

Physical linkage of a human immunoglobulin heavy chain variable region gene segment to diversity and joining region elements

(antibody genes/lymphocyte development)

HARRY W. SCHROEDER, JR.*¹, MICHAEL A. WALTER[†], MARTEN H. HOFKER[†], ALLEN EBENS*,
KO WILLEMS VAN DIJK[‡], LINDA C. LIAO[†], DIANE W. COX[†], ERIC C. B. MILNER[‡],
AND ROGER M. PERLMUTTER*[§]

*Howard Hughes Medical Institute and the Departments of Medicine (Medical Genetics) and Biochemistry, University of Washington, Seattle, WA 98195;
[†]Research Institute, Hospital for Sick Children, Toronto, ON, M5G 1X8 Canada; and [‡]Virginia Mason Research Center, Seattle, WA 98101

Communicated by Ray D. Owen, July 29, 1988 (received for review May 23, 1988)

ABSTRACT Antibody genes are assembled from a series of germ-line gene segments that are juxtaposed during the maturation of B lymphocytes. Although diversification of the adult antibody repertoire results in large part from the combinatorial joining of these gene segments, a restricted set of antibody heavy chain variable (V_H), diversity (D_H), and joining (J_H) region gene segments appears preferentially in the human fetal repertoire. We report here that one of these early-expressed V_H elements (termed V_{H6}) is the most 3' V_H gene segment, positioned 77 kilobases on the 5' side of the J_H locus and immediately adjacent to a set of previously described D_H sequences. In addition to providing a physical map linking human V_H , D_H , and J_H elements, these results support the view that the programmed development of the antibody V_H repertoire is determined in part by the chromosomal position of these gene segments.

Antibody genes are assembled from a series of discontinuous germ-line gene segments that are juxtaposed during B-lymphocyte development (1–3). For human antibody heavy chains, a series of several hundred variable (V_H) region gene segments (4, 5), at least five diversity (D_H) region gene segments (6, 7), and six functional joining (J_H) region gene segments (6) contribute to the generation of a diverse repertoire of antigen combining sites. V_H gene sequences can be grouped into families based on amino acid or nucleotide sequence similarity (8–10). Members of each family are 80% or more identical at the nucleotide level. Five human V_H gene families have been defined in this manner (4, 11, 12).

Although combinatorial joining of germ-line gene segments provides an important mechanism for diversification of the antibody repertoire, a limited set of germ-line V_H gene segments is preferentially expressed during murine fetal B-cell development (13–15). This observation may partially explain the fact that B-cell precursors reactive with specific antigens appear at characteristic times during the development of the mammalian immune system (16). The mechanism responsible for preferential expression of particular V_H gene segments during fetal life is unknown; however, it is provocative that the V_H family (V_{H7183}) that is most frequently represented in murine fetal B-cell transcripts is positioned at the centromeric end of the V_H locus near the constant region genes (14). This observation has led to the suggestion that chromosomal order may in part regulate the timing of rearrangement of V_H gene sequences (13, 14, 17).

In man, as in mouse, a subset of the total V_H repertoire is utilized in the products of fetal B-lineage cells (18, 19). Here we report that a unique V_H gene segment (V_{H6})[¶] that con-

tributes to the fetal heavy chain repertoire is the V_H gene segment most proximal to the gene encoding constant region of the μ chain (C_μ) in the human heavy chain locus and is positioned 77 kilobases (kb) on the 5' side of the J_H region. In addition, we have isolated genomic clones that link the V_H , D_H , and J_H gene segments.

MATERIALS AND METHODS

Cosmid Libraries and Library Screening. Construction of a human genomic DNA library in the cosmid vector pTL5 has been described (20). We screened 2×10^5 cosmids with the 15P1 V_{H6} cDNA probe (18) by standard methods (20, 21). Four overlapping cosmids (c3p1, c12.2, c17p1, and c17p3) were isolated and mapped preliminarily by using *EcoRI* and *Cla I* digestion (Boehringer Mannheim). Additional mapping was performed by using a partial digestion method (22) that involves probing with synthetic oligomers (5'-GCTGAAGC-CAGTTACCTTCGGAAAAAGAGT-3' and 5'-CATGCTGT-CCAGGCAGGTAGATGACGACCA-3') complementary to sequences flanking the *Cla I* site of the pTL5 vector. A second cosmid library was prepared with partially digested *Mbo I* fragments in the C2RB vector (23) and was screened by using the μ -switch probe (see below) by standard techniques (20, 21). Cosmid cos $\mu 6$ was linearized by digestion with *Sal I* and mapped by using a partial digestion method (22) with labeled *Sal I*-*BamHI* or *Sal I*-*Nru I* fragments of pBR322 as probes (24).

DNA Probes. The V_H -containing *EcoRI* fragments from clones 51P1, 56P1, 58P2, and 15P1 (18) were used as probes for V_{H1} , V_{H3} , V_{H4} , and V_{H6} gene families, respectively. Probes for V_{H2} (12) and V_{H5} (5) were generously supplied by T. Honjo (Kyoto University) and Fred Alt (Columbia University). A 6-kb *BamHI*-*HindIII* fragment spanning the J_H region (6) was provided by P. Leder (Harvard University). A 2.7-kb *BamHI*-*Pst I* subclone (pCW101) containing the 5' region of the constant region of the δ chain (C_δ) gene was provided by M. Belle White (University of Wisconsin). A μ -switch-region probe (pSM) containing a 2.2-kb *Sac I* fragment was provided by R. Wall (University of California-Los Angeles). The *Cla*9.5 probe was obtained by purifying the 9.5-kb *Cla I* fragment from cosmid c17p3 (above). Probes

Abbreviations: D_H , J_H , and V_H , heavy chain diversity, joining, and variable regions, respectively; FIGE, field inversion gel electrophoresis; PFGE, pulsed-field gel electrophoresis; C_μ and C_δ , μ chain and δ chain constant regions, respectively.

[¶]To whom reprint requests should be addressed at Howard Hughes Medical Institute SL-15, University of Washington, Seattle, WA 98195.

^{¶¶}The sequence reported in this paper is being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J04097).

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.

were labeled with ^{32}P to specific activities of $>10^9$ cpm/ μg by using primer-extension procedures (25). In addition, three oligonucleotide probes (5'-CGCACAGTAATACACGGC-CGTGTC, 5'-ACAGTAATACACGGCTGTGTC, and 5'-TGACACAGTAATACACAGCCGTGTC) corresponding to the conserved 3' portion of functional human V_H gene segments were labeled with polynucleotide kinase and used to screen for related sequences in the cosmid DNAs at low stringency ($1 \times \text{SSC}$ at 42°C ; $1 \times \text{SSC} = 0.15 \text{ M NaCl}/0.015 \text{ M sodium citrate, pH 7.0}$). The locations of three D_H gene segments (D_4 , D_1 , and D_2) were assigned by Southern blot analysis with an oligonucleotide probe (5'-GTGGGGCTCG-TGTCAGTGTG) encoding the terminal 2 nucleotides of the 5' nonamer recognition signal, the spacer, and the heptamer recognition signal from the sequence reported by Siebenlist *et al.* (7).

DNA Sequencing. Sequence was determined on both strands of pUC19 (26) subclones by the dideoxynucleotide chain-termination method with specific oligonucleotide primers (27).

Genomic Blotting. DNA samples were digested with restriction enzymes according to the manufacturers instructions and subjected to electrophoresis, blotting, and hybridization at 43°C as described (28). High-stringency washes were performed at 65°C in $0.1 \times \text{SSC}/0.1\%$ sodium dodecyl sulfate (SDS) for 2–4 hr. For pulsed-field and field-inversion gel electrophoresis (PFGE and FIGE, respectively), DNA was prepared in low-melting-point agarose with 10^6 cells per plug. Peripheral blood leukocyte DNA was digested with 15–20 units of *Bss*HII or *Eag* I per plug overnight as recommended (New England Biolabs). PFGE was performed at 250 V and 4°C for 24 hr. The apparatus was modified from that of Hardy *et al.* (29) by moving the electrodes 10 cm from the gel, which resulted in straighter lanes. FIGE was performed at 10 V/cm and 14°C in a submarine apparatus (Bethesda Research Laboratories) by using a Hoefer PC 750 controller.

RESULTS

The V_{H6} Gene Segment Is Unique. In a sample of 14 functional V_H elements present in transcripts derived from human fetal B-lineage cells, 13 sequences could be classified

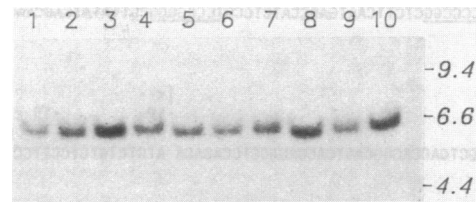


FIG. 1. V_{H6} gene family identified by genomic blot hybridization. DNA from 10 unrelated individuals (numbered arbitrarily) was digested with *Bgl* II and the blot was analyzed with the 15P1 V_H probe (prepared by *Eco*RI–*Ava* I digestion). The positions of size markers are shown at the right in kb.

by nucleotide sequence similarity as V_{H1} , V_{H3} , or V_{H4} -like, while one sequence, 15P1, was $<75\%$ related to all other known human V_H gene segments (19). Fig. 1 shows that a probe derived from this V_H region detects a single monomorphic restriction fragment in DNAs from various individuals. Since this pattern differs from that observed with other human V_H probes (4, 11, 12), clone 15P1 defines another V_H gene family, designated V_{H6} (4).

The structure of the human V_{H6} locus was determined by isolating overlapping cosmid clones with the 15P1 cDNA as a probe. The inserts derived from these clones together span 75 kb of germ-line DNA (Fig. 2). The sequence of the single V_{H6} element contained in these clones is presented in Fig. 3 and includes typical motifs important for transcriptional regulation (30, 31), a leader exon, an 86-base-pair intervening sequence, the V_H coding exon, and the 3' heptamer and nonamer sequences that participate in gene rearrangement (3). The V_{H6} coding region is identical to that determined for the cDNA clone 15P1. This result confirms that the restriction fragment shown in Fig. 1 does indeed contain the functional V_{H6} gene segment.

In man, V_H gene segments derived from various families are frequently interspersed in close proximity with one another (4, 11, 12, 32). Two approaches were therefore taken to find additional V_H elements that might be located adjacent to the V_{H6} gene segment. (i) Probes derived from all five human V_H families were used to screen for related sequences in the insert DNAs by low-stringency hybridization and washing (42°C and $2 \times \text{SSC}$). Only the V_{H6} gene segment was

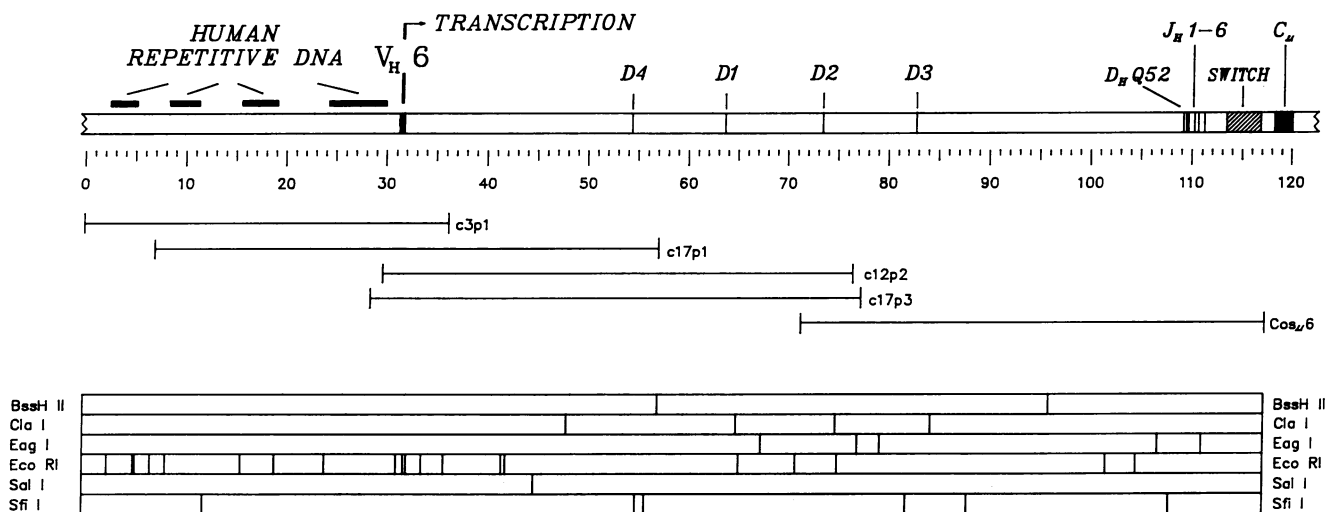


FIG. 2. Restriction map of the human heavy chain locus showing the physical linkage of V_{H6} , D_H , J_H , C_μ , and C_δ . The relative positions of four overlapping cosmid clones containing the V_{H6} gene segment with respect to $\text{cos}\mu 6$ containing the human J_H and μ -switch regions are shown above a scale in kb. The positions of sites for six restriction enzymes are shown with vertical lines at the bottom of the figure. Blocks of repetitive sequences, identified by probing with labeled human genomic DNA, are found in the region on the 5' side of the V_{H6} gene segment as shown. $\text{cos}\mu 6$ also contains repetitive sequences which have not yet been localized. The $\text{Cla} 9.5$ probe is the 3' most *Cla* I fragment in cosmid 17.3 (see text for description).

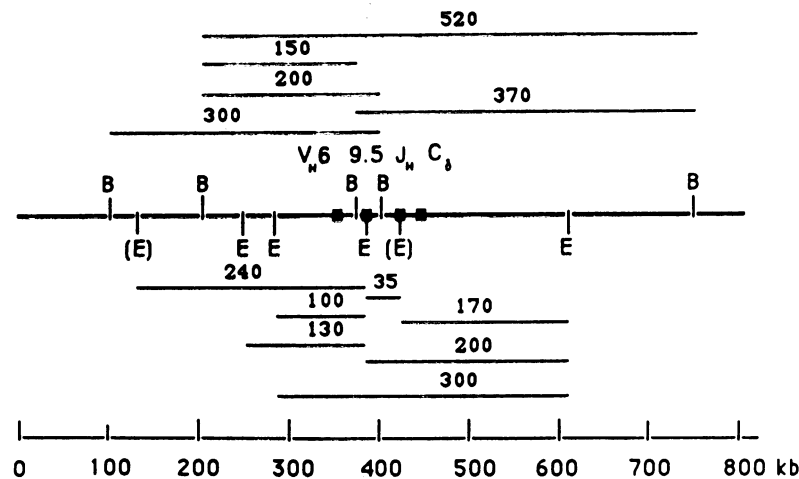


FIG. 5. Long-range map of the human V_H - D_H - J_H region. The data shown in Fig. 4 were used to assemble a genomic restriction map. Lines above the map indicate positions of hybridized *Bss*HII fragments, lines below indicate positions of hybridized *Eag* I fragments. B, *Bss*HII; E, *Eag* I. Due to variable methylation, not all *Bss*HII and *Eag* I sites are indicated. Sites that were frequently variably cut are indicated in parenthesis. Absolute sizes are approximate, except where the presence of restriction sites has been directly confirmed by cloning (Fig. 2).

the <80 kb of DNA separating the V_H6 gene segment from the J_H gene segment should contain multiple D_H sequences. In man, a single D_H segment (DHQ52) has been positioned 100 base pairs on the 5' side of the J_H1 segment (6). An additional set of four germ-line D_H sequences (D1-4) has been identified on a 33-kb segment of unlinked DNA, defined by overlapping genomic clones (7). To test for the presence of D_H gene segments adjacent to the V_H6 gene segment, we used an oligonucleotide complementary to the heptamer and 5' spacer of the human D1-4 elements. Three putative D_H segments were identified by hybridization with this oligomer (Fig. 2). The sequence of the 3'-most putative D_H from cosmid c17p3 was determined (data not shown) and found to be identical to that of D2 reported by Siebenlist *et al.* (7). The restriction map of cosmid clones 12p2 and c17p3 agrees with that reported by Siebenlist *et al.* (7) throughout a 20-kb overlapping region.

To verify that the V_H6 - D_H region is in fact linked to the J_H locus, a μ -switch-region probe was used to screen a second cosmid library. A single clone was isolated (cos μ 6) that spans the distance from the 3' end of the V_H6 -containing cosmid c17p3 to the 5' end of the C_H -switch region. A complete restriction map of the V_H6 , D_H , and J_H regions in man, deduced from our overlapping clones and spanning 120 kb of germ-line DNA, is presented in Fig. 2. Identical *Bam*HI and *Hind*III restriction sites were identified in the region of overlap of c17p3 and cos μ 6 (data not shown). The map is in agreement with results of long-range mapping studies (Fig. 5 and ref. 4) and is in good agreement with the published map of the human D_H and J_H regions (6, 7, 33). An additional *Eco*RI site immediately on the 3' side of D3 reported by Siebenlist *et al.* (7) is polymorphic in the human population (data not shown) and is not present in cos μ 6. Note that the human V_H6 , D_H , and J_H gene segments are all positioned in the same transcriptional orientation, the expected result given the predominance of deletional rearrangements in the heavy chain locus (3). Finally, since the V_H6 gene segment is positioned within 20 kb of the D4 gene segment and since no additional V_H gene sequences are found in this region, we can state with some confidence that the human V_H6 gene segment is the most 3' V_H sequence in the human genome and is located 77 kb on the 5' side of the J_H cluster.

DISCUSSION

The human V_H6 gene segment was first identified as a component of the quite restricted human fetal antibody

repertoire (18, 19). We have now demonstrated that this gene segment, similar to those that are expressed early in murine fetal immune development (13-15), is positioned at the 3' end of the V_H locus, adjacent to the C_H gene. Berman *et al.* (4) have provided a lower-resolution PFGE map of the V_H locus, which is in agreement with the cloning studies reported here. While these results are consistent with the view that chromosomal position directs the early development of the V_H repertoire, the V_H6 gene segment was not the most frequently encountered sequence in human fetal samples. Instead, members of the much larger human V_H3 gene family were repeatedly identified in developing human B lymphocytes (18). In addition, the V_H5 gene family has been positioned 160 kb on the 5' side of the J_H locus and seemingly on the 3' side of any V_H3 elements (12). Despite this relatively J_H -proximal location, no V_H5 sequences were found in the early human repertoire (18, 19). Thus factors other than chromosomal position almost certainly contribute to the biased utilization of particular V_H gene segments in developing human B lymphocytes.

Although physical linkage of V_H , D_H , and J_H gene segments has not been achieved in any other mammalian species, three D_H gene families (DFL, DSP, and DQ52) containing a total of 12 elements have been identified in an 80-kb region on the 5' side of the murine J_H locus (33). In addition, Matsuda *et al.* (34) have characterized human genomic D_H gene segments and have linked the D1-4 and J_H regions on a series of cosmid clones analogous to those reported here. Striking parallels are seen in the content and organization of D_H gene segments in man as compared with mouse. The human D_H region identified here and independently by Matsuda *et al.* (34) spans \approx 70 kb and is bounded at the 3' end by the DHQ52 gene segment, which is closely related to a similarly positioned and identically named D gene segment in the mouse. The murine DFL family and members of the human DFL equivalent (D1-D4) are at the 5' end of this part of the D locus in both species (Fig. 2 and ref. 33). Some members of the DSP family in the mouse are interspersed with DFL members, but the majority are clustered in the middle of the locus. No human DSP-like elements have been reported, although analysis of functional heavy chain transcripts supports the existence of these sequences in the genome (18). Studies by ourselves and others indicate that there are many as yet undefined germ-line D_H gene segments in man, some of which almost certainly reside in the region between V_H6 and J_H but were not detected with the D_H oligomer as a result of sequence variability in the region

surrounding the heptamer recognition sequence. For example, the functionally rearranged V_H6 gene segment in clone 15P1, which by virtue of its 3' location uses a D_H sequence that must be presumed to lie within the 77-kb D_H region defined here, nevertheless includes D_H sequences that do not correspond to any existing germ-line element. Other D_H sequences are apparently distributed in more 5' positions within the V_H locus (34).

We have reported (18, 19) that the earliest-appearing V_H gene segments in man and mouse are structurally quite similar and have now demonstrated that at least some of these gene segments occupy similar C_μ-proximal positions within the V_H loci of both species. These results indicate that the earliest events in antibody diversification have been conserved in man and mouse to a remarkable degree and further suggest that selection of appropriate V_H elements during early repertoire assembly may be crucially important for satisfactory development of humoral immunity (17). From this perspective, it will be interesting to identify human antibodies of defined specificities that employ the V_H6 gene segment.

We thank F. Alt, T. Honjo, P. Leder, R. Wall, and M. Belle White from the laboratory of F. Blattner for probes, and our colleagues for helpful discussions. This work was supported by the Howard Hughes Medical Institute and by grants from the Natural Sciences and Engineering Research Council of Canada, the American Cancer Society, and the March of Dimes foundation. M.A.W. was supported by a studentship and M.H.H. was supported by a fellowship from the Medical Research Council of Canada.

1. Honjo, T. (1983) *Annu. Rev. Immunol.* **1**, 499–528.
2. Tonegawa, S. (1983) *Nature (London)* **302**, 575–581.
3. Yancopoulos, G. D. & Alt, F. W. (1986) *Annu. Rev. Immunol.* **4**, 339–368.
4. Berman, J. E., Mellis, S. J., Pollock, R., Smith, C. L., Suh, H., Heinke, B., Kowal, C., Surti, U., Chess, L., Cantor, C. R. & Alt, F. W. (1988) *EMBO J.* **7**, 727–738.
5. Walter, M. A. & Cox, D. W. (1988) *Am. J. Hum. Genet.* **42**, 446–451.
6. Ravetch, J. V., Siebenlist, U., Korsmeyer, S., Waldmann, T. & Leder, P. (1981) *Cell* **27**, 583–591.
7. Siebenlist, U., Ravetch, S., Korsmeyer, S., Waldmann, T. & Leder, P. (1981) *Nature (London)* **294**, 631–635.
8. Brodeur, P. H. & Riblet, R. (1984) *Eur. J. Immunol.* **14**, 922–930.
9. Dildrop, R. (1984) *Immunol. Today* **5**, 85–86.
10. Kabat, E. A. Wu, T. T., Reid-Miller, M., Perry, H. M. & Gottesman, K. S. (1987) *Sequences of Proteins of Immunological Interest* (Dept. of Health and Human Services, Washington, DC), 4th Ed..
11. Lee, H. L., Matsuda, F., Kinashi, T., Kodaira, M. & Honjo, T. (1987) *J. Mol. Biol.* **195**, 761–768.
12. Shen, A., Humphries, C., Tucker, P. & Blattner, F. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8563–8567.
13. Perlmutter, R. M., Kearney, J. F., Chang, S. P. & Hood, L. E. (1985) *Science* **227**, 1597–1601.
14. Yancopoulos, G. D., Desiderio, S. V., Paskin, M., Kearney, J. F., Baltimore, D. & Alt, F. W. (1984) *Nature (London)* **311**, 727–733.
15. Lawler, A. M., Lin, P. S. & Gearhart, P. J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 2454–2457.
16. Klinman, N. R. & Press, J. L. (1985) *Transplant. Rev.* **24**, 41–83.
17. Perlmutter, R. M. (1987) *Curr. Topics Microbiol. Immunol.* **135**, 95–110.
18. Schroeder, H. W., Jr., Hillson, J. L. & Perlmutter, R. M. (1987) *Science* **238**, 791–793.
19. Perlmutter, R. M., Schroeder, H. W., Jr., & Hillson, J. L. (1988) in *B Cell Development*, UCLA Symposia on Molecular and Cellular Biology, New Series, eds., Witte, O., Howard, M. & Klinman, N. (Liss, New York), Vol. 85, in press.
20. Garvin, A. M., Pawar, S., Marth, J. D. & Perlmutter, R. M. (1988) *Mol. Cell. Biol.* **8**, 3058–3064.
21. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
22. Rackwitz, H. R., Zehetner, G., Murialdo, H., Chai, J. H., Poustka, A., Frischauf, A. & Lehrach, H. (1985) *Gene* **40**, 259–266.
23. Bates, P. F. & Swift, R. A. (1983) *Gene* **26**, 137–146.
24. Hofker, M. H., van Ommen, G. J. B., Bakker, E., Burmeister, M. & Pearson, P. L. (1986) *Hum. Genet.* **74**, 270–274.
25. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
26. Messing, J. (1983) *Methods Enzymol.* **101**, 20–78.
27. Strauss, E., Kobori, J., Siu, G. & Hood, L. (1986) *Anal. Biochem.* **154**, 353–360.
28. Perlmutter, R. M., Klotz, J. L., Bond, M. W., Nahm, M., Davie, J. M. & Hood, L. (1984) *J. Exp. Med.* **159**, 179–192.
29. Hardy, D. A., Bell, J. I., Long, E. O., Lindsten, T. & McDevitt, H. O. (1986) *Nature (London)* **323**, 453–455.
30. Parslow, T. G., Blair, D. L., Murphy, W. J. & Granner, D. K. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 2650–2654.
31. Mason, J. O., Williams, G. T. & Neuberger, M. S. (1985) *Cell* **41**, 479–487.
32. Kodaira, M., Kinashi, T., Umemura, I., Matsuda, F., Nome, T., Ono, Y. & Honjo, T. (1986) *J. Mol. Biol.* **190**, 529–541.
33. Wood, C. & Tonegawa, S. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3030–3034.
34. Matsuda, F., Lee, K. H., Nakai, S., Sato, T., Kodaira, M., Zong, S. Q., Ohno, H., Fukuhara, S. & Honjo, T. (1988) *EMBO J.* **7**, 1047–1051.