SOMATIC RECOMBINATION AND MOSAIC STRUCTURE OF IMMUNOGLOBULIN GENES*

SUSUMU TONEGAWA

Basel Institute for Immunology, Basel, Switzerland

I. Antibody Gene Paradox

One of the most distinctive characteristics of immunoglobulin molecules is seen in the extent of the diversity within a single organism; it is estimated that an individual mouse has the capacity to synthesize one to a hundred million different immunoglobulins. In the past, the myeloma proteins helped greatly in the elucidation of the structural basis of this enormous diversity. Out of these studies a surprising structural feature of immunoglobulins emerged by the middle of the 1960s. Namely, both the light and heavy chains are composed of two distinctive regions, the amino terminal variable (V) and the carboxyl terminal constant (C) regions (1,2). While the constant region specifies about a dozen types or classes of chains, the variable regions are extremely heterogeneous in the primary structure. This latter heterogeneity accounts for the major proportion of immunoglobulin diversity. Given these results, it seemed logical to some to believe that immunoglobulin chains are encoded in thousands of different structural genes. However, this rather straightforward interpretation of the structural results failed to explain the genetic results, namely, that the serologically identifiable phenotypes in some of the C regions (allotypes) segregate as if they were under the control of a single gene (3).

In order to resolve the apparent paradox, Dreyer and Bennett proposed in 1965 that the V and C regions are encoded in two separate DNA segments and that one of the multiple V-coding DNA segments is combined with the single-copy C-coding DNA segment in immunoglobulin-producing cells (4). This hypothesis

*Lecture delivered January 17, 1980.

ISBN 0-12-312075-6

seemed to resolve the apparent paradox described above and, in addition, to provide the structural basis for the "one lymphocyte clone–one antibody specificity" principle required by Burnet's clonal selection theory (5). However, the hypothesis violated two basic dogmas that had been widely accepted by biologists. First, the idea that the V and C regions are encoded in two separate DNA segments was in direct conflict with the "one gene–one polypeptide chain" dogma formulated by Beadle and Tatum and supported by numerous, subsequent molecular and biological studies on the structure and expression of procaryotic genes. Second, the proposal that the V- and C-coding DNA sequences are rearranged in lymphocytes violated the concept of genome constancy in development and differentiation, which was supported by Gurdon's nuclear transfer experiments (6). It was generally believed that differentiation is a manifestation of differential expression of genes whose configuration in the genome is rigidly kept constant throughout ontogeny. It is thus not surprising that, when proposed, the Dreyer and Bennett hypothesis did not appeal to many biologists except for those immunologists who were confronted with the apparent paradox and were struggling to resolve it. While in 1970 the hypothesis was extended and modified by Gally and Edelman to incorporate the problem of the genetic origin of antibody diversity (7), the lack of direct evidence prevented the theory from being widely accepted, particularly by molecular biologists.

II. New Approach I—Restriction Enzymes

Discovery of restriction enzymes in the late 1960s and the subsequent development of numerous nucleic acid techniques, particularly the *in vitro* DNA recombinant method, opened entirely new and direct approaches to problems associated with the complex higher cell genomes, such as those of immunoglobulin genes. In order to answer the question of whether DNA sequences are reorganized during development, one of course needs a way to study arrangement of these genes in the genome. Restriction enzymes are very useful for this purpose because these enzymes recognize and cleave DNA at the sites charac-

terized by specific short sequences. When total DNA isolated from a mammalian genome is digested completely with a restriction enzyme, it gives a mixture of millions of DNA fragments of a fixed composition that can be fractionated by size by agarose gel electrophoresis. The DNA fragment carrying a specific gene can be identified by a method developed by Edward Southern (8). Briefly, the DNA fragments in the agarose gel are denatured with alkali and transferred *in situ* to a piece of nitrocellulose filter which retains the single-stranded DNA. The filter is incubated with a radioactive mRNA corresponding to the gene in question or a complementary DNA biochemically synthesized from the mRNA, washed extensively, and exposed to an X-ray film. The cellular DNA fragment containing the gene in question is detected in the autoradiogram by virtue of its hybridization with the mRNA or the complementary DNA. We reasoned simple-mindedly that if immunoglobulin gene sequences are rearranged during differentiation, the distribution of the flanking enzyme cleavage sites would be different before and after the rearrangement and consequently the gene would be detected on fragments of different sizes when embryo DNA and DNA of proper myeloma cells are compared.

Mouse immunoglobulins are under the control of three unlinked sets of genes: genes for κ-type light chains, genes for λ-type light chains, and genes for heavy chains. The λ chain genes are further divided into two subsets, one for the λ_I subtype and the other for the λ_{II} subtype. A characteristic feature of mouse λ chains is that they constitute less than 5% of the total light chains. Accordingly, the V region diversity of mouse λ chains is much more limited than those of the κ chains or the heavy chains. We took advantage of this situation and initially concentrated on the hopefully simpler λ chain gene system for elucidation of the basic principles and then went on to study presumably more complex κ and heavy chain genes for various specific phenomena unique to these gene systems.

Figure 1 shows the autoradiogram of the so-called Southern gel blot of three types of DNAs digested by a restriction enzyme, *Eco*RI (9). The hybridization probe used is that of the λ_I-type light chain gene. The embryo DNA gave three bands. Further

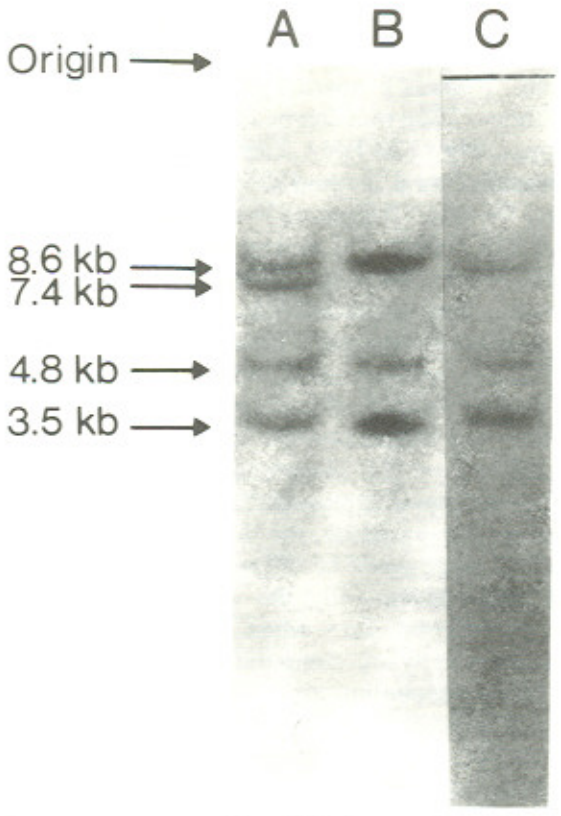

FIG. 1. λ_I gene sequence containing DNA fragments in embryo and myeloma cells. High-molecular-weight DNAs extracted from 13-day-old BALB/c embryos (B), myelomas H2020 (a λ_I chain producer) (A), and MOPC321 (a κ chain producer) (C) were digested to completion with *Eco*RI, electrophoresed on a 0.9% agarose gel, transferred to nitrocellulose membrane filters, and hybridized with a nick-translated *Hha*I fragment of the plasmid B1 DNA.

studies using hybridization probes containing the V sequence alone or the C sequence alone established that the 8.6-kb fragment contains only the $C\lambda_I$ gene sequence, while the 3.5-kb fragment contains only the $V\lambda_I$ gene sequence. The 4.8-kb fragment contains a $V\lambda_{II}$ gene sequence and was detected by virtue of its cross-hybridization with the $V\lambda_I$ gene probe. The DNA of

the λ_I chain-secreting myeloma, H2020, gave a fourth band of 7.4 kb, in addition to the three bands common to those of embryo DNA. This fragment contains both the $V\lambda_I$ and $C\lambda_I$ gene sequences. Myeloma MOPC321 secretes a κ chain and no λ-type light chain. The pattern of the DNA is essentially identical with that of embryo. The patterns of other nonlymphatic tissues such as liver, kidney, brain, and muscle, were all embryo-type (data not shown). These results suggested that the $V\lambda_I$ and $C\lambda_I$ sequences were widely separated in the embryo genome, while in the λ_I myeloma cells the two sequences are on a single *Eco*RI DNA fragment in one of the two homologous chromosomes on which the λ_I gene is located. This change in organization of the λ_I gene seems to be a consequence of a specific recombination event that occurred during differentiation of lymphocyte precursor cells.

III. New Approach II—Gene Cloning

In order to confirm the implications suggested by this type of experiment we cloned each of the four DNA fragments detected by the λ_I gene probe by the *in vitro* recombination method. I will concentrate on the characteristics of the three λ_I gene clones, namely, DNA clones of the embryonic V gene, the embryonic C gene, and the myeloma V + C gene. The position of the respective gene sequences in these DNA clones was determined by examining under an electron microscope the R-loop structures formed by a cloned DNA fragment and purified λ_I chain mRNA. The embryonic V clone gave a small R-loop corresponding to the V-coding part of the mRNA and a single-strand tail composed of the C-coding part of the mRNA (Fig. 2A). The embryonic C gene clone gave a double-loop structure. The smaller loop is an R-loop and contains the C gene sequence (Fig. 2B), while the larger loop is composed of a 1.2-kb-long, double-stranded DNA that separates the C gene sequence away from a short sequence complementary to the V–C junction region of the mRNA. This sequence, which we refer to as the *J DNA segment* or the *J sequence*, is present in the mRNA and forms an independent R-loop that is too small to be visualized under the microscope. The single-stranded tail protruding from the contact between the two

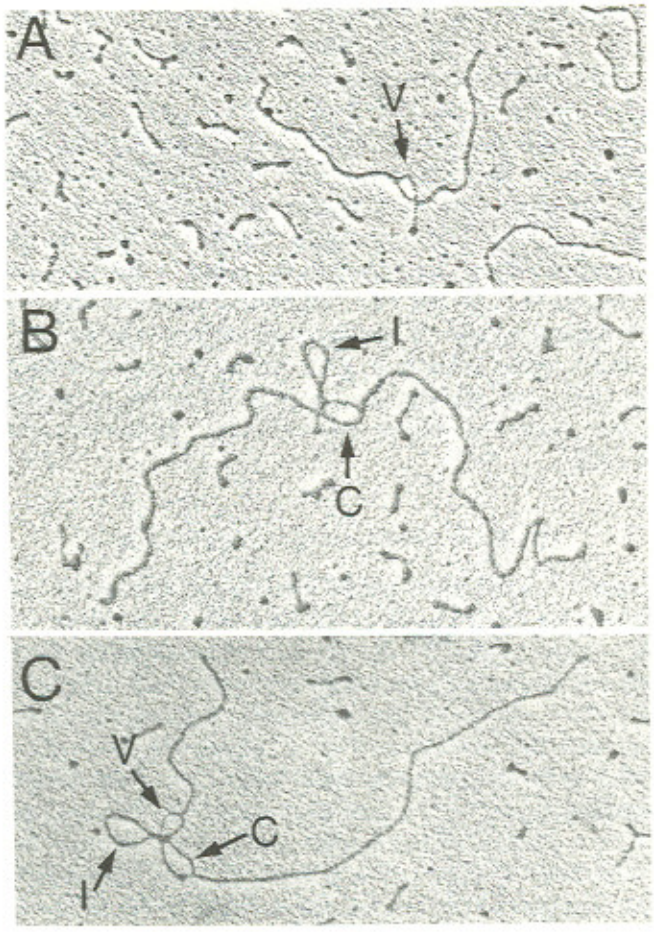

FIG. 2. R-loop structures of three λ_I-chain gene clones. (A) Embryonic $V\lambda_I$ clone, Ig99λ. (B) Embryonic $C\lambda_I$ clone, Ig25λ. (C) Myeloma H2020 $V\lambda_I$ + Cλ clone, Ig303λ. V and C indicate the respective R-loops. I indicates the 1.2-k. intron which appears as a double-stranded DNA loop.

loops contains the V-coding part of the mRNA. The myeloma clone gave the triple-loop structure: the V gene R-loop, C gene R-loop, and again a 1.2-kb-long, double-stranded DNA loop (Fig. 2C).

Heteroduplex analysis of these three λ_1 gene clones indicated that one end of the myeloma clone is entirely homologous to the embryonic V gene clone up to the position of the V gene, while the other end of the myeloma clone is homologous to the embryonic C gene clone starting at the position of the J DNA segment. We characterized the three λ_1 gene clones further by DNA sequencing and came up with the configuration of λ_1 chain genes illustrated in Fig. 3 (9). In embryo cells the hydrophobic signal peptide (residues -19 to -5) is encoded in a DNA segment separated from the DNA segment coding for the major body of the V region (residues -4 to 96). At an unknown distance away from the two DNA segments a short DNA segment coding for the V–C junction region (residues 97 to 110), and therefore referred to as a J DNA segment, is present and another DNA segment coding for the C region (residues 111 to 214) lies

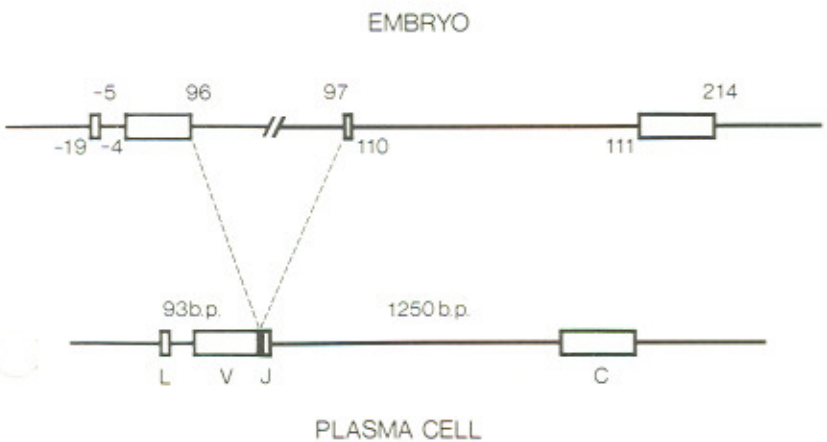

FIG. 3. Organization of the mouse λ_1 light chain gene in embryo (top) and λ_1 chain-secreting myeloma cells (bottom). The numbers indicate the encoded amino acid residues. The boxes represent coding sequences. Adapted from Brack *et al.* (9).

1.25 kb downstream of the J DNA segment. In the λ_1 chain-synthesizing plasma cells, the V and J DNA segments are fused to form a continuous sequence that codes for the entire V region as defined by protein sequence analysis. Except for a few base changes in the V DNA segment, which is attributed to an independent somatic event, no other difference in the sequence organization has been observed between the embryo and the plasma cell DNA. In the plasma cell the entire region is transcribed and the transcript is processed to generate a mature mRNA. We refer to the protein-encoding sequences as *exons* (10) and the transcribed, but non-protein-encoding sequences between exons as *introns* (10).

Figure 4 summarizes the organization of κ-type light chain gene sequences in embryo and κ chain-synthesizing myelomas. Unlike in the λ chain gene where both the V and J sequences

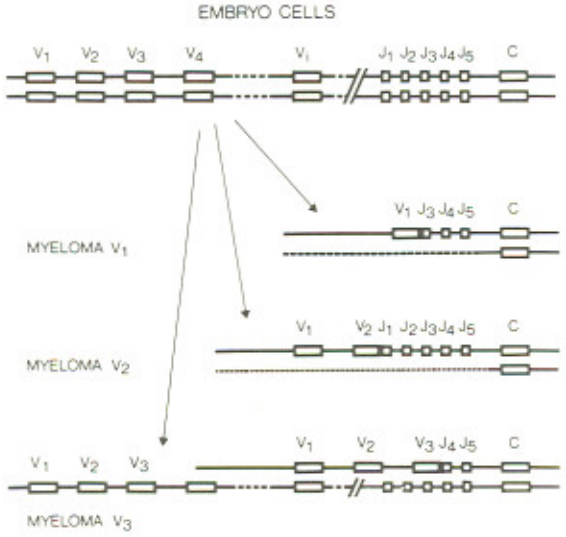

FIG. 4. Organization of mouse κ light chain genes in embryo and κ chain-secreting myeloma cells. Unlike the λ_1 chain gene case, the DNA sequence is often rearranged in the homologous chromosome in the vicinity of the Cκ gene (myelomas V_1 and V_2). But this is not always the case, as illustrated in myeloma V_3.

are unique, there are a few hundred different Vκ DNA copies clustered (11,12) and five different copies of Jκ DNA segments also clustered and located a few kilobases upstream of the unique Cκ DNA (13,14). The essential features of the somatic rearrangement of κ chain genes are analogous to those of the λ chain genes, namely, one of the multiple Vκ DNA segments joins with one of the five Jκ DNA segments and consequently an active transcription unit is formed. Those J DNA segments remaining between the J DNA segment selected for joining with a V sequence and the C DNA segment are apparently cotranscribed and spliced out during RNA processing.

One additional feature of rearrangement which was noticed in the κ chain gene system is a second type of recombination which is present in many, but not all, κ chain-synthesizing myelomas (12). The recombination occurs in the 5′ proximity of the Cκ DNA segment and involves a series of sequences with as yet unidentified locations in the genome. Neither V nor J DNA segments participate in this type II recombination (P. Kennedy, unpublished results). Since in many κ myelomas the unique Cκ DNA segments are either on type I or type II recombinants, we postulate that the two types of recombinants represent the two homologous chromosomes as illustrated in Fig. 4. While we do not yet know the physiological meaning, if any, of the type II recombination, one possible interpretation is that the type II recombination is an active process by which creation of an active, complete gene by type I recombination is prevented in one of the two homologs. It has been known for some time that, unlike all other known autosomal genes, only one of the two homologous chromosomes is expressed for Ig gene loci in a given lymphocyte, a phenomenon called *allelic exclusion* (15).

IV. Mechanism of Somatic Recombination in Immunoglobulin Genes

In order to study the fine structure of the V–J recombination sites, we determined the nucleotide sequences of the J cluster. In so doing we observed two stretches of short sequences that are conserved in the 5′-flanking regions of the five Jκ DNA

segments (13). One conserved sequence lies immediately before the J DNA segment and is composed of a heptameric palindromic sequence CACTGTG. The other conserved sequence is further upstream and is composed of a T-rich nonamer, GGTTTTTGT, or its closely related variant (Table I). These conserved sequences are invertedly repeated in the 3′-noncoding regions of a variety of embryonic V-coding DNA segments (Table I). Because of this, any given pair of an embryonic V DNA segment and a J DNA segment could form a stem-loop structure near the recombination sites. An example of such a structure is shown in Fig. 5. Based on these results and a series of Southern gel blot experiments described elsewhere (13), we believe that the V–J recombination occurs by looping out and deletion of the entire DNA segment separating the V and the J DNA segments. Curiously enough, the inverted repeat structures shown here have some resemblance to those found at the ends of prokaryotic insertion elements. For instance, Ohtsubo and Grindley determined sequences of IS1 elements and found that some 30 bases at one end are invertedly repeated on the other end (16,17). Furthermore, a run of Ts surrounded by Gs is also present in the stem of IS1 as well as in the equivalent stems of all other prokaryotic insertion elements. Although these structural similarities may be coincidental, they could reflect a common evolutionary origin or a common enzymatic mechanism for the two recombination systems, or both.

A striking feature of these conserved sequences is the regularity in the length of the spacer between the heptamer and the nonamer. As summarized in Table I, in almost all cases the spacers are 12 or 23, plus or minus one, base pairs long. Since a DNA double-helix completes a turn every 10.4 base pairs, the foregoing indicates that the internal boundaries of the heptamer and the nonamer of a given DNA segment are oriented in near the same direction relative to the axis of the helix, regardless of the length of the spacer. These features of DNA sequences around the V–J joining sites suggest that the κ and λ chain gene recombinations are mediated by the same or very similar enzymes and that the recombinase or recombinases contain two binding proteins: one recognizing the heptamer and nonamer one

TABLE I

TWO CONSERVED SEQUENCES NEAR THE V–J JOINING SITES[a]

	V		Spacer (bp)		J		Spacer (bp)	
κ	Vκ21C	CACAGTG	11	ACAAAAACC	Jκ1	GGTTTTTGT	23	CACTGTG
	Vκ21B	–––––––	12	–––––––––	Jκ2	A–––––––	23	––G––––
	Vκ41	–––––––	12	–––T–––––	Jκ3	––G–––––	21	––––––A
	Vκ2	–––––––	12	–––T–––––	Jκ4	–––––––––	24	–––––––
					Jκ5	–––––––––	23	–––––––
λ	$V\lambda_1$	–––A–––	22	T–––G–––A	$J\lambda_1$	––––––––C	12	–––A–––

[a] The sequences were taken from the following references: Vκ21C (13), Vκ41 (40), Vκ2 (41), Jκ1 to Jκ2, (13,14), $V\lambda_1$ and $J\lambda_1$ (42). Vκ21B is an unpublished observation of G. Heinrich.

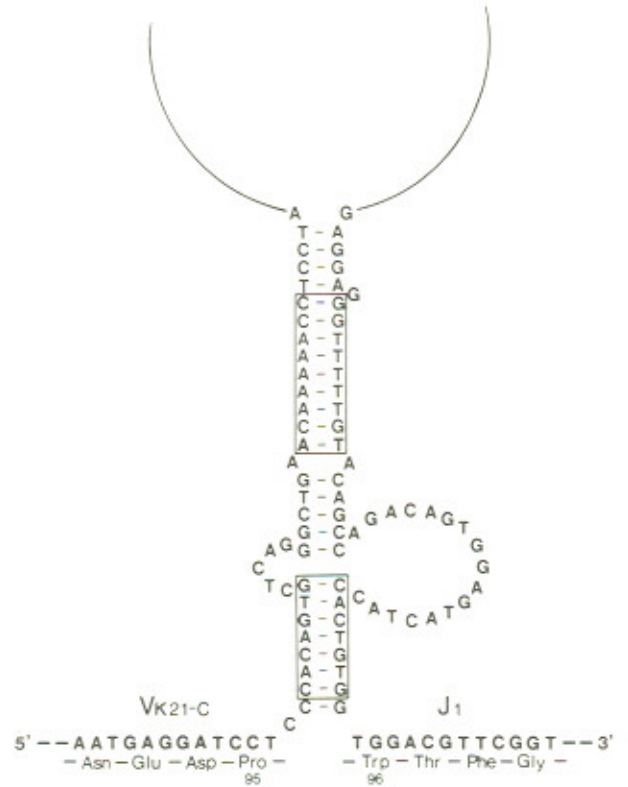

FIG. 5. An example of inverted repeat stem-loop structures that could be formed between 3′ noncoding regions of embryonic V DNAs and 5′ flanking regions of J DNAs. The conserved heptamer and nonamer

$$\begin{matrix} CACTGTG \\ GTGACAC \end{matrix} \text{ and } \begin{matrix} GGTTTTTGT \\ CCAAAAACA \end{matrix}$$

are surrounded by boxes. The numbers at the bottom indicate the amino acid positions.

turn apart, and the other two turns apart. The two proteins hold the two recombining partners (V and J) together and cut and rejoin the strands in the vicinity of the heptamers. Since the recognition sequences of one partner are complementary to those of the other, it is possible that intrastrand base pairing (see Fig. 5) occurs, however transiently, in the enzyme–DNA complex

and facilitates the ligation reaction. Recently, P. Early and L. Hood and their co-workers independently came up with a similar model of the recombinase based on the results given in Table I, in addition to their own results of heavy chain genes (P. Early and L. Hood, personal communication).

The available protein sequences around the V–J joining sites suggest that the recombinase can cut and join the DNA strands of a given pair of V and J at slightly different positions so that different triplet codons are reconstituted at the recombination site. For instance, in the Vκ21-C and J_1 pair shown in Fig. 5, the residues 95 and 96 would be Pro-Trp, if the recombination occurs in one of the two ways indicated by 1 and 2 in Fig. 6. In contrast, recombinations 3 and 4 generate Pro(95)-Arg(96) and Pro(95)-Pro(96), respectively. The latter two ways of recombination can account for the J peptides of four known κ chains synthesized by MOPC173, PC6684, and MPC11 and PC7940, respectively (13). Note that none of the five J DNA segments identified thus far can by itself encode Arg or Pro at position 96. The residue 96 is at the margin of one of the three subregions in the V region that are directly involved in determination of complementary region to the antigen. It thus could contribute to the size and shape of the antigen-combining site. If the above hypothesis is correct, the antibody diversity can be amplified by V–J joinings in two different ways, namely by various combinations of the V and J DNA segments and by modulation of the

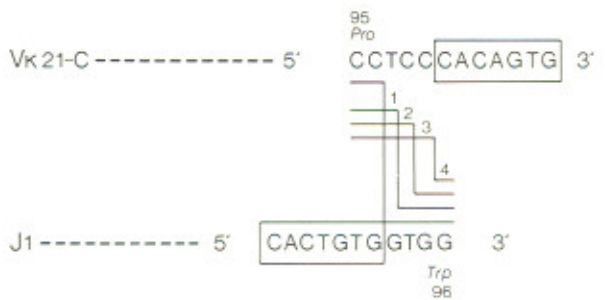

FIG. 6. Four alternative ways of recombination between the embryonic Vκ21-C DNA and the J_1 DNA segments. The conserved pentameric sequences are in boxes.

joining sites. Superimposed with these two somatic modes of antibody diversification is somatic point mutation in the V genes (18,19).

V. Mosaic Structure of Immunoglobulin Genes

Earlier amino acid sequencing studies indicated that immunoglobulin chains consist of homology units (20–22). These homology units compose structurally identifiable protein domains (23,24). The light chain consists of the V and C domains and the γ_1 class heavy chain consists of the V domain and three C domains, CH1, CH2, and CH3. Amino acid sequence studies have shown that each of the domains is also a homology unit. Namely, the amino acid sequences of all six domains are homologous to each other to a varying, but significant extent. Domains carry unique and independent functions. For instance, the V regions of both light and heavy chains are directly involved in recognition and binding of antigens, while the C region domains exert various effector functions such as binding with complement and membranes. As I have already mentioned, the V and C regions of light chains are encoded in separate exons in the active gene. In addition, the signal peptide which seems to carry a unique function of promoting transmembrane movement of secretory proteins is also encoded in a separate exon. This correlation between a functional protein unit and an exon has been even more spectacularly shown for the γ_1 heavy chain gene. Introns break coding exactly at the boundary of the four domains, V, CH1, CH2, and CH3 (Fig. 7) (25). In addition, a short peptide referred to as a hinge region and located between the CH1 and CH2 domains is encoded in yet another exon. The hinge region shows no sequence homology with other domains and has an important function of holding the four Ig chains together by disulfide bonds. As shown in Fig. 7, recent studies showed that the other heavy chain genes have similar structures (26–28; R. Maki and S. Tonegawa, unpublished results).

Based on these correlations between an exon and a functional unit of a protein, we speculated that many higher cell genes were

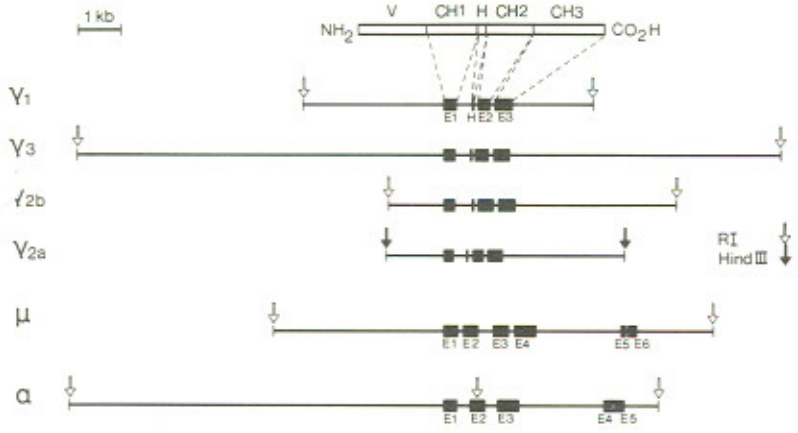

FIG. 7. Embryonic DNA clones carrying various heavy chain C region genes. At the top, correlation of the four exons, E1, H, E2, and E3, to the protein domains, CH1, CH2, and CH3, and the hinge region (H) is illustrated. The E5 and E6 of the Cμ gene and E4 and E5 of the Cα gene are involved in the membrane-bound form of the respective chains (R. Maki, unpublished results).

created in evolution by assembly of DNA segments, each coding for a polypeptide chain of some functional use. The assembly is accomplished on the DNA level by translocation of an exon together with parts of its flanking introns. This "exon shuffling" has at least three advantages in speeding up evolution (10,29). First, new proteins can be produced by bringing together amino acid sequences that have already evolved separately to fold up and to perform some functions, rather than by adding or deleting a stretch of "random" base sequences to an existing sequence. Second, multiple copies of a ready-made protein block can be made by usual gene duplication mechanisms and these copies can be recombined with other blocks to generate a spectrum of proteins having partly common and partly distinct functions. Third, the mechanisms for selecting the exon DNA sequences need not be precise, since the edges of the insertion or deletion could be located anywhere in the flanking introns, and if sequences were inserted into the intron, the exact position of insertion would not matter either.

VI. Exon Shuffling and the Heavy Chain Switch Phenomenon

Whether the above type of genome organization played a role in generation of some higher cell genes is as yet a matter of speculation. However, in light of the genuine possibility that exon shuffling might have indeed occurred in evolution, we might question whether an essentially analogous process takes place to create new genes in development and differentiation of individual organisms.

One phenomenon that may be nicely explained by exon-shuffling is the so-called "switch" of immunoglobulin heavy chain C regions. In mice a heavy chain consists of a V region attached to one of eight different C regions, each of which seems to be encoded by a separate, unique DNA segment. Each of these C regions appears to share the same set of V regions, namely, a given V gene can be expressed with any one of the eight C regions. The most immature B lymphocytes carry a μ class heavy chain (30–32). Upon encountering an antigen and often with the stimulation of proper T cells, the small B lymphocytes divide and undergo differentiation and become blast cells and eventually plasma cells. During this process, the constant region of the heavy chain synthesized switches from the μ class to any one of the several other classes, such as γ_1, γ_{2a}, γ_{2b}, α, and so on, without alteration in the V region. This phenomenon may be explained if a V-coding exon once coexpressed with the Cμ exons is translocated in the neighborhood of, for example, the Cγ_{2b} exons and forms a new transcription unit. Recent studies on the structure of heavy chain genes indicate that this is indeed what happens during the switch process (26,33).

Using a Cγ_{2b} gene probe, we screened two libraries of chimeric phages that contain either embryo DNA or DNA from a γ_{2b} chain-secreting myeloma (MOPC141), partially predigested with *Eco*RI. In this way we obtained several isolates of embryonic Cγ_{2b} gene clones and rearranged myeloma DNA clones containing the complete γ_{2b} gene (26). We also identified and isolated

the germ line V gene for the MOPC141 V_H region. Positions of the coding sequences of the three DNA clones and regions of sequence homology between various pairs of these clones have been determined by R-loop analysis and by heteroduplex analysis. As schematically indicated in Fig. 8, the 5′ portion of the myeloma DNA clone is homologous with the germ line V gene clone up to the 3′ end of the V gene. The 3′ position of the same myeloma clone is homologous with the embryonic $C\gamma_{2b}$ gene clone. This latter homology starts at a site within the intron between the V and $C\gamma_{2b}$ genes on the myeloma clone and extends to the 3′ ends of the two DNA clones. Thus, the 2.6-kb portion of the intron of the myeloma clone seems to be derived neither from the embryonic V nor from the $C\gamma_{2b}$ gene clone. This situation is in contrast to that of light chain genes where the sequences in the two types of embryonic DNA clones can account for the entire sequence contained in the corresponding myeloma DNA clone (see above).

In order to determine the embryonic origin of the 2.6 kb portion of the intron on the myeloma DNA clone that was unaccounted for by the embryonic $C\gamma_{2b}$ gene clone, we dissected a 1.3-kb

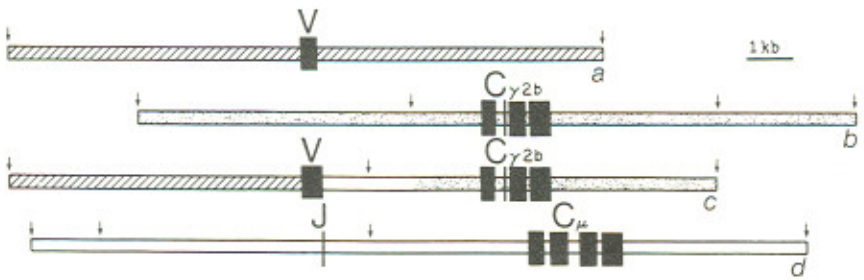

FIG. 8. Four DNA clones containing the whole or part of the $Ig\gamma_{2b}$ class heavy chain gene. (a) embryonic DNA clone containing the germ line V gene from which the V gene arose (clone PJ14); (b) embryonic $C\gamma_{2b}$ gene clone MEP3; (c) complete γ_{2b} gene clone M141-P21 isolated from myeloma MOPC141; (d) embryonic Cμ gene clone MEP203. Arrows indicate *Eco*RI cleavage sites. Filled boxes are exons. The regions of sequence homology among the four DNA clones are indicated by bars of different types of shading.

fragment from this portion of the intron and isolated another embryonic clone using this DNA fragment as the hybridization probe. The new embryonic DNA clone was shown to be homologous to the myeloma DNA clone in exactly that part of the intron in which the myeloma clone was not homologous to the embryonic $C\gamma_{2b}$ gene clone (Fig. 8). Further analysis of the embryonic clone established three important facts (26). First, the clone contains the $C\mu$ gene composed of four domain-encoding exons as shown in Fig. 8. Second, the clone contains a cluster of four heavy chain J DNA segments. One of these J DNA segments codes for the J peptide of the MOPC141 γ_{2b} chain and is located exactly at the 5′ end limit of the homology between this embryo clone and the myeloma clone. Third, DNA sequencing studies show that the putative recognition sequences for the recombinase present in the 5′-flanking regions of the light chain J DNA segments are also present in equivalent positions of the heavy chain J segments (H. Sakano and S. Tonegawa, unpublished results).

These results suggest that at least two recombination events are necessary for generation of the complete γ_{2b} gene from embryonic DNA sequences as shown in Fig. 9. One of them is analogous to the light chain V–J joining, except that the J DNA segment used for joining is located in the 5′-flanking region of the $C\mu$ gene, rather than the $C\gamma_{2b}$ gene. The second recombination occurs between a pair of sites, one located between the J DNA segment and the $C\mu$ gene, and the other in the 5′-flanking sequence of the $C\gamma_{2b}$ gene. The temporal order of the two recombinations during B cell differentiation cannot be determined from these results. However, the attractive possibility, in light of the heavy chain "switch" phenomena, is that the V–J recombination occurs first, followed by $C\mu$–$C\gamma_{2b}$ switch recombination. Since this V–J joining is analogous to that of light cha. genes, it is very reasonable to assume that this event generates an active, complete μ chain gene. The second rearrangement thus replaces the $C\mu$-coding exons of the active μ chain gene with the $C\gamma_{2b}$-coding exons by a recombination that occurs within the intron. Recent studies of α-type heavy chains in our laboratories and in Lee Hood's laboratory indicate that a complete

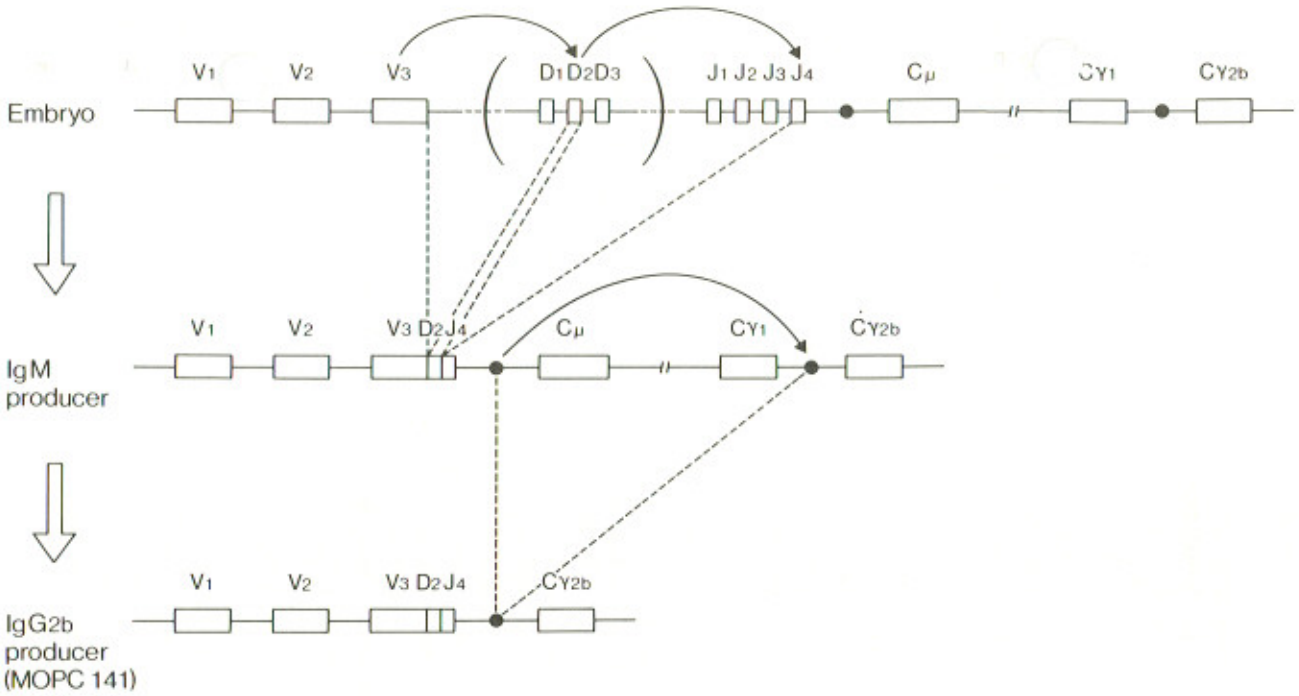

FIG. 9. Somatic generation of a complete μ gene active in mature B cells and of a complete γ_{2b} gene active in a γ_{2b} myeloma cell, MOPC141. Occurrence of the D DNA segments coding for the third hypervariable region of heavy chain V regions is predicted by recent unpublished results of P. Early and L. Hood (personal communication) and H. Sakano and myself (unpublished). The position and number of the D DNA segments are unknown. The embryonic V, D, and J DNA segments join to generate a complete μ gene active in IgM-producing B cells. Subsequently, an additional recombination replaces the Cμ gene with the $C\gamma_{2b}$ gene to generate a complete $C\gamma_{2b}$ gene.

α chain gene has a structure analogous to that of the γ_{2b} gene. It thus appears that shuffling of exons is a general event that accompanies the heavy chain switch.

Is the mechanism of the switch recombination similar to that of the V–J joining? Recently, Richard Maki has mapped the $C\gamma_1$ gene to the 5′ side of the $C\gamma_{2b}$ gene and has shown by Southern gel blotting analysis that the $C\gamma_1$ gene is totally absent in the $C\gamma_{2b}$ chain-secreting myeloma cells. This result suggests that the switch recombination also occurs by deletion of the sequences intervening the two recombination sites. Hitoshi Sakano has recently determined the nucleotide sequences at the γ_{2b} gene switch recombination sites and found no sequence that allows the stem-loop structure like the one that can be formed at the V–J recombination sites. Instead, he observed a short, direct sequence homology in the vicinity of the switch recombination site.

VII. Conclusions and Prospects

The structural analysis of immunoglobulin genes demonstrated that DNA sequences in the higher cell genome undergo controlled reorganization during differentiation of lymphocytes. The sequence reorganization seems to be vital to three key events that occur during B cell differentiation. The first is to amplify the diversity of the V region sequences, the second is to let immature B cells commit expression of one of several hundred V genes, and the third is to switch the C_H region from one class to another.

How generally the higher organism uses reorganization of DNA sequences as a means for cell determination in differentiation is an intriguing question. Recent studies revealed several cases where somatic DNA rearrangement was correlated with control of gene expression. Thus, the alternative expression of flagellar antigens in *Salmonella*, referred to as phase variation, has been shown to be regulated by an invertible sequence located adjacent to one of the flagellin structural genes (34). Mating type interconversion of homothallic yeast has been linked with alternative insertion of two different mobile DNA sequences (35). Pathogenic African trypanosomes evade the immune system of

their mammalian hosts by the sequential expression of alternative cell surface glycoproteins (36). Recent studies suggest that DNA rearrangement is responsible for this phenomenon. In addition, DNA rearrangement has been implicated for control of gene expression in several other eukaryotes, including *Drosophilas* '37) and maize (38,39), although in these cases the rearrangement does not seem to be linked with the "normal" development of the organisms as much as in other cases.

One interesting distinction between some of the aforementioned examples and the immunoglobulin gene system can be seen in the mode of DNA rearrangement. Namely, while the phase variation in *Salmonella* and mating-type interconversion in yeast are mediated by basically *reversible* inversion or transposition, the rearrangement in the immunoglobulin genes seems to be *irreversible* deletion and loss of defined DNA sequences. Most changes in development and differentiation of higher organisms seem to occur in a forward direction and do not revert under normal conditions. This apparent unidirectionality in differentiation is assured by the events occurring at the DNA level in case of immunoglobulin genes. We have yet to find the second gene(s) whose behavior reflects an apparently irreversible developmental phenomenon. However, no more than a dozen gene systems have been studied in details that match that of immunoglobulin genes. It seems clear that future studies will reveal more examples of rearrangement-dependent gene control.

ACKNOWLEDGMENTS

I wish to thank various people who at one time or another collaborated with me in the various experiments cited in this article. They include Nobumichi Hozumi, Gaston Matthyssens, Ora Bernard, Minoru Hirama, Rita Lenhard-Schuller, Christine Brack, Allan Maxam, Walter Gilbert, Richard Maki, and Yitoshi Sakano.

REFERENCES

1. Hilschmann, N., and Craig, L. C., *Proc. Natl. Acad. Sci. U.S.A.* **53**, 1403 (1965).
2. Putnam, F. W., and Easley, C. W., *J. Biol. Chem.* **240**, 1626 (1965).
3. Oudin, J., *J. Exp. Med.* **112**, 107 (1960).

4. Dreyer, W. J., and Bennett, J. C., *Proc. Natl. Acad. Sci. U.S.A.* **54,** 864 (1965).
5. Burnet, F. M., "The Clonal Selection Theory of Acquired Immunity." Vanderbilt Univ. Press, Nashville, Tennessee, 1959.
6. Gurdon, J. B., *Sci. Am.*, December (1968).
7. Gally, J. A., and Edelman, G. M., *Nature (London)* **227,** 341 (1970).
8. Southern, E. M., *J. Mol. Biol.* **98,** 503 (1975).
9. Brack, C., Hirama, M., Schuller, R., and Tonegawa, S., *Cell* **15,** 1–14 (1978).
10. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O., and Gilbert, W., *Proc. Natl. Acad. Sci. U.S.A.* **74,** 3171–3175 (1978).
11. Tonegawa, S., Hozumi, N., Matthyssens, G., and Schuller, R., *Cold Spring Harbor Symp. Quant. Biol.* **41,** 877–889 (1976).
12. Lenhard-Schuller, R., Hohn, B., Brack, C., Hirama, M., and Tonegawa, S., *Proc. Natl. Acad. Sci. U.S.A.* **75,** 4709–4713 (1978).
13. Sakano, H., Hüppi, K., Heinrich, G., and Tonegawa, S., *Nature (London)* **280,** 288–294 (1979).
14. Max, E. E., Seidman, J. G., and Leder, P., *Proc. Natl. Acad. Sci. U.S.A.* **76,** 3450–3454 (1979).
15. Pernis, B., Chiappino, G., Kelus, A. S., and Gell, P. G. H., *J. Exp. Med.* **122,** 853 (1965).
16. Ohtsubo, H., and Ohtsubo, E., *Proc. Natl. Acad. Sci. U.S.A.* **75,** 615–619 (1978).
17. Grindley, N. D. F., *Cell* **13,** 419–426 (1978).
18. Weigert, M. G., Cesari, I. M., Yonkovich, S. J., and Cohn, M., *Nature (London)* **228,** 1045 (1970).
19. Tonegawa, S., *Proc. Natl. Acad. Sci. U.S.A.* **73,** 203 (1976).
20. Hill, R. L., Delaney, R., Fellows, R. E., and Lebovitz, H. E., *Proc. Natl. Acad. Sci. U.S.A.* **56,** 1762 (1966).
21. Singer, S. J., and Doolittle, R. F., *Science* **153,** 13 (1966).
22. Edelman, G. M., Cunningham, B. A., Gall, W. E., Gottlieb, P. D., Rutishauser, U., and Waxdal, M., *Proc. Natl. Acad. Sci. U.S.A.* **63,** 78 (1969).
23. Amzel, L., Poljak, R., Saul, F., Varga, J., and Richards, F., *Proc. Natl. Acad. Sci. U.S.A.* **71,** 1427 (1974).
24. Padlan, E. A., Segal, D. M., Cohen, G. H., and Davies, D. R., *in* "The Immune System: Genes, Receptors, Signals" (E. Sercarz, A. Williamson, and C. Fox, eds.), p. 7. Academic Press, New York, 1974.
25. Sakano, H., Rogers, J. H., Hüppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R., and Tonegawa, S., *Nature (London)* **277,** 627 (1979).
26. Maki, R., Traunecker, A., Sakano, H., Roeder, W., and Tonegawa, S., *Proc. Natl. Acad. Sci. U.S.A.* **77,** 2138 (1980).
27. Early, P. W., Davis, M. M., Kaback, D. B., Davidson, N., and Hood, L., *Proc. Natl. Acad. Sci. U.S.A.* **76,** 857 (1979).
28. Calame, K., Rogers, J., Early, P., Davis, M., Livant, D., Wall, R., and Hood, L., *Nature (London)* **284,** 452 (1980).
29. Crick, F., *Science* **204,** 264 (1979).

30. Raff, M. D., Megan, M., Owen, J. J. T., and Cooper, M. D., *Nature (London)* **259,** 224 (1976).
31. Gathings, W. E., Lawton, A. R., and Cooper, M. D., *Eur. J. Immunol.* **7,** 804 (1977).
32. Burrows, P., LeJeune, M., Kearney, J. F., *Nature (London)* **280,** 838 (1979).
33. Davis, M. M., Calame, K., Early, P. W., Livant, D. L., Joho, R., Weissman, I. L., and Hood, L., *Nature (London)* **283,** 733 (1980).
34. Zieg, J., Hilmen, M., and Simon, M., *Cell* **15,** 237 (1978).
35. Hicks, J. B., Strathern, J. N., and Herskowitz, I., *in* "DNA Insertion Elements, Plasmids, and Episomes" (A. Bukhari, J. Shapiro, and S. Adhya, eds.), pp. 457–462. Cold Spring Harbor Lab., Cold Spring Harbor, New York, 1977.
36. Hoeijmakers, J. H. J., Frasch, A. C. C., Bernards, A., Borst, P., and Cross, G. A. M., *Nature (London)* **284,** 78 (1980).
37. Gehring, W. J., and Paro, R., *Cell* **19,** 897 (1980).
38. McClintock, B., *Proc. Natl. Acad. Sci. U.S.A.* **36,** 344 (1950).
39. McClintock, B., *Cold Spring Harbor Symp. Quant. Biol.* **16,** 13 (1952).
40. Seidman, J. G., Max, E. E., and Leder, P., *Nature (London)* **280,** 370 (1979).
41. Seidman, J. G., Leder, A., Edgell, M. H., Polsky, F., Tilghman, S. M., Tiemeier, D. C., and Leder, P., *Proc. Natl. Acad. Sci. U.S.A.* **75,** 3881 (1978).
42. Bernard, O., Hozumi, N., and Tonegawa, S., *Cell* **15,** 1133 (1978).