Volume 219, number 2, 405–409 FEB 04949 July 1987

Characterization of a cDNA coding for sex steroid-binding protein of human plasma

Benito G. Que and Philip H. Petra

Departments of Ob/Gyn and Biochemistry, University of Washington, Seattle, WA 98195, USA

Received 5 May 1987; revised version received 30 May 1987

A cDNA (912 nucleotides) coding for human plasma sex steroid-binding protein (SBP) was characterized from a phage clone previously isolated by screening a Charon 21A human liver cDNA library with rat androgen binding protein (ABP) cDNA. The deduced amino acid sequence from the cDNA indicated that the insert was a partial clone coding for 281 amino acids starting with residue 92 (glycine) encompassing the alternating leucyl residues and the carboxyl-end 373 (histidine) as previously reported [(1986) Biochemistry 25, 7584]. The potential polyadenylation signal sequence ATTAAA is present as part of the 3'-coding region and the stop codon TAA. Both are followed by a short 20 untranslated nucleotides and a poly(A) tract of 49 nucleotides. Significant homologous sequences (76%) at the DNA level exist between human SBP and rat ABP which might suggest the possibility that both evolved from a common primordial gene. Demonstration of the presence of an SBP cDNA in a human liver cDNA library provides the first evidence that liver is the site of SBP biosynthesis.

Steroid-binding protein; cDNA; Amino acid sequence; (Human plasma)

1. INTRODUCTION

Human plasma sex steroid-binding protein (hSBP) is a steroid hormone transporting protein with high affinity for testosterone and estradiol [1,2]. It regulates the metabolic clearance rate of testosterone in plasma [3,4] and may also be involved in specific transport of the steroid into target cells via receptor-mediated endocytosis [5]. Human SBP is a dimeric glycoprotein which binds one molecule of steroid and is composed of two identical subunits [6,7]. The monomer contains 373 amino acids, 2 disulfide bonds, and 3 oligosaccharide chains [8]. The molecular mass of the dimer based on the amino acid sequence and a 14% carbohydrate content is 93400 Da [9]. The site of biosynthesis remains unknown although liver has been suggested as a possibility from studies on Hep-G2 cells [10].

Recently, a related protein, the androgen-binding protein (ABP) of rat testis, was cloned and the cDNA was sequenced [11,12]. Its deduced amino acid sequence was found to be 68% homologous to that of hSBP [13]. A positive phage clone was isolated from a human liver Charon 21A cDNA library using rat ABP cDNA as probe (Yarbrough, D. and Joseph, D., unpublished). In an attempt to gain insight into the site of biosynthesis, regulation, gene structure, and the steroid-binding site of hSBP, we have isolated and characterized the cDNA insert from that clone and we report here that it codes for hSBP.

Correspondence address: P.H. Petra, Departments of Ob/Gyn and Biochemistry, University of Washington, Seattle, WA 98195, USA

Published by Elsevier Science Publishers B.V. (Biomedical Division)

2. MATERIALS AND METHODS

The phage clone was obtained from Drs Yarbrough and Joseph. Restriction endonucleases, M13mp18 and M13mp19 RF vectors, Klenow fragment and T_4 ligase were purchased from BRL. Dideoxysequencing kits were purchased from New England Biolab. Radiochemical ^{32}P-nucleotides and ^{35}S-dATP were bought from New England Nuclear. High titer stock of pure Charon 21A recombinant positive phage was prepared as described [14] except that the host cell used was *E. coli* LE392. Phage DNA was prepared according to published procedure [15], and M13-dideoxy sequencing experiments were as described [14,16].

3. RESULTS

One positive clone coding for human sex steroid-binding protein (SBP) was isolated and plaque purified. As shown in fig.1, the recombinant phage released three fragments upon partial *Eco*RI digestion having sizes of about 910, 570 and 340 base pairs. These were subcloned into phosphatase-treated M13mp18 or M13mp19 RF vectors previously digested with *Eco*RI. As shown in fig.2, the cDNA sequence contains a coding region (843 nucleotides) followed by a stop codon (TAA), a non-coding region of 20 nucleotides and a poly(A) tail of 49 nucleotides. The internal *Eco*RI site is located at the position corresponding to amino acid residues Glu-205–Phe-206 (boxed in fig.2). The poly(A) tail is preceded by the putative polyadenylation signal sequence ATTAAA (boxed in fig.2) which is part of the 3′-coding region and the termination codon. This is analogous to cDNA coding for human blood clotting factor X in which ATTAAA is present in the 3′-coding region [17] and the cDNA coding for the β-subunit of human chorionic gonadotropin in which the stop codon (TAA) is present in the ATTAAA sequence [18]. The amino acid sequence deduced from the cDNA was in complete agreement with the published sequence [8]. The cDNA clone coded from the residue 92 glycine and extended through the carboxyl-end 373 histidine. As shown in fig.3, the corresponding amino acid was about 67% sequence identity with rat ABP amino acid sequence [13] while at the nucleotide level, it achieved higher homology (76%) due to several single base substitutions (silent mutation).

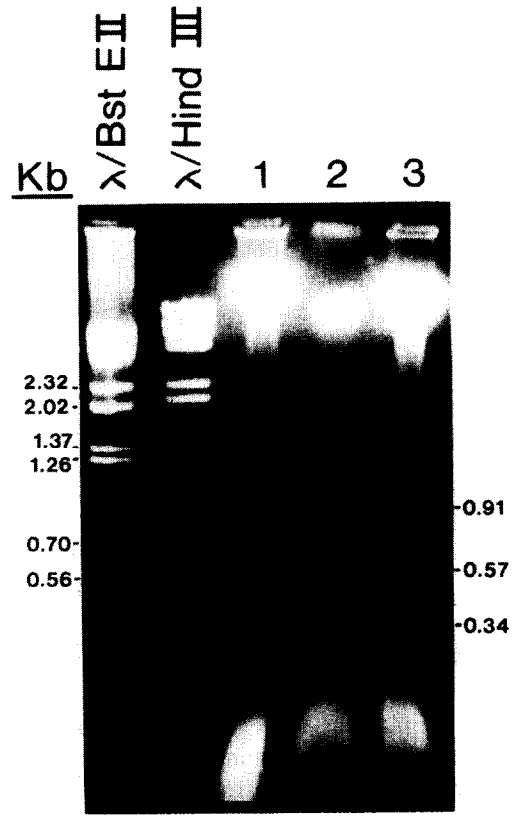

Fig.1. Analyses of recombinant phage DNA in 0.8% agarose containing ethidium bromide with corresponding size markers (λDNA digested with *Bst*EII or *Hin*dIII). Lanes: 1, undigested phage clone DNA; 2, complete *Eco*RI digestion of 10 μg of recombinant λDNA; 3, partial *Eco*RI digestion of 10 μg of recombinant λDNA.

Fig.2. Nucleotide and amino acid sequences for human sex steroid-binding protein. Solid triangles represent glycosylation sites. The first box represents the internal *Eco*RI site; the second, the alternating leucine sequence; and the third, the polyadenylation signal ATTAAA.

```
     92                                                  100
     G   P   R   L   D   D   G   R   W   H   Q   V   E   V
 CG GGA CCA CGG CTG GAT GAT GGG AGA TGG CAC CAG GTG GAA GTC
                110                                     120
 K   M   E   G   D   S   V   L   L   E   V   D   G   E   E
AAG ATG GAG GGG GAC TCT GTG CTG CTG GAG GTG GAT GGG GAG GAG
                                    130
 V   L   R   L   R   Q   V   S   G   P   L   T   S   K   R
GTG CTG CGC CTG AGA CAG GTC TCT GGG CCC CTG ACC AGC AAA CGC
                140                                     150
 H   P   I   M   R   I   A   L   G   G   L   L   F   P   A
CAT CCC ATC ATG AGG ATT GCG CTT GGG GGG CTG CTC TTC CCC GCT
                                    160
 S   N   L   R   L   P   L   V   P   A   L   D   G   C   L
TCC AAC CTT CGG TTG CCG CTG GTT CCT GCC CTG GAT GGC TGC CTG
                170                                     180
 R   R   D   S   W   L   D   K   Q   A   E   I   S   A   S
CGC CGG GAT TCC TGG CTG GAC AAA CAG GCC GAG ATC TCA GCA TCT
                                    190
 A   P   T   S   L   R   S   C   D   V   E   S   N   P   G
GCC CCC ACT AGC CTC AGA AGC TGT GAT GTA GAA TCA AAT CCC GGG
                200                                     210
 I   F   L   P   P   G   T   Q   A   E   F   N   L   R   D
ATA TTT CTC CCT CCA GGG ACT CAG GCA GAA TTC AAT CTC CGA GAC
                                    220
 I   P   Q   P   H   A   E   P   W   A   F   S   L   D   L
ATT CCC CAG CCT CAT GCA GAG CCC TGG GCC TTC TCT TTG GAC CTG
                230                                     240
 G   L   K   Q   A   A   G   S   G   H   L   L   A   L   G
GGA CTC AAG CAG GCA GCA GGC TCA GGC CAC CTC CTT GCT CTT GGG
                                    250
 T   P   E   N   P   S   W   L   S   L   H   L   Q   D   Q
ACA CCA GAG AAC CCA TCT TGG CTC AGT CTC CAC CTC CAA GAT CAA
                260                                     270
 K   V   V   L   S   S   G   S   G   P   G   L   D   L   P
AAG GTG GTG TTG TCT TCT GGG TCG GGG CCA GGG CTG GAT CTT CCC
                                    280
 L   V   L   G   L   P   L   Q   L   K   L   S   M   S   R
CTG GTC TTG GGA CTC CCT CTT CAA CTG AAG CTG AGT ATG TCC AGG
                290                                     300
 V   V   L   S   Q   G   S   K   M   K   A   L   A   L   P
GTG GTC TTG AGC CAA GGG TCG AAG ATG AAG GCC TTG GCC TTG CCT
                                    310
 P   L   G   L   A   P   L   L   N   L   W   A   K   P   Q
CCT TTA GGC TTG GCT CCC CTC CTT AAC CTC TGG GCC AAG CCT CAA
                320                                     330
 G   R   L   F   L   G   A   L   P   G   E   D   S   S   T
GGG CGT CTC TTC CTT GGG GCT TTA CCA GGA GAA GAC TCT TCC ACC
                                    340
 S   F   C   L   N   G   L   W   A   Q   G   Q   R   L   D
TCT TTT TGC CTG AAT GGC CTT TGG GCA CAA GGT CAG AGG TTG GAT
                350 ▲                                   360
 V   D   Q   A   L   N   R   S   H   E   I   W   T   H   S
GTG GAC CAG GCC TTG AAC AGA AGC CAT GAG ATC TGG ACT CAC AGC
                  ▲                 370
 C   P   Q   S   P   G   N   G   T   D   A   S   H  ***
TGC CCC CAG AGC CCA GGC AAT GGC ACT GAC GCT TCC CAT TAA AGC

TCCACCTTAGAACCCCA49
```

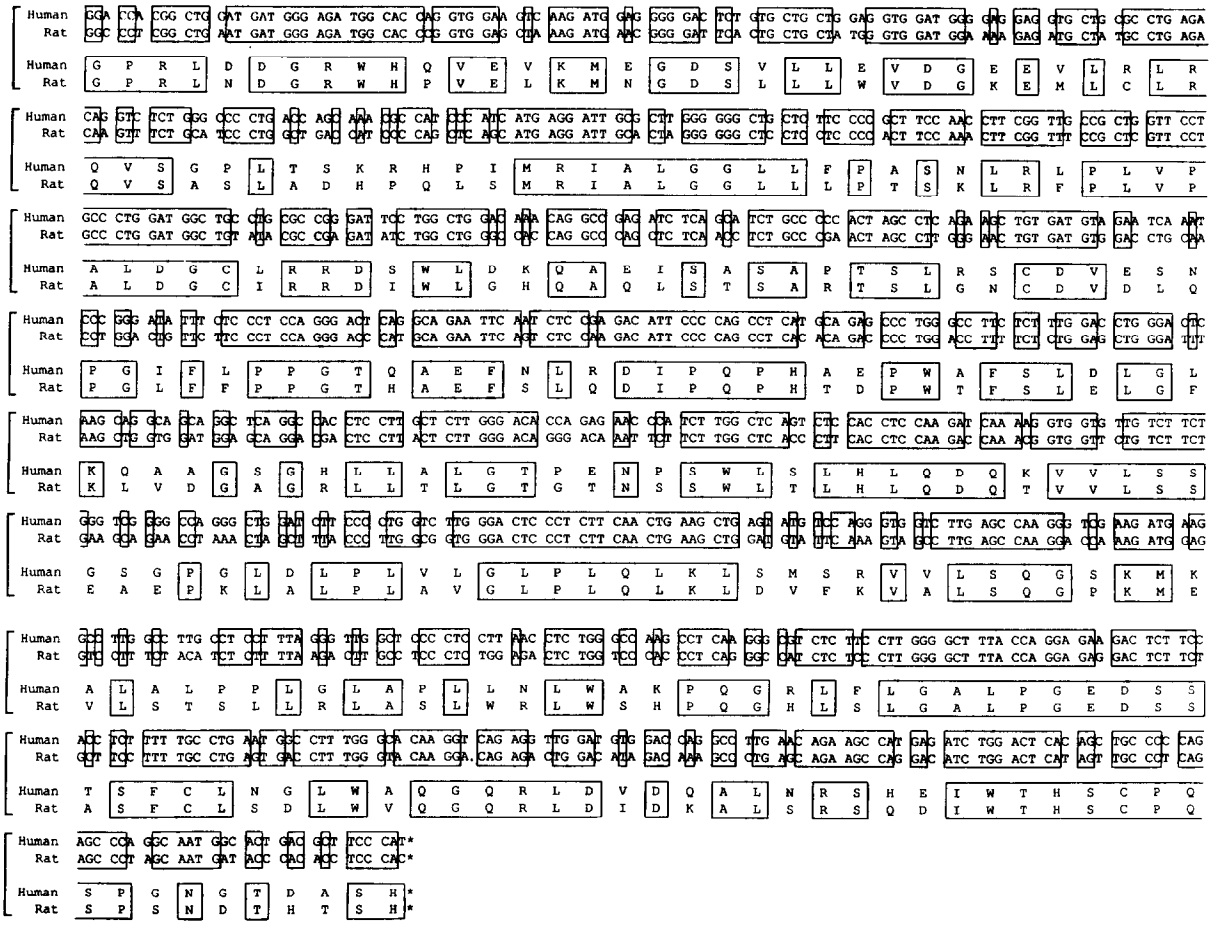

Fig.3. Comparison of the nucleotide and amino acid sequences between hSBP and rat ABP beginning at Gly-92 [8,12,13].

4. DISCUSSION

In this paper we report the first characterization of a cDNA coding for hSBP. Since it was isolated from a liver cDNA library, the results provide the first experimental evidence that liver is the site of SBP biosynthesis. Furthermore, we have recently detected the presence of mRNA for SBP in human liver by RNA dot blot and Northern blot analyses (unpublished). The results, however, do not necessarily mean that the liver is the only site of biosynthesis.

The cDNA clone codes for 75% of the hSBP amino acid sequence and is in complete agreement with it. The previously reported 68% homology at the amino acid level between rat ABP and hSBP [13] translates to a 76% homology at the nucleotide level due to the presence of silent mutations. The data strongly suggest that the two proteins have arisen from a common ancestral gene. Similarity between those two proteins emphasizes the importance of determining the extent of homology between hABP and hSBP which we have postulated to be very similar if not identical [9]. The information will be important for defining their relationship and provides answers as to why they exist in different compartments. We have recently found that the hSBP sequence is homologous to the carboxyl-terminal domain (residues 249–635) of human protein S [19], a vitamin K-dependent cofactor [20] involved in blood coagulation (manuscript in preparation).

Although much has been learned in recent years about the chemical structure of hSBP, its physiological role remains unclear. The definition of that role is now likely through the availability of cDNA probes which will allow cloning of the gene and description of its genomic structure and regulation. As a result it is hoped that new approaches toward solution of clinical disorders associated with abnormal levels of plasma SBP will be found. Furthermore, site-directed mutagenesis of the cDNA for SBP will permit the identification of amino acid residues involved in steroid binding. For instance, the cDNA we describe in this report contains the coding region for the alternating leucine sequence located at positions 267–281 (boxed in fig.2) which we believe can participate in steroid binding [8,9]. That hypothesis can now be tested by expressing the wild type as well as the mutated clones. In addition, since the clone does not have the coding sequence containing the only tyrosine residue in the molecule (Tyr-57), we will be able to test whether or not that residue is involved in steroid binding.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Drs D. Yarbrough and D. Joseph for providing the phage clone. This research was supported by NIH grant no.HD13956.

REFERENCES

[1] Mercier, A., Alfsen and Baulieu, E. (1965) Proc. 2nd Symp. Steroid Hormones, Ghent, Belgium, Excerpta Medica Int. Congr. Ser. no.101, p.212.
[2] Rosenbaum, W., Christy, N.P. and Kelly, W.G. (1966) J. Clin. Endocr. Metab. 26, 1399–1403.
[3] Vermeulen, A.L., Verdonck, L., Van der Straeten, M. and Orie, M. (1969) J. Clin. Endocr. Metab. 29, 1470–1480.
[4] Petra, P.H., Stanczyk, F.Z., Namkung, P.C., Fritz, M.A. and Novy, M.J. (1985) J. Steroid Biochem. 22, 739–746.
[5] Bordin, S. and Petra, P.H. (1980) Proc. Natl. Acad. Sci. USA 77, 5678–5682.
[6] Petra, P.H. (1979) J. Steroid Biochem. 11, 245–252.
[7] Petra, P.H., Kumar, S., Hayes, R., Ericsson, L.H. and Titani, K. (1986) J. Steroid Biochem. 24, 45–49.
[8] Walsh, K.A., Titani, K., Kumar, S., Hayes, R. and Petra, P.H. (1986) Biochemistry 25, 7584–7590.
[9] Petra, P.H., Namkung, P.C., Titani, K. and Walsh, K.A. (1986) in: Binding Proteins of Steroid Hormones (Forest, M.G. and Pugeat, M. eds) vol.149, pp.15–30, John Libbey, London.
[10] Khan, M.S., Knowles, B.B., Aden, D.P. and Rosner, W. (1981) J. Clin. Endocr. Metab. 53, 448–449.
[11] Joseph, D.R., Hall, S.H. and French, F.S. (1985) J. Androl. 6, 392–395.
[12] Joseph, D.R., Hall, S.H. and French, F.S. (1986) in: Binding Proteins of Steroid Hormones (Forest, M.G. and Pugeat, M. eds) vol.149, pp.123–135, John Libbey, London.
[13] Petra, P.H., Titani, K., Walsh, K.A., Joseph, D.R., Hall, S.H. and French, F.S. (1986) in: Binding Proteins of Steroid Hormones (Forest, M.G. and Pugeat, M. eds) vol.149, pp.137–142, John Libbey, London.
[14] Que, B.G. and Davie, E.W. (1986) Biochemistry 25, 1525–1528.
[15] Kaslow, D.C. (1986) Nucleic Acids Res. 14, 6767.
[16] Que, B.G. and Petra, P.H. (1986) Biochem. Biophys. Res. Commun. 137, 620–625.
[17] Leytus, S.P., Chung, D.W., Kisiel, W., Kurachi, K. and Davie, E.W. (1984) Proc. Natl. Acad. Sci. USA 81, 3699–3702.
[18] Fiddes, J.C. and Goodman, H.M. (1980) Nature 286, 684–687.
[19] Hoskin, J., Norman, D.K., Beckman, R.J. and Long, G.L. (1987) Proc. Natl. Acad. Sci. USA 84, 349–353.
[20] DiScipio, R.G. and Davie, E.W. (1979) Biochemistry 18, 899–904.