Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 7708–7712, October 1986
Biochemistry

Complete amino acid sequence of human thyroxine-binding globulin deduced from cloned DNA: Close homology to the serine antiproteases

(plasma proteins/thyroid hormone/structural homology/serpin)

IRWIN L. FLINK*, THOMAS J. BAILEY*, THOMAS A. GUSTAFSON*, BRUCE E. MARKHAM*, AND EUGENE MORKIN*†‡

Departments of *Internal Medicine and †Pharmacology, University of Arizona College of Medicine, Tucson, AZ 85724

Communicated by Herbert E. Carter, July 11, 1986

ABSTRACT **Antibodies directed against thyroxine-binding globulin (TBG) have been used to screen a human liver λgt11 expression library. A 1.46-kilobase clone was identified which encodes nearly the complete amino acid sequence, beginning at amino acid 17 of the mature protein. To complete the protein sequence, the cDNA clone was used to identify a genomic clone coding for TBG in a human X chromosome library. The overlapping recombinant clones contained an open reading frame coding for 415 amino acids followed by a polyadenylylation signal (AATAAA) located 275 nucleotides from a TAG termination codon. Beginning at residue 21, the deduced amino acid sequence agrees closely with the known NH_2-terminal sequence of the mature peptide. The preceding 20 amino acid residues are hydrophobic in character and presumably represent a leader sequence. Four glycosylation sites were identified, corresponding to the number determined for the purified protein. DNA blot hybridization revealed a single-copy gene, which by chromosomal analysis was found to be located on the long arm of the X chromosome. Unexpectedly, the nucleotide sequence of TBG is closely homologous to those encoding the plasma serine antiproteases α_1-antichymotrypsin and α_1-antitrypsin. However, there is little overall homology between TBG and transthyretin (prealbumin), the other major thyroxine-binding protein of human plasma.**

In most vertebrate species, thyroxine-binding globulin (TBG) has been reported to be the major thyroxine-binding protein (1). It is a glycoprotein (molecular mass about 54 kDa) that is synthesized in the liver as a polypeptide of 45 kDa (2). TBG represents a relatively rare translation product of total liver mRNA, accounting for only about 0.018% of total acid-precipitable radioactivity (2). Efforts to obtain detailed information about the structure of TBG and its thyroxine-binding site have been hampered by the presence of multiple electrophoretic forms and the inability to crystallize the protein (3, 4).

In this report, we describe the use of serum containing polyclonal antibodies directed against human TBG to isolate cDNA clones coding for TBG from a human liver λgt11 expression library (5). To complete the coding sequence, one of these clones was used to identify a genomic clone from a human X chromosome library. Residues 21–40 of the deduced amino acid sequence correspond closely to the NH_2-terminal amino acid sequence of mature TBG reported by Cheng (6). Amino acids 1–20 are very hydrophobic in nature and presumably represent the leader sequence of the propeptide. Unexpectedly, TBG shows a high degree of homology to α_1-antichymotrypsin and α_1-antitrypsin, and appears to belong to the serpin superfamily, most of which are serine proteinase inhibitors (7, 8).

METHODS

Antibody Screening of Recombinant Phage. A human liver cDNA library of $\approx 1.2 \times 10^6$ independent recombinants was constructed in λgt11 (9) and kindly provided by S. L. C. Woo (Department of Cell Biology, Baylor College of Medicine, Houston, TX). The library was screened with anti-TBG antibodies (Calbiochem) essentially as described by Davis and Young (5). Phage were adsorbed to *Escherichia coli* strain Y1090, plated at a density of 25,000 plaque-forming units per 130-mm Luria (L) plate and grown at 42°C for 4 hr. Nitrocellulose filters (Schleicher & Schuell, 0.45 μm) were soaked in 10 mm isopropyl β-D-thiogalactoside and air-dried at room temperature. Duplicate filters were overlaid on plaques and β-galactosidase fusion protein synthesis was induced for 2 hr at 37°C. Filters were keyed to plates and washed for 10 min with Tris-buffered saline (TBS: 50 mM Tris·HCl, pH 8.1/150 mM NaCl) and 30 min with 20% fetal bovine serum in TBS. Filters were incubated overnight at 5°C with antibody diluted 1:250 in TBS containing 20% fetal bovine serum. Unbound antibody was removed by washing the filters successively in TBS for 10 min, 0.05% Nonidet P-40 in TBS for 3 min, and finally TBS for 10 min. ^{125}I-labeled protein A (ICN, 31.9 μCi/μg; 1 Ci = 37 GBq) was diluted to 8×10^4 cpm/ml in TBS containing 20% fetal bovine serum. Duplicate filters were placed opposite to each other and incubated at room temperature for 1 hr with gentle shaking in a 150-mm glass Petri dish with 25 ml of radioactive protein A. Filters were washed three times as described above to remove unbound protein A and then were autoradiographed 24–72 hr with Kodak XR-5 film at −70°C with a Dupont Cronex Lightning Plus intensifying screen.

Isolation of Recombinant λ Bacteriophages Containing the TBG Gene. A human X chromosome library (10) was supplied by K. E. Davies (Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Oxford, UK). The library was screened by the plaque-hybridization technique (11) using λcTBG8 as a probe. The cDNA clone was labeled with [α-^{32}P]dCTP by nick-translation as described (12).

Isolation of Phage DNA. Candidate positive signals obtained in the first high-density screen were taken through four successive rounds of antibody screening at progressively

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.

Abbreviations: TBG, thyroxine-binding globulin; kb, kilobase(s). Standard one-letter amino acid abbreviations are used to specify sequences.

‡To whom reprint requests should be addressed.

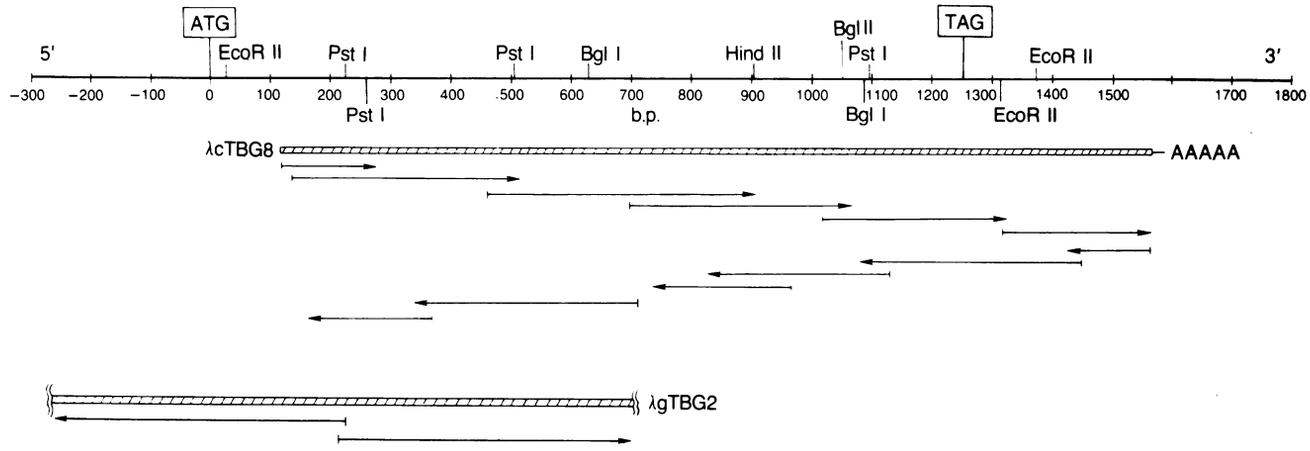

FIG. 1. Partial restriction map and sequencing strategy for cDNA clone λcTBG8 and for genomic clone λgTBG2, which was obtained from a human X chromosome library. ATG and TAG represent initiation and termination codons, respectively. Length is given in base pairs, and restriction sites are indicated. Arrows indicate sequences determined for fragments obtained using synthetic oligodeoxynucleotide sequences of the inserts themselves as primers.

lower plaque densities. The resulting, repeatedly positive, and well-isolated phage plaques were picked, amplified to yield high-titer plate stocks, and used for the large-scale isolation of phage. Lysogens of each positive clone were constructed according to Maniatis *et al.* (13) in *E. coli* strain BNN103 and grown to OD_{600} 0.5 at 30°C in L broth. The culture was induced by incubating for 15 min at 42°C with constant shaking. The induced culture was incubated at 38°C for an additional 1.5–3 hr with vigorous shaking. Bacteriophage λgt11 was purified by successive centrifugation on a 5–40% (vol/vol) glycerol step gradient followed by CsCl equilibrium centrifugation. The phage band was collected and dialyzed. DNA was extracted with phenol and ethanol-precipitated at −70°C.

Subcloning and Nucleotide Sequence Analysis. Purified phage DNA was digested with various restriction endonucleases, and the resultant fragments were subcloned into M13mp10 or M13mp11 (14) at appropriate restriction sites. After transformation, the single-stranded recombinant phages were used as templates for sequence determination by the dideoxy chain-termination method (15). In most experiments, specific sequencing primers were prepared from the determined sequence and used to analyze the adjacent regions. These primers, 17 nucleotides in length, were synthesized by an automatic DNA synthesizer (Applied Biosystems, Foster City, CA; model 380A) and purified by electrophoresis.

DNA Blot Analysis. Human placental DNA digested with restriction endonucleases *Eco*RI, *Hin*dIII, and *Hin*dIII/*Bam*HI was purchased from Oncor (Gaithersburg, MD). Samples of each digest, containing about 10 μg of DNA, were electrophoresed in 1% agarose gels and electroblotted onto Nytran paper (Schleicher & Schuell). Filter hybridization was carried out with ^{32}P-labeled insert DNA from clone cλTBG8 (16).

RESULTS

A human liver λgt11 cDNA library was screened with antibody against human TBG. Two clones expressed a hybrid protein that reacted with the antibody. The clone containing the longer DNA insert, ≈1.46 kilobases (kb) long and designated λcTBG8, was sequenced in its entirety according to the strategy shown in Fig. 1. The smaller clone was partially sequenced and found to represent a fragment of the larger clone. Both strands of the inserted DNA fragment of λcTBG8 were subjected to sequence analysis.

The overlapping recombinant clones provided a cDNA sequence consisting of 1432 nucleotides, and a stretch of 23 adenylate residues corresponding to the 3′ poly(A) terminus of the mRNA. This poly(A) tract is preceded by the hexanucleotide AATAAA (nucleotides 1854–1859) which is usually found in eukaryotic mRNAs (17). To obtain the additional nucleotide sequence corresponding to the NH_2-terminal region of the protein, the clone λcTBG8 was used as a probe to identify a 1.9-kb genomic clone coding for TBG in a human X chromosome library. The partial nucleotide sequence of this clone (λgTBG2), together with the cDNA sequence, is shown in Fig. 2.

We assume that the TBG translation initiation triplet is the ATG at nucleotides 331–333 because this ATG triplet is preceded almost immediately by a TAA stop codon at positions 316–318 without an intervening intron splice site (GT/AG) and occurs in an open reading frame that furnishes an amino acid sequence overlapping with that deduced from the cDNA insert. A possible transcription initiation site is located 235 bases upstream from this ATG and is preceded by sequences resembling consensus "TATA" and CAAT sequences (18) at positions 65–69 and 19–22, respectively. In the 3′ direction, the open reading frame continues for 414 codons before the termination triplet TAG is found at nucleotides 1576–1578.

As is a common feature in known secreted proteins (19), the NH_2-terminal region of the deduced TBG amino acid sequence is hydrophobic, and this region probably serves as a signal peptide (20) that is cleaved off in the secretion process of mature TBG. By comparison with the known NH_2-terminal amino acid sequence of TBG (6), the final protein secreted into the plasma would begin with the alanine residue (position 1) of the deduced sequence.

```
                    5         10        15        20
Deduced  A S P E G K V T A C H S S Q  P  N A  T  L Y/
Ref. 6   A S P E G K V T A D S S S Q (P) X A (S) L Y/
```

Sixteen of the 19 residues found experimentally agree with the deduced amino acid sequence. One of the 20 positions in the amino acid sequence could not be determined, and two positions were mistakenly thought to represent serine residues, probably because the correct amino acid was present in low yield. The amino acid sequence assignment for position 10 was reported to be aspartic acid, but our nucleotide sequence indicates that this should be cysteine. Comparison of the deduced amino acid composition (Table 1) of human

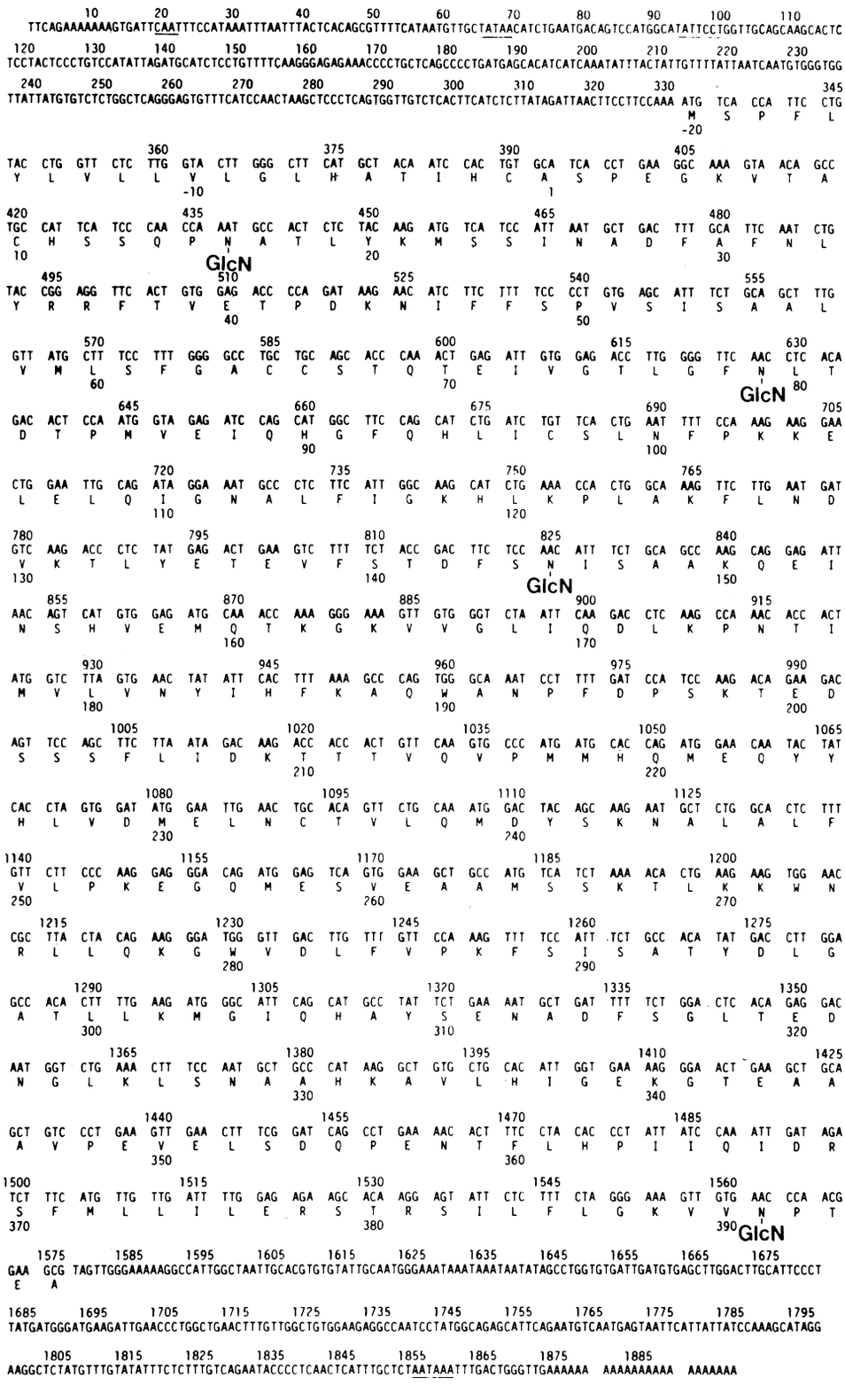

FIG. 2. Nucleotide sequence corresponding to cDNA for human TBG and the immediate upstream genomic sequence. The derived amino acid sequence is indicated in the single-letter code. The amino acids in the mature protein are numbered 1–395, and those in the putative signal peptide are numbered −20 to −1. The four glycosylation sites are located at positions 16, 79, 145, and 391 and are indicated by GlcN. In the 5′ untranslated region, sequences corresponding to consensus CAAT and TATA sequences (18), as well as a possible transcriptional start site, are underlined. In the 3′ untranslated region, a possible polyadenylylation signal sequence is underlined.

TBG also shows good agreement with that obtained by direct analysis of the purified protein (21).

If this proposal is correct, the mature human TBG contains 395 amino acids, which would give a molecular mass of 44,180 Da. This value is reasonably close to a recent estimate of 45 kDa for human TBG, which was obtained by $NaDodSO_4$/polyacrylamide gel electrophoresis of the purified protein (2).

Four potential N-linked glycosylation sites in TBG were identified at positions 16, 79, 145, and 391 of the amino acid sequence (Fig. 2). This number of sites agrees exactly with the value determined chemically for the protein (22).

Analysis of human placental DNA by hybridization with human TBG cDNA clone λcTBG8 under stringent conditions indicates that the TBG gene consists of a single copy (Fig. 3). Since we were able to isolate a positive genomic clone from an X chromosome library, there must be a single copy of the TBG gene located on the X chromosome. Preliminary data (J. Trent, personal communication) indicate that λcTBG8 hybridizes to the long arm of the X chromosome.

Table 1. Amino acid composition of human TBG

Amino acid	No. of residues	
	This work	Ref. 21
Aspartic acid*	38	38
Threonine	27	26
Serine	32	31
Glutamic acid*	44	44
Proline	17	16
Glycine	17	20
Alanine	29	30
Valine	26	29
Methionine	14	13
Isoleucine	22	19
Leucine	43	40
Tyrosine	9	10
Phenylalanine	23	23
Histidine	12	12
Lysine	28	30
Arginine	6	6
Cysteine	5	5
Tryptophan	3	3
Total	395	395

Results from Gershengorn *et al.* (21) have been normalized to 395 residues.

*Asparagine and glutamine values from the deduced sequence are included with aspartic acid and glutamic acid, respectively.

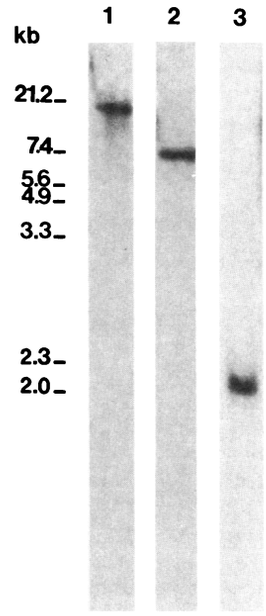

FIG. 3. DNA blot hybridization analyses with TBG cDNA clone λcTBG8 and human placental DNA. DNA was cut with *Eco*RI (lane 1), *Bam*HI (lane 2), or *Bam*HI/*Hin*dIII (lane 3) prior to electrophoresis, transfer, and hybridization. The migration distances of unlabeled DNA markers on the same gel are indicated at left.

DISCUSSION

Recombinant cDNA clones coding for TBG have been isolated from a human liver expression library by screening with antibody against TBG. Comparison with the known NH_2-terminal sequence of human TBG revealed that the cDNA clones code for residues 17–395 of the mature protein. The additional sequence information was obtained from a genomic clone isolated from an X chromosome library.

Comparison of the nucleotide and deduced amino acid sequences of human TBG with those for human transthyretin (prealbumin) (23) showed little overall homology. This is somewhat surprising, since both proteins exhibit high-affinity binding of thyroxine. However, a search of the nucleotide sequences contained in the BIONET GenBank DNA database (release no. 39) revealed an unexpectedly close homology of TBG with human α_1-antichymotrypsin (58%) (24) and α_1-antitrypsin (53%) (25) and a more distant homology with antithrombin III (27%) (26). The amino acid sequence comparisons among these proteins also support the hypothesis that they share a common genetic ancestry. Thus, the amino acid sequence of TBG is about 37% homologous with both α_1-antitrypsin and α_1-antichymotrypsin, which is similar to the level of homology between the two antiproteases themselves (42%).

The target protease for TBG, if it exists, is unknown. However, the reactive centers of the antiprotease family usually can be identified by examining the aligned sequences. In the sequence alignment displayed in Fig. 4, the reactive-site residues Met-Ser in α_1-antitrypsin and the Leu-Ser residues in α_1-antichymotrypsin (indicated by arrow) may correspond to the Leu-Ser residues at positions 352–353 of the TBG sequence. For α_1-antitrypsin and α_1-antichymotrypsin, the mechanism of inhibition has been shown to involve complex formation with the protease, accompanied by proteolytic attack on the Met-Ser or Leu-Ser bonds present at their active sites (27).

The inhibitory specificity of the antiproteases is primarily defined by a single amino acid at the reactive center of the molecule (8). This residue acts as a "bait" for the appropriate serine proteinase—that is, methionine (or valine) for elastase, leucine for chymotrypsin, and arginine for thrombin. On this basis, the antiprotease activity of TBG, if any exists, would be antichymotryptic in type.

With respect to this possible antiprotease activity, it is interesting that carboxypeptidase digestion of purified TBG has been reported to release leucine and alanine in order of decreasing yields (21), suggesting that these amino acids might represent the COOH-terminal and penultimate residues of the protein, respectively. Although the deduced amino acid sequence shows alanine to be the COOH-terminal residue, leucine would be produced as a second COOH terminus if proteolytic cleavage occurred at the Leu-Ser bond. Hydrolysis of the latter bond might be expected if TBG, acting as an antiprotease, had encountered its target enzyme prior to isolation. In other antiproteases, the cleaved peptide fragment tends to remain tightly bound to the surface of the protein, even in the presence of denaturing agents (27), which may have led to detection of both COOH-terminal residues.

The crystallographic structure of the closely related antiprotease α_1-antitrypsin has been determined (28). The polypeptide chain is arranged almost exclusively in well-defined secondary structural elements, including one large and two smaller regions of β-sheet. Parts of two of these sheets are arranged to form a barrel located at one end of the elliptical molecule. It is clear from comparison of the amino acid sequences of TBG and α_1-antitrypsin that residues buried in the three-dimensional structure that are particularly important for the integrity of the molecule, such as those comprising the large β-sheet, are highly conserved. If the overall structures of TBG and α_1-antitrypsin are similar, the barrel structure may represent the thyroxine-binding site. In transthyretin, the other major thyroxine-binding protein of plasma, the four subunits form two β-barrels, each of which serves as a thyroxine-binding site (29).

TBG deficiency in humans was described by Tanaka and Starr (30), who noted a complete absence of TBG binding of radiolabeled thyroxine in the plasma electrophoretogram. Subsequently, careful study of a number of families with absent or low TBG have provided evidence for X-linked inheritance (31). In the present work, we were able to identify a clone from an X chromosome library that contains sequences coding for TBG. Furthermore, preliminary data indicate that λcTBG8 hybridizes to the long arm of the X chromosome. Genes coding for the other proteinase inhibitors are believed to be located on autosomal chromosomes.

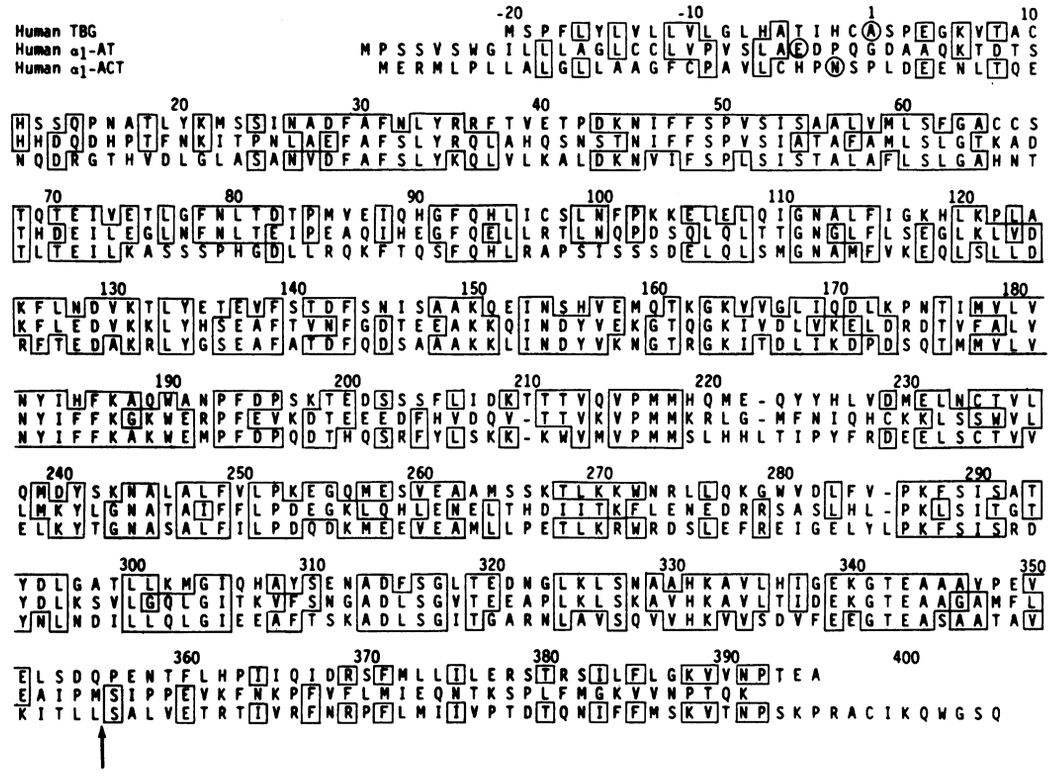

FIG. 4. Comparison of the amino acid sequences of human TBG, α_1-antitrypsin (α_1-AT), and α_1-antichymotrypsin (α_1-ACT). Positions −20 to −1 refer to the amino acid residues of the TBG propeptide. Positions 1–395 represent the mature protein. Arrow indicates the cleaved peptide bond at the target sites of α_1-AT and α_1-ACT. In TBG, Leu-Ser at positions 352–353 also may represent a target site.

The genes for α_1-antitrypsin and α_1-antichymotrypsin have been localized to chromosome 14q near the immunoglobulin γ heavy chain locus (32, 33), whereas the gene for antithrombin III maps to a distinct locus at 1q23 (34).

Although deficiency of α_1-antitrypsin may lead to chronic obstructive pulmonary disease, neonatal hepatitis, and juvenile and adult cirrhosis, individuals with TBG deficiency are euthyroid and exhibit no known endocrine abnormalities (35, 36). The availability of a human cDNA probe for TBG should help to further clarify the significance of TBG deficiency states in humans.

This investigation was supported by research Grants 5 P01 HL20984-09 and R01 HL35751-01 from the National Institutes of Health and by the Gustavus and Louise Pfeiffer Research Foundation. Computer resources used in these studies were provided by the BIONET National Computer Resource for Molecular Biology, whose funding is provided by the Biomedical Research Technology Program, Division of Research Resources, National Institutes of Health (IU42 RR-01685-02).

1. Refetoff, S., Robin, N. I. & Fang, V. C. (1970) *Endocrinology* **86**, 793–805.
2. Bartalena, L., Tera, J. R. & Robbins, J. (1984) *J. Biol. Chem.* **259**, 13605–13609.
3. Marshall, J. S., Pensky, J. & Williams, S. (1973) *Arch. Biochem. Biophys.* **156**, 456–462.
4. Gartner, R., Henze, R., Horn, K., Pickardt, C. R. & Scriba, P. C. (1981) *J. Clin. Endocrinol. Metab.* **52**, 657–664.
5. Davis, R. W. & Young, R. A. (1983) *Science* **222**, 778–782.
6. Cheng, S.-Y. (1977) *Biochem. Biophys. Res. Commun.* **79**, 1212–1218.
7. Hunt, L. T. & Dayhoff, M. O. (1980) *Biochem. Biophys. Res. Commun.* **95**, 864–871.
8. Travis, J. & Salvesen, G. S. (1983) *Annu. Rev. Biochem.* **52**, 655–709.
9. Chandra, T., Stackhouse, R., Kidd, V. J. & Woo, S. L. C. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1845–1848.
10. Davies, K. E., Young, B. D., Elles, R. G., Hills, M. E. & Williamson, R. (1981) *Nature (London)* **293**, 374–376.
11. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
12. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
13. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
14. Messing, J., Crea, R. & Seeburg, P. H. (1981) *Nucleic Acids Res.* **9**, 309–321.
15. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
16. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
17. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214.
18. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349–383.
19. Kreil, G. (1981) *Annu. Rev. Biochem.* **50**, 317–348.
20. Silhavy, T. J., Benson, S. A. & Emr, S. D. (1983) *Microbiol. Rev.* **47**, 313–344.
21. Gershengorn, M. C., Cheng, S.-Y., Lippoldi, R. E., Lord, R. S. & Robbins, J. (1977) *J. Biol. Chem.* **252**, 8713–8718.
22. Zinn, A. B., Marshall, J. S. & Carlson, D. M. (1978) *J. Biol. Chem.* **253**, 6761–6767.
23. Kanda, Y., Goodman, D., Canfield, R. & Morgan, F. (1974) *J. Biol. Chem.* **249**, 6768–6805.
24. Chandra, T., Stackhouse, R., Kidd, V. J., Robson, K. J. H. & Woo, S. L. C. (1983) *Biochemistry* **22**, 5055–5061.
25. Colau, B., Chuchana, P. & Bollen, A. (1983) *DNA* **3**, 327–330.
26. Bock, S. C., Wion, K. L., Vehar, G. A. & Lawn, R. M. (1982) *Nucleic Acids Res.* **10**, 8113–8125.
27. Mori, M. & Travis, J. (1983) *J. Biol. Chem.* **258**, 12749–12752.
28. Loebermann, H., Tokuoka, R., Deisenhofer, J. & Huber, R. (1984) *J. Mol. Biol.* **177**, 531–556.
29. Blake, C. C. F., Geisow, M. J., Swan, I. D. A., Rerat, C. & Rerat, B. J. (1974) *J. Mol. Biol.* **88**, 1–12.
30. Tanaka, S. & Starr, P. (1959) *J. Clin. Endocrinol. Metab.* **19**, 84–91.
31. Refetoff, S., Robin, N. I. & Alper, C. A. (1972) *J. Clin. Invest.* **51**, 848–867.
32. Cox, D. W., Markovie, V. D. & Teshima, I. E. (1982) *Nature (London)* **297**, 428–430.
33. Rabin, M., Watson, M., Breg, W. R., Kidd, V., Woo, S. L. C. & Ruddle, F. H. (1985) *Cytogenet. Cell Genet.* **40**, 728 (abstr.).
34. Bock, S. C., Harris, J. F., Skolnick, M. H., Schwartz, C., Balazs, I., Olson, S. & Trent, J. M. (1984) *Cytogenet. Cell Genet.* **37**, 423 (abstr.).
35. Refetoff, S. (1979) in *Endocrinology*, ed. DeGroot, L. J. (Grune & Stratton, New York), Vol. 1, pp. 347–356.
36. Burr, W. A., Ramsden, D. B. & Hoffenberg, R. (1980) *Q. J. Med.* **49**, 295–313.