an appropriate probe. The heteroduplex molecule, formed between the embryonic J_H clone and the Q52 J_H clone, provides the most direct evidence for the involvement of deletion in D–J_H joining. As the size of the cloned DNA fragments corresponds well to those of the fragments detected by Southern gel blotting analysis, the deletion cannot be attributed to cloning artefacts (see Fig. 1 for clone Q52 J_H).

D–J_H joining and allelic exclusion

Unlike all previous autosomal gene studies, only the gene on one of the two homologous chromosomes is active in a given cell for immunoglobulin chains[31–33]. This 'allelic exclusion' is manifested in several ways for light chains. A common form is that one copy of the chromosome contains a rearranged, complete gene, and the other retains the germ-line configuration[4]. Alternatively, both alleles are rearranged, but one abortively in several different forms (refs 5, 18, 33–36, 45–48 and P. Kennedy and S.T., unpublished). At the heavy-chain locus, rearrangements on both copies of the chromosome seem to be very common (Fig. 1), and lack of rearrangement as observed in light chains does not seem to be a common form of allelic exclusion. The demonstration that the heavy-chain locus includes D DNA segments suggests a unique way in which this locus can be abortively rearranged. As shown here for myeloma QUPC52, the unexpressed chromosome contains a D–J segment. Although allelic exclusion at the heavy-chain locus could be manifested by other forms of abortive rearrangement observed in the light-chain loci, the existence of D segments may provide additional forms. One might imagine that in addition to D–J joining V–D joining can occur.

The analysis of myeloma DNA indicates that rearrangement in the vicinity of J DNA segments is much more frequent in heavy-chain genes than in light-chain genes. This distinction also applies to normal B cells: little, if any J_H was observed in the germ-line configuration[37], whereas unrearranged $J\kappa$ is observed in significant amounts[38]. The reason for this difference is not clear. However, it is reasonable to assume that the frequency of joining of two separate DNA segments increases as the number of segments increases and/or as the distance between them decreases. If there are many D segments (see the accompanying article[17]) and if they are close to J segments as D_{Q52} is, D–J joining could account for more frequent abortive rearrangement in the heavy-chain locus.

We thank Lisa Sultzman, Lena Angman, Patricia Riegert and André Traunecker for technical assistance, Richard Maki and William Roeder for reading the manuscript, and Michael Potter for myelomas. The Basel Institute for Immunology was founded and is supported by F. Hoffmann-La Roche & Co.

Received 7 January; accepted 11 March 1981.

1. Hozumi, N. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3628–3632 (1976).
2. Brack, C. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5652–5656 (1977).
3. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3171–3175 (1978).
4. Brack, C., Hirama, M., Schuller, R. & Tonegawa, S. *Cell* **15**, 1–14 (1978).
5. Lenhard-Schuller, R., Hohn, B., Brack, C., Hirama, M., & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4709–4713 (1978).
6. Bernard, O., Hozumi, N. & Tonegawa, S. *Cell* **15**, 1133–1144 (1978).
7. Seidman, J. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 3881–3885 (1978).
8. Seidman, J. G., Max, E. E. & Leder, P. *Nature* **280**, 370–375 (1979).
9. Sakano, H., Hüppi, K., Heinrich, G. & Tonegawa, S. *Nature* **280**, 288–294 (1979).
10. Max, E. E., Seidman, J. G. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3450–3454 (1979).
11. Maki, R., Traunecker, A., Sakano, H., Roeder, W. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2138–2142 (1980).
12. Early, P. *et al. Cell* **19**, 981–992 (1980).
13. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. *Nature* **286**, 676–683 (1980).
14. Bernard, O. & Gough, N. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3630–3634 (1980).
15. Honjo, T. *et al. Cold Spring Harb. Symp. quant. Biol.* (in the press).
16. Schilling, J., Clevinger, B., Davie, J. M. & Hood, L. *Nature* **283**, 35–40 (1980).
17. Kurosawa, Y. *et al. Nature* **290**, 565–570 (1981).
18. Cory, S., Adams, J. M. & Kemp, D. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4943–4947 (1980).
19. Maki, R., Kearney, J., Paige, C. & Tonegawa, S. *Science* **209**, 1366–1369 (1980).
20. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
21. Yoshida, T. H., Imai, H. T. & Potter, M. *J. natn. Cancer Inst.* **41**, 1083–1089 (1968).
22. Hengartner, H., Meo, T. & Müller, E. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4494–4498 (1978).
23. Tiemeier, D. C., Enquist, L. & Leder, P. *Nature* **263**, 526–528 (1976).
24. Kabat, E. A., Wu, T. T. & Bilofsky, H. in *Sequences of Immunoglobulin Chains* (NIH, Bethesda, 1979).
25. Seidman, J. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 6022–6026 (1980).
26. Cory, S. & Adams, J. M. *Cell* **19**, 37–51 (1980).
27. Honjo, T. & Kataoka, T. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2140–2144 (1978).
28. Yaoita, Y. & Honjo, T. *Nature* **286**, 850–853 (1980).
29. Rabbitts, T. H. *et al. Nature* **283**, 351–356 (1980).
30. Cory, S. *et al. Nature* **285**, 450 (1980).
31. Pernis, B. *et al. J. exp. Med.* **122**, 853–875 (1965).
32. Cebra, J. S., Colberg, J. E. & Dray, S. *J. exp. Med.* **123**, 547–558 (1966).
33. Max, E. E. *et al. Cell* **21**, 793–799 (1980).
34. Seidman, J. G. & Leder, P. *Nature* **276**, 790–795 (1978).
35. Perry, R. P. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 1937–1941 (1980).
36. Seidman, J. G. & Leder, P. *Nature* **286**, 779–783 (1980).
37. Nottenberg, C. & Weissman, I. L. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
38. Joho, R. & Weissman, I. L. *Nature* **284**, 179–181 (1980).
39. Wahl, G. M., Stern, M. & Stark, G. R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3683–3687 (1979).
40. Sakano, H. *et al. Nature* **277**, 627–633 (1979).
41. Benton, W. D. & Davis, R. W. *Science* **196**, 180–182 (1977).
42. Maxam, A. & Gilbert, W. *Meth. Enzym.* **65**, 498–502 (1980).
43. Newell, N., Richards, J. E., Tucker, P. W. & Blattner, F. *Science* **209**, 1128–1132 (1980).
44. Coleclough, C., Cooper, D. & Perry, R. P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1422–1426 (1980).
45. Altenburger, W., Steinmetz, M. & Zachau, H. G. *Nature* **287**, 603–607 (1980).
46. Choi, E., Kuehl, M. & Wall, R. *Nature* **286**, 776–779 (1980).
47. Schnell, H. *et al. Nature* **286**, 170–173 (1980).
48. Steinmettz, M., Zachau, H. G. & Mach, B. *Nucleic Acids Res.* **6**, 3213–3229 (1979).

Identification of D segments of immunoglobulin heavy-chain genes and their rearrangement in T lymphocytes

Yoshikazu Kurosawa, Harald von Boehmer, Werner Haas, Hitoshi Sakano, André Trauneker & Susumu Tonegawa

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

The finding that the diversity (D) and joining (J_H) but not the variable (V_H) DNA segments of mouse immunoglobulin heavy-chain genes are joined in the DNA of some cloned cytolytic T cells, led to identification and sequencing of three different D DNA segments. Two segments identified on the embryo DNA carry on both the 5' and 3' sides two sets of characteristic sequences separated by a 12-base pair spacer, which have been implicated as recognition signals for a recombinase. The third segment, identified in a form joined with a J_H DNA segment in a T cell, carries the recognition signal on the 5' side. These results support the 12/23-base pair model for somatic generation of immunoglobulin V genes, and rule out the possibility that the cytolytic T cells use assembled V_H, D and J_H sequences to encode their antigen receptors.

THE variable region of an immunoglobulin chain is encoded in multiple DNA segments scattered along a chromosome of a germ-line genome[1–10]. These DNA segments are assembled into a continuous stretch with concomitant deletion of the spacer

0028–0836/81/160565—06$01.00

© 1981 Macmillan Journals Ltd

sequences during differentiation of B lymphocytes. The somatic rearrangement of DNA sequences is considered to have key roles in both somatic amplification of antibody diversity and determination of the specificity of lymphocyte clones. In mouse light chains of both λ and κ types, two DNA segments, V and J, are separate in the germ-line genome and are joined in the lymphocytes which express these DNA segments[1–5]. The V and J segments code for about 95 residues at the NH_2-terminal and about 13 residues at the CO_2H-terminal ends of the V region, respectively[1,3–5]. In contrast, DNA sequencing analysis of heavy-chain genes suggests that the V_H region is encoded by three DNA segments referred to as V_H, D and J_H, and somatic joining of the three DNA segments is necessary to generate a complete heavy-chain V gene[6–8]. The D DNA segment encodes the third hypervariable region, which determines in part the shape and size of the antigen-combining site[11,12].

The abovementioned sequence rearrangement of immunoglobulin genes should occur during the development of B lymphocytes before the appearance of the first detectable immunoglobulins. Does an analogous or similar rearrangement occur in the precursors of T lymphocytes, which share a common stem cell with B lymphocytes? This question is particularly important because of the possible use of the immunoglobulin V gene pool by T cells for coding their antigen-specific receptors[13–16]. One approach to this question is to analyse DNA of cloned T cells by the Southern gel blotting procedure using a J DNA sequence as the hybridization probe. A V–J or V–D–J joining will be monitored by a difference in the size of DNA fragments detected in embryo (or kidney) and T-cell DNA. We have therefore applied this technique to several cloned cytolytic T cells, and show here that sequence rearrangements frequently occur in T cells in the vicinity of the J_H sequences. Similar findings have been reported by two other groups with T-lymphoma lines[17] and cytolytic T-cell lines[18]. However, our further analysis of two T-cell clones by recombinant DNA techniques demonstrates that the rearrangements observed in these T-cell clones are D–J joinings rather than V–D–J joinings. This led to the identification of two D DNA segments on the germ-line genome. One of the three D DNA segments described here was found to be identical to the D segment that was independently identified in our laboratory during the parallel analysis of an abortively rearranged sequence observed in the myeloma QUPC-52 (ref. 19).

T-cell DNA sequence is frequently rearranged in the vicinity of the J_H cluster

We have previously identified a cluster of four J_H DNA segments[7] on a 6.4-kilobase *Eco*RI fragment cloned from BALB/c embryos (clone MEP203)[6]. A 1.1-kilobase *Sac*I–*Eco*RI fragment derived from the 3′ end (relative to orientation of transcription) of this fragment (a J_H probe) would detect, when used as a hybridization probe on Southern blots of cellular DNA digested with *Eco*RI, sequence rearrangements that have occurred within the 6.4 kilobase *Eco*RI fragment. The results of various cytolytic T-cell clones and three thymomas are shown in Fig. 1*a*. The cytolytic T-cell clones, except B6.1, were confirmed for their functional activity before extraction of DNA. The embryo or kidney DNAs of five different mouse strains—BALB/c, C3H, C57BL/6, AKR and DBA—all gave the expected fragment of 6.4 kilobases previously identified. The DNA of three cytolytic T-cell clones—DAS 1, D.FL.2, D.FL.13—and two thymomas—BW 5147 and T5—each gave a single band indistinguishable from that of embryos or kidneys, suggesting that no gross sequence rearrangement has occurred in these cells within the 6.4-kilobase *Eco*RI fragment. [The faint bands observed above the 6.4-kilobase band in D.FL.2 and D.FL.13 DNA are due to incomplete digestion as demonstrated by separate experiments (data not shown).] In contrast, DNA of six cytolytic T cells—C.SP.2, C.FL.1, B6.1, D.FL.1, D.FL.12 and D.FL.16—gave one additional band of 5.5, 5.3 or 4.0 kilobases. This indicates that some of the two or more copies of chromosome 12 [some T cells are polyploid (H.v.B. and W.H., unpublished)] in these T cells have undergone sequence rearrangement involving the 6.4-kilobase embryonic fragment. In one cytolytic T-cell clone, D.FL. 4, and one thymoma, WC2, no 6.4-kilobase embryonic fragment was observed. The presence of two non-embryonic bands, of 5.5 and 5.3 kilobases, in D.FL.4 suggests that sequence rearrangement has occurred on both (assuming diploidy) chromosomal copies in different ways. In contrast, WC2 seems to contain only one copy of the 6.4-kilobase sequence and sequence rearrangement has occurred in this fragment. Thus, we have observed sequence rearrangement in 8 out of 13 cytolytic T-cell or thymoma clones. In the majority, rearrangement occurred only in some but not all of the copies of chromosome 12. The sizes of the rearranged fragments seem to fall into three categories—5.5, 5.3 and 4.0 kilobases.

Table 1 Germ-line and myeloma D sequences

Germline D			
$D_{SP2.1}$	TC	TACTATGGTAAC	
$D_{SP2.2}$	TC	TACTATGATTACG	AC
D_{Q52}	C	AACTGGGAC	
Myeloma D			
S107		TACTACGGTAGTA	
M603		TACTACGGTAGTA	C
M167	G	GACTACGGTAATA	GCTACTTTG
MC101	AGG	GACTACGTTAGTA	GGTACGACCC
MPC11	GGGATT	TACTACAATAGTA	GCCC
M141	CGTCTCAATTTAT	TACTACGGTCGTA	GCGACAAATACTTCACTT
M173-II	TGCCCCTCTAC	TACTACAGTTACA	GCGGGGGTT
H76			CCGGGGGTCCCC

Three germ-line D sequences identified in the present study are compared with the D region sequences of the eight complete V_H genes present in the indicated myelomas. For $D_{SP2.2}$ and D_{Q52} the sequences between the palindromic heptamers are shown. The $D_{SP2.1}$ has been identified only in a form joined with the J_{H3} segment and therefore its exact 3′ boundary is unknown. For S107 (ref. 8), M603 (ref. 8), M167 (ref. 8), MC101 (ref. 27) and M141 (ref. 7) the DNA sequences of both the complete (rearranged) and the embryonic (unrearranged) V_H genes are known. These sequences and the embryonic J_H sequences[7] were used to deduce the 5′ and the 3′ boundaries of these D regions. For MPC11 (ref. 28), M173-II (H.S. and Y.K., unpublished) and H76 (ref. 9), only the DNA sequences of the complete V_H genes are known and therefore the 5′ boundaries are ambiguous. The 3′ boundaries were deduced as above. All sequences are aligned for the maximal homology. The D regions of S107 and M603 contain a 13-mer (underlined) which is a perfect palindrome around the central G. The homology present between each of the other D sequences and the 13-mer is also indicated by underlining. The homology between H76 and M173-II D sequences is indicated by double underlining.

Two DNA clones from two cytolytic T cells

We chose B6.1 and C.SP2 for further studies and cloned the rearranged DNA fragments of 5.5 and 5.3 kilobases, respectively, using a λ_{WES} vector[20]; we refer to these fragments as B6B1 (short-term B6) and SP2B1 (short-term SP2), respectively. Electron microscopic examination of the heteroduplex molecules formed between either of these T-cell DNA clones and the 6.4-kilobase embryonic *Eco*RI fragment (clone ME183–8) allowed mapping of the recombination sites. The B6 and embryo DNA fragments gave a single-stranded DNA loop flanked by double-stranded DNA arms (Fig. 2*a*). In contrast, the SP2 and embryo DNA fragments gave Y-shaped heteroduplexes. Measurement of the various parts of the heteroduplex molecules indicated that the B6 DNA fragment has undergone a 1.2-kilobase deletion in a region immediately adjacent (5′ side) to the J_{H2} sequence. In contrast, the entire sequence 5′ to J_{H3} is

©1981 Nature Publishing Group

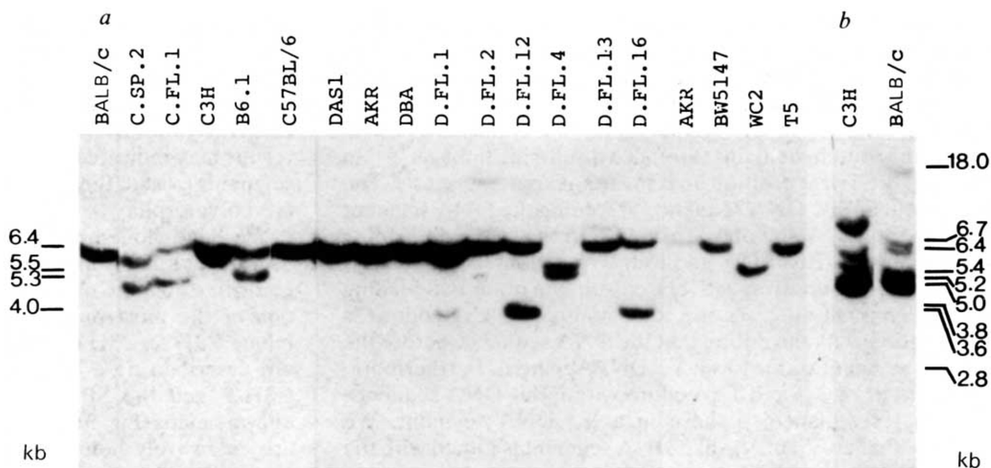


Fig. 1 *a*, Hybridization of T-cell and thymoma DNA with the J_H probe. High molecular weight DNA was prepared from various cloned cytolytic T cells and thymomas by the procedure described previously[29], digested to completion with *Eco*RI, and fractionated on 0.8% agarose gel by electrophoresis. The DNA was denatured and transferred *in situ* to nitrocellulose filters (Schleicher and Schuell, BA85) essentially as described previously[30]. The filters were incubated for hybridization with nick-translated J_H probe (1.1-kilobase *Sac*I–*Eco*RI fragment indicated in Fig. 3*A*) according to the published procedures[2]. Characteristics of the various cytolytic cells have been described previously[31,32]. C.SP.2 and C.FL.1 are derived from C3H, B6.1 from C57BL/6, DAS 1 from AKR/DBA F1, all D.FL clones from DBA, and thymomas were all derived from AKR. For comparison, DNA from BALB/c embryos and kidneys of C3H, C57BL/6, AKR and DBA were also analysed. These DNA are indicated by strain names. Numbers and arrows to the left indicate sizes of DNA fragments in kilobases (kb). *b*, Hybridization of BALB/c embryo and C3H kidney DNA with SP2D-J probe. A nitrocellulose filter, prepared as described for *a*, was incubated with the nick-translated SP2D-J probe, 0.8-kilobase *Bgl*II–*Hin*dIII fragment indicated in Fig. 3*C*. The sizes of the BALB/c DNA fragments detected are shown in kilobases. The 3.8-, 3.6- and 2.8-kilobase bands could be seen in the original autoradiogram, but are too faint to be reproduced.

replaced with a sequence of unknown origin in the SP2 DNA fragment. The points of sequence divergence between the T and embryo cell DNA fragments were further determined by comparison of restriction enzyme cleavage sites (Fig. 3*A–C*). The results confirm the conclusion drawn by heteroduplex mapping.

The SP2 fragment contains a joined D–J DNA segment

We determined the DNA sequence of the 0.8-kilobase *Bgl*II–*Hin*dIII fragment (Fig. 3*Ca*) of the SP2 fragment containing the site of recombination according to the strategy shown in Fig. 3*Cb*. This sequence was aligned with the previously determined sequence of the embryonic fragment in the J_{H3}-coding region (Fig. 4). As suggested by the heteroduplex molecules (Fig. 2*b*) and the restriction enzyme maps (Fig. 3*A*, *C*), the sequences of the two DNA fragments match well starting with the first base of the Phe codon TTT located in the 5′ edge of the J_{H3} DNA segment. On the SP2 fragment the Phe codon is preceded by an Asn codon, AAC, which is in turn preceded by a nonamer, TACTATGGT, that codes for a trimeric peptide, Tyr-Tyr-Gly, when read in phase with the J_{H3}. This nonameric nucleotide sequence or its close variants appear in the D regions of seven out of eight myeloma V_H genes sequenced to date (Table 1). Furthermore, the trimeric peptide, Tyr-Tyr-Gly, encoded by the nonamer has been found in the D regions of a majority of myeloma V regions for which amino acid sequence data are available[11]. These data strongly suggest that a D DNA segment is joined with the J_{H3} sequence on the SP2 fragment.

Further evidence for this notion comes from the feature of the nucleotide sequence located immediately 5′ to the putative D DNA segment on the SP2 fragment. Based on the DNA sequencing analysis of the various germ-line DNA segments, it has been predicted that a nonamer,

GGTTTTTGT
CCAAAAACA

and a palindromic heptamer,

CACAGTG
GTGTCAC

separated by a 12-base pair spacer or closely related variants of these sequences, precedes a germ-line D DNA segment[7,8]. As shown in Fig. 4, the putative D segment joined with the J_{H3} on the SP2 fragment is preceded by two sequences,

GATTTTTGT and TACTGTG
CTAAAAACA ATGACAC

separated by a 12-base pair spacer. These results partially fulfill the predictions of the 12/23-base pair spacer model[7,8], and in turn support the conclusion that the sequence joined with the J_{H3} on the SP2 fragment is indeed a D DNA segment. We refer to this D DNA segment as $D_{SP2.1}$.

No V_H–D–J_H joining on the SP2 fragment

Is a germ-line V_H DNA segment also joined on the SP2 fragment? A visual inspection of the SP2 DNA sequence in the region preceding the $D_{SP2.1}$ segment indicates that no germ-line V_H or V_H-like sequence is attached to the D DNA segment. First, there are one, two or four translation termination codon(s) in the three different reading frames of the 300-base pair (an approximate size of a germ-line V_H segment) region 5′ to the $D_{SP2.1}$ segment. Thus, this region cannot constitute part of a

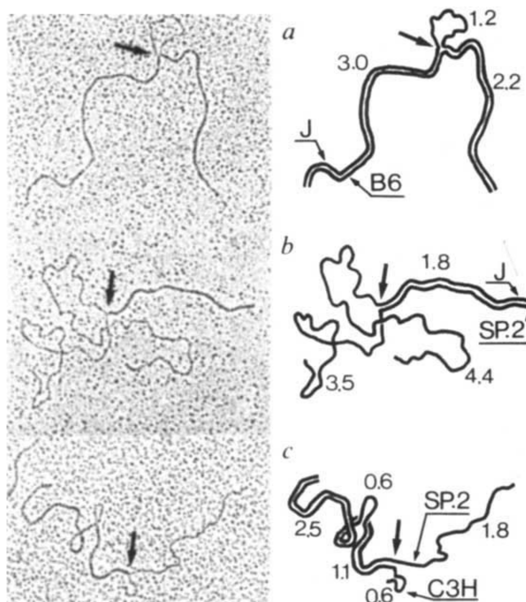


Fig. 2 Heteroduplex molecules formed between cloned DNA fragments. Cloned DNA was digested with *Eco*RI and examined under an electron microscope as described previously[2]. Thick arrows indicate points at which heteroduplexes branch out. The numbers indicate lengths of various parts of the molecules in kilobases. In this experiment we used clone ME184-8 instead of MEP203[6,7] as a source of the 6.4-kilobase embryonic *Eco*RI fragment. Clone ME184-8 was prepared by Richard Maki and contains, in a phage vector, λ_{WES}, the 6.4-kilobase *Eco*RI fragment carrying the four J_H DNA segments. *a*, 6.4-kilobase ME184-8 (J in short-form) versus 5.3-kilobase B6B1 (B6); *b*, 6.4-kilobase ME184-8 (J) versus 5.5-kilobase SP2B1 (SP2); *c*, 5.5-kilobase SP2B1 (SP2) versus 5.0-kilobase C3H-2 (C3H).

©1981 Nature Publishing Group

germ-line V_H DNA segment. Second, some amino acid residues are highly conserved among the V_H regions. In particular, Cys residues are invariably present in all species studied to date at positions 22 and 92 and are the only Cys residues in the V_H region[10]. These two Cys residues play an important part in folding the protein domains through a disulphide linkage[21,22]. In addition, the Trp at position 36 is invariant across species[11]. The 300-base pair SP2 DNA sequence preceding the $D_{SP2.1}$ segment contains three Cys and one Trp codons in the frame defined by the J_{H3} peptide. Eight Cys and four Trp codons, and four Cys and one Trp codons, respectively, exist in the other two reading frames. Thus, in all cases the total number of Cys codons is incompatible with the notion that the DNA sequence before the D DNA segment comprises a V_H DNA segment. Furthermore, the locations of Cys and Trp codons within this DNA sequence bear no resemblance to those in a V_H DNA segment. We conclude that no V_H or V_H-like DNA segment is joined with the D–J_H DNA on the SP2 fragment.

Identification of a $D_{SP2.1}$-like D DNA segment in a germ-line genome

Identification of a D–J_H joining in the DNA of a cytolytic T-cell clone allowed the identification, cloning and characterization of germ-line D DNA segments. We dissected out the 0.8-kilobase *Bgl*II–*Hin*dIII fragment (Fig. 3*C*) from the SP2 clone and used it as a hybridization probe (SP2DJ probe) for analysis of *Eco*RI-digested total DNAs of both BALB/c and C3H embryos (Fig. 1*b*). As this probe contains the J_{H3} and its 3′-flanking sequence, it detects the 6.4-kilobase embryonic *Eco*RI fragment carrying

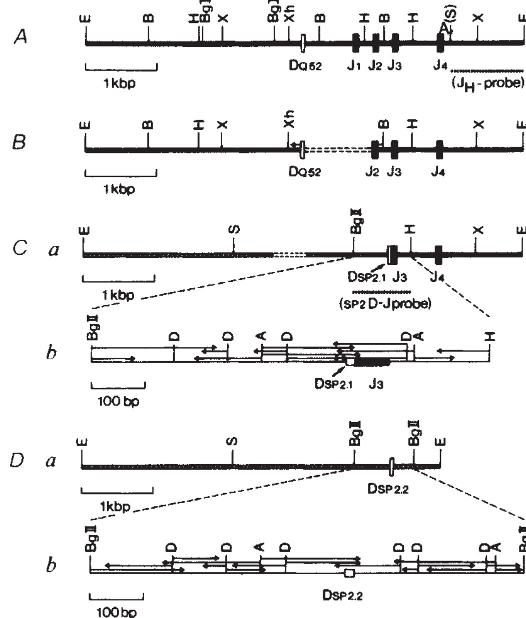


Fig. 3 Restriction enzyme cleavage maps of the *Eco*RI inserts of various DNA clones. *A*, The 6.4-kilobase *Eco*RI fragment of MEP203[6,7]; *B*, B6B1; *C*, SP2B1; *D*, C3H-2. Clone MEP203, which was described previously[6,7], is an embryonic DNA clone and contains three *Eco*RI fragments, one of which is the 6.4-kilobase fragment carrying the four J_H segments. Cleavage sites were determined by single and double digestions of the cloned DNA by various restriction enzymes. Homologous regions are indicated by bars of two types of shading. The broken lines in *B* indicate the 0.6-kilobase region that appeared as a deletion loop in the heteroduplex molecules formed between the SP2B1 and C3H-2 fragments. The filled boxes correspond to the four J_H DNA segments previously identified on the MEP203 6.4-kilobase fragment[6,7]. The open boxes are *D* segments identified by DNA sequencing. Sequencing strategy and finer restriction maps of the 0.8-kilobase *Bgl*II–*Hin*dIII fragment from SP2B1 and the 0.85-kilobase *Bgl*II–*Bgl*II fragment from C3H-2 are shown in *Cb* and *Db*, respectively. Horizontal arrows indicate the directions and extent of sequencing. The sequence across the 1.2-kilobase deletion of the B6B1 fragment was determined using the *Xba*I–*Bam*HI fragment (*B*). The J_H probe used in Fig. 1*a* and the SP2D probe used in Fig. 1*b* are indicated in *A* and *Ca*, respectively. Abbreviations used are: E, *Eco*RI; B, *Bam*HI; H, *Hin*dIII; Bgl, *Bgl*I; X, *Xba*I; Xh, *Xho*I; S, *Sac*I; BglI, *Bgl*II; D, *Dde*I; A, *Ava*II; kbp, kilobase pairs; bp, base pairs.

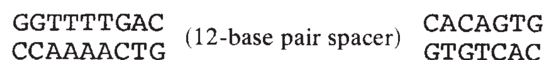
the four germ-line J_H segments. In addition, the probe detected one (5.0 kilobase), four (18, 6.7, 5.4 and 5.2 kilobase) and three (3.8, 3.6 and 2.8 kilobase) bands of high, medium and low intensity, respectively, in BALB/c DNA. A similar, but distinctive hybridization pattern was obtained in C3H DNA. These results may indicate that a number of $D_{SP2.1}$ or $D_{SP2.1}$-like DNA segments exist in the germ-line genomes of mice, some of which are polymorphic.

We have cloned most of these DNA segments in phage λ vectors and determined their structure and organization in the germ-line genome (Y.K., unpublished). Some characteristics of one of the most prominent fragments, that of 5.0 kilobases (clone C3H-2, C3H in short-term) isolated from C3H embryos, are described here. The heteroduplex formed between clone C3H-2 and the SP2 fragment (Fig. 2*c*) and the restriction enzyme maps (Fig. 3*C*, *D*) indicate that the two DNA fragments are extensively homologous in the 3.6-kilobase region which starts with the 5′ *Eco*RI end and ends somewhere near the position of $D_{SP2.1}$ on the SP2 fragment. However, that the homology is imperfect is clearly shown by the presence of a 0.6-kilobase insert (or deletion) in the C3H (or SP2) clone. Nevertheless, the above result suggests that the clone C3H may contain a $D_{SP2.1}$-like DNA segment at the 3′ edge of the homology. We therefore prepared a finer restriction map around this region and determined the entire DNA sequence of the 0.85-kilobase *Bgl*II–*Bgl*II fragment, according to the strategy shown in Fig. 3*Db*.

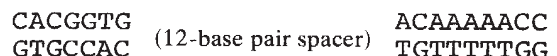
As predicted, the sequence homology between the SP2 and C3H fragments is extensive, starting with the 5′-*Bgl*II site and ending with the triplet AAC, the last codon of $D_{SP2.1}$ on the SP2 fragment (Fig. 4). There are 28 single-base substitutions or deletions (or insertions) in the 485-base pair homology region, two of which are within the D DNA segments. The D DNA segment on the embryonic clone C3H is flanked by two sets of the putative recognition sequences with 12-base pair spacers. This confirms the universality of the 12/23-base pair spacer model[7,8]. Although the C3H-2 D DNA segment and $D_{SP2.1}$ are highly homologous, the imperfectness of the homology in the 5′-flanking regions as well as in the coding segments themselves clearly indicate that they are two different copies of the D segments that occupy independent positions in the germ-line genome. We therefore refer to the C3H-2 D DNA segment as $D_{SP2.2}$.

The B6 fragment contains another joined D–J sequence

Figure 5 shows the sequence across the 1.2-kilobase deletion of the B6 fragment and sequences of the embryonic fragment that cover the 5′ and 3′ boundaries of the deletion. The sequence homology between the B6 and the embryonic fragments ends with the C·G base pair immediately to the left of the line Ld and resumes with a T·A base pair immediately to the right of line Rd. The latter base pair corresponds to the second letter of the Phe codon TTT located near the 5′ terminus of the previously identified J_{H2} sequence[7,8]. The single base pair C·G present on the B6 fragment and flanked by the two lines, Ld and Rd, is unaccounted for by the embryonic sequence (see below). Ld is preceded by a putative enzyme recognition sequence,



both on the B6 and MEP203 fragments. On the latter fragment Ld is followed by nine base pairs, which are in turn followed by another putative signal sequence,



Thus, the 9-base pair sequence on the MEP203 fragment fulfills the requirement imposed on a D DNA segment by the 12/23-base pair spacer rule[7,8]. We therefore conclude that the B6 deletion is due to the joining of the J_{H2} sequence with a D DNA

©1981 Nature Publishing Group

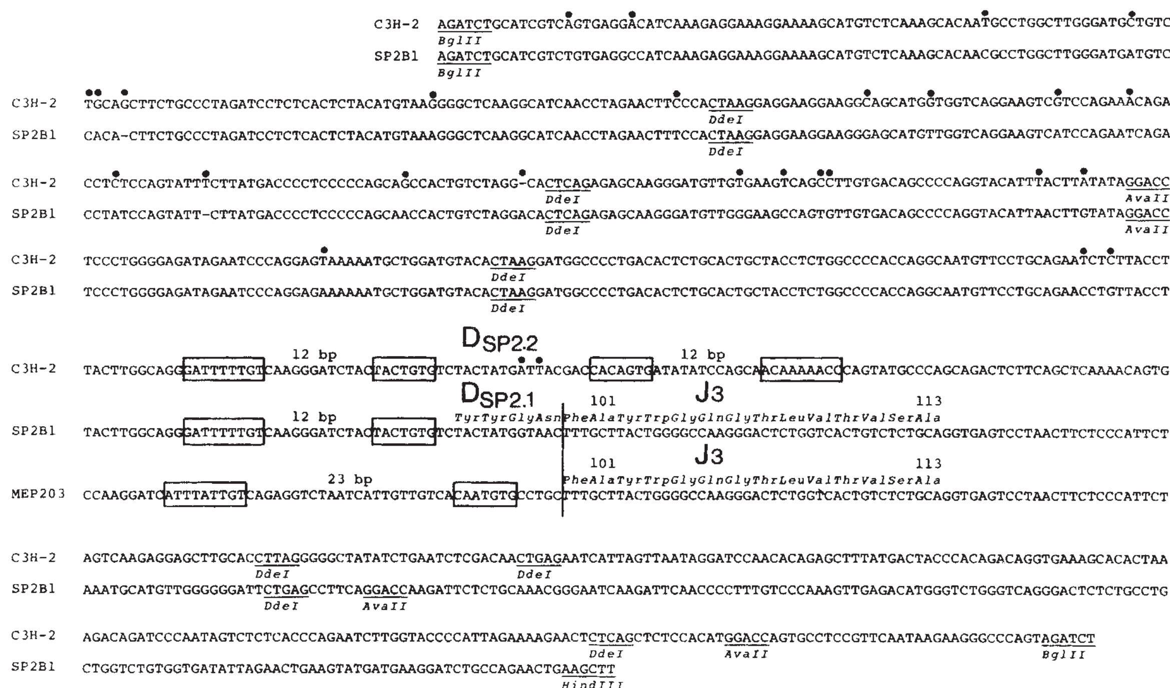


Fig. 4 Nucleotide sequences of three DNA clones: C3H-2, SP2B1 and MEP203 (refs 6, 7). All sequences were determined by the method of Maxam and Gilbert[33]. The sequence of the 0.85-kilobase *Bgl*II–*Bgl*II fragment of clone C3H-2 is aligned for the maximal homology with the sequence of the 0.8-kilobase *Bgl*II–*Hin*dIII fragment of clone SP2B1. Also aligned with the SP2B1 sequence is the sequence of the J_{H3}-containing region of clone MEP203, which was determined previously in our laboratory[7]. The SP2B1 fragment contains a D segment ($D_{SP2.1}$) fused with the J_{H3} segment. Amino acid residues encoded by the fused $D_{SP2.1}$ and J_{H3} on the SP2B1 fragment and those by the unfused J_{H3} on the MEP203, numbered according to Kabat *et al.*[11], are indicated in italics. The SP2B1 sequence is extensively homologous to the C3H-2 sequence from the 5′ *Bgl*II site to the $D_{SP2.1}$–J_{H2} boundary, where the homology switches to the MEP203 sequence. The dots and dashes indicate base substitutions and deletions within the homologous regions. The putative recognition sequences for recombinations are boxed. C3H-2 contains a D segment, $D_{SP2.2}$.

segment located 1.2 kilobases upstream. This D DNA segment has been independently identified during analysis of a rearranged fragment isolated from myeloma QUPC52 and is therefore referred to as D_{Q52} (see the accompanying article[19]).

As mentioned above, one base pair, C·G, is inserted on the B6 fragment at the site of recombination. Based on repeated sequencing of the relevant region, we believe that error is unlikely. This suggested the possible involvement of an intermediate recombination, the extra base pair arising from an independent D DNA segment. However, as described in the accompanying article, no additional D sequence has been identified between D_{Q52} and J_{H2}. Thus, if one is to pursue this possibility, it is necessary to invoke a scheme that allows joinings of multiple D DNA segments, the order of which in the assembled gene is independent of that in the germ-line genome. Although such a scheme is possible, one alternative is that the act of recombination introduced the extra base pair at the site of joining.

Structure and diversity of D sequences

In the digest of BALB/c mouse embryo DNA (Fig. 1*b*) the SP2D probe detected at least eight *Eco*RI fragments, each of which is likely to contain at least one copy of a $D_{SP2.1}$-like DNA segment. For the 5.0-kilobase fragment, this was demonstrated directly by DNA sequencing, which led to the identification of $D_{SP2.2}$. Our recent result indicates that the 5.0-kilobase band contains three *Eco*RI fragments each of which hybridizes with the SP2DJ probe (Y.K., unpublished). Thus, assuming that positive hybridization with the probe signifies the presence of at least one D segment, the minimum number of $D_{SP2.1}$-like segments is 10 (8 bands, of which one is triple).

If $D_{SP2.1}$-like DNA segments are abundant, as the Southern gel blot suggests, this is in agreement with the frequent appearance of the $D_{SP2.1}$-like sequences in the D regions of the assembled V genes. The D-coding regions of seven out of eight assembled V_H genes sequenced to date contain a palindromic 13-mer, TACTACGGTAGTA (ref. 8), or its variants (Table 1). The $D_{SP2.1}$ and $D_{SP2.2}$ also contain sequences similar to the 13-mer (Fig. 4, Table 1). In addition, the D regions of most mouse V_H regions for which amino acid sequences are available contain a tripeptide Tyr-Tyr-Gly that is encoded by the $D_{SP2.1}$ and $D_{SP2.2}$ segments[11]. Even the D_{Q52} segment which did not cross-hybridize with the $D_{SP2.1}$-like segments (Y.K., unpublished) shows significant homology, not only in the coding region (Table 1), but also in the flanking regions (Figs 4, 5). Thus, many, if not all D segments are related to the D_{SP2} segments, both evolutionarily and structurally.

Many D regions have extensions on one or both side(s) of the 13-mer (Table 1). The only striking characteristic of the sequences within these extensions is that the 3′ extension of the M173-II region shares a heptamer, CGGGGGT, with the H76 D region, the only D region apparently having no homology with the core 13-mer (Table 1). Each of these D regions may be encoded by its own D segment. Alternatively, some D regions, particularly those with relatively long extensions, may be encoded in two or more discrete D segments which are joined in the complete V genes. Two observations argue for the presence of D–D joinings. First, one and four base pairs at the presumed D_{Q52}–J_{H2} joining sites of the B6B1 and Q52 clones were in fact encoded neither by the D_{Q52} nor J_{H2} DNA segment (ref. 19 and Fig. 5). Second, we have been unable to detect a D_{M141} band in Southern gel blots of embryo DNA when this sequence of 44 base pairs (Table 1) was dissected out from the complete V_{M141} gene and used as the hybridization probe. In the same hybridization conditions the J_{H4} band was easily detected using the corresponding probe of nearly the same length (H.S., unpublished). If D–D joinings indeed occur, we predict that some D segments carry putative recognition sequences with 23-base pair spacers on at least one side.

©1981 Nature Publishing Group

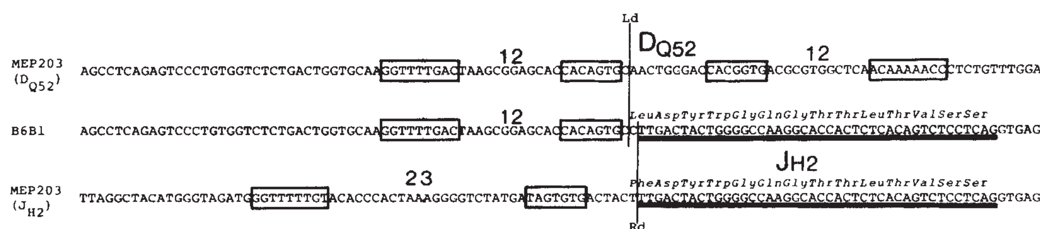
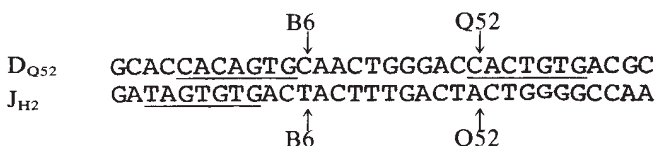


Fig. 5 Nucleotide sequences of the recombination regions of clone MEP203 and clone B6B1. The sequence across the 1.2-kilobase deletion of the B6B1 clone is compared with the two sequences of the embryonic fragment, MEP203, that cover the 5′ (top)[19] and the 3′ (bottom)[7] boundaries of the deleted segment. The two vertical lines, Ld and Rd, designate the left (5′ side) and the right (3′ side) boundaries of the deletion as determined by DNA sequencing, respectively. The other markings are as in Fig. 4.

Somatic diversification of heavy chains by V_H–D–J_H joinings

Amino acid sequence data[12] strongly suggest that the three types of germ-line V gene segments, V_H, D and J_H, combine in various, albeit not all possible, combinations to form complete V genes that are active in myelomas and presumably in B lymphocytes. The presence of the common 'recognition' sequences on the joining ends of all three gene segments supports this 'random' assortment. Furthermore, in both κ light- and heavy-chain genes it has been established that a given J DNA segment can assort with different V DNA segments[4,5,23].

Another somatic mechanism that might be used to amplify the coding capacity of immunoglobulin genes has been previously suggested by us[4] and by Max *et al.*[5], based on a proposed flexibility in the joining enzyme(s) with respect to the exact recombination sites on a given pair of joining gene segments. Although several previous observations in the κ-gene system support this hypothesis[4,5,7,24,25], none of the previous studies provides direct evidence for the hypothesis, because no two joinings examined took place between the same pair of gene segments. The present study, when combined with that described in the accompanying paper[19], presents such a case. Thus, the D–J joinings observed in myeloma QUPC52 and T-cell B6 both used the D_{Q52} and J_{H2}, and yet the exact recombination sites differ both on the D and J sides in the two joinings as shown below,



where the arrows indicate the recombination sites and the underlined sequences are the palindromic heptamers. Provided that these recombinants represent the intermediates of successful V–D–J joinings, we can conclude that modulation of exact recombination sites is another source of antibody diversity that is generated somatically.

T-cell antigen receptors are not encoded in assembled V_H, D and J_H segments

Our gel-blotting experiments using the J_H hybridization probe demonstrated that the DNA sequences are often, but perhaps more importantly not always, rearranged in the vicinity of the J_H cluster in various clones of cytolytic T cells and in thymomas. Two other groups reported similar findings and implicated them in the somatic formation of the complete structural genes coding for the antigen receptors of T cells[17,18]. The present results clearly rule out such a simple interpretation of the gel-blotting experiments. The recombinations in both SP2 and B6 cells involve D and J sequences, but no V_H sequence. In fact, neither recombinant can code for a V-like polypeptide chain in the region located 5′ to the recombination sites. We also emphasize that the recombination does not always occur among T-cell clones; we observed no indication of rearrangement around the J_H cluster in four well characterized clones of functional cytolytic T cells. We thus conclude that if the V_H DNA pool is shared by T and B cells to encode the antigen receptors, the four J_H segments identified to date are not used for rearrangement, at least in cytolytic T cells. It may be that T-cell receptors are encoded in an entirely different set of genes, or alternatively T and B cells share the same or an overlapping set of germ-line V_H genes, and another set of J-like and C-like DNA segments exists elsewhere in the genome, with which a V DNA segment joins specifically in T cells. Why, then, are there D–J_H joinings in T cells? One possibility is that they occurred in the common precursor cell which gave rise to B and T cells and have no active role in T-cell function. As discussed in the accompanying article[19], D–J_H joinings might be common on the unexpressed copy of chromosome 12 in myelomas, and one recent study[26] suggests that this might also be the case in natural B lymphocytes. Alternatively, the hypothetical T-cell recombinase may mediate these joinings due to lack of absolute specificity for the gene segments encoding the T-cell receptors.

We thank Lena Angman, Lisa Sultzman, Patricia Riegert and Gholamreza Dastoornikoo for technical assistance, Michael Julius and Bonnie Blomberg for discussion, Michael Potter for myelomas, Richard Maki for DNA clone ME184-8, Marcus Nabholz for T-cell clone DAS 1 and A. Coutinho for thymomas. The Basel Institute for Immunology was founded and is supported by F. Hoffmann-La Roche & Co., Ltd, Basel, Switzerland.

Received 7 January; accepted 11 March 1981.

1. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74,** 3171–3175 (1978).
2. Brack, C., Hirama, M., Schuller, R. & Tonegawa, S. *Cell* **15,** 1–14 (1978).
3. Bernard, O., Hozumi, N. & Tonegawa, S. *Cell* **15,** 1133–1144 (1978).
4. Sakano, H., Hüppi, K., Heinrich, G. & Tonegawa, S. *Nature* **280,** 288–294 (1979).
5. Max, E. E., Seidman, J. G. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **76,** 3450–3454 (1979).
6. Maki, R., Traunecker, A., Sakano, H., Roeder, W. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **77,** 2138–2132 (1980).
7. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. *Nature* **286,** 676–683 (1980).
8. Early, P. *et al. Cell* **19,** 981–992 (1980).
9. Bernard, O. & Gough, N. M. *Proc. natn. Acad. Sci. U.S.A.* **77,** 3630–3634 (1980).
10. Rabbitts, T. H., Matthyssens, G. & Hamlyn, P. H. *Nature* **284,** 238–243 (1980).
11. Kabat, E. A., Wu, T. T. & Bilofsky, H. in *Sequences of Imounoglobulin Chains* (NIH, Bethesda, 1979).
12. Schilling, J., Clevinger, B., Davie, J. M. & Hood, L. *Nature* **283,** 35–40 (1980).
13. Binz, H. & Wigzell, H. *Contemporary Topics Immun.* **7,** 113–177 (1977).
14. Rajewsky, K. & Eichmann, K. *Contemporary Topics Immun.* **7,** 69–112 (1977).
15. Tada, T. & Okumura, K. *Adv. Immun.* **28,** 1–87 (1979).
16. Julius, M. H., Cosenza, H. & Augustin, A. A. *Nature* **267,** 437–439 (1977).
17. Cory, S., Adams, J. M. & Kemp, D. *Proc. natn. Acad. Sci. U.S.A.* **77,** 4943–4947 (1980).
18. Forster, A., Hobart, M., Hengartner, H. & Rabbitts, T. H. *Nature* **286,** 897–899 (1980).
19. Sakano, H., Kurosawa, Y., Weigert, M. & Tonegawa, S. *Nature* **290,** 562–565 (1981).
20. Leder, P., Tiemeier, D. & Enquist, L. *Science* **196,** 175–177 (1977).
21. Edelman, G. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **63,** 78–85 (1969).
22. Poljak, R. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **70,** 3305–3310 (1973).
23. Weigert, M., Gatmaitan, L., Loh, E., Schilling, J. & Hood, L. *Nature* **276,** 785–790 (1978).
24. Seidman, J. G., Max, E. E. & Leder, P. *Nature* **280,** 370–375 (1979).
25. Max, E. E., Seidman, J. G., Miller, H. & Leder, P. *Cell* **21,** 793–799 (1980).
26. Nottenberg, C. & Weissman, I. *Proc. natn. Acad. Sci. U.S.A.* **77,** in the press (1980).
27. Honjo, T. *et al. Cold Spring Harbor Symp. quant. Biol.* (in the press).
28. Zakut, R., Cohen, J. & Givol, D. *Nucleic Acids Res.* **8,** 3591–3601 (1980).
29. Hozumi, N. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **73,** 3628–3632 (1976).
30. Southern, E. M. *J. molec. Biol.* **98,** 503–517 (1975).
31. Haas, W., Mathur-Rochat, J., Pohlit, H., Nabholz, M. & von Boehmer, H. *Eur. J. Immun.* **10,** 828–834 (1980).
32. von Boehmer, H. *et al. Eur. J. Immun.* **9,** 592–597 (1979).
33. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74,** 560–564 (1980).

©1981 Nature Publishing Group