Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 1261–1265, March 1986
Biochemistry

Primary structure of the human melanoma-associated antigen p97 (melanotransferrin) deduced from the mRNA sequence

(tumor-associated antigen/amino acid sequence/transferrin superfamily)

TIMOTHY M. ROSE*, GREGORY D. PLOWMAN*†, DAVID B. TEPLOW‡, WILLIAM J. DREYER‡, KARL ERIK HELLSTRÖM*†, AND JOSEPH P. BROWN*†

*ONCOGEN, 3005 First Avenue, Seattle, WA 98121; †Department of Pathology, University of Washington Medical School, Seattle, WA 98195; and ‡Division of Biology, California Institute of Technology, Pasadena, CA 91125

Communicated by Hans Neurath, October 21, 1985

ABSTRACT **p97 is a cell-surface glycoprotein that is present in most human melanomas but only in trace amounts in normal adult tissues. To determine the structure of this tumor-associated antigen and to identify its functional domains, we have purified and cloned p97 mRNA and determined its nucleotide sequence. The mRNA encodes a 738-residue precursor, which contains the previously determined N-terminal amino acid sequence of p97. After removal of a 19-residue signal peptide, the mature p97 molecule comprises extracellular domains of 342 and 352 residues and a C-terminal 25-residue stretch of predominantly uncharged and hydrophobic amino acids, which we believe acts as a membrane anchor. Each extracellular domain contains 14 cysteine residues, which form seven intradomain disulfide bridges, and one or two potential N-glycosylation sites. Protease digestion studies show that the three major antigenic determinants of p97 are present on the N-terminal domain. The domains are strikingly homologous to each other (46% amino acid sequence homology) and to the corresponding domains of human serum transferrin (39% homology). Conservation of disulfide bridges and of amino acids thought to compose the iron binding pockets suggests that p97 is also related to transferrin in tertiary structure and function. We propose that p97 be renamed melanotransferrin to denote its original identification in melanoma cells and its evolutionary relationship to serotransferrin and lactotransferrin, the other members of the transferrin superfamily.**

p97 is a tumor-associated antigen that was first identified in human melanoma by using monoclonal antibodies (1–3). It has been studied extensively with regard to its expression in normal and neoplastic tissues and is present in most human melanomas and in certain fetal tissues, but it is found only in trace amounts in normal adult tissues (4–6). p97 has been used as a target for diagnostic imaging of melanomas in human clinical trials (7).

p97 is a monomeric cell surface sialoglycoprotein, with an apparent molecular weight as determined by $NaDodSO_4$/polyacrylamide gel electrophoresis of slightly less than 97,000 (4). Use of monoclonal antibodies has defined three major antigenic sites, which are present on a stable M_r 40,000 tryptic fragment (4). Subsequent work has shown that at least two other independently characterized human melanoma-associated antigens, gp95 (3) and gp87 (8), are identical to p97.

The N-terminal amino acid sequence of p97 is homologous to transferrin and, like transferrin, p97 binds iron (9). Analysis of somatic cell hybrids and by *in situ* hybridization has shown that the p97 gene, like the genes for transferrin and the transferrin receptor, is located on chromosomal region 3q21-3q29 (10, 11). These observations suggest that p97 plays a role in iron metabolism. To determine the structure of p97 and identify functional and antigenic domains, we have cloned and sequenced p97 mRNA. The availability of cloned p97 cDNA will allow us to study the regulation of the expression of p97 and to develop animal models to study the usefulness of such tumor-associated antigens in tumor therapy.

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.

METHODS

Polysome Immunopurification. Polysomes prepared from SK-MEL 28 melanoma cells (12) by magnesium precipitation (13) were purified by affinity chromatography using three monoclonal antibodies specific for p97 as described (14). p97-enriched mRNA was isolated by elution with EDTA and purified by affinity chromatography on oligo(dT)-cellulose (Bethesda Research Laboratories).

cDNA Cloning with Oligo(dT) as Primer. For first-strand cDNA synthesis, p97-enriched mRNA and oligo(dT) (Collaborative Research, Waltham, MA) were incubated with reverse transcriptase (Molecular Genetic Resources). The second strand was synthesized by incubation with the large fragment of *Escherichia coli* DNA polymerase (Bethesda Research Laboratories), and the double-stranded cDNA was digested with S1 nuclease (gift from D. Durnam). The cDNA was then dC-tailed with terminal deoxynucleotidyltransferase (Bethesda Research Laboratories), hybridized with *Pst* I-digested dG-tailed pBR322 (Bethesda Research Laboratories) (15), and used to transform $CaCl_2$-treated *E. coli* RR1.

cDNA Cloning with Synthetic Primers. cDNA was prepared as described above using SK-MEL 28 mRNA and synthetic oligonucleotide primers. The cDNA was dG-tailed, ligated with *Eco*RI-cut λgt10 (16) and an oligonucleotide (AATTC-CCCCCCCCCC) bridge, packaged (17), and plated on *E. coli* C_{600} rK^- mK^+ *hfl*.

RNA Blot Analysis. SK-MEL 28 mRNA was denatured, electrophoresed on a 0.7% agarose/2.2 M formaldehyde gel and transferred to nitrocellulose. The filters were probed with a nick-translated p973a2f1 cDNA insert.

Screening of cDNA Libraries. DNA from colonies of transformed bacteria was bound to paper (18) and screened by differential hybridization with cDNA probes synthesized on p97-enriched and unenriched mRNA templates. Libraries in λgt10 were screened for p97 inserts by plaque hybridization (19) with genomic exon fragments as probes. Probes were radiolabeled with [^{32}P]TTP (New England Nuclear; 3200 Ci/mmol; 1 Ci = 37 GBq) by nick-translation to a specific activity of 5–10 × 10^8 cpm/μg.

DNA Sequence Analysis. cDNA inserts were excised and subcloned into the plasmid vector pEMBL18$^+$ (20) for subsequent propagation and restriction mapping. cDNA was also subcloned into the M13mp18 phage cloning vector (21) and sequenced using the dideoxy chain-termination method

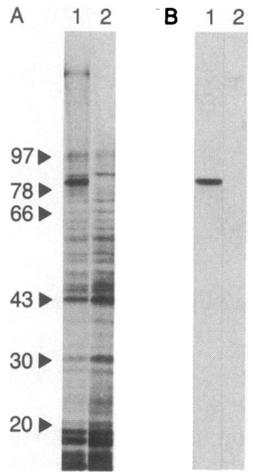

FIG. 1. Cell-free translation of p97 mRNA. Five nanograms of mRNA was translated in the reticulocyte lysate system, and the translation products were analyzed by $NaDodSO_4$/PAGE and autoradiography. (*A*) Total translation products (0.5 μl), 6-day exposure. (*B*) Translation products (5 μl) immunoprecipitated with anti-p97 serum, 1-day exposure. Lanes: 1, p97-enriched mRNA; 2, unenriched mRNA.

of Sanger *et al.* (22). M13 clones containing large inserts were sequenced by generating deletions using DNase I (23) or exonuclease III (24) and by using synthetic 21-mer oligonucleotide primers.

RESULTS AND DISCUSSION

Purification of p97 mRNA. Polysomes bearing p97 nascent chains were purified by incubation with three IgG2a monoclonal antibodies (96.5, 118.1, 133.2) specific for distinct epitopes of p97 (2, 4, 5, 10) followed by affinity chromatography on protein A-Sepharose. In a typical experiment, 150 A_{260} units of polysomes yielded 260 ng of p97-enriched mRNA, 0.23% of the total mRNA. When translated in *Xenopus* oocytes (25) and assayed for p97 as described (5, 10), p97-enriched mRNA yielded 80 pg of p97/ng of mRNA, whereas p97-unenriched mRNA yielded only 0.44 pg of p97/ng of mRNA, showing that p97 mRNA activity had been enriched 180-fold. The yield of p97 mRNA activity was 42%. Translation in the reticulocyte lysate system (26) showed that p97-enriched mRNA coded for a major polypeptide with an apparent M_r by $NaDodSO_4$/PAGE of 84,000, which was not detectable in the translation products of unenriched mRNA, and was immunoprecipitated by antiserum specific for p97 (Fig. 1). We concluded that it was the unglycosylated precursor of p97.

Isolation of cDNA Clones for Human p97. p97-enriched mRNA was used as template for oligo(dT)-primed cDNA synthesis. The cDNA was cloned in pBR322, and the resulting library was screened with cDNA probes. A 243-base-pair clone, p973a2f1, was identified that hybridized to p97-enriched cDNA but not detectably to unenriched cDNA and, in addition, selected p97 mRNA in hybrid-selection translation experiments (data not shown). A polyadenylylation signal (AATAA) and a poly(A) tract were present at the 3′ end of the cDNA. Nick-translated p973a2f1 hybridized 100-fold more strongly to p97-enriched mRNA than to unenriched melanoma mRNA and not detectably to fibroblast mRNA (data not shown). RNA blot analysis with the cloned cDNA as a probe identified an mRNA of approximately 4 kilobases that was present in SK-MEL 28 melanoma cells and absent from fibroblasts (data not shown).

Attempts to obtain cDNA clones extending more than 1 kilobase from the polyadenylylation site were unsuccessful, possibly due to a region of high G+C content (>80%) with extensive secondary structure. Genomic cloning was used to circumvent this problem. Four overlapping genomic clones were isolated from libraries of λ L47.1 containing size-fractionated SK-MEL 28 DNA enriched for a specific p97 restriction fragment (unpublished data). Restriction fragments that hybridized to the 4-kilobase p97 mRNA on RNA blots were sequenced and p97 exons were identified by a computer-assisted homology search between the predicted coding sequences and the amino acid sequence of human and chicken transferrin (11, 27, 28). Three synthetic oligonucleotides based on p97 genomic exon sequences were used to prime cDNA synthesis on SK-MEL 28 mRNA and the cDNA was cloned into λgt10. Three overlapping cDNA clones (10a1, 1j1, 2f1) spanning 2368 nucleotides of the p97 mRNA, including the entire coding region, were identified by using p97 exon-specific fragments as probes (Fig. 2).

Structure of p97. The p97 cDNA sequence is shown in Fig. 3. An open reading frame of 2214 nucleotides extends from the first ATG, around which the sequence conforms with the consensus initiation sequence determined by Kozak (29), to the TGA at position 2215. The most 5′ cDNA clone contains an additional 60 nucleotides upstream of the initiating ATG. The 3′ noncoding region of p97 mRNA, which was not obtained as a cDNA clone, was identified as a single genomic exon containing 1667 nucleotides (unpublished data). Residues 20–32 of the predicted amino acid sequence are identical

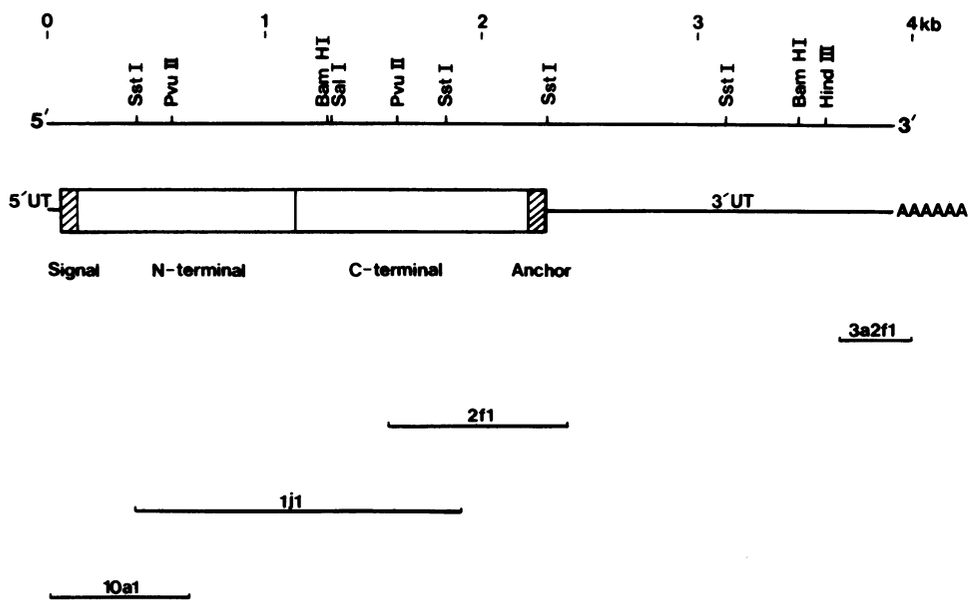

FIG. 2. Structure of p97 mRNA. The p97 cDNA clone p973a2f1 was isolated from an oligo(dT)-primed p97-enriched melanoma cDNA library in pBR-322, whereas cDNA clones p972f1, p971j1, and p9710a1 were isolated by priming with p97 exon-specific oligonucleotides and cloned in λgt10. The structure of the coding and noncoding regions of the mRNA are indicated as is the duplicated domain structure of the p97 precursor. kb, Kilobase(s).

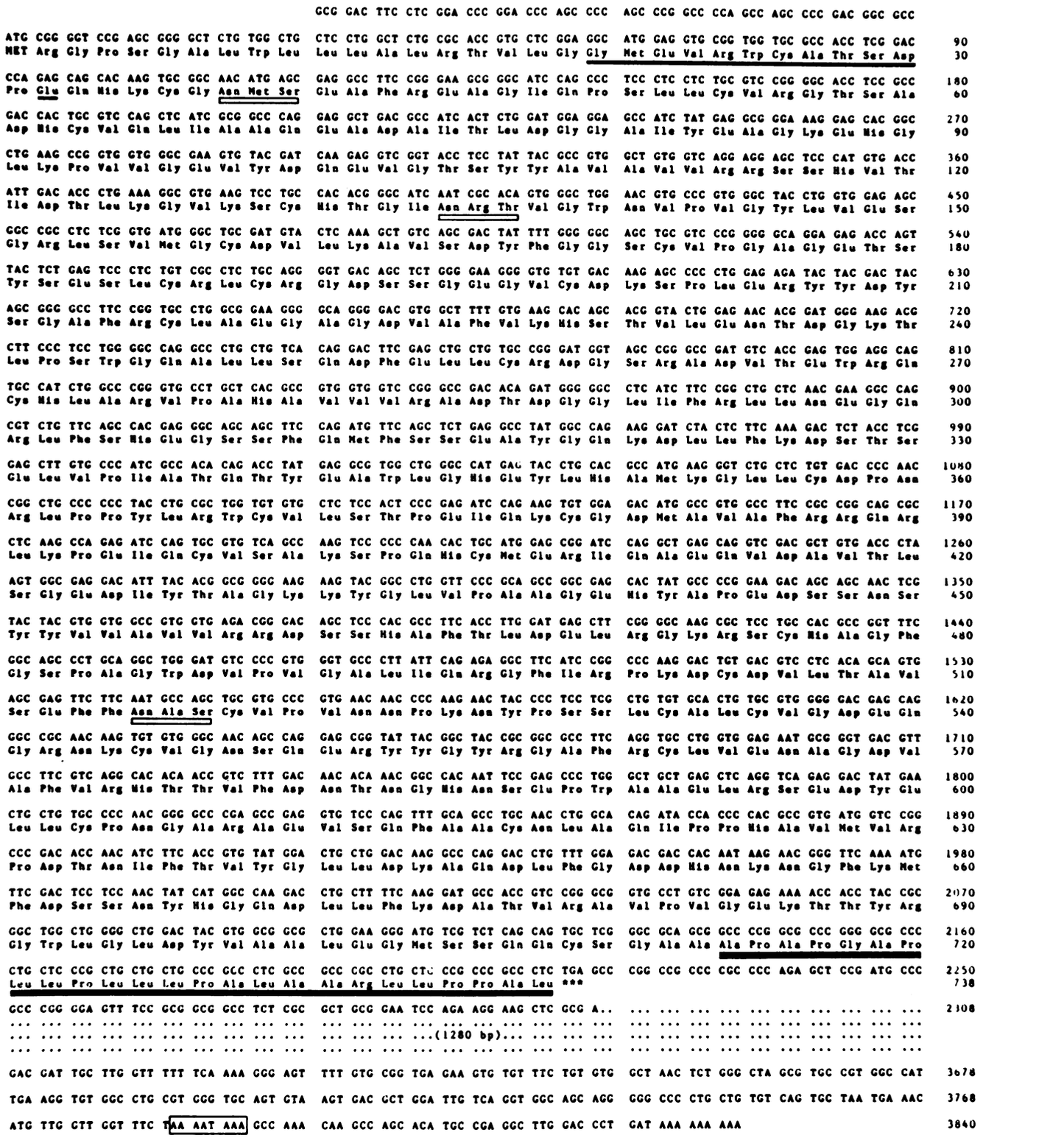

FIG. 3. Nucleotide sequence of the human p97 precursor cDNA and the deduced amino acid sequence. The N-terminal amino acid residues determined previously by protein sequencing, which are identical to those predicted from the nucleotide sequence, are underlined. The potential glycosylation sites at amino acids 38, 135, and 515 (open bars) and the membrane anchor region at the C terminus (solid bar) are indicated. One polyadenylylation signal AATAAA was detected at position 3785, 47 base pairs upstream of a polyadenylylated tract.

to the known N-terminal amino acid sequence of p97 (9), proving the identity of the cloned cDNA. Furthermore, the predicted molecular weight of the precursor is 80,196, in good agreement with the observed molecular weight of the *in vitro* translation product.

The amino acid sequence of the p97 precursor comprises four structural domains. Since residue 20 of the precursor sequence corresponds to the N terminus of mature p97, amino acids 1–19 constitute a signal peptide, a conclusion that is supported by its length and hydrophobic nature. Amino acids 20–361 and 362–713 comprise homologous domains of 342 and 352 amino acids. Potential N-glycosylation sites occur at positions 38 and 135 in the N-terminal domain and position 515 in the C-terminal domain. Finally,

we believe that amino acids 714–738, a region of predominantly uncharged and hydrophobic residues, anchor p97 in the cell membrane (30) and may extend into the cytoplasm.

The domain structure of p97 is supported by protease digestion experiments. Digestion of p97 with trypsin, papain (4), or thrombin produced a glycosylated antigenic fragment of $M_r \approx 40{,}000$. The fragment was purified from a thrombin digest of p97 that had been metabolically labeled with [^{35}S]methionine or [^{35}S]cysteine and sequenced as described (9). Cysteine residues were identified at positions 7 and 17, and methionine residues were identified at positions 2 and 20. Identical results were obtained with intact p97 and are in complete agreement with the N-terminal sequence of p97 predicted from the cDNA sequence. We conclude that the M_r 40,000 protease-resistant fragment corresponds to the N-terminal domain of p97. We have been unable to isolate the C-terminal domain of p97, possibly because it is protease sensitive.

Homology of p97 with Transferrin. A search of the amino acid sequence library of the Protein Identification Resource (release 5.0) (31) as described by Wilbur and Lipman (32) showed that p97 is strikingly homologous to three members of the transferrin superfamily; human serum transferrin, human lactotransferrin, and chicken transferrin (37%–39% homology; Fig. 4). Since human and chicken transferrin show 50% homology to each other, p97 must have diverged from serum transferrin more than 300 million years ago, when the mammalian and avian lineages diverged (34). Conservation of cysteine residues within the domains of p97 and between the transferrin family members is notable. p97 has 14 cysteine residues located in homologous positions in each domain. Human transferrin contains all of these cysteines in homologous positions in both domains, while human lactotransferrin and chicken transferrin lack only two of these cysteine residues (in their C-terminal domains). Unlike p97, these proteins contain 4–7 additional cysteines in their C-terminal domains, which have no corresponding member in the N-terminal domain. Human transferrin also contains 2 extra cysteines unique to its N-terminal domain. The positions of most of the disulfides in human serum transferrin, lactotransferrin, and chicken transferrin have been determined directly (27, 33, 35–38), and one can thus predict the positions of seven disulfide bonds in each domain of p97.

The amino acid homology between domains of p97 (46%—achieved by insertion of 7 gaps of 9 residues) is more striking than that seen in human transferrin (43%—16 gaps, 45 residues) or chicken transferrin (35%—12 gaps, 49 residues). Given the extensive sequence homology between p97 and transferrin and the apparently similar folding patterns, based on the conservation of cysteines, we believe that if the present low-resolution x-ray structure of transferrin (39) can be refined it may be possible to deduce the three-dimensional structure of p97.

Function of p97. Its membership in the transferrin superfamily, its ability to bind iron (9), and its common chromosomal localization with transferrin and the transferrin receptor (10, 11) all support a role for p97 in iron transport. The iron binding pocket of transferrin is thought to contain two or three tyrosines, one or two histidines and a single bicarbonate-binding arginine (33). Conservation of these amino acids in p97 support its proposed role in iron metabolism (Fig. 4). Since p97 is a membrane-bound transferrin-like molecule and has no homology with the transferrin receptor (40), its role in cellular iron metabolism may differ from that provided by circulating serum transferrin and the cellular transferrin receptor. Expression of the cloned p97 cDNA in eukaryotic cells will allow experimental testing of its functional properties. Because of its strong homology with members of the transferrin superfamily and because it was first isolated from melanoma cells, we propose renaming p97 melanotransferrin.

Obtaining a full-length cDNA for p97 has allowed a detailed analysis of the primary structure of a human membrane-bound tumor-associated antigen. p97 should therefore serve

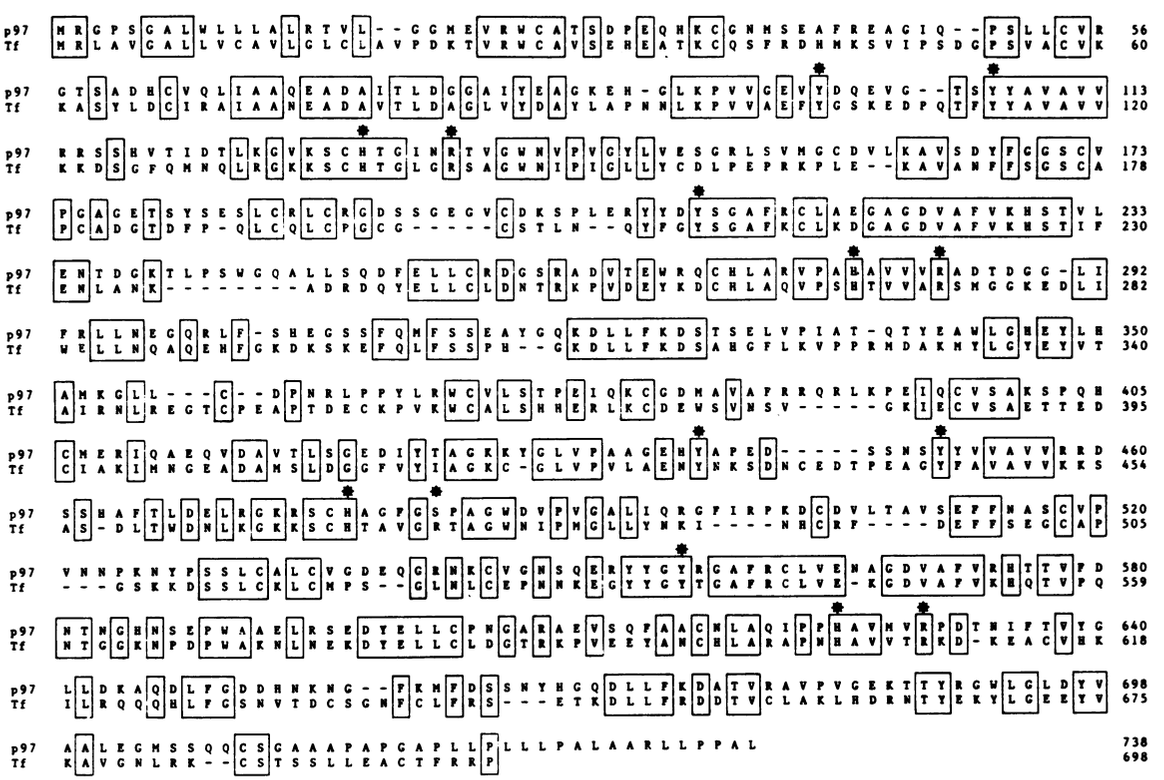

FIG. 4. Comparison of the predicted amino acid sequence of the p97 precursor and that of human serotransferrin (11, 30). Conserved residues are boxed. Tyrosine, histidine, and arginine residues implicated in iron binding of transferrin are indicated (33).

as a useful model for this class of proteins and it will be of interest to see which other tumor-associated differentiation antigens have structural or functional homology with known proteins.

We are grateful to Drs. Stanley McKnight, Richard Gelinas, Gregory Barsh, Steven Henikoff, and David Lee for their advice and encouragement and to Toni Monsey, Cynthia Green, and Jan Conitz for their excellent technical assistance. This investigation was supported by National Cancer Institute Grants CA 34777 and 38415 and by National Institutes of Health Grant GM 07266.

1. Woodbury, R. G., Brown, J. P., Yeh, M.-Y., Hellström, I. & Hellström, K. E. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2183–2187.
2. Brown, J. P., Wright, P. W., Hart, C. E., Woodbury, R. G., Hellström, K. E. & Hellström, I. (1980) *J. Biol. Chem.* **255**, 4980–4983.
3. Dippold, W. G., Lloyd, K. O., Li, L. T., Ikeda, H., Oettgen, H. F. & Old, L. J. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6114–6118.
4. Brown, J. P., Nishiyama, K., Hellström, I. & Hellström, K. E. (1981) *J. Immunol.* **127**, 539–546.
5. Brown, J. P., Woodbury, R. G., Hart, C. E., Hellström, I. & Hellström, K. E. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 539–543.
6. Garrigues, H. J., Tilgen, W., Hellström, I., Franke, W. & Hellström, K. E. (1982) *Int. J. Cancer* **29**, 511–515.
7. Larson, S. M., Carrasquillo, J. A., Krohn, K. A., Brown, J. P., McGuffin, R. W., Ferens, J. M., Graham, M. M., Hill, L. D., Beaumier, P. L., Hellström, K. E. & Hellström, I. (1983) *J. Clin. Invest.* **72**, 2101–2114.
8. Khosravi, M. J., Dent, P. B. & Liao, S.-K. (1985) *Int. J. Cancer* **35**, 73–80.
9. Brown, J. P., Hewick, R. M., Hellström, I., Hellström, K. E., Doolittle, R. F. & Dreyer, W. J. (1982) *Nature (London)* **296**, 171–173.
10. Plowman, G. D., Brown, J. P., Enns, C. A., Schroder, J., Nikinmaa, B., Sussman, H. H., Hellström, K. E. & Hellström, I. (1983) *Nature (London)* **303**, 70–72.
11. Yang, F., Lum, J. B., McGill, J. R., Moore, C. M., Naylor, S. L., van Bragt, P. H., Baldwin, W. D. & Bowman, B. H. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 2752–2756.
12. Carey, T. E., Takahashi, T., Resnick, L. A., Oettgen, H. F. & Old, L. J. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 3278–3282.
13. Palmiter, R. D. (1974) *Biochemistry* **13**, 3606–3615.
14. Korman, A. J., Knudsen, P. J., Kaufman, J. F. & Strominger, J. L. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1844–1848.
15. Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Nabaer, S. P., Chick, W. L. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3727–3731.
16. Davis, R. W., Botstein, D. & Roth, J. R. (1980) in *A Manual for Genetic Engineering: Advanced Bacterial Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
17. Grosveld, F. G., Dahl, H. M., de Boer, E. & Flavell, R. A. (1981) *Gene* **13**, 227–237.
18. Taub, F. & Thompson, B. (1982) *Anal. Biochem.* **126**, 222–230.
19. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
20. Dente, L., Cesareni, G. & Cortese, R. (1983) *Nucleic Acids Res.* **11**, 1645–1655.
21. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33**, 103–119.
22. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
23. Hong, G. F. (1982) *J. Mol. Biol.* **158**, 539–549.
24. Henikoff, S. (1984) *Gene* **28**, 351–359.
25. Gurdon, J. B. (1974) in *The Control of Gene Expression in Animal Development* (Harvard Univ. Press, Cambridge, MA).
26. Pelham, H. R. & Jackson, R. J. (1976) *Eur. J. Biochem.* **67**, 247–256.
27. MacGillivray, R. T. A., Mendez, E., Sinha, S. K., Sutton, M. R., Lineback-Zins, J. & Brew, K. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2504–2508.
28. Jeltsch, J. M. & Chambon, P. (1982) *Eur. J. Biochem.* **122**, 291–295.
29. Kozak, M. (1980) *Nucleic Acids Res.* **8**, 127–142.
30. Davis, N. G., Boeke, J. D. & Model, P. (1985) *J. Mol. Biol.* **181**, 111–121.
31. Dayhoff, M. O., Schwartz, R. M., Chen, H. R., Hunt, L. T., Barker, W. C. & Orcutt, B. C. (1981) *Nature (London)* **290**, 8.
32. Wilbur, W. J. & Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 726–730.
33. Metz-Boutigue, M.-H., Jolles, J., Mazurier, J., Schoentgen, F., Legrand, D., Spik, G., Montreuil, J. & Jolles, P. (1984) *Eur. J. Biochem.* **145**, 659–676.
34. Goodman, M., Moore, G. W. & Matsuda, G. (1975) *Nature (London)* **253**, 603–608.
35. Mazurier, J., Metz-Boutigue, M.-H., Jolles, J., Spik, G., Montreuil, J. & Jolles, P. (1983) *Experientia (Basel)* **39**, 135–141.
36. MacGillivray, R. T. A., Mendez, E., Shewale, J. G., Sinha, S. K., Lineback-Zins, J. & Brew, K. (1983) *J. Biol. Chem.* **258**, 3543–3553.
37. Williams, J., Elleman, T. C., Kingston, J. B., Wilkins, A. G. & Kuhn, K. A. (1982) *Eur. J. Biochem.* **122**, 297–303.
38. Williams, J. (1974) *Biochem. J.* **141**, 745–752.
39. Gorinsky, B., Horsburgh, C., Lindley, P. F., Moss, D. S., Parkar, M. & Watson, J. L. (1979) *Nature (London)* **281**, 157–158.
40. Schneider, C., Owen, M. J., Banville, D. & Williams, J. G. (1984) *Nature (London)* **311**, 675–678.