Proc. Natl. Acad. Sci. USA
Vol. 76, No. 7, pp. 3450–3454, July 1979
Genetics

Sequences of five potential recombination sites encoded close to an immunoglobulin κ constant region gene

(J segment/intervening sequence/palindromes/antibody diversity)

EDWARD E. MAX, J. G. SEIDMAN, AND PHILIP LEDER

Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20205

Contributed by Philip Leder, May 7, 1979

ABSTRACT **Immunoglobulin κ chain gene formation involves site-specific somatic recombination between one of several hundred germ-line variable region genes and a joining site (or "J segment") encoded close to the constant region gene. We have cloned and determined the nucleotide sequence of major portions of the recombination region of the mouse κ gene and discovered a series of five such J segments spread out along a segment of DNA 2.4 kilobases from the κ constant region gene. These J segments encode the 13 COOH-terminal amino acids of the variable region, probably including amino acids involved in the antigen combining site and in heavy/light chain contacts. The J segments also display striking sequence homology to one another in both their coding and immediately flanking sequences. Major elements of a short palindrome—CAC($\frac{T}{A}$)GTG—are preserved adjacent to the recombination sites of both variable and J region genes and constitute inverted repeats at both ends of the sequences to be joined. These palindromes can be written as a hypothetical stem structure that draws variable and J regions together, providing a possible molecular basis for the DNA joining event. Four of the J segments that we have discovered encode amino acid sequences already found in myeloma proteins. By altering the frame of recombination, we can account for additional light chain amino acid sequences, suggesting that the V/J joining event might generate antibody diversity somatically both by using different combinations of variable and J region genes and by using alternative joining frames.**

Formation of a complete immunoglobulin gene involves a recombination event that seems to be critical to the development of immunoglobulin diversity as well as to the activation of immunoglobulin genes (1–4). Such recombination takes place between a germ-line variable (V) region gene and a distant segment of DNA, the J region, that encodes the 13 amino acids conventionally associated with the carboxy end of the V region. Evidently, the J segment contains two important signals: one for DNA recombination and one for RNA splicing. This has been demonstrated directly for mouse λ (5–7) and κ (8–10) light chain genes by using cloned segments of DNA derived from embryonic and antibody-producing myeloma cells. Even after recombination, however, light chain genes remain in three discrete coding segments separated from one another by two intervening sequences of DNA. Early studies of heavy chain genes suggest that this organization is a general pattern related to immunoglobulin domains (11, 12).

We have found that a J-region gene located 3.7 kilobases (kb) from the κ constant (C) region gene is joined in a myeloma cell to the germ-line gene corresponding to the V region of the MOPC-41 light chain (10). Except for the recombination event, the germ-line and expressed sequences are identical, suggesting that no further somatic alteration was involved in creating this gene (10). Moreover, we noted that the germ-line κ J-region had several interesting coding and flanking sequence homologies to a λ J-region segment described by Bernard *et al.* (5). Here we extend our sequence information in the region surrounding this germ-line κ J segment. We have found five potential J-region genes, regularly spaced at intervals of 309–354 base pairs, on the 5′ side of the C region gene. Four of these correspond to amino acid sequences found in myeloma light chains. It is clear that, by altering the frame of recombination between germ-line V and J-region genes, other light chain sequences could be accounted for. Moreover, the amino acid sequence encoded by the light chain J and its heavy chain analogue might influence the character and conformation of the complementarity-determining region of an antibody (the antigen-combining site) (13). Various combinations of V and J regions could, therefore, be an important source of immunoglobulin diversity.

Finally, the five J-region sequences retain regions of homology that may be critical to both DNA recombination and RNA splicing. In particular, they retain major elements of the palindromic sequence CAC($\frac{T}{A}$)GTG that occurs in homologous positions adjacent to the recombination sites in germ-line κ and λ V-region genes. This sequence forms the basis for a hypothetical stem structure that may be drawn between all V- and J-region gene sequences thus far determined and that may represent an intermediate in the recombination between V- and J-region genes.

MATERIALS AND METHODS

Recombinant DNA Techniques. The cloned, 16-kb embryonic mouse κ C region fragment and derivative fragments used in these sequence studies have been described (9). All experiments were carried out in accordance with the National Institutes of Health Guidelines on Recombinant DNA Research at an EK2/P2 level of containment.

Sequencing Technique. The partial chemical degradation method of Maxam and Gilbert (14) was used for all sequencing. The restriction enzyme sites, sequencing strategy, overlaps, and extent of bidirectional sequencing are shown in Fig. 1. Our experience with thin (0.3 mm) gels suggests that sequences determined in a single direction are subject to <5% error. We have tried to use the most reliable portions of each run and have sequenced critical regions twice.

Computer Applications. The sequences have been set in format by using a program, PUBTRANS, written by Jacob V. Maizel, Jr. Homology and dyad symmetry searches were carried out by using the program of Korn and Queen (15).

RESULTS AND DISCUSSION

Sequence of Germ-Line DNA Fragment Containing Five Potential J-Region Genes and the κ C Region Gene. We have found (10) that the recombination event that formed the MOPC-41 light chain gene occurred 3.7-kb from the germ-line κ C region gene, adjacent to a J segment that carries codons for

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: V, variable; kb, kilobase(s); C, constant.

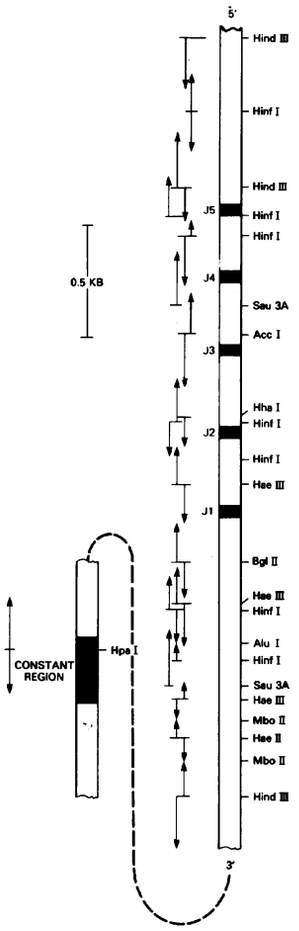

FIG. 1. Strategy for determining sequence of germ-line J- and C-region genes. DNA for sequencing was derived from a 6.5-kb *Eco*RI/*Bam*HI fragment cloned in pBR322. This plasmid was initially cleaved with *Hin*dIII, *Hin*dIII/*Xba* I/*Bam*HI, or *Hin*dIII/*Eco*RI and the resulting fragments were isolated by electrophoresis on polyacrylamide gels followed by electrophoretic elution. These fragments, as well as others generated by the indicated further cleavages, were end-labeled with ^{32}P by treatment with bacterial alkaline phosphatase followed by T4 polynucleotide kinase and [γ-^{32}P]ATP. The double-end-labeled fragments were cleaved at internal restriction sites to obtain single-end-labeled fragments which were subjected to the partial chemical degradation sequence analysis described by Maxam and Gilbert (14). Final electrophoresis on four separate 0.3-mm polyacrylamide/urea gels yielded autoradiograms that were readable up to about 250 base pairs from the labeled end.

amino acids 96–108 of the MOPC-41 light chain. We have now determined the sequence of extensive portions of the region surrounding this segment. The extent of the determined sequence (about 3 kb) and a major portion of the sequence itself is shown in Fig. 2. The sequence reveals a rather striking pattern: beginning with the J segment used to form the active MOPC-41 gene and repeating every 309–354 bases for the next 1.2 kb, a J-region segment appears, identifiable by its homology to the J expressed in MOPC-41. We number these regions J1 through J5, starting at the 3′ side (closest to the C region). Each J region encodes a 13 amino acid sequence; and four of the sequences—J5, J4, J2 and J1—correspond to J segments found in the myeloma light chains of MOPC-41, MOPC-21, MOPC-149, and MOPC-511, respectively (Table 1). The remaining sequence, J3, is discussed below.

No further J regions in the sequenced segment were detected by several computer searches for homology to the five known J regions. It is possible that other κ J regions are encoded in unsequenced portions of this fragment or in fragments not yet cloned. However, the latter seems unlikely because we have not been able to detect other J-containing *Eco*RI fragments by using this J-containing segment as a probe in *in situ* hybridization experiments (unpublished data). Furthermore, no other J regions are present in the 1.5-kb segment on the 3′ side of the last regularly occurring J segment. In addition, within 150 base pairs to the 5′ side of J5 there is a region of reiterated DNA that may represent a border of an otherwise unique sequence. This work has been done with BALB/c mice; analysis of κ light chain amino acid sequences (22) and mRNA precursors (23) in NZB myelomas suggests the existence of multiple J-region genes in NZB mice as well.

The Recombination Event: Clues Provided by Inverted Repeat Sequences Adjacent to the V- and J-Region Recombination Site. The structural homology shared among the five J-segments (Fig. 3) suggests a common ancestral gene. The homology in the coding sequence is striking. Excluding the J3 sequence, 10 of the 13 codons are unchanged except in a few third-base positions (see below). It is likely that the preserved coding sequence is of significance for the configuration of the immunoglobulin molecule, but it may play some further role in DNA or RNA joining as well. Two other regions of preserved homology are of interest. These are the G- and T-rich sequence on the 5′ side of each J segment (underlined in Fig. 3 and also partially preserved in the V-region genes) and the short palindrome—CAC($^{T}_{A}$)GTG—present with some minor variations adjacent to each J-region recombination site (marked with an arrow in Fig. 3). The palindromic sequence is present adjacent to each germ-line V-region recombination site thus far sequenced (marked with an arrow in Fig. 3). These palindromes are complementary inverted repeat sequences and therefore can be visualized as forming an inversion loop or stem structure drawing the V and J sequences together. An example of such a stem structure, drawn between the J5 sequence and the VK41 sequence, is shown in Fig. 4. Similar structures can be drawn between each sequenced κ J and κ V gene as well as between the λ J and its V-region sequence (6, 7).

Although the significance of these palindromic sequences remains to be evaluated, some signal must ensure reasonably precise recombination between V- and J-region segments. If joining V and J does in fact involve the recognition of these short palindromes, it is significant that they do not become a part of the recombined sequence as does, for example, the integration core sequence of phage λ (24). From our studies on the sequences of the MOPC-41 recombinant and its precursors (10), we have deduced the specific nucleotides that were joined by the recombination event that formed the gene. These are shown by asterisks in Fig. 4. Both lie at the base of the hypothetical stem structure; the stem is eliminated by the recombination event. Different nucleotide pairs would be joined in the case of alternative frames of recombination (see below). The lower panel of Fig. 4 shows diagrammatically how both strands of the double helix might interact to form such a stem structure: in one region each strand interacts with its usual complementary partner, whereas in an adjacent region each strand interacts with a distal sequence of the same strand. Presumably, this structure, requiring the disruption of normal interactions of double-stranded DNA, would be thermodynamically unstable and would require special enzymes for formation. Given its completely hypothetical nature (e.g., the relative order of the V and J/C genes is not known), the major advantage of this structure would be in allowing recombination to be accomplished by several known enzymatic activities. Joining the DNA strands at the base of such a stem could be achieved by a nicking enzyme and DNA ligase, or a nicking-closing enzyme, or by bypassing the stem with a DNA polymerase. Any of these mechanisms would eliminate the stem while joining the appropriate V/J sequences. These short palindromic repeats are also reminiscent of the ends of translocatable drug-resistance elements (25), the insertion sequence ISI (26), and the integratable human virus Ad2 (27, 28).

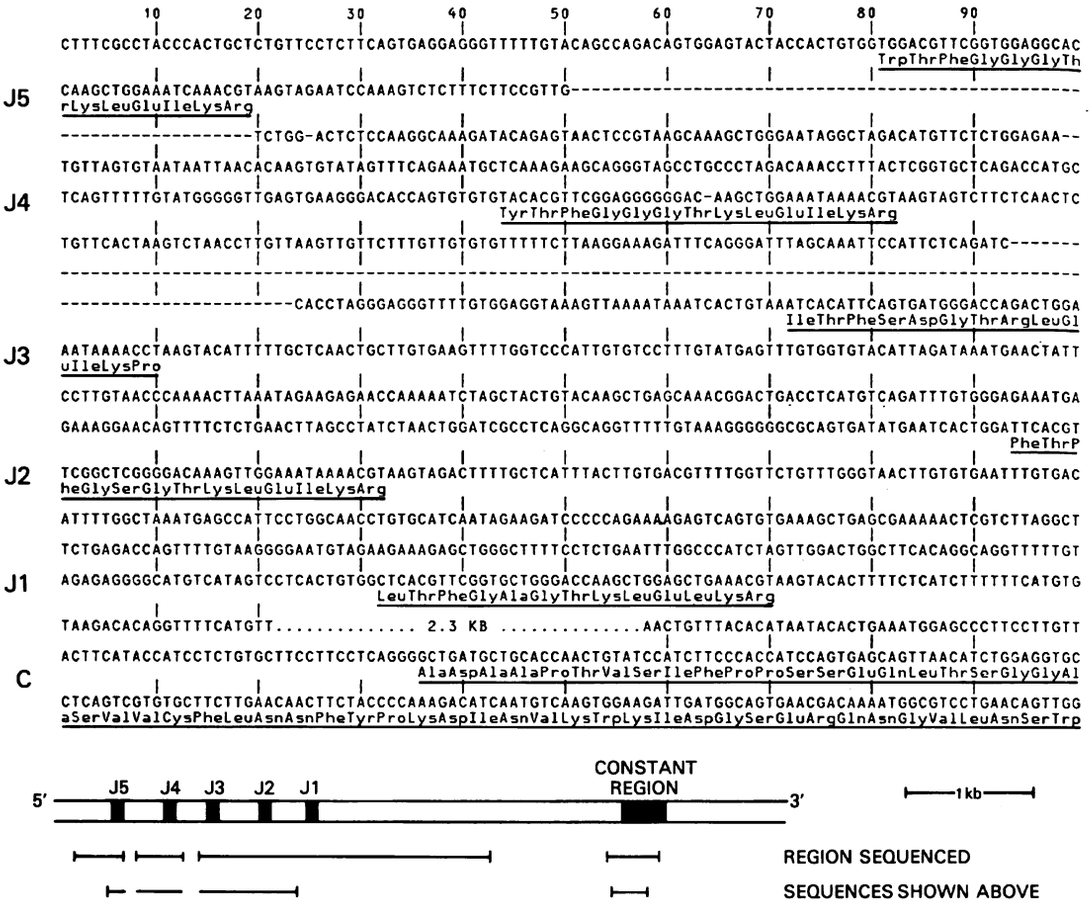

FIG. 2. Sequence of region containing κ J- and C-region genes. Coding regions of the genes are indicated by the amino acids below the nucleotide sequence. The dashes indicate undetermined nucleotides. The lengths of these unsequenced regions are estimated by restriction analysis. The diagram below indicates the extent of the sequence determined and shows which portion of the known sequence is given above. The J regions were identified by their homology with the J of the expressed MOPC-41 gene (J5). They are numbered from the J closest to the C region.

V/J Recombination as Mechanism for Generating Diversity. How can the set of five J-region sequences that we have found be reconciled with the existence of more than five different amino acid sequences in the J region of known κ chains? Table 1 lists the amino acid sequences corresponding to our germ line J regions as well as all the known BALB/c J regions derived from published amino acid sequence studies (29, 30). There are four J regions that correspond to germ-line genes and five that do not correspond to any germ-line J. It is apparent that each of these five non-germ-line sequences may be derived from one of the germ-line regions by changing the codon at position 96. We can explain the non-germ-line sequences by assuming that the recombination enzymes allow some flexibility in the frame of recombination. For instance, the recombination event that joined VK41 with J5 occurred between the second and third nucleotides of codon 95 (CCG):

VK41	J5	recombined	
CCT CCC × TGG TGG	→	CCG TGG.	[10]

If the same sequences recombined between the first and second nucleotides of codon 96, the resulting sequence (CCT CGG) would code for an arginine in position 96, corresponding to the non-germ-line J sequence shown as sequence 2 in Table 1. Table 2 indicates all the codons that could be generated at position 96 by shifting the recombination frame between VK41 and the four germ-line J-region genes. In addition to three of the known non-germ-line J regions, several other J regions that have not yet been observed could be generated. If this model is correct, these J sequences may be found as more light chains are sequenced. If we make an additional assumption—namely, that other VK genes have different codons at position 96—then all the known J-region amino acid sequences could be generated.

Role of the J Segment in the Antibody Molecule. From what is known of the three-dimensional structure of M603, a phosphocholine-binding IgA, it appears that amino acid substitutions in position 96 (which is a part of the complementarity-determining region) would alter the properties of the antigen-binding site (13). Furthermore, positions 96 and 98 of the J region are important for contacts between heavy and light chains, and this might also affect the configuration of the antigen-binding site (31). Analogous positions play a similar role

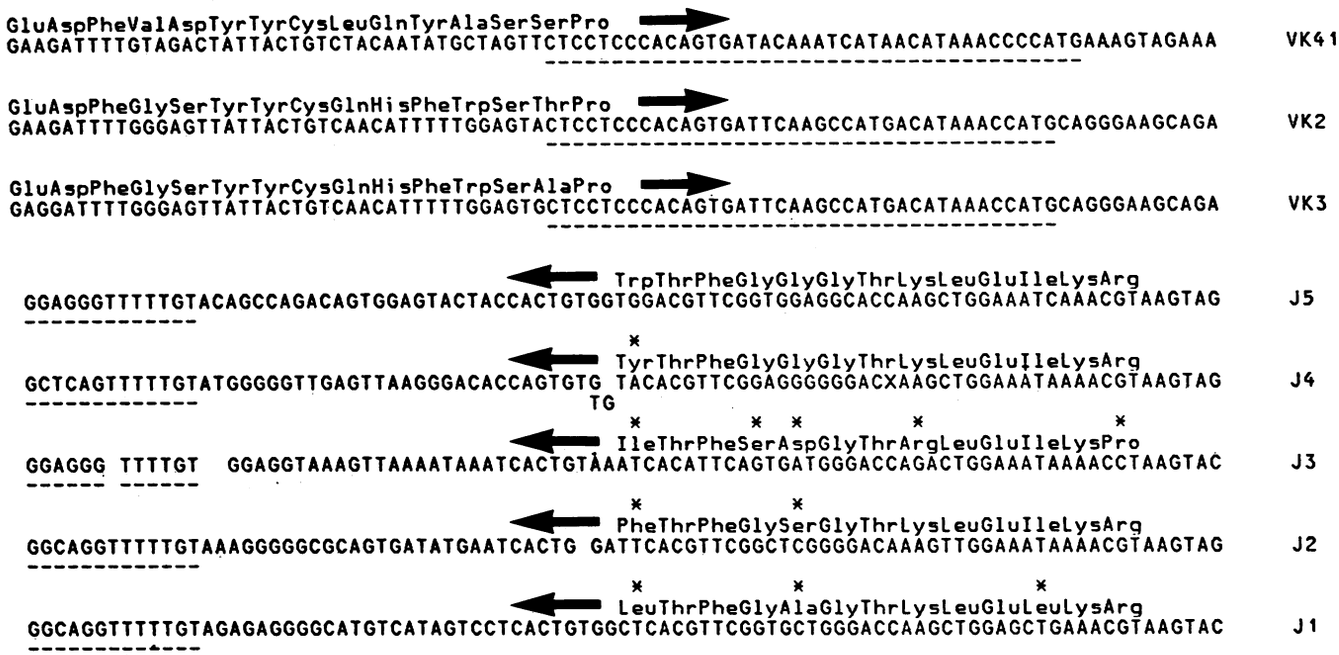

FIG. 3. Nucleotide sequences at the recombination sites of several germ-line V- and J-region genes. K2 and K3 are closely related κ V regions (*Upper*) identified and cloned on the basis of cross hybridization to a cDNA probe specific for the expressed MOPC-149 gene (3). More extensive sequences are published elsewhere for K2 and K3 (3) and for VK41 (10). The J region (*Lower*) sequences of the present work have been aligned to facilitate comparison with each other and with the V regions. Homologous regions are underlined. The septanucleotide (boldface arrows) appearing (with minor variations) in the V regions as well as in the J regions may be important in the V/J recombination event and is discussed in the text. The asterisks identify amino acids that differ from the J5 sequence.

in the heavy chain (32). It seems, therefore, that the different J regions would affect the diversity not only of the primary structure of the antibody molecule but also of the antigen-combining site itself.

The Peculiar J3 Sequence Violates the GT/AG Excision Rule. The sequence encoded by J3 does not correspond to any published BALB/c myeloma sequence. It has alterations in three amino acid positions—99, 103, and 108—which are invariant in all known BALB/c J regions, as well as an aspartate in position 104, which also has not been observed among known BALB/c J regions. Moreover, all the other J-region genes end in the dinucleotide G-T, consistent with the splicing out of an intervening sequence beginning with a G-T and ending with an A-G. Almost all known intervening sequences have the form G-T. . .A-G (33–35). The J3 sequence, however, has a C-T instead of a G-T at the 5′ border of the intervening sequence. The significance of this alteration is not known, but if it interferes with RNA processing, the sequence would obviously not be expressed. If this were the case, it is further possible that the selective forces that act on an expressed sequence no longer maintain the functional sequence of J3. Hence, this J sequence may represent some intermediate stage in the evolutionary deterioration of a J gene.

Table 1. Germ-line J-region gene translations and observed J-region amino acid sequences

Amino acid translation of germ-line BALB/c J-region genes	Amino acid sequences of J regions found in κ chains of BALB/c mice (96 … 108)
	1. W T F G G G T K L E I K R*
	2. R - - - - - - - - - - -
J4 Y - - - - - - - - - - -*	3. Y - - - - - - - - - - -*
	4. P - - - - - - - - - - -
J3 I - - S D - - R - - - - P*	
J2 F - - - S - - - - - - - -*	5. F - - - S - - - - - - - -*
	6. W - - - S - - - - - - -
	7. I - - - S - - - - - - -
J1 L - - - A - - - - - L - -*	8. L - - - A - - - - - L - -*
	9. I - - - A - - - - - L - -

W = Trp, T = Thr, F = Phe, G = Gly, K = Lys, L = Leu, E = Glu, I = Ile, R = Arg, Y = Tyr, P = Pro, S = Ser, D = Asp, A = Ala. The myeloma proteins used as examples are identified and referenced as follows: 1, MOPC 41 (16); 2, MOPC 173 (17); 3, MOPC 21 (18); 4, MOPC 11 (19); 5, MOPC 149 (J. G. Seidman, unpublished results); 6, MOPC 321 (20); 7, X 24 (21); 8, MOPC 511 (E. Apella, personal communication); 9, TEPC 601 (21).
* Germ-line sequence.

Table 2. Codon diversity created by alternative frames of recombination between germ-line, V-, and J-region genes

Germ-line genes	Junction sequence	Codons generated	
VK41	C C C	TGG - Trp*	(1)
		CGG - Arg*	(2)
J5	T G G	CCG - Pro*	(4)
		CCC - Pro*	(4)
VK41	C C C	TAC - Tyr*	(3)
		CAC - His	
J4	T A C	CCC - Pro*	(4)
VK41	C C C	TTC - Phe*	(5)
		CTC - Leu	
J2	T T C	CCC - Pro	
VK41	C C C	CTC - Leu*	(8)
		CCC - Pro	
J1	C T C		

* Accounts for a light chain of known sequence, corresponding to number in Table 1 as indicated by numbers in parentheses.

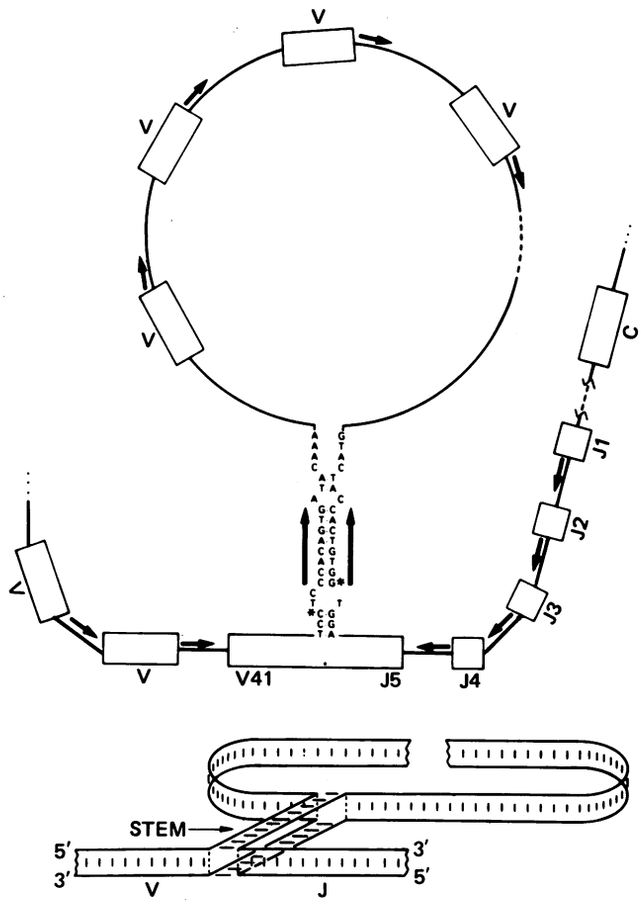

FIG. 4. Hypothetical stem structure formed between inverted repeats located next to germ-line V- and J-region genes. Above is a diagram of hypothetical intermediate in the chromosomal rearrangement κ V-, J-, and C-region genes. The distance between V- and J/C-region genes is not known and this is indicated by the broken line. It must be emphasized that the relative order of the V and J/C genes is also not known. However, preliminary evidence (unpublished results) suggests that the recombination event is accompanied by a deletion of sequences 5′ to the recombining J segment. Each V- and J-region gene is bordered by a palindrome that is also an inverted repeat sequence (boldface arrow) located on the 3′ side of the V genes and on the 5′ side of the J genes. Each of these sequences can be written as a complementary stem structure as shown in the figure in which the actual sequence of MOPC-41 recombinant is used as an example. The bases marked with asterisks are those actually joined to form the recombinant (10). The resulting amino acid codon sequence is indicated. Similar stem structures which draw V- and J-regions together can be written for each κ and λ V- and J-sequence now known (3, 5, 7). Below is a three-dimensional representation of the hypothetical stem intermediate. V- and J-coding sequences interact with their opposite strands in a normal DNA duplex, but the inverted repeat adjacent to the V gene is drawn as a stem interacting with its complement located on the same strand, presumably many thousands of bases away adjacent to a J sequence.

We are grateful to Dr. Jacob V. Maizel, Jr., for his assistance in establishing the computer programs used in these studies and to Dr. David Davies for his helpful comments regarding the potential significance of the J segment in the immunoglobulin molecule. We are also grateful to Terri Broderick for her expert assistance in the preparation of this manuscript.

1. Dreyer, W. J. & Bennett, J. C. (1965) *Proc. Natl. Acad. Sci. USA* **54,** 864–868.
2. Hozumi, N. & Tonegawa, S. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 3628–3632.
3. Seidman, J. G., Leder, A., Edgell, M. H., Polsky, F., Tilghman, S. M., Tiemeier, D. C. & Leder, P. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 3881–3885.
4. Rabbitts, T. H. & Forster, A. (1978) *Cell* **13,** 319–327.
5. Bernard, O., Hozumi, N. & Tonegawa, S. (1978) *Cell* **15,** 1133–1144.
6. Brack, C., Hirama, M., Lenhard-Schuller, R. & Tonegawa, S. (1978) *Cell* **15,** 1–14.
7. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 1485–1489.
8. Seidman, J. G., Leder, A., Nau, M., Norman, B. & Leder, P. (1978) *Science* **202,** 11–17.
9. Seidman, J. G. & Leder, P. (1978) *Nature (London)* **276,** 790–795.
10. Seidman, J. G., Max, E. E. & Leder, P. (1979) *Nature (London)*, in press.
11. Early, P. W., Davis, M. M., Kaback, D. B., Davidson, N. & Hood, L. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 857–861.
12. Sakano, H., Rogers, J. H., Hüppi, I., Brack, C., Traunecker, A., Maki, R., Wall, R. & Tonegawa, S. (1979) *Nature (London)* **277,** 627–633.
13. Padlan, E. A., Davies, D. R., Rudikoff, S. & Potter, M. (1976) *Immunochemistry* **13,** 945–949.
14. Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
15. Queen, C. L. & Korn, J. (1979) *Methods Enzymol.* **64,** in press.
16. Gray, W. F., Dreyer, W. J. & Hood, L. (1967) *Science* **155,** 465–467.
17. Schiff, C. & Fougereau, M. (1975) *Eur. J. Biochem.* **59,** 525–537.
18. Svasti, J. & Milstein, C. (1972) *Biochem. J.* **128,** 427–444.
19. Smith, G. P. (1978) *Biochem. J.* **171,** 337–347.
20. McKean, D., Potter, M. & Hood, L. (1973) *Biochemistry* **12,** 749–759.
21. Rao, D. N., Rudikoff, S. & Potter, M. (1978) *Biochemistry* **17,** 5555–5559.
22. Weigert, M., Gatmaitan, L., Loh, E., Schilling, J. & Hood, L. (1978) *Nature (London)* **276,** 785–790.
23. Perry, R. P., Kelley, D. E. & Schibler, U. (1979) *Proc. Natl. Acad. Sci. USA*, in press.
24. Landy, A. & Ross, W. (1977) *Science* **197,** 1147–1160.
25. Kleckner, N. (1977) *Cell* **11,** 11–23.
26. Ohtsubo, H. & Ohtsubo, E. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 615–619.
27. Garon, C. F., Berry, K. W. & Rose, J. A. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 2391–2395.
28. Wolfson, J. & Dressler, D. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 3054–3057.
29. McKean, D. J. & Potter, M. (1979) *ICN-UCLA Symposium on T- and B-Cell Lymphocytes: Recognition and Function*, in press.
30. Kabat, E. A., Wu, T. T. & Bilofsky, H. (1979) *Sequences of Immunoglobulin Chains*, in press.
31. Padlan, E. A., Davies, D. R., Pecht, I., Givol, D. & Wright, C. (1976) *Cold Spring Harbor Symp. Quant. Biol.* **41,** 627–637.
32. Rao, D. N., Rudikoff, S., Krutzsch, H. & Potter, M. (1979) *Proc. Natl. Acad. Sci. USA*, **76,** 2890–2894.
33. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4853–4857.
34. Catterall, J. F., O'Malley, B., Robertson, M. A., Staden, R., Tanaka, Y. & Brownlee, G. G. (1978) *Nature (London)* **275,** 510–513.
35. Konkel, D., Tilghman, S. & Leder, P. (1978) *Cell* **15,** 1125–1132.