Volume 222, number 1, 186–190 FEB 05149 September 1987

Human protein S cDNA encodes Phe-16 and Tyr 222 in consensus sequences for the post-translational processing

Hans K. Ploos van Amstel, A. Linda van der Zanden, Pieter H. Reitsma and Rogier M. Bertina

Hemostasis and Thrombosis Research Unit, Dept of Hematology, Leiden University Hospital, Leiden, The Netherlands

Received 29 July 1987

Partial cDNAs coding for human protein S were isolated from a pUC9 human liver cDNA library. Together, the overlapping clones span a (partial) 5'-non-coding region, and the complete protein S coding and 3'-untranslated regions. The derived amino acid sequence deviates at five positions from two previously reported protein S sequences. Two of these differences (Phe instead of Leu at position −16 and Tyr instead of Asp at position 222) are found in regions that are important for the post-translational modification of protein S, the γ-carboxylation of glutamic acid and the hydroxylation of asparagine, respectively.

Protein S; cDNA; Carboxylation; Hydroxylation

1. INTRODUCTION

Human protein S is a vitamin K-dependent glycoprotein [1] circulating in the plasma both 'free' and bound to C4b-binding protein, a component of the complement system [2]. In its free form protein S serves as a cofactor of activated protein C [3], thereby enhancing both the inactivation of the coagulation factors Va [4] and VIIIa [5] and the stimulation of fibrinolysis by activated protein C [6]. The important physiological role of protein S in the anticoagulant pathway is illustrated by the finding that hereditary protein S deficiency is associated with an increased risk of developing thrombotic disease at a relatively young age [7].

Before the mature protein appears in the circulation four post-translational modifications occur. The signal and propeptide are cleaved off, key glutamic acids are γ-carboxylated, specific aspartic acid and asparagines are β-hydroxylated and carbohydrate side chains are added.

The reported sequences of human protein S [8,9] deviate at seemingly crucial positions from the postulated consensus sequences for (1) the γ-carboxylation of glutamic acids [10] and (2) the β-hydroxylation of asparagine in the fourth EGF-domain [11].

This report describes the isolation of human protein S cDNAs. Our data indicate differences with earlier reported sequences at five positions including two in the consensus sequences.

Correspondence address: J.K. Ploos van Amstel, Hemostasis and Thrombosis Research Unit, Dept of Hematology, University Hospital, Bldg 1, C2-R, PO Box 9600, 2300 RC Leiden, The Netherlands

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under the accession number Y00692

2. MATERIALS AND METHODS

2.1. *Materials*

Restriction enzymes were obtained from Promega Biotec (Madison, USA). Nitrocellulose BA-85 was purchased from Schleicher and Schuell (Dassel, FRG). The nick-translation kit, M13 sequencing kit, [γ-^{32}P]dATP (>3000 Ci/mmol) and [α-^{35}S]dATP (>600 Ci/mmol) were obtained from Amersham International (Amersham, England).

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

[α-^{32}P]dCTP (>3000 Ci/mmol) was obtained from New England Nuclear (Boston, USA).

2.2. *cDNA cloning*

An amplified human liver cDNA library in plasmid pUC9 (kindly provided by H. Pannekoek, Central Laboratory of the Netherlands Red Cross Blood Transfusion Service, Amsterdam, The Netherlands), that contained 3×10^5 independent recombinants with an average length insert of 1500 nucleotides, was screened by in situ hybridization [12].

A pilot screening of 2×10^4 recombinants was performed with a [γ-^{32}P]dATP radiolabelled 27-nucleotide-long oligomer, the sequence of which was derived from a positive clone isolated from a human liver cDNA expression library in λgt11 screened with immunopurified polyclonal anti protein S IgG (Ploos van Amstel et al., in press). One positive clone, designated pSUL1 (fig.1) was obtained and used as a probe to screen an additional 2×10^5 recombinants of the pUC9 library at high density. Eight different positive clones were isolated, designated pSUL1-8 (fig.1). The coding region of the clone with the largest insert, pSUL5, was used to screen an additional 1×10^5 recombinants, resulting in the isolation of pSUL9 (fig.1). The 5′-*PstI*/*XbaI* fragment of pSUL9 was used to rescreen the filters. Two positive clones, pSUL10 and 11 could be isolated (fig.1).

Labelling of the cDNA fragments was performed by nick-translation [13] using [α-^{32}P]dCTP.

2.3. *Nucleotide sequence determination*

Restriction fragments of the positive clones were subcloned in M13 mp18 and mp19. Sequencing was performed with the dideoxy chain termination reaction [14] using [α-^{35}S]dATP. Furthermore, oligonucleotides were synthesized (Cyclone, Brunswick, USA) and used as specific primers in the sequencing reactions. The sequencing strategy is outlined in fig.1.

3. RESULTS AND DISCUSSION

Screening 3×10^5 recombinants from an amplified human liver cDNA library in pUC9 with a human protein S specific oligonucleotide and cDNA fragments resulted in the isolation of eleven protein S partial cDNAs, designated pSUL1-11 (fig.1). Determination of the nucleotide sequence (the strategy is outlined in fig.1) showed that, together, the clones encompass a 120-nucleotide-long 5′-non-coding region, the region encoding the protein S precursor and the complete 3′-untranslated region, as indicated by the presence of two polyadenylation signals and a poly(A) tail (not shown). In fig.2 the nucleotide sequence is shown of the 5′-non-coding region and the part of the cDNA coding for the human protein S leader peptide. The 5′-non-coding region is 12 nucleotides longer than the sequence reported by Hoskins et al. [9] whereas the sequence reported by Lundwall et al. [8] starts at amino acid position −13. Northern blot analysis of human liver RNA showed a protein S mRNA species of approx. 4 kb (not shown), suggesting a longer 5′-non-coding region than the one given in fig.2. The 5′-non-coding region is followed by an open reading frame of 2028 nucleotides encoding the protein S precursor. The derived amino acid sequence shows a leader peptide (residues −41 to −1) (fig.2), the vitamin K-dependent domain (1–41), a region sensitive to thrombin cleavage (42–75), four EGF domains (76–248) and a carboxy terminal loop (249–635) that shows significant homology with

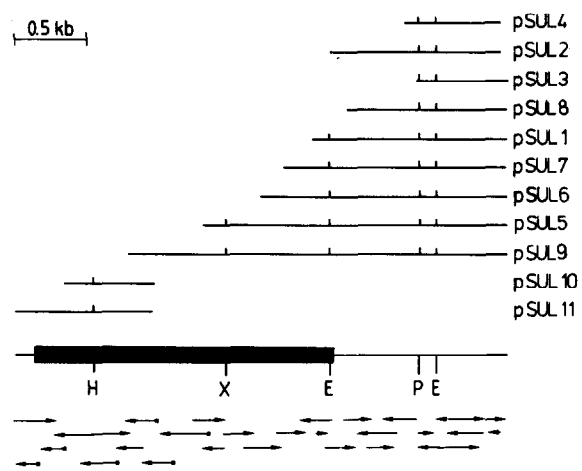

Fig.1. Partial restriction map and sequencing strategy of the human protein S cDNAs. The coding region is indicated by the black box. Arrows indicate the direction and extension of the dideoxy chain termination reactions. Closed squares indicate sequencing reactions primed by specific oligonucleotides. E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; X, *Xba*I.

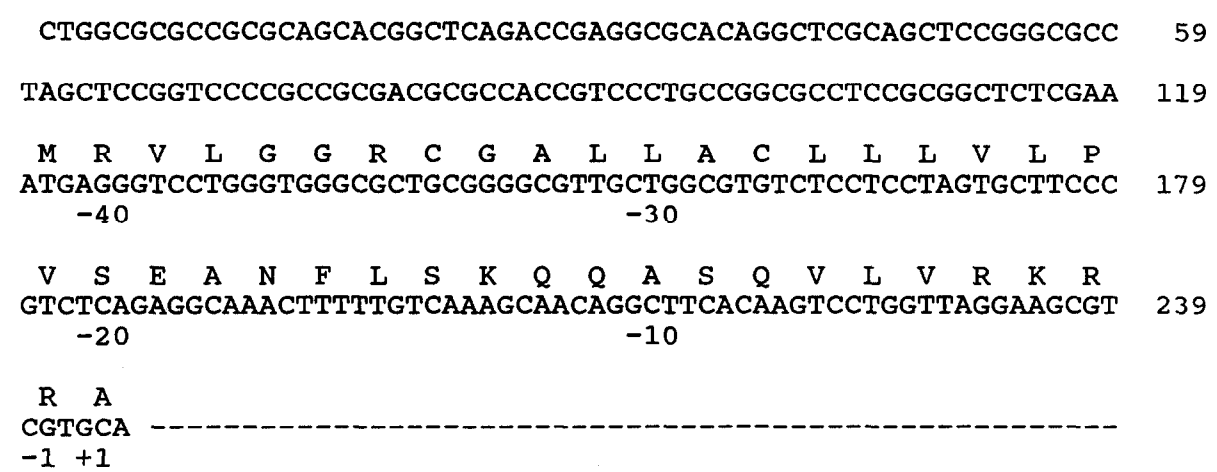

Fig.2. Nucleotide sequence and the derived amino acid sequence of the 5'-region of clone pSUL11.

the androgen binding protein of the rat [15]. The protein S coding region is followed by a 3'-untranslated region 1134-nucleotides long ending in a poly(A) tail.

The derived amino acid sequence deviates at five positions (table 1) from the previous reported sequences [8,9]. All differences involve thymine and never adenine. The amino acids of these positions predicted from our cDNA sequence are identical to those for bovine protein S in the corresponding positions [16]. The differences found at position −16 and 222 are of special interest because they both are noted to be involved in the post-translational modification of the protein.

Table 1

Codon and amino acid differences between reported protein S cDNAs and derived amino acid sequences

Residue	Protein S			
	Bovine	Human		
	(1)	(2)	(3)	(4)
−31	CTG Leu	–	CCT Pro	TTG Leu
−16	TTT Phe	–	CTT Leu	TTT Phe
180	CCA Pro	CTA Leu	CCA Pro	CCA Pro
222	TAC Tyr	CAC His	TAC Tyr	TAC Tyr
304	GAT Asp	TAT Tyr	GAT Asp	GAT Asp

(1) Dahlbäck et al. [16]; (2) Lundwall et al. [8]; (3) Hoskins et al. [9]; (4) This report

The Phe at position −16 is conserved in the propeptides of all the known vitamin K-dependent proteins (fig.3) and is crucial for the γ-carboxylation of the glutamic acids. Jorgenssen et al. [10] showed for factor IX that substitution of Phe −16 by Ala destroys the recognition site of the carboxylase thereby preventing the enzyme from carboxylating the adjacent glutamic acids. As a consequence the molecule cannot adopt the metal-dependent conformation, which is essential for its biological activity. We may therefore assume that a Leu at position −16 [9] would result in a non-carboxylated protein S, which does not agree with the presence of carboxylated protein S in human plasma [1].

A Tyr at position 222 fits well in the consensus sequence (fig.4) proposed by Stenflo et al. [11] as the minimum requirement for the β-hydroxylation of aspartic acid and asparagine residues in the EGF domains. Interestingly, structure analysis of EGF domains [17] indicate that the Tyr/Phe residue lies opposite to the modified Asp/Asn.

Human protein S was reported to contain three hydroxylated residues [8] whereas bovine protein S has four hydroxylated residues [8,11]. It was hypothesized by Stenflo et al. [11] that in human protein S the His at position 222 would disturb the consensus sequence and might be the reason for

	-16						-10				-6				-1	+1
Prothrombin	V	F	L	A	P	Q	Q	A	R	S	L	L	Q	R	V	R	R	A
Factor VII	V	F	V	T	E	E	E	A	H	G	V	L	H	R	R	R	R	A
Factor IX	V	F	L	D	H	E	N	A	N	K	I	L	N	R	P	K	R	Y
Factor X	L	F	I	R	R	E	Q	A	N	N	I	L	A	R	V	I	R	A
Protein C	V	F	S	S	S	E	R	A	H	Q	V	L	R	I	R	K	R	A
Protein S	N	F	L	S	K	Q	Q	A	S	Q	V	L	V	R	K	R	R	A

Fig.3. Comparison of the amino acid sequences for the propeptide regions of the human vitamin K-dependent proteins. The amino acid sequences of prothrombin, factor VII, factor IX, factor X and protein C were obtained from Jorgenssen et al. [10], of protein S from this report. The amino termini of the mature proteins are indicated by +1.

the absence of one of the four hydroxylated residues as compared with bovine protein S. However, the presence of a Tyr at position 222 in both our sequence and that reported by Hoskins et al. [9] does not support this hypothesis. Sequence comparison for the four EGF domains of bovine and human protein S (fig.4) shows such a high degree of homology, that it is not clear why human protein S should contain just three hydroxylated residues. Whether a polymorphic population of human protein S exists (containing either three or four hydroxylated residues) or whether other factors are responsible for the partial hydroxylation of human protein S remains to be established.

91																106	
	M	T	C	K	D	G	Q	A	T	F	T	C	I	C	K	S	Bovine
	M	S	C	K	D	G	K	A	S	F	T	C	T	C	K	P	Human
132																147	
	Q	I	C	E	N	T	P	G	S	Y	H	C	S	C	K	N	Bovine
	Q	I	C	D	N	T	P	G	S	Y	H	C	S	C	K	N	Human
174																189	
	A	V	C	K	N	I	P	G	D	F	E	C	E	C	A	E	Bovine
	A	V	C	K	N	I	P	G	D	F	E	C	E	C	P	E	Human
213																228	
	Q	L	C	V	N	Y	P	G	G	Y	S	C	Y	C	D	G	Bovine
	Q	L	C	V	N	Y	P	G	G	Y	T	C	Y	C	D	G	Human
	-	-	C	-	N D	-	-	-	-	Y F	-	C	-	C	-	-	Consensus

Fig.4. Comparison of the amino acid sequences for the EGF domains of bovine [16] and human protein S (this report). The numbers indicate the residue positions of the mature proteins. The consensus sequence (bottom) for β-hydroxylation is based on a large survey of EGF domains [11].

ACKNOWLEDGEMENTS

We thank Mrs M.J. Mentink for her help in preparing the manuscript. This work was supported by a grant from the Trombose Stichting Nederland (no. 86.015).

REFERENCES

[1] DiScipio, R.G. and Davie, E.W. (1979) Biochemistry 18, 899–904.
[2] Dahlbäck, B. and Stenflo, J. (1981) Proc. Natl. Acad. Sci. USA 78, 2512–2516.
[3] Bertina, R.M., Van Wijngaarden, A., Reinalda-Poot, J., Poort, S.R. and Bom, V.J.J. (1985) Thromb. Haemost. 53, 268–272.
[4] Walker, F.J. (1980) J. Biol. Chem. 255, 5521–5524.
[5] Gardiner, J.E., McGarn, M.A., Berridge, C.W., Fulchner, C.A., Zimmerman, T.S. and Griffin, J.H. (1984) Circulation 70, suppl.II, 205.
[6] De Fouw, N.J., Haverkate, F., Bertina, R.M., Koopman, J., Van Wijngaarden, A. and Van Hinsbergh, V.M.M. (1986) Blood 67, 1189–1192.
[7] Bertina, R.M. (1985) Haemostasis 15, 241–246.
[8] Lundwall, A., Dackowski, W., Cohen, E., Shaffer, M., Mahr, A., Dahlbäck, B., Stenflo, J. and Wydro, R. (1986) Proc. Natl. Acad. Sci. USA 83, 6716–6720.
[9] Hoskins, J., Norman, D.K., Beckman, R.J. and Long, G.L. (1987) Proc. Natl. Acad. Sci. USA 84, 349–353.
[10] Jorgenssen, M.J., Cantor, A.B., Furie, B.C., Brown, C.L., Shoemaker, C.B. and Furie, B. (1987) Cell 48, 185–191.
[11] Stenflo, J.H., Lundwall, A. and Dahlbäck, B. (1987) Proc. Natl. Acad. Sci. USA 84, 368–372.
[12] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, NY.
[13] Rigby, P.W.J., Diekmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237–254.
[14] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[15] Joseph, D.R., Hall, S.H. and French, F.S. (1987) Proc. Natl. Acad. Sci. USA 84, 339–343.
[16] Dahlbäck, B., Lundwall, A. and Stenflo, J. (1986) Proc. Natl. Acad. Sci. USA 83, 4199–4203.
[17] Cooke, R.M., Wilkinson, A.J., Baron, M., Pastore, A., Tappin, M.J., Campbell, I.D., Gregory, H. and Sheard, B. (1987) Nature 327, 339–341.