

Somatic Reorganization of Immunoglobulin Genes during Lymphocyte Differentiation

S. TONEGAWA, H. SAKANO, R. MAKI, A. TRAUNECKER, G. HEINRICH, W. ROEDER,
AND Y. KUROSAWA

Basel Institute for Immunology, CH-4005 Basel 5, Switzerland

Immunoglobulin molecules consist of two identical light (L) chains and two identical heavy (H) chains. Both light- and heavy-chain polypeptides have two regions, the aminoterminal half, called the variable (V) region, and the carboxyterminal half, called the constant (C) region. In embryonic cells, the two regions are encoded in widely separated DNA segments. Complete immunoglobulin genes are created by somatic recombination that occurs during differentiation of lymphocyte precursor cells (Tonegawa et al. 1977b). The organization of the gene sequences before and after somatic recombination has been studied extensively first for the λ -type light chain of the mouse (Brack and Tonegawa 1977; Tonegawa et al. 1977a, 1978; Bernard et al. 1978; Brack et al. 1978). These studies established that in the embryonic genome, the major portion of the conventionally defined V region is encoded in a DNA segment (V DNA) that is located some distance away from a DNA segment (J DNA) coding for the rest of the V region. Although the distance between the V and C DNA segments is unknown, the J DNA segment was mapped at 1.2 kb toward the 5' side, relative to the direction of transcription, of the C DNA segment. Somatic recombination takes place at the 3' end of the V DNA segment and the 5' end of the J DNA segment. Consequently, in λ_1 chains synthesizing B lymphocytes or plasma cells, the DNA segment coding for the entire V region is much closer to the C DNA segment. An analogous somatic recombination also takes place in the DNA sequences coding for κ -type light chains (Lenhard-Schuller et al. 1978; Seidman and Leder 1978; Seidman et al. 1979), although the two recombinations, one for λ and the other for κ , do not seem to take place within the same lymphocyte precursor.

Since the heavy-chain polypeptide is also composed of a V and C region, it is reasonable to assume that what has been established at the DNA level for the light chains will also be true for heavy chains. However, the expression of heavy chains has some features that make it more complex than the expression of light chains. Unlike for the light chains, for the heavy chains of the mouse there exist eight different C regions, each appearing to share the same set of V regions. In addition, it appears that as the lymphocytes differentiate, the C region of the heavy chain synthesized switches from one class to another without alteration of the V region.

Recent studies on heavy-chain gene clones showed that, unlike the light-chain genes, at least two recombinations are necessary to generate the heavy-chain gene

(Davis et al. 1980; Maki et al. 1980b; Sakano et al. 1980). One of them occurs at or near the 3' end of the V DNA segment and in the vicinity of the J DNA segment. In contrast to the light-chain genes, where the J DNA segments are located in the 5' flanking regions of their respective C genes, in the heavy-chain gene system the J DNA is located only in the 5' flanking region of the germ-line C_μ gene. The second recombination occurs between a pair of sites, one located between the J DNA segment and the C_μ gene and the other in the 5' flanking sequence of a different constant gene. The C_μ exons are thus replaced by the exons of another C gene. This recombination is referred to as "switch-recombination" because it precedes and seems to be a key event required for the heavy-chain switch.

In this paper we summarize our recent studies on the structure and somatic DNA rearrangements of immunoglobulin genes.

V-J Joining in Immunoglobulin Light-chain Genes

Figure 1 summarizes the organization of κ -type light-chain genes in embryonic and κ -chain-synthesizing myeloma cells. Unlike in the λ -chain gene, where both the V and J sequences are unique, there are a few hundred different V_κ DNA copies clustered and five different copies of J_κ DNA segments also clustered and located about 3 kb upstream of the unique C_κ DNA. 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 V_κ DNA segments joins with one of the five J_κ 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.

To study the fine structure of the V-J recombination sites, we determined the nucleotide sequences of the J cluster (Sakano et al. 1979a) (Fig. 2A). The distance between adjacent J segments is rather constant and the five are arranged as follows: J_1 -310 bp- J_2 -246 bp- J_3 -286 bp- J_4 -299 bp- J_5 . Figure 2B compares the five J sequences in the coding and flanking regions. Sequences are highly conserved in the coding regions and around the RNA-splicing sites. In contrast, sequences in the rest of the flanking regions are rather diverse. Nevertheless, two, short, conserved sequences are in the 5' noncoding regions. One is a nanomer, GGT TTTTGT, located

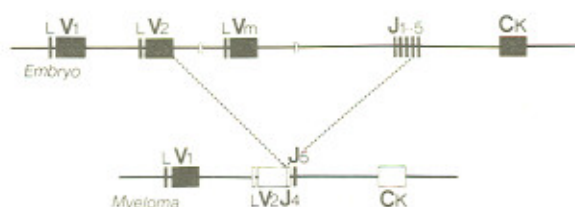


Figure 1. Organization of mouse κ light-chain genes in embryo and κ -chain-producing myeloma cells. To generate an active gene, recombination occurs between one of a few hundred embryonic V DNAs and one of five J DNAs. Hydrophobic leader peptides are encoded by separate exons (L). Coding sequences that are expressed into a protein sequence in myeloma cells are shown by open boxes.

about 30 bp from the J segments, and the other is a palindromic hexamer interrupted by an AT base pair at the center of symmetry, CACTGTG, immediately preceding the J segments. The same or closely related sequences are also present at the equivalent positions in the λ_1 -chain gene clone. These conserved sequences are repeated inverted in the 3' noncoding regions of a variety of embryonic V-coding DNA segments (Table 1). 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 Figure 3. Based on these results and a series of Southern gel blot experiments described elsewhere (Sakano et al. 1979a), 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 sequence (IS) elements. For instance, Ohtsubo and Ohtsubo (1978) and Grindley (1978) determined sequences of IS1 elements and found that some 30 bases at one end are repeated inverted on the other end. Furthermore, a run of T's surrounded by G's is also present in the stem of IS1 as well as in the equivalent stems of many other transposable elements, including bacteriophage Mu (Kamp et al. 1979), and of invertible elements of the *Salmonella* flagellin gene (Zieg and Simon 1980). Although these structural similarities may be coincidental, they could reflect a common evolutionary origin, or a common enzymatic mechanism for these 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 nanomer. As summarized in Table 1, in almost all cases the spacers are 12 bp or 23 bp long, plus or minus 1 bp. The spacers of all V_{κ} segments correspond to the shorter spacer (12 bp), whereas those of $V_{\lambda 1}$ and $V_{\lambda H}$ correspond to the longer one. All J_{κ} spacers are the longer ones, whereas the $J_{\lambda 1}$ spacer is the shorter one. It should be emphasized that although the lengths of the spacers are conserved in a given type of gene, their sequences have highly diverged. 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, contains two binding proteins—one recognizing the heptamer and nanomer separated by the short spacer and the other the heptamer and nanomer separated by the long spacer. A similar model has been proposed independently by Early et al. (1980a). 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 intra-strand base pairing (see Fig. 3) occurs, however transiently, in the enzyme-DNA complex and facilitates the ligation reaction.

The available protein sequences (Table 2) 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 (Weigert et al. 1978; Max et al. 1979; Sakano et al. 1979a). For instance, in the $V_{\kappa 21-C}$ and J_1 pair shown in Figure 3, 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 Figure 4. In contrast, recombination events 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 MOPC-173, PC6684, MPC-11, and PC7940, respectively (Table 1). Note that none of the five J DNA segments identified thus far can by itself encode Arg or Pro at position 96. Since residue 96, the most variable residue in the κ light chains, is at the margin of one of the three hypervariable (HV) regions, it 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 joining sites.

Correlation of Exons and Functional Protein Units

Amino acid sequencing studies indicated that immunoglobulin peptides consist of homology units, each of which is similar in size (about 110 residues) and homologous to the light-chain C region (Edelman et al. 1969). These homology units compose structurally identifiable protein domains (Poljak et al. 1973; Padlan et al. 1973). The light chain consists of the V and C domains, and the $\gamma 1$ -class heavy chain consists of the V domain and three C domains, C_{H1} , C_{H2} , and C_{H3} . 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, whereas the C domains exert various effector functions such as binding with complement and membranes. As already mentioned, the V and C regions of light chains are encoded in separate exons in the active gene. In addition, the hydrophobic leader peptide, which seems to carry a unique function of secreting the proteins, is also encoded in a separate exon (Seidman et al. 1978; Tonegawa et al. 1978; G. Heinrich, unpubl.).

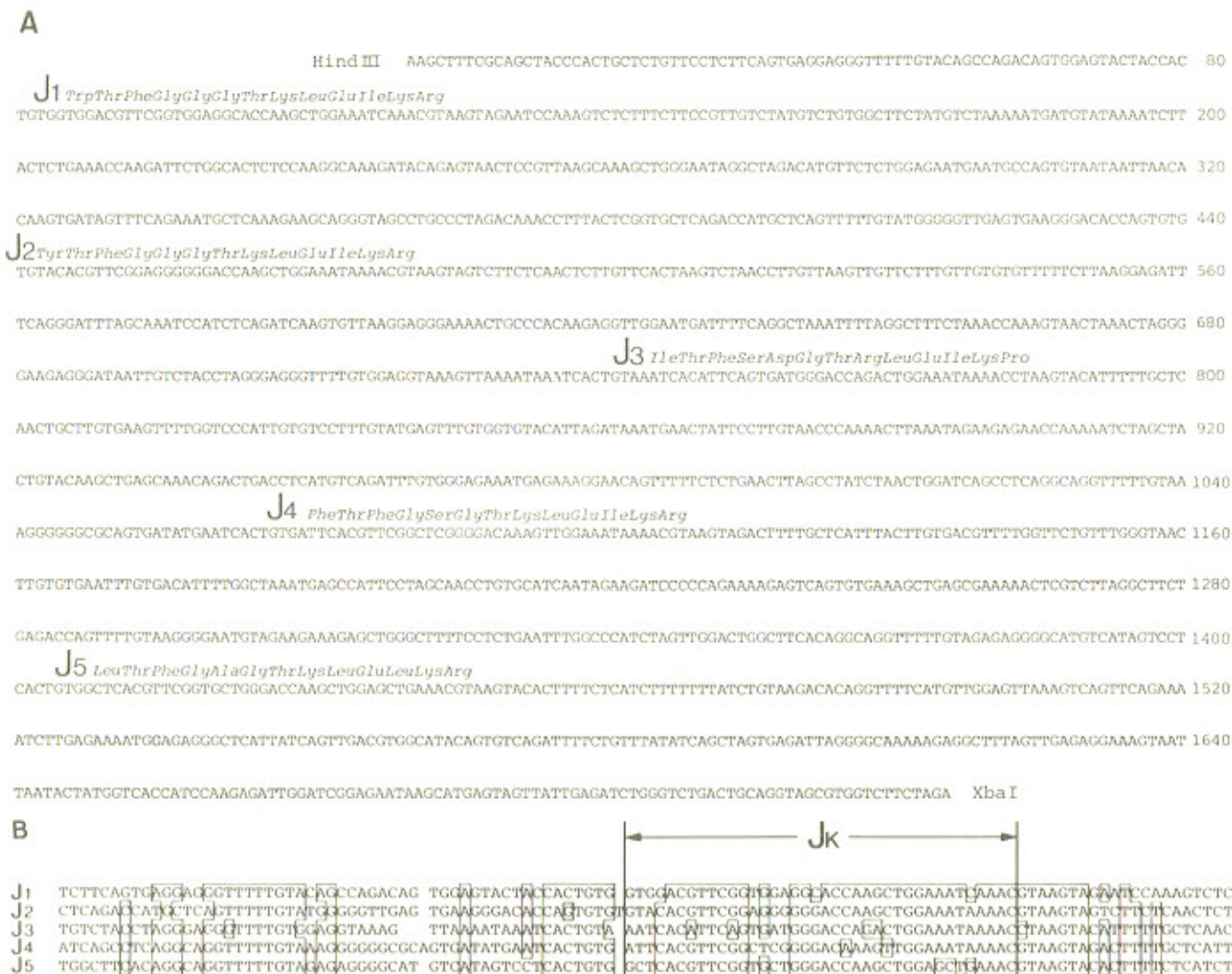


Figure 2. (A) Complete nucleotide sequence of the 1.7-kb *Hind*III-*Xba*I fragment containing the J_k cluster (Sakano et al. 1979a). By comparing published J_k peptide sequences with the nucleotide sequence determined here, we identified five J DNA sequences. The amino acid sequences encoded by these J DNA segments are shown in italics. (B) Comparison of nucleotide sequences of the five J coding sequences and the flanking regions. Nucleotides common in at least four J DNAs are outlined. Vertical lines indicate coding ends by J DNA segments.

Table 1. Conserved Sequences of Embryonic J and V DNAs

J DNA	Nanomer	Spacer (bp)	Heptamer	Reference
J _{κ1}	GGTTTTGT	23	CACTGTG	a, b
J _{κ2}	AGTTTTGT	23	CAGTGTG	a, b
J _{κ3}	GGGTTTTGT	21	CACTGTA	a, b
J _{κ4}	GGTTTTGT	24	CACTGTG	a, b
J _{κ5}	GGTTTTGT	23	CACTGTG	a, b
J _{λ1}	GGTTTTGC	12	CACAGTG	c
J _{λ11}	AGTTTAGT	22	GACTGTG	d, e
J _{λ12}	GGTTTTGT	23	TAGTGTG	d, e
J _{λ13}	ATTATTGT	21	CACAATG	d
J _{λ14}	GGTTTTGT	23	CAATGTG	d
		22	TATTGTG	d
		31	TACTATG	d
V DNA	Heptamer	Spacer (bp)	Nanomer	Reference
V _{κ21-c}	CACAGTG	11	ACAAAAACC	a
V _{κ21-a}	CACAGTG	12	ACAAAAACC	f
V _{κ41}	CACAGTG	12	ACATAAACC	g
V _{κ2}	CACAGTG	12	ACATAAACC	h
V _{λ1}	CACAATG	22	TCAAGAACA	c
V _{λ11}	CACAATG	23	ACAAAGAACA	i
V _{λ141}	CACAGTG	23	ACAAATACC	d
V _{λ107}	CACAGTG	23	ACAAAACC	e
V _{λ26}	CACAGTG	22	ACACAAACC	j
V _{λ52}	CACAGTG	22	ACACAAACC	j

Two blocks of conserved sequences found in the 3' flanking region of embryonic V DNAs and in the 5' flanking region of embryonic J DNAs. The numbers between the two sequences indicate the length of the spacer between them in base pairs. The bases different from those of the basic sequences CACTGTG and GGTTTTTGT, are underlined. Only the sequences on sense strands are shown in the table. The sequences were taken from the following references: a, Sakano et al. (1979a); b, Max et al. (1979); c, Bernard et al. (1978); d, Sakano et al. (1980); e, Early et al. (1980a); f, G. Heinrich (pers. comm.); g, Seidman et al. (1979); h, Seidman et al. (1978); i, Tonegawa et al. (1978); j, G. Matthysens and T. H. Rabbitts (pers. comm.).

This correlation between a functional protein unit and an exon has been even more spectacularly shown for the $\gamma 1$ heavy-chain gene. Introns break coding exactly at the boundary of the four domains, V, C_{H1}, C_{H2}, and C_{H3} (Honjo et al. 1979; Sakano et al. 1979b). In addition, a short peptide, referred to as a hinge region and located between the C_{H1} and C_{H2} domains, is encoded in yet another exon. The hinge region shows no sequence homology with other domains and has the important function of holding the four immunoglobulin chains together by disulfide bonds.

Both electron microscopy and DNA sequencing studies showed that the $\gamma 2b$ heavy chain has a similar gene structure (Tucker et al. 1979; Maki et al. 1980b; Yamawaki-Kataoka et al. 1980). Our recent studies on $\gamma 2a$ and $\gamma 3$ heavy-chain gene clones demonstrated that these genes also have similar mosaic gene structure. The μ - and α -class heavy chains are known to have no hinge region and are composed of four and three C-region domains, respectively. In accordance with these domain structures, the C _{μ} (Calame et al. 1980; Gough et al. 1980; Maki et al. 1980b) and C _{α} (Early et al. 1979) genes are split into four and three exons, respectively, of approximately the same size. The exon-intron structure of all mouse heavy-chain C genes, except for those of ϵ and δ , are illustrated schematically in Figure 5.

It is thus clear that in immunoglobulin genes there exists an excellent correlation between exons and func-

tional units of polypeptide chains. As described in a later section of this paper and depicted in Figure 5, the correlation also applies to a hydrophobic peptide at the carboxyl terminus of the membrane-bound forms of the μ - and α -class heavy chains.

Differentiation of B Lymphocytes

The organization and expression of immunoglobulin heavy-chain genes is expected to be more complex than those of the light-chain genes. Three sources of complexity are evident. First, heavy chains appear in several different classes or subclasses, each characterized by a unique C region. For instance, mouse heavy chains are classified into five different classes, namely, μ , δ , γ , α , and ϵ . The γ -class chains are further subdivided into four different subclasses, namely, $\gamma 1$, $\gamma 2a$, $\gamma 2b$, and $\gamma 3$. Second, heavy chains of all classes and subclasses seem to share a common pool of V regions. In other words, a given V_H gene seems to be expressed in association with any one of the several C_H genes. Third, expression of various classes or subclasses of heavy-chain genes is developmentally controlled (Fig. 6).

The first recognizable cells to begin development in the B-cell lineage contain free μ -class heavy chains in the cytoplasm (Raff et al. 1976; Burrows et al. 1979; Levitt and Cooper 1980). In these μ -only cells, V-J rearrangement has occurred for the heavy chains, but no

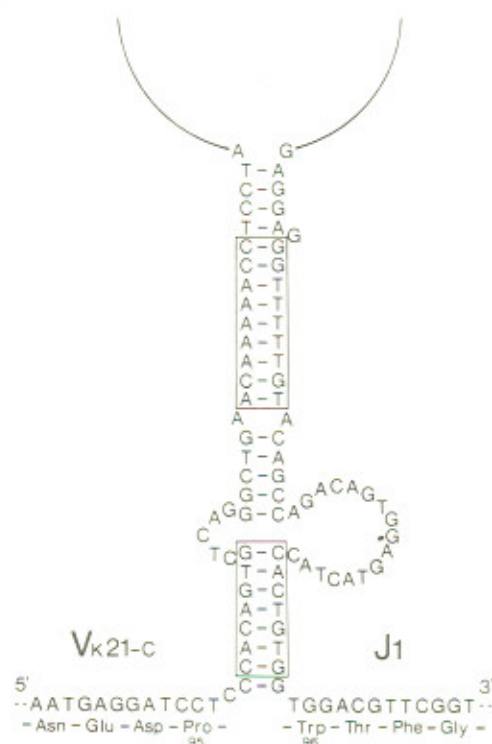


Figure 3. Inverted-repeat stem structure that could be formed between the 3' noncoding region of embryonic V DNA ($V_{\kappa 21-C}$) and the 5' noncoding region of J DNA (J_1) (Sakano et al. 1979a). Sequences for the antisense strand have been omitted. Conserved base pairs in the stem region, CACTGTG and GGTGTTTGT, are surrounded. Amino acid sequences encoded by the V and J DNAs are shown, and the residue numbers are also indicated at the bottom.

DNA rearrangement has been detected for the light chains (Maki et al. 1980a). In mice, these cells, referred to as pre-B cells, are first detectable in the livers of 12-day-old fetuses. In adult animals, pre-B cells are present exclusively in bone marrow. Large pre-B cells have a

high mitotic rate and give rise to smaller pre-B cells that rarely divide; these in turn give rise to immature B cells bearing complete IgM molecules on the cell surface. The immature, IgM-bearing B cells in mice are resting, small lymphocytes that bear few, if any, additional surface receptors that characterize the mature, virgin B lymphocytes. These receptors include IgD molecules that share the same idiotype with the IgM present on the same cells (Salsano et al. 1974; Fu et al. 1975; Goding and Layton 1976). Upon stimulation with antigen and often with the help of T cells, the IgM-IgD double-bearing B cells undergo further maturation, which culminates in plasma cells actively secreting immunoglobulin molecules.

At least two apparently parallel pathways are followed during this process. In one, the B cells mature to IgM-secreting plasma cells. In the other, the synthesis of a new class of heavy chain, such as γ , α , or ϵ , which apparently carries the same V region as the μ and δ chains of the progeny cells, is initiated with a concomitant disappearance of the δ chain (Cooper et al. 1977; Pernis et al. 1977). This phenomenon has been referred to as "class switch." During this process, the cells undergo a transient state in which the IgM and the newly acquired class of immunoglobulin coexist within a single cell and eventually mature to plasma cells that synthesize and secrete the latter-class immunoglobulin exclusively. Within a clone, all classes of heavy chains that appear during the various stages of B-cell differentiation share the same light chain that appeared first in the immature B cells.

Additional Exons Encode the Carboxyterminal Peptides of Membrane-bound μ and α Heavy Chains

As described in the previous section, IgM molecules in the resting B lymphocytes are anchored in the cell-surface membrane and function as antigen receptors. Upon triggering with proper antigen or stimulation with

Table 2. J Peptide Sequences of κ Light Chains

Group	Myeloma	J_{κ} peptide	J_{κ} DNA
I	M141 M63 B32 2880 7210 M70 M41 A17 1229 6308 C101 T111 X44 7769	W T F G G G T K L E I K R	$J_{\kappa 1}$
II	M173 6684	R - - - - -	$J_{\kappa 1}$
III	M21 A22 9245 3741	Y - - - - -	$J_{\kappa 2}$
IV	M11 7940	P - - - - -	$J_{\kappa 1}/J_{\kappa 2}$
V	M321 7043	P - - - S - - - - -	$J_{\kappa 4}$
VI	T124	- - - S - - - - -	$J_{\kappa 4}$
VII	X24	I - - - S - - - - -	$J_{\kappa 4}$
VIII	LPC1 M511 4045 3154 H8 7175 2485 7183	L - - - A - - - - -	$J_{\kappa 5}$
IX	T601	I - - - A - - - - -	$J_{\kappa 5}$
X	none	I - - S D - - R - - - P	$J_{\kappa 3}$
XI	2413	- - - - - D - - - E -	?

κ light chains of BALB/c and NZB myelomas are classified into ten groups according to the J peptide sequences determined by McKean et al. (1978), Weigert et al. (1978), and Rudikoff and Potter (1974). J_{κ} DNAs of MOPC-141, LPC1, and HOPC-8 κ chains were assigned based on the R-loop mapping data of their DNA clones, although their amino acid sequences are unknown. Assignment of J_{κ} DNAs for J_{κ} peptides belonging to groups II and IV are tentative because the V-J joining sites seem to be unfixed (see text). Nucleotide sequences of BALB/c J_{κ} DNA for NZB myeloma 2413 has not been assigned. No J_{κ} peptide so far sequenced corresponds to $J_{\kappa 3}$ DNA. Amino acid residues are shown by one-letter codes (IUPAC-IUB Commission 1969).

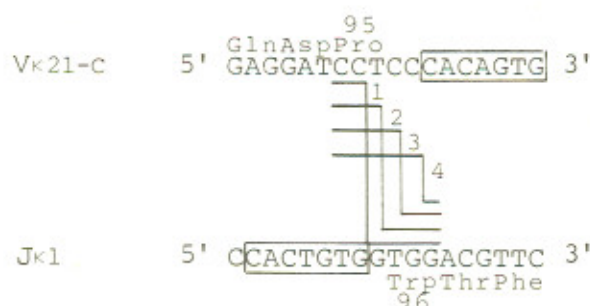


Figure 4. Variable recombination sites between $V_{\kappa 21-C}$ and $J_{\kappa 1}$.

mitogens, the lymphocytes proliferate and mature to Ig-secreting blast cells and eventually to plasma cells (step d in Fig. 6). It has been known that the μ chain of membrane-bound IgM is slightly larger than that of secreted IgM (Melcher and Uhr 1976). This difference was thought to be due to the presence of an extra hydrophobic peptide in the C region of the membrane-bound μ chains (Bergman and Haimovich 1978; Vassalli et al. 1979).

During analysis of R loops formed by the embryonic C_{μ} gene clone (clone MEP203) and μ -chain mRNA isolated from MOPC-104E myeloma cells, we noticed two different types of hybrid structures. The majority (>95%) of the hybrid molecules exhibited four R loops, presumably corresponding to the four domains of the secreted μ chain (Fig. 7A). The rest of the hybrid molecules contained, in addition to the four R loops observed in the majority type, two more exons located to the 3' side of the first four (Fig. 7B). Since MOPC-104E cells not only actively secrete IgM molecules, but also carry a small amount of them on the membrane (L.

Forni, pers. comm.), the aforementioned results suggested to us that the two types of R loops were formed by two different mRNA molecules coding for the secreted and membrane-bound forms of μ chains. This interpretation was borne out by two types of experiments.

First, we analyzed R loops formed by the embryonic C_{μ} gene clone and partially purified μ mRNA from a B-cell hybridoma, GCL28. This hybridoma was prepared by fusing mouse splenic B cells with a Syrian hamster B lymphoma line, GD36AG1, and has been shown to bear IgM predominantly on the cell surface (Raschke 1978). In agreement with the interpretation raised above, an overwhelming majority of the hybrid molecules exhibited the six-exon structure that is indistinguishable from the minority type observed with the MOPC-104 μ mRNA. Second, we analyzed the sizes of the GCL28 and MOPC-104E μ mRNAs by gel electrophoresis followed by blotting procedures (Alwine et al. 1977), using as the hybridization probes two genomic DNA fragments dissected out from the embryonic C_{μ} gene clone. These were a 4.2-kb *Xba*I-*Xba*I fragment containing exons 1 to 4 and a 1.5-kb *Xba*I-*Xba*I fragment carrying exon 6. As shown in Figure 8, GCL28 cells contain two types of μ mRNA, 2.6 kb and 2.4 kb, but the former is much more abundant, indicating that this μ mRNA codes for the membrane-bound μ chain. In contrast, only the 2.4-kb μ mRNA was detected in MOPC-104E (we believe that this myeloma also contains the 2.6-kb mRNA, but the level is too low to be detected in the experiment illustrated), indicating that this mRNA codes for the secreted μ chain. In agreement with the R-loop analysis, only the 2.6-kb mRNA hybridized with the probe containing exon-6 sequences (Fig. 8).

We have recently observed an analogous structure for the C_{α} gene. α mRNA from MOPC-315 gave two types of R-loop structures: The majority type contains three

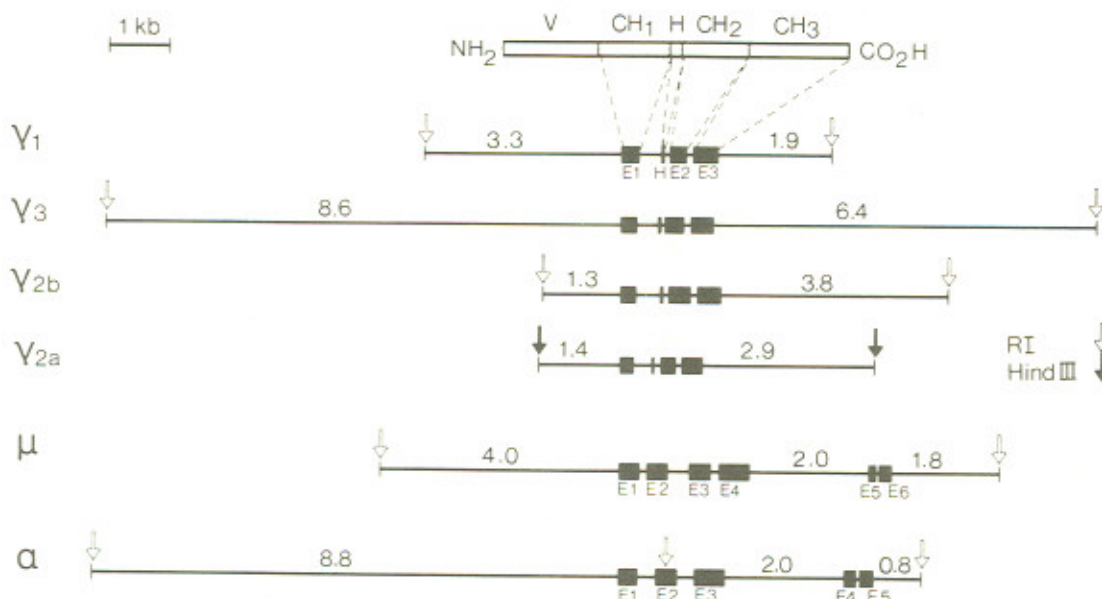


Figure 5. Correspondence of exons to functional units of Ig chains. The numbers indicate lengths (in kb) of DNA between two flanking exons or between an exon and an enzyme cleavage site.

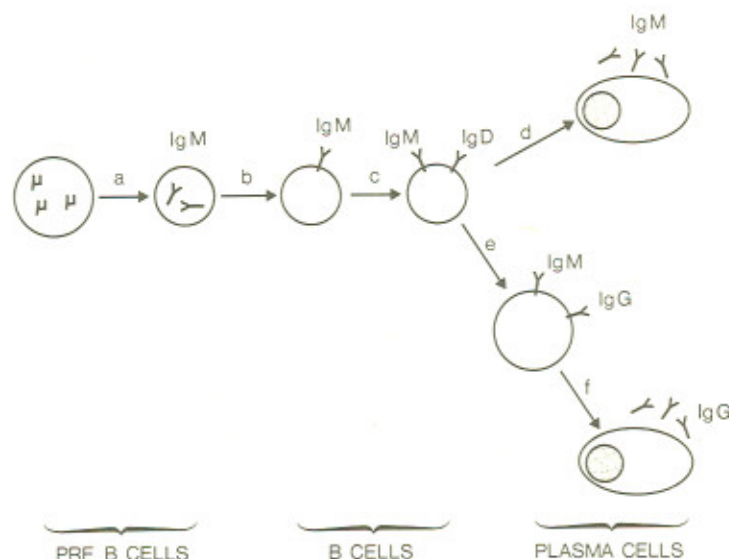


Figure 6. Differentiation of B lymphocytes. See text for explanation.

exons and the minority type the same three exons plus two more exons located 2 kb downstream (Fig. 9). This myeloma contains two types (2 kb and 2.3 kb) of α mRNA (data not shown), and the cells secrete, as well as bear, IgA molecules.

The above results strongly suggest that the membrane-bound forms of both μ and α chains are encoded in part by exons that are shared by the secreted form of the respective chain and in part by exons unique to it. The latter exons most probably encode the extra peptide that renders the otherwise secreted immunoglobulin molecules hydrophobic and able to be anchored in the cell-surface membrane. That this is indeed true has been shown by DNA sequence studies of both μ cDNAs (Rogers et al. 1980) and genomic coding regions (Early et al. 1980b). Thus, here again we have an example of a correlation between an exon and a peptide having localized specific function. Differential processing of a primary RNA transcript leading to generation of multiple forms of mRNA has precedents in DNA tumor viruses (for review, see Tooze 1980). In these cases, it is not clear whether synthesis of different forms of mRNA are actively regulated during the life cycle of viruses or in the transformed cells. In contrast, synthesis of the two forms of immunoglobulin heavy chains, particularly that of μ chains, is clearly regulated during lymphocyte differentiation. Immunoglobulin genes are the first case in which regulation of RNA processing was invoked for control of gene expression in development.

Somatic Rearrangement of Heavy-chain Genes

To study the sequence reorganization of heavy-chain genes that is expected to take place during B-cell differentiation, we chose the γ_{2b} heavy-chain gene expressed in myeloma MOPC-141. Using a $C_{\gamma_{2b}}$ gene probe, we screened two libraries of chimeric phages that

contain either embryonic DNA or DNA from a γ_{2b} -chain-secreting myeloma (MOPC-141), partially digested with *EcoRI*. In this way, we obtained several isolates of embryonic $C_{\gamma_{2b}}$ gene clones and rearranged myeloma DNA clones containing the complete γ_{2b} gene (Maki et al. 1980b). We also identified and isolated the germ-line V gene for the MOPC-141 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 indicated schematically in Figure 10, the 5' portion of the myeloma DNA 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 gene 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 Fig. 1).

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 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. 10). Further analysis of the embryonic clone established three important facts (Maki et al. 1980b; Sakano et al. 1980). First, the clone contains the C_μ gene composed of four domain-encoding exons as shown in Figure 10. Second,

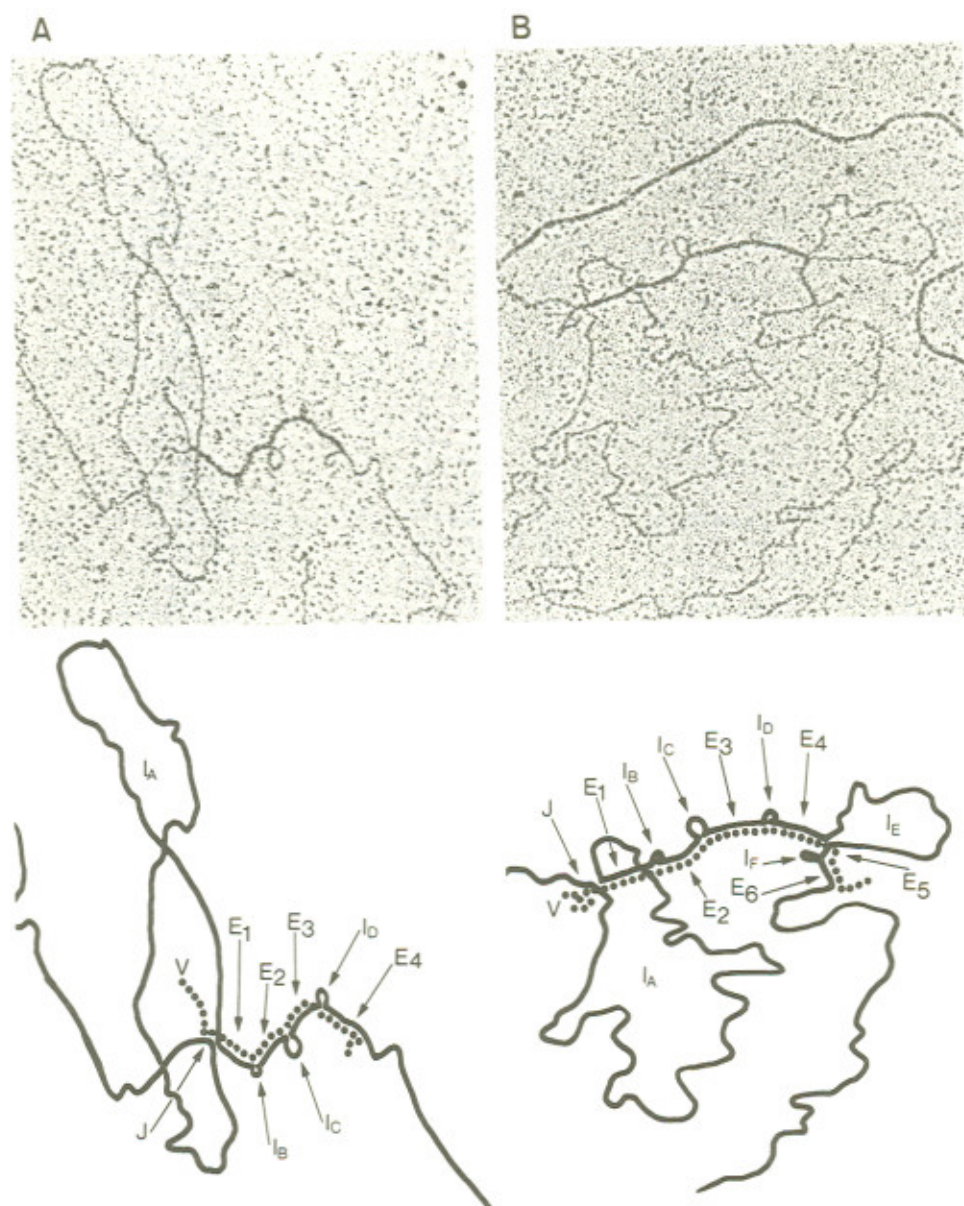


Figure 7. DNA coding for two different forms of μ -chain mRNA. (A) DNA-RNA hybrid formed between an embryonic C_μ gene clone (clone MEP203) DNA and the major μ -chain mRNA from MOPC-104E. (B) Hybrid formed between clone MEP203 DNA and the minor μ -chain mRNA from MOPC-104E.

the clone contains a cluster of four heavy-chain J DNA segments (Fig. 11). One of these J DNA segments codes for the J peptide of the MOPC-141 γ_{2b} chain and is located exactly at the 5' end of the homology between this embryo clone and the myeloma clone (Fig. 12). Third, DNA sequencing studies showed 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 (Figs. 11 and 12).

These results suggest that at least two recombination events are necessary for generation of the complete γ_{2b} gene from embryonic DNA sequences. 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_μ 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_μ gene and the other in the 5' flanking sequence of the $C_{\gamma_{2b}}$ gene. Since this V-J joining is analogous to that of light-chain genes, it is very reasonable to assume that this event generates an active, complete μ -chain gene. The second rearrangement thus replaces the C_μ -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 an α -type gene (Davis et al. 1980) indicate that a complete α -chain gene has a structure analogous to that of the γ_{2b} gene. In addition, a sequencing

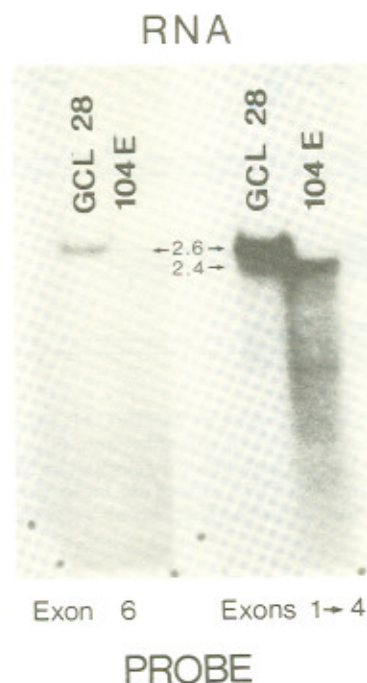


Figure 8. Analysis of two forms of μ -chain mRNA by gel blot hybridization. Poly(A)⁺ RNA was purified by two passages through oligo(dT) cellulose and electrophoretically separated on a 1% agarose gel containing methyl mercury. The RNA was transferred to diazobenzylmethyl paper and hybridized with 1×10^7 cpm of nick-translated DNA containing exons 1-4 or exon 6, essentially as described by Alwine et al. (1977). The 28S (5 kb) and 18S (2.1 kb) rDNAs were used as size markers. Sizes are given in kbp.

analysis of a part of a γ 1 gene suggests that this gene may also have a structure analogous to that of the γ 2b gene (Kataoka et al. 1980). Below we describe the fine sequence structure of the two types of recombination as studied by DNA sequencing.

Somatic Generation of Complete Heavy-chain V Genes

Involvement of three different germ-line gene segments (V, D, and J). For the analysis of the V-J joining, we first constructed restriction enzyme cleavage maps (Fig. 13A) for DNA clones MEP203 (embryonic J + C μ), M141-P21 (complete γ 2b), and PJ14 (embryonic V) and sequenced the appropriate regions (Sakano et al. 1980). Figure 12 shows these three sequences: As shown, the coding of a V gene by the myeloma clone (M141-P21) continues up to Ser (125), whereas that of the embryonic clone (PJ14) ends prematurely with Ser (97). The nucleotide sequence of the embryonic J (MEP203) is also compared. Unexpectedly, coding by the J begins with Leu (112) and ends with Ser (125). Thus, it appears that the 14-residue-long peptide comprising the third hypervariable region (HV3) beginning with Val (98) and ending with Thr (111) is encoded neither in the germ-line V gene nor in the J DNA segment. This situation is in contrast to that of light-chain genes described in the previous section (also see Bernard et al. 1978; Sakano et al. 1979a; Seidman et al. 1979). In the latter cases, no such coding gap was observed. We suggest that the HV3 of this heavy chain is encoded in one or more discrete segments. A similar observation has also been made in a complete immunoglobulin α heavy-chain gene by Early et al. (1980a). These DNA segments coding for HV3 have been referred to as D DNA segments, D standing for diversity (Schilling et al. 1980). The V, D, and J DNA segments must be assembled by recombination to generate the complete heavy-chain V genes.

The DNA sequence of the clone PJ14 demonstrated that coding by the germ-line V_{M141} gene ends with the second base of the Ser (97) codon AGC (Fig. 12). As shown in Table 3, the amino acid sequences immediately preceding the Ser (97) are highly conserved among various V regions. Particularly noteworthy is the univer-

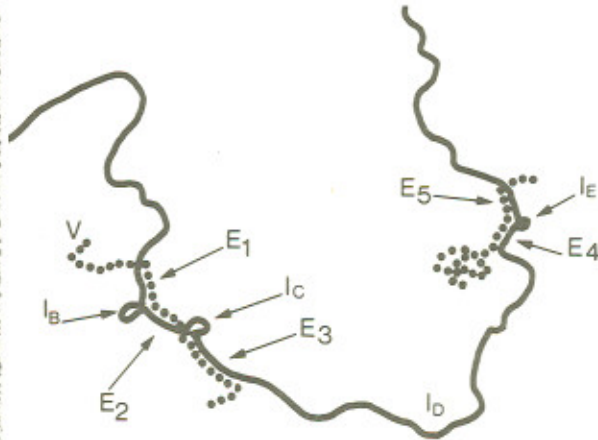
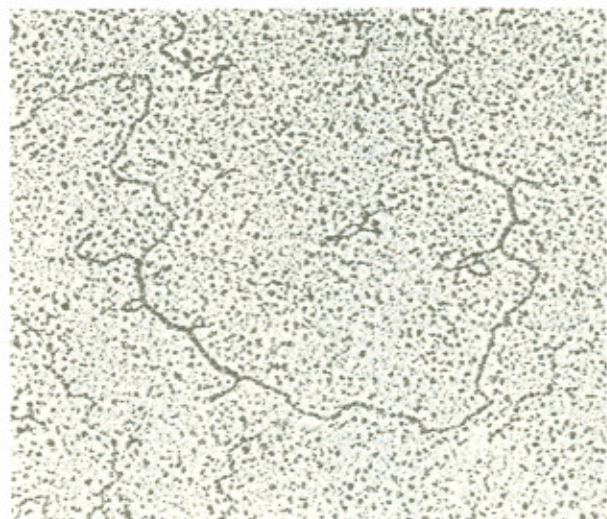


Figure 9. DNA-RNA hybrids formed between an embryonic C μ gene clone DNA and the minor α -chain mRNA from MOPC-315. In this picture, two molecules of α mRNA hybridized; one (probably the major one) hybridized to E1 and E2, the other (the minor one) hybridized to E4 and E5. The 5' region of the second mRNA, presumably complementary with E1 to E3, appears as a bush.

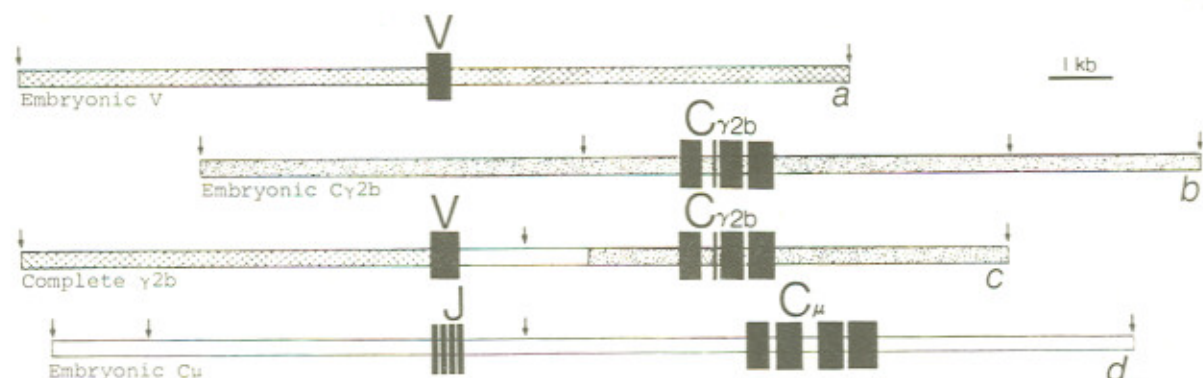


Figure 10. Four heavy-chain DNA clones used for the analysis of the complete $\gamma 2b$ gene of MOPC-141. (a) Embryonic DNA clone containing the germ-line V gene from which the V gene arose (clone PJ14); (b) embryonic $C_{\gamma 2b}$ gene clone MEP3; (c) complete $\gamma 2b$ gene clone M141-P21 isolated from myeloma MOPC-141; (d) embryonic C_{μ} gene clone MEP203. (↓) *EcoRI* cleavage sites; (■) exons. The regions of sequence homology among the four DNA clones are indicated by bars of different types of shading. Isolation and characterization of these clones have been described (Maki et al. 1980; Sakano et al. 1980).

sal Cys at position 95. Tyr (93), Tyr (94), Ala (96), and Arg (97) are also well conserved. We therefore believe that the coding by most, if not all, germ-line V genes ends in the immediate vicinity of triplet 97. If so, coding gaps of various lengths exist in the majority of the cases listed in Table 3. As suggested above, these residues must be encoded elsewhere in the genome. The gaps in the heavy chains of four myelomas, A4, E109, U61, and A47N, are the shortest; only the first letter of the Gly (98) is unaccounted for. Similarly, only several more bases are necessary to code for the HV3 of the J558 α chain. As discussed above, the 3' end of a given germ-line V_{κ} gene may not be fixed precisely and can shift within a range of several nucleotides (Max et al. 1979; Sakano et al. 1979a). If the same applies to the heavy-chain genes, the heavy chains with shorter HV3, such as those of A4, E109, U61, and J558, could be fully encoded by the respective germ-line V gene and the J DNA segment. It thus appears that in some cases the germ-line V DNA segment directly joins with the J DNA segment, as is the case in the light-chain genes. However, it is also possible that the assembly of the V_H gene always occurs in two steps but that in some cases the D segments are very short or do not even contribute to protein-encoding segments at all. At present, the location and the number of the D DNA segments in the germ-line genome are unknown. They may be at the 3' end of each of the multiple germ-line V genes or may be clustered between the V DNA segments and the J DNA segments. In either case, we assume that the order of the V, D, and J DNA segments in the germ-line genome is the same as in the completely assembled V genes, so that successive joining of the three types of DNA segments is possible by the looping-out of the intervening spacers (Fig. 15).

Exact boundaries of J_H DNA segments are unfixed. As previously described for the light-chain J DNAs, the exact 5' boundary of J_H DNA segments seems to be unfixed (Sakano et al. 1980). For instance,

in the case of J_{H4} , the coding of the MOPC-141 $\gamma 2b$ chain begins with the second base of the Leu (112) triplet UUG as indicated in Figures 11 and 12. In contrast, coding of the MOPC-21 $\gamma 1$ chain and the MOPC-173 $\gamma 2a$ chain by the same J_{H4} DNA segment seems to start with a Tyr triplet UAU and another Tyr triplet UAC, respectively (Table 3, Fig. 11). The first letters of the two Tyr triplets are seven and ten bases ahead (i.e., 5' side) of the first base used for the coding of the MOPC-141 J region. Similar examples for J_{H1} , J_{H2} , and J_{H3} are indicated in Figure 11 and Table 3.

The sequences that might be recognized by a putative recombinase. Sequences similar to the heptamer and nanomer preceding the light-chain J DNA segments are also present 5' to all four J_H DNA segments (Sakano et al. 1980), although deviation from the basic sequences seems to be slightly higher for the J_H segments than for the J_L ones (Table 1). As shown in Figure 12 and summarized in Table 1, the same or similar sequences are also present in the 3' flanking regions of the germ-line V_{M141} DNA. In the case of J_{H3} , two sequences, CA-CATG and CAATGTG, are related closely to the basic heptamer sequence CACTGTG. Similarly, two sequences, TATTGTG and TACTATG, are present near the 5' end of J_{H4} (Fig. 12). The presence of the multiple heptamers could contribute to modulation of the recombination site on J DNAs. Both the germ-line V_{M141} gene and all J_H segments show the recognition sequences separated by longer (23 bp) spacers. If the D segment contains recognition sequences with shorter (12 bp) spacers on each side, then V-D and D-J joinings would be mediated by recombinases having components equivalent to those acting on the light-chain genes. In this way, all recombinations leading to assembly of immunoglobulin V genes would follow a 12- and 23-bp spacer rule. This rule prohibits some unwanted recombinations, such as those between V_{λ} and J_{κ} and V_{κ} and J_{λ} . However, some other unwanted recombinations, such as those between V_H and J_{λ} , D and J_{κ} , and V_{λ} and



Figure 11. Nucleotide sequence of the 1.4-kb region of the embryonic clone MEP203 containing four J_H DNA segments (Sakano et al. 1980). The sequencing strategy is shown in Fig. 2A. The amino acid sequences encoded by these J DNA segments are shown in italics. The sequences closely related to the heptamer CACTGTG and nanomer GGTTTTGT present in the 5' region of each J DNA segment are underlined. The vertical lines indicate determined (solid lines) or predicted (broken lines) sites where coding by the J DNA segments begins in the various heavy-chain genes present in the indicated myelomas. The sites were determined by comparing the J DNA sequences with the sequences of the rearranged V DNA cloned from the respective myeloma (for details, see Sakano et al. 1980). The putative sites for the heavy-chain genes of S107, M167, M315, A4, J539, M173, and M21 were deduced by comparing the J peptide amino acid sequence of each heavy chain with the DNA sequence of the corresponding J DNA segment (Table 3). Predicted joining sites on J_{H1} DNA for M167 and S107 were confirmed by Early et al. (1980) by their direct sequencing of cDNAs for heavy-chain mRNAs prepared from respective myelomas.

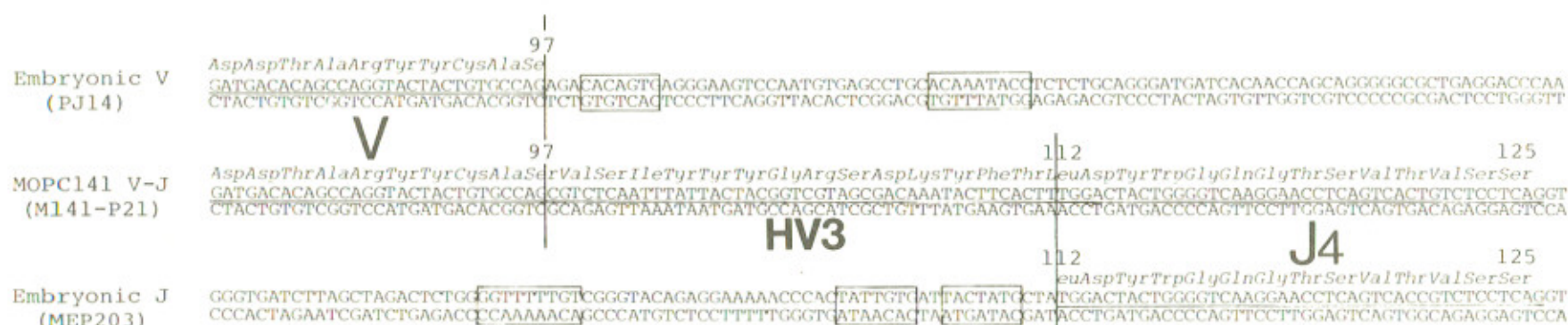


Figure 12. Nucleotide sequence around the V-J joining region of MOPC-141 $\gamma 2b$ gene (M141-P21) (Sakano et al. 1980). Sequences of its germ-line components (embryonic V; PJ14 and embryonic J; MEP203) are compared and possible recombination sites are indicated by vertical lines. Note that the sequence coding for HV3 is not present either on the embryonic V sequence or on the J sequence. Related sequences to two blocks of conserved sequences, CACTGTG and GGTTTTTGT, found near the V-J joining sites are boxed. Amino acid sequences encoded by these DNAs are in italics. Numbers indicate the positions of amino acid residues.

Table 3. Amino Acid Sequences of Mouse Immunoglobulin Heavy Chains

Myeloma	End of V	HV3 ^a	J	J _H DNA
M104E	***YYCAR	DY	dWYFDVWGAGTTVTSS	
J558	***-----	D	rY-----	
T15	***-----	DYYGS	sY-----	
M603	***-----	NYYGST	-----	
M167	***---T-	DADYGDSYF	g-----	J _{H1}
T601	***-----	LGYY	g-----	
S107	***-----	DYYGS	sY-----	
H8	***-----	DYYGN	sY-----	
M511	***-----	DGDYGS	sY-----	
X24	***YYCAR	LGYYG	YFDYWGQGTTTLTVSS	
M315	***---G	DNDHL	-----	
Hdex4	***-----	DK	n-----	J _{H2}
Hdex5	***-----	DS	n-----	
X44	***YYCAR	LHYYGYA	AYWGQGLTLTVSA	
J539	***-----	LHYYGYN	-----	
A4	***---TT		sWF-----	J _{H3}
E109	***H---TT		g-----	
U61	***---TT		g-----	
A47N	***---ST		g-----P	
M173	***YYCAR	SP	YYAMDYWGQGTSTVSS	
M21	***-----	HGNYPW	-----	J _{H4}
M141	***---S	VSIYYGRSDKYFT	l-----	

Amino acid sequences of mouse heavy chains around the carboxyl ends of the V regions. Amino acid sequences of mouse heavy chains for residues 90-113 are shown by one-letter codes (numbering is after Kabat et al. 1979). According to the J-region peptides, the chains are classified into four groups. The first amino acid residues in the J peptides are shown in small letters whenever the first letters of their codons are not included in the corresponding J DNA segments. The amino acid sequence of the MOPC-141 γ 2b chain was deduced from the nucleotide sequences determined by Sakano et al. (1980). Other sequences were taken from Bourgeois et al. (1972), Francis et al. (1974), Milstein et al. (1974), Rudikoff and Potter (1974, 1976), Vrana et al. (1978), Kehry et al. (1979), Rao et al. (1979), and Schilling et al. (1980). Nucleotide sequences for J_H DNAs are shown in Fig. 7.

^a HV3 stands for third hypervariable region.

V_κ could occur. This may be the reason why the three gene families are distributed in three different chromosomes.

Sequences at the Switch-recombination Sites

As described previously, the complete γ 2b gene clone from MOPC-141 (M141-P21) has homology with the J region of the embryonic C_μ gene clone (MEP203). We considered this homology on the γ 2b gene as a trace of the μ -gene expression before the μ - γ 2b switch and assume that the recombination for the switch took place between a pair of sites, one in the 5' flanking region of the C_{γ2b} gene and the other between the J DNA and the C_μ gene. To understand this novel somatic recombination better, we constructed restriction enzyme cleavage maps (Fig. 13B) for three DNA clones, M141-P21 (complete γ 2b), MEP3 (embryonic C_{γ2b}), and MEP203 (embryonic C_μ), and determined nucleotide sequences around the recombination sites (Fig. 14) (Sakano et al. 1980). The maps show that switch-recombination occurs between a pair of sites, one located within the 750-bp HindIII-HindIII segment of the germ-line C_μ gene clone and the other within the 200-bp EcoRI-BamHI segment of the germ-line C_{γ2b} gene clone, and generates the 650-bp HindIII-BamHI segment present in the complete γ 2b gene clone (see Fig. 13B). As expected, the 5' portion of the M141-P21 sequence is the same as the sequence of the germ-line C_μ gene clone MEP203, and the 3' portion is identical to the sequence of the germ-line C_{γ2b}

gene clone MEP3. The exact μ - γ 2b switch-recombination sites are indicated by a vertical line in Figure 14. We propose to refer to the μ - γ 2b switch sites on the germ-line C_μ gene and the germ-line C_{γ2b} gene as S_{μ(γ2b)} and S_{γ2b(μ)}, respectively.

Are there any characteristic sequences around the μ - γ 2b switch sites that might be considered as part of the recognition signal for the recombinase? As shown in Figure 14, two short blocks of sequences, TCCTGG and AGA, present in front of (i.e., toward the 5' side of) S_{μ(γ2b)} are also present near the S_{γ2b(μ)} site in the equivalent positions. The two sets of sequences are in the same orientation relative to the direction of transcription and thus different from those sequences conserved near the V-J joining sites not only in sequence content but also in relative orientation. Although it is not clear whether these sequences are indeed recognized by the recombinase, the results suggest that the enzymatic mechanism for the μ - γ 2b switch-recombination is quite different from that for the V-J joining.

Recently, Davis et al. (1980) isolated a complete α heavy-chain gene clone from MOPC-603 that carries both V_H and C_α genes. Heteroduplex analysis of this clone against the germ-line C_α and C_μ gene clones suggested that the complete α heavy-chain gene is also generated by a switch-recombination that involves the 5' flanking region of the germ-line C_μ gene. The S_{μ(α)} site was mapped between two HindIII sites, at 2.2 kb and 1.3 kb from the C_μ gene (see Fig. 13Bc). Thus, the μ - α and μ - γ 2b switch-recombinations seem to use different sites



Figure 14. Nucleotide sequences around the $\mu(\gamma 2b)$ switch sites in clones MEP203, MEP3, and M141-P21 (Sakano et al. 1980). Sequencing strategy is shown in Fig. 2B. The recombination sites are indicated by a vertical line designated S. The two blocks of sequences commonly present in the 5' side of the recombination sites of embryonic μ and $\gamma 2b$ gene clones are in boxes.

in the 5' flanking region of the germ-line C_{μ} gene. This suggests that there may be a number of different switch sites in this region, possibly at least one for each of the six heavy-chain classes or subclasses. Recently, this notion has been confirmed by heteroduplex analysis of various heavy-chain gene clones from different myelomas (A. Trautnecker and S. Tonegawa, unpubl.).

DISCUSSION

Two Types of Somatic Recombination

Structural analysis of immunoglobulin genes demonstrated that two different types of somatic DNA rearrangements are necessary for the generation of complete immunoglobulin genes, namely, V-J or V(D)J joining both for light- and heavy-chain genes and switch-recombination for heavy-chain genes (see Fig. 15 for organization of heavy-chain genes). These two types of recombination are clearly different in several aspects, although both of them apparently accompany the deletion of DNA segments (Honjo and Kataoka 1978; Sakano et al. 1979a; Coleclough et al. 1980; Corey and Adams 1980; Rabbitts et al. 1980). Below, we discuss the two types of recombination with respect to function, structure, and evolution.

V-J or V(D)J joining. Ontogenically, V(D)J or V-J joining takes place early in B-cell differentiation. The earliest recognizable cells in the B-lymphocyte lineage already contain a complete μ chain but no light chain. Thus, it would be expected that the heavy-chain V(D)J joining occurs before pre-B cells emerge. Indeed, our recent analysis of two types of cell lines that exhibit characteristics of pre-B cells indicates that in these cells V(D)J joining had occurred in the heavy-chain genes but not necessarily in the light-chain genes. Thus, it appears that assembly of the heavy-chain V gene segments precedes that of light-chain V gene segments (Maki et al. 1980a), although both events should be completed before the immature B cells with complete IgM molecules emerge from pre-B cells. The molecular mechanism for this asynchronous assembly of V_L and V_H genes is unknown. It may be that the two chains have their own recombinase, despite the apparent similarity in the presumed recognition sequences, and that synthesis and/or activity of these enzymes is regulated in the early stages of B-cell differentiation. Alternatively, the two gene systems may use a common recombinase, and the higher-order structure of the DNA in chromatin is controlled developmentally so that gene segments for a given type of chain become accessible to the recombinase at a defined stage of B-cell differentiation.

V-J joining generates a complete V gene and commits the B-cell precursor to expression of that particular V gene during the course of differentiation that follows. It is thus a key event in the determination of the antigen specificity of the descendant B cells. The antibody diversity encoded in the germ-line V genes can be amplified by this process in three different ways. First, differ-

ent combinations of a germ-line V_L gene and a germ-line V_H gene are selected for recombination from the respective gene pool in different B-cell clones. Second, for a given chain, different combinations of V, D, and J DNA segments are recombined. In the case of the heavy chain, in which a separate D DNA segment participates in the recombination, it is certain that the antibody diversity is amplified by the random assortment. In contrast, it is not yet proved beyond a doubt that the J DNA segment contributes to antibody specificity. However, since the amino acid residue encoded by the first codon of J DNA segments is either within or immediately adjacent to the HV3, a J peptide may affect as well the shape of the antigen-combining site. Finally, the exact joining ends of the V_{κ} , J_{κ} , and J_H DNA segments seem to be different from one recombination event to another. We expect that a similar flexibility applies to the joining ends of V_H and the hypothetical D_H DNA segments.

Although the aforementioned three mechanisms can enormously amplify the germ-line-encoded repertoire of antibody diversity, they do not account for the somatic diversification observed in the rest of the V region, namely, in HV1 and HV2, which also participate in antigen recognition. On the basis of the difference observed in the diversity of the mouse V_{H1} regions and their germ-line V gene (one copy per haploid), it was concluded previously (Tonegawa 1976; Bernard et al. 1978; Brack et al. 1978) that somatic mutations must occur in all three hypervariable regions. This concept is being verified in mouse κ light chains (G. Heinrich, unpubl.). Our recent analysis of germ-line V_{κ} genes coding for members of the $V_{\kappa 21}$ group (Potter 1977) strongly suggests that V_{κ} regions grouped into a single subgroup are encoded in V genes that are derived from a common germ-line V gene by introduction of somatic mutations mainly, but perhaps not entirely, into the hypervariable regions. How and when these mutations occur in the germ-line V genes is totally unknown. Kabat et al. (1978) proposed the "mini-gene theory" to explain the apparently independent assortment of the three hypervariable regions and the four framework regions. Although this proposal was borne out for HV3 of heavy chains (D region) and the fourth framework regions of both light and heavy chains (J_L and J_H), the available evidence strongly argues against the idea that the same applies to the rest of the V regions. Alternatively, Seidman et al. (1979) promoted unequal recombinations occurring between closely related germ-line V genes as a primary source of V gene diversification, both in evolution and in development. Unfortunately, no evidence has been obtained that supports this hypothesis.

How did the V-J or V(D)J joining arise in evolution? Three lines of evidence argue for the hypothesis that V, D, and J DNA segments were contiguous in the primordial gene coding for one immunoglobulin domain unit and that they were separated by insertion of an extraneous DNA sequence. First, a complete V region is similar in size to one C-region domain unit and exhibits residual homology. In particular, the J region is homologous to

the carboxyterminal region of the C_L region (Hill et al. 1966). Second, as pointed out earlier, the spacers separating the germ-line V, D, and J DNA segments and prokaryotic insertion elements are somewhat similar in terminal structure. Finally, in both immunoglobulin genes and the prokaryotic insertion elements, an imprecise excision is a major manifestation of the recombination events. If the hypothesis is correct, then V-J or V(D)J joining may be considered as a reversal of an ancient DNA insertion event that occurred accidentally within a primordial V gene but was fixed and subsequently exploited under the pressure to increase the antibody repertoire of the organism. In contrast to the heavy-chain genes, V and D DNA segments of light-chain genes are contiguous in the germ-line genome. We thus believe that the hypothetical splitting of the primordial V gene occurred in two separate steps: Splitting of J DNA segments occurred before the V_L and V_H genes diverged, whereas that of D DNA segments occurred after they diverged. It has been shown that antibody diversity is much more limited in amphibians than in mammals (Du Pasquier and Wabl 1978). It would be interesting to see whether V_H and D_H segments are contiguous in lower vertebrates.

Switch-recombination. In contrast to the V-J or V(D)J joining, switch-recombinations occur after the pre-B cells develop into B cells bearing complete Ig molecules on the surface. Although the IgM-IgD double-bearing cells are the most common progenitors for switched cells, there is evidence suggesting that IgM single-bearing cells can also switch under an appropriate circumstance (Zan-Bar et al. 1979).

A switch-recombination replaces the C_μ exons of a complete μ -chain gene with exons coding for the C regions of other classes or subclasses except for the C_δ (see the following for expression of δ genes). The purpose here is to change the effector functions localized in the heavy-chain C regions without alteration of antigen specificity determined by the V region. Thus, in contrast to the V-J or V(D)J joining that occurs at the margins of protein-encoding DNA segments, switch-recombination takes place in regions that do not encode proteins. It can occur, a priori, anywhere within the 7.5-kb intron separating the J DNA segments and the C_μ exons on the one hand, and anywhere within a several kilobase region located 5' to the respective C exon of other classes on the other, as long as the transcription unit on the recombined DNA can encompass the entire V and C exons. This suggests the possibility that switch-recombination sites may be scattered widely in these regions. Indeed, the $S_{\mu(\gamma 2b)}$ site for the MOPC-141 $\gamma 2b$ chain and the $S_{\mu(\alpha)}$ site for the MOPC-603 α chain are at least 500 bp apart. Furthermore, recent analysis in our laboratory as well as by others (see Honjo et al.; Hood et al.; both this volume) of several heavy-chain gene clones showed that there is at least one switch site for each class or subclass in the intron between the J and C_μ exons and that in some cases there are multiple sites for a single class or subclass. In the region approximately 4.5 kb 5' to the C_μ

gene, a pentameric sequence, CTGAG, or its close variants, is repeated in tandem more than 100 times (Dunnick et al. 1980; Kataoka et al. 1980; Sakano et al. 1980). The same pentameric sequences are prominent in the 5' flanking region of the C_α gene (H. Sakano, unpubl.). Since μ - α sites fall within these regions (Davis et al. 1980; and our unpubl. results), this direct homology between C_μ and C_α DNAs might play an essential role in μ - α switch-recombination. Sequence homology between the switch regions of C_μ and $C_{\gamma 2b}$ genes is more limited, but we did observe short blocks of directly repeated sequences. Whether the switch to various classes or subclasses is actively controlled at the level of the nucleotide sequence will be clarified by cloning and sequence analysis of various heavy-chain genes.

All classes and subclasses of C_H genes are considered to have arisen by duplication from a primordial C_H gene and its flanking sequences (Dayhoff 1972). We believe that the switch-recombination regions have most likely arisen from the duplicated 5' flanking sequences by making use of the sequence homology among them. On the basis of the correlation between exons and functional units within an immunoglobulin light chain (Bernard et al. 1978; Tonegawa et al. 1978) and a heavy chain (Sakano et al. 1979b), we previously speculated that some higher genes were created in evolution by assembly of exons, each encoding a polypeptide chain of some functional use. The assembly was thought to be mediated by homology in the flanking intron sequences. The idea has also been discussed by other authors (Darnell 1978; Crick 1979). Although the hypothesis has yet to be tested experimentally, we note that a striking similarity exists between the postulated process and the switch-recombination, although the latter occurs in development rather than in evolution.

Somatic Reorganization of DNA Sequences as Determination in Differentiation

How the higher organism generally uses reorganization of DNA sequences as a means for cell determination in differentiation is an intriguing question. Recent studies revealed several cases in which 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 (Zieg and Simon 1980). Mating-type interconversion of homothallic yeast has been linked with alternative insertion of two different mobile DNA sequences (Klar et al., this volume). Pathogenic African trypanosomes evade the immune systems of their mammalian hosts by the sequential expression of alternative cell-surface glycoproteins (Hoeijmakers et al. 1980). Recent studies suggest that DNA rearrangement is responsible for this phenomenon. One interesting distinction between some of the aforementioned examples and the immunoglobulin gene system can be seen in the mode of DNA rearrangement. Whereas the phase variation in

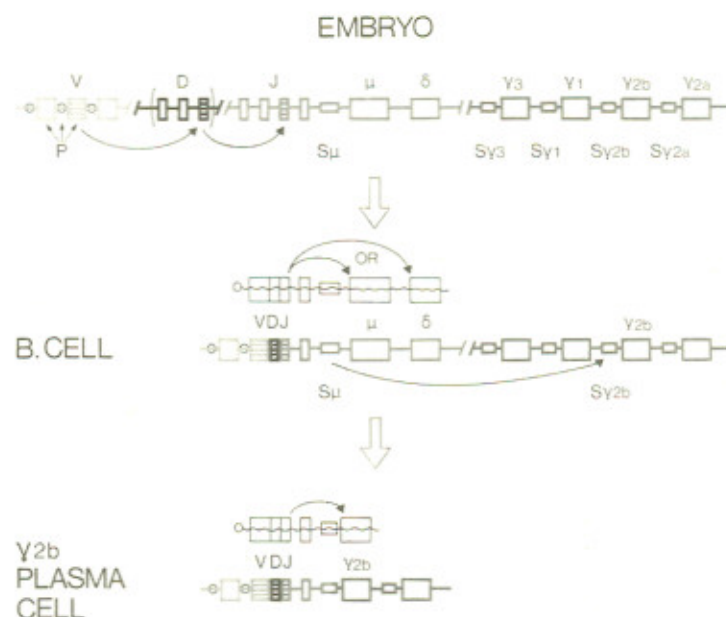


Figure 15. Arrangement of various parts of Ig heavy-chain gene segments in embryos, B cells (IgM-IgD double producers), and $\gamma 2b$ -chain-synthesizing plasma cell. The order and spacing of the four C_γ genes were determined by analysis of a large number of overlapping DNA fragments cloned in Charon 4A phage vector. Distances between the genes are: $C_{\gamma 3}$ -35 kb- $C_{\gamma 1}$ -22 kb- $C_{\gamma 2b}$ -15 kb- $C_{\gamma 2a}$ (W. Roeder, unpubl.). S_μ , $S_{\gamma 3}$, etc., designate approximate positions of switch-recombination sites. (○) Presumed locations of promoters; (—) primary transcripts, (↔) either DNA recombination or RNA splicing events. The number and locations of D DNA segments are hypothetical.

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. Many 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 the case of immunoglobulin genes. We have yet to find a second example of a gene(s) whose behavior reflects an apparently irreversible developmental phenomenon. However, no more than a few gene systems have been studied in the same detail as the immunoglobulin genes. Further studies are necessary before we can decide whether the strategy employed by the immunoglobulin genes is restricted to this gene system.

Control of RNA Processing in Cell Differentiation

Analysis of the μ - and α -chain genes coding for the secreted and the membrane-bound forms of the respective chain revealed that gene activity is regulated at the level of RNA processing in cell differentiation. One can envisage two major possibilities with respect to the mechanism. In one, the control occurs at the poly(A) attachment sites. In this hypothesis, the entire C gene, including the extra exons, can be transcribed into a single mRNA, both in immunoglobulin-bearing and in immunoglobulin-secreting cells. Although in the former cells poly(A) attached predominantly to a site 3' to the last exon (exon 6 and exon 5 for a μ and α chain, respectively), in the latter cells it attached to a site 3' to exon 4 (μ) or exon 3 (α). In the other hypothesis, regulation at the level of RNA-splicing sites must be invoked. Obviously, further analysis is necessary to determine which of these possibilities is correct.

Our recent studies on expression of δ -chain genes

strongly suggest that control at the level of RNA processing is not unique to the membrane versus the secreted forms of immunoglobulin chains. Unlike γ -, α -, or ϵ -chain genes, the δ -chain gene is mostly expressed simultaneously with the μ -chain gene, and no μ - δ switch occurs when B cells mature to plasma cells. Recently, F. Blattner and co-workers (pers. comm.) mapped the C_δ gene a few kilobases 3' to exon 6 of the C_μ gene. We have recently analyzed organization of μ - and δ -chain genes in a B-cell hybridoma line, GCL28, that bears both IgM and IgD molecules on the surface. These analyses indicate that both chains share the same copy of the assembled V gene and that no sequence arrangement exists in the entire region encompassing the complete V gene, C_μ gene, and C_δ gene. On the basis of these results, we conclude, as was suggested earlier (Tonegawa et al. 1978), that in IgM-IgD double-bearing cells, the primary transcript containing both μ - and δ -gene sequences is processed such that two mature mRNAs of a composition V, D, J, C_μ and V, D, J, C_δ are generated (R. Maki and A. Trautenecker, unpubl.). The conclusions drawn from these studies are illustrated schematically in Figure 15.

CONCLUDING REMARKS

Molecular analysis of immunoglobulin genes has contributed in several aspects to our understanding of higher-cell genes in differentiation and evolution. In particular, it revealed the dynamic nature of DNA in both of these processes. Proteins are encoded in the genome in discrete pieces that often carry their own functions. These building blocks are shuffled and appear in different combinations in ontogeny and perhaps in phylogeny. During cell differentiation, expression of higher-cell genes can be controlled both by reorganization of the DNA sequences and by differential use of the RNA transcripts.

ACKNOWLEDGMENTS

We are thankful for the technical assistance of Ms. P. Riegert, Ms. L. Sultzman, Ms. L. Gibson, Ms. C. Olsson, Ms. L. Angman, Ms. D. Norman, and Mr. G. Dastoor-nikoo.

REFERENCES

- ALWINE, J. C., D. J. KEMP, and G. R. STARK. 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzylmethyl-paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci.* **74**: 5354.
- BERGMAN, Y. and J. HAIMOVICH. 1978. B lymphocytes contain three species of μ chains. *Eur. J. Immunol.* **8**: 876.
- BERNARD, O., N. HOZUMI, and S. TONEGAWA. 1978. Sequences of mouse immunoglobulin light chain genes before and after somatic changes. *Cell* **15**: 1133.
- BRACK, C. and S. TONEGAWA. 1977. Variable and constant parts of the immunoglobulin light chain gene of a mouse myeloma cell are 1250 nontranslated bases apart. *Proc. Natl. Acad. Sci.* **74**: 5652.
- BRACK, C., M. HIRAMA, R. SCHULLER, and S. TONEGAWA. 1978. A complete immunoglobulin gene is created by somatic recombination. *Cell* **15**: 1.
- BOURGEOIS, A., M. FOUGEREAU, and C. DEPREVAL. 1972. Sequences of amino acids of the NH₂-terminal region of a mouse clonal immunoglobulin heavy chain. *Eur. J. Immunol.* **24**: 446.
- BURROWS, P., M. LEJEUNE, and J. F. KEARNEY. 1979. Evidence that murine pre-B cells synthesize μ heavy chains but no light chains. *Nature* **280**: 838.
- CALAME, K., J. ROGERS, P. EARLY, M. DAVIS, D. LIVANT, R. WALL, and L. HOOD. 1980. Mouse C_{μ} heavy chain immunoglobulin gene segment contains three intervening sequences separating domains. *Nature* **284**: 452.
- COLECLOUGH, C., D. COOPER, and R. P. PERRY. 1980. Rearrangement of immunoglobulin heavy chain genes during B-lymphocyte development as revealed by studies of mouse plasmacytoma cells. *Proc. Natl. Acad. Sci.* **77**: 1422.
- COOPER, M. D., J. F. KEARNEY, P. M. LYDYARD, C. E. GROSSI, and A. R. LAWTON. 1977. Studies of generation of B-cell diversity in mouse, man and chicken. *Cold Spring Harbor Symp. Quant. Biol.* **41**: 139.
- CORY, S. and J. M. ADAMS. 1980. Deletions are associated with somatic rearrangement of immunoglobulin heavy chain genes. *Cell* **19**: 37.
- CRICK, F. 1979. Split genes and RNA splicing. *Science* **204**: 264.
- DARNELL, J. E. 1978. Implications of RNA-RNA splicing in evolution of eukaryotic cells. *Science* **202**: 1257.
- DAVIS, M. M., K. CALAME, P. W. EARLY, D. L. LIVANT, R. JOHO, I. L. WEISSMAN, and L. HOOD. 1980. An immunoglobulin heavy-chain gene is formed by at least two recombination events. *Nature* **283**: 733.
- DAYHOFF, M. O. 1972. *Atlas of protein sequence and structure*, vol. 5. National Biomedical Research Foundation, Washington, D.C.
- DUNNICK, W., T. H. RABBITTS, and C. MILSTEIN. 1980. An immunoglobulin deletion mutant with implications for the heavy-chain switch and RNA splicing. *Nature* **286**: 669.
- DUPASQUIER, L. and M. R. WABL. 1978. Antibody diversity in amphibians: Inheritance of isoelectric focussing antibody patterns in isogenic frogs. *Eur. J. Immunol.* **8**: 428.
- EARLY, P. W., M. M. DAVIS, D. B. KABACK, N. DAVIDSON, and L. HOOD. 1979. Immunoglobulin heavy chain gene organization in mice: Analysis of a myeloma genomic clone containing variable and α constant regions. *Proc. Natl. Acad. Sci.* **76**: 857.
- EARLY, P., H. HUANG, M. DAVIS, K. CALAME, and L. HOOD. 1980a. An immunoglobulin heavy chain variable region gene is generated from three segments of DNA; V_H, D and J_H. *Cell* **19**: 981.
- EARLY, P., J. ROGERS, M. DAVIS, K. CALAME, M. BOND, R. WALL, and L. HOOD. 1980b. Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. *Cell* **20**: 331.
- EDELMAN, G. M., B. A. CUNNINGHAM, W. GALL, P. GOTTLIEB, U. RUTIDHAUSER, and M. WAXDAL. 1969. The covalent structure of an entire γ G immunoglobulin molecule. *Proc. Natl. Acad. Sci.* **63**: 78.
- FRANCIS, S. H., R. G. Q. LESLIE, L. HOOD, and H. N. EISEN. 1974. Amino-acid sequence of the variable region of the heavy (alpha) chain of a mouse myeloma protein with antihapten activity. *Proc. Natl. Acad. Sci.* **71**: 1123.
- FU, S. M., R. J. WINCHESTER, and H. G. KUNKEL. 1975. Similar idiotype specificity for the membrane IgD and IgM of human B lymphocytes. *J. Immunol.* **114**: 250.
- GODING, J. M. and L. E. LAYTON. 1976. Antigen-induced collapsing of IgM and IgD-like receptors on murine B cells. *J. Exp. Med.* **144**: 852.
- GOUGH, N. M., D. J. KEMP, B. M. TYLER, J. M. ADAMS, and S. CORY. 1980. Intervening sequences divide the gene for the constant region of mouse immunoglobulin μ chains into segments, each encoding a domain. *Proc. Natl. Acad. Sci.* **77**: 554.
- GRINDLEY, N. D. F. 1978. DNA sequences at the integration sites of the insertion element IS1. *Cell* **13**: 419.
- HILL, R. L., R. DELANEY, R. E. FELLOWS, and H. E. LEBOVITZ. 1966. The evolutionary origins of the immunoglobulins. *Proc. Natl. Acad. Sci.* **56**: 1762.
- HOEIJMAKERS, J. H. J., A. C. C. FRASCH, A. BERNARDS, P. BORST, and G. A. M. CROSS. 1980. Novel expression-linked copies of the genes for variant surface antigens in trypanosomes. *Nature* **284**: 78.
- HONJO, T. and T. KATAOKA. 1978. Organization of immunoglobulin heavy chain genes and allelic deletion model. *Proc. Natl. Acad. Sci.* **75**: 2140.
- HONJO, T., M. OBATA, Y. YAMAWAKI-KATAOKA, T. KATAOKA, T. KAWAKAMI, N. TAKAHASHI, and Y. MANO. 1979. Cloning and complete nucleotide sequence of mouse immunoglobulin γ_1 chain gene. *Cell* **18**: 559.
- IUPAC-IUB Commission on Biochemical Nomenclature. 1969. A one-letter notation for amino acid sequences. Tentative rules. *Biochem. J.* **113**: 1.
- KABAT, E. A., T. T. WU, and H. BILOFSKY. 1978. Variable region genes for immunoglobulin framework are assembled from small segments of DNA—A hypothesis. *Proc. Natl. Acad. Sci.* **75**: 2429.
- . 1979. *Sequences of immunoglobulin chains*. National Institutes of Health, Bethesda, Maryland.
- KAMP, D., L. T. CHOW, T. R. BROKER, D. KWON, D. ZIPSER, and T. KAHMANN. 1979. Site-specific recombination in phage Mu. *Cold Spring Harbor Symp. Quant. Biol.* **43**: 1159.
- KATAOKA, T., T. KAWAKAMI, N. TAKAHASHI, and T. HONJO. 1980. Rearrangement of immunoglobulin γ_1 -chain genes and mechanism for heavy-chain class switch. *Proc. Natl. Acad. Sci.* **77**: 919.
- KEHRY, M., C. SIBLEY, J. FUHRMANN, J. SCHILLING, and L. HOOD. 1979. Amino acid sequence of a mouse immunoglobulin μ chain. *Proc. Natl. Acad. Sci.* **76**: 2932.
- LENHARD-SCHULLER, R., B. HOHN, C. BRACK, M. HIRAMA, and S. TONEGAWA. 1978. DNA clones containing mouse immunoglobulin κ chain genes isolated by in vitro packaging into phage λ coats. *Proc. Natl. Acad. Sci.* **74**: 4709.
- LEVITT, D. and M. D. COOPER. 1980. Mouse pre-B cells synthesize and secrete μ heavy chains but not light chains. *Cell* **19**: 617.
- MAKI, R., J. KEARNEY, C. PAIGE, and S. TONEGAWA. 1980a. Immunoglobulin gene rearrangement in immature B cells. *Science* **209**: 1366.

- MAKI, R., A. TRAUNECKER, H. SAKANO, W. ROEDER, and S. TONEGAWA. 1980b. Exon shuffling generates an immunoglobulin heavy chain gene. *Proc. Natl. Acad. Sci.* 77: 2138.
- MAX, E. E., J. G. SEIDMAN, and P. LEDER. 1979. Sequences of five recombination sites encoded close to an immunoglobulin κ constant region gene. *Proc. Natl. Acad. Sci.* 76: 3450.
- McKEAN, D. J., M. BELL, and M. POTTER. 1978. Mechanisms of antibody diversity: Multiple genes encode structurally related mouse κ variable regions. *Proc. Natl. Acad. Sci.* 75: 3913.
- MELCHER, U. and J. W. UHR. 1976. Cell surface immunoglobulin. XVI. Polypeptide chain structure of mouse IgM and IgD-like molecule. *J. Immunol.* 116: 409.
- MILSTEIN, C., K. ADETUGBO, N. J. COWAN, and D. S. SECKER. 1974. Clonal variants of myeloma cells. *Prog. Immunol.* 2: 157.
- OHTSUBO, H. and E. OHTSUBO. 1978. Nucleotide sequence of an insertion element, IS1. *Proc. Natl. Acad. Sci.* 75: 615.
- PADLAN, E. A., D. M. SEGAL, T. F. SPANDE, D. R. DAVIES, S. RUDIKOFF, and M. POTTER. 1973. Structure of 4.5 Å resolution of a phosphorylcholine-binding Fab. *Nat. New Biol.* 245: 165.
- PERNIS, B., L. FORNI, and A. L. LUZZATI. 1977. Synthesis of multiple immunoglobulin classes by single lymphocytes. *Cold Spring Harbor Symp. Quant. Biol.* 41: 175.
- POLJAK, R., L. AMZEL, H. AVEY, B. CHEN, R. PHIZACKERLEY, and F. SAUL. 1973. Three-dimensional structure of the Fab' fragment of a human immunoglobulin at 2.8-Å resolution. *Proc. Natl. Acad. Sci.* 70: 3305.
- POTTER, M. 1977. Antigen-binding myeloma proteins of mice. *Adv. Immunol.* 25: 141.
- RABBITS, T. H., A. FORSTER, W. DUNNICK, and D. L. BENTLEY. 1980. The role of gene deletion in the immunoglobulin heavy chain switch. *Nature* 283: 351.
- RAFF, M. D., M. MEGAN, J. J. T. OWEN, and M. D. COOPER. 1976. Early production of intracellular IgM by B lymphocyte precursors in the mouse. *Nature* 259: 224.
- RAO, D. N., S. RUDIKOFF, H. KRUTZSCH, and M. POTTER. 1979. Structural evidence for independent joining region gene in immunoglobulin heavy chains from anti-galactan myeloma proteins and its potential role in generating diversity in complementarity-determining regions. *Proc. Natl. Acad. Sci.* 76: 2890.
- RASCHKE, W. C. 1978. Expression of murine IgM, IgD and Ia molecules on hybrids of murine LPS blasts with a Syrian hamster B lymphoma. *Curr. Top. Microbiol. Immunol.* 81: 70.
- ROGERS, J., P. EARLY, C. CARTER, K. CALAME, M. BOND, L. HOOD, and R. WALL. 1980. Two mRNAs with different 3' ends encode membrane-bound and secreted forms of immunoglobulin μ chain. *Cell* 20: 303.
- RUDIKOFF, S. and M. POTTER. 1974. Variable region sequence of the heavy chain from a phosphorylcholine-binding myeloma protein. *Biochemistry* 13: 4033.
- . 1976. Size differences among immunoglobulin heavy chains from phosphorylcholine-binding proteins. *Proc. Natl. Acad. Sci.* 73: 2109.
- SAKANO, H., K. HUPPI, G. HEINRICH, and S. TONEGAWA. 1979a. Sequences at the somatic recombination sites of immunoglobulin light chain genes. *Nature* 280: 288.
- SAKANO, H., R. MAKI, Y. KUROSAWA, W. ROEDER, and S. TONEGAWA. 1980. Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy chain genes. *Nature* 286: 676.
- SAKANO, H., J. H. ROGERS, K. HUPPI, C. BRACK, A. TRAUNECKER, R. MAKI, R. WALL, and S. TONEGAWA. 1979b. Domains and the hinge region of an immunoglobulin heavy chain are encoded in separate DNA segments. *Nature* 277: 627.
- SALSANO, R., S. S. FRÖLAND, J. B. NATVIG, and T. E. MICHAELSEN. 1974. Same idiotype of B-lymphocyte membrane IgD and IgM. Formal evidence for monoclonality of chronic lymphocyte leukemia cells. *Scand. J. Immunol.* 3: 841.
- SCHILLING, J., B. CLEVERING, J. M. DAVIE, and L. HOOD. 1980. Amino acid sequence of homogeneous antibodies to dextran and DNA rearrangements in heavy chain V-region gene segments. *Nature* 283: 35.
- SEIDMAN, J. G. and P. LEDER. 1978. The arrangement and rearrangement of antibody genes. *Nature* 276: 790.
- SEIDMAN, J. G., E. E. MAX, and P. LEDER. 1979. A κ -immunoglobulin gene is formed by site-specific recombination without further somatic mutation. *Nature* 280: 370.
- SEIDMAN, J. G., A. LEDER, M. NAU, B. NORMAN, and P. LEDER. 1978. Antibody diversity. *Science* 202: 11.
- SMITH, H. O. and M. L. BIRNSTIEL. 1976. A simple method for DNA restriction mapping. *Nucleic Acids Res.* 3: 2387.
- TONEGAWA, S. 1976. Reiteration frequency of immunoglobulin light chain genes: Further evidence for somatic generation of antibody diversity. *Proc. Natl. Acad. Sci.* 73: 203.
- TONEGAWA, S., C. BRACK, N. HOZUMI, and R. SCHULLER. 1977a. Cloning of an immunoglobulin variable region gene from mouse embryo. *Proc. Natl. Acad. Sci.* 74: 5652.
- TONEGAWA, S., N. HOZUMI, G. MATTHYSSENS, and R. SCHULLER. 1977b. Somatic changes in the content and context of immunoglobulin genes. *Cold Spring Harbor Symp. Quant. Biol.* 41: 877.
- TONEGAWA, S., A. MAXAM, R. TIZARD, O. BERNARD and W. GILBERT. 1978. Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain. *Proc. Natl. Acad. Sci.* 75: 1485.
- TOOZE, J., ed. 1980. *The molecular biology of tumor viruses: DNA tumor viruses*, 2nd Ed., Part 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- TUCKER, P. W., K. B. MARCU, M. NEWELL, J. RICHARDS, and F. R. BLATTNER. 1979. Sequence of the cloned gene for the constant region of murine γ_{2b} immunoglobulin heavy chain. *Science* 206: 1303.
- VASSALLI, P., R. TEDGHI, B. LISOWSKA-BERNSTEIN, A. TARTAKOFF, and J. JATON. 1979. Evidence for hydrophobic region with heavy chains of mouse B lymphocyte membrane-bound IgM. *Proc. Natl. Acad. Sci.* 76: 5525.
- VRANA, M., S. RUDIKOFF, and M. POTTER. 1978. Sequence variation among heavy chains from insulin-binding myeloma proteins. *Proc. Natl. Acad. Sci.* 75: 1957.
- WEIGERT, M., L. GATMAITAN, E. LOH, J. SCHILLING, and L. HOOD. 1978. Rearrangement of genetic information may produce immunoglobulin diversity. *Nature* 276: 785.
- YAMAWAKI-KATAOKA, Y., T. KATAOKA, N. TAKAHASHI, M. OBATA, and T. HONJO. 1980. Complete nucleotide sequence of immunoglobulin γ_{2b} chain gene cloned from newborn mouse DNA. *Nature* 283: 786.
- ZAN-BAR, I., S. STROBER, and E. S. VITETTA. 1979. The relationship between surface immunoglobulin isotype and immune function of murine B lymphocytes. *J. Immunol.* 123: 925.
- ZIEG, J. and M. SIMON. 1980. Analysis of the nucleotide sequence of an invertible controlling element. *Proc. Natl. Acad. Sci.* 77: 4196.