

Complete amino acid sequence of human placental 17 β -hydroxysteroid dehydrogenase deduced from cDNA

H. Peltoketo, V. Isomaa, O. Mäentausta and R. Vihko

Biocenter and Department of Clinical Chemistry, University of Oulu, SF-90220 Oulu, Finland

Received 4 August 1988; revised version received 7 September 1988

cDNA clones for 17 β -hydroxysteroid dehydrogenase (17-HSD; EC 1.1.1.62) were isolated from a placental λ gt11 expression library using polyclonal antibodies against placental 17-HSD. The largest cDNA contained 1325 nucleotides, consisting of a short 5'-noncoding segment, a coding segment of 987 nucleotides terminated by a TAA codon, and a 329 nucleotide long 3'-noncoding segment. The open reading frame encoded a polypeptide of 327 amino acid residues with a predicted M_r of 34853. The amino acid sequence of 23 N-terminal amino acids determined from purified 17-HSD agreed with the sequence deduced from cDNA. The deduced amino acid sequence also contained two peptides previously characterized from the proposed catalytic area of placental 17-HSD.

Estrogen; 17-Hydroxysteroid dehydrogenase; Molecular cloning; Nucleotide sequence; Amino acid sequence

1. INTRODUCTION

17 β -Hydroxysteroid dehydrogenase (17-HSD; EC 1.1.1.62) catalyzes interconversion of neutral and phenolic 17-hydroxy and 17-ketosteroids. It is present in steroid-forming tissues and also in certain steroid target tissues such as human endometrium [1,2]. In the latter tissue the activity of this enzyme is regulated by progestins [3–5], and it possibly has a role in regulating the exposure of the endometrium to estrogen action [1,2,4]. A similar role for this enzyme in breast tissue has been proposed [6]. Hence detailed information of the properties and regulation of this enzyme may be central for our understanding of the hormone dependency of endometrial and breast cancer.

17-HSD has been purified to homogeneity from term placental tissue in this [7] and other [8,9] laboratories. In its native form, placental 17-HSD is composed of two similar if not identical subunits

having an M_r of about 34000 (see [10]). Preliminary data suggest that human endometrium contains a protein showing similar immunological properties as the one purified from placental tissue [7].

As part of our studies directed towards understanding the regulation of 17-HSD in steroid target tissues, we report the cloning of the cDNA of human placental 17-HSD and provide the complete amino acid structure of the enzyme deduced from the cDNA.

2. MATERIALS AND METHODS

2.1. Materials

A human placental cDNA library in the λ gt11 expression vector was obtained from Clontech (Palo Alto, CA). Restriction endonucleases, proteinase K, DNase I, RNase A, and an M13 cloning kit were obtained from Boehringer Mannheim (Mannheim, FRG). A SequenaseTM sequencing kit was purchased from the United States Biochemical Corporation (Cleveland, OH) and [³⁵S]dATP was from Amersham (Amersham, England). Sequence-derived oligonucleotide primers (17 bases) were synthesized with a DNA synthesizer (Applied Biosystems, Foster City, CA). Nitrocellulose filters were obtained from Schleicher and Schuell (Dassel, FRG).

Correspondence address: H. Peltoketo, Biocenter and Department of Clinical Chemistry, University of Oulu, SF-90220 Oulu, Finland

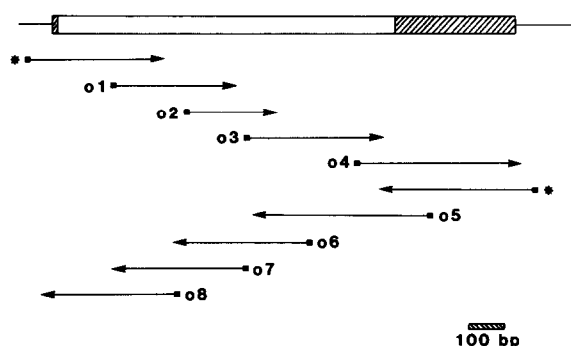


Fig.1. The sequencing strategy. The white area represents the coding sequence of 17-HSD and the hatched areas the 5'- and 3'-untranslated regions. M13 universal primer (*) and sequence specific oligonucleotides (o1-o8) were used to determine the complete sequence of both strands. The bar gives the approximate scale.

2.2. 17-HSD antiserum

Polyclonal antiserum against highly purified placental 17-HSD was raised in rabbits [7]. Antiserum was absorbed with *E. coli* strain Y1090r⁻ lysate to remove antibodies that recognize *E. coli* antigens [11].

2.3. Antibody screening of λ gt11 library

About 5×10^5 plaque-forming units (pfu) of recombinant phages from a placental cDNA library were screened with an anti-17-HSD-antiserum, diluted 1:200 in Tris-buffered saline containing 20% fetal calf serum [11], with the exception that positive plaques were detected by using protein A-peroxidase and 4-chloro-1-naphthol as a substrate. Plaques were further purified by three to five cycles of screening at low plaque density with antiserum until all phages produced positive signals. Eleven positive phage clones were amplified, digested with *Eco*RI, the sizes of the inserts were determined and they were purified.

2.4. cDNA sequence analyses

The *Eco*RI inserts from the λ gt11 clones were subcloned into an M13mp18 vector [12]. The DNA sequence was determined by the dideoxy chain termination method of Sanger [13] using SequenaseTM-enzyme [14] and the universal primer as well as synthetic oligonucleotides for priming (fig.1). The two longest inserts, one later found to have the whole coding area and the other lacking about 40 bases from the 5'-end, were sequenced.

2.5. Protein sequencing

Amino acid sequence analysis of 23 N-terminal amino acids from purified placental 17-HSD was performed with an Applied Biosystems model 477A pulse liquid protein/peptide sequencer with a 120A PTH analyzer (Applied Biosystems, Foster City, CA).

3. RESULTS AND DISCUSSION

A human placental λ gt11 library was screened with antiserum raised against human placental 17-HSD. As a result of screening 5×10^5 individual plaques, 15 clearly positive plaque clones were identified. Eleven plaques remained positive upon three to five rounds of rescreening and they were amplified and digested with *Eco*RI. The sizes of the inserts varied from 0.8 to 1.3 kb and the two longest cDNA inserts (1.2 and 1.3 kb) were sequenced. To compare cDNA-deduced amino acid sequence to the sequence of native placental protein, 23 amino acids from the N-terminus of the purified enzyme were analyzed and the sequences were found to be identical.

The 1.3 kb insert, named λ HSD2A, contained an open reading frame of 987 bp from the initiation codon ATG to the stop codon TAA, and additionally 9 nucleotides of the 5'- and 329 nucleotides of the 3'-noncoding segment (fig.2). The sequence AATAAA, thought to be required for polyadenylation [15], was located 16 bp upstream from the poly(A) tract. The 1.2 kb insert, named λ HSD5B, contained a sequence from nucleotide 40 to nucleotide 1252 of λ HSD2A, suggesting that these two cDNAs were derived from the same RNA-pieces. The codon ATG, preceding the codon of the first amino acid residue, is obviously the initiator codon, since the sequence surrounding it conforms to Kozak's consensus sequence [16].

Three peptide sequences derived from placental 17-HSD have been published previously. The sequence of five N-terminal amino acids determined by Burns et al. [17] differs in the second amino acid; according to our data it is arginine instead of glutamate. Nicolas and Harris [18] analyzed a tryptic heptadecapeptide labeled to a cysteine residue. This peptide, suggested to be part of the coenzyme binding site in 17-HSD [18], is located between amino acid residues 51 and 67. The peptide of 20 amino acid residues sequenced by Murdock et al. [19] is located between residues 204 and 223 (fig.2). The hydropathicity profile of the

Fig.2. Nucleotide and deduced amino acid sequence of the subunit of human placental 17-HSD. The numbering of amino acids is based on the analysis of the N-terminal sequence of the purified protein (underlined). Two known internal peptides are also underlined (solid line). The potential cAMP-dependent kinase phosphorylation site (basic-basic-X-Ser) is indicated with a dot. The putative polyadenylation signal site (AATAAA) is marked with a dashed line.

AGT	CTC	ACC	ATG	GCC	CGC	ACC	GTG	GTG	CTC	ATC	ACC	GGC	TGT	TCC	TCG	GGC	ATC	GGC	CTG	60
				Ala	Arg	Thr	Val	Val	Leu	Ile	Thr	Gly	Cys	Ser	Ser	Gly	Ile	Gly	Leu	16
CAC	TTG	GCC	GTA	CGT	CTG	GCT	TCA	GAT	CCA	TCC	CAG	AGC	TTC	AAA	GTG	TAT	GCC	ACG	TTG	120
His	Leu	Ala	Val	Arg	Leu	Ala	Ser	Asp	Pro	Ser	Gln	Ser	Phe	Lys	Val	Tyr	Ala	Thr	Leu	36
AGG	GAC	CTG	AAA	ACA	CAG	GGC	CGG	CTG	TGG	GAG	GCG	GCC	CGG	GCC	CTG	GCA	TGC	CCT	CCG	180
Arg	Asp	Leu	Lys	Thr	Gln	Gly	Arg	Leu	Trp	Glu	Ala	Ala	Arg	Ala	Leu	Ala	Cys	Pro	Pro	56
GGA	TCC	CTG	GAG	ACG	TTG	CAG	CTG	GAC	GTA	AGG	GAC	TCA	AAA	TCC	GTG	GCC	GCT	GCC	CGG	240
Gly	Ser	Leu	Glu	Thr	Leu	Gln	Leu	Asp	Val	Arg	Asp	Ser	Lys	Ser	Val	Ala	Ala	Ala	Arg	76
GAA	CGC	GTG	ACT	GAG	GGC	CGC	GTG	GAC	GTG	CTG	GTG	TGT	AAC	GCA	GGC	CTG	GGC	CTG	CTG	300
Glu	Arg	Val	Thr	Glu	Gly	Arg	Val	Asp	Val	Leu	Val	Cys	Asn	Ala	Gly	Leu	Gly	Leu	Leu	96
GGG	CCG	CTG	GAG	GCG	CTG	GGG	GAG	GAC	GCC	GTG	GCC	TCT	GTG	CTG	GAC	GTG	AAT	GTA	GTA	360
Gly	Pro	Leu	Glu	Ala	Leu	Gly	Glu	Asp	Ala	Val	Ala	Ser	Val	Leu	Asp	Val	Asn	Val	Val	116
GGG	ACT	GTG	CGG	ATG	CTG	CAG	GCC	TTC	CTG	CCA	GAC	ATG	AAG	AGG	CGC	GGT	TCG	GGA	CGC	420
Gly	Thr	Val	Arg	Met	Leu	Gln	Ala	Phe	Leu	Pro	Asp	Met	Lys	Arg	Arg	Gly	Ser	Gly	Arg	136
GTG	TTG	GTG	ACC	GGG	AGC	GTG	GGA	GGA	TTG	ATG	GGG	CTG	CCT	TTC	AAT	GAC	GTT	TAT	TGC	480
Val	Leu	Val	Thr	Gly	Ser	Val	Gly	Gly	Leu	Met	Gly	Leu	Pro	Phe	Asn	Asp	Val	Tyr	Cys	156
GCC	AGC	AAG	TTC	GCG	CTC	GAA	GGC	TTA	TGC	GAG	AGT	CTG	GCG	GTT	CTG	CTG	CTG	CCC	TTT	540
Ala	Ser	Lys	Phe	Ala	Leu	Glu	Gly	Leu	Cys	Glu	Ser	Leu	Ala	Val	Leu	Leu	Leu	Pro	Phe	176
GGG	GTC	CAC	TTG	AGC	CTG	ATC	GAG	TGC	GGC	CCA	GTG	CAC	ACC	GCC	TTC	ATG	GAG	AAG	GTG	600
Gly	Val	His	Leu	Ser	Leu	Ile	Glu	Cys	Gly	Pro	Val	His	Thr	Ala	Phe	Met	Glu	Lys	Val	196
TTG	GGC	AGC	CCA	GAG	GAG	GTG	CTG	GAC	CGC	ACG	GAC	ATC	CAC	ACC	TTC	CAC	CGC	TTC	TAC	660
Leu	Gly	Ser	Pro	Glu	Glu	Val	Leu	Asp	Arg	Thr	Asp	Ile	His	Thr	Phe	His	Arg	Phe	Tyr	216
CAA	TAC	CTC	GCC	CAC	AGC	AAG	CAA	GTC	TTT	CGC	GAG	GCG	GCG	CAG	AAC	CCT	GAG	GAG	GTG	720
Gln	Tyr	Leu	Ala	His	Ser	Lys	Gln	Val	Phe	Arg	Glu	Ala	Ala	Gln	Asn	Pro	Glu	Glu	Val	236
GCG	GAG	GTC	TTC	CTC	ACC	GCT	TTG	CGC	GCC	CCG	AAG	CCG	ACC	CTG	CGC	TAC	TTC	ACC	ACC	780
Ala	Glu	Val	Phe	Leu	Thr	Ala	Leu	Arg	Ala	Pro	Lys	Pro	Thr	Leu	Arg	Tyr	Phe	Thr	Thr	256
GAG	CGC	TTC	CTG	CCC	CTG	CTG	CGG	ATG	CGC	CTG	GAC	GAC	CCC	AGC	GGC	TCC	AAC	TAC	GTC	840
Glu	Arg	Phe	Leu	Pro	Leu	Leu	Arg	Met	Arg	Leu	Asp	Asp	Pro	Ser	Gly	Ser	Asn	Tyr	Val	276
ACC	GCC	ATG	CAC	CGG	GAA	GTG	TTC	GGC	GAC	GTT	CCG	GCA	AAG	GCC	GAG	GCT	GGG	GCC	GAG	900
Thr	Ala	Met	His	Arg	Glu	Val	Phe	Gly	Asp	Val	Pro	Ala	Lys	Ala	Glu	Ala	Gly	Ala	Glu	296
GCT	GGG	GGC	GGG	GCC	GGG	CCT	GGG	GCA	GAG	GAC	GAG	GCC	GGG	CGC	AGT	GCG	GTG	GGG	GAC	960
Ala	Gly	Gly	Gly	Ala	Gly	Pro	Gly	Ala	Glu	Asp	Glu	Ala	Gly	Arg	Ser	Ala	Val	Gly	Asp	316
CCT	GAG	CTC	GGC	GAT	CCT	CCG	GCC	GCC	CCG	CAG	TAA	AGG	CTT	CCT	CAG	CCG	CTG	TCT	CCC	1020
Pro	Glu	Leu	Gly	Asp	Pro	Pro	Ala	Ala	Pro	Gln	***									327
GCG	CCC	TTC	TTT	GTC	CCC	TGG	GTC	TGT	GTG	GTC	CCT	GGG	GAT	GGG	GCG	GCG	GTA	GCA	GCT	1080
GTG	GGT	GGC	TAA	TTA	AGA	TAG	ATC	GCG	TTA	GCC	AGT	TTT	ACC	AGC	GCA	GCT	AGG	CGC	GAT	1140
GGC	GTC	GCC	TGT	AAT	GCC	AGC	GCT	TTG	GGA	GGC	GGA	GGC	AGG	AGG	ATC	GCT	CAA	GCC	CCG	1200
GAG	TTG	GAG	ACC	AGC	CAG	AGC	AAC	ACA	GTG	AGA	CCC	CCA	TCT	CTA	CAA	AAA	TAA	AGA	AAA	1260
TTT	AAA	AAT	CAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	1320
AAA	AA																			1325

deduced amino acid sequence, showing a hydrophobic pocket in this area (fig.3), supports the conclusion of Murdock et al. [19] that this sequence contains the substrate binding site. This kind of structure is shared by several proteins interacting with steroid hormones [20].

The amino acid sequence deduced from cDNA is predicted to contain 327 amino acid residues with a calculated M_r of 34853. This value is in agreement with values obtained in SDS-PAGE using purified enzyme (see [10]). Of the 327 amino acids, there are 6 cysteine residues, 40 acid residues, 32 basic residues, and 20 aromatic residues (table 1). The amino acid composition calculated from the cDNA-deduced sequence agrees well with that obtained by direct chemical analysis of the purified protein [17,21,22].

The amino acid sequence of 17-HSD we have deduced from cDNA does not contain potential Asn-X-Ser/Thr N-glycosylation sites [23], confirming the finding that 17-HSD is not a glycoprotein [10,24]. One potential cyclic AMP-dependent phosphorylation site (base-base-X-Ser) [25] is presented in the 17-HSD sequence at position 131-134 (fig.1). Varying degrees of possible phosphorylation of 17-HSD may explain the microheterogeneity found in isoelectric focusing of the purified enzyme [26].

Table 1

The amino acid composition of the subunit of 17-HSD

	Residues per polypeptide deduced from cDNA	Burns [17]	Jarabak [21] ^a	Nicolas [22] ^a
Ala	38	36	54	54
Arg	23	22	37	34-36
Asn	5			
Asp	17	21	40	54-56
Cys	6	6	10	12
Gln	8			
Glu	23	28	51	68-70
Gly	32	31	50	48
His	7	7	11	14
Ile	4	4	7	20
Leu	42	41	66	66
Lys	9	10	20	30
Met	6	4	5	12
Phe	13	13	20	24
Pro	20	19	31	40
Ser	19	19	34	40
Thr	16	16	29	32-34
Trp	1	1	2	4
Tyr	6	6	9	16
Val	32	32	47	52
N =	327	316	523	624 ± 4

^a Residues per enzyme molecule assuming that 17-HSD is composed of two identical subunits

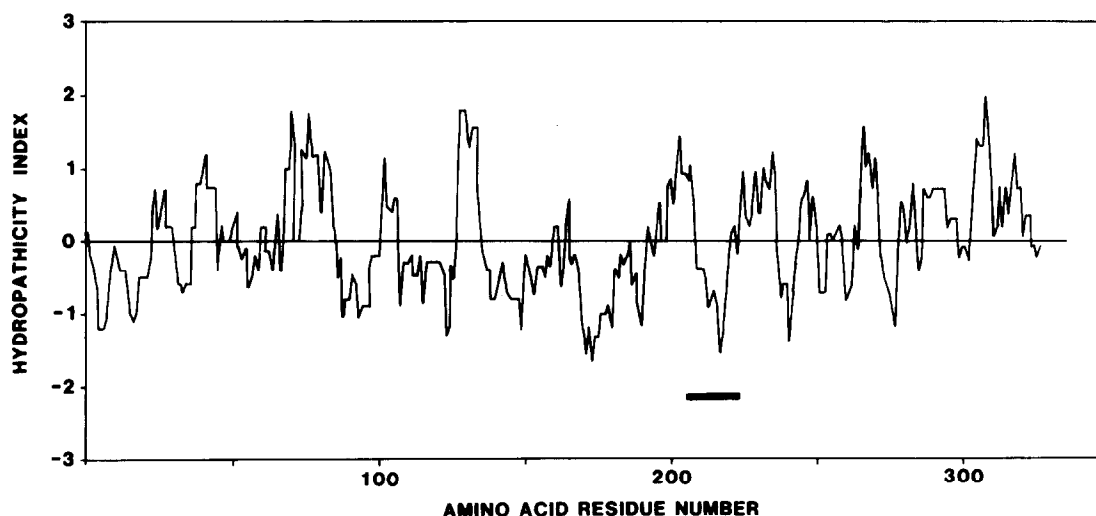


Fig.3. Hydropathicity profile of the deduced amino acid sequence of human placental 17-HSD, determined according to Hopp and Woods [27]. The area thought to be involved in substrate binding is indicated by a bar. Negative values indicate hydrophobic regions of the protein.

A computer homology search of the cDNA of 17-HSD and the amino acid sequence of the derived protein was performed against The Genetic Sequence Data Bank (GenBank) and The National Biomedical Research Foundation (NBRF) Protein Data Bank. No significant homology with other proteins or DNA sequences was found.

Our cloning of the cDNA of human placental 17-HSD will be followed by studies on the gene structure, which should facilitate studies on the hormonal regulation of this enzyme in a number of hormone-responsive tissues, such as human endometrium and breast epithelium. In these tissues, 17-HSD may have important physiological regulator functions, and it may also be important for our understanding of the behavior of cancer in these tissues.

Acknowledgements: We thank Dr Annika Linnala-Kankkunen (Department of Biochemistry, University of Kuopio) for analysing the N-terminal sequence of the subunit of 17-HSD. We also thank Mr Pekka Kallunki (Department of Biochemistry, University of Oulu) for preparing the synthetic oligonucleotides used for the DNA sequencing, and Ms Birgitta Grekula and Ms Anne Karjalainen for skillful technical assistance. This work was supported by the Research Council for Medicine of the Academy of Finland.

REFERENCES

- [1] Pollow, K., Lübbert, H., Boquoi, E., Kreutz Jeske, R. and Pollow, B. (1975) *Acta Endocrinol.* 79, 134–145.
- [2] Tseng, L., Gusberg, S.B. and Gurpide, E. (1977) *Ann. NY Acad. Sci.* 286, 190–198.
- [3] Tseng, L. and Gurpide, E. (1975) *Endocrinology* 97, 825–833.
- [4] Tseng, L. (1980) in: *Advances in Steroid Hormone Research*, vol.IV (Thomas, J.A. and Singhal, R.L. eds) pp.329–361, Urban and Schwarzenberg, Baltimore.
- [5] Neumannova, M., Kauppila, A., Kivinen, S. and Vihko, R. (1985) *Obstet. Gynecol.* 66, 695–700.
- [6] Fournier, S., Brihmat, F., Durand, J.C., Sterkers, N., Martin, P.M., Kuttan, F. and Mauvais-Jarvis, P. (1985) *Cancer Res.* 45, 2895–2899.
- [7] Mäentausta, O., Peltoketo, H., Isomaa, V. and Vihko, R., submitted.
- [8] Karavolas, H.J., Baedeker, M.L. and Engel, L.L. (1970) *J. Biol. Chem.* 245, 4948–4952.
- [9] Chin, C.C. and Warren, J.C. (1973) *Steroids* 22, 373–378.
- [10] Pons, M., Nicolas, J.C., Bossioux, A.M., Descomps, B. and Crastes de Paulet, A. (1977) *J. Steroid Biochem.* 8, 345–358.
- [11] Huynh, T.V., Young, R.A. and Davis, R.W. (1985) in: *DNA Cloning: A Practical Approach* 1 (Glover, D.M. ed.) pp.49–78, IRL Press, Oxford.
- [12] Messig, J. (1983) *Methods Enzymol.* 101, 20–78.
- [13] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
- [14] Tabor, S. and Richardson, C.C. (1987) *Proc. Natl. Acad. Sci. USA* 84, 4767–4771.
- [15] Proudfoot, N.J. and Brownlee, G.G. (1973) *Nature* 252, 359–362.
- [16] Kozak, M. (1986) *Cell* 44, 283–292.
- [17] Burns, D.J.W., Engel, L.L. and Bethune, J.L. (1972) *Biochemistry* 11, 2699–2703.
- [18] Nicolas, J.C. and Harris, J.I. (1973) *FEBS Lett.* 29, 173–176.
- [19] Murdock, G.L., Chin, C.C. and Warren, J.C. (1986) *Biochemistry* 25, 641–646.
- [20] Carlstedt-Duke, J., Strömstedt, P.-E., Persson, B., Cederlund, E., Gustafsson, J.-Å and Jörnvall, H. (1988) *J. Biol. Chem.* 263, 6842–6846.
- [21] Jaraback, J. and Street, M.A. (1971) *Biochemistry* 10, 3831–3834.
- [22] Nicolas, J.C., Pons, M., Descomps, B. and Crastes de Paulet, A. (1972) *FEBS Lett.* 23, 175–179.
- [23] Bause, E. (1983) *Biochem. J.* 209, 331–336.
- [24] Burns, D.J.W., Engel, L.L. and Bethune, J.L. (1971) *Biochem. Biophys. Res. Commun.* 44, 786–792.
- [25] Krebs, E.G. and Beavo, J.A. (1979) *Annu. Rev. Biochem.* 48, 923–959.
- [26] Engel, L.L. and Groman, E.V. (1974) *Recent Prog. Horm. Res.* 30, 139–169.
- [27] Hopp, T.P. and Woods, K.R. (1981) *Proc. Natl. Acad. Sci. USA* 78, 3824–3828.