

Joining of immunoglobulin heavy chain gene segments: Implications from a chromosome with evidence of three D-J_H fusions

(recognition elements/Abelson murine leukemia virus/N gene segment/terminal deoxynucleotidyltransferase)

FREDERICK W. ALT* AND DAVID BALTIMORE

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

Contributed by David Baltimore, March 8, 1982

ABSTRACT A chromosomal segment with a unique structure around the immunoglobulin heavy chain joining region (J_H) has been molecularly cloned from an Abelson murine leukemia virus-transformed cell line. Attached to J_H3 in the cloned DNA, in inverted sequence, is the DNA from J_H1 to the J_H2 recognition sequence. The inverted segment is attached at its other end to the 5' recognition sequence of a diversity segment (D). To form this structure, three joining events must have occurred on the same chromosome. One of these events could have been a normal D-J_H joining but the others must have been irregular events including ones that result in inversions. One of the joining events left fused recognition elements from J_H2 and a D whose sequence shows that, during joining, reciprocal joinings of the recognition elements must occur to fuse the heptameric elements back to back. Because joined D and J_H undergo deletion of terminal coding sequence during recombination but the joined heptameric recognition sequences do not contain the deleted sequence, joining must be a nonreciprocal event. Also, extra nucleotides are inserted between D and J_H as part of the joining process; it is suggested that this added sequence is a product of the activity of terminal deoxynucleotidyltransferase at the D/J_H (and probably the V_H/D) joints and that it represents a new element of heavy chain gene structure, the N region.

The variable region of immunoglobulin light and heavy chains is encoded in multiple germ-line DNA elements which are rearranged somatically to form the complete variable region gene (1-5). Formation of a complete heavy chain variable region gene involves at least two recombinational events: joining of a variable gene segment (V_H) to a diversity segment (D) and joining of the D to a joining segment (J_H) produces the complete V_HDJ_H heavy chain variable region several kilobases (kb) 5' from the most proximal constant region (C_H) gene (4, 5). The recombination processes involved with variable region gene formation are apparently mediated by a set of highly conserved recognition sequences which consist of a palindromic heptamer and a characteristic nonamer separated by a spacer region (4, 5). A complete recognition sequence, starting with the heptamer, lies flush with the 3' border of each V_H and D and the 5' border of each D and J_H (4-7). The spacer region is characteristically 23 base pairs (bp) long for V_H and J_H recognition sequences (4, 5) and 12 bp for the 3' and 5' D recognition sequences (6, 7). Apparently, the recombination process can only occur between recognition sequences containing 12- or 23-bp spacers, (4, 5). Thus, the V_H-to-J_H joining process for heavy chains appears to be obligately mediated by V_H/D and D/J_H joinings.

Most of our current understanding of this joining process has

come from studies that compared the organization and structure of the various elements in myeloma cell DNA to those in embryonic DNA. We have been exploring the possibility of using Abelson murine leukemia virus (A-MuLV)-transformed cells as a model system for studying aspects of the early stages of B-lymphoid cell differentiation. These studies have shown that some A-MuLV transformants continue immunodifferentiative events during growth in culture (8-11). Of particular interest are lines derived from transformation of early fetal liver cells. In culture, such lines usually exhibit continued heavy chain rearrangement at or near the J_H region, in most cases probably in the form of continued rearrangement of a preexisting D-J_H complex (8). In this report we describe a novel series of rearrangements at a single J_H of such a fetal liver line and discuss the possible significance of these rearrangements in terms of the general mechanism of V_H-D-J_H recombination.

EXPERIMENTAL

The 38B9 cell line was derived by A-MuLV transformation of 13-day murine fetal liver cells (8). It was apparent that the 38B9 line underwent J_H-associated rearrangements in culture because the parental line contained four distinct, nonembryonic J_H-containing *Eco*RI fragments (8) whereas subclones derived from the line contained one or two of these fragments and often additional J_H-containing *Eco*RI fragments of novel size (not shown). To define the events of this rearrangement process we have molecularly cloned rearranged J_H-associated DNA fragments from this and other lines. Below, we describe a single *Eco*RI fragment containing a uniquely rearranged J_H derived from a 38B9 subclone.

The 38B9-7 subclone had two J_H-containing *Eco*RI fragments, one of 5.2 kb that comigrated with a fragment found in the parental clone and a new fragment of 5.6 kb. Total *Eco*RI-digested DNA from this subclone was cloned into the *Eco*RI site of Charon 16A λ phage, and recombinants containing DNA that hybridized to a J_H-specific probe were purified and amplified as described (9). The 5.6-kb insert from one such clone was subcloned into the *Eco*RI site of pBR322 to yield plasmid pJ_H38B9-7.1 (the insert of which is hereafter referred to as "insert 7.1"). Comparison of a partial restriction map of insert 7.1 and the 6.2-kb *Eco*RI embryonic DNA fragment that contains the J_H indicated that a new sequence had been rearranged into J_H3 (Fig. 1). To analyze the nature of this rearrangement further, we prepared a specific probe from the rearranged portion of the clone

Abbreviations: V, variable segment; D, diversity segment; J, joining segment; C, constant segment; kb, kilobase(s); bp, base pair(s); A-MuLV, Abelson murine leukemia virus.

* Present address: Dept. of Biochemistry, College of Physicians and Surgeons, Columbia Univ., New York, NY 10027.

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.

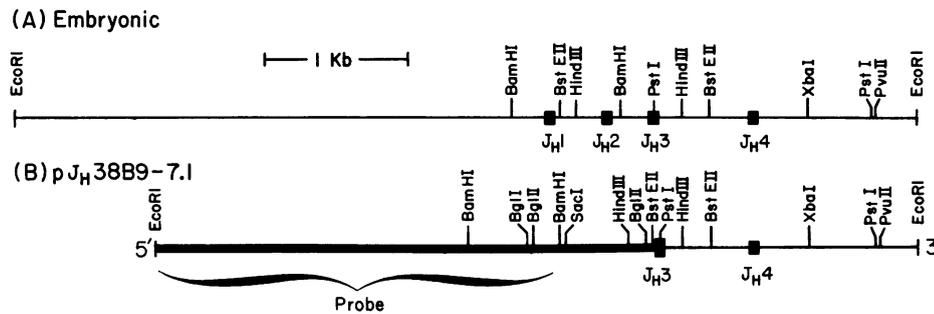


FIG. 1. Partial restriction map of clone pJ_H38B9-7.1. Genomic DNA from cellular subclone 38B9-7.1 was digested to completion with *EcoRI* and cloned into the *EcoRI* site of Charon 16A λ phage. Phage containing inserts that hybridized to a J_H-specific probe were then purified and amplified, and their inserts subcloned into the *EcoRI* site of pBR322. For restriction mapping, the subcloned inserts were excised from pBR322 by digestion with *EcoRI*, labeled with ³²P by nick-translation, and digested with various combinations of enzymes, and the products were analyzed by electrophoresis through agarose gels. All details of these methods have been reported (9). No sites are indicated between the most 5' *Bam*HI and *EcoRI* sites in either map. (A) Partial restriction map of 6.2-kb embryonic *EcoRI* fragment which contains J_H (5). (B) Partial restriction map of clone pJ_H38B9-7.1; thick line indicates the region of nonidentity between the two clones.

consisting of the 5'-terminal 2.0- and 0.7-kb *Bam*HI fragments of insert 7.1. This probe hybridized to a set of embryonic *EcoRI* fragments that was nearly identical to the set containing the previously described families of genomic Ds (12). This result suggested that insert 7.1 contained a D in the absence of a V_H because appending a V_H to a D would delete the 5' flanking sequence that provides the hybridization signal from a D.

To analyze the rearranged portion of this clone further, we determined the sequence of the 800 bases directly 5' to the J_H3

joint as described in Fig. 2. A new sequence was rearranged into J_H3 at a point 3 bases 3' from the beginning of the J_H3 coding sequence (point A in Fig. 2). The J_H1-J_H2 region of the chromosome had been inverted and attached to J_H3. Specifically, the J_H1 segment starting at a point 5 bases from the start of its coding sequence and the entire J_H1-J_H2 spacer sequence had been inverted and linked to J_H3 via the 4-base-pair (bp) sequence $\begin{matrix} T-A-C-C \\ A-T-C-C \end{matrix}$ (point A). The inverted segment extended through the inverted J_H2 nonamer recognition sequence and

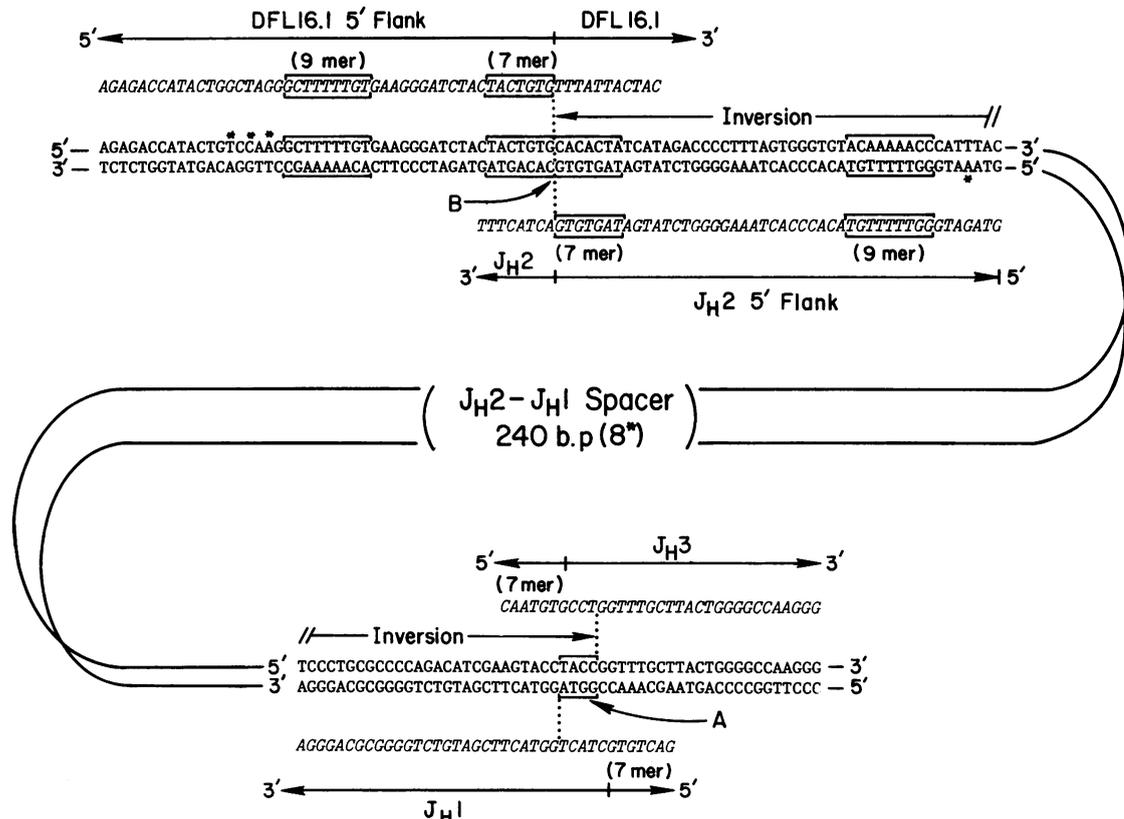


FIG. 2. Partial nucleotide sequence of insert 7.1. The 250-bp *Hind*III-*Bgl* II fragment spanning the J_H3 joint and the adjacent 5' 600-bp *Bgl* II-*Bam*HI fragment were labeled with [γ -³²P]ATP at either end and subjected to sequence determination by the method of Maxam and Gilbert (13) as described (9). Only sequence around the novel joints that are found in insert 7.1 is indicated. Point A (lower part) indicates the 4-bp insert at the joint of J_H3 and the inverted J_H1 sequence. Point B indicates the fusion between the 5' flanking sequence of J_H2 and a D element. For comparison, part of the published embryonic sequences of J_H1, -2, and -3 plus their immediate 5' flanking sequence (5) and the 5' flanking sequence of D_{FL16.1} (12) are shown in italics above and below the corresponding sequences in insert 7.1. The 5'-to-3' orientation of the embryonic sequences, relative to the normal direction of transcription, is indicated. Differences between the sequence of insert 7.1 and the corresponding embryonic segments—including eight differences in the J_H2-J_H1 spacer region—are indicated with asterisks. Characteristic recognition heptamers and nonamers in insert 7.1 are bracketed.

ended abruptly at the terminus of the inverted J_H2 heptamer recognition sequence $\begin{matrix} T-A-G-T-G-T-G \\ A-T-C-A-C-A-C \end{matrix}$ (point B). In place of the inverted J_H2 sequence, at this point the noninverted 5' flanking sequence of a D began. It could be recognized because it differed in only 3 of 46 bases from the published 5' flanking sequence of $D_{FL16.1}$ (12). The D sequence began with its 5' heptamer recognition sequence $\begin{matrix} T-A-C-T-G-T-G \\ A-T-G-A-C-A-C \end{matrix}$ joined precisely on its 3' side to the 3' side (with respect to its embryonic configuration) of the J_H2 heptamer yielding the fused sequence 5' $T-A-C-T-G-T-G-C-A-C-A-C-T-A$ 3' $A-T-G-A-C-A-C-G-T-G-T-G-A-T$ 5' (point B).

DISCUSSION

Model of rearrangement

A schematic version of the structure described above is shown in Fig. 3G. Although this structure could have been generated by various rearrangement events, to explain it we have chosen only to consider events mediated by the D and J_H recognition sequences and therefore subject to the 12-bp/23-bp spacer rule (4, 5). Given this assumption, insert 7.1 must have resulted from

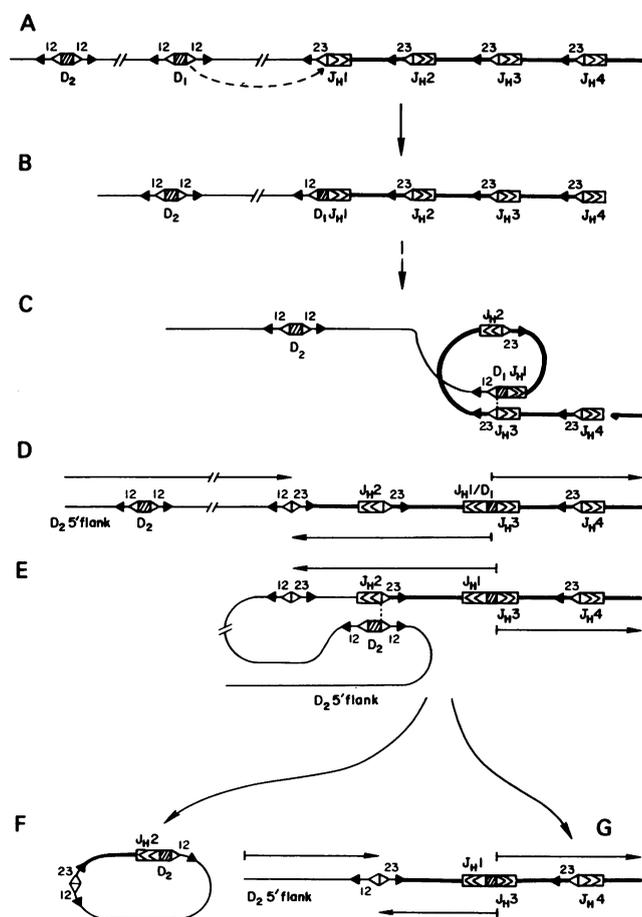


FIG. 3. One possible mechanism for the generation of insert 7.1. The structure of the embryonic J_H region with associated Ds is illustrated in A; the structure of insert 7.1 is shown in G. Intermediate structures or by-products generated during the proposed pathway are illustrated in B-F. \square , J_H in germ-line configuration; \boxtimes , J_H in inverted configuration; \square , D; \triangleleft and \triangleright , 5' and 3' recognition heptamers, respectively; \blacktriangleleft and \blacktriangleright , 5' and 3' recognition nonamers, respectively. Spacer size (in bp) between the heptamer and nonamer signals is indicated above or below the line as 12 or 23. Arrows pointing from left to right indicate sequence in its embryonic configuration; arrows pointing from right to left indicate inverted sequence. Distances are not drawn to scale.

at least three events—one each occurring between a D and J_H s 1, 2, and 3. Although a unique pathway of rearrangement cannot be derived from the data shown here, we will outline one series of rearrangements that leads to the final structure and present further details about individual steps or alternative pathways below.

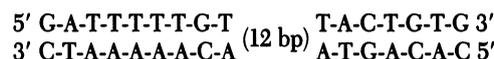
The three steps involved in generating insert 7.1 could be as follows. (i) Starting with the embryonic configuration (Fig. 3A), a normal D- J_H joining event fuses a D (indicated as D_1) to J_H1 (Fig. 3B). (ii) The D_1 - J_H1 structure then loops around (Fig. 3C) and uses the 5' D_1 recognition sequence to promote fusion of the D_1 - J_H1 to J_H3 (Fig. 3D). Such a recombination event inverts the DNA segment from J_H1 to the end of the J_H3 heptameric recognition sequence. In the process, the inverted J_H1 joins to J_H3 by a 4-bp sequence $\begin{matrix} T-A-C-C \\ A-T-G-G \end{matrix}$ derived from D_1 and the J_H3 heptameric recognition sequence becomes fused to the 5' D_1 heptameric sequence (Fig. 3D). (iii) The final rearrangement fuses the 5' D heptameric sequence of a second D (indicated as D_2) to the inverted heptameric sequence of J_H2 (Fig. 3E). This rearrangement event generates two products: one is a circle containing a fusion of D_2 to J_H2 (Fig. 3F); the other is the observed structure in insert 7.1 (Fig. 3G).

Types of D- J_H recombination

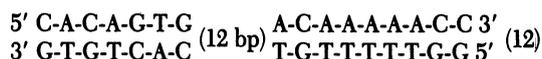
The proposed mechanism for the generation of insert 7.1 involves three types of D- J_H recombinations which we will term "normal," "inverted," and "pseudonormal" (Fig. 4).

Normal D- J_H Joining. This type of joining involves the fusion of the 3' side of a D to a J_H (6). During this process the intervening DNA sequence is presumably excised as a circle (see below) by fusion of 3' D and J_H recognition sequences. The first proposed rearrangement in the formation of insert 7.1 was a normal joining of D to J_H1 (Fig. 3).

Inverted D- J_H Joining. We use the term "inverted D- J_H joining" to describe the process by which a D uses its 5' recognition sequence to mediate joining with a J_H (Fig. 4). This type of recombination theoretically should be possible because a 5' recognition element



read in the inverted direction is nearly identical to a 3' recognition element



and therefore could be appropriately recognized by recombination proteins. Whereas joining events in the normal orientation lead to deletion of the intervening sequence, such an event in the inverted configuration would invert the DNA sequence lying between the recombination points.

In our example, inverted recombination occurs between J_H3 and the 5' side of a D already joined on its 3' side to J_H1 . Thus, the 4-bp sequence $\begin{matrix} T-A-C-C \\ A-T-G-G \end{matrix}$ which links J_H1 and J_H3 (Fig. 2) is presumably the core sequence of a D that was deleted on both its 5' and 3' sides during the two recombination events it underwent. The occurrence of the sequence $\begin{matrix} G-G-T-A \\ C-C-A-T \end{matrix}$ in the core of many Ds (12) supports this proposal. Deletion of bases from both the D and J_H , as observed in the J_H1 -D- J_H3 joint in insert 7.1, apparently is a normal part of this recombination system (3-7, 12, 14, 15). Curiously, the $\begin{matrix} T-A-C-C \\ A-T-G-G \end{matrix}$ sequence is a direct repeat of the final 4 bases of the attached J_H1 segment, but we assume this to be coincidental. Although pathways leading to the generation of insert 7.1 could be postulated in which in-

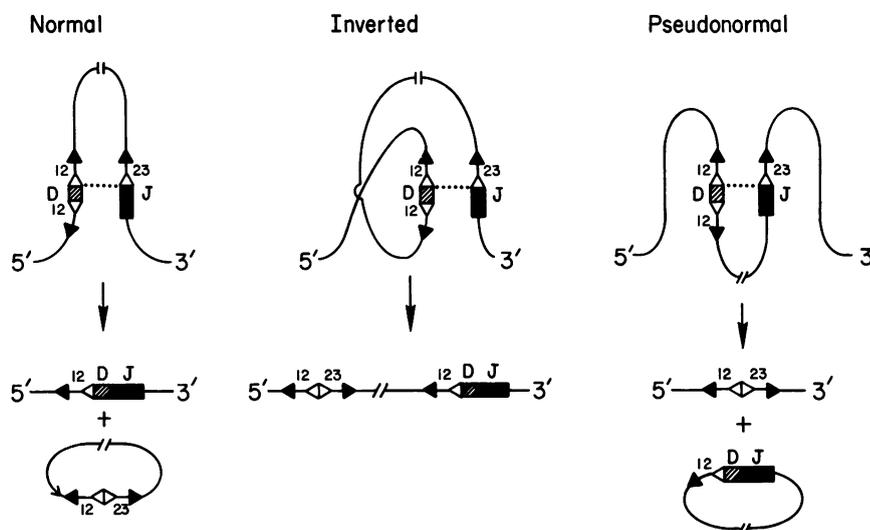


FIG. 4. Forms of D-J_H recombination. Details are discussed in the text. Symbols are as in Fig. 3.

verted joining was the first or last step, all models subject to the 23-bp/12-bp joining rule require an inverted joining event.

Inverted joining is possible because of the bifunctional nature of Ds which must fuse with both a J_H and a V_H to mediate the formation of a complete V_HDJ_H heavy chain variable region gene (4-7). No evidence for such inverted D-J_H (or V_H-D) joining has been reported and it remains to be determined whether such inverted recombination events can mediate the construction of complete V_HDJ_H genes. It could be that the joining process is organized in such a way as to avoid primary inverted joints but that secondary inverted joints (Fig. 3) occur with some frequency because joining of the 3' side of a D to a J_H "activates" the ability of the 5' side to join or because of the proximity of a 5' D recognition element to J_H recognition elements after an initial D joining has occurred.

Pseudonormal D-J_H Recombination. We use this term to describe the process by which the 5' side of a D is linked to a previously inverted J_H (Fig. 4). The joining is referred to as pseudonormal because it involves D and J_H recognition elements oriented in their normal configuration, but the 5' D recognition sequence (rather than the usual 3' one) couples to a previously inverted J_H recognition sequence. The reverse orientation of the process would cause the D-J_H complex to be lost from the chromosome as a circle; the fused recognition elements would be retained. It seems unlikely that pseudonormal joining plays any significant biological role. It should also be noted that other recombination pathways can be constructed that lead to insert 7.1 but involve only normal and inverted D-J_H joining (although one normal joint would have to involve a previously inverted D).

Retention of fused recognition elements

An important aspect of the structure of insert 7.1 is that it retains the reciprocal product of a normal joining event—fused back-to-back recognition heptamers. Recognition elements at a D-J_H joint usually are lost from the cell's DNA. The ligation of the fused recognition elements after the D-J_H joining in insert 7.1 suggests that these elements are also ligated during a normal D-J_H joining. Thus, if the recombination process were intrachromosomal, the intervening sequence would be lost as a circle (Fig. 4). Intrachromosomal recombination must have been involved in the generation of insert 7.1 because inverted joinings between sister chromatids would have resulted in one dicentric chromosome and another lacking a centromere. In addition, the

general absence of sequence 5' to J_H in cells that have undergone D-J_H joining (8, 16) and of Ds in cells with two V_H-D-J_H rearrangements (unpublished data) suggests that sister chromatid D → J_H or V_H → DJ_H joinings are extremely rare events.

Deletion and insertion during joining

It is notable that, although the point of recombination between D and J_H is variable and almost never occurs at the boundaries of both sequences (4-7, 9, 12, 15), in the fused recognition heptamers the joint is a perfect back-to-back fusion of the heptamers (Fig. 2). Similarly, perfectly fused V_κ and J_κ recognition sequences are often found as reciprocal joining fragments in cells that have undergone V_κ-to-J_κ joining (17, 18). Again, however, the V_κ-to-J_κ joining is almost never precise (2, 3, 14). These observations suggest that D-to-J_H (V_L-to-J_L, etc.) joining is not the result of a reciprocal recombination event but is a multistep process. In Fig. 5 we present such a multistep process. Although its details could easily be different, the overall format of the process is likely to occur in this fashion.

At stage I, the partners are shown as being nicked DNA strands because all four chains must be opened before recombination can occur. At stage II, the recognition signals are joined. Because the products of a joining event are not reciprocal, joining of the signals must be a separate event from the joining of the coding elements. The production of a covalently linked circle by this joining event is indicated because the recovery of the joined signals in insert 7.1 implies the occurrence of such a product.

At stage II, the coding sequences do not join but they must be held by protein (dashed box) so that their proximity is retained. D or J_H coding sequence must be removed to account for the loss of bases at the joint (6, 7, 12). This is most easily diagrammed as the activity of a double-strand exonuclease although other possibilities could be imagined.

One of the most puzzling characteristics of the D-J_H joining process is that at the joint there are often extra nucleotides present that do not occur in any known D or in any of the Js. It has been suggested that they may arise from D-D joining (12), but no evidence of such a process has yet been found. The structure at stage III suggests that these nucleotides could be added to the 3' ends available after exonuclease digestion. An enzyme capable of such an addition and present in many A-MuLV transformants is terminal deoxynucleotidyltransferase (19). This enzyme, found at highest concentration in thymus but also present

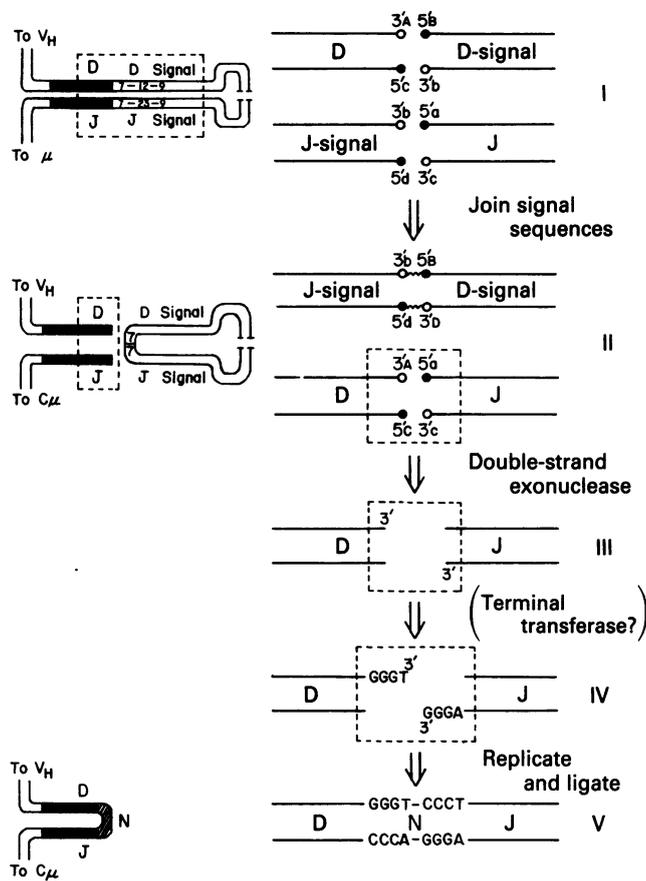


FIG. 5. Model for the D-J_H recombination reaction: generation of DNA element N.

in bone marrow (20), polymerizes random deoxynucleotides at 3' ends but shows a clear preference for dG residues (unpublished data). Remarkably, the extra nucleotides at the D-J joints are often rich in dG residues at the 3' end of both the coding strand and the noncoding strand (see figure 11 in ref. 12 and figure 7 in ref. 9, where the sequence G-T-G-G-G-G-C-C links J_H3 to D_{Q52}).

To finish the joining process, a DNA polymerase and ligase would have to replicate the added bases and finally ligate the structure. It is curious that, when no bases are added between D and J_H, there is often an exact overlap of bases from D and J_H at the joint (see figure 12 in ref. 12)—it is possible that, when even one base of overlap is present, the DNA polymerase begins replicative polymerization.

The process outlined in Fig. 5 generates a new element in variable region structure, the N region (for nucleotides), which can occur between D and J_H (Fig. 5, V) and also between V_H and D. (The V_H-D joint could have N_L; the D-J_H joint could have N_R.) If this proposal is correct, it would complete our knowledge of the origin of the various DNA elements that go into production of a variable region. N regions are not detected in light chain variable region sequences, suggesting that cells that are joining κ segments should lack terminal transferase, as has been observed (unpublished data).

Allelic exclusion

We have previously proposed that heavy chain allelic exclusion might be related in part to a high rate of nonproductive rearrangement but would ultimately rely on a cessation of the rearrangement process upon production of a heavy chain protein

(or C _{μ} region) (8, 9). This model relied on the assumption that D-J_H rearrangements were intermediate in the V_H gene assembly process. The data in this paper support the proposal that D-J_H rearrangements are not abortive because, once formed, a D-J_H can clearly undergo further recognition sequence-mediated recombination.

The rearrangements involved in the generation of insert 7.1 also indicate that a D joining to one J_H does not render the other J_Hs inactive with respect to recombination. However, it should be noted that, once a V_HDJ_H joining has occurred, functional or not, that allele will probably be terminally rearranged because no further Ds would be available to mediate further V_H joining (unpublished data). The possibility of continued rearrangement could be more significant in the generation of light chain genes, in which a nonfunctional V_LJ_L joint potentially could be replaced by joining of an upstream V_L to a downstream J_L.

We thank Elise Thomas for expert technical assistance and Dr. Michael Boss and Ms. Nancy Andrews for their advice and assistance. We are also indebted to Drs. Y. Kurosawa and S. Tonegawa for making available unpublished data and for their stimulating discussions—particularly their observation that no bases are added to joints between Ds and Js that contain overlapping bases. This work was supported by Grant MV-34M from the American Cancer Society, Grant CA14051 (core grant to S. E. Luria) from The National Cancer Institute, and by a contribution from The Whitehead Charitable Foundation. D.B. is an American Cancer Society Research Professor.

1. Brack, C., Hiram, M., Lenhard-Schuller, R. & Tonegawa, S. (1978) *Cell* 15, 1-14.
2. Max, E., Seidman, J. & Leder, P. (1979) *Proc. Natl. Acad. Sci. USA* 76, 3450-3454.
3. Sakano, H., Huppi, K., Heinrich, G. & Tonegawa, S. (1979) *Nature (London)* 280, 288-294.
4. Early, P., Huang, H., Davis, M., Calame, K. & Hood, L. (1980) *Cell* 19, 981-992.
5. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. (1980) *Nature (London)* 286, 676-683.
6. Sakano, H., Kurosawa, Y., Weigert, M. & Tonegawa, S. (1981) *Nature (London)* 290, 562-565.
7. Kurosawa, Y., Boehmer, H., Haas, W., Sakano, H., Trauneker, A. & Tonegawa, S. (1981) *Nature (London)* 290, 565-570.
8. Alt, F. W., Rosenberg, N., Lewis, S., Thomas, E. & Baltimore, D. (1981) *Cell* 27, 381-390.
9. Alt, F. W., Rosenberg, N., Enea, V., Siden, E. & Baltimore, D. (1982) *Mol. Cell Biol.* 2, 386-400.
10. Alt, F. W., Rosenberg, N. E., Thomas, E., Casanova, R. & Baltimore, D. (1982) *Nature (London)* 296, 325-330.
11. Alt, F. W., Rosenberg, N. E., Lewis, S., Casanova, R. J. & Baltimore, D. (1981) in *B Lymphocytes in the Immune Response*, eds. Klinman, N., Mosier, D., Scher, I. & Vitetta, E. S. (Elsevier/North-Holland, New York), pp. 33-41.
12. Kurosawa, Y. & Tonegawa, S. (1981) *J. Exp. Med.* 155, 201-208.
13. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.
14. Weigert, M., Perry, R., Kelley, D., Hunkapiller, T., Schilling, J. & Hood, L. (1980) *Nature (London)* 283, 497-499.
15. Gough, N. M. & Bernard, O. (1981) *Proc. Natl. Acad. Sci. USA* 78, 509-513.
16. Cory, S., Adams, J. M. & Kemp, D. J. (1980) *Proc. Natl. Acad. Sci. USA* 77, 4943-4947.
17. Steinmetz, M., Altenburger, W. & Zachau, H. G. (1980) *Nucleic Acids Res.* 8, 1709-1720.
18. Hoctl, J., Muller, C. R. & Zachau, H. G. (1982) *Proc. Natl. Acad. Sci. USA* 79, 1383-1387.
19. Bollum, F. J. (1974) in *The Enzymes*, ed. Boyer, R. D. (Academic, New York), Vol. 10, pp. 145-171.
20. Silvertone, A. E., Rosenberg, N., Sato, V. L., Scheid, M. P., Boyse, E. A. & Baltimore, D. (1978) in *Differentiation of Normal and Neoplastic Hematopoietic Cells*, Cold Spring Harbor Conference on Cell Proliferation, eds. Clarkson, B., Marks, P. A. & Till, J. E. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), Vol. 5, Book A, pp. 433-453.