Amino Acid Sequence of the Sex Steroid Binding Protein of Human Blood Plasma†

Kenneth A. Walsh,*,‡ Koiti Titani,‡ Koji Takio,§ Santosh Kumar, Rutherford Hayes,⊥ and Philip H. Petra‡,⊥

Department of Biochemistry, Department of Obstetrics/Gynecology, and Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195

Received April 24, 1986; Revised Manuscript Received June 23, 1986

ABSTRACT: The amino acid sequence of the sex steroid binding protein (SBP) from human plasma has been determined. The SBP subunit consists of a 373-residue polypeptide chain containing two disulfide bonds and three oligosaccharide chains. The sequence was solved primarily by analysis of peptides derived by cleavage at either lysyl or methionyl residues. In our preparations, approximately half of the protein molecules have the amino-terminal sequence Arg-Pro-Val-Leu-Pro; the other half lack Arg-Pro and begin with the valine. Preparations of Hammond et al. [Hammond, G. L., Robinson, P. A., Sugino, H., Ward, D. N., & Finne, J. (1986) *J. Steroid Biochem. 24*, 815] have an additional leucine at the amino terminus, making a total of 373 residues in the chain. Oligosaccharide chains are placed at Thr-7 and at Asn residues 351 and 367. The two disulfide bonds connect Cys-164 to Cys-188 and Cys-333 to Cys-361. The reported heterogeneity of preparations of the molecule may result in part from the amino-terminal microheterogeneity, in part from variations in the oligosaccharide moieties, and possibly in part from rearrangements involving cyclic imide formation in two Asn-Gly sequences. Certain hydrophobic segments are suggested as possible components of the steroid-binding sites. The protein shows no homology either with the cDNA-derived sequences of the estrogen and glucocorticoid receptors found by others to be homologous with each other or with any other protein sequence in the 1986 data base.

It is now accepted that the early steps in the expression of steroid hormonal activity in mammalian cells involve the participation of unique steroid-binding proteins. For a number of years we have been developing a protein model for the sex steroid binding protein of plasma (SBP),[1] in the hope of elucidating the molecular nature of the steroid-binding process (Petra, 1979). The method of purification, chemical characterization, and details of the molecular organization of the human, monkey, baboon, and rabbit proteins have been reported (Petra & Lewis, 1980; Petra et al., 1983, 1986a,b). In each case the native protein is a dimeric structure with apparently identical subunits, each dimer binding only one molecule of steroid (Turner et al., 1984). Two possible molecular models accounting for this stoichiometry have been proposed (Petra et al., 1983; Namkung & Petra, 1986).

In order to understand the molecular nature of the steroid-binding process, both the primary and tertiary structures of SBP need to be available. Herein we describe the determination of the amino acid sequence of human SBP.

Materials and Methods

Human SBP was prepared as described previously (Petra & Lewis, 1980), and reduced and S-carboxymethylated as described by Takio et al. (1983).

Fragments were generated by cleavage with cyanogen bromide as described by Petra et al. (1986a). Peptides M1, M4, and M6 in the present report correspond to fractions V, III, and IV, respectively, of Petra et al. (1986a). Peptides M2 and M3 were isolated from their fraction VII by HPLC on an Altex Ultrapore RPSC column with an acetonitrile gradient. The isolation of Cys-containing peptides T9, T18, and M6-T23 (CBIV-T23) are also detailed in that report.

S-(Carboxymethyl)-SBP was digested at glutamyl (and occasional aspartyl) residues by protease V8 from *Staphylococcus aureus* (Miles) in 0.1 M NH_4HCO_3, pH 8, or at lysyl residues by *Achromobacter* protease I [a gift of Dr. T. Masaki, Ibaraki, Japan; Masaki et al. (1981)] by incubation in 50 mM Tris-HCl, pH 9.0, containing 2 M urea at 37 °C for 6 h. Selective cleavage at arginyl residues with 40 μg of trypsin was accomplished in 3 h at 37 °C after citraconylation of 40 nmol of SBP as described by Reimann et al. (1984). Lysyl residues were then decitraconylated at pH ~1.5. Each digest was lyophilized and dissolved in 6 M guanidine hydrochloride containing 10 mM phosphate, pH 6, and fractionated by exclusion chromatography on tandem TSK columns (LKB) [cf. Titani et al. (1986)]. Pooled fractions were desalted and subfractionated by reversed-phase HPLC.

Asn–Gly bonds were cleaved with 2 M NH_2OH at pH 9–9.5 at 45 °C for 5–6 h in the presence of 6 M guanidine hydrochloride and then desalted and subfractionated by reversed-phase HPLC. Asp–Pro bonds were selectively cleaved in 70% formic acid at 37 °C for 24 h (Landon, 1977) and Asp bonds in 2 M formic acid at 110 °C for 4 h (Titani et al., 1986). For subdigests of larger peptides, α-chymotrypsin, *S. aureus* V8 protease, or trypsin were used.

Mixtures of peptides were purified to homogeneity with acetonitrile gradients on SynChropak RP-P or RP-8 columns. With some larger peptides, an Altex Ultrapore RPSC column was employed. Amino acid compositions were determined in 16-h hydrolysates and amino sugars in 2-h hydrolysates either by reversed-phase HPLC with a Waters Picotag system (Bidlingmeyer et al., 1984) or by ion-exchange chromatography in a Dionex D-500 system. Most amino acid sequences were determined with a Beckman 890C sequencer. The

† This work was supported by National Institutes of Health Grants GM-15731 (K.A.W.) and HD-13956 (P.H.P.).

‡ Department of Biochemistry.
§ Howard Hughes Medical Institute.
⊥ Department of Obstetrics/Gynecology.

[1] Abbreviations: SBP, human plasma sex steroid binding protein; *S*-CM, *S*-carboxymethyl; HPLC, high-precision liquid chromatography.

0006-2960/86/0425-7584$01.50/0

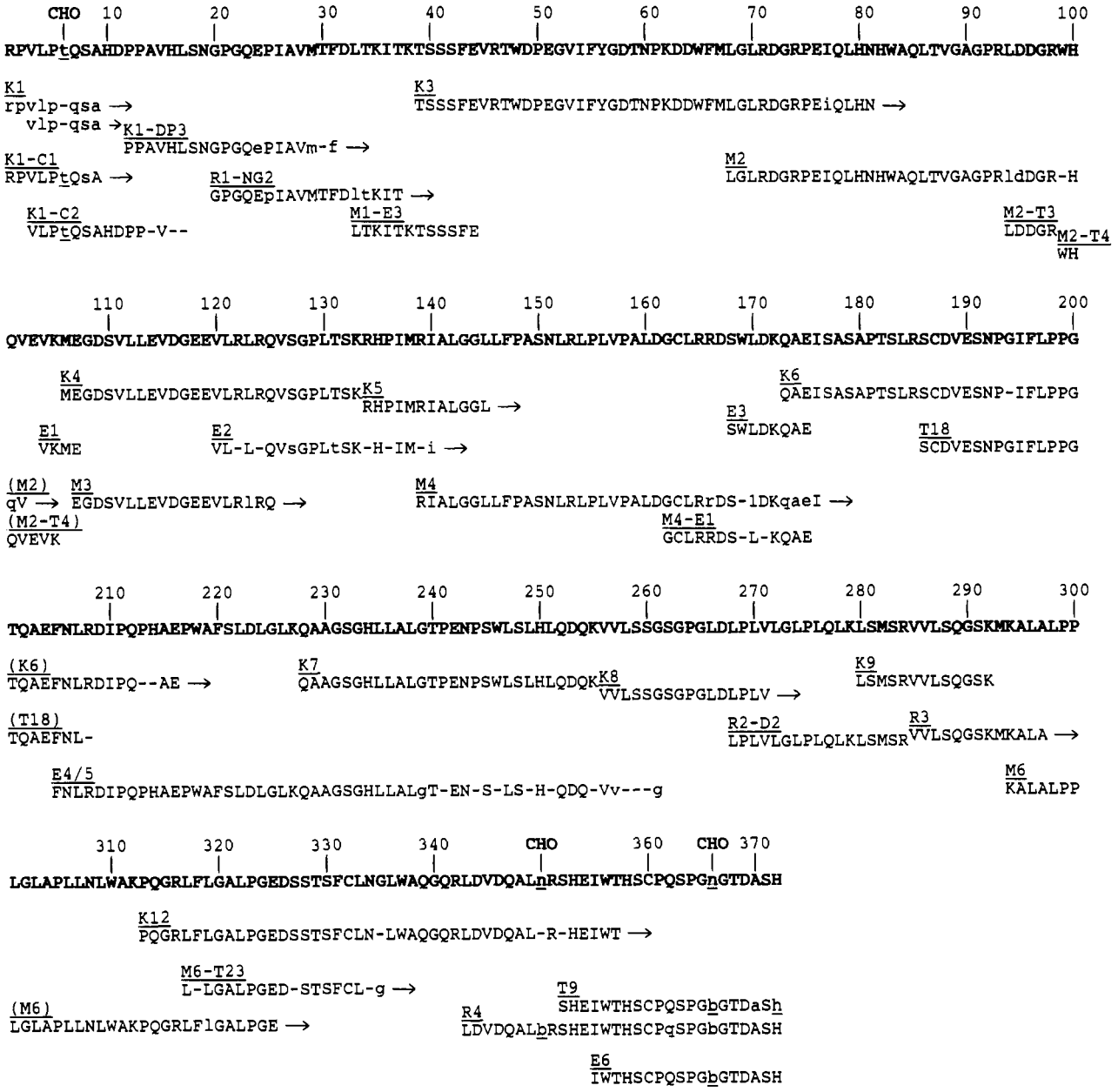

FIGURE 1: Detailed summary of the proof of sequence of the human SBP isolated in our laboratory (Figure 4 summarizes the sequence and adds an additional leucine at the N-terminus). Sequences derived by Edman degradation of specific peptides (names are underlined) are given below the summary sequence (bold type). Prefixes K, M, E, and R refer to peptides derived by cleavage at lysyl, methionyl, glutamyl, and arginyl residues. The prefix T denotes cleavage by trypsin. Subpeptides are identified by hyphenated suffixes where, in addition to the above abbreviations, C, DP, NG, and D refer to chymotryptic subdigestion, Asp–Pro cleavage, Asn–Gly cleavage, or aspartyl cleavage, respectively. Within the specified sequences lower-case letters denote tentative identification of phenylthiohydantoins, whereas those underlined are deduced from amino acid compositions. Those not identified are indicated by dashes or by an arrow for a long unidentified sequence. The nomenclature E4/5 denotes an overlap peptide with an uncleaved glutamyl residue. Its shorter amino-terminal portion (E4, residues 205–242) was also isolated but not used in the proof of sequence.

phenylthiohydantoins were identified by complementary HPLC systems (Glajch et al., 1985; Ericsson et al., 1977). Sequences of three fragments (M4, M6, K12) were analyzed with an Applied Biosystems sequencer using a program adapted from Hunkapiller et al. (1983).

Analysis of the sequence for internal homology and for homology with other proteins followed the computerized programs of Dayhoff et al. (1983).

RESULTS

The overall strategy of analysis of the amino acid sequence is outlined in Figure 1. In general, a set of peptides generated by cleavage at lysyl residues (peptides K1–K12) were aligned by overlapping fragments resulting from cleavage at either methionine (M2–M6), glutamic acid (E1–E6), or arginine (R1–R4) residues. In some cases (indicated by hyphenated acronyms), subdigests were fractionated on HPLC so that internal sequences could be determined.

Each of three primary peptide mixtures (cleaved at Lys, Arg, or Glu) was first fractionated according to size on tandem TSK columns (Figures 2 and 3). Peptides in pooled fractions were readily purified by reversed-phase HPLC using acetonitrile gradients. The amino acid compositions of the purified products of cleavage at lysyl residues are listed in Table I, together with other key peptides from the amino and carboxyl termini. Of the lysyl cleavage products, only K1 and K12 contained hexosamines (Table I), indicating that the single O-linked and the two N-linked oligosaccharide chains reported by Avvakumov et al. (1983) must be found in these glycopeptides.

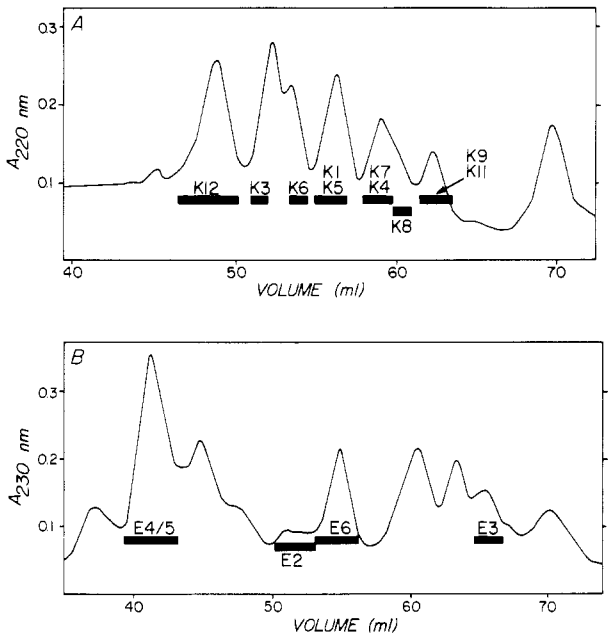

FIGURE 2: Primary separation of mixtures of peptides resulting from cleavage of *S*-CM-SBP: (A) at Lys (25 nmol of protein) with *Achromobacter* protease I; (B) at Glu (occasionally at Asp) with *S. aureus* protease V8. Each of these separations took place in 6 M guanidine hydrochloride, 10 mM phosphate (pH 6.0) at a flow rate of 0.5 mL/min on a tandem series of one TSK-SW3000 and two TSK-SW2000 columns. Each pooled fraction was desalted and subfractionated with an acetonitrile gradient (in dilute aqueous trifluoroacetic acid) and reversed-phase HPLC. Peptides from the digest at Lys residues were purified on an Ultrapore RPSC column, except K4, K7, and K12, which were separated on a SynChropak RP-P column. Peptides from the Glu digest were purified on a SynChropak RP-8 column, except E4/5, which was separated on a SynChropak RP-P column. If more than one peptide was obtained from a single pooled TSK fraction, they are listed from the top down in order of increasing hydrophobicity (their elution order on reversed-phase HPLC). Peptide E1 (residues 104–107, Figure 1) was isolated from a separate digest of *S*-CM-SBP that was applied directly to a Synchropak RP-P column, where it eluted just after the breakthrough fraction.

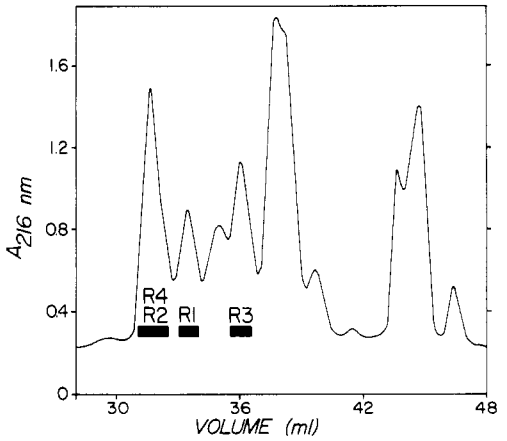

FIGURE 3: Primary separation of the mixture of peptides resulting from cleavage of 40 nmol of *N*-citraconyl-*S*-CM-SBP at Arg with trypsin. A combination of two TSK-SW2000 columns was used with the same solvent system as in Figure 2. Pooled fractions were desalted and subfractionated on a SynChropak RP-8 column as in Figure 2.

Analysis of Amino Acid Sequence. Difficulties in the analysis were encountered at the outset when amino-terminal analysis of the protein revealed two sequences in equal quantities, apparently corresponding to Arg-Pro-Val-Leu-Pro-?-Gln- and Val-Leu-Pro-?-Gln-. This same pair of amino-terminal sequences was observed upon analysis of the mixed fragments K1. In each case the pattern of phenylthiohydantoins fitted predictions for the indicated mixture, suggesting that a longer chain, RPVLP?Q-, was accompanied by another chain two residues shorter, but otherwise identical.

Interpretation of these data was complicated by (1) the unidentified residue at position 6 of the longer sequence (which appears to be a glycosylated threonine, vide infra), (2) the high proline content (which tends to result in incomplete Edman degradation), (3) an acid-labile Asp–Pro bond at residues 11–12 (that causes a "background" of artifactually broken chains to rise more rapidly than usual during the degradation), and (4) an Asn-Gly sequence at residues 19–20 (where the yield of phenylthiohydantoins drops, e.g., by 70% in the degradation of K1-DP3). The latter two problems were turned into advantages by cleaving the mixed peptides K1 at the Asp–Pro bond with formic acid and the mixed arginyl peptides R1 at Asn–Gly with hydroxylamine. These procedures generated pure peptides K1-DP3 and R1-NG2, respectively (Figure 1), which overlapped each other, providing the sequence of residues 12–37.

The true amino termini of the two forms of the protein were demonstrated in small, pure peptides K1–C1 and K1–C2, isolated from a chymotryptic subdigest of K1 by cleavage at Leu-17. The amino acid composition of K1-C2 (residues 3–17, Table I) indicated that an unplaced residue of threonine must be at residue 6 (of the longer form). SBP is known to have one O-linked oligosaccharide chain containing *N*-acetylgalactosamine (Avvakumov et al., 1983). This is consistent with our finding of galactosamine in hydrolysates of K1 and with the lack of a phenylthiohydantoin corresponding to residue 6 (glycosylated phenylthiohydantoins resist extraction during routine Edman degradations).

The amino-terminal sequence was linked to peptide K3 by the sequence of a subpeptide M1-E3 obtained by cleavage of M1 at Glu-44 and Asp-32. Overlapping sets of peptides provided a continuous sequence to Lys-255, although the overlaps in the region of residues 103–108 were only dipeptides. Peptides K6 and K7 each had an amino-terminal glutamine that partially cyclized to the pyroglutamyl form. During the Edman degradation, the initial yields were only 52% (using 2.5 nmol of K6) and 36% (6.6 nmol of K7), whereas the repetitive yields of later cycles were 94% and 95%, respectively.

Peptides K7–K9 and M6 were difficult to align for several reasons that relate to the hydrophobicity of this segment of the molecule and the difficulty of cleaving the Met–Ser bond at residues 282–283. The hydrophobicity of K8 resulted in severe loss of the peptide from the spinning cup during Edman degradation beyond Asp-267. The hydrophobic nature of this region may also account for our inability to isolate, after cleavage of Glu bonds, a fragment corresponding to residues 243–325 that would have overlapped K7-K8-K9. The absence of a small cyanogen bromide fragment corresponding to residues 283–293 made it necessary to seek an alternate overlap of K9 and M6. These difficulties were overcome by first generating fragments R2 and R3 by cleavage at Arg residues 208 and 284 (Figure 3). The sequence of R3 overlapped K9 and M6. R2 (residues 209–284) was treated with dilute acid to generate R2-D2 and permit sequencing through the hydrophobic segment of K8, overlapping K9. The overlap of K7 and K8 is the weakest segment in the proof. The alignment depends on the incomplete sequence Val-Val---Gly in cycles 52–57 of peptide E4/5.

The carboxyl-terminal cyanogen bromide fragment must be M6 as it alone lacked homoserine. Its partial sequence and that of the overlapping glycopeptide K12 provided most of the sequence through Th-357. An Asn-Gly sequence at residues

Table I: Amino Acid Compositions of Selected Peptides[a]

Fragment	K1[b]	K2[c]	K3[b]	K4	K5[b]	K6[b]	K7	K8[b]	K9	K10[c]	K11[b]	K12	K1-C2	R4	T9[b,d]	E6
Residues	1-35[e]	36-38	39-105	106-133	134-172	173-227	228-255	256-279	280-292	293-294	295-312	313-372	3-17	343-372	352-372	355-372
Asx (D/N)	3.0 (3)		7.7 (9)	1.8 (2)	4.0 (4)	5.3 (5)	1.5 (2)	1.3 (1)			1.1 (1)	6.9 (7)	1.0 (1)	4.3 (5)	1.7 (2)	2.2 (2)
Glx (E/Q)	3.0 (3)		6.6 (7)	4.8 (5)		7.4 (7)	3.6 (4)	1.3 (1)	1.1 (1)			6.9 (7)	1.2 (1)	2.9 (3)	1.8 (2)	1.2 (1)
CMC (C)					1.1 (1)	0.8 (1)						0.6 (2)		0.8 (1)	0.6 (1)	0.7 (1)
Ser (S)	1.8 (2)		3.1 (3)	2.7 (3)	1.6 (2)	5.4 (6)	2.6 (3)	2.7 (3)	3.3 (4)			6.9 (7)	1.2 (1)	5.1[f] (4)	3.2 (4)	3.4[f] (3)
Gly (G)	2.0 (2)		5.8 (7)	3.4 (3)	3.1 (3)	4.4 (3)	3.1 (3)	4.1 (4)	1.4 (1)		1.1 (1)	6.7 (7)	1.2[f]	2.9[f] (2)	1.9 (2)	2.3 (2)
His (H)	1.8 (2)		2.4 (3)		0.6 (1)	1.4 (1)	2.0 (2)					2.3 (3)	1.7 (2)	2.9 (3)	2.7 (3)	1.8 (2)
Arg (R)	0.6[e] (1)		4.6 (5)	2.0 (2)	5.2 (5)	2.6 (2)			1.1 (1)			2.9 (3)		1.1 (1)		
Thr (T)	2.7 (3)	(1)	3.2 (4)	1.0 (1)		2.1 (2)	0.9 (1)					2.5 (3)	1.0 (1)	1.9 (2)	1.7 (2)	1.7 (2)
Ala (A)	3.0 (3)		2.2 (2)		3.1 (3)	5.6 (6)	2.9 (3)				3.9 (4)	3.7 (4)	2.1 (2)	2.0 (2)	1.2 (1)	1.0 (1)
Pro (P)	5.6[e] (6)		4.2 (4)	1.2 (1)	4.1 (4)	7.0 (7)	2.0 (2)	3.1 (3)			3.2 (3)	3.8 (4)	3.1 (3)	1.9 (2)	2.0 (2)	2.1 (2)
Tyr (Y)			1.4 (1)													
Val (V)	3.3 (3)		5.2 (5)	3.6 (4)	1.8 (1)	1.4 (1)		3.8 (3)	1.7 (2)			1.2 (1)	1.9 (2)	1.1 (1)		
Met (M)	0.9 (1)		0.6 (1)	0.3 (1)	0.8 (1)				0.7 (1)	(1)						
Ile (I)	0.9 (1)	(1)	1.7 (2)		1.9 (2)	2.9 (3)						1.0 (1)		1.1 (1)	0.7 (1)	0.8 (1)
Leu (L)	3.0 (3)		4.6 (5)	4.6 (5)	8.7 (9)	6.0 (6)	5.6 (6)	7.6 (8)	2.0 (2)		6.7 (7)	6.8 (7)	2.1 (2)	2.6 (2)		
Phe (F)	1.0 (1)		2.9 (3)		1.2 (1)	2.8 (3)						2.3 (2)				
Lys (K)	1.0 (1)	(1)	1.8 (2)	1.1 (1)	1.0 (1)	1.2 (1)	0.9 (1)	1.0 (1)	1.4 (1)	(1)	1.0 (1)					
Trp (W)			N.D. (4)		N.D. (1)	N.D. (1)	N.D. (1)				N.D. (1)	N.D. (2)		N.D. (1)	N.D. (1)	N.D. (1)
GalN[g]	+		-	-	-	-	-	-	-		-	-				
GlcN[g]	-		-	-	-	-	-	-	-		-	++				
Yield (%)	38[h]	--	15	21	24	27	48[h]	38	43	--	35	31	17	15	--	25

[a] Residues per molecule by Picotag amino acid analysis unless otherwise indicated or (in parentheses) from the sequence (Figure 1). [b] Amino acid analysis by ion-exchange chromatography. [c] Not isolated. [d] Recalculated from analysis of Petra et al. (1986a). [e] Equal mixture of residues 1–35 and 3–35. [f] Carbohydrate tends to distort the Ser/Gly values in Picotag analyses. [g] GalN (galactosamine) and GlcN (glucosamine) observed in 2-h hydrolysates. [h] Peptides K1 and K7 were also purified in one HPLC step from 1.7 nmol of protein with yields of 56 and 62%, respectively.

334–335 caused a marked drop in the yield of the Edman degradation at that point [cf. Hermodson et al. (1973)] in both K12 and a tryptic subpeptide (M6–T23).

The remainder of the sequence was provided by carboxyl-terminal peptides from each of three digests. The only tryptic peptide lacking Lys or Arg was T9. Its amino acid composition (Table I) indicated that an unidentified phenylthiohydantoin corresponding to residue 366 must be Asx or His (Figure 1). Analysis of E6 placed His at the carboxyl terminus and, as a consequence, Asx at residue 366. The placement of an N-glycosylated Asn at that locus is consistent with the lack of a phenylthiohydantoin at residue 366 in both sequences and the presence of glucosamine in K12. The composition and sequence of R4 confirmed the C-terminal His and indicated a second N-glycosylated Asn at residue 350. To provide a final confirmation of the oligosaccharide placements, peptide K12 was digested with trypsin at Arg-316, Arg-342, and Arg-351. The four products were separated by HPLC, identified by amino acid analyses, and analyzed for hexosamines. The peptides corresponding to residues 343–351 and 352–372 contained glucosamine; those corresponding to residues 313–316 and 317–342 did not. Both proposed N-linked oligosaccharides are found in typical Asn-X-Ser(Thr) sequences.

Placement of Disulfides. SBP was treated with CNBr as before but without prior reduction and alkylation of the cystine residues. The fragmentation pattern [cf. Petra et al. (1986a)] was indistinguishable on TSK columns, whether the disulfides were reduced or not. It is concluded that the two Cys residues within M6 (residues 332 and 360) must be connected by a disulfide bond. Similarly Cys-163 and Cys-187, both in M4, must be linked by a disulfide bond.

Search for Homology. No internal homology was evident in the SBP sequence upon examination by the RELATE program of Dayhoff et al. (1983). Nor was a homologous relationship found to other proteins with the SEARCH program and the 3557 sequences in the Feb 28, 1986, protein data base of the National Biomedical Research Foundation. That data base included two proteins that interact with steroids, a δ-isomerase (Benson et al., 1971) and a prostatic steroid-binding protein (Parker et al., 1983). The data base was supplemented with the recently reported sequences of the human glucocorticoid receptor (Hollenberg et al., 1985), the human estrogen receptor (Greene et al., 1986), and the erythroblastosis viral oncogene product v-erbA (Weinberger et al., 1985).

Discussion

In a previous paper (Petra et al., 1986a) we described peptides generated from human SBP by the action of trypsin or cyanogen bromide. The present paper relates those peptides to others isolated after cleavage at lysine, arginine, or glutamic acid residues. Together, the products of these five digests provide the overlapping set of information that leads to the amino acid sequence in Figure 1. The lysine-specific protease from *Achromobacter* was particularly useful in this analysis, and 10 peptides from that digest (or their subdigestion products) provided nearly 70% of the sequence. Although a tripeptide and a dipeptide were not isolated from this digest (K2 and K10, Table I), this simply reflects the ease with which very small peptides can be overlooked in salt-containing, flow-through fractions on reversed-phase HPLC. It is also apparent that the Met–Thr bond following Met-29 and the Met–Ser bond following Met-282 both resisted cleavage with cyanogen bromide, as observed in many other proteins.

The most difficult problem in the analysis related to the "ragged" amino terminus, where there is an equimolar mixture of polypeptide chains differing only in their starting points at residues 1 or 3, as recognized earlier during Edman degradation of the whole protein (Petra et al., 1983). Moreover, each of the primary digests yielded a fraction containing the corresponding mixture from the amino terminus that resisted subfractionation into its two component peptides. Both separation of the two components and Edman degradation of the mixture were made more difficult by the oligosaccharide moiety on Thr-6. The Edman degradation was further compromised by the high proline content of the first 21 residues and by the Asn-Gly sequence at residues 19–20. The sequence was finally solved with smaller, separable subdigestion products by taking advantage of the sensitivity of the Asn–Gly linkage to hydroxylamine, of the Asp–Pro bond at residues 10–11 to acid, and of the Leu-17 bond to chymotryptic digestion.

The carboxyl terminus of the protein also presented difficulties, in part because the two oligosaccharide chains tend to interfere with quantitative Edman degradation. Ultimately, the proof of that sequence depended upon the complete agreement between degradation data and amino acid compositions of three peptides, each containing 30 or fewer residues.

The weakest point in the proof of sequence is the alignment of K7 and K8 in the vicinity of residue 255, where an extended analysis of peptide E4/5 provides the only overlapping data. All other overlaps are more convincing. Most of the phenylthiohydantoins (83%) were identified at least twice, and in complementary HPLC systems. Nineteen other residues were identified as their phenylthiohydantoins in only one HPLC system but confirmed either by the composition of a small peptide or by the specificity of cleavage. However, His-82, Arg-134, and Trp-337 were each identified only once, as were the sequences of residues 146–162, 257–267, and 298–312. Although we have no reason to suspect errors in any of these identifications, we prefer the duplicate proof that characterizes the remainder of the molecule.

The amino acid composition based on the sequence is in reasonable agreement with that determined in hydrolysates (Table II), although these data differ from earlier estimates by Mickelson et al. (1978) and with the revised composition of Turner et al. (1984). It is probable that the tyrosine and

Table II: Amino Acid Composition of Human SBP

	Hammond et al. (1986)[a]	acid hydrolysis[b]		sequence (Figure 4)[c]
		24 h	48 h	
Ala	25.2	27.5	25.8	25
Cys	6.1	4.3	3.2	4
Asx	32.0	31.4	32.4	34
Glx	34.1	34.7	37.8	35
Phe	10.3	11.2	11.7	10
Gly	33.4	35.5	37.7	34
His	11.5	12.0	11.9	12
Ile	9.8	9.5	9.4	10
Lys	12.4	12.0	12.1	12
Leu	53.9	58.0	58.0	59
Met	4.2	5.7	4.7	6
Pro	34.0	33.8	34.6	34
Arg	20.0	18.6	17.9	19
Ser	28.3	28.7	27.8	33
Thr	14.2	14.2	13.6	15
Val	18.3	19.0	19.8	20
Trp	ND	11.7[d]	ND	10
Tyr	2.2	2.7	1.8	1

[a] Recalculated ratios of Hammond et al. (1986). [b] Data for 24- and 48-h hydrolysates are from two different preparations of SBP. Each set of data is normalized to 58 leucine/mol (as found in the sequence of our material, Figure 1). [c] The molecular weight of the polypeptide chain is calculated from the sequence to be 40 499. Inclusion of 14% carbohydrate (Turner et al., 1984) gives a subunit molecular weight of 47 000. [d] Recalculated from Turner et al. (1984) for M_r 40 499.

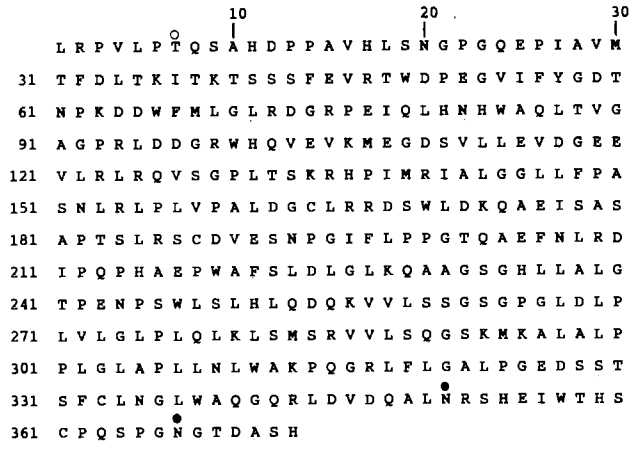

FIGURE 4: Summary of the amino acid sequence of human SBP, including the amino-terminal leucine reported by Fernlund and Laurell (1981) and Hammond et al. (1986), which was lacking in our preparations. The numbering of the residues thus differs by one from that in the proof of sequence in Figure 1. The open circle above Thr-7 denotes an O-linked oligosaccharide chain; the closed circles above Asn-351 and Asn-367 denote N-linked oligosaccharide chains.

phenylalanine values in the hydrolysates were distorted by amino sugars during analysis on ion-exchange columns.

It has been known for some time that the native human protein was composed of two similar subunits (Petra, 1979; Cheng et al., 1983; Suzuki & Sinohara, 1984). We have reported that they were virtually identical (Petra et al., 1984, 1986a), and this has been confirmed recently by Hammond et al. (1986). The data reported here verify the conclusion since only one amino acid sequence was found for the dimer, except for the absence of the amino-terminal Arg-Pro sequence in ~50% of the subunits.

Hammond et al. (1986) also report heterogeneity at the amino terminus where their sequence LRPVLPXQ- was mixed with a component (~25%) that lacked the terminal leucine. Fernlund and Laurell (1981) have reported the presence of a leucine residue at the amino terminus. The amino-terminal sequence of the protein isolated in our laboratory lacked that leucine entirely, and ~50% of our protein lacked in addition the penultimate Arg-Pro. Otherwise our amino-terminal sequence and that of Hammond et al. are identical for 25 residues, including the carbohydrate attachment site that Hammond et al. postulated to be Asn and that we identify as Thr. It is assumed that somewhat more proteolytic degradation occurred during the preparation or storage of our protein than that of Hammond et al. Our summary of the sequence in Figure 4 (and subsequent discussion) includes the leucine that was observed only in the preparations of Hammond et al. (1986).

A plausible explanation for the microheterogeneity of SBP implicates variability in the oligosaccharide chains of the SBP subunit (Petra et al., 1986b; Hammond et al., 1986). Another possibility involves the rearrangement of the Asn–Gly bonds at residues 20 and 335 (Figure 4) to yield cyclic imide intermediates leading to free α- or β-carboxyl groups (Bornstein & Balian, 1977). Combinations of these side reactions, the carbohydrate variability, and the amino-terminal differences could generate a set of minor variants of the SBP subunit. Random pairing would yield different active dimeric SBP molecules that would exhibit the different mobilities observed both during isoelectric focusing and on SDS-PAGE [e.g., Petra et al. (1983)].

The two proposed Asn-linked carbohydrate moieties (residues 351 and 367, Figure 4) are each found two residues from the amino-terminal side of a hydroxy amino acid, as expected from studies of other glycoproteins. These placements and that of a Thr-linked carbohydrate at residue 7 are in accord with studies by Avvakumov et al. (1983) of the carbohydrate moieties of SBP. They separated from a pronase digest two glycopeptide fractions, one of which appeared to contain only Asn and carbohydrate moieties. The other contained O-linked carbohydrate and threonine, as well as several other amino acids. However, the yields of these amino acids indicate the presence of peptide contaminants in addition to a fragment surrounding Thr-7. They noted a high proline content, and we find several prolines in the vicinity of Thr-7. Indeed, proline-rich areas are often found in the vicinity of O-linked carbohydrate (Aubert et al., 1976).

Examination of the amino acid sequence for hydropathy (Kyte & Doolittle, 1982) and for hydrophilicity (Hopp & Woods, 1981) reveals several general features of the molecule. The carboxyl-terminal 35 amino acid residues are predominantly hydrophilic, even without consideration of the two oligosaccharide chains. This whole segment is surely a surface feature of the molecule. To a lesser extent, the amino-terminal region is also hydrophilic, but this is most clearly identified by its oligosaccharide attachment. Three markedly hydrophobic regions, each of 9–13 residues, are evident at residues 141–149, 269–277, and 298–311 (Figure 4). Nearly 50% of these three regions consists of leucine residues.

An unusual pattern of alternating leucyl residues is found between residues 267 and 281 (Figure 4), namely LDLPLVLGLPLQLKL. Analysis of the sequence for predominant secondary structure, by the method of Chou and Fasman (1978), revealed that residues 269–275 are predicted to form a β-structure and residues 277–290 an α-helical segment. Interestingly, such a structure would place four leucines on one side of the hypothetical β-structure and three leucines (residues 277, 281, and 288) on one face of the α-helix. Such a structure has the potential to create a "steroid sandwich". However, it is not possible at this time to distinguish between putative roles of hydrophobic segments as apolar faces for the steroid-binding site and as hydrophobic strands buried in the interior of the subunits. Affinity-labeling studies or three-dimensional analyses would serve to explore these alternatives. Alternatively, it may be useful to take advantage of the single tyrosine at residue 57 as an anchoring site for a reporter group. Other hydrophobic residues of potential interest are the 20 aromatic residues, but they appear to be distributed in a random pattern throughout the molecule from Phe-32 to Trp-357.

The cDNA-derived amino acid sequences of other steroid-binding proteins appear to bear no similarity to SBP. It has been reported recently that domains of the human glucocorticoid receptor (Hollenberg et al., 1985) and human estrogen receptor (Greene et al., 1986) are homologous to each other and interestingly to the erythroblastosis viral oncogene product v-erbA (Weinberger et al., 1985). However, this homologous relationship does not appear to extend to SBP.

The availability of the amino acid sequence should now simplify efforts to clone the SBP gene, to understand the role of the protein in regulation, and eventually to achieve high resolution of its three-dimensional structure by X-ray diffraction. Comparison of the sequence of SBP with those of other steroid-binding proteins should lead to a description of steroid-binding domains and an understanding of the very tight binding of steroids that characterizes these protein/hormone complexes.

Added in Proof. After submission of this paper, we attended the First International Symposium on Binding Proteins in Lyon

(April 28, 1986) where the cDNA sequence of a rat testis androgen binding protein was reported by Drs. D. R. Joseph, S. H. Hall, and F. S. French. The androgen-binding protein has 373 residues, an amino-terminal Leu-Arg, a carboxyl-terminal Ser-His, and 68% sequence identity with SBP (without gaps), as will be detailed in a joint communication in the forthcoming proceedings of that symposium.

ACKNOWLEDGMENTS

We thank Roger D. Wade for performing the amino acid analyses, Lowell Ericsson for help with the homology search, Mary K. Woods for assistance with the manuscript, and Dr. Hans Neurath for his support and advice.

Registry No. Protein (human blood plasma sex steroid binding), 104291-03-4.

REFERENCES

Aubert, J.-P., Biserte, G., & Loucheux-Lefebre, M.-H. (1976) *Arch. Biochem. Biophys. 175*, 410.

Avvakumov, G. V., Matveentseva, I. V., Akhrem, L. V., Strel'chyonok, O. A., & Akhrem, A. A. (1983) *Biochim. Biophys. Acta 760*, 104.

Benson, A. M., Jarabak, R., & Talalay, P. (1971) *J. Biol. Chem. 246*, 7514.

Bidlingmeyer, B. A., Cohen, S. A., & Tarvin, T. L. (1984) *J. Chromatogr. 336*, 93.

Bornstein, P., & Balian, G. (1977) *Methods Enzymol. 47*, 132.

Cheng, C. Y., Musto, N. A., Gunsalus, G. L., & Bardin, C. W. (1983) *J. Steroid Biochem. 19*, 1379.

Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45.

Dayhoff, M. O., Barker, W. C., & Hunt, L. T. (1983) *Methods Enzymol. 91*, 524.

Ericsson, L. H., Wade, R. D., Gagnon, J., MacDonald, R. M., Granberg, R. R., & Walsh, K. A. (1977) in *Solid Phase Methods in Protein Sequence Analysis* (Previero, A., & Coletti-Previero, M.-A., Eds.) p 137, Elsevier/North-Holland Press, New York.

Fernlund, P., & Laurell, C. (1981) *J. Steroid Biochem. 14*, 545.

Glajch, J. L., Gluckman, J. C., Charikofsky, J. G., Minor, J. M., & Kirkland, J. J. (1985) *J. Chromatogr. 318*, 23.

Greene, G. L., Gilna, P., Waterfield, M., Baker, A., Hort, Y., & Shine, J. (1986) *Science (Washington, D.C.) 231*, 1150.

Hammond, G. L., Robinson, P. A., Sugino, H., Ward, D. N., & Finne, J. (1986) *J. Steroid Biochem. 24*, 815.

Hermodson, M. A., Ericsson, L. H., Neurath, H., & Walsh, K. A. (1973) *Biochemistry 12*, 3146.

Hollenberg, S. M., Weinberger, C., Ong, E. S., Cerelli, G., Oro, A., Lebo, R., Thompson, E. B., Rosenfeld, M. G., & Evans, R. M. (1985) *Nature (London) 318*, 635.

Hopp, T. P., & Woods, K. R. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 3824.

Hunkapiller, M. W., Hewick, R. M., Dreyer, W. J., & Hood, L. E. (1983) *Methods Enzymol. 91*, 399.

Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105.

Landon, M. (1977) *Methods Enzymol. 47*, 145.

Masaki, T., Tanabe, M., Nakamura, K., & Soejima, M. (1981) *Biochim. Biophys. Acta 660*, 44.

Mickelson, K. E., Teller, D. C., & Petra, P. H. (1978) *Biochemistry 17*, 1409.

Namkung, P. C., & Petra, P. H. (1986) First International Symposium on Binding Proteins of Steroid Hormones, Lyon, France, Abstract 1.

Parker, M. G., White, R., Hurst, H., Needham, M., & Tilly, R. (1983) *J. Biol. Chem. 258*, 12.

Petra, P. H. (1979) *J. Steroid Biochem. 11*, 245.

Petra, P. H., & Lewis, J. (1980) *Anal. Biochem. 105*, 165.

Petra, P. H., Stanczyk, F. Z., Senear, D. F., Namkung, P. C., Novy, M. J., Ross, J. B. A., Turner, E. E., & Brown, J. A. (1983) *J. Steroid Biochem. 19*, 699.

Petra, P. H., Titani, K., Ericsson, L. H., Takio, K., & Walsh, K. A. (1984) Seventh International Congress on Endocrinology, Quebec City, Abstract 2738.

Petra, P. H., Kumar, S., Hayes, R., Ericsson, L. H., & Titani, K. (1986a) *J. Steroid Biochem. 24*, 45.

Petra, P. H., Namkung, P. C., Senear, D. F., McCrae, D. A., Rousslang, K. W., Teller, D. C., & Ross, J. B. A. (1986b) *J. Steroid Biochem. 25*, 191.

Reimann, E. M., Titani, K., Ericsson, L. H., Wade, R. D., Fischer, E. H., & Walsh, K. A. (1984) *Biochemistry 23*, 4185.

Suzuki, Y., & Sinohara, H. (1984) *J. Biochem. (Tokyo) 96*, 751.

Takio, K., Smith, S. B., Walsh, K. A., Krebs, E. G., & Titani, K. (1983) *J. Biol. Chem. 258*, 5531.

Titani, K., Kumar, S., Takio, K., Ericsson, L. H., Wade, R. D., Ashida, K., Walsh, K. A., Chopek, M. W., Sadler, J. E., & Fujikawa, K. (1986) *Biochemistry 25*, 3171.

Turner, E. E., Ross, J. B. A., Namkung, P. C., & Petra, P. H. (1984) *Biochemistry 23*, 492.

Weinberger, C., Hollenberg, S. M., Rosenfeld, M. G., & Evans, R. M. (1985) *Nature (London) 318*, 670.