Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes

Hitoshi Sakano, Richard Maki, Yoshikazu Kurosawa, William Roeder & Susumu Tonegawa

Basel Institute for Immunology, 487 Grenzacherstrasse, Postfach CH 4005, Basel 5, Switzerland

At least two types of somatic recombination are necessary for the generation of a complete immunoglobulin γ2b gene from germ-line DNA sequences. The first type of recombination consists of the assembly of three separate DNA segments, each encoding a different part of the variable region. The second type of recombination replaces the exons coding for the constant region of the μ chain with those coding for the same region of the γ2b chain. The DNA sequencing studies suggest that the two types of recombination operate by different mechanisms.

COMPLETE, active immunoglobulin genes are created by somatic recombination that occurs during the differentiation of lymphocyte precursor cells[1-3]. The organization of the gene sequences before and after somatic recombination has been studied extensively for both λ- and κ-type light chains in the mouse[3-10]. These studies established that in the embryonic genome, the major portion of the conventionally defined variable (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. The J DNA segment has been mapped a few kilobases upstream of (that is, 5′ side, with respect to the direction of transcription) the single-copy DNA segment (C DNA) coding for the constant (C) region. In the myeloma cells in which the light-chain gene in question is active, the V DNA segment and the J DNA segment are contiguous as a consequence of a recombination that apparently accompanies deletion of the DNA sequence occurring between the V and J DNA segments in the germ-line genome[7]. The entire region of the chromosome containing the V DNA segment, J DNA segment, J–C intron and C DNA segment is transcribed into a single RNA molecule from which a mature messenger RNA is generated by RNA splicing[11,12]. The V–J joining step is considered to be a key event in the activation of the immunoglobulin gene. It may also contribute to the increase of the organism's capacity to synthesize a large number of different antibodies[7,13-15].

As the heavy-chain polypeptide is also composed of a V and C region, one might assume that what has been established at the DNA level for the light chains would also be true for heavy chains. However, the expression of heavy chains has some features which make it considerably more complex than that of light chains. For instance, unlike the light chains, there are in the mouse at least eight different C regions, each of which seems to share the same set of V regions, that is, a given V gene can be expressed with more than one heavy-chain class or subclass. In addition, it seems 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[16-18].

Recently, we reported on the structure of the complete γ2b gene isolated from a myeloma, MOPC141, an IgG2b secretor[19]. These studies showed that, unlike the light-chain genes, at least two recombinations are necessary to generate the heavy-chain gene. 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 used for the MOPC141 γ2b gene does not lie in the 5′-flanking region of the germ-line Cγ2b gene, but is located in the 5′-flanking region of the germ-line Cμ gene. This recombination between V DNA and J DNA near the Cμ gene probably allowed the μ gene to be activated in the precursor B cells to MOPC141 myeloma. The second recombination occurs between a pair of sites, one located between the J DNA segment and the Cμ gene and the other in the 5′-flanking sequence of the Cγ2b gene, and replaces the Cμ-coding exons with the Cγ2b exons. This recombination is referred to as 'switch recombination', for it seems to be a key event preceding the heavy-chain switch. A similar gene structure was also observed for a complete α-chain gene[20] and for a γ1-chain gene[21].

We report here on the two types of recombination, studied by DNA sequencing. The sequences around the V–J joining and the switch recombination sites have different features, suggesting that at least two distinct enzyme systems are involved in the generation of the active immunoglobulin γ2b gene. In addition, sequence analysis of the V DNA of MOPC141 and its germ-line components, the embryonic V DNA and J DNA, revealed that the third hypervariable region is encoded in a separate DNA segment(s) in the germ-line genome.

Mouse DNA inserts in the DNA clones used in this study are listed in Fig. 1. The homology among these DNA fragments has been determined by heteroduplex analysis and restriction enzyme mapping. The coding regions were mapped by R-loop analysis using mRNAs from MOPC141 (γ2b) or MOPC104E (μ). Isolation and electron microscopic characterization of clones M141–P21 (complete γ2b), MEP203 (embryonic J plus Cμ) and MEP3 (embryonic Cγ2b) have been described elsewhere[19]. The isolation of the embryonic V-gene clone, PJ14, is described below.

Identification of J DNA for MOPC141

We have previously identified the putative V–J joining site on the complete γ2b gene clone (M141–P21) and on an embryonic Cμ gene clone (MEP203) by heteroduplex analysis and R-loop mapping[19]. The electron microscopic studies indicated that the putative J DNA for the MOPC141 heavy-chain gene lies 3.6 kilobases 5′ to the Cγ2b gene on M141–P21 and 5 kilobases 5′ to the Cμ gene on MEP203. To analyse the structural features of heavy-chain V–J joining sites, we determined the nucleotide sequences of the appropriate regions of M141–P21 and MEP203. Although the amino acid sequence of the MOPC141 heavy chain is unknown, identification of the J DNA segment coding for this chain was possible because J-peptide sequences are highly conserved. As shown in Fig. 2, M141–P21 carries a DNA sequence in the vicinity of the putative V–J joining sites which can encode a peptide that is identical to the J peptides of

0028–0836/80/330676—08$01.00

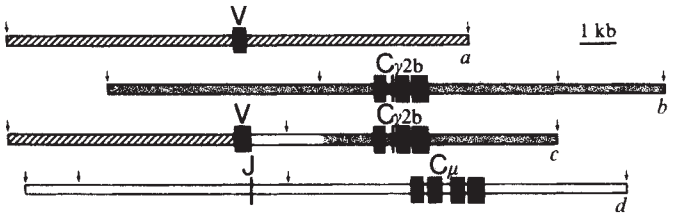

Fig. 1 Mouse DNA inserts in the clones used in the study. Embryonic V_{M141} gene clone PJ14 (*a*), embryonic Cγ2b gene clone MEP3 (*b*), complete γ2b gene clone M141–P21 (abbreviated to M141) (*c*) and embryonic Cμ gene clone MEP203 (*d*) are shown. Arrows indicate *Eco*RI cleavage sites. Filled boxes are exons identified by R-loop mapping. The regions of homology among the four DNA clones as determined by heteroduplex analysis are indicated by bars of different types of shading.

MOPC21 and MOPC173 heavy chains (see Table 1). We also analysed the MEP203 insert to determine whether the same J DNA segment is present in the expected position. Our previous study on clones MEP203 and M141–P21 suggested that the J_{M141} DNA segment and the sequence that follows this segment are derived from a region 5′ to the Cμ gene. In fact, the sequence of MEP203 showed that the J DNA segment and its 3′ non-coding sequence found on M141–P21 are also present in the region about 5 kilobases 5′ to the Cμ gene on MEP203 (Fig. 2). Comparison of these two J_{M141} DNA sequences, one on M141–P21 and the other on MEP203, indicates that the germ-line J_{M141} DNA can encode the M141 J peptide starting with the second nucleotide of the Leu codon at position 112.

In γ2b heavy-chain peptides, the boundary of V and C regions has been thought to be around Ala 126 (ref. 22). The J_{M141} DNA segment can fully encode Ser 125, but only partially Ala 126. We found a triplet, GGT, instead of the Ala codon GCN. Because, in almost all cases studied thus far, an intron starts with the doublet GT (ref. 23), we tentatively conclude that the coding of the J_{M141} DNA ends with the first letter G in the triplet GGT.

Four heavy-chain J DNA segments around 8 kilobases 5′ to the germ-line Cμ gene

Previous studies showed that most, if not all, κ-light-chain J DNAs are tightly linked[7,13]. To examine whether the same applies to the heavy-chain J DNA segments, we isolated DNA clones containing the rearranged V gene from myelomas MOPC603, MOPC315, HOPC8 and MOPC173 and analysed the heteroduplexes formed between each of these DNA clones and the embryonic μ DNA clone MEP203. It is expected that the divergency point observed on a Y-shaped heteroduplex molecule corresponds to the J DNA segment used for expression in the respective myeloma cells. In this way, we mapped four different J DNA segments in the region 5′ to the Cμ gene, at 1.1 (MOPC173-II J), 1.6 (HOPC8-II J), 1.9 (MOPC315 J) and 2.2 (MOPC603 J) kilobases, 5′ to the *Eco*RI site located 3.5 kilobases from Cμ DNA on MEP203. As MEP203 carries a 3-kilobase deletion at 1.5 kilobases 5′ to the Cμ gene (see below), the *Eco*RI site is 6.5 kilobases from the Cμ gene on the mouse chromosome. Using μ-chain mRNA from MOPC104E in R-loop analysis, we also identified the J_{M104E} DNA segment at 2.3 kilobases 5′ to the *Eco*RI site (data not shown). As the MOPC104E μ chain and the MOPC603 α chain seem to have the same J peptide (Table 1), this J DNA segment is probably the same as that detected by heteroduplex analysis using the MOPC603-derived V-gene clone.

Figures 3*Aa* and 4 show the restriction map and the DNA sequence of the 1.5-kilobase J-rich region, respectively. So far, 22 heavy chains have been sequenced in the J-peptide regions. They can be classified into four different groups according to the amino acid sequences of the J peptides (Table 1). Comparison of these J-peptide sequences with the DNA sequence allowed identification of the corresponding four J DNA segments. These four segments can account for all heavy-chain J peptides sequenced so far (see Table 1) and, therefore, seem to encode the J peptides of all different classes or subclasses. We designate these J DNAs as J_{H1}, J_{H2}, J_{H3} and J_{H4}, from 5′ towards 3′. No additional J-like DNA sequence has been observed in the 1.5-kilobase region sequenced thus far. There is one unexpressed J DNA, Jκ3 (ref. 7), in a κ-light-chain gene, and there may also be some in the heavy-chain gene.

Isolation of the germ-line V-gene clone for the MOPC141 heavy chain

To determine the recombination site on the germ-line V gene, we attempted to identify and clone the *Eco*RI DNA fragment carrying the germ-line V gene for the MOPC141 heavy chain. First, we analysed embryonic DNA by Southern hybridization using the V DNA probe purified from M141–P21. This probe is a 257-base pair *Ava*II–*Hga*I fragment encoding residues 9–94 of the M141 heavy chain (Fig. 3*Ab*). In the *Eco*RI digest of mouse embryo DNA, about 10 hybridizable bands were detected (data not shown). Detection of 5–10 bands by a single V-gene probe has previously been reported for both κ light chains[5,9] and heavy chains[24–26]. This is due to cross-hybridization among a subset of germ-line V genes that have presumably diverged more recently. As the length of the 5′-noncoding region of the V_H gene on M141–P21 is 6.5 kilobases, the size of the *Eco*RI fragment carrying the germ-line V_H DNA for the M141 heavy chain should be larger than 6.5 kilobases. We separated the *Eco*RI digest of embryonic DNA on a preparative agarose gel, eluted the DNA from the 6.5–23-kilobase region and used it for DNA cloning. Using the V-region probe, we isolated some 30 clones from 2 separate experiments and classified them into 5 different types according to the size of the *Eco*RI insert. Heteroduplex molecules formed between the clones of each type and M141–P21, a complete γ2b gene clone, were examined by electron microscopy.

Although the *Eco*RI inserts of these clones were all hybridizable with the V-gene probe, only the clones containing a 14-kilobase insert formed Y-shaped heteroduplex molecules with M141–P21, in which the homology extended from the 5′ *Eco*RI end to a site 6.9 kilobases away from it. R-loop analysis using the MOPC141 γ2b mRNA indicated that the V-coding DNA segments map at or around the heteroduplex fork point. We further analysed one of these 14-kilobase clones, PJ14, by restriction mapping and DNA sequencing. Evidence that the 14-kilobase clone contains the germ-line V gene for the MOPC141 γ2b chain is obtained from restriction enzyme maps. The cleavage sites for *Bam*HI, *Kpn*I, *Xba*I, *Sac*I and *Hin*dIII in the 5′-flanking region of the V gene on M141–P21 were all found at the corresponding positions on the embryonic V-gene clone, PJ14 (data not shown).

The third hypervariable region is encoded separately

The DNA sequences around the V-coding regions of PJ14 and M141–P21 are shown in Fig. 2. The sequences include a 46-base pair exon presumably coding for the hydrophobic signal peptide, an 81-base pair intron, the entire V-gene exon and the 3′-flanking regions. As in all κ- and λ-chain genes studied so far (refs 4, 6, 8 and G. Heinrich, unpublished results), the V coding begins with the residue at position −4 within the signal peptide. Also, as in the light-chain genes, the coding by the embryonic clone PJ14 ends prematurely with Ser 97, whereas that of the myeloma clone, as mentioned above, continues up to Ser 125. In the 0.6-kilobase stretch of DNA starting with the *Hin*dIII site in the 5′-noncoding region and ending with the Ser 97 codon, the

Table 1 Amino acid sequences of mouse immunoglobulin heavy chains around the carboxyl ends of the V regions

Myeloma	End of V	HV3	J	J DNA
M104E	··· YYCAR	DY	d W Y F D V W G A G T T V T V S S	
J558	··· YYCAR	D	r Y – – – – – – – – – – – – – – –	
T15	··· YYCAR	DYYGS	s Y – – – – – – – – – – – – – – –	
M603	··· YYCAR	NYYGST	– – – – – – – – – – – – – – –	
M167	··· YYCTR	DADYGDSYF	g – – – – – – – – – – – – – – –	J_{H1}
T601	··· YYCAR	LGYY	g – – – – – – – – – – – – – – –	
S107	··· YYCAR	DYYGS	s Y – – – – – – – – – – – – – – –	
H8	··· YYCAR	DYYGN	s Y – – – – – – – – – – – – – – –	
M511	··· YYCAR	DGDYGS	s Y – – – – – – – – – – – – – – –	
M315	··· YYCAG	DNDHL	Y F D Y W G Q G T T L T V S S	
X24	··· YYCAR	LGYYG	– – – – – – – – – – – – – – –	J_{H2}
Hdex4	··· YYCAR	DK	n – – – – – – – – – – – – – – –	
Hdex5	··· YYCAR	DS	n – – – – – – – – – – – – – – –	
A4	··· YYCTT		s W F A Y W G W G T L V T V S A	
E109	··· HYCTT		g – – – – – – – – – – – – – – –	
U61	··· YYCTT		g – – – – – – – – – – – – – – –	J_{H3}
A47N	··· YYCST		g – – – – – – – – – P – – –	
X44	··· YYCAR	LHYYGYA	– – – – – – – – – – – – – – –	
J539	··· YYCAR	LHYYGYN	– – – – – – – – – – – – – – –	
M141	··· YYCAS	VSIYYYGRSDKYFT	1 D Y W G W G T S V T V S S	
M21	··· YYCAR	HGNYPW	Y A M – – – – – – – – – – – – –	J_{H4}
M173	··· YYCAR	SP	Y Y A M – – – – – – – – – – – – –	

Amino acid sequences of 22 mouse heavy chains[27,32,34–40] for residues 90–113 (numbering is after Kabat *et al.*[22]) are shown by one-letter codes[41]. According to the J-region peptides, the chains are classified into four groups. The amino acid sequence of the MOPC141 γ2b chain was deduced from the nucleotide sequences determined in this study. 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.

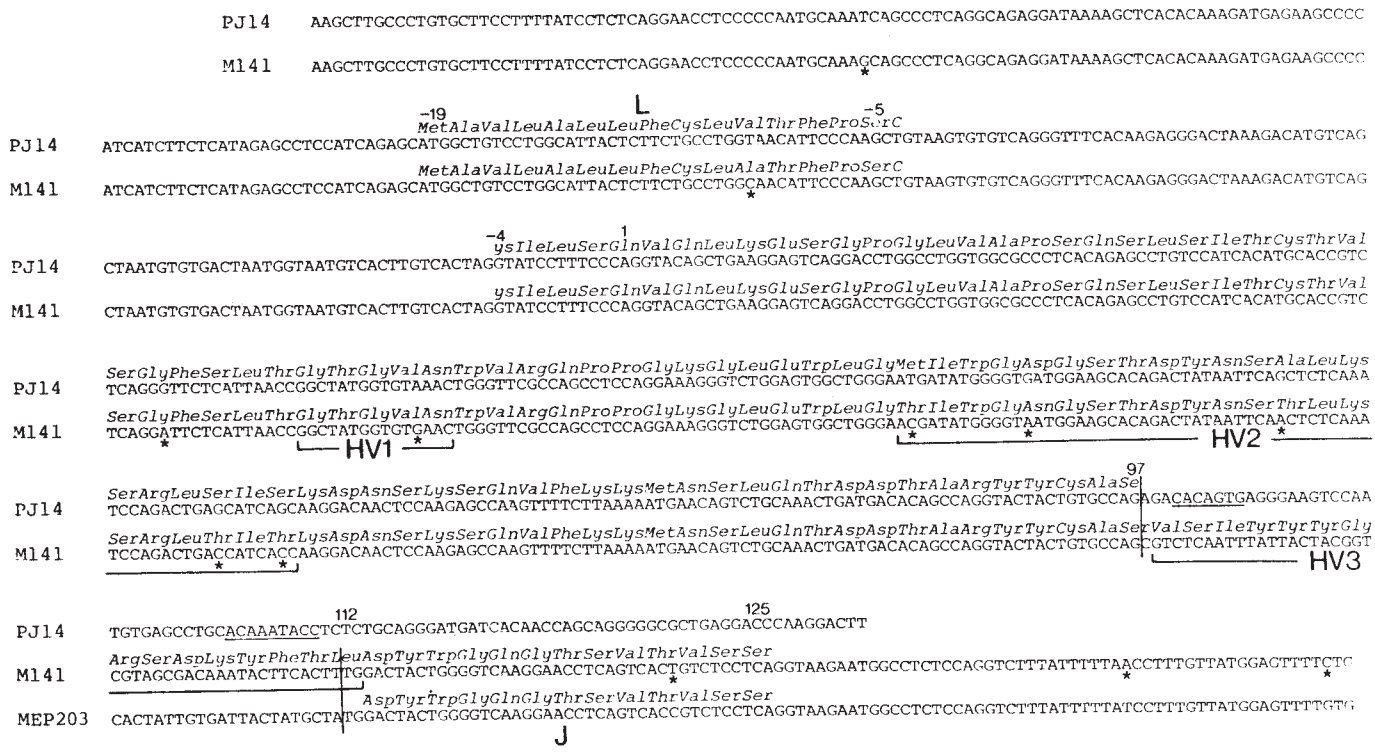

Fig. 2 Nucleotide sequences of MOPC141 V_H DNA (M141), its embryonic V_H DNA (PJ14) and the embryonic J DNA (J_4) coding for the J peptide of the MOPC141 γ2b chain (MEP203). Sequencing strategy is shown in Fig. 3*A*. Sequence determination was carried out according to the method of Maxam and Gilbert[43]. The amino acid sequence of the complete MOPC141 V region predicted by the nucleotide sequence of clone M141 is shown in italics. The amino acid sequences predicted by the germ-line V DNA segment in clone PJ14 and the J_4 DNA segment in clone MEP203 are also shown. The nucleotide differences observed between the two clones, M141 and PJ14, are shown by asterisks. Vertical lines indicate the possible recombination sites. The two conserved sequences in the 3′-flanking region of the embryonic V DNA are underlined. HV1, HV2 and HV3 designate the three hypervariable regions.

sequences of the two DNA clones differ in 9 base pairs. Of these differences, 6 are within the first and the second hypervariable (HV) regions. Similar, limited, single-base differences have previously been observed in the hypervariable regions of a germ-line V gene and its somatic variants for the light chains of both λ and κ types and were taken as direct evidence for somatic diversification of antibody genes[6]. Three other differences may be a consequence of the similar somatic mechanism or may have occurred during the propagation of the myeloma cells.

Figure 2 also shows the DNA sequence around the J_{M141} DNA segment of the embryonic clone MEP203. As described before, coding begins within the Leu 112 triplet and ends with the Ser 125 triplet. It thus seems that the 14-residue peptide comprising the third hypervariable region (HV3) beginning with Val 98 and ending with Thr 111 is encoded in neither the germ-line V gene nor the J DNA segment. This situation is in contrast to that of light-chain genes previously studied[6–8], in which no such coding gap was observed. We suggest that the third hypervariable region of this heavy chain is encoded in one or more discrete DNA segments (D DNA segments for 'diversity'[27]) that lie elsewhere in the germ-line genome. The V, D and J DNA segments must be assembled by recombination to generate the complete V gene present in the myeloma cells.

The exact joining ends of V and J DNA segments are unfixed

Close examination of the amino acid sequences around the third hypervariable region and the nucleotide sequences around the 3′ end of the germ-line V_H gene and around the 5′ end of the J_H DNA segments suggest several interesting features of the DNA joining events. As previously suggested for the light-chain J DNA segments[7,13–15], the exact 5′ boundary of J_H DNA segments seems to be unfixed (Table 1). For instance, in the case of J_{H4}, the coding of the MOPC141 γ2b chain begins with the second base of the Leu 112 triplet UUG as indicated in Fig. 2. In contrast, coding of the MOPC21 γ1 chain and the MOPC173 γ2a chain by the same J_{H4} DNA segment seems to start with a Tyr triplet UAU and another Tyr triplet UAC, respectively. The first letters of the two Tyr triplets are 7 and 10 bases ahead (5′ side) of the first base used for the coding of the MOPC141 J region. Similar examples for J_{H1}, J_{H2} and J_{H3} are indicated in Fig. 4 and Table 1.

The DNA sequence of the clone PJ14 demonstrated that coding by the germ-line V_{M141} gene ends with the second letter of the Ser 97 codon AGC (Fig. 2). As shown in Table 1, the amino acid sequences immediately preceding Ser 97 are highly conserved among various V regions. Particularly noteworthy is

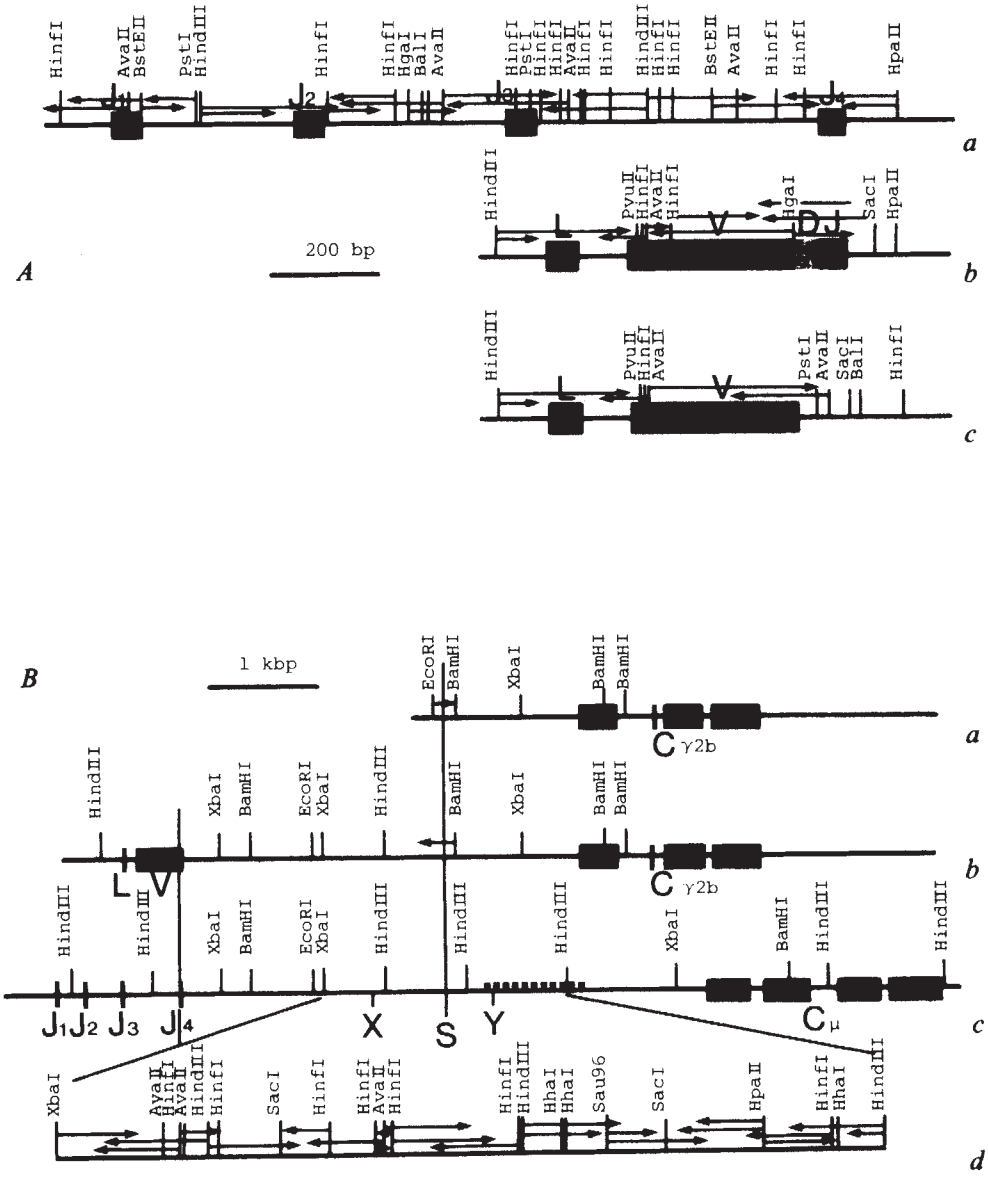

Fig. 3 Restriction enzyme cleavage maps of part of the mouse DNA inserts in clones MEP 203 (*Aa*, *Bc*, *Bd*), MEP3 (*Ba*), M141–P21 (*Ab*, *Bb*) and PJ14 (*Ac*). Cleavage sites were determined by the end-labelling method of Smith and Birnstiel[42], by the separation of single and double digests of DNA fragments by various restriction enzymes, and/or by DNA sequence determination. Horizontal arrows indicate the extent and direction of sequence determination. The filled boxes are exons identified by R-loop mapping and/or DNA sequencing. Vertical line S in *B* indicates the switch recombination sites for MOPC141 heavy-chain genes. X and Y in *Bc* are sequences similar to the Sμ(γ2b) sequence. Four J DNA segments, J_1, J_2, J_3 and J_4, were identified by heteroduplex analysis of rearranged V-gene clones isolated from MOPC603, MOPC315, HOPC8 and MOPC173, respectively. The J DNA segment for MOPC104E was identified as J_1 by R-loop mapping with μmRNA of MOPC104E. The region of clone MEP203, composed of repetitions of CTGAG or sequences closely related to it, is indicated by a broken line in *Bc*. bp, base pairs; kbp, kilobase pairs.

Fig. 4 Nucleotide sequence of the 1.4-kilobase region of the embryonic clone MEP203 containing four J_H DNA segments. The sequencing strategy is shown in Fig. 3*Aa*. The J_H DNA segments are referred to as J_1, J_2, J_3 and J_4 from 5′ to 3′. The amino acid sequences encoded by these J DNA segments are shown in italics. The sequences closely related to the palindromic heptamer, CACTGTG, and present in front of every J DNA segment, are underlined. The vertical lines indicate determined (solid lines) or predicted (broken lines) sites where coding by the J DNA segments begin 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. (See Fig. 2 for the M141 γ2b chain gene. The others are based on unpublished sequence data.) We have identified, isolated and sequenced two rearranged V genes, one from HOPC8 and the other from MOPC173. The V regions encoded by these rearranged V genes are different from the published ones, and are therefore referred to as H8-II and M173-II. It is not clear whether these V genes are part of the active heavy-chain genes of the respective myelomas. 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 1).

the universal 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 the 97th triplet. If so, coding gaps of various lengths exist in the majority of the cases listed in Table 1. 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 Gly 98 is unaccounted for. Similarly, only several more bases are necessary to code for the HV3 of the J558 α chain. Previous studies on κ-chain genes indicated that the 3′ end of a given germ-line V gene may not be precisely fixed and could shift within a range of several nucleotides[7]. 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 seems 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 always occurs in two steps, but that in some cases the D segments are very short or do not even contribute to protein encoding at all.

In either case, by modulation of the joining sites and by various combinations of these DNA segments, diversity could be generated in the third hypervariable region of heavy-chain molecules. However, the location and 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 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.

Sequence characteristics around the V and J joining sites

We have previously reported that the five Jκ DNA segments and the one Jλ DNA segment are preceded by two blocks of short, conserved sequences: a palindromic hexamer interrupted by an AT base pair at the centre of symmetry, CACTGTG, and a nanomer, GGTTTTTGT (ref. 7). Similar sequences are also present 5′ to all four J_H DNA segments, although deviation from the basic sequences seems to be slightly higher for the J_H segments than for the J_L segments (Table 2). We also reported previously that germ-line Vκ and Vλ DNA clones contain inverted complements of the nanomer and heptamer in the 3′-noncoding regions. As shown in Fig. 2 and summarized in Table 2, the same or similar sequences are also present in the equivalent positions of the germ-line V_{H141} gene clone. Based on the fact that these sequences are highly conserved, we propose that they constitute recognition signals for the putative recombinase. The finding of similar sequences near the heavy-chain gene DNA segments strengthens the validity of this hypothesis.

In the case of J_{H3}, two sequences, CACAATG and CAATGTG, are closely related to the basic heptamer sequence, CACTGTG. Similarly, two sequences, TATTGTG and TACTATG, are present near the 5′ end of J_{H4} (Fig. 4). In these cases it is not clear which of the two sequences constitutes part of the proposed recognition signal.

The presence of the closely related conserved sequences in the corresponding positions of both light- and heavy-chain germ-line V genes and their J DNA segments suggests that all recombination events necessary for the generation of complete, somatic V genes are carried out by the same or a similar mechanism. We predict that the similar recognition sequences are also present on both sides of the D DNA segments.

An additional feature of the conserved signal sequences is the striking regularity in the length of the spacer between the heptamer and the nanomer. As summarized in Table 2, in almost all cases the spacers are 12 or 23 base pairs long. One exception is one of the two alternative cases of J_{H4}, where the spacer is 31 base pairs long. As a DNA double helix completes a turn every 10.4 base pairs[28], the above indicates that the internal boundaries of the heptamer and the nanomer of a given DNA segment are orientated in nearly the same direction relative to the axis of the helix, regardless of the length of the spacer. The spacers of all Vκ segments correspond to one turn, and those of $V\lambda_I$ and $V\lambda_{II}$ to two turns. The spacer of V_{M141} is two turns long. All Jκ spacers are two turns whereas the $J\lambda_I$ spacer is one turn. All spacers of the four J_H segments are two turns. The 31-base pair spacer, one of the two alternative spacers of J_{H4}, could be considered as an exceptional three turns. It should be emphasized that although the lengths of the spacers are conserved in a given type of gene segment, their sequences have diverged widely.

A model of the joining enzymes

The above mentioned features around the joining sites suggest some predictions about the nature of the recombinases. In both λ and κ light-chain V–J joinings, one partner in the recombination displays the heptamer and nanomer one turn of the helix apart from the recombinase, and the other displays these sequences two turns apart. We predict that the recombinase contains two DNA-binding proteins: one recognizing the heptamer and nanomer one turn apart, and the other recognizing them two turns apart. As the two partners carry essentially identical recognition sequences, the two DNA-binding proteins would be structurally closely related.

The germ-line V_{M141} gene and all J_H segments show the recognition sequences separated by two-turn spacers. If the D segment contains recognition sequences with one-turn 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 the assembly of immunoglobulin V genes would follow a one-turn/two-turn spacer rule.

The two recombinase components hold the two recombining partners together and cut and rejoin the strands in the vicinity of the heptamers. As the recognition sequences of one partner are complementary to those of the other, it is possible that intra-strand base pairing occurs, however transiently, in the enzyme–DNA complex and facilitates the ligation reaction.

This rule prohibits some unwanted recombinations, such as those between Vλ and Jκ, and Vκ and Jλ. However, some other unwanted recombinations, such as those between V_H and Jλ, D and Jκ, and Vλ and Vκ, could occur. This may be the reason that the three gene families are distributed in three different chromosomes.

The switch recombination sites for the complete γ2b gene of MOPC141

Our previous electron microscopic studies[19] demonstrated that the complete γ2b gene clone M141–P21 shows some homology to the germ-line Cμ gene clone MEP203 and some to the germ-line Cγ2b gene clone MEP3. As shown in Fig. 1, the homology between clone M141–P21 and MEP203 is 2.4 kilobases long and extends, on M141–P21, from the 3′ end of the V DNA segment to a site 1.2 kilobases 5′ to the Cγ2b gene, and, on MEP203, from the J_4 DNA segment to a site 2.5 kilobases 5′ to the Cμ gene. The remaining part of the 3.6-kilobase V–Cγ2b intron of the complete γ2b gene and the sequence that follows this part are entirely homologous to the germ-line Cγ2b gene clone. The restriction enzyme maps of the three DNA clones shown in Fig. 3*B* confirm the results of the electron microscopic studies. Thus, the cleavage sites present in the first 2.2-kilobase portion of the V–Cγ2b intron are also present at the corresponding positions of the germ-line Cμ gene clone, whereas the cleavage sites located in the rest of the intron and the DNA stretch that follows are also found in the corresponding regions of the germ-line Cγ2b gene clone. The maps show that switch recombination occurs between a pair of sites, one located within the 750-base pair *Hin*dIII–*Hin*dIII segment of the germ-line Cμ gene clone and the other within the 200-base pair *Eco*RI–*Bam*HI segment of the germ-line $C\gamma_{2b}$ gene clone, and generates the 650-base pair *Hin*dIII–*Bam*HI segment present in the complete γ2b gene clone (see Fig. 3*B*).

To understand this novel somatic event better, we determined the nucleotide sequences around the switch recombination sites (Fig. 5*a*). As expected, the 5′ portion of the M141–P21 sequence is the same as the sequence of the germ-line Cμ gene clone, MEP203, whereas the 3′ portion is identical to the sequence of the germ-line Cγ2b gene clone, MEP3. The exact μ–γ2b switch recombination sites are indicated by a vertical line in Fig. 5*a*. We propose to refer to the μ–γ2b switch sites on the germ-line Cμ gene and the germ-line Cγ2b gene as Sμ(γ2b) and Sγ2b(μ), respectively.

Table 2 The two conserved blocks of sequences near the V–J or V–D–J joining sites

J DNA segments	Nanomers		Heptamers	Ref(s)
Jκ1	GGTTTTTGT	23	CACTGTG	7, 13
Jκ2	AGTTTTTGT	23	CAGTGTG	7, 13
Jκ3	GGGTTTTGT	21	CACTGTA	7, 13
Jκ4	GGTTTTTGT	24	CACTGTG	7, 13
Jκ5	GGTTTTTGT	23	CACTGTG	7, 13
JλI	GGTTTTTGC	12	CACAGTG	6
J_{H1}	AGTTTTAGT	22	GACTGTG	*
J_{H2}	GGTTTTTGT	23	TAGTGTG	*
J_{H3}	ATTTATTGT	21	CACAATG	*
		23	CAATGTG	*
J_{H4}	GGTTTTTGT	22	TATTGTG	*
		31	TACTATG	*
Basic sequence	GGTTTTTGT		CACTGTG	
V DNA segments	**Heptamers**		**Nanomers**	**Ref.**
Vκ21C	CACAGTG	11	ACAAAAACC	7
Vκ21B	CACAGTG	12	ACAAAAACC	†
VK41	CACAGTG	12	ACATAAACC	8
VK2	CACAGTG	12	ACATAAACC	5
$V\lambda_I$	CACAATG	22	TCAAGAACA	6
$V\lambda_{II}$	CACAATG	23	ACAAGAACA	4
V_{H141}	CACAGTG	23	ACAAATACC	*
Basic sequence	CACAGTG		ACAAAAACC	

Two blocks of conserved sequences found in the 5′-flanking region of J DNA segments and in the 3′-flanking region of embryonic V DNA segments are compared. The numbers between the two types of sequences indicate the distance between them in base pairs. The bases different from those of the basic sequences in the corresponding positions are underlined. The refs from which the sequences were taken are listed (* this paper; † G. Heinrich, personal communication).

Sequences around the μ–γ2b switch sites

Are there any characteristic sequences around the μ–γ2b switch sites that might be considered as part of the recognition signal for the recombinase? As shown in Fig. 5*a*, two short blocks of sequences, $\frac{\text{TCCTGG}}{\text{AGGACC}}$ and $\frac{\text{AGA}}{\text{TCT}}$, present in front of (towards the 5′ side of) Sμ(γ2b) are also present near the Sγ2b(μ) site in the equivalent positions. The two sets of sequences are in the same orientation relative to the direction of transcription, and thus, differ from those sequences conserved near the V–J joining sites not only in the sequence content but also in the relative

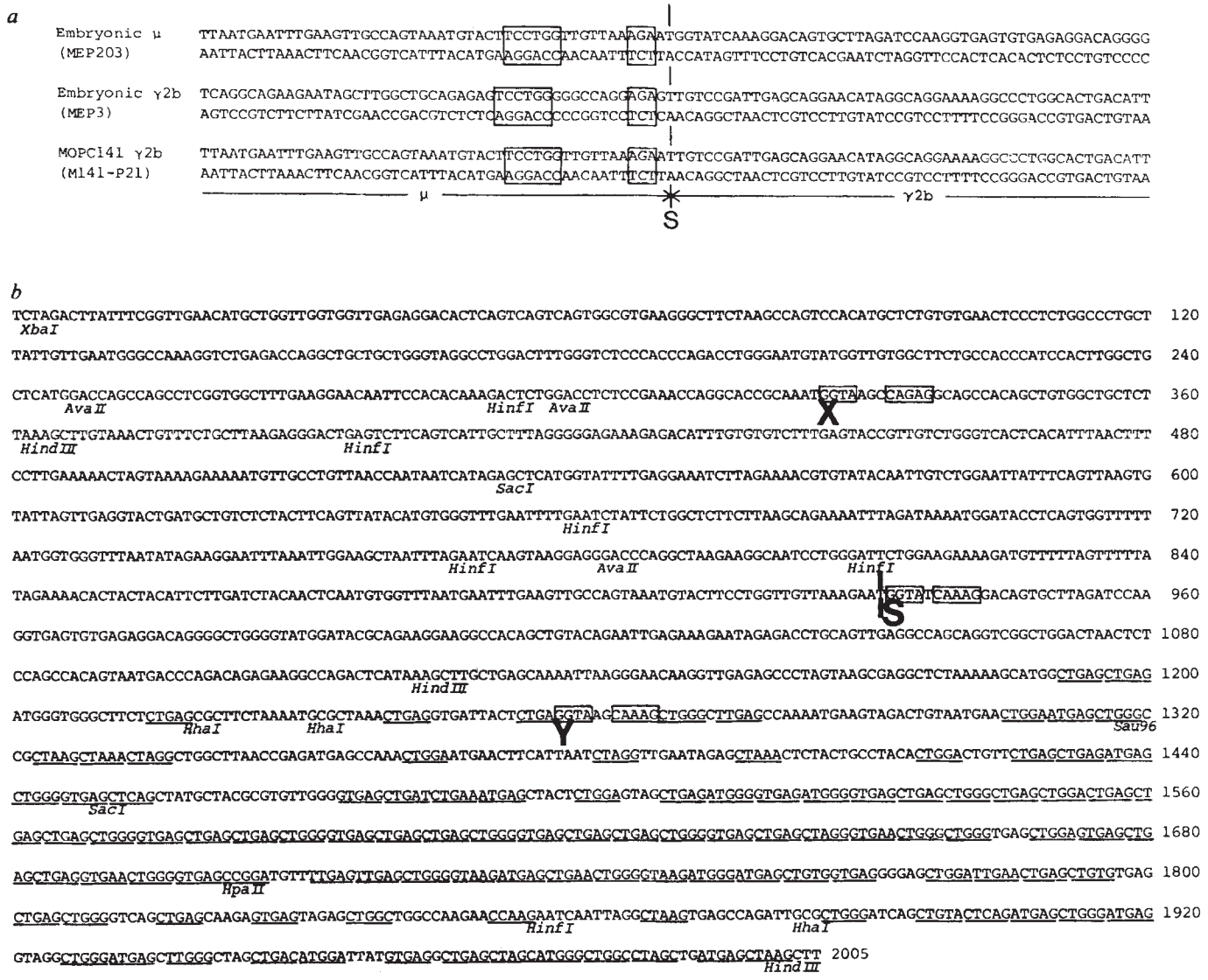

Fig. 5 *a*, Nucleotide sequences around the μ(γ2b) switch recombination sites in clones MEP203, MEP3 and M141–P21. Sequencing strategy is shown in Fig. 3*B*. The recombination sites are indicated by the vertical line designated by S. The two blocks of sequences commonly present in the 5′ side of the recombination sites of the embryonic μ and γ2b gene clones are in boxes. *b*, The nucleotide sequence of 5′-flanking region of Cμ DNA surrounding the switch recombination site for M141 γ2b gene. The 1.8-kilobase *Xba*I–*Hin*dIII segment (Fig. 3*Bd*) was sequenced according to the strategy shown in Fig. 3*Bd*. Two regions (X and Y) homologous to the M141 γ2b switch site (S) are indicated and common sequences GGTA and CAPuG are in boxes. Pentameric sequence CTGAG or its related sequences are underlined.

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–γ2b switch recombination is quite different from that for V–J joining.

Sequences of the 5′-flanking region of the germ-line Cμ gene

Recently, Davis *et al.*[20] isolated a complete α heavy-chain gene clone from MOPC603 that carries both V_H and Cα genes. Heteroduplex analysis of this clone against the germ-line Cα and Cμ 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 Cμ gene. The Sμ(α) site was mapped between the two *Hin*dIII sites at 2.2 and 1.3 kilobases, respectively, from the Cμ gene (see Fig. 3*Bc*). Thus, the μ–α and μ–γ2b switch recombinations seem to use different sites in the 5′-flanking region of the germ-line Cμ gene. This suggests that there may be several different switch sites in this region, possibly one at least for each of the six heavy-chain classes or subclasses, with the possible exception of the δ chain. (The behaviour of B cells suggests that μ–δ switches probably do not occur[29,30].)

We therefore extensively sequenced the 5′-noncoding region of the germ-line Cμ gene clone, MEP203. Fig. 3*Bd* shows the restriction map of the 2.2-kilobase region preceding the Cμ gene, and its complete nucleotide sequence is shown in Fig. 5*b*. We searched for sequences that had a resemblance to those found near the Sμ(γ2b) site and found two, which are referred to as X and Y and are located, respectively, around 0.6 kilobases upstream and 0.3 kilobases downstream of the Sμ(γ2b) site (Fig 5*b*). All three sequences share a tetranucleotide, GGTA, and a pentanucleotide, CAPuAG. However, the two sequences, TCCTGG and AGA, commonly present in the regions 5′ to the Sμ(γ2b) and Sγ2b(μ) sites, are absent in the vicinities of X and Y. It is premature to predict that X and Y are indicative of switch recombination sites. Cloning and sequence analysis of complete heavy-chain genes of various classes and subclasses will clarify

the matter. (Recently, we learned that the Y sequence is indeed used for the recombination of the γ1 gene sequenced by Kataoka *et al.*[21] (N. Takahashi and T. Honjo, personal communication).)

In the 0.8-kilobase region extending from the 5′ side of the Y sequence to the *Hin*dIII site, a pentameric sequence, CTGAG, or its close variants are repeated in tandem. As the Sμ(α) site reported by Davis *et al.*[20] falls within this region, involvement of the repeated sequence in the μ–α switch recombination can be invoked. Indeed, our recent DNA sequencing study on the germ-line Cα gene clone indicates that the same pentameric sequences are prominent in the region where the Cα(μ) site had been mapped by Davis *et al.* (H.S., unpublished observations).

Clone MEP203 carries a deletion of about 3 kilobases which seems to have occurred during the cloning or subsequent propagation of the clone in *Escherichia coli*. Restriction enzyme maps of several independent Cμ gene clones isolated by our laboratory and another laboratory[20] suggest that most clones carry deletions of various lengths that map between the two *Hin*dIII sites surrounding the Y sequence. As shown in Fig. 5*b*, this is the region where the pentameric sequences are repeated. Our recent DNA sequencing studies on two independent Cμ gene clones carrying different deletions confirmed that these deletions occur within the pentamer-rich regions.

The two types of somatic recombination

The present studies deal with two types of somatic recombination required for the generation of complete heavy-chain genes, namely V–D–J or V–J joining and switch recombination. The two types of recombination are clearly different in several aspects, although they both occur somatically during lymphocyte differentiation. Although V–D–J or V–J joining should precede the synthesis of the free μ chain detected in pre-B cells[31], switch recombination almost certainly occurs after these cells mature to small B cells bearing complete IgM molecules on the cell surface. V–D–J or V–J joining takes place at the margins of two protein-encoding DNA segments and generates complete V genes. In contrast, switch recombination occurs in the sequences located away from the protein-encoding DNA segments and in effect replaces the Cμ exons of the complete μ-chain gene with exons coding for the C regions of other heavy-chain classes or sub-classes, except, perhaps, for the δ class.

In V–D–J or V–J joining, the recombinant DNAs are to be translated in the recombined regions. This imposes certain restrictions on the precise cutting and joining sites because the encoded polypeptide chains would have to fold properly to be able to function as immunoglobulin subunits. Indeed, this type of recombination seems to occur within a short (several base pairs) region adjacent to the conserved palindromic heptamer present at the margin of every V and J DNA segment studied to date. However, the recombination site does not seem to be unique in a given gene segment and this flexibility, however limited, seems to have been exploited during evolution to increase the genetic capacity of the organism's antibody repertoire. Switch recombination can *a priori* occur anywhere within the 7.5-kilobase intron separating the J DNA segments and the Cμ exons, and anywhere within a several-kilobase region located 5′ to the respective C exon of other classes, as long as the transcription unit on the recombined DNA can encompass the V and C exons and the transcript can be processed properly to the mature mRNA. This suggests that switch recombination sites may be scattered widely in these regions. Indeed, the Sμ(γ2b) site for the MOPC141 γ2b chain and the Sμ(α) site for the MOPC603 α chain are at least 500 base pairs apart. Furthermore, our recent studies demonstrated that in the *Eco*RI digests of various myeloma DNA, the sizes of the DNA fragments detected by the 3-kilobase *Xba*I–*Xba*I fragment carrying the switch region (see Fig. 3*Bc*) are diverse, even among those myelomas secreting the heavy chains of the same class or subclass (R.M., unpublished observations). This suggests that one or both switch recombination sites are not class or subclass specific.

We previously proposed that the V–J joining can be considered as a reversal of an ancient, accidental insertion of an IS-like DNA element carrying invertedly repeated sequences at the margins[7]. In contrast, the evolution of immunoglobulin chains suggests that the C exons of various heavy-chain classes or subclasses arose by duplication of ancient Cμ exons and their flanking sequences[44]. The switch recombination sites have probably arisen from these duplicated flanking sequences by making use of the sequence homology among them. Whether the switch to various classes or subclasses is actively controlled at the level of the DNA recombinations remains to be determined.

Early *et al.*[33] have independently reported the analysis of M603 α heavy-chain V DNA and its germ-line V DNA and J DNA. They found that the 18-base pair D DNA segment in M603 V DNA is not accounted for by either germ-line V DNA or J DNA. They also identified two heavy-chain J DNAs by nucleotide sequencing that correspond to our J_{H1} and J_{H2}. Our determined or predicted joining sites on J_{H1} DNA for M603, M167 and S107 heavy-chain genes are in agreement with their direct sequencing of cDNAs for heavy-chain mRNAs prepared from respective myelomas.

We thank Dr Michael Potter for myelomas and Ms L. Sultzman, Mr G. Dastoornikoo, Ms L. Gibson, Mrs P. Riegert and Mr A. Traunecker for technical assistance.

Received 5 May; accepted 23 June 1980.

1. Hozumi, N. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **73,** 3628–3632 (1976).
2. Tonegawa, S., Hozumi, N., Matthyssens, G. & Schuller, R. *Cold Spring. Harb. Symp. quant. Biol.* **41,** 877–889 (1976).
3. Brack, C., Hirama, M., Schuller, R. & Tonegawa, S. *Cell* **15,** 1–14 (1978).
4. Tonegawa, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **74,** 3171–3175 (1978).
5. Seidman, J. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **75,** 3881–3885 (1978).
6. Bernard, O., Hozumi, N. & Tonegawa, S. *Cell* **15,** 1133–1144 (1978).
7. Sakano, H., Hüppi, K., Heinrich, G. & Tonegawa, S. *Nature* **280,** 288–294 (1979).
8. Seidman, J. G., Max, E. E. & Leder, P. *Nature* **280,** 370–375 (1979).
9. Lenhard-Schuller, R., Hohn, B., Brack, C., Hirama, M. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **74,** 4709–4713 (1978).
10. Seidman, J. G. & Leder, P. *Nature* **276,** 790–795 (1978).
11. Gilmore-Herbert, M. & Wall, R. *Proc. natn. Acad. Sci. U.S.A.* **75,** 342–345 (1978).
12. Schibler, U., Marcu, K. B. & Perry, R. P. *Cell* **15,** 1495–1509 (1978).
13. Max, E. E., Seidman, J. G. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **76,** 3450–3454 (1979).
14. Weigert, M., Gatmaitan, L., Loh, E., Schilling, J. & Hood, L. *Nature* **276,** 785–790 (1978).
15. Weigert, M. *et al. Nature* **283,** 497–499 (1980).
16. Pernis, B., Forni, L. & Amante, L. *Ann. N.Y. Acad. Sci.* **190,** 420–431 (1971).
17. Pernis, B., Forni, L. & Luzzati, A. L. *Cold Spring Harb. Symp. quant. Biol.* **41,** 175–183 (1976).
18. Cooper, M. D., Kearney, J. F., Lydyard, P. M., Grossi, C. E. & Lawton, A. R. *Cold Spring Harb. Symp. quant. Biol.* **41,** 139–145 (1976).
19. Maki, R., Traunecker, A., Sakano, H., Roeder, W. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **77** (in the press).
20. Davis, M. M. *et al. Nature* **283,** 733–739 (1980).
21. Kataoka, T., Kawakami, T., Takahashi, N. & Honjo, T. *Proc. natn. Acad. Sci. U.S.A.* **77,** 919–923 (1980).
22. Kabat, E. A., Wu, T. T. & Bilofsky, H. in *Sequences of Immunoglobulin Chains* (NIH, Bethesda, 1979).
23. Breathnack, R., Benoist, C., O'Hare, K., Cannon, F. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **75,** 4853–4857 (1978).
24. Rabbitts, T. H., Matthyssens, G. & Hamlyn, P. H. *Nature* **284,** 238–243 (1980).
25. Davis, M., Early, P., Calame, K., Livant, D. & Hood, L. in *Eucaryotic Gene Regulation* (eds Axel, R., Maniatis, T. & Fox, C. F.) 393–406 (ICN–UCLA Symp., Academic, New York, 1979).
26. Cory, S. & Adams, J. M. *Cell,* **19,** 37–51 (1980).
27. Schilling, J., Clevinger, B., Davie, J. M. & Hood, L. *Nature* **283,** 35–40 (1980).
28. Wang, J. C. *Proc. natn. Acad. Sci. U.S.A.* **76,** 200–203 (1979).
29. Salsano, P., Froland, S. S., Natvig, J. B. & Michaelsen, T. E. *Scand. J. Immun.* **3,** 841–846 (1974).
30. Fu, S. M., Winchester, R. J. & Kunkel, H. G. *J. Immun.* **114,** 250–261 (1975).
31. Burrows, P., LeJeune, M. & Kearney, J. F. *Nature* **280,** 838–840 (1974).
32. Rao, D. N., Rudikoff, S., Krutzsch, H. & Potter, M. *Proc. natn. Acad. Sci. U.S.A.* **76,** 2890–2894 (1979).
33. Early, P. *et al. Cell* **19,** 981–992 (1980).
34. Kehry, M., Sibley, C., Fuhrman, J., Schilling, J. & Hood, L. *Proc. natn. Acad. Sci. U.S.A.* **76,** 2932–2936 (1979).
35. Vrana, M., Rukikoff, S. & Potter, M. *Proc. natn. Acad. Sci. U.S.A.* **75,** 1957–1961 (1978).
36. Francis, S. H., Leslie, R. G. Q., Hood, L. & Eisen, H. N. *Proc. natn. Acad. Sci. U.S.A.* **71,** 1123–1127 (1974).
37. Milstein, C., Adetugbo, K., Cowan, N. J. & Secker, D. S. *Prog. Immun.* **2,** 157–168 (1974).
38. Bourgeois, A., Fougereau, M. & Depreval, C. *Eur. J. Biochem.* **24,** 446–455 (1972).
39. Rudikoff, S. & Potter, M. *Proc. natn. Acad. Sci. U.S.A.* **73,** 2109–2112 (1976).
40. Rudikoff, S. & Potter, M. *Biochemistry* **13,** 4033–4038 (1974).
41. IUPAC–IUB Commission *J. Biochem.* **113,** 1 (1969).
42. Smith, H. O. & Birnstiel, M. L. *Nucleic Acids Res.* **3,** 2387–2398 (1976).
43. Maxam, A. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74,** 560–564 (1978).
44. Hill, R. L., Delaney, R., Fellows, R. E. Jr & Lebovitz, H. E. *Proc. natn. Acad. Sci. U.S.A.* **56,** 1762–1769 (1966).