Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 6283–6287, July 1994
Biochemistry

Three distinct human thymopoietins are derived from alternatively spliced mRNAs

(thymic proteins/nuclear proteins/alternative splicing)

CRAFFORD A. HARRIS, PAULA J. ANDRYUK, SCOTT CLINE, H. KAREN CHAN, ANAN NATARAJAN*, JOHN J. SIEKIERKA, AND GIDEON GOLDSTEIN

Immunobiology Research Institute, Annandale, NJ 08801

Communicated by Edward A. Boyse, March 28, 1994

ABSTRACT **Thymopoietin (TP) was originally isolated as a 5-kDa 49-aa protein from bovine thymus in studies of the effects of thymic extracts on neuromuscular transmission and was subsequently observed to affect T-cell differentiation and function. We now report the isolation of cDNA clones for three alternatively spliced mRNAs that encode three distinct human T-cell TPs. Proteins encoded by these mRNAs, which we have named TPα (75 kDa), TPβ (51 kDa), and TPγ (39 kDa), contain identical N-terminal regions, including sequences nearly identical to that of the originally isolated 49-aa protein, but divergent C-terminal regions. TP mRNAs are expressed in many tissues, most abundantly in adult thymus and fetal liver of the tissues so far examined. Distinct structural domains and functional motifs in TPs α, β, and γ suggest that the proteins have unique functions and may be directed to distinct subcellular compartments.**

Thymopoietin (TP) was originally isolated as a 5-kDa polypeptide from bovine thymus on the basis of its ability to affect neuromuscular transmission when injected into mice (1). Subsequently, purified 5-kDa TP was found to affect certain immunological functions, inducing differentiation of prothymocytes to thymocytes, as measured by changes in expression of Thy-1 and other cell surface phenotypic markers (2), and enhancing allogeneic responses of peripheral T cells (3). The 5-kDa TP and a synthetic peptide corresponding to aa 32–36 of TP, Arg-Lys-Asp-Val-Tyr (thymopentin) gave similar results in both neurophysiological and immunological assays, suggesting that thymopentin may act as a TP mimetic (4, 5). Furthermore, in clinical studies with thymopentin being used as a putative immunoregulatory drug, promising findings have been obtained in double-blind, placebo-controlled studies of subjects with rheumatoid arthritis (6), recurrent herpes simplex (7), atopic dermatitis (8), and asymptomatic human immunodeficiency virus infection (9).

To better define the cellular function(s) of TP and to provide a molecular basis for the mechanism of action of thymopentin, we isolated and characterized cDNAs encoding this protein. Zevin-Sonkin *et al.* (10) previously isolated a bovine cDNA that encodes the originally determined 49-aa bovine TP sequence (11, 12) at the 5′ end of a larger open reading frame. Here we report the isolation, using a probe derived from the bovine sequence of Zevin-Sonkin *et al.* (10), of human TP cDNA clones that encode three distinct human TP proteins, derived from three alternatively spliced mRNAs, each of which contains a sequence at the N terminus that is homologous to the originally determined bovine 49-aa TP sequence. TPα (75 kDa), TPβ (51 kDa), and TPγ (39 kDa), contain some domains in common and some unique domains, suggesting that these proteins may serve distinct functions and may be directed to distinct subcellular locations.†

MATERIALS AND METHODS

Isolation of Human TP cDNA and Genomic Clones. Overlapping partial cDNA clones were isolated from a library prepared from thymus mRNA (Clontech) and a library prepared from the T-cell line Jurkat (Stratagene) in multiple rounds of screening. The initial probe used was an oligonucleotide containing sequences encoding bovine TP aa 1–42 from the published partial bovine cDNA of Zevin-Sonkin *et al.* (10). In subsequent rounds of screening, probes derived from the initial human clones were used. The genomic clone λSHG-1 was isolated from a λFIX II library (Stratagene). Library screening methods were essentially as described (13).

Sequence Determination and Analysis. All sequences reported, with the exception of the 3′ untranslated sequences of TPs β and γ, were determined in at least two independent clones and on both strands of at least one clone. Some sequencing was done by LARK Sequencing Technologies (Houston). Sequences were compared to release 7.0 of the Entrez:Sequences collection of data bases from the National Center for Biotechnology Information, using the modified FASTA (14) program in MACVECTOR 4.1 (Kodak). Protein sequences were searched for motifs in release 9 of the Prosite data base (15) by using MACPATTERN (16). Hydropathy analysis was by the method of Engelman *et al.* (17) as implemented in MACVECTOR.

Northern Blot Analysis. Poly(A)$^+$ RNA from the human T-cell line CEM (American Type Culture Collection) was prepared by acid guanidinium thiocyanate/phenol/chloroform (18) extraction using RNAzol (Cinna/Biotecx Laboratories, Friendswood, TX), followed by selection on oligo(dT) columns (13). Poly(A)$^+$ RNA from human tissues and blots of human tissue mRNAs were purchased from Clontech. Glyoxylated poly(A)$^+$ RNAs were separated in 1.2% agarose gels, blotted to nylon membranes, and hybridized and washed as described (13). Sizes of mRNAs were determined by comparison to RNA size markers (GIBCO/BRL).

Probes for detection of TP mRNAs were overlapping oligodeoxynucleotides that were radiolabeled by extension of 3′ ends to generate the complete double-stranded sequence. Oligonucleotide sequences were as follows: α/β/γ, nt 1–87 (Fig. 1) (sense) and 156–64 (antisense); α-specific, nt 1488–1587 (sense) and 1587–1571 (antisense); β-specific, nt 849–898 (antisense) and 929–879 (antisense); β/γ-specific, nt 1286–1330 (sense) and 1365–1316 (antisense). The glyceraldehyde-3-phosphate dehydrogenase (GAPDH) probe was obtained from Clontech.

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.

Abbreviation: TP, thymopoietin.

*Present address: Terumo Corp., Somerset, NJ 08873.

†The sequences reported in this paper have been deposited in the GenBank data base (accession nos. U09086–U09088).

FIG. 1. *(Legend appears at the bottom of the opposite page.)*

Expression of Recombinant Human TPs in *Escherichia coli*. Recombinant TPs were expressed in *E. coli* BL21(DE3) using inducible T7 RNA polymerase-dependent pET vectors (ref. 19; Novagen). Plasmid pET17b-hTPα was constructed by PCR amplification of the open reading frame from λT.32 and an overlapping fragment from λT.153 and insertion into pET-17b. This construct encodes TPα as a fusion protein with three additional amino acids, Met-Ala-Ser, at the N terminus. The vectors pET3a-hTPβ and pET3a-hTPγ were constructed by PCR amplification of the open reading frames from λT.17 and λT.206, respectively, and insertion into pET-3a.

Protein Immunoblotting. Mammalian cell extracts were prepared by lysis of cells in 1% (vol/vol) Nonidet P-40/20 mM Tris·HCl, pH 7.5/150 mM NaCl/1 mM EDTA/0.1 mM EGTA/0.5 mM dithiothreitol, plus the following protease inhibitors (Boehringer Mannheim): aprotinin, 10 μg/ml; pepstatin, 0.3 mM; Pefabloc, 0.1 mM; and E-64, 1 μg/ml. After passage through a 27½-gauge needle to reduce viscosity and centrifugation to remove insoluble material, sample buffer (20) was added. *E. coli* extracts were prepared by direct lysis in sample buffer. Proteins were separated by SDS/PAGE and transferred to nitrocellulose, and TPs were detected with affinity-purified rabbit antiserum 81912 raised against a synthetic peptide containing TP aa 1–19; this antiserum also recognizes a synthetic TP containing aa 1–52, which corresponds to the originally described 5-kDa TP species (unpublished work). Blots were developed with peroxidase-linked goat anti-rabbit immunoglobulin (Pierce) and an enhanced chemiluminescence system (Amersham).

RESULTS

Analysis of TP α, β, and γ Sequences. Sequencing of overlapping partial cDNA clones revealed the existence of three distinct alternatively spliced human TP mRNAs encoding three distinct TP proteins that contain both unique and common domains (Fig. 1). The three proteins, which we have named TPs α, β, and γ, have predicted molecular masses of 75 kDa, 51 kDa, and 39 kDa, respectively. TPs α, β, and γ have identical N-terminal domains through Glu^{187}, and aa 1–49 are homologous to the originally determined 49-aa sequence of purified bovine thymopoietin (11, 12). After Glu^{187}, TPα diverges from TPs β and γ. TPγ differs from TPβ only in lacking the β-specific domain containing aa 221–329 of TPβ. In regions where the amino acid sequences of TPs α, β, and γ are identical, their nucleotide sequences are identical as well, consistent with their originating by alternative splicing of transcripts from a single gene. This was confirmed by sequencing of genomic clones (unpublished work).

Hydropathy analysis (17) revealed that TPs α, β, and γ lack an N-terminal hydrophobic signal sequence typical of secreted polypeptides. However, the analysis revealed that TPs β and γ contain a possible hydrophobic transmembrane domain near their C termini (Fig. 1), suggesting possible association of these proteins with cellular membranes. Analysis of TP sequences for motifs in the Prosite data base (15) revealed several potential phosphorylation sites for protein kinases, including KTYDAASY, aa 619–626 of TPα, a sequence that matches a consensus sequence for phosphorylation by some tyrosine kinases [(K or R)$X_{2\text{ or }3}$(D or E)$X_{2\text{ or }3}$Y] (21).

Comparison of the sequences of TPs α, β, and γ to the Entrez:Sequences collection of data bases did not reveal any compelling similarities to previously known protein or nucleic acid sequences other than TP.

TPs α, β, and γ Are the Major TP Species Expressed in a Human T-Cell Line. Recombinant TPs α, β, and γ expressed in *E. coli* comigrated with the three major TPs detected in extracts of the T-cell line CEM with an antiserum against TPα/β/γ aa 1–19 (Fig. 2*A*). An additional, less abundant ≈43-kDa immunoreactive protein has not been further characterized. In some experiments another less abundant immunoreactive protein, slightly larger than TPα, was also detected (data not shown).

TP mRNAs in T-Cell Lines. Three major TP mRNAs, estimated to be 4.4 kb, 4.1 kb, and 4.0 kb, were detected in CEM cells with an oligonucleotide probe containing sequences encoding aa 1–52 of the TPα/β/γ common region (Fig. 2*B*, left and right lanes). Only the 4.0-kb mRNA was detected with an α-specific probe, only the 4.4-kb mRNA was detected with a β-specific probe, and the 4.1-kb mRNA was detected with a β/γ-specific probe but not with α-specific or β-specific probes (Fig. 2*B*). This suggests that the 4.0-kb mRNA encodes TPα, the 4.4-kb mRNA encodes TPβ, and the 4.1-kb mRNA encodes TPγ. The smaller size of TPα mRNA relative to TPβ and TPγ mRNAs despite its encoding the largest protein indicates that it contains shorter 3′ and/or 5′ untranslated regions; the complete 3′ and 5′ untranslated regions have not been cloned.

Expression of TP mRNAs in Adult and Fetal Tissues. TP mRNAs were detected in all tissues examined, with highest expression in adult thymus and in fetal liver (Fig. 3). In some tissues, TP mRNAs of slightly different sizes than the thymus mRNAs were resolved when electrophoresis times were extended (Fig. 3*A*). Whether such differences result from different 5′ or 3′ untranslated regions or additional distinct patterns of alternative splicing of coding exons is not known. Expression of TPs α, β, and γ in many tissues, with particularly high expression in thymus, was also observed in rodents, and initial analysis of rat TP cDNAs suggests a high level of sequence conservation between rat and human TPα (unpublished observations).

DISCUSSION

Three major TPs, α (75 kDa), β (51 kDa), and γ (39 kDa), were demonstrable in extracts of the T-cell line CEM by immunoblotting with antibodies to a synthetic peptide corresponding to aa 1–19 predicted from the cDNAs reported here. The TP α, β, and γ mRNAs that we described minimally encode

FIG. 1. (*on opposite page*). Nucleotide and predicted amino acid sequences and schematic diagram of human TPα (*A*) and of human TPs β and γ (*B*). Sequences were determined from overlapping partial cDNA and genomic clones. The TPα sequence was determined from the cDNA clones λT.32 and λT.153, and confirmed in the genomic clone λSHG-1. The TPβ sequence was determined from the cDNA clones λT.6, λT.17, and λT.209. The TPγ sequence was determined from the cDNA clone λT.206. As the 5′ and 3′ end sequences have not been determined, the sequences have been numbered so that amino acid +1 is the N-terminal proline of mature TP and nucleotide +1 is the first nucleotide of the proline codon. Stars below Glu^{187} indicate the end of the N-terminal region that is common to TPs α, β, and γ. TPγ differs from TPβ only in that it lacks an alternatively spliced 109-aa central domain (shaded). Nucleotide and amino acid numbers in parentheses are for TPγ, after subtraction of the β-specific region. A sequence in TPα resembling a consensus sequence for tyrosine phosphorylation preceeding Tyr^{626} is underlined. A hydrophobic, potentially membrane-spanning, domain in TPs β and γ is doubly underlined. Nucleotide 1792 in TPα is C in λT.153, but G in λSHG-1, which changes aa 598 from Gln to Glu. Nucleotide 579 is C in λT.6 and in λSHG-1, but T in λT.206, in both cases encoding Leu. The first ≈60 nt of λT.206 sequence (TPγ) do not match other cDNA or genomic sequences (data not shown). We assume that this sequence is the result of a cloning artifact and that the TPβ and TPγ mRNA sequences have identical 5′ untranslated regions. The poly(A) tract at the end of the β/γ sequence is part of an *Alu* repeat in the 3′ untranslated region and is not the 3′ end of the mRNA.

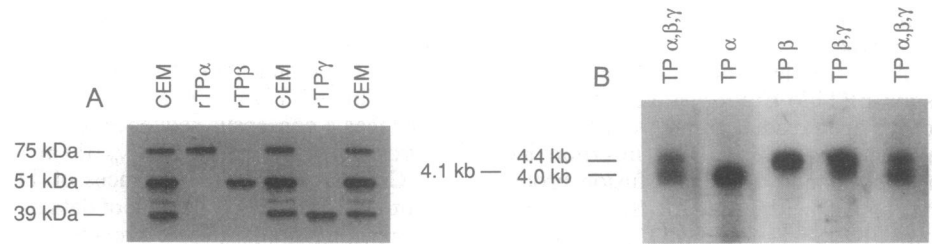

FIG. 2. Expression of TP proteins and RNAs in CEM cells. (*A*) TP α, β, and γ cDNAs expressed in *E. coli* produce 75-kDa, 51-kDa, and 39-kDa proteins that comigrate with the major TP proteins expressed in CEM cells. Lysate proteins of the human T-cell line CEM or of *E. coli* strains expressing recombinant TPs were separated by SDS/PAGE and immunoblotted with an antiserum raised against aa 1–19 of the common N-terminal region of TPs α, β, and γ. The molecular masses of TPs α, β, and γ predicted from the cDNAs were confirmed by comparison with marker proteins in separate experiments (data not shown). (*B*) Thymopoietin mRNAs in CEM cells. TP mRNAs in CEM cells were identified on Northern blots by probes containing sequences from the N-terminal common region that is present in TPs α, β, and γ or probes specific for TPα, for TPβ, or for TPs β and γ, as described in *Materials and Methods*. The ≈4.1-kb TPγ mRNA and the ≈4.0-kb TPα mRNA are not well resolved in the left and right lanes.

these three proteins. This was shown by comigration of recombinant TPs α, β, and γ expressed in *E. coli* with the CEM proteins. No immunoreactive proteins corresponding in size to the originally purified 5-kDa 49-aa bovine TP (1, 11, 12) were detected by immunoblotting of extracts of CEM, other cell lines, or tissues (data not shown). These extracts were prepared rapidly in lysis solutions containing SDS and/or protease inhibitors, reagents that were not used in the original isolations (1, 11, 12). This suggests that the originally purified 5-kDa protein may have been a proteolytic fragment of the presently described larger TPs, the fragment being active in the bioassays used for its isolation.

The first 49 aa predicted by the human TP α, β, and γ cDNAs are closely similar to the sequence determined for the originally purified bovine 5-kDa TP, differing at only 5 amino acids (11, 12). Comparison of the human cDNA derived sequence with the bovine cDNA of Zevin-Sonkin *et al.* (10) reveals only 1 amino acid difference (at position 13) between human and bovine in the first 49 aa. In contrast, a previously reported 48-aa sequence for the human 5-kDa TP differs at 12 amino acids (22), and this sequence now appears to be incorrect.

The sequences predicted from the human TP α, β, and γ cDNAs are similar to the sequence predicted from the bovine cDNA reported by Zevin-Sonkin *et al.* (10) from aa 1 through aa 81, differing only at positions 13 (Asp in human, Glu in bovine) and 53–56 (Pro-Ala-Gly-Thr in human, Ala-Thr-Ser-Ala in bovine); but beyond aa 81 there is no further homology, either in nucleotide or in amino acid sequence. Sequencing of the human TP$\alpha/\beta/\gamma$ gene in a genomic clone revealed that the DNA sequence encoding aa 81 lies in the middle of an exon with no nearby potential splice donor sites (unpublished work), indicating that a TP containing C-terminal sequence similar to the bovine sequence reported by Zevin-Sonkin *et al.* is unlikely to be produced from the human TP$\alpha/\beta/\gamma$ gene.

The existence of shared and unique domains in TPs α, β, and γ resulting from alternative splicing suggests that these proteins may have both shared and unique activities and may be localized to different subcellular compartments. TPs α, β, and γ lack classical N-terminal hydrophobic signal sequences for secretion; this suggests that they may be largely localized intracellularly and may have important intracellular functions. However, early observations of the effects of TP on neuromuscular transmission (1) along with an analysis of neuromuscular transmission in thymectomized and thymus-grafted rats (23) provided evidence for TP secretion. Preliminary analysis of conditioned media from human and mouse T-cell lines by TP immunoassay is consistent with the pres-

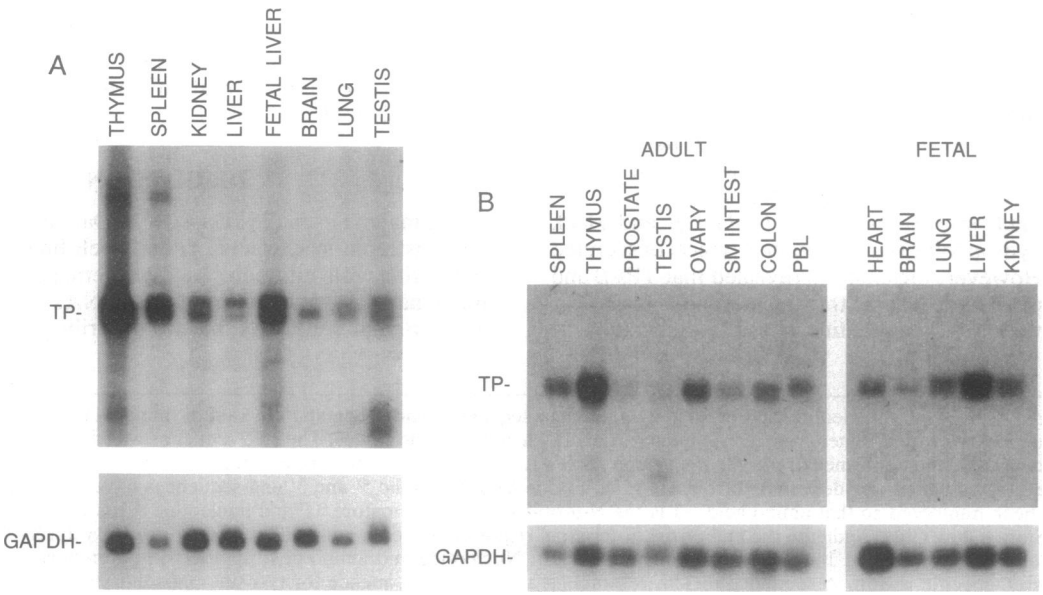

FIG. 3. Expression of TP mRNAs in human adult and fetal tissues. TP mRNAs were detected with the $\alpha/\beta/\gamma$ probe described in *Materials and Methods*. The glyceraldehyde-3-phosphate dehydrogenase (GAPDH) probe was used to control for any differences in sample loading and transfer. (*A*) Glyoxylated poly(A)$^+$ RNAs were separated in agarose gels and blotted to nylon. (*B*) Multiple tissue RNA blots were obtained from Clontech.

ence of extracellular TP, which remains to be characterized (unpublished observations). Extracellular TP could possibly be secreted by an alternative secretion pathway such as that used by interleukin 1 or the fibroblast growth factors, which also have no classical signal sequences (24). TPs β and γ do contain a hydrophobic domain near their C termini, which may function as a transmembrane signal–anchor domain. Western blot analysis of crude plasma membrane preparations derived from T-cell lines failed to indicate the presence of TP (unpublished observations); however, preliminary analysis of TP subcellular localization by immunofluorescence microscopy (M. A. Talle, personal communication) suggests that TPs β and γ are localized to the nuclear membrane. The net positive charge on the N-terminal side of the putative transmembrane domain along with a net negative charge on the C-terminal side suggests an orientation where the N-terminal domain of the protein, and therefore the bulk of the protein, is intranuclear (25). Interestingly, in these same experiments, TPα (which lacks an obvious transmembrane domain) appears to be localized within the nucleus. In addition, examination of the TPα amino acid sequence reveals a short region of basic amino acids (aa 189–195) suggestive of a nuclear localization domain (26).

Examination of TP α, β, and γ sequences for additional motifs revealed potential phosphorylation sites for several protein kinases. Of particular interest is a consensus sequence for tyrosine phosphorylation in TPα at Tyr^{626}. This is intriguing in that phosphorylation on tyrosine often serves to regulate proteins involved in controlling cell growth, activation, or differentiation.

TP mRNA expression was detected in all tissues examined, suggesting that some TP function(s) may be important in many or all cell types. However, TP mRNA expression was highest in adult thymus and in fetal liver, a major fetal site for production of T-cell precursors. This suggests that TPs may play especially important roles in T-cell development and function, as also suggested by earlier reports that TP promotes prothymocyte differentiation (2) and enhances the allogeneic response of peripheral T cells (3).

In summary, TPs appear to be proteins without strong sequence similarity to other proteins in current data bases. The combinatorial arrangement of structural domains and functional motifs arising from alternative splicing suggests that TPs may have multiple functions extending beyond those previously proposed. The high level of expression of TPs in adult thymus and fetal liver is consistent with TPs playing especially important roles in T-cell development and function, yet their expression at lower levels in many other tissues suggests that one or more of their functions is important in most cell types.

We thank Duane Mitchell, Michael Peck, and Mohan Viswanathan, summer student interns, for their contributions to this work; Diane Setcavage, Michael Culler, Ponniah Shenbagamurthi and coworkers for generating antibodies to synthetic peptide fragments of thymopoietins; and Mary Anne Talle, Scott Wadsworth, Janet Andres, and Maureen Morgan for communicating unpublished results and for much helpful advice.

1. Goldstein, G. (1974) *Nature (London)* **247**, 11–14.
2. Basch, R. S. & Goldstein, G. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 1474–1478.
3. Sunshine, G. H., Basch, R. S., Coffey, R. G., Cohen, K. W., Goldstein, G. & Hadden, J. W. (1978) *J. Immunol.* **120**, 1594–1599.
4. Goldstein, G., Scheid, M. P., Boyse, E. A., Schlesinger, D. H. & Wauwe, J. V. (1979) *Science* **204**, 1309–1310.
5. Ranges, G. E., Goldstein, G., Boyse, E. A. & Scheid, M. P. (1982) *J. Exp. Med.* **156**, 1057–1064.
6. Kantharia, B. K., Goulding, N. J., Hall, N. D., Davies, J., Maddison, P. J., Bacon, P. A., Farr, M., Wojtulewski, J. A., Englehart, K. M., Liyanage, S. P. & Cox, N. L. (1989) *Br. J. Rheumatol.* **28**, 118–123.
7. Bolla, K., DeMaubeuge, J., Liden, S., Bonerandi, J. & Sundal, E. (1988) *Int. J. Immunother.* **4**, 219–227.
8. Leung, D. Y. M., Hirsch, R. L., Schneider, L., Moody, C., Takaoka, R., Li, S. H., Meyerson, L. A., Mariam, S. G., Goldstein, G. & Hanifin, J. M. (1990) *J. Allergy Clin. Immunol.* **85**, 927–933.
9. Conant, M. A., Calabrese, L. H., Thompson, S. E., Poiesz, B. J., Rasheed, S., Hirsch, R. L., Meyerson, L. A., Kremer, A. B., Wang, C.-C. & Goldstein, G. (1992) *AIDS* **6**, 1335–1339.
10. Zevin-Sonkin, D., Ilan, E., Guthmann, D., Riss, J., Theodor, L. & Shoham, J. (1992) *Immunol. Lett.* **31**, 301–310.
11. Schlesinger, D. H. & Goldstein, G. (1975) *Cell* **5**, 361–365.
12. Audhya, T., Schlesinger, D. H. & Goldstein, G. (1981) *Biochemistry* **20**, 6195–6200.
13. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
14. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
15. Bairoch, A. (1993) *Nucleic Acids Res.* **21**, 3097–3103.
16. Fuchs, R. (1991) *Comput. Appl. Biosci.* **7**, 105–106.
17. Engelman, D. M., Steitz, T. A. & Goldman, A. (1986) *Annu. Rev. Biophys. Biophys. Chem.* **15**, 321–353.
18. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159.
19. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990) *Methods Enzymol.* **185**, 60–89.
20. Laemmli, U. K. (1970) *Nature (London)* **227**, 658–680.
21. Patschinsky, T., Hunter, T., Esch, F. S., Cooper, J. A. & Sefton, B. M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 973–977.
22. Audhya, T., Schlesinger, D. H. & Goldstein, G. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 3545–3549.
23. Goldstein, G. & Hofmann, W. F. (1969) *Clin. Exp. Immunol.* **4**, 181–189.
24. Rubartelli, A. & Sitia, R. (1991) *Biochem. Soc. Trans.* **19**, 255–259.
25. Hartmann, E., Rapoport, T. A. & Lodish, H. F. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5786–5790.
26. Kalderon, D., Roberts, B. L., Richardson, W. D. & Smith, A. E. (1984) *Cell* **39**, 499–509.