LAH # 298

Isolation of complementary DNA clones encoding the human lymphocyte glycoprotein T1/Leu-1

Nancy H. Jones*, Martha L. Clabby*, Deno P. Dialynas†, Huei-Jen S. Huang‡, Leonard A. Herzenberg‡ & Jack L. Strominger*†

* Division of Tumor Virology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA
† Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138, USA
‡ Department of Genetics, Stanford University, Stanford, California 94305, USA

The T1/Leu-1/CD5 molecule, a human T-cell surface glycoprotein of relative molecular mass (M_r) 67,000[1-3], has been implicated in the proliferative response of activated T cells[4] and in T-cell helper function[5]. A similar involvement in T-cell proliferation has been reported for Ly-1 (refs 6-8), the murine homologue of T1 (ref. 9). Here we report the complete amino-acid sequence of the T1 precursor molecule deduced from complementary DNA clones. The protein contains a classical signal peptide; a 347-amino-acid extracellular segment; a transmembrane region; and a 93-amino-acid intracellular segment. The extracellular segment contains many cysteine residues and is composed of two related structural domains separated by a proline/threonine-rich region. The T1 molecule has structural features characteristic of other receptor molecules.

We purified the T1 molecule from the lymphoblastoid tumour cell line HPB-ALL using lectin and immunoaffinity chromatographies (Fig. 1*a*). The purified protein had a lower M_r (~56,000, Fig. 1*a*) than that previously described for the T1 molecule (67,000; refs 1, 2). But when the same immunoaffinity column was used to purify the T1 molecules in small scale from surface radioiodinated cells, we obtained a band of M_r 67,000 (Fig. 1*b*, lane 1). When both molecules were treated with endoglycosidase F (Endo F), removing all N-linked oligosaccharides[10], there are identical shifts in mobility on SDS-polyacrylamide gels (Fig. 1*b*, lanes 2, 4). When the M_r 67,000 material was compared with the M_r 56,000 material by peptide mapping (Fig. 1*c*), we found extensive homology, indicating that the M_r 56,000 form is derived from the M_r 67,000 molecule.

It is likely that the M_r 56,000 T1 results from proteolysis during the isolation procedure because protease inhibitors were not added until after membranes were made from frozen cells. In contrast, radioiodinated viable cells from which the M_r 67,000 T1 was isolated were lysed and solubilized in the presence of protease inhibitors. When these cells are lysed in the absence of protease inhibitors some of the M_r 67,000 material is cleaved

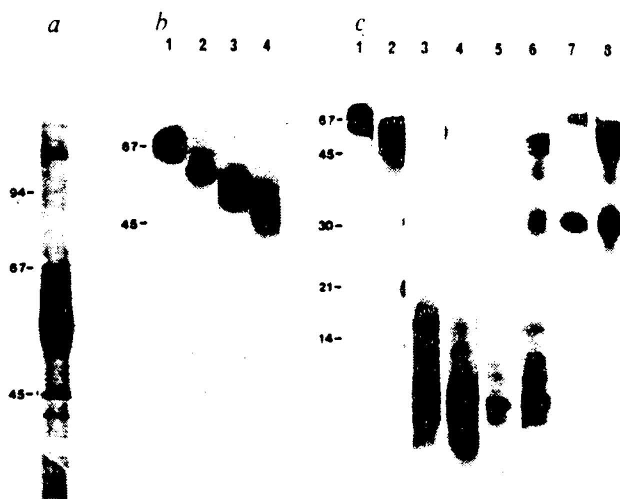

Fig. 1 Purification and analysis of T1. *a*, Silver-stained SDS-polyacrylamide gel of purified T1 (0.1 μg). *b*, SDS-polyacrylamide gel electrophoresis of Endo F cleavage products of ^{125}I-labelled T1. Lane 1, control M_r 67,000 T1; 2, endo F-digested M_r 67,000 T1; 3, control M_r 56,000 T1; 4, endo F-digested M_r 56,000 T1. *c*, One-dimensional peptide mapping of ^{125}I-labelled T1 using 10 μg of papain (lanes 3, 4), chymotrypsin (lanes 5, 6) or V8 protease (lanes 7, 8). Lanes 3, 5 and 7, M_r 67,000 T1; lanes 4, 6 and 8, M_r 56,000 T1. Undigested M_r 67,000 T1 and 56,000 T1 are in lanes 1 and 2, respectively.

Methods. *a*, Frozen HPB-ALL cells (5×10^{10} cells) were thawed, hand homogenized in a Potter Elvehjem homogenizer and crude membranes prepared by hypotonic lysis and differential centrifugation as described[25]. The membranes were solubilized overnight at 4 °C in 500 ml extraction buffer (10 mM Tris, *p*H 7.5, 150 mM NaCl, 1% Nonidet P-40 (NP-40), 1 mM $MnCl_2$, 1 mM phenylmethylsulphonyl fluoride, 1 mM $CaCl_2$, 1 U ml^{-1} aprotinin and 1 mM iodoacetamide) and centrifuged for 60 min at 100,000*g*. The supernatant was applied sequentially to lentil, ricin and wheat-germ lectin affinity columns (EY laboratories) which were eluted with lysis buffer containing 3% (w/v) α-methylmannoside, 10% (w/v) galactose or 5% (w/v) *N*-acetylglucosamine, respectively. Sodium deoxycholate (DOC) was added to the pooled lectin column eluates (0.5% final concentration) and applied to a column of anti-T1 monoclonal antibody 10.2 coupled to cyanogen bromide-activated Sepharose CL-4B (Pharmacia). The column was washed with two column volumes of the following buffers: (1) extraction buffer; (2) 0.5 M LiCl, 10 mM Tris, *p*H 7.5, 0.5% NP-40; (3) 10 mM Tris, 0.1% DOC; and eluted with 50 mM diethylamine, *p*H 11.5, 0.1% DOC. The eluted protein was precipitated with 5 volumes of cold acetone, resuspended in 500 μl water and 1 μl analysed on SDS-polyacrylamide gels that were silver stained to access yield and purity[26]. Yield = 50 μg. *b*, Purified T1 protein (56,000 M_r) (3 μg) was ^{125}I-labelled with Iodobeads (Pierce) according to manufacturers instructions. HPB-ALL cells (5×10^7) were surface-iodinated with lactoperoxidase as described[27], solubilized with 1 ml extraction buffer and T1 protein (M_r 67,000) was purified on an anti-T1 antibody column as described above. An aliquot of each ^{125}I-labelled T1 preparation (with 10 μg soybean trypsin inhibitor as carrier) was acetone-precipitated and the pellet resuspended in 50 μl 50 mM EDTA, 0.5% NP-40, 0.1 M Na phosphate, *p*H 6.1, 1% 2-mercaptoethanol. Endo F (NEN) (1 U) was added to half of each sample and all samples were incubated 4 h at 37 °C. Samples were then analysed by SDS-polyacrylamide gel electrophoresis. *c*, An aliquot of each ^{125}I-labelled T1 preparation was acetone-precipitated as above and one-dimensional peptide mapping was performed on the soluble samples and analysed on a 7-19% gradient SDS-polyacrylamide gel as described[28].

to a band of a M_r ~56,000 (data not shown). This proteolytic cleavage is highly specific as only a single band of M_r 56,000 is seen and because other molecules purified from the same membrane preparations including HLA-A, -B, -C antigens; T-cell receptor α- and β- chains; T6 molecules; and T8 molecules are not degraded.

The 34 N-terminal amino acids of the M_r 56,000 T1 molecule were determined by sequencing the purified molecules on a gas-phase sequenator[11]. The sequence H_2N-Arg-Leu-Ser-Trp-Tyr-Asp-Pro-Asp-Phe-Gln-Ala-Arg-Leu-Thr-Arg-Ser-Asn-Ser-Lys-X-Gln-Gly-Gln-Leu-Glu-Val-Tyr-Leu-Lys-Asp-Gly-Trp-X-Met was obtained twice. We synthesized two sets of non-overlapping oligonucleotide probes consisting of a 23mer of 128-fold degeneracy corresponding to amino acids 4-11 and a 17mer of 256-fold degeneracy corresponding to amino acids 21-26.

A cDNA library was constructed in the phage λgt10 using poly(A)$^+$RNA from the human tumour T-cell line HPB-MLT as previously described[12], and we obtained one clone that hybridized with both probes. This clone, pT1-1, has an insert of ~1.0 kilobase (kb). A second clone of 2.1 kb, pT1-2, was obtained from the same cDNA library using the insert of pT1-1 as a hybridization probe. It overlaps the 3′ end of the pT1-1 cDNA insert by 637 nucleotides, but does not contain the 5′ sequence to which either of the oligonucleotide probes would hybridize.

The partial restriction maps for each insert (Fig. 2*a*) and the 2,320-base pair (bp) combined nucleotide sequences of both (Fig. 2*b*) are shown. The combined nucleotide sequence has a long open reading frame that contains a stretch of nucleotides starting at base 145 that corresponds to the N-terminal amino-acid sequence obtained from the purified T1 protein. There are three potential translation initiation ATG triplets at amino acid positions −8, −22 and −24 (bases 121-123, 79-81 and 73-75, respectively). The initiation site was tentatively placed at nucleotides 73-75 because this is the first ATG triplet that appears downstream from an in-frame termination codon TGA (nucleotides 49-51), and as it is flanked by nucleotides that fulfill criteria for a translation initiation site[13]. The open reading frame ends with the termination codon TAA at nucleotides 1,558-1,560. The 3′ untranslated region is not complete as neither a poly(A) tail nor a polyadenylation signal (AATAAA) is present.

To determine the size of the messenger RNA encoding the T1 molecule, poly(A)$^+$RNA from the HPB-ALL cell line was analysed on Northern blots using the nick-translated insert from pT1-1 (Fig. 3*a*, lane 1). Two mRNAs were detected at 3.6 and 2.7 kb. Thus, the 2,320-bp nucleotide sequence presented here lacks at least 400 bp. An identical hybridization pattern is seen when the insert from pT1-2 is used as a probe (data not shown). No hybridization is seen with mRNA from the T1-negative B lymphoblastoid cell line JY (Fig. 3*a*, lane 2).

The predicted amino-acid sequence (Fig. 2*b*) demonstrates several features of the T1 glycoprotein. The mature protein consists of 471 amino-acid residues with M_r 52,167. The 23 amino-acid residues between the initiation codon at residue −24 and the Arg residue at position 1 are characteristic of a classic signal sequence: they are hydrophobic, of the appropriate length and the sequence terminates at position −1 with a residue containing a small side chain (glycine). The presence of this signal sequence next to the experimentally determined N-terminal amino-acid sequence indicates the M_r 56,000 T1 probably resulted from a proteolytic cleavage from the C-terminus of the intact T1 molecule, possibly near a cluster of Arg and Lys residues at residues 379-391. A second hydrophobic region at amino acid residues 348-378 represents a putative 31-amino acid transmembrane region. To the C-terminal side of residue 378 there is a stretch of basic residues that is typical of sequences found on the cytoplasmic side of a transmembrane region.

There are 5 possible sites for N-linked glycosylation, 2 N-terminal to the transmembrane region and 3 C-terminal. Only the 2 N-terminal sites are used, as the purified M_r 56,000 T1 molecule, which contains the same number of glycans as the M_r 67,000 molecule (Fig. 1*b*), probably lacks most of the region C-terminal to the transmembrane region. As there is no precedent for glycosylation occurring on cytoplasmic residues, it is likely that residues 1-347 are extracellular and residues 379-471 are intracellular. The extracellular region is relatively cysteine-rich (22 cysteines) and contains a 20-residue stretch with 17 threonines and prolines (amino acids 114-133). The Thr/Pro-rich region forms an extended peptide that separates two homologous domains. Alignment of amino-acid residues 1-113 from the first domain with residues 242-347 from the second domain reveals a significant 30% amino-acid homology when alignment is optimized by proposing a single 6 amino-acid deletion between residues 290 and 291 (Fig. 4*a*). This strongly suggests that the domains arose by a gene duplication event. A model of T1 structure is shown in Fig. 4*b*. The nucleotide and

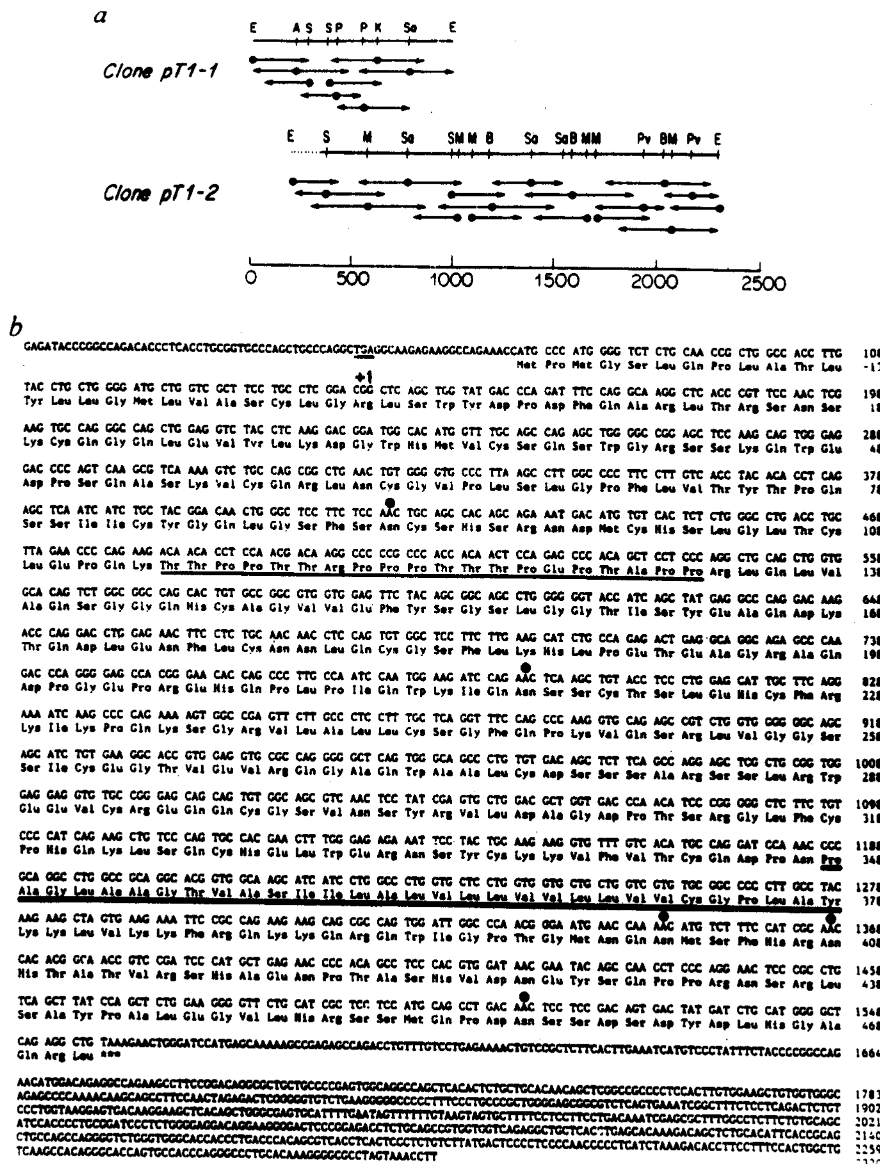

Fig. 2 Nucleotide and predicted amino-acid sequences of T1. *a*, Partial restriction map and sequencing strategy of cDNA inserts pT1-1 and pT1-2. Arrows, direction and extent of sequencing. Dashed line at the 5′ end of the pT1-2 insert, 185-bp aberrant sequence that probably represents the ligation of two cDNAs during cloning. This sequence fails to detect T1-specific RNA. *b*, Nucleotide- and amino-acid sequences of T1. Nucleotides and amino acids are numbered at the right. The start of the mature protein sequence is denoted +1. An in-frame upstream stop codon and the threonine/proline-rich region are underlined. Black bar, transmembrane region; black circles, sites of possible asparagine-linked glycosylation.

Methods. A cDNA library was prepared in λgt10[12] and screened with phosphorylated oligonucleotide probes as described[22]. Each cDNA insert was sequenced from the restriction sites shown using the dideoxy-chain termination technique of Sanger *et al.*[29]. Sequencing was done on both strands. The sequences were analysed using Microgenie Software (Beckman) according to the program of Queen and Korn[14].

protein sequences of T1 were compared with the sequences in the Genbank and National Biomedical Research Foundation databases[14,15]. There are no significant similarities with other known sequences.

To determine the number of T1 genes present in the human genome, we performed Southern blot analyses[16] on HPB-ALL DNA using the insert of pT1-2 as a probe (Fig. 3*b*). Digestion of HPB-ALL DNA with *Eco*RI yields a single fragment of ~12–15 kb. Digestion with *Bgl*II yields two fragments of ~6 and 2.5 kb and digestion with *Pst*I, which has three cleavage sites in the cDNA sequence, produces four fragments of 0.7, 1.4, 1.6 and 3.5 kb. These data suggest that the T1 molecule is encoded by a single-copy gene.

To provide further evidence that the cDNA clones presented here encode the T1 protein, DNA from secondary L-cell transfectants selected for T1 expression by fluorescence activated cell sorting (FACS) was screened using the insert from pT1-1 (Fig. 3*c*). *Hin*dIII digestion of DNA from the three transfectants (Fig. 3*c*, lanes 3–5) and from JM, a T1$^+$ human T-cell line (Fig. 3*c*, lane 2), yields a major band of 4.7 kb in all cases which is amplified 20–50-fold in the T1-selected transfectants. A weaker band of 2.8 kb seen with the transfectants occurs in the JM cell line after longer exposure times (data not shown). The faint bands above the 4.7-kb band could not be detected from JM and may represent a small portion of T1 DNA that was either integrated into the chromosome at a different location from most T1 DNA or is present in double minute chromosomes[17].

The T1 molecule is expressed on most peripheral T cells[1–3] and on a subset of B cells (2–5%) in lymph nodes[18]. The T1 molecule is thought to be the human equivalent of the murine Ly-1 molecule[9]. The insert of pT1-1 has been used to clone cDNAs encoding Ly-1 (H.-J.S.H. *et al.*, in preparation). These cDNAs have been expressed in L cells which then bind αLy-1 antibodies, thus confirming that the two molecules are homologous. Studies by other workers have suggested that Ly-1 and T1 are directly involved in sustaining the proliferation of activated T cells[4–8]. Indeed, Ly-1 may be the receptor for some as yet undetermined ligand.

The high cysteine content, long cytoplasmic domain suitable for signal transduction and phosphorylation seen following incubation of HPB-ALL cells with ^{32}P-orthophosphate (data not shown) are structural features of other transmembrane receptor molecules, including the low-density lipoprotein receptor[19], the epidermal growth factor receptor[20], the insulin receptor[21], the interleukin-2 receptor[22,23] and the transferrin receptor[24]. The availability of a cDNA clone encoding T1 should help to elucidate the function and regulation of this molecule that is apparently involved in the growth of T lymphocytes.

We thank Joyce Culgin and Maryellen Thomas for the preparation of the manuscript; Dr John Hansen for the gift of 10.2

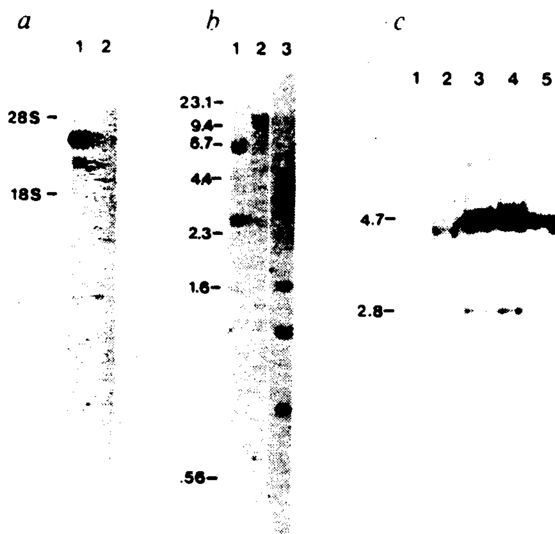

Fig. 3 *a*, Northern blot analysis of T1 mRNA. Poly (A)$^+$ mRNA (2 μg) from the T cell line HPB-ALL (lane 1) or the B lymphoblastoid cell line JY (lane 2) was electrophoresed through a 1% agarose/formaldehyde gel and transferred to nitrocellulose as described[30]. The filter was hybridized overnight with nick-translated pT1-1 insert (specific activity 10^8 c.p.m. μg^{-1}) and washed as described[30]. Sizes estimated from HPB-ALL 28S (5.2 kb) and 18S (2.0 kb) ribosomal RNAs. *b*, Southern blot analysis of T1 genomic DNA. HPB-ALL cellular DNA (10 μg) was digested with *Bgl*II (lane 1), *Eco*R1 (lane 2) or *Pst*I (lane 3) and separated through a 1% agarose gel. The DNA in the gel was denatured, blotted onto nitrocellulose and the filter hybridized overnight with nick-translated pT1-2 insert[30]. Size markers, λ phage DNA digested with *Hin*dIII. *c*, Southern blot analysis of T1-amplified L cell transfectants. DNA (15 μg) from murine L cells (lane 1) or the T-cell line JM (lane 2) and DNA (5 μg) from three independent T1-L cell transfectants (lanes 3–5) were digested with *Hin*dIII, electrophoresed on an 0.8% agarose gel and analysed by Southern blots using the nick-translated insert of pT1-1 as probe[16,30]. L cells were co-transfected with the herpes simplex thymidine kinase gene and total human DNA from the T-cell JM and T1-positive transfectants were selected using FACS[31]. Secondary transfectants were obtained by transfecting the DNA from three different T1 positive primary transfectants into L cells. These cells were amplified for T1 expression by several rounds of sorting the most positive 0.5% of cells as described[17].

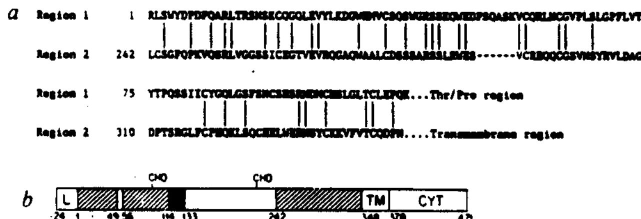

Fig. 4 *a*, Amino-acid sequence alignment of two homologous regions of T1. Region 1 spans residues 1–113 and region 2 residues 242–347. Dashes, 6-amino-acid deletion in region 2. *b*, Model of T1 structure. L, Leader or signal peptide; TM, transmembrane region; CYT, cytoplasmic region; CHO, positions of putative N-linked glycosylation; hatched regions, the two homologous domains.

monoclonal antibody; Dr David Andrews and William Lane for N-terminal amino-acid sequencing and oligonucleotide synthesis; Mary-Ann Lane for use of DNA-analysis program; and Dr James Woody (Naval Medical Research Institute) for providing HPB-ALL cells for protein purification. This work was supported by NIH AI15669 (J.L.S.), GM-17367, GM-28428 (L.A.H.), NCI 5R32 CA09031 (N.H.J.), CA04681 (H.-J.S.H.) and Damon Runyon–Walter Winchell Cancer Fund (D.P.D.).

Received 18 June; accepted 16 July 1986.

1. Engleman, E. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 1791–1795 (1981).
2. Martin, P. J., Hansen, J. A., Nowinski, R. C. & Brown, M. A. *Immunogenetics* **11**, 429–439 (1980).
3. Reinherz, E. L., Kung, P. C., Goldstein, G. & Schlossman, S. *J. Immun.* **123**, 1312–1317 (1979).
4. Ledbetter, J. A. *et al. J. Immun.* **135**, 2331–2336 (1985).
5. Thomas, Y. *et al. J. Immun.* **133**, 724–728 (1984).
6. Hollander, N., Pillemer, E. & Weissman, I. L. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1148–1151 (1981).
7. Stanton, T., Stevens, T. L., Ledbetter, J. A. & Wofsy, D. *J. Immun.* **136**, 1734–1737 (1986).
8. Lodgberg, L. & Shevach, E. M. *Eur. J. Immun.* **15**, 1007–1013 (1985).
9. Ledbetter, J. A. *et al. J. exp. Med.* **153**, 310–323 (1981).
10. Elder, J. H. & Alexander, S. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4540–4544 (1982).
11. Hewick, R. M., Hunkapillar, M. W., Hood, L. E. & Dreyer, W. J. *J. biol. Chem.* **256**, 7990–7997 (1981).
12. Dialynas, D. P. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 2619–2623 (1986).
13. Kozak, M. *Nucleic Acids Res.* **12**, 857–872 (1984).
14. Queen, C. & Korn, L. J. *Nucleic Acids Res.* **12**, 581–599 (1984).
15. Lipman, D. J. & Pearson, W. R. *Science* **227**, 1435–1441 (1985).
16. Southern, E. *J. molec. Biol.* **98**, 503–517 (1975).
17. Kavathas, P. & Herzenberg, L. A. *Nature* **306**, 385–387 (1983).
18. Caligaris-Cappio, F., Gobbi, M., Bofill, M. & Janossy, G. *J. exp. Med.* **155**, 623–628 (1982).
19. Yamamoto, T. *et al. Cell* **39**, 27–38 (1984).
20. Ullrich, A. *et al. Nature* **309**, 418–425 (1984).
21. Ullrich, A. *et al. Nature* **313**, 756–761 (1985).
22. Leonard, W. J. *et al. Nature* **311**, 626–631 (1984).
23. Nikaido, T. *et al. Nature* **311**, 631–635 (1984).
24. McClelland, A., Kuhn, L. C. & Ruddle, F. H. *Cell* **39**, 267–274 (1984).
25. Turner, M. J. *et al. J. biol. Chem.* **250**, 4512–4519 (1975).
26. Morrissey, J. H. *Analyt. Biochem.* **117**, 307–310 (1981).
27. Collins, J. J. *et al. Virology* **86**, 205–213 (1978).
28. Cleveland, D. W., Fischer, S. G., Kirschner, M. W. & Laemmli, U. K. *J. biol. Chem.* **252**, 1102–1106 (1977).
29. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
30. Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
31. Kavathas, P. & Herzenberg, L. A. *Proc. natn. Acad. Sci. U.S.A.* **80**, 524–528 (1983).