Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 5360–5363, August 1987
Genetics

B-cell variant of mouse T200 (Ly-5): Evidence for alternative mRNA splicing

(leukocyte surface glycoprotein/cDNA sequence/genomic structure)

MATTHEW L. THOMAS*†, PAMELA J. REYNOLDS*, AILEEN CHAIN*, YINON BEN-NERIAH‡§, AND IAN S. TROWBRIDGE*

*Department of Cancer Biology, The Salk Institute for Biological Studies, P. O. Box 85800, San Diego, CA 92138; and ‡Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142

Communicated by Lewis Thomas, April 20, 1987

ABSTRACT **The complete cDNA sequence of mouse T200 glycoprotein from the pre-B-cell line 70Z/3 has been determined. The deduced protein sequence differs from that previously reported for a T-cell form of the molecule [Saga, Y., Tung, J.-S., Shen, F.-W. & Boyse, E. A. (1986) *Proc. Natl. Acad. Sci. USA* 83, 6940–6944] by the insertion of 139 amino acid residues in the amino-terminal region of the molecule. RNA transfer blotting using a cDNA probe encoding this sequence established that the predominant T200 mRNA species from B cells and cytotoxic T-cell clones but not T-helper cell clones or thymocytes contain all or part of the insert. Overlapping genomic clones of murine T200 were isolated and analysis of the intron–exon structure at the 5′ end of the gene provides evidence that variants of T200 glycoprotein are generated by alternative mRNA splicing.**

Mouse T200 glycoprotein is a major high molecular weight leukocyte cell-surface molecule bearing the Ly-5 alloantigenic determinant (1–4). It is expressed on all hematopoietic cells except mature erythrocytes and their immediate progenitors (5). Biochemical analysis has established that the mature glycoprotein exhibits cell-type-specific variation in structure (2, 6, 7). B cells express a M_r 220,000 variant whereas thymocytes express a M_r 180,000 species. Other leukocytes express multiple forms of the glycoprotein within that size range. Antigenic variants of the glycoprotein have been defined by monoclonal antibodies (8–12). Recently, T200 cDNAs have been isolated from a rat thymocyte library and a mouse T-cell library (13, 14). The complete 1152-amino acid sequence of a mouse T-cell form of the molecule has been deduced (15). RNA transfer blotting has shown that different leukocytes express different sizes of T200 mRNA (16). Furthermore, S1 nuclease protection mapping has indicated that an additional stretch of nucleotides is present in the 5′ region of T200 mRNA from a macrophage and a B-cell line that is not found in thymocyte T200 mRNA (15). Here we report the isolation of two T200 cDNAs from the mouse chemically induced leukemia pre-B-cell line 70Z/3 (17) that span the entire coding region of the molecule. Comparison of the deduced amino acid sequences of T200 from the pre-B-cell line with that of the T-cell line (15) indicates that the B-cell form has an insert of 139 amino acids close to the amino terminus of the molecule. Analysis of the genomic structure of T200 suggests that these structural variants of the glycoprotein are generated by alternative splicing of T200 mRNA.

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.

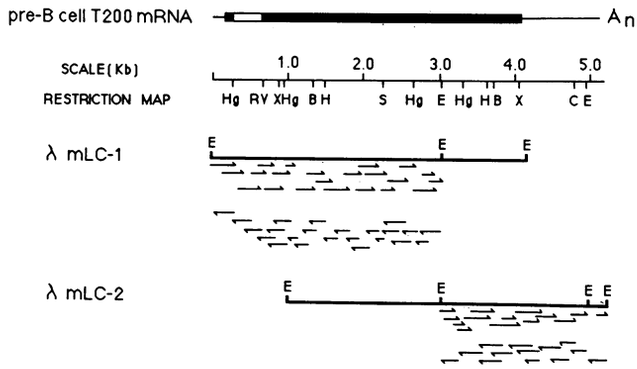


FIG. 1. Analysis of mouse T200 cDNA clones derived from the pre-B-cell line 70Z/3. Two T200 cDNA clones isolated from the λgt11 library derived from 70Z/3 cells are shown together with a partial restriction map, and the sequencing strategy used to obtain the complete sequence of the coding region. Above the restriction map is a diagrammatic representation of the structure of the pre-B-cell T200 mRNA predicted by the DNA sequence of the two clones. The 5′- and 3′-untranslated regions are represented by thin lines and the coding region is represented by the thick line. The region encoding the 139-amino acid insert not present in the basic thymocyte form of T200 is indicated by the open area. B, *Bam*HI; C, *Cla* I; E, *Eco*RI; H, *Hin*dIII; Hg, *Hgi*A1; RV, *Eco*RV; S, *Sma* I; X, *Xba* I.

MATERIALS AND METHODS

Isolation of T200 Genomic λ Clones and cDNAs. Initially, a 4.4-kilobase (kb) rat L-CA cDNA probe derived from overlapping rat thymocyte cDNAs pLC-1 and pLC-2 by ligation at the unique *Xba* I site (13) was used to screen 2×10^6 plaques from a recombinant bacteriophage library from the murine plasmacytoma MOPC 41 (18). Two overlapping T200 λ genomic clones were isolated. To obtain further overlapping genomic clones, the library was rescreened, first with probes from the 3′ or 5′ ends of the two original T200 λ genomic clones and then with probes derived from the 70Z/3 T200 cDNAs described below. The clones were characterized by the Cos mapping method of Rackwitz *et al.* (19) in combination with standard restriction mapping methods. T200 cDNA clones were obtained by screening 1×10^6 plaques derived from a λgt11 library from the mouse chemically induced leukemia pre-B-cell line 70Z-3 (20) with a mixture of two 750-base-pair (bp) *Hin*dIII fragments from one of the original mouse T200 λ genomic clones; one fragment contained the exon encoding the membrane-spanning region,

†Present address: Department of Pathology, Washington University, St. Louis, MO 63110.
§Present address: Hebrew University, Hadessa Medical School, Jerusalem, Israel 91010.

the other contained a second exon encoding a segment of the extracellular domain of T200 glycoprotein (corresponding to amino acid residues 396–448 of the T200 sequence shown in Fig. 2).

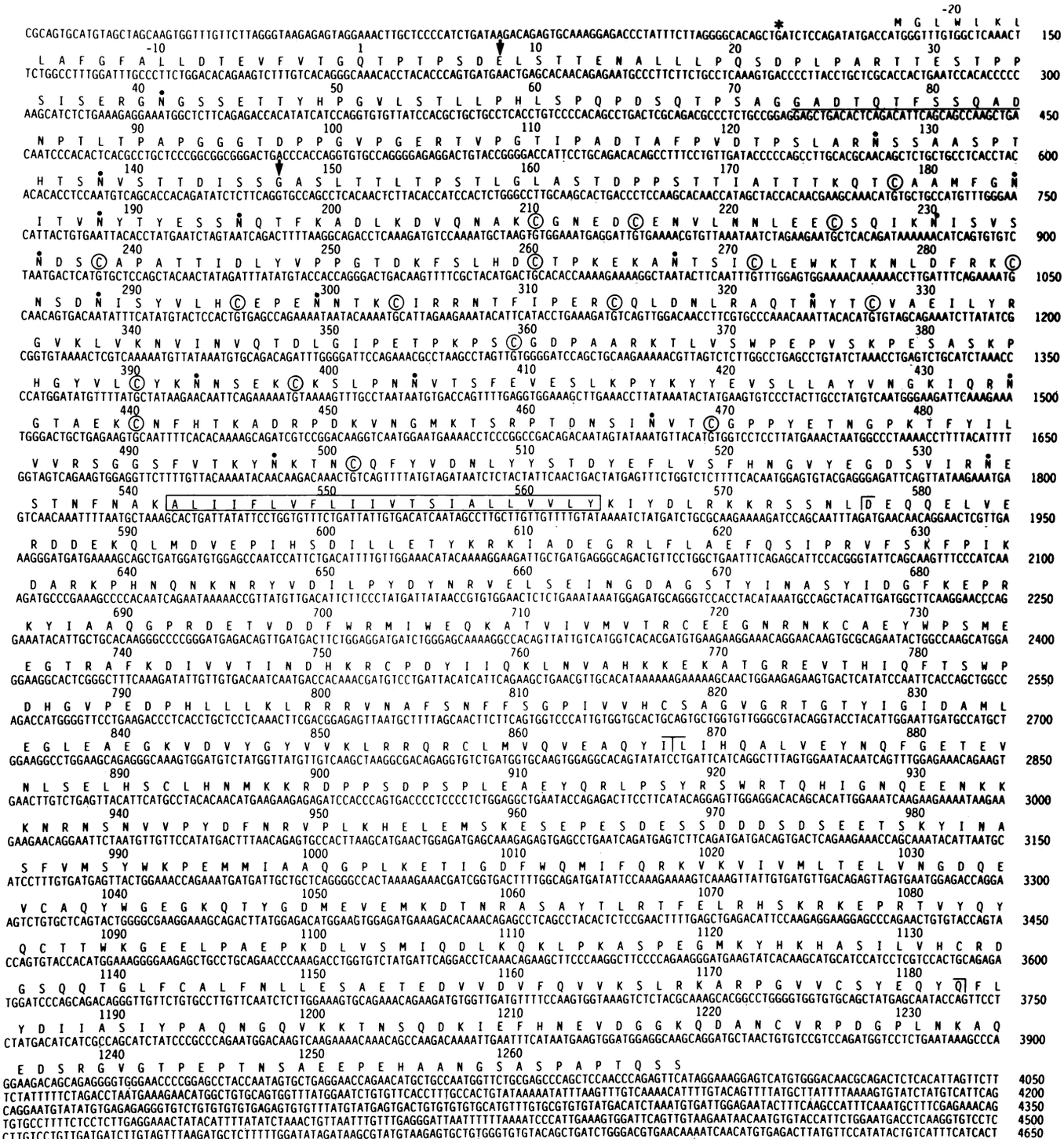


FIG. 2. Complete nucleotide and predicted primary sequence of the 70Z/3 T200 variant. The sequences obtained from the two clones λ mLC-1 and λ mLC-2 were combined to obtain the complete nucleotide [4991 bp excluding the poly(A) tail] and amino acid sequence of the 70Z/3 T200 glycoprotein. The predicted protein sequence is initiated at a methionine corresponding to nucleotide position 131 and terminates at position 4003. The putative transmembrane spanning region is boxed and cysteine residues in the extracellular domain are circled. Potential sites of N-linked glycosylation in the extracellular domain are indicated by dots over the asparagine residues. The first termination codon upstream from the coding region is marked with an asterisk. The first amino acid of the insert sequence (glutamic acid) and the first residue after the insert (glycine) are marked by arrows. The sequence homologous to the rat L-CA tryptic peptide recognized by monoclonal antibody OX-22 is underlined (21). The two homologous subdomains of the cytoplasmic region spanning residues 577–867 and 868–1182 are also indicated. Two poly(A) signal sequences are also underlined. Amino acids are identified by the single-letter code.

DNA Sequencing. Genomic and cDNA fragments were subcloned into M13mp18 and overlapping deletions of each fragment were obtained by the method of Dale *et al.* (22). DNA sequencing was by the dideoxy chain-termination procedure of Sanger *et al.* (23).

RNA Transfer Blotting Analysis. A blot of poly(A)$^+$ RNA from mouse lymphoid cells immobilized on Zetaprobe (Bio-Rad) shown in figure 1 of ref. 15 was rehybridized with a 400-bp *Hgi*A1/*Eco*RV fragment from the 5′ end of λmLC-1 (see Fig. 1) according to the method of Gatti *et al.* (24).

RESULTS AND DISCUSSION

Nucleotide and Predicted Amino Acid Sequence of T200 cDNA from the Pre-B-Cell Line 70Z/3. Two T200 cDNA clones were isolated from the 70Z/3 pre-B-cell cDNA library, which overlapped to span 5.1 kb (Fig. 1). The 5′ end of λMLC-1 and the 3′ end of λMLC-2 were sequenced in both directions to yield the complete coding region of T200 glycoprotein derived from the 70Z/3 cell line (Figs. 1 and 2). The nucleotide sequence predicts a protein of 1291 amino acids. Based on hydrophobicity analysis and the known orientation of rat L-CA (T200) in the cell membrane (13), it is inferred that the 70Z/3 T200 molecule consists of a leader sequence of 23 amino acids, an external domain of 541 amino acids, a transmembrane region of 22 amino acids, and a cytoplasmic domain of 705 amino acids. Comparison of the 70Z/3 T200 cDNA sequence with the sequence of T200 from a T-cell cDNA library reported by Saga *et al.* (15) showed these to be identical over most of their length except for 11 isolated nucleotide differences that now need to be verified or revised. However, a major difference in the two sequences was an insertion of 417 nucleotides at the 5′ end of the 70Z/3 T200 cDNA, which was not present in the T-cell cDNA (Figs. 1 and 2). It has been shown that the 70Z/3 cell line expresses a high molecular weight (M_r, 220,000) form of T200 glycoprotein characteristic of B cells on its cell surface (6) and that the T200 mRNA from this cell line is significantly larger than that of thymocytes and helper T-cell lines (16, 25). The extra nucleotide sequence in the T200 cDNA from the 70Z/3 cell encodes an additional 139 amino acids (residues 8–146) and thus provides an explanation for the structural differences between the B-cell and thymocyte forms of the molecule. The amino acid sequence of the insert is rich in serine and threonine residues, which are potential sites for O-linked oligosaccharides (Fig. 2). In addition, there are two potential acceptor sites for N-linked oligosaccharides in the insert. It is likely, therefore, that as well as the increase in size due to the extra polypeptide segment (M_r, 14,000), the B-cell form of the glycoprotein contains multiple oligosaccharides not present in the thymocyte form that contribute to the molecular weight difference (M_r, 40,000) between the two species.

T200 mRNAs from Cytotoxic T-Cell Lines Contain Sequences from a B-Cell Insert. To establish that the extra nucleotide sequence identified in the 70Z/3 cDNA was present in the predominant T200 mRNA species of the pre-B-cell line and to determine whether other leukocytes expressed related T200 mRNA species, a 400-bp *Hgi*A1/*Eco*RV fragment encoding most of the insert (Fig. 1) was isolated and used for RNA transfer blot analysis. In previous studies (16), T200 mRNA from various mouse lymphoid tissues and a series of cytotoxic and helper T-cell lines had been shown to differ in size. As shown in Fig. 3, when the same mRNA blot was hybridized with the *Hgi*A1/*Eco*RV fragment of the pre-B-cell T200 cDNA, the probe selectively hybridized to the T200 mRNA species from 70Z/3 cells and the four cytotoxic T-cell lines. These mRNAs are larger than T200 mRNAs from thymocytes and T-helper cell lines that did not hybridize to the probe. T200 mRNA from lymph node cells

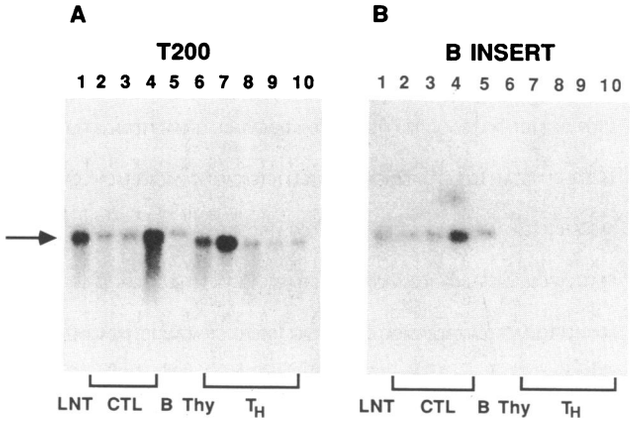
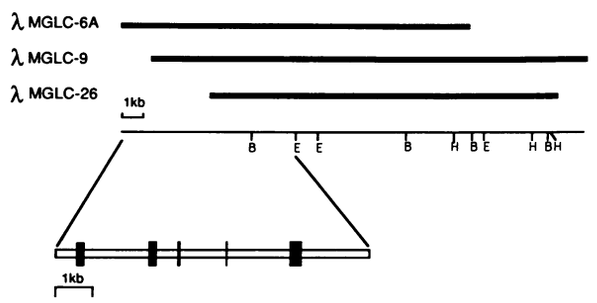


FIG. 3. RNA transfer blot analysis of T200 mRNA from various mouse leukocytes. RNA was fractionated on a 0.5% agarose formaldehyde gel and blotted onto Zetaprobe (Bio-Rad). (*A*) The filter was hybridized with a 3.7-kb *Eco*RI probe from the rat thymocyte cDNA clone pLC-2. Lanes 1: RNA was from lymph node T cells (LNT) isolated by depleting B cells with the monoclonal antibody J11D plus complement (16); 2–4, RNA from the cytotoxic T-lymphocyte (CTL) clones B3, 8.2, and H7, respectively; 5, RNA from the pre-B-cell line 70Z/3 (B); 6, from thymocyte RNA (Thy); 7–10, T-helper (T_H) RNA from K3, BB5, K31C7, and K321C9, respectively. (*B*) After decay of the bound radioactivity, the blot was rehybridized with the 400-bp *Hgi*A1/*Eco*RV fragment from the 5′ end of λ mLC-1, which includes most of the insert sequence not found in the basic thymocyte form of T200. Arrow indicates position of 28S RNA.

hybridized weakly to the B-cell insert, suggesting that only a fraction of these mRNA species contain sequences related to the B-cell T200 insert. These results confirm that the T200 mRNA species in cytotoxic and helper T cells differ and establish that T200 mRNA from cytotoxic cells contains sequences related to the B-cell insert.

Evidence That the Variant Forms of Murine T200 Arise by Alternative mRNA Splicing. To clarify the genetic basis for the murine T200 variants, the genomic structure of T200 was investigated. Overlapping λ genomic T200 clones spanning a total of 75 kb were isolated and a partial restriction map was constructed. This analysis suggests there is a single T200 gene. The restriction map at the 5′ end of the gene is shown in Fig. 4*A*. Restriction fragments were isolated from genomic clones that hybridized to probes from the 5′ end of the 70Z/3 λMLC-1 cDNA. These fragments were then sequenced to determine the exon–intron structure at the 5′ end of the T200 gene (Fig. 4*B*). The positions of five exons encoding sequences found at the 5′ end of the 70Z/3 T200 cDNA have been identified. Two of these encode amino acid residues 51–99 and 100–146 of the predicted 70Z/3 T200 protein sequence, which represents the carboxyl-terminal region of the B-cell insert. Thus, the nucleotide insert found in the 70Z/3 T200 cDNA is encoded by at least three separate exons, which are spliced out in the basic thymocyte form of the molecule. As described by Saga *et al.* (26), cDNAs have been isolated from a library derived from the B-leukemic cell line I.29 that predict two other variants of mouse T200 glycoprotein. These forms of the molecule contain smaller inserts encoded by either only one or both of the 3′ exons of the 70Z/3 B-cell insert. Similar variants of T200 glycoprotein in rat (21) and human (27) have also been identified. The largest form of T200 in human cells is homologous to that of the 70Z/3 T200 molecule except that the human form contains an additional 69 nucleotides inserted at base-pair position 351 (Fig. 1). It is not known whether a larger form of T200 glycoprotein than that predicted by the nucleotide sequence of the cDNAs isolated from the 70Z/3 library is also expressed in the mouse.

A.



B.

```
                52                    99
              Gly Val Leu           Pro  Pro
ttctaacctcag  GT GTG TTA ......  CCA CCA G  gtttgggggctc

               101                   146
              Gly Val Pro           Ser  Ser
ttctgattgcag  GT GTG CCA ......  TCT TCA G  gtgtgaccatta

               148                   170
              Gly Ala Ser           Thr  Ile
atttccttgcag  GT GCC AGC ......  ACC ATA G  gtgacaatctta

               172                   178
              Ala Thr Thr           Thr  Cys
tctcaattacag  CT ACC ACA ......  ACA TGT G  gtaagtttgtcg

               180                   250
              Ala Ala Met           Pro  Pro
          ccag CT GCC ATG ......  CCA CCA G  gtgaatgtcaat
```

FIG. 4. (*A*) Partial restriction map of λ genomic clones derived from a MOPC 41 library. Enlarged area indicates the intron–exon structure for the 5′ *Eco*RI fragment. Exons are indicated by solid rectangles. (*B*) Nucleotide sequences of the intron–exon junctions. Intron sequences are in lowercase letters.

The results of this analysis, together with complementary data reported by Saga *et al.* (26), provide a genetic basis for the structural and antigenic heterogeneity of T200 glycoprotein in the mouse. At least four variants of murine T200 glycoprotein can be generated by alternative splicing of exons at the 5′ end of the gene and additional variants may exist. It should now be possible to define more precisely when during ontogeny and upon which leukocyte subpopulations the various forms of T200 glycoprotein are expressed. It seems likely, as these cell-type-specific variants have been conserved between species, that they are of functional significance. One possibility is that the presence of the various inserts that are likely to display multiple oligosaccharides modulates the interaction of T200 glycoprotein with other cell-surface molecules and thus influences leukocyte interactions with other cells. It is of interest that the monoclonal antibody OX-22, which recognizes an antigenic determinant restricted to a high molecular weight form of rat L-CA, distinguishes two T-helper cell populations in the rat (28, 29). A tryptic peptide of rat L-CA with which OX-22 antibody reacts has been identified (21) and the homologous sequence in murine T200 lies in the extra sequence found in the 70Z/3 form of the molecule (see Fig. 2). Evidence has also been obtained for two distinct subpopulations of T-helper cells in the mouse (30), and it is possible that these murine T-helper subsets will also express different forms of T200 glycoprotein.

We would like to thank our colleagues for many useful suggestions, particularly Jeff Arriza and Glen Evans. We would also like to thank Ami Koide for typing the manuscript. M.L.T. was a Special Fellow of the Leukemia Society of America. The work was supported by Grant CA 17733 from the National Cancer Institute.

1. Trowbridge, I. S., Ralph, P. & Bevan, M. J. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 157–161.
2. Trowbridge, I. S. (1978) *J. Exp. Med.* **148,** 313–323.
3. Komuro, K., Itakura, K., Boyse, E. A. & John, M. (1975) *Immunogenetics* **1,** 452–456.
4. Omary, M. B., Trowbridge, I. S. & Scheid, M. P. (1980) *J. Exp. Med.* **151,** 1311–1316.
5. Scheid, M. P. & Triglia, D. (1979) *Immunogenetics* **9,** 423–433.
6. Tung, J.-S., Scheid, M. P., Pierotti, M. A., Hammerling, U. & Boyse, E. A. (1981) *Immunogenetics* **14,** 101–106.
7. Watson, A., Dunlap, B. & Bach, F. H. (1981) *J. Immunol.* **127,** 38–42.
8. Coffman, R. L. & Weissman, I. L. (1981) *Nature (London)* **289,** 681–683.
9. Kincade, P. W., Lee, G., Watanabe, T., Sun, L. & Scheid, M. P. (1981) *J. Immunol.* **127,** 2262–2268.
10. Kung, J. T., Sharrow, S. O., Mage, M. G. & Paul, W. E. (1982) *J. Immunol.* **129,** 81–86.
11. Coffman, R. L. (1982) *Immunol. Rev.* **69,** 5–23.
12. Lefrancoise, L., Puddington, L., Machamer, C. E. & Bevan, M. J. (1985) *J. Exp. Med.* **162,** 1275–1293.
13. Thomas, M. L., Barclay, A. N., Gagnon, J. & Williams, A. F. (1985) *Cell* **41,** 83–93.
14. Shen, F.-W., Saga, Y., Litman, G., Freeman, G., Tung, J.-S., Cantor, H. & Boyse, E. A. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 7360–7363.
15. Saga, Y., Tung, J.-S., Shen, F.-W. & Boyse, E. A. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 6940–6944.
16. Lefrancoise, L., Thomas, M. L., Bevan, M. J. & Trowbridge, I. S. (1986) *J. Exp. Med.* **163,** 1337–1342.
17. Paige, C. J., Kincade, P. W. & Ralph, R. (1978) *J. Immunol.* **121,** 641–648.
18. Evans, G. A., Margulies, D. H., Camerini-Otero, R. D., Ozato, K. & Seidman, J. G. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 1994–1998.
19. Rackwitz, H.-R., Zehetner, G., Frischauf, A.-M. & Lehrach, H. (1984) *Gene* **30,** 195–200.
20. Ben-Neriah, Y., Bernard, S. A., Paskind, M., Daley, G. Q. & Baltimore, D. (1986) *Cell* **44,** 577–586.
21. Barclay, A. N., Jackson, D. I., Willis, A. C. & Williams, A. F. (1987) *EMBO J.* **6,** 1259–1264.
22. Dale, R. M. K., McClure, B. A. & Houchins, J. P. (1985) *Plasmid* **13,** 31–40.
23. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
24. Gatti, R. A., Concannon, P. & Salser, W. (1984) *Biotechniques* **2,** 148–155.
25. Raschke, W. C. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 161–165.
26. Saga, Y., Tung, J.-S., Shen, F.-W. & Boyse, E. A. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 5364–5368.
27. Ralph, S. J., Thomas, M. L., Morton, C. C. & Trowbridge, I. S. (1987) *EMBO J.* **6,** 1251–1257.
28. Spickett, G. P., Brandon, M. R., Mason, D. W., Williams, A. F. & Woollett, G. R. (1983) *J. Exp. Med.* **158,** 795–810.
29. Arthur, R. P. & Mason, D. (1986) *J. Exp. Med.* **163,** 774–786.
30. Mosmann, T. R., Cherwinski, H., Bond, M. W., Giedlin, M. A. & Coffman, R. L. (1986) *J. Immunol.* **136,** 2348–2357.