Volume 220, number 1, 129-135 FEB 04980 August 1987

A cDNA coding for human sex hormone binding globulin

Homology to vitamin K-dependent protein S

Sten Gershagen, Per Fernlund and Åke Lundwall

Department of Clinical Chemistry, University of Lund, Malmö General Hospital, S-214 01 Malmö, Sweden

Received 16 March 1987

Affinity purified antibodies to human sex hormone binding globulin (SHBG) were used in screening a human liver cDNA library, constructed in the expression vector λgt11. One clone, identified as producing recombinant SHBG, carried a cDNA insert of 1.1 kb. The nucleotide sequence of the insert had an open reading frame coding for 356 amino acid residues. The coding sequence was followed by a short 3′-region of 19 non-translated nucleotides and a poly(A) tail. Confirmation that the cDNA clone represented human SHBG was obtained by the finding of a complete agreement in amino acid sequence with several peptide fragments generated from purified SHBG by proteolytic cleavage. The primary structure of SHBG shows a considerable homology to that of protein S, a vitamin K-dependent protein with functions in the coagulation system.

Sex hormone binding globulin; Sex hormone; Carrier protein; Plasma protein; Blood coagulation

1. INTRODUCTION

Sex hormone binding globulin (SHBG) is an important carrier of sex steroids in plasma [1]. It is thought to be synthesized in the liver [2] and the concentration of SHBG in plasma is dependent on several hormones, being increased by estrogens and thyroid hormones and decreased by androgens. Human SHBG is a dimeric protein with a molecular mass of approx. 100 kDa. It is composed of two identical or very similar, non-covalently associated protomers. Only one steroid molecule is bound per dimeric protein, but little else is known about the mechanisms for the steroid binding [3-5].

The protomers of SHBG vary in size, two main size classes having usually been observed, one heavier and one lighter occurring at an approximate ratio of 10:1 [6,7]. However, it has recently been demonstrated that this only represents the prevailing pattern, occurring in 80% of the population. In the remaining 20% there are three different sizes of the SHBG protomer [8].

Previous reports on the carbohydrate structure [4,9,10] and partial amino acid sequences of human SHBG [3,7] have recently been complemented with the complete amino acid sequence of the human SHBG protomer [4]. None of these reports, however, explains the observed variation in size of the SHBG protomer.

To gain further knowledge of the structure of SHBG and its possible variation, and to permit studies of its gene structure and of its biosynthesis in various organs, we have isolated a cDNA clone coding for SHBG from a human liver cDNA library.

Correspondence address: S. Gershagen, Dept Clinical Chemistry, Malmö General Hospital, S-21401 Malmö, Sweden

2. MATERIALS AND METHODS

2.1. *Reagents*

Chemicals used for amino acid analysis were obtained from Beckman Instruments, and for protein

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

sequence determination from Applied Biosystems. Solvents for HPLC were from Rathburn. TPCK-trypsin was obtained from Sigma and *Staphylococcus aureus* protease V8 from Miles. A λgt11 screening kit was purchased from Promega Biotechniques. The Klenow fragment of DNA polymerase 1 was obtained from Boehringer Mannheim. The restriction enzymes *Eco*RI and *Stu*I were from Amersham, who also supplied α-[^{35}S]thio-dATP (22 TBq/mmol). Deoxynucleotides, dideoxynucleotides, T_4 DNA polymerase and the replicative form of the phage M13 mp18 were purchased from Pharmacia. An antiserum to human SHBG has been described [11]. It was affinity purified on a column of agarose coupled with SHBG, purified as described [12].

2.2. *Isolation of cDNA clones and DNA sequencing*

A λgt11 cDNA library prepared from human liver poly(A)$^+$ RNA was kindly provided by Dr Savio L. Woo. The library was screened according to the procedures described by Young and Davis [13,14], using affinity purified rabbit anti-human SHBG as first antibody and an alkaline phosphatase labelled goat anti-rabbit immunoglobulin for detection. Recombinant phages were purified from plate lysates by centrifugation in a CsCl gradient, and the DNA was extracted by standard procedures [15]. Sequencing of DNA was carried out on material subcloned in M13 mp18 according to Sanger et al. [16] as modified by Biggin et al. [17]. Nucleotide sequences were aligned and analyzed with computer programs kindly provided by R. Staden, Medical Research Council Unit, Cambridge [18,19] and the Protein Identification Resource [20].

2.3. *Isolation and sequencing of peptide fragments*

Human SHBG (1.4 mg), purified as described [12], was dialyzed at 4°C against 6 M guanidine HCl and 5 mM EDTA in 0.5 M Tris-HCl buffer, pH 8.0. The dialyzed sample was flushed with N_2 and reduced by adding 3.1 μmol of dithiothreitol followed by incubation at 37°C for 4 h. Alkylation was done with 7.8 μmol of solid, recrystallized, iodoacetic acid and the mixture was left for 2 h in the dark at room temperature with gentle agitation. The sample was dialyzed against 0.1 M ammonium bicarbonate buffer, pH 7.5, followed by evaporation. The residue was dissolved in 200 μl of the ammonium bicarbonate buffer and treated with 28 μg of trypsin or *S. aureus* protease V8 for 2 h at 37°C. The sample was evaporated to dryness, which was repeated twice after the addition of redistilled water to remove salt thoroughly.

High-performance liquid chromatography (HPLC) was performed at room temperature on a μBondapak column (0.39 × 30 cm) eluted at a flow rate of 0.8 ml/min with a linear gradient from solvent A : solvent B (95 : 5, v/v) to solvent A : solvent B (30 : 70, v/v) developed over 1 h. Solvent A was 0.08% trifluoroacetic acid in water and solvent B acetonitrile. Fractions corresponding to peptide peaks were pooled and lyophilized. One half of the material from each isolated peptide was subjected to amino acid analysis and the other to sequence determination. Amino acid sequence and composition of peptides were determined on an Applied Biosystems gas-phase sequencer and a Beckman 6300 amino acid analyzer, respectively, as described [21].

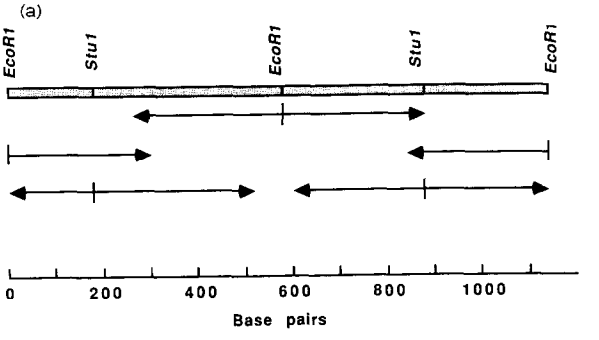

Fig.1. Restriction map and sequencing strategy (a) and nucleotide sequence with predicted amino acid sequence (b) of human SHBG cDNA. (a) Restriction endonuclease cleavage sites are shown above the linear map; arrows, DNA fragments sequenced. (b) Nucleotide sequence of the cDNA presented in the 5′ to 3′ orientation. The longest translational open reading frame is shown above the nucleotide sequence. Numbers to the right of the sequences indicate amino acid residue (upper) and nucleotide (lower). Number 18 is assigned to the first amino acid residue for reasons given in the text. Amino acid sequences also determined in peptides (generated from purified SHBG by proteolytic cleavage) are underlined.

(b)

```
    Val His Ser Ala Ala Gln Thr Thr Leu Ile Ala Val Met Thr Phe Asp Leu Thr Lys    36
 CT GTC CAC TCG GCG GCT CAG ACA ACT CTT ATC GCT GTC ATG ACC TTT GAC CTC ACC AAG    59

Ile Thr Lys Thr Ser Ser Ser Phe Glu Val Arg Thr Trp Asp Pro Glu Gly Val Ile Phe    56
ATC ACA AAA ACC TCC TCC TCC TTT GAG GTT CGA ACC TGG GAC CCA GAG GGA GTG ATT TTT   119

Tyr Gly Asp Thr Asn Pro Lys Asp Asp Trp Phe Met Leu Gly Leu Arg Asp Gly Arg Pro    76
TAT GGG GAT ACC AAC CCT AAG GAT GAC TGG TTT ATG CTG GGA CTT CGA GAC GGC AGG CCT   179

Glu Ile Gln Leu His Asn His Trp Ala Gln Leu Thr Val Gly Ala Gly Pro Arg Leu Asp    96
GAG ATC CAA CTG CAC AAT CAC TGG GCC CAG CTT ACG GTG GGT GCT GGA CCA CGG CTG GAT   239

Asp Gly Arg Trp His Gln Val Glu Val Lys Met Glu Gly Asp Ser Val Leu Leu Glu Val   116
GAT GGG AGA TGG CAC CAG GTG GAA GTC AAG ATG GAG GGG GAC TCT GTG CTG CTG GAG GTG   299

Asp Gly Glu Glu Val Leu Arg Leu Arg Gln Val Ser Gly Pro Leu Thr Ser Lys Arg His   136
GAT GGG GAG GAG GTG CTG CGC CTG AGA CAG GTC TCT GGG CCC CTG ACC AGC AAA CGC CAT   359

Pro Ile Met Arg Ile Ala Leu Gly Gly Leu Leu Phe Pro Ala Ser Asn Leu Arg Leu Pro   156
CCC ATC ATG AGG ATT GCG CTT GGG GGG CTG CTC TTC CCC GCT TCC AAC CTT CGG TTG CCG   419

Leu Val Pro Ala Leu Asp Gly Cys Leu Arg Arg Asp Ser Trp Leu Asp Lys Gln Ala Glu   176
CTG GTT CCT GCC CTG GAT GGC TGC CTG CGC CGG GAT TCC TGG CTG GAC AAA CAG GCC GAG   479

Ile Ser Ala Ser Ala Pro Thr Ser Leu Arg Ser Cys Asp Val Glu Ser Asn Pro Gly Ile   196
ATC TCA GCA TCT GCC CCC ACT AGC CTC AGA AGC TGT GAT GTA GAA TCA AAT CCC GGG ATA   539

Phe Leu Pro Pro Gly Thr Gln Ala Glu Phe Asn Leu Arg Asp Ile Pro Gln Pro His Ala   216
TTT CTC CCT CCA GGG ACT CAG GCA GAA TTC AAT CTC CGA GAC ATT CCC CAG CCT CAT GCA   599

Glu Pro Trp Ala Phe Ser Leu Asp Leu Gly Leu Lys Gln Ala Ala Gly Ser Gly His Leu   236
GAG CCC TGG GCC TTC TCT TTG GAC CTG GGA CTC AAG CAG GCA GCA GGC TCA GGC CAC CTC   659

Leu Ala Leu Gly Thr Pro Glu Asn Pro Ser Trp Leu Ser Leu His Leu Gln Asp Gln Lys   256
CTT GCT CTT GGG ACA CCA GAG AAC CCA TCT TGG CTC AGT CTC CAC CTC CAA GAT CAA AAG   719

Val Val Leu Ser Ser Gly Ser Gly Pro Gly Leu Asp Leu Pro Leu Val Leu Gly Leu Pro   276
GTG GTG TTG TCT TCT GGG TCG GGG CCA GGG CTG GAT CTG CCC CTG GTC TTG GGA CTC CCT   779

Leu Gln Leu Lys Leu Ser Met Ser Arg Val Val Leu Ser Gln Gly Ser Lys Met Lys Ala   296
CTT CAG CTG AAG CTG AGT ATG TCC AGG GTG GTC TTG AGC CAA GGG TCG AAG ATG AAG GCC   839

Leu Ala Leu Pro Pro Leu Gly Leu Ala Pro Leu Leu Asn Leu Trp Ala Lys Pro Gln Gly   316
CTT GCC CTG CCT CCC TTA GGC CTG GCT CCC CTC CTT AAC CTC TGG GCC AAG CCT CAA GGG   899

Arg Leu Phe Leu Gly Ala Leu Pro Gly Glu Asp Ser Ser Thr Ser Phe Cys Leu Asn Gly   336
CGT CTC TTC CTG GGG GCT TTA CCA GGA GAA GAC TCT TCC ACC TCT TTT TGC CTG AAT GGC   959

Leu Trp Ala Gln Gly Gln Arg Leu Asp Val Asp Gln Ala Leu Asn Arg Ser His Glu Ile   356
CTT TGG GCA CAA GGT CAG AGG CTG GAT GTG GAC CAG GCC CTG AAC AGA AGC CAT GAG ATC  1019

Trp Thr His Ser Cys Pro Gln Ser Pro Gly Asn Gly Thr Asp Ala Ser His
TGG ACT CAC AGC TGC CCC CAG AGC CCA GGC AAT GGC ACT GAC GCT TCC CAT TAA AGCTCCA  1080

CCTAAGAACCCCAAAAAAAAAAAAAAAAAA
```

3. RESULTS

3.1. *Isolation and sequencing of a human SHBG cDNA clone*

Approximately 1.2×10^6 recombinants, in a λgt11 human liver cDNA library were screened. One clone, designated λSHBG-1, was isolated, the DNA of which was digested with *Eco*RI and then analyzed by agarose gel electrophoresis. In addition to the vector bands, two bands were seen, one representing a fragment of approximately 560 nucleotides, the other of 450. The *Eco*RI fragments were isolated and subcloned in bacteriophage M13 mp18. The ends of the subcloned DNA were sequenced, yielding approximately 90% of the sequence of the whole cDNA insert. From analysis of the partial sequence, two suitably located *Stu*I cleavage sites were found. By combined *Eco*RI and *Stu*I digestion, fragments were generated that yielded enough data to complete the sequence of the λSHBG-1 cDNA insert. Fig.1a gives an outline of the sequencing strategy. The nucleotide and derived amino acid sequences for the cloned SHBG cDNA are given in fig.1b. The length is 1110 base pairs, including a short poly(A) tail of 18 nucleotides. The coding region, encoding 356 amino acids, extends from the beginning of the insert to a stop codon, TAA, at position 1071–1073. The stop codon is also part of the polyadenylation signal, ATTAAA. A stretch of 19 non-translated nucleotides separates the coding region from the poly(A) tail.

3.2. *Comparison of the deduced amino acid sequence with peptide fragments of SHBG*

A trypsin digest of reduced and alkylated human SHBG was separated by HPLC. Nine pure peptides and two fractions, each a mixture of two peptides, were obtained. Similarly, six pure peptides were isolated from a digest of reduced and alkylated SHBG obtained with the *S. aureus* protease V8. For both digests, the yield of each peptide was about 10%, which probably should not be regarded as low when considering the small scale on which the isolations had to be carried out.

The isolated peptides were sequenced and all of them could be identified as parts of the amino acid sequence deduced from the nucleotide sequence of the cDNA. The parts of the deduced sequence corresponding to the peptides are underlined in fig.1b. The peptide sequence data accounted for altogether 49% of the entire deduced amino acid sequence (fig.1b) and there were no peptide sequences that could not be found in the deduced sequence.

The SHBG cDNA contains one internal *Eco*RI restriction site (fig.1b), and this site is not overlapped by any of the DNA sequences (fig.1a). It is, however, overlapped by one long tryptic peptide (fig.1b) thus excluding the presence of more than one *Eco*RI cleavage site.

Comparison of the deduced amino acid sequence with the recently published protein sequence of the SHBG protomer [4] shows a complete agreement from residue 27 in the published sequence to its end. The amino acid residues in fig.1b have therefore been numbered to coincide with the published sequence, the only discrepancy being in the amino-terminal part; the first nine amino acids, which should correspond to amino acid residues 18–26 in the published sequence, are completely different.

3.3. *Homology to vitamin K-dependent protein S*

An alignment of human SHBG with bovine and human protein S [22,23] is shown in fig.2. The alignment shown was obtained by first aligning human SHBG separately with human and bovine protein S by means of a computer program [20]. All three sequences were then aligned with each other by making a few arbitrary adjustments which increased residue correspondence. The computer alignments with human and bovine protein S gave the maximal scores 252 and 264, respectively, when 40 was used as a penalty for a break. The mean scores obtained with 100 random sequences, generated from an amino acid composition identical to that of human SHBG, were 11.4 and 11.7 for alignments with human and bovine protein S, respectively, with standard deviations 6.2 and 4.7.

4. DISCUSSION

Except for the nine amino acid residues coded for by the 5′-end, the amino acid sequence deduced from the cloned cDNA is in complete agreement with the recently published sequence of human SHBG determined by protein chemical methods [4]. Although the first 29 nucleotides of the cDNA (coding for nine amino acids) might

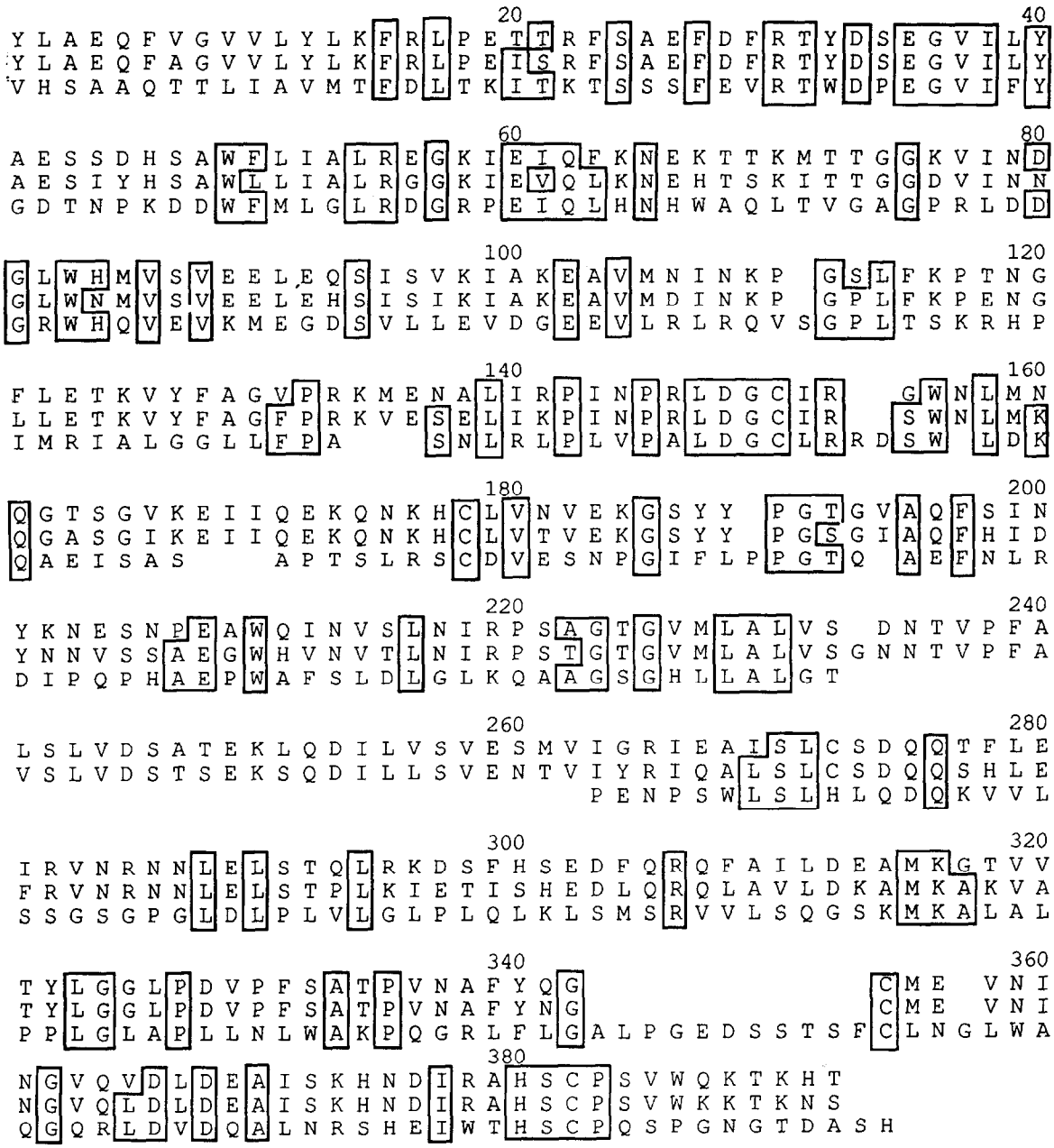

Fig.2. Amino acid sequence comparison between human SHBG (lower) and the carboxy-terminal portions of human (middle) and bovine (upper) protein S. Amino acid residues matched in SHBG and in at least one of the protein S species are boxed. Spaces represent gaps inserted to maximize the number of matches in the comparison.

represent a cloning artefact, another explanation that cannot be ruled out is that the variant sequence represents an SHBG subunit partly different from that sequenced by Walsh et al. [4].

It is a well established fact that SHBG, which is a dimeric protein, contains subunits differing slightly in structure [3,6,7] and it has recently been shown that the SHBG subunits are not uniform in

all humans, 20% of the population having an extra SHBG subunit which is heavier than the usual two subunits [8]. The structure differences between the SHBG subunits may be due to different degrees of glycosylations or other posttranslational modifications, but a difference in amino acid sequence is also a possibility. Such a variation in primary structure might result from an alternative splicing, for instance.

If SHBG were purified from pooled plasma containing a minor amount of a subunit with a sequence slightly different from that of the predominant form of the protein, sequence analysis would probably not detect the variant sequence. In contrast, cloning of a cDNA would give one unique species, one that could represent a variant polypeptide. The resolution of the question of whether variation in the primary structure of the SHBG subunit exists must await the isolation of additional SHBG cDNA clones, studies of the SHBG gene and/or the sequencing of individual subunits of SHBG.

When the amino acid sequence of human SHBG was reported SHBG was found not to be homologous to any other protein, thus implying it to be the single member of a new protein family [4]. Shortly afterwards, however, androgen binding protein (ABP) from rat testis was shown to have a high degree of homology to human SHBG [24]. Our comparison of the sequence of human SHBG with the sequences of bovine and human protein S, two proteins recently cloned and sequenced [22,23,25], reveals a homology of human SHBG to protein S, not unexpected in the light of the observation [24] that rat ABP seems to be related to the carboxy-terminal region of protein S.

Protein S is an important cofactor in the anticoagulation system [26] and has connections with the complement system [27]. Structurally, protein S has been subdivided into four different regions, the Gla region, the thrombin-sensitive region, the growth factor region, and the large carboxy-terminal region [22,23]. The homology between human SHBG and protein S is confined to the carboxy-terminal region of protein S, and SHBG is almost comparable in size to this region. It is noteworthy that the four half-cystine residues in SHBG seem to correspond to four of the five half-cystine residues present in the carboxy-terminal region of protein S. These four residues in protein S form disulphide loops within the carboxy-terminal region. By analogy, we suggest that the corresponding half-cystines are disulphide bonded in SHBG, i.e. Cys_{164} to Cys_{188} and Cys_{333} to Cys_{361}.

The lowest degree of homology between SHBG and protein S is seen in the parts corresponding to residues 240 to 290 in SHBG (fig.2). This is noteworthy since this part of the SHBG molecule has been suggested to be responsible for the steroid binding property of SHBG [4]. It contains unusual structural features not recognizable in protein S, such as a long stretch of alternating leucine residues (residues 267–281) and a sequence repeat (residues 277–292 being a repeat of residues 248–263). It is too early to speculate whether this reflects a more pronounced difference in function between these parts of the molecules since the steroid binding site of SHBG is not yet accurately mapped and the function of the carboxy-terminal region of protein S is largely unknown. Nonetheless, the finding of amino acid sequence homology between human SHBG and protein S from both the human and the cow throws new light on these proteins and may help in the elucidation of their function.

ACKNOWLEDGEMENTS

We thank Dr Savio L. Woo for his gift of a human liver λgt11 cDNA library and Professor P.H. Petra for letting us read his pre-publication manuscript on the SHBG sequence. The skillful technical assistance of Ingrid Dahlkvist is gratefully acknowledged. This work was supported by grants from the Swedish Medical Research Council (project no.5913). S.G. received grants from the Fertility Research Institute and Sterling Winthrop Scandinavia.

REFERENCES

[1] Anderson, D.C. (1974) Clin. Endocrinol. 3, 69–79.
[2] Kahn, M.S., Knowles, B.B., Aden, D.P. and Rosner, W. (1981) J. Clin. Endocrinol. Metab. 53, 448–449.
[3] 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.

[4] Walsh, K.H., Titani, K., Takio, K., Kumar, S., Hayes, R. and Petra, P.H. (1986) Biochemistry 25, 7584–7590.
[5] Örstad, A., Lulka, M.F., Eide, B., Petra, P.H. and Ross, J.B.A. (1986) Biochemistry 25, 2686–2692.
[6] Cheng, C.Y., Musto, N.A., Gunsalus, G.L. and Bardin, C.W. (1983) J. Steroid Biochem. 19, 1379–1389.
[7] Hammond, G.L., Robinson, P.A., Sugino, H., Ward, D.N. and Finne, J. (1986) J. Steroid Biochem. 24, 815–824.
[8] Gershagen, S., Henningsson, K. and Fernlund, P. (1987) J. Biol. Chem., in press.
[9] Avvakumov, G.V., Matveentseva, J.V., Akhrem, L.V., Shelchyonok, O.A. and Akhrem, A.A. (1983) Biochim. Biophys. Acta 760, 104–110.
[10] Cheng, C.Y., Musto, N.A., Gunsalus, G.L., Frick, J. and Bardin, C.W. (1985) J. Biol. Chem. 260, 5631–5640.
[11] Fernlund, P., Gershagen, S., Kjellberg, M. and Rannevik, G. (1984) Scand. J. Clin. Lab. Invest. 44, 25–31.
[12] Fernlund, P. and Laurell, C.-B. (1981) J. Steroid Biochem. 14, 545–552.
[13] Young, R.A. and Davis, R.W. (1983) Proc. Natl. Acad. Sci. USA 80, 1194–1198.
[14] Young, R.A. and Davis, R.W. (1983) Science 222, 778–782.
[15] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[16] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[17] Biggin, M.D., Gibson, T.T. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963–3965.
[18] Staden, R. (1982) Nucleic Acids Res. 10, 2951–2961.
[19] Staden, R. (1982) Nucleic Acids Res. 10, 4731–4751.
[20] Orcutt, B.C., Dayhoff, M.O., George, D.G. and Barker, W.C. (1984) Protein Identification Resource, National Biomedical Research Foundation, Georgetown University Medical Center, Washington, DC 20007, USA.
[21] Dahlbäck, B., Lundwall, Å. and Stenflo, J. (1986) J. Biol. Chem. 261, 5111–5115.
[22] Dahlbäck, B., Lundwall, Å. and Stenflo, J. (1986) Proc. Natl. Acad. Sci. USA 83, 4199–4203.
[23] Lundwall, Å., Dackowski, W., Cohen, E., Shaffer, M., Mahr, A., Dahlbäck, B., Stenflo, J. and Wydro, R. (1986) Proc. Natl. Acad. Sci. USA 83, 6716–6720.
[24] Joseph, D.R., Hall, S.H. and French, F.S. (1987) Proc. Natl. Acad. Sci. USA 84, 339–343.
[25] Hoskins, J., Norman, K.D., Beckmann, R.J. and Long, G.L. (1987) Proc. Natl. Acad. Sci. USA 84, 349–353.
[26] Walker, F.J. (1981) J. Biol. Chem. 256, 11128–11131.
[27] Dahlbäck, B. (1984) Semin. Thromb. Hemostasis 10, 139–148.