Volume 215, number 1, 100–104 FEB 04635 May 1987

The cDNA-deduced primary structure of human sex hormone-binding globulin and location of its steroid-binding domain

G.L. Hammond, D.A. Underhill, C.L. Smith, I.S. Goping, M.J. Harley, N.A. Musto*, C.Y. Cheng* and C.W. Bardin*

Department of Obstetrics and Gynecology, University of Western Ontario, 375 South Street, London, Ontario N6A 4G5, Canada and **The Population Council, 1230 York Avenue, New York, NY 10021, USA*

Received 23 February 1987

We have sequenced a cDNA for sex hormone-binding globulin (SHBG) isolated from a phage λgt11 human liver cDNA library. The library was screened with a radiolabeled rat androgen-binding protein (ABP) cDNA, and the abundance of SHBG cDNAs was 1 in 750 000 plaques examined. The largest human SHBG cDNA (1194 base-pairs) contained a reading frame for 381 amino acids. This comprised 8 amino acids of a signal peptide followed by 373 residues starting with the known NH_2-terminal sequence of human SHBG, and ending with a termination codon. The predicted polypeptide M_r of SHBG is 40 509, and sites of attachment of one *O*-linked (residue 7) and two *N*-linked oligosaccharide (residues 351 and 367) chains were identified. Purified SHBG was photoaffinity-labeled with Δ^6-[^{3}H]testosterone and cleaved with trypsin. The labeled tryptic fragment was isolated by reverse-phase HPLC, and its NH_2-terminal sequence was determined. The results suggest that a portion of the steroid-binding domain of SHBG is located between residue 296 and the 35 predominantly hydrophilic residues at the C-terminus of the protein.

cDNA cloning; Amino acid sequence; Sex hormone-binding globulin; Steroid-binding domain

1. INTRODUCTION

SHBG is a glycoprotein in human serum that binds testosterone and estradiol with high affinity and is thought to regulate their entry into cells [1].

Correspondence address: G.L. Hammond, Department of Obstetrics and Gynecology, University of Western Ontario, 375 South Street, London, Ontario N6A 4G5, Canada

Abbreviations: SHBG, sex hormone-binding globulin, also known as testosterone-estradiol-binding globulin (TeBG) and sex-steroid-binding protein (SBP); ABP, androgen-binding protein; Δ^6-testosterone, 17β-hydroxyandrosta-4,6-dien-3-one; SSPE, 150 mM NaCl/10 mM $NaH_2PO_4H_2O$/1 mM EDTA (pH 7.4); SSC, 150 mM NaCl/15 mM sodium acetate (pH 7.4); Denhardt's solution, 0.02% Ficoll/0.02% polyvinylpyrrolidone/0.02% bovine serum albumin (Pentax, fraction V) in H_2O (w/v)

Human SHBG has been purified to apparent homogeneity, but analyses of its amino acid composition and carbohydrate content have produced conflicting data [2–7]. In its native form, SHBG is composed of two identical subunits which appear to associate to form a single steroid-binding site [2,6–8]. The molecular composition of the SHBG steroid-binding domain has not, however, been determined; despite the fact that it can be photoaffinity-labeled with Δ^6-[^{3}H]testosterone [9].

It is known that SHBG is produced by hepatocytes [10], and its synthesis is thought to be stimulated by thyroid hormone [10] and estrogens [11], and decreased by androgens [11]. The Sertoli cells of the testis also produce a protein that is physicochemically and immunochemically closely related to plasma SHBG [12], and is generally referred to as testicular androgen-binding protein (ABP). The production of ABP by Sertoli cells is

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

induced by testosterone and follicle-stimulating hormone [13], and its synthesis therefore appears to be regulated differently from that of SHBG. This may be attributed to the presence of different genes for these two proteins, but may also be the result of cell-specific differences in the expression of a single gene. In order to begin to address these questions we have isolated a cDNA for human SHBG by taking advantage of its ability to hybridize with a rat ABP cDNA isolated from a testicular expression library [14]. This has enabled us to predict its amino acid sequence and to locate a region of the molecule that appears to be involved in steroid binding.

2. MATERIALS AND METHODS

2.1. *cDNA cloning*

A cDNA for rat ABP [14] was used as a probe to screen a λgt11 human liver cDNA library that was kindly provided by S.L.C. Woo (Baylor College of Medicine, Houston, TX). The rat ABP cDNA comprised two *Eco*RI fragments that had been subcloned into the *Eco*RI site of pBR322 [14,15]. The library was screened using *E. coli* (strain Y1088) as the host bacteria at a density of 3×10^4 plaques/150 mm culture plate. Nitrocellulose filters were used to transfer DNA, and were prehybridized at 37°C in a buffer (5 × SSPE, 5 × Denhardt's solution, 0.1% SDS) containing 40% deionized formamide and 100 μg denatured salmon sperm DNA/ml. Both rat ABP cDNA *Eco*RI fragments were nick-translated in the presence of ^{32}P-labeled dCTP [15], and the resultant probes were mixed and hybridized with the filters for 16 h at 37°C in prehybridization buffer (2×10^6 cpm/ml). The filters were washed 3 times at room temperature in 2 × SSC/0.1% SDS, once for 30 min in 1 × SSC/0.1% SDS at 37°C, and twice for 30 min in 0.2 × SSC/0.1% SDS at 37°C. Autoradiography was performed overnight at −80°C using XAR film (Eastman Kodak, Rochester, NY) and Dupont Cronex^R HI-Plus XH intensifying screens (DuPont Canada, Markham, Ontario).

2.2. *DNA sequencing*

When the largest SHBG cDNA was excised from the λgt11 cloning vector two *Eco*RI fragments were obtained and were subcloned into the *Eco*RI site of pBR322 for amplification. After excision from the plasmid with *Eco*RI, both fragments were isolated by electroelution from a preparative polyacrylamide gel [15], and digested with appropriate restriction enzymes (fig.1). The resultant fragments were subcloned into M13mp18 and M13mp19 bacteriophage vectors to produce single-stranded templates for sequencing by the dideoxy-chain termination method [16].

2.3. *Photolabeling of SHBG and preparation of tryptic fragments*

Human SHBG was purified [9,17] and photolabeled with Δ^6-[^{3}H]testosterone [9], and the product was analysed by SDS-PAGE and fluorography [9]. About 200 μg photolabeled SHBG was equilibrated with trypsin buffer (50 mM Tris, 30 mM KCl, 2 mM $MgCl_2$, pH 7.5, at 22°C), 20 μg *N*-tosyl-L-phenylalanine chloromethyl-treated trypsin was added, and digestion was performed at 37°C for 24 h. Preliminary experiments indicated that digestion was completed using an enzyme:substrate ratio of 1:10 within 16 h under these conditions. The tryptic digest was fractionated by HPLC as follows: the sample was lyophilized, resuspended in 0.1% trifluoroacetic acid, and injected onto a Vydac C-18 reverse-phase HPLC column (4.6 × 250 mm (i.d.), particle size 5 μm). The initial solvent (A) consisted of 5% acetonitrile and 95% water containing 0.1% trifluoroacetic acid (v/v), and the second solvent (B) was 95% acetonitrile and 5% water containing 0.1% trifluoroacetic acid (v/v). Elution was performed using a linear gradient from 0 to 80% solvent B at a flow rate of 1 ml/min over 90 min, and the effluent was monitored by UV absorbance at 206 nm. An aliquot (5 μl) of each fraction (0.25 ml) collected from the HPLC column was counted, and the radiolabeled tryptic fragment was detected in a fraction that eluted in the presence of 41% acetonitrile.

2.4. *NH_2-terminal sequence analysis*

The fraction containing the photolabeled tryptic fragment was reconstituted in 0.1% SDS, and the volume was reduced to 100 μl under nitrogen. The sample was subjected to Edman degradation in the presence of Polybrene [18] in a gas-phase sequenator (Applied Biosystems, Foster City, CA)

equipped with automatic converting flask and microprocessor. Phenylthiohydantoin-amino acids were determined by reverse-phase HPLC.

3. RESULTS AND DISCUSSION

As a result of screening 1.5×10^6 individual plaques from a male human liver cDNA library with a rat ABP cDNA probe, we were only able to isolate two SHBG cDNAs. Their abundance therefore appears to be very low when compared to the number of cDNAs for corticosteroid-binding globulin (CBG) we have isolated from the same library [14]. This difference may reflect differences in the relative amounts of mRNAs for these proteins in the liver, and contribute to the 10–50-fold higher plasma concentrations of CBG [19] when compared to SHBG [20,21].

The largest of the two SHBG cDNAs isolated contained two *Eco*RI fragments (0.55 and 0.64 kb in length), while the other only contained a single 0.55 kb insert. The 0.55 kb fragments from both clones hybridize with each other, and represent the 3′-end of the SHBG mRNA (fig.1). Both the 0.64 and 0.55 kb *Eco*RI fragments of the SHBG cDNA we have sequenced (fig.1) hybridize with the corresponding *Eco*RI fragments of rat ABP cDNA [14]. In fact, when the rat ABP [22] and human SHBG cDNA sequences are compared there is only a 23% mismatch in nucleotide sequences over a 1153 nucleotide overlap, and 68% identity with respect to the amino acid sequences of the mature proteins (not shown).

The human SHBG cDNA we have sequenced is not full length, but it does contain the entire coding region for the mature SHBG peptide, as well as a portion of the leader sequence (fig.2). This was ascertained by comparing the predicted amino acid sequence from the cDNA with the known NH_2-terminal sequence of human SHBG [6] underlined in fig.2. The cDNA-deduced amino acid sequence also agrees with the sequence recently obtained by direct chemical methods [23], and confirms the presence of a Leu at the NH_2-terminus of the protein [4,6]. Residue no.7 from the amino-terminus appears to be the site of attachment of an oligosaccharide [6,23], and the cDNA confirms the presence of a threonine in this position [23]. It is therefore likely that this is the site of attachment of the single *O*-linked oligosaccharide chain predicted for each SHBG subunit [24]. The amino acid sequence of SHBG we have deduced from its cDNA also contains two Asn-X-Ser/Thr (where X is not a Pro) consensus sequences [25] for the attachment of *N*-linked oligosaccharide chains (fig.2), and these both appear to be utilized [23,24].

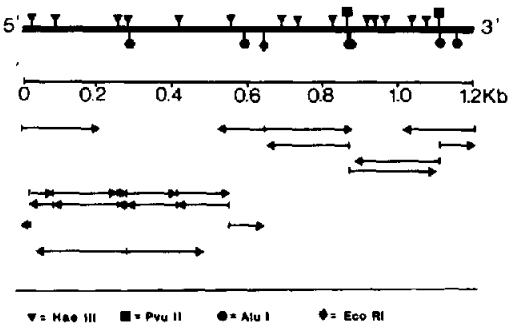

Fig.1. Partial restriction map of human liver SHBG cDNA and a diagrammatic summary of the sequencing strategy. Restriction fragments were subcloned into M13mp18 and M13mp19, and single-stranded templates were sequenced by the dideoxy chain termination method [16] in the directions shown by the arrows.

Although the 3′-end of the human SHBG cDNA ends with a poly(A)-rich region, it does not appear to represent the poly(A) tail of the mRNA; (i) because it is followed by a single G that is not part of the *Eco*RI linker, and (ii) because it is not preceded by the usual eucaryotic polyadenylation signals (AATAAA or AAGAAA). We therefore assume that this poly(A)-rich region was used as the site for oligo(dT)-primed reverse transcription of the SHBG mRNA, which most probably contains a longer 3′-untranslated region than is represented by the cDNA.

The use of Δ^6-[^{3}H]testosterone enabled us to photolabel an amino acid within the steroid-binding site of SHBG, and to isolate a tryptic fragment generated by cleavage at residue 295 (Lys). The NH_2-terminal sequence of the photolabeled fragment includes the only other lysine before the C-terminus of the molecule, but there are two arginine residues that may be targets for trypsin at positions 343 and 352. Preliminary analyses of the secondary structure of SHBG, however, indicate that both arginines are located within a predominantly hydrophilic region of the molecule; extending from residue 338 to the C-terminus of

```
-8 Gln His Thr Arg Gln Gly Trp Ala Leu Arg Pro Val Leu Pro Thr Gln Ser Ala His Asp  12
 1 CAG CAC ACC CGC CAG GGA TGG GCC CTG AGA CCT GTT CTC CCC ACC CAG AGT GCC CAC GAC

   Pro Pro Ala Val His Leu Ser Asn Gly Pro Gly Gln Glu Pro Ile Ala Val Met Thr Phe  32
61 CCT CCG GCT GTC CAC CTC AGC AAT GGC CCA GGA CAA GAG CCT ATC GCT GTC ATG ACC TTT

   Asp Leu Thr Lys Ile Thr Lys Thr Ser Ser Ser Phe Glu Val Arg Thr Trp Asp Pro Glu  52
121 GAC CTC ACC AAG ATC ACA AAA ACC TCC TCC TCC TTT GAG GTT CGA ACC TGG GAC CCA GAG

   Gly Val Ile Phe Tyr Gly Asp Thr Asn Pro Lys Asp Asp Trp Phe Met Leu Gly Leu Arg  72
181 GGA GTG ATT TTT TAT GGG GAT ACC AAC CCT AAG GAT GAC TGG TTT ATG CTG GGA CTT CGA

   Asp Gly Arg Pro Glu Ile Gln Leu His Asn His Trp Ala Gln Leu Thr Val Gly Ala Gly  92
241 GAC GGC AGG CCT GAG ATC CAA CTG CAC AAT CAC TGG GCC CAG CTT ACG GTG GGT GCT GGA

   Pro Arg Leu Asp Asp Gly Arg Trp His Gln Val Glu Val Lys Met Glu Gly Asp Ser Val 112
301 CCA CGG CTG GAT GAT GGG AGA TGG CAC CAG GTG GAA GTC AAG ATG GAG GGG GAC TCT GTG

   Leu Leu Glu Val Asp Gly Glu Glu Val Leu Arg Leu Arg Gln Val Ser Gly Pro Leu Thr 132
361 CTG CTG GAG GTG GAT GGG GAG GAG GTG CTG CGC CTG AGA CAG GTC TCT GGG CCC CTG ACC

   Ser Lys Arg His Pro Ile Met Arg Ile Ala Leu Gly Gly Leu Leu Phe Pro Ala Ser Asn 152
421 AGC AAA CGC CAT CCC ATC ATG AGG ATT GCG CTT GGG GGG CTG CTC TTC CCC GCT TCC AAC

   Leu Arg Leu Pro Leu Val Pro Ala Leu Asp Gly Cys Leu Arg Arg Asp Ser Trp Leu Asp 172
481 CTT CGG TTG CCG CTG GTT CCT GCC CTG GAT GGC TGC CTG CGC CGG GAT TCC TGG CTG GAC

   Lys Gln Ala Glu Ile Ser Ala Ser Ala Pro Thr Ser Leu Arg Ser Cys Asp Val Glu Ser 192
541 AAA CAG GCC GAG ATC TCA GCA TCT GCC CCC ACT AGC CTC AGA AGC TGT GAT GTA GAA TCA

   Asn Pro Gly Ile Phe Leu Pro Pro Gly Thr Gln Ala Glu Phe Asn Leu Arg Asp Ile Pro 212
601 AAT CCC GGG ATA TTT CTC CCT CCA GGG ACT CAG GCA GAA TTC AAT CTC CGA GAC ATT CCC

   Gln Pro His Ala Glu Pro Trp Ala Phe Ser Leu Asp Leu Gly Leu Lys Gln Ala Ala Gly 232
661 CAG CCT CAT GCA GAG CCC TGG GCC TTC TCT TTG GAC CTG GGA CTC AAG CAG GCA GCA GGC

   Ser Gly His Leu Leu Ala Leu Gly Thr Pro Glu Asn Pro Ser Trp Leu Ser Leu His Leu 252
721 TCA GGC CAC CTC CTT GCT CTT GGG ACA CCA GAG AAC CCA TCT TGG CTC AGT CTC CAC CTC

   Gln Asp Gln Lys Val Val Phe Ser Ser Gly Ser Gly Pro Gly Leu Asp Leu Pro Leu Val 272
781 CAA GAT CAA AAG GTG GTG TTC TCT TCT GGG TCG GGG CCA GGG CTG GAT CTG CCC CTG GTC

   Leu Gly Leu Pro Leu Gln Leu Lys Leu Ser Met Ser Arg Val Val Leu Ser Gln Gly Ser 292
841 TTG GGA CTC CCT CTT CAG CTG AAG CTG AGT ATG TCC AGG GTG GTC TTG AGC CAA GGG TCG

   Lys Met Lys Ala Leu Ala Leu Pro Pro Leu Gly Leu Ala Pro Leu Leu Asn Leu Trp Ala 312
901 AAG ATG AAG GCC CTT GCC CTG CCT CCC TTA GGC CTG GCT CCC CTC CTT AAC CTC TGG GCC

   Lys Pro Gln Gly Arg Leu Phe Leu Gly Ala Leu Pro Gly Glu Asp Ser Ser Thr Ser Phe 332
961 AAG CCT CAA GGG CGT CTC TTC CTG GGG GCT TTA CCA GGA GAA GAC TCT TCC ACC TCT TTT

   Cys Leu Asn Gly Leu Trp Ala Gln Gly Gln Arg Leu Asp Val Asp Gln Ala Leu Asn Arg 352
1021 TGC CTG AAT GGC CTT TGG GCA CAA GGT CAG AGG CTG GAT GTG GAC CAG GCC CTG AAC AGA

   Ser His Glu Ile Trp Thr His Ser Cys Pro Gln Ser Pro Gly Asn Gly Thr Asp Ala Ser 372
1081 AGC CAT GAG ATC TGG ACT CAC AGC TGC CCC CAG AGC CCA GGC AAT GGC ACT GAC GCT TCC

   His ---
1141 CAT TAA AGCTCCACCTAAGAACCCCCTTTGAAAGTTAAAAAAAAAAAAAAAG
```

Fig.2. The amino acid sequence of human SHBG deduced from the sequence of its cDNA. The numbering of amino acids is based on the known NH_2-terminal sequence [6], and this is underlined (solid). The potential positions of one *O*-linked carbohydrate (○) and two *N*-linked carbohydrate chains (●) are identified. The dashed underline indicates the NH_2-terminal amino acid sequence of the tryptic fragment that was photolabeled with Δ^6-[^{3}H]testosterone. Gaps in this sequence represent amino acids for which an ambiguous signal was obtained.

the protein and including two possible *N*-linked carbohydrate chains at residues 351 and 367. Although this is obviously an exposed surface region of the molecule, the presence of two carbohydrate chains may protect the arginines from trypsin, and the labeled fragment could include the C-terminus of the protein. Nevertheless, because the steroid-binding site is likely to be located within a hydrophobic pocket, the photolabeled amino acid is probably not located in this hydrophilic C-terminal region. Further studies are required to identify the amino acid that is photolabeled with Δ^6-[^{3}H]testosterone, and to define the boundaries of the steroid binding site, but it is anticipated that the availability of a cDNA will enable us to address these and other questions related to the structure and function of this protein.

ACKNOWLEDGEMENTS

This work was supported by grants from the Medical Research Council of Canada (MA-9567), The Melon Foundation and the National Institutes of Health (HD 13541). We also wish to thank Bobbie Lucas for typing the manuscript. G.L.H. is the recipient of an Ontario Cancer Research and Treatment Foundation Research Scholarship.

REFERENCES

[1] Siiteri, P.K., Murai, J.T., Hammond, G.L., Nisker, J.A., Raymoure, W.J. and Kuhn, R.W. (1982) Recent Prog. Horm. Res. 38, 457–510.
[2] Rosner, W. and Smith, R.N. (1975) Biochemistry 14, 4813–4820.
[3] Petra, P.H. (1979) J. Steroid Biochem. 11, 245–252.
[4] Fernlund, P. and Laurell, C.-B. (1981) J. Steroid Biochem. 14, 545–552.
[5] Iqbal, M.J. and Johnson, M.W. (1979) J. Steroid Biochem. 10, 535–540.
[6] Hammond, G.L., Robinson, P.A., Sugino, H., Ward, D.N. and Finne, J. (1986) J. Steroid Biochem. 24, 815–824.
[7] Petra, P.H., Stanczyk, F.Z., Senear, D.F., Namkung, P.C., Novy, M.J., Ross, J.B.A., Turner, E. and Brown, J.A. (1983) J. Steroid Biochem. 19, 699–706.
[8] Petra, P.H., Kumar, S., Hayes, R., Ericsson, L.H. and Titani, K. (1986) J. Steroid Biochem. 24, 45–49.
[9] Cheng, C.Y., Musto, N.A., Gunsalus, G.L. and Bardin, C.W. (1983) J. Steroid Biochem. 19, 1379–1389.
[10] Rosner, W., Aden, D.P. and Khan, M.S. (1984) J. Clin. Endocrinol. Metab. 59, 806–808.
[11] Anderson, D.C. (1974) Clin. Endocrinol. 3, 69–96.
[12] Cheng, C.Y., Frick, J., Gunsalus, G.L., Musto, N.A. and Bardin, C.W. (1984) Endocrinology 114, 1395–1401.
[13] Tindall, D.J. and Means, A.R. (1980) in: Advances in Steroid Hormone Research, vol.IV (Thomas, J.A. and Singhal, R.L. eds) pp.295–327, Urban and Schwarzenberg, Baltimore.
[14] Hammond, G.L., Reventos, J., Musto, N.A., Gunsalus, G.L. and Bardin, C.W. (1986) in: Gene Regulation of Steroid Hormones III (Roy, A. and Clark, J. eds) pp.155–165, Springer, New York.
[15] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1983) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.
[16] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[17] Cheng, C.Y., Bardin, C.W., Nagendranath, N., Ecobar, N., Han, A.C., Musto, N.A. and Gunsalus, G.L. (1985) Int. J. Androl. 8, 1–12.
[18] Hunkapiller, M.W., Hewick, R.M., Dreyer, W.J. and Hood, L.E. (1983) Methods Enzymol. 91, 399–413.
[19] Robinson, P.A., Langley, M.S. and Hammond, G.L. (1985) J. Endocrinol. 104, 259–267.
[20] Hammond, G.L., Langley, M.S. and Robinson, P.A. (1985) J. Steroid Biochem. 23, 451–460.
[21] Cheng, C.Y., Bardin, C.W., Musto, N.A., Gunsalus, G.L., Cheng, S.L. and Ganguly, M. (1983) J. Clin. Endocrinol. Metab. 56, 68–75.
[22] Joseph, D.R., Hall, S.H. and French, F.S. (1987) Proc. Natl. Acad. Sci. USA 84, 339–343.
[23] Walsh, K.A., Titani, K., Kumar, S., Hayes, R. and Petra, P.H. (1986) Biochemistry 25, 7584–7590.
[24] Avvakumov, G.V., Matveentseva, I.V., Akhrem, L.V., Strel'chyonok, O.A. and Akrem, A.A. (1983) Biochim. Biophys. Acta 760, 104–110.
[25] Bause, E. (1983) Biochem. J. 209, 331–336.