Furthermore, the antibodies they prepared were T cell clone-specific and they could show that these antibodies blocked activation of the T cell clone in a clone-specific fashion. The receptor identified by these experiments was composed of two polypeptide subunits, designated α and β, that are held together by a disulfide bond. These studies were critical in the sense that the receptor was finally identified, its overall structure defined, and its predicted structural variability confirmed. However, the paucity of the protein on the T cell surface and the absence of the secreted form of the receptor made it very difficult to obtain further information about the structure of this molecule, especially its amino acid sequence.

α AND β GENES

In the meantime, molecular biologists were attempting to identify the genes coding for the T cell receptor. This turned out to be a much more difficult endeavor than the cloning of immunoglobulin genes. Although T cell lines and hybridomas expressing a homogeneous receptor were becoming available, these cells were more difficult to grow than myelomas and the amount of the receptor made was at least two orders of magnitude less than the amount of immunoglobulin produced by a myeloma cell. In 1984, Mark Davis and his coworkers at Stanford University and Tak Mak and his coworkers at the University of Toronto independently made a breakthrough (66–68). Their experimental strategy depended on two assumptions. First that mRNAs coding for the α and β polypeptide chains are present in a T cell hybridoma or T cell tumor, but are absent in B cell tumors. And second, that the α and β chain genes are rearranged in T cells in a manner similar to the immunoglobulin genes in B cells. Thus, they made a library from the fraction of T cell-cDNA that did not hybridize with B cell-derived mRNA and tested each T cell-specific cDNA clone for rearrangement of the corresponding gene in T cells. As the source of T cells, Davis' group used a hybridoma obtained by fusing a mouse helper T cell recognizing antigen plus self class II MHC molecules with a T cell tumor, while Mak's group used a human T cell tumor. The two groups came up with one common class of cDNA clones that satisfied the above criteria. The nucleotide sequence showed that the corresponding polypeptide chain is significantly (30–35%) homologous to immunoglobulin chains. Furthermore, the cDNA clones contained sequences homologous to V and C regions in the correct orientation. Thus it seemed certain that the gene represented by this class of cDNA clones codes for one of the two subunits of the T cell receptor. That this gene encodes the β subunit was soon confirmed by determination of the partial amino acid sequence of the human β chain (69).

In my laboratory at MIT Haruo Saito and I collaborated with David Kranz and Herman Eisen to isolate both α and β cDNA clones from another type of T cell, namely a cytotoxic T cell clone specific for class I MHC molecules. In 1984, using a modified subtractive cDNA library method we identified two classes of cDNA clones that also satisfied the criteria for a T cell receptor gene (70). One class of these clones represented the β subunit. Taken together with the earlier finding by Davis' group this demonstrated that the two major classes of T cells, helper T cells and cytotoxic T cells, employ the same set of genes at least for the β subunit. The same conclusion was drawn subsequently for the α subunit. This point is significant because the two types of T cell are specific for two distinct subclasses of MHC gene products. Thus, the same T cell receptor genes mediate recognition of both class I and class II MHC.

The polypeptide chains encoded by the other class of rearranging T cell-specific cDNAs isolated by Saito and myself was also homologous to immunoglobulin chains by 30–35%. These cDNA clones were, however, clearly distinct from β cDNA clones because the polypeptide chains encoded by the two sets of cDNAs were homologous by only 30–35% not only in the V regions but

also in the C regions. Since only two subunits, α and β, were known for the T cell receptor, we initially proposed that this second class of cDNAs represented the α gene (70). However, even before the work was published, a question arose about the assignment of this gene as encoding the α chain. The putative α cDNAs do not carry codons for N-linked glycosylation sites, while unpublished studies from Charlie Janeway's laboratory at Yale University and Jim Allison at Berkeley indicated that both α and β subunits of at least some T cells carry N-linked carbohydrates. While it was still possible that the apparent discrepancy in glycosylation could be explained by differences in the type of t cell or in mouse strains used, continued screening of our subtracted cDNA library yielded within a few weeks a third class of clones whose genes also rearrange specifically in T cells (71). This gene not only was as homologous to immunoglobulin genes as the first two classes of T cell-specific genes, but also had two potential sites for N-linked glycosylation, and therefore was a better candidate for the α gene. This proposition was soon confirmed by comparing its nucleotide sequence with the partial amino acid sequence of the human α subunit (72). Furthermore the α gene was also cloned at about the same time from a helper hybridoma (73).

Once cDNA encoding the α and β chains were identified, it was straightforward to determine the organization of the corresponding genes in genomic DNA. These studies demonstrated that both α and β genes are organized in the germline genome and rearranged in T cells in a way remarkably similar to the immunoglobulin genes (74–77). Thus, the organism inherits the genetic information for these polypeptide chains as V_α and J_α gene segments or V_β, D_β, and J_β gene segments, and a random assembly of these gene segments occurs exclusively during T cell development to generate a diversity comparable to that of immunoglobulins for receptors expressed on the surface of mature T cells. Even the presumed recognition sequences for the site specific recombinase, the so-called heptamers and nonamers with a 12 or 23 base pair spacer, seem to be common for both the immunoglobulin and T cell receptor genes (Fig. 3).

The complete primary structure of a T cell receptor can be deduced from the nucleotide sequences of the α and β cDNA clones. Its comparison with the primary structure of an immunoglobulin molecule suggests that the external part of the receptor is composed of four compact, immunoglobulin-like globular domains associated in two non-covalently bound pairs V_α–V_β and C_α–C_β, and further stabilized by an interchain disulfide bond between the C domain and the transmembrane region. This extracellular part of the receptor is anchored on the membrane lipid bilayer through two transmembrane peptides, one each from the α and β chains (Fig. 6) (70).

Determining the structure and organization of genes encoding the T cell receptors settled the issue of their relationship with immunoglobulins and accounted for the genetic origins of their diversity. However these studies did not illuminate the mechanism by which these receptors can accomplish the dual recognition of an antigen and a MHC determinant. This last issue is particularly tantalizing because recent studies using a technique for injecting T cell receptor genes into cloned, functional T cells confirmed that the $\alpha\beta$ heterodimer alone is sufficient to mediate this dual specificity (78). In order to understand how the
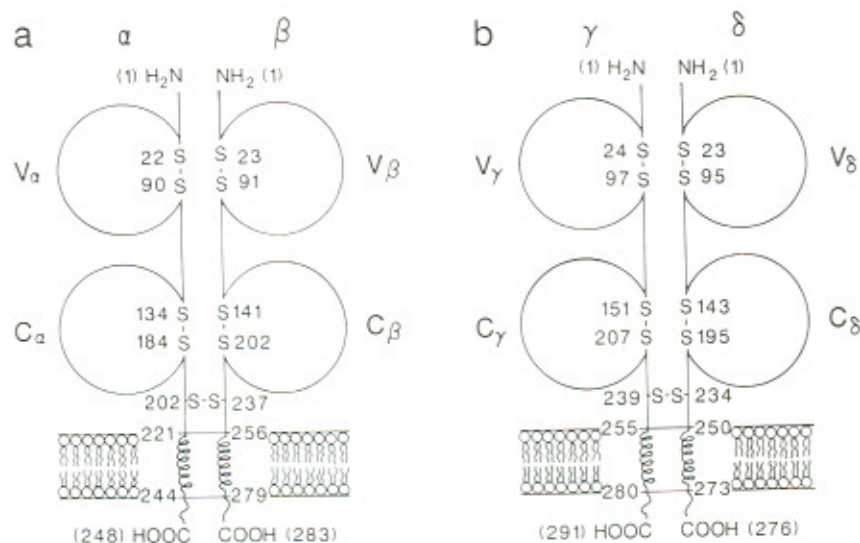
Fig. 6. Diagram showing the subunit structures of T cell receptors $\alpha\beta$ (panel a) and $\gamma\delta$ (panel b) as deduced from the nucleotide sequences of cDNA clones. The $\alpha\beta$ receptor is from an alloreactive cytotoxic mouse T cell clone, 2C and the $\gamma\delta$ receptor from a mouse thymocyte hybridoma, KN6, prepared by Osami Kanagawa of Lilly Research Laboratories, La Jolla, California. Intra- and inter-chain disulfide bonds are indicated. The receptors are thought to be anchored on the membrane lipid bilayer by transmembrane peptides. The invariant CD3 complex associated with the heterodimers are not shown.

heterodimer simultaneously recognizes the two determinants much more information is needed as to the structure of the receptor and of its compound ligand. It seems that the ultimate solution has to come from X-ray crystallographic analysis of the receptor protein.

A NEW T CELL RECEPTOR, $\gamma\delta$

Since it was established that the third T cell-specific rearranging gene discovered was for the α subunit, the second one becomes an orphan. This gene is so closely related to the other two genes, however, that it seems certain that it must have some role in recognition by T cells. Nevertheless, previous immunochemical studies did not reveal any polypeptide chain that might be considered as a candidate for the protein product of this gene. The γ gene is also assembled somatically from V, J, and C gene segments and shares a number of characteristics with the α and β genes as well as with immunoglobulin genes (79, 80).

A number of possibilities were considered initially as firsthand roles for the putative γ chain. For instance, it was thought that the γ chain may be a subunit for a second T cell receptor coexpressed with the $\alpha\beta$ heterodimer. This hypothesis is in line with the two receptor model of dual recognition of antigen

and MHC by T cells. Another possibility proposed was that there may be a switch in the subunits of the T cell receptor during T cell development. A model was proposed in which a T cell receptor composed initially of a $\gamma\beta$ heterodimer switches to an $\alpha\beta$ heterodimer as T cells differentiate in the thymus (81, 82). This model, which seemed to be supported by the time course kinetics of appearance of α-, β-, and γ-specific RNA in developing thymocytes, was an attempt to explain an apparent dilemma in the intrathymic selection of the T cell repertoire (for instance see ref. 83).

Subsequent studies carried out in my and several other laboratories, however, revealed a number of features of the γ gene and its expression which are not necessarily consistent with these hypotheses. First, the γ gene is not rearranged in some T cell clones or hybridomas. Furthermore, even in many of those T cells in which this gene is rearranged, the joining of the V and J gene segments does not allow the J region codons to be translated in phase with the V region codons (84–86). Thus, the γ gene product does not seem to be universally expressed in conventional, $\alpha\beta$ receptor-positive cytotoxic and helper T cells. Second, the γ polypeptide chain is expressed on the surface of a small (less than 0.5%) subset of peripheral T cells as a component of a heterodimer referred to as $\gamma\delta$ (87–89). The majority of these T cells exhibit none of the CD4 or CD8 glycoproteins of conventional $\alpha\beta$ receptor T cells on their surface, and therefore belong to a distinct cell population. Third, like the $\alpha\beta$ heterodimer, the $\gamma\delta$ heterodimer is associated relatively tightly with another glycoprotein, CD3 (87). The invariant CD3 protein complex contains a subunit that seems to play a critical role in the transmission of the signal received by the variable heterodimer into the cell (90). Thus, the similarity between the $\gamma\delta$- and $\alpha\beta$-receptor heterodimer includes both their structure and transmembrane signal transmission. Fourth, $\gamma\delta$-bearing cells are relatively abundant in the CD4$^-$, CD8$^-$ fraction of fetal and adult thymocytes (91–94). For instance, thymocytes of 16-day old fetal mice which are mostly double negative (i.e. CD4$^-$ and CD8$^-$) cells are a relatively rich source of $\gamma\delta$-bearing cells. Since double negative thymocyte populations contain precursors for mature, functional, $\alpha\beta$-bearing T cells (95), a natural question that arises is whether $\gamma\delta$-bearing thymocytes are precursors for $\alpha\beta$-bearing T cells. Fifth, another major site of expression of the $\gamma\delta$-receptor is epidermal tissues. It was recently found by two groups that this tissue contains Thy-1$^+$ (another cell surface marker shared by all types of T cells), CD3$^+$, CD4$^-$ and CD8$^-$ cells bearing $\gamma\delta$-heterodimers (96, 97). Unlike conventional T cells, these cells look more like dendritic cells and are therefore referred to as dendritic epidermal cells (DEC). Finally, the δ gene also undergoes rearrangement. D, J, and C gene segments for δ polypeptide chains have recently been mapped within the α gene family between V_α and J_α gene segments (98). The nested configuration of α and δ gene segments is intriguing and provokes curiosity about the possible relationship of gene organization with the regulation of the rearrangement and expression of the two types of genes. Another intriguing question is whether and to what extent the pool of V_α and V_δ gene segments overlap.

Despite the rapid progress made in the characterization of the γ and δ genes

and their products, the most intriguing problem, namely the physiological role of the $\gamma\delta$ receptor-bearing cells is currently unknown. One can only speculate on this matter using the currently available information. As to the ligand of this new receptor, it is probably correct to emphasize the fact that the receptor shares with the immunoglobulin B cell and the $\alpha\beta$ T cell receptors the same genetic basis for somatic diversification. It therefore is likely that the ligand in question will also exhibit structural diversity. In this respect it is interesting that recent studies by us and others suggest that at least part of the ligand is encoded in MHC (99, 100). The effector function of the cells bearing $\gamma\delta$ receptors has not yet been defined with certainty, but recent studies using human and mouse $\gamma\delta$ cell clones suggest that many of these cells have cytotoxic capacity (89, 100, 101). The finding of a new type of T cell with an apparently distinct set of T cell receptors in epidermal tissues stimulates one's imagination. It may be that occurrence of this type of T cell is not restricted to the outer epithelial tissues but extends to all epithelial layers including the lining of various internal organs (C. Janeway, personal communication). If so, these cells may have evolved to protect the part of the body that is most vulnerable to infection, namely its external and internal epithelial surfaces that are in direct contact with the environment. However, the prominence of $\gamma\delta$ cells in the thymus suggests an intrathymic role of these cells as well. An exciting possibility that has not been ruled out is a role for these cells in the intrathymic selection of appropriate $\alpha\beta$ T cells.

CONCLUDING REMARKS

Use of restriction enzymes and recombinant DNA methods allowed resolution of a long-standing and central issue in immunology, the genetic origins of antibody diversity. It turned out that an organism does not inherit even a single complete gene for antibody polypeptide chains. Rather, the genetic information is transmitted in germline as no more than several hundred gene segments. Through a series of specialized somatic recombinations occurring specifically during the differentiation of B lymphocytes, these gene segments are assembled into tens of thousands of complete genes. Somatic hypermutation occurring in these assembled genes further diversifies antibody polypeptide chains, so that B cells displaying immunoglobulin receptors having a better fit to a given antigen can be selected in a later phase of B cell differentiation. Thus in the immune system, organisms have exploited two major processes for modification of DNA, recombination and mutation, as a means to diversify somatically the limited amount of inherited genetic information in order to cope with the vastly diverse antigen universe.

Why has somatic diversification been necessary in the evolution of the immune system? Microorganisms and substances produced by them are the primary source of biologically relevant antigens against which vertebrates need to produce antibodies for survival. Since the generation time of microorganisms is several orders of magnitude shorter than that of vertebrates, the former can produce genetic variants much faster than the latter. Thus if genetic alterations in

the germline genome were to be the only source of antibody diversity, vertebrates would be unable to deal effectively with the rapidly changing world of antigens. Somatic diversification allows the individual organism to generate a virtually limitless number of lymphocyte variants. Like organisms in an ecosystem, these lymphocytes are subject to selection by antigens and the fittest will survive. Thus, as Jerne and Burnet were aware, the individual immune system can be conceived of as a kind of Darwinian microcosm.

The molecular biological approach played an even more fundamental role in the analysis of the T cell receptor in that very little structural information existed prior to the cloning of the receptor genes. It was demonstrated that the polypeptide chains composing the receptor protein are encoded by genes that share a common ancestor with the immunoglobulin genes. Like immunoglobulins, T cell receptors are diversified by somatic recombination, but unlike immunoglobulins, these receptor proteins have not been observed to undergo further diversification by somatic mutation. The reason for this difference is unknown, but it seems likely that the explanation will be as follows. First, unlike immunoglobulins, T cell receptors function exclusively as cell surface receptors which are specialized for interacting with cell-bound antigens. Since both ligand and receptor are distributed in two dimensional space when a T cell interacts with an antigen-presenting cell, and as T cells have mechanisms for transiently adhering to other cells, these receptor-ligand interactions occur under conditions of high local concentration. Thus, improvement of the affinity beyond the one accomplished by somatic rearrangement may not be necessary in T cell recognition.

Second, the ligand consists in part of an essentially invariant component, self MHC. As the T cell receptor is selected both during development and during immunization for self MHC recognition, the extreme variability available via somatic mutation may not only be unnecessary but even disadvantageous. Finally, and probably more important, T cells appear to be selected early in development for self tolerance, the inability to recognize self antigen. Somatic mutation during antigenic stimulation, as occurs in B lymphocytes, could lead to the development of autoreactivity. While autoreactive B cells depend upon the additional presence of autoreactive helper T cells to generate autoimmunity, this is not true of autoreactive T cells, which can directly produce disease. Thus, Ehrlich's famous concept of "honor autotoxins", originally developed for antibodies, is probably critical only for T cells. It will be interesting to examine autoreactive T cell receptors for any evidence of post-thymic somatic diversification.

Finally, it is interesting to notice that during the fifteen years in which I have studied the immune system, the role of molecular genetics in immunological research has altered radically. When I started investigating the problem of antibody diversity, there was abundant information about the structure and function of antibody molecules, while virtually nothing was known about their genes. By contrast, in the most recent study on T cell recognition no gene product was known at all when the rearranging gene, γ was discovered. From the structure of the gene and its rearranging behavior, it was deduced to be a receptor gene, and this discovery has led directly to new insights into T cell

development and T cell biology. This short history of research in one area, lymphocyte receptors, is yet another witness to the power of DNA technology, and to the ability of this approach not only to explain known biological phenomena, but also to contribute to the discovery of new biological systems.

ACKNOWLEDGEMENT

REFERENCES

1. Porter, R. R. (1973). *Science* **180**:713–716.
2. Edelman, G. M. (1973). *Science* **180**:830–840.
3. Hilschmann, N. and Craig, L. C. (1965). *Proc. Natl. Acad. Sci. USA* **53**:1403–1409.
4. Hood, L. and Talmage, D. W. (1970). *Science* **168**:325–334.
5. Lederberg, J. (1959). *Science* **129**:1649–1653.
6. Jerne, N. K. (1971). *Eur. J. Immuno.* **1**:1–14.
7. Cohn, M., Blomberg, B., Geckeler, W., Raschka, W., Riblet, R. and Weigert, M. (1974). in *"The Immune System: Genes, Receptors, Signals"* (E. E. Sercarz, A. R. Williamson and C. F. Fox, eds.), Academic Press, New York and London, p. 89.
8. Gelderman, A. H., Rake, A. V. and Britten, R. J. (1971). *Proc. Natl. Acad. Sci. USA* **68**:172–176.
9. Bishop, J. O. (1971). *Biochem. J.* **126**:171–185.
10. Tonegawa, S., (1976). *Proc. Natl. Acad. Sci. USA* **73**:203–207.
11. Weigert, M., Cesari, M., Yonkovich, S. J. and Cohn, M. (1970). *Nature* **228**:1045–1047.
12. Campbell, A., (1962). *Advan. Genet.* **11**:101.
13. Dryer, W. J. and Bennett, J. (1965). *Proc. Natl. Acad. Sci. USA* **54**:864–869.
14. Danna, J. J., Sack, Jr., G. H. and Nathans, D. (1973). *J. Mol. Biol.* **78**:363.
15. Southern, E. (1975). *J. Mol. Biol.* **98**:503.
16. Hozumi, N. and Tonegawa, S. (1976). *Proc. Natl. Acad. Sci., USA* **73**:3628–3632.
17. Thomas, M., White, R. L. and Davis, R. N. (1976). *Proc. Natl. Acad. Sci. USA* **73**:2294–2298.
18. Tonegawa, S., Brack, C., Hozumi, N. and Schuller, R. (1977). *Proc. Natl. Acad. Sci. USA* **74**:3518–3522.
19. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. and Gilbert, W. (1978). *Proc. Natl. Acad. Sci. USA* **75**:1485–1489.
20. Brack, C., Hirami, M., Lenhard-Schuller, R. and Tonegawa, S. (1978). *Cell*, **15**:1–14.
21. Lenhard-Schuller, R., Hohn, B., Brack, C., Hirama, M. and Tonegawa, S. (1978). *Proc. Natl. Acad. Sci. USA* **75**:4709–4713.
22. Brack, C. and Tonegawa, S. (1977). *Proc. Natl. Acad. Sci. USA* **74**:5652–5656.
23. Bernard, O., Hozumi, N. and Tonegawa, S. (1978). *Cell* **15**:1133–1144.
24. Weigert, M., Gatmaitan, L., Loh, E., Schilling, J. and Hood, L. (1978). *Nature* **276**, 785–789.
25. Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M. and Perry, H. (1983). U.S. Dept. of Health and Human Services Publication.
26. Seidman, J. G., Leder, A., Nau, M., Norman, B. and Leder, P. (1978). *Science* **202**:11–16.
27. Cory, S., Tyler, B. M. and Adams, J. M. (1981). *J. Mol. Appl. Genet.* **1**:103–116.
28. Davis, M., Calame, K., Early, P., Livant, D., Joho, R., Weisman, I. and Hood, L. (1980). *Nature* **283**:733–738.

29. Maki, R., Traunecker, A., Sakano, H., Roeder, W. and Tonegawa, S. (1980). *Proc. Natl. Acad. Sci. USA* **77**:2138–2142.
30. Early, P., Huang, H., Davis, M., Calame, K. and Hood, L. (1980). *Cell* **19**:981–992.
31. Sakano, H., Maki, R., Kurasawa, Y., Roeder, W. and Tonegawa, S. (1980). *Nature* **286**:676–683.
32. Sakano, H., Kurosawa, Y., Weigert, M. and Tonegawa, S. (1981). *Nature* **290**, 562–565.
33. Kurosawa, Y., von Boehmer, H., Haas, W., Sakano, H., Traunecker, A. and Tonegawa, S. (1981). *Nature* **290**:565–570.
34. Kurosawa, Y. and Tonewaga, S. (1982). *J. Exp. Med.* **155**:201–218.
35. Wood, C. and Tonegawa, S. (1983). *Proc. Natl. Acad. Sci. USA* **80**:3030–3034.
36. Sakano, H., Huppi, K., Heinrich, G. and Tonegawa, S. (1979). *Nature* **280**:288–294.
37. Max, E. E., Seidman, J. G. and Leder, P. (1979). *Proc. Natl. Acad. Sci. USA* **76**:3450–3454.
38. Halligan, B. D. and Desiderio, S. V. (1987). *Proc. Natl. Acad. Sci. USA* **84**:7019–7023.
39. Aguitera, R. J., Akira, S., Okagaki, K. and Sakano, H. *Cell* (in press).
40. Alt, F. and Baltimore, D. (1982). *Proc. Natl. Acad. Sci. USA* **79**:4118–4122.
41. Azuma, T., Igras, V., Reilly, E. and Eisen, H. N. (1984). *Proc. Natl. Acad. Sci. USA* **81**:6139–6143.
42. Burnet, F. M. (1959). *The Clonal Selection Theory of Acquired Immunity.* London: Cambridge Univ.
43. Bothwell, A. L. M., Paskind, M., Reth, M., Imanishi-Kari, T., Rawesky, K. and Baltimore, D. (1981). *Cell* **24**:624.
44. Crews, S., Griffin, J., Huang, H., Calame, K. and Hood, L. (1981). *Cell* **25**:59–66.
45. Givol, D., Zakut, R., Effron, K., Rechavi, G., Ram, D. and Cohen, J. B. (1981). *Nature* **292**:426–430.
46. Selsing, E. and Storb, U., (1981). *Nucl. Acids Res.* **9**:5725–5735.
47. Gearhart, P. J., Johnson, N. D., Douglas, R. and Hood, L. (1981). *Nature* **291**:29–33.
48. Kim, S., Davis, M., Sinn, E., Patten, P. and Hood, L. (1981). *Cell* **27**:573–580.
49. Askonas, B. A. and Williamson, A. R. (1972). *Eur. J. Immunol.* **2**:487–493.
50. Berek, C., Griffiths, G. M. and Milstein, C. (1985). *Nature* **316**:412–418.
51. McKean, D., Hüppi, K., Bell, M., Standt, L., Gerhard, W. and Weigert, M. (1984). *Proc. Natl. Acad. Sci. USA* **81**:3180–3184.
52. Okumura, K., Julius, M. H., Tsu, T., Herzenberg, L. A. and Herzenberg, L. A. (1976). *Eur. J. Immunol.* **6**:467–472.
53. Sablitzberg, M., Kocks, C. and Rajewsky, K. (1985). *EMBO J.* **4**:345–350.
54. Sickevitz, M., Kocks, C., Rajewsky, K. and Dildrop, R. (1987). *Cell* **48**:757–770.
55. Wysocki, L. J., Manser, T. and Gefter, M. L. (1986). *Proc. Natl. Acad. Sci. USA* **83**:1847–1851.
56. Maki, R., Kearney, J., Paige, C. and Tonegawa, S. (1980). *Science* **209**:1366–1369.
57. Perry, R. P., Kelley, D. E., Coleclough, C. and Kearney, J. F. (1981). *Proc. Natl. Acad. Sci. USA* **78**:247–251.
58. Reth, M. G., Ammirati, P. A., Jackson, S. J. and Alt, F. W. (1985). *Nature* **317**:353–354.
59. Eisen, H. N. and Siskind, G. W. (1964). *Biochemistry* **3**:996–1000.
60. Rosenthal, A. S. and Shevach, E. M. (1973). *J. Exp. Med.* **138**:1194–1212.
61. Katz, D. H., Hamaoka, T., Dorf, M. E. and Benacerraf, B. (1973). *Proc. Natl. Acad. Sci. USA* **70**:2624–2628.
62. Zinkernagel, R. M. and Doherty, P. C. (1975). *J. Exp. Med.* **141**:1427–1436.
63. Allison, J. P., McIntyre, B. W. and Bloch, D. (1982). *J. Immunol.* **129**:2293–2300.
64. Meuer, S. C., Acuto, O., Hussey, R. E., Hodgdon, J. C., Fitzgerald, K. A., Schlossman, S. F. and Reinherz, E. L. (1983). *Nature* **303**:808–810.
65. Haskins, K., Kubo, R., White, J., Pigeon, M., Kappler, J. and Marrack, P. (1983). *J. Exp. Med.* **157**:1149–1169.
66. Hedrick, S. M., Cohen, D. I., Nielsen, E. A. and Davis, M. M. (1984). *Nature* **308**:149–153.
67. Hedrick, S. M., Nielsen, E. A., Kavaler, J., Cohen, D. I. and Davis, M. M. (1984). *Nature* **308**:153–158.
68. Yanagi, Y., Yoshikai, Y., Leggett, K., Clark, S. P., Aleksander, I. and Mak, T. W. (1984). *Nature* **308**:145–149.
69. Acuto, O., Fabbi, M., Smart, J., Poole, C., Protentis, J., Royer, H. D., Schlossman, S. and Reinherz, E. L. (1984). *Proc. Natl. Acad. Sci. USA* **81**:3851–3855.
70. Saito, H., Kranz, D. M., Takagaki, Y., Hayday, A. C., Eisen, H. N. and Tonegawa, S. (1984). *Nature* **309**:757–762.
71. Saito, H., Kranz, D. M., Takagaki, Y., Hayday, A. C., Eisen, H. and Tonegawa, S. (1984). *Nature* **312**:36–39.

72. Hannum, T. H., Kappler, J. W., Trowbridge, I. S., Marrack, P. and Freed, J. H. (1984). *Nature* **312**:65–67.
73. Chien, Y-h., Becker, D. M., Lindsten, T., Okamura, M., Cohen, D. J. and Davis, M. M. (1984). *Nature* **312**:31–35.
74. Chien, Y-h., Gascoigne, N., Kavaler, J., Lee, N. E. and Davis, M. M. (1984). *Nature* **309**:322–326.
75. Siu, G., Kronenberg, M., Strauss, E., Haars, R., Mak, T. W. and Hood, L. (1984). *Nature* **311**:344–350.
76. Hayday, A. C., Diamond, D. J., Tanigawa, G., Heilig, J. S., Folsom, V., Saito, H. and Tonegawa, S. (1985). *Nature* **316**:828–832.
77. Winoto, A., Mjolsness, S. and Hood, L. (1985). *Nature* **316**:832–836.
78. Dembić, Z., Haas, W., Weiss, S., McCubray, J., Kiefer, H., von Boehmer, H. and Steinmetz, M. (1986). *Nature* **320**:232–238.
79. Kranz, D. M., Saito, H., Heller, M., Takagaki, Y., Haas, W., Eisen, H. N. and Tonegawa, S. (1985). *Nature* **313**:752–755.
80. Hayday, A. C., Saito, H., Gillies, S. D., Kranz, D. M., Tanigawa, G., Eisen, H. N. and Tonegawa, S. (1985). *Cell* **40**:259–269.
81. Raulet, D. H., Garman, R. D., Saito, H. and Tonegawa, S. (1985). *Nature* **314**:103–107.
82. Pernis, B. and Axel, R. (1985). *Cell* **41**:13–16.
83. Schwartz, R. (1984). in *Fundamental Immunology* (ed. William Paul), Raven Press, NY. pp. 379–438.
84. Reilly, E. B., Kranz, D. M., Tonegawa, S. and Eisen, H. N. (1986). *Nature* **321**:878–880.
85. Rupp, F., Frech, G., Hengartner, H., Zinkernagel, R. M. and Joho, R. (1986). *Nature* **321**:876–878.
86. Heilig, J. S. and Tonegawa, S. (1987). *Proc. Natl. Acad. Sci. USA* **84**:8070–8074.
87. Brenner, M. B., McLean, J., Dialynas, D. P., Strominger, J. L., Smith, J. A., Owen, F. L., Seidman, J. G., Ip, S., Rosen, F. and Krangel, M. S. (1986). *Nature* **322**:145–149.
88. Lanier, L. L., Federspiel, N. A., Ruitenberg, J. J., Phillips, J. H., Allison, J. P., Littman, D. and Weiss, A. (1987). *J. Exp. Med.* **165**:1076–1094.
89. Borst, J., Van de Griend, R. J., van Oostveen, J. W., Ang, S-L., Melief, C. J., Siedman, J. G. and Bolhuis, R. L. H. (1987). *Nature* **325**:683–688.
90. Wauve, J. P. van, Mey, J. R. de and Goosens, J. G. (1980). *J. Immunol.* **124**:2708–2812.
91. Bank, I., DePinho, R. A., Brenner, M. B., Cassimeris, J., Alt, F. W. and Chess, L. (1986). *Nature* **322**:179–81.
92. Lew, A. M., Pardoll, D. M., Maloy, W. L., Fowlkes, B. J., Kruisbeck, A., Cheng, S-F., Germain, R. N., Bluestone, J. A., Schwartz, R. H. and Coligen, J. E. (1986). *Science* **234**:1401–1405.
93. Nakanishi, N., Maeda, K., Ito, K., Heller, M. and Tonegawa, S. (1987). *Nature* **325**:720–723.
94. Pardoll, D. M., Fowlkes, B. J., Bluestone, J. A., Kruisbeck, A., Maloy, W. L., Coligan, J. E. and Schwartz, R. H. (1987). *Nature* **326**:79–81.
95. Fowlkes, B. J. (1984). Thesis, George Washington University Graduate School of Arts and Sciences, Washington DC.
96. Koning, F., Stingl, G., Yokoyama, W. M., Maloy, W. L., Tschachler, E., Shevach, E. M. and Coligan, J. E. (1987). *Science* **236**:834–837.
97. Kuziel, W. A., Takashima, A., Bonyhadi, M., Bergstresser, P. R., Allison, J. P., Tigelaar, R. E. and Tucker, P. W. (1987). *Nature* **328**:263–266.
98. Chien, Y-h., Iwashima, M., Kaplan, K. B., Elliott, J. F. and Davis, M. M. (1987). *Nature* **327**:677–682.
99. Maeda, K., Nakanishi, N., Rogers, B. L., Haser, W. G., Shitara, K., Yoshida, H., Takagaki, Y., Augustin, A. A. and Tonegawa, S. (1987). *Proc. Natl. Acad. Sci. USA* **84**:6536–6540.
100. Matis, L. A., Cron, R. and Bluestone, J. A. (1987). *Nature* **330**:262–264.
101. Moingeon, P., Jitsukawa, S., Faure, F., Troalen, F., Triebel, F., Graziani, M., Forestier, F., Bellet, D., Bohunon, C. and Hercend, T. (1987). *Nature* **325**:723–726.
102. Gillies, S. D., Morrison, S. L., Oi, V. T. and Tonegawa, S. (1983). *Cell* **33**:717–728.
103. Banerji, J., Olson, L. and Schaffner, W. (1983). *Cell* **33**:729–739.
104. Satow, Y., Cohen, G. H., Padlan, E. A. and Davies, D. R. (1986). *J. Mol. Biol.* **190**:593–604.
105. Novotny, J., Bruccoleri, R. E., Newell, J., Murphy, D., Haber, E. and Karplus, M. (1983). *J. Biol. Chem.* **258**:14433–14437.
106. Amit, A. G., Mariuzza, R. A., Phillips, S. E. V. and Poljak, R. (1986). *Science* **233**:747–752.
107. Sheriff, S., Silverton, E. W., Padlan, E. A., Cohen, G. H., Smith-Gill, S. J., Finzel, B. C. and Davies, D. R. (1987). *Proc. Natl. Acad. Sci. USA* **84**:8075–8079.