Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 4199–4203, June 1986
Biochemistry

Primary structure of bovine vitamin K-dependent protein S

BJÖRN DAHLBÄCK, ÅKE LUNDWALL, AND JOHAN STENFLO

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

Communicated by Jan G. Waldenström, February 18, 1986

ABSTRACT **Protein S is a vitamin K-dependent plasma protein that functions as a cofactor to activated protein C in the inactivation of coagulation factors Va and VIIIa. The nucleotide sequence of a full-length cDNA clone, obtained from a bovine liver library, was determined and the amino acid sequence was deduced. In addition, 95% of the structure was determined by protein sequencing. Protein S consists of 634 amino acids in a single polypeptide chain and has one asparagine-linked carbohydrate side chain. The cDNA sequence showed that the protein has a leader sequence, 41 amino acid residues long. The amino-terminal part of the molecule containing γ-carboxyglutamic acid is followed by a region, residues 42–75, with two peptide bonds that are very sensitive to cleavage by thrombin. Residues 76–244 have four cysteine-rich repeat sequences, each about 40 residues long, that are homologous to the precursor of mouse epidermal growth factor. In contrast to the other vitamin K-dependent plasma proteins, the carboxyl-terminal part of protein S is not homologous to the serine proteases.**

Protein S, which has been purified from human and bovine plasma, is a vitamin K-dependent protein that, unlike the other vitamin K-dependent clotting factors, is not a serine protease (1–4). It functions as a cofactor when activated protein C, a regulator of blood clotting, degrades factors Va and VIIIa by limited proteolysis (5, 6). The physiological importance of the protein S cofactor function is indicated by a predisposition to venous thrombosis in patients with hereditary protein S deficiency (7, 8). The concentration of protein S in human plasma is around 25 μg/ml, half of which exists in a noncovalent complex with the regulatory complement protein C4b-binding protein (C4BP) (9, 10), suggesting a role, as yet unknown, for protein S in the regulation of the complement system (11).

Protein S is a single-chain molecule of molecular weight ≈80,000 (1–4). Like the other vitamin K-dependent plasma proteins, it contains a γ-carboxyglutamic acid (Gla)-rich amino-terminal domain (1–3, 9), close to which is located another modified residue, β-hydroxyaspartic acid (12, 13). Protein S is very sensitive to proteolysis by thrombin (4), which cleaves the protein first at Arg-70 and then at Arg-52 (14). In the thrombin-cleaved protein, the Gla-rich domain (residues 1–45) is linked to the large carboxyl-terminal fragment by a disulfide bond (14). Thrombin-cleaved protein S, which has no cofactor activity, has a lower affinity for Ca^{2+} than has the intact protein (4, 15) and has also lost its affinity for negatively charged phospholipid surfaces (16). We now report the amino acid sequence of bovine protein S. Protein sequencing provided 95% of the sequence, and we have also sequenced a cDNA that encodes the entire protein S molecule. The half-cystine residues have been tentatively paired.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Chemicals used for amino acid analysis were from Beckman; those for protein sequence determination, from Applied Biosystems (Foster City, CA); and those for HPLC, from Ratburn. The same enzymes were used for the proteolytic degradation of protein S as previously reported (14). The restriction enzyme *Pst* I was from New England Biolabs; the Klenow fragment of DNA polymerase I was from Boehringer Mannheim; [γ-^{32}P]ATP (3000 Ci/mmol; 1 Ci = 37 GBq), [α-^{32}P]dCTP (3000 Ci/mmol), [α-[^{35}S]thio]dATP (600 Ci/mmol), and the nick-translation kit were from Amersham. Deoxynucleotides, dideoxynucleotides, T4 DNA polymerase, polynucleotide kinase, and the replicative form of the phage M13 mp8 were from Pharmacia-PL Biochemicals. Nitrocellulose filters (HATF 082) were from Millipore. A bovine liver cDNA library in pBR322, containing 10^5 different recombinants, and synthetic oligonucleotides, were kindly provided by R. M. Wydro (Integrated Genetics, Inc., Framingham, MA). The 17-nucleotide universal primer was a kind gift from P. Lind (Department of Cell Research, Wallenberg Laboratory, Uppsala).

Isolation of cDNA Clones and DNA Sequencing. Mixed oligonucleotide probes encoding bovine protein S were radiolabeled with [γ-^{32}P]ATP, by use of polynucleotide kinase, to a specific activity of 0.5–1.0 μCi/pmol. Approximately 25 μCi of the labeled probe was used for screening 20 nitrocellulose filters in 20 ml of hybridization solution, with a modified Grunstein–Hogness protocol (17). Hybridizing clones were colony-purified, and the plasmid DNA was isolated. cDNA inserts were subcloned in vector M13 mp8 and sequenced by the modified dideoxy sequencing technique of Biggin *et al.* (18). The cDNA inserts of positive clones were labeled by nick-translation and used in a second round of screening. The sequence of the full-length clone was obtained from subclones generated from sonicated DNA, as described by Bankier and Barrell (19). Nucleotide sequences were aligned and analyzed with computer programs (20, 21) kindly provided by R. Staden (Medical Research Council Unit, Cambridge, UK).

Protein Sequence Determination. Protein S fragments were generated by enzymatic and chemical cleavages and the fragments were purified by gel filtration and reversed-phase high-performance liquid chromatography (HPLC) as described (14). Trypsin cleavage was performed on the reduced and carboxymethylated protein, after modification of lysine residues by succinylation. The Glu-specific enzyme from *Staphylococcus aureus* V8 was used in an attempt to generate large fragments of the intact protein. After an initial gel filtration, the fragments were reduced and carboxymethylated and again separated by gel filtration and/or HPLC. The V8 enzyme was also used for digestion of larger fragments obtained by cyanogen bromide or trypsin cleavage. Fragments formed by digestion with thrombin or chymotrypsin and by partial acid hydrolysis were prepared as described

Abbreviations: EGF, epidermal growth factor; Gla, γ-carboxyglutamic acid.

(14). Cyanogen bromide fragments were prepared from reduced and carboxymethylated protein S and from intact protein S. Cyanogen bromide cleavage of the intact protein was performed to determine the disulfide bond arrangement in the carboxyl-terminal part of protein S. The lyophilized cyanogen bromide fragments were dissolved in 0.1 M NH_4HCO_3. After centrifugation, the clear supernatant was chromatographed on a column (1.5 × 90 cm) packed with Sephadex G-75 (Pharmacia) in 0.1 M NH_4HCO_3. A major UV-absorbing peak was eluted close to the void volume and was further purified by HPLC. The pellet was dissolved in 0.05 M Tris Cl/0.1 M NaCl, pH 7.5, containing 6 M guanidine hydrochloride and was chromatographed on a column (1.5 × 90 cm) packed with Sephacryl S-200 in the same buffer. A major UV-absorbing peak was followed by a smaller peak. The material in the latter peak was further purified by HPLC. Amino acid sequence was determined on a Beckman 890C sequencer or on an Applied Biosystems gas-phase sequencer, and the amino acid composition of peptides on a Beckman 6300 amino acid analyzer, as described (14).

RESULTS AND DISCUSSION

Amino Acid and Nucleotide Sequence of Bovine Protein S. The amino acid sequence determination of protein S fragments generated nine nonoverlapping sequence segments covering 600 amino acids, out of a total of 634 (Fig. 1). The order of the segments in the sequence could be tentatively arranged. Approximately 130 fragments were analyzed either in the spinning cup or the gas-phase sequencer. Most of protein S was sequenced at least twice, though Fig. 2 shows only those fragments necessary to deduce the sequence. The parts of protein S that were not covered by amino acid sequence data were residues 127–136, 314, 473–482, 515–519, 540–543, and 566–569. In addition, the two overlaps 146–147 and 391–392 were missing. The positions of the 11 Gla residues and that of the single hydroxyaspartic acid residue have already been reported (14).

Two mixed oligodeoxynucleotide probes were synthesized; probe 1 was a 23-mer (5′ GTCATRAANCCRTCYTCRTTRCA 3′; Y = C or T, R = A or G, N = A, C, G, or T) and probe 2 a 20-mer (5′ ACYTCCATRCANCCYTGRTA 3′)

```
                                  Met Arg Val Leu Gly Gly Arg Thr Gly Thr Leu Leu Ala Cys Leu Ala Leu Val Leu Pro Val Leu Glu Ala Asn Phe
CCGCCGCTTCTGCCCAAGCCTCCGCCCGTTTCGCC ATG AGG GTC CTA GGT GGG CGC ACC GGG ACG CTG CTG GCA TGC CTC GCC CTA GTG CTT CCC GTC TTG GAG GCT AAC TTT  113

Leu Ser Arg Gln His Ala Ser Gln Val Leu Ile Arg Arg Arg Arg Ala Asn Thr Leu Leu Gla Gla Thr Lys Lys Gly Asn Leu Gla Arg Gla Cys Ile Gla Gla   20
TTG TCG AGG CAA CAT GCT TCA CAA GTC CTG ATT AGG AGA CGC CGT GCA AAT ACA TTG CTT GAA GAA ACG AAA AAG GGC AAT CTT GAA AGA GAA TGC ATT GAA GAA  218

Leu Cys Asn Lys Gla Gla Ala Arg Gla Ile Phe Gla Asn Asn Pro Gla Thr Glu Tyr Phe Tyr Pro Lys Tyr Leu Gly Cys Leu Gly Ser Phe Arg Ala Gly Leu   55
CTG TGC AAT AAA GAA GAA GCC AGG GAA ATC TTT GAA AAT AAC CCG GAA ACG GAA TAT TTT TAT CCA AAA TAT TTA GGT TGT CTT GGC TCT TTC AGA GCT GGA TTA  323

Phe Thr Ala Ala Arg Leu Ser Thr Asn Ala Tyr Pro Asp Leu Arg Ser Cys Val Asn Ala Ile Ser Asp Gln Cys Asn Pro Leu Pro Cys Asn Glu Asp Gly Phe   90
TTC ACT GCT GCT CGT CTG TCA ACT AAT GCT TAC CCT GAC CTG AGG AGC TGT GTC AAT GCC ATT TCG GAC CAG TGT AAT CCT CTG CCA TGC AAT GAA GAT GGA TTT  428

Met Thr Cys Lys Asp Gly Gln Ala Thr Phe Thr Cys Ile Cys Lys Ser Gly Trp Gln Gly Glu Lys Cys Glu Ser Asp Ile Asn Glu Cys Lys Asp Pro Val Asn  125
ATG ACC TGC AAA GAT GGC CAA GCG ACA TTC ACT TGC ATT TGT AAA TCA GGT TGG CAA GGA GAA AAG TGT GAA TCT GAT ATA AAT GAA TGC AAA GAT CCT GTA AAT  533

Ile Asn Gly Gly Cys Ser Gln Ile Cys Glu Asn Thr Pro Gly Ser Tyr His Cys Ser Cys Lys Asn Gly Phe Val Met Leu Ser Asn Lys Lys Asp Cys Lys Asp  160
ATA AAT GGA GGT TGC AGC CAG ATT TGT GAA AAC ACA CCT GGA AGT TAC CAC TGT TCC TGT AAA AAT GGT TTT GTT ATG CTT TCA AAT AAA AAG GAC TGC AAA GAT  638

Val Asp Glu Cys Val Leu Lys Pro Ser Ile Cys Gly Thr Ala Val Cys Lys Asn Ile Pro Gly Asp Phe Glu Cys Glu Cys Ala Glu Gly Tyr Lys Tyr Asn Pro  195
GTG GAT GAA TGC GTT TTA AAG CCA AGC ATT TGT GGC ACA GCT GTG TGC AAG AAC ATC CCA GGA GAC TTT GAA TGT GAA TGT GCT GAA GGC TAC AAA TAC AAT CCC  743

Val Ser Lys Ser Cys Asp Asp Val Asp Glu Cys Ala Glu Asn Leu Cys Ala Gln Leu Cys Val Asn Tyr Pro Gly Gly Tyr Ser Cys Tyr Cys Asp Gly Lys Lys  230
GTA TCA AAG TCT TGT GAC GAT GTG GAT GAA TGC GCT GAG AAC TTG TGT GCT CAA CTT TGT GTC AAT TAC CCT GGA GGT TAC TCT TGT TAC TGT GAC GGA AAG AAA  848

Gly Phe Lys Leu Ala Gln Asp Gln Lys Ser Cys Glu Ala Val Pro Val Cys Leu Pro Leu Asp Leu Asp Lys Asn Tyr Glu Leu Leu Tyr Leu Ala Glu Gln Phe  265
GGA TTC AAA CTT GCC CAA GAT CA G AAG AGT TGT GAG GCT GTT CCA GTG TGC CTT CCT TTG GAC CTT GAC AAA AAT TAT GAA TTG CTT TAC TTG GCA GAG CAG TTT  953

Val Gly Val Val Leu Tyr Leu Lys Phe Arg Leu Pro Glu Thr Thr Arg Phe Ser Ala Glu Phe Asp Phe Arg Thr Tyr Asp Ser Glu Gly Val Ile Leu Tyr Ala  300
GTA GGG GTT GTT TTG TAT TTA AAA TTT CGT TTG CCA GAA ACT ACC AGA TTT TCA GCT GAA TTT GAT TTC CGG ACA TAT GAT TCA GAA GGT GTT ATC CTG TAT GCA 1058

Glu Ser Ser Asp His Ser Ala Trp Phe Leu Ile Ala Leu Arg Glu Gly Lys Ile Glu Ile Gln Phe Lys Asn Glu Lys Thr Thr Lys Met Thr Thr Gly Gly Lys  335
GAA TCT TCT GAT CAC TCA GCT TGG TTC CTG ATT GCG CTT CGT GAG GGA AAG ATT GAA ATT CAG TTC AAG AAT GAA AAG ACA ACC AAA ATG ACA ACT GGA GGC AAA 1163

Val Ile Asn Asp Gly Leu Trp His Met Val Ser Val Glu Glu Leu Glu Gln Ser Ile Ser Val Lys Ile Ala Lys Glu Ala Val Met Asn Ile Asn Lys Pro Gly  370
GTT ATT AAT GAT GGT TTA TGG CAT ATG GTC TCT GTG GAA GAA TTA GAA CAG AGT ATT AGT GTA AAA ATA GCT AAA GAA GCT GTA ATG AAT ATA AAT AAG CCT GGA 1268

Ser Leu Phe Lys Pro Thr Asn Gly Phe Leu Glu Thr Lys Val Tyr Phe Ala Gly Val Pro Arg Lys Met Glu Asn Ala Leu Ile Arg Pro Ile Asn Pro Arg Leu  405
AGC CTT TTT AAG CCC ACC AAT GGG TTT CTA GAA ACC AAA GTA TAC TTT GCA GGA GTA CCT CGG AAA ATG GAA AAT GCA CTC ATT AGA CCG ATT AAC CCT CGT CTA 1373

Asp Gly Cys Ile Arg Gly Trp Asn Leu Met Asn Gln Gly Thr Ser Gly Val Lys Glu Ile Ile Gln Glu Lys Gln Asn Lys His Cys Leu Val Asn Val Glu Lys  440
GAT GGA TGT ATT CGA GGC TGG AAT TTG ATG AAT CAA GGA ACT TCA GGA GTA AAG GAA ATC ATT CAA GAA AAA CAA AAT AAG CAT TGT CTT GTC AAT GTG GAG AAG 1478

Gly Ser Tyr Tyr Pro Gly Thr Gly Val Ala Gln Phe Ser Ile Asn Tyr Lys Asn Glu Ser Asn Pro Glu Ala Trp Gln Ile Asn Val Ser Leu Asn Ile Arg Pro  475
GGT TCC TAC TAT CCT GGT ACT GGA GTT GCT CAG TTT AGC ATA AAT TAT AAG AAT GAA TCC AAT CCT GAG GCT TGG CAA ATC AAT GTG TCC TTG AAT ATT CGC CCA 1583

Ser Ala Gly Thr Gly Val Met Leu Ala Leu Val Ser Asp Asn Thr Val Pro Phe Ala Leu Ser Leu Val Asp Ser Ala Thr Glu Lys Leu Gln Asp Ile Leu Val  510
TCA GCG GGC ACC GGT GTT ATG TTG GCC TTG GTT TCC GAT AAC ACA GTG CCC TTT GCC TTG TCC TTG GTG GAT TCC GCC ACT GAA AAG CTT CAG GAT ATC CTG GTA 1688

Ser Val Glu Ser Met Val Ile Gly Arg Ile Glu Ala Ile Ser Leu Cys Ser Asp Gln Gln Thr Phe Leu Glu Ile Arg Val Asn Arg Asn Asn Leu Glu Leu Ser  545
TCT GTT GAA AGT ATG GTA ATA GGT CGG ATA GAG GCC ATA AGT CTG TGT TCC GAT CAG CAA ACC TTT CTG GAA ATC AGA GTC AAC AGA AAC AAT TTG GAA CTA TCG 1793

Thr Gln Leu Arg Lys Asp Ser Phe His Ser Glu Asp Phe Gln Arg Gln Phe Ala Ile Leu Asp Glu Ala Met Lys Gly Thr Val Val Thr Tyr Leu Gly Gly Leu  580
ACT CAA CTT AGA AAG GAT AGC TTC CAC TCT GAA GAC TTT CAA AGA CAA TTT GCC ATC TTG GAT GAA GCA ATG AAA GGA ACA GTG GTC ACT TAC CTG GGT GGC CTT 1898

Pro Asp Val Pro Phe Ser Ala Thr Pro Val Asn Ala Phe Tyr Gln Gly Cys Met Glu Val Asn Ile Asn Gly Val Gln Val Asp Leu Asp Glu Ala Ile Ser Lys  615
CCA GAT GTT CCA TTC AGT GCC ACA CCA GTG AAT GCC TTT TAT CAA GGC TGT ATG GAA GTG AAC ATT AAT GGT GTA CAG GTG GAT TTG GAT GAA GCC ATT TCT AAA 2003

His Asn Asp Ile Arg Ala His Ser Cys Pro Ser Val Trp Gln Lys Thr Lys His Thr
CAT AAT GAT ATT AGA GCT CAC TCG TGT CCA TCA GTT TGG CAG AAG ACA AAG CAT ACT TAAGGCATTTTTTCTCTGCTGATAATACCTTTTTCCTGTGTGTAATTATACTTATGTTTCAA 2122

TAATAGCTGAAGAATTTTACCTACAATGTGCATATCTTGATTATTTTGTGGTACTTTAACTTTCCTGAAATTTTAAAAAGGTCCTTTTTCAAGAAAACAAGATTCTCTTGTGATATGAATCATATTAAAAATGTTCTT 2260

ACCTCTGTTGCTGTCTAGAAATTAATGAAACCTATAAAAATTTTTAATTTGAAATTTTTGTGACAAATGACATTTCTTCTTTTTTTATGTTTGTAAAAGTAAAGTTTAATTTTATCATCATGaaaaaaaaaaaaaa
```

FIG. 1. Nucleotide sequence of cDNA coding for bovine protein S and amino acid sequence of protein S. The amino acid sequence derived from the cDNA sequence was in complete agreement with the determined amino acid sequence, with one exception (see text). Asp-95 is hydroxylated to β-hydroxyaspartic acid. Asn-458 is glycosylated (arrowhead). The arrow denotes the start of the mature plasma protein. The polyadenylylation signal is underlined.

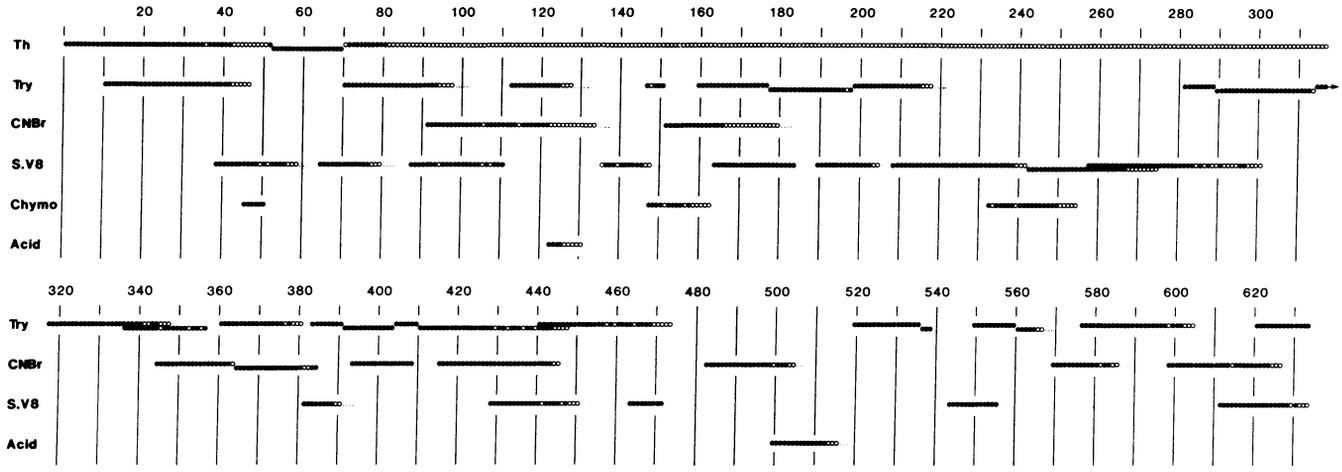

FIG. 2. Fragments used in the derivation of the amino acid sequence. ●, Residue identified during sequencing; ○, unidentified residue; Th, fragments obtained by thrombin cleavage of protein S; Try, fragments obtained by cleavage by trypsin; CNBr, fragments obtained by cleavage with cyanogen bromide; S.V8, fragments obtained by cleavage with *S. aureus* V8 protease; Chymo, fragments obtained by cleavage with chymotrypsin; Acid, fragments obtained by partial acid hydrolysis.

complementary to mRNA coding for residues 85–92 and residues 594–600, respectively. The bovine liver cDNA library was plated on nitrocellulose filters, and a total of 10^5 recombinants were screened with the two oligonucleotide probes. One clone, pBLS-200, was identified using probe 1. Sequence analysis showed it to have an insert that coded for amino acid residues 44–104. With probe 2, six clones were identified, but two had inserts less than 150 nucleotides long and were not examined further. The other four clones contained identical inserts coding for the 105 carboxyl-terminal amino acids and 76 nucleotides of the 3′ nontranslated region. One of the four identical clones, pBLS-400, was used together with pBLS-200 to rescreen the library. Fifteen colonies gave a positive hybridization signal with pBLS-400, one with pBLS-200, and one with both. The latter, clone pBLS-2400, carried a cDNA insert of ≈2.4 kilobases, suggesting it might contain the entire protein S coding sequence. In order to determine the nucleotide sequence, 60 μg of plasmid DNA was digested with *Pst* I. The cDNA insert was isolated, self-ligated, sonicated, and end-repaired. Subfragments of 300–600 base pairs were isolated and ligated into the *Sma* I site of the replicative form of the phage M13 mp8, which was subsequently used to transfect *Escherichia coli* K-12 JM101 cells. Forty-two white plaques, picked at random, were used for preparation of template DNA and sequenced, yielding a continuous sequence of 2379 nucleotides (Fig. 3). On average, each nucleotide was determined four times.

In order to keep an open reading frame, an extra nucleotide had to be introduced around position 595–600. This area was sequenced several times on both strands and unambiguously gave the sequence TGTAAAATGGT (positions 591–601). Protein sequencing in this region gave the sequence -Cys-Lys-Asn-Gly-, which indicates that one nucleotide (A) had been artifactually deleted at position 594–598, presumably during the preparation of the cDNA library.

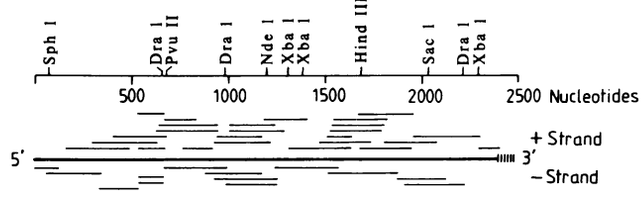

FIG. 3. Restriction map and sequencing strategy for the full-length cDNA clone pBLS-2400.

The amino acid sequence of protein S and the nucleotide sequence of the cDNA clone are given in Fig. 1. The amino-terminal Ala of protein S is coded for by nucleotides 159–161. At the 5′ side of this residue, the cDNA codes for 41 amino acid residues. The methionine residue in position −41 is coded for by nucleotides 36–38. We propose that this is the initiation site for translation, since there is no other methionine in the region. Further, the position corresponds to putative initiation sites of translation in factor IX and protein C (22).

At position 2061 there is a stop codon, TAA. The amino acids coded for at the 5′ side of the stop codon are identical to the carboxyl-terminal amino acid sequence determined by protein sequencing. Thus, the prepro-form of protein S is coded for by nucleotides 36–2060, yielding a molecule of 675 amino acid residues. At the 3′ side of the coding sequence, there are 321 nontranslated nucleotides and then a poly(A) tail. Two putative polyadenylylation signals were found, the sequence ATTAAA located at position 2246 and the sequence AGTAAA 20 nucleotides upstream of the poly(A) tail.

The mature protein S has 634 amino acid residues and its amino acid composition is $Cys_{34}Asp_{32}Asp(OH)_1Asn_{44}Thr_{31}Ser_{44}Glu_{42}Gla_{11}Gln_{24}Pro_{27}Gly_{43}Ala_{36}Val_{43}Met_{11}Ile_{34}Leu_{51}Tyr_{21}Phe_{27}Trp_6His_8Lys_{44}Arg_{20}$. [Asp(OH) is *erythro*-β-hydroxyaspartic acid and Gla is γ-carboxyglutamic acid.] The molecular weight calculated from the amino acid sequence was 70,641. The amino acid compositions of the cyanogen bromide fragments indicated that only the fragment containing residues 416–482 contains *N*-acetylglucosamine. There is no *N*-acetylgalactosamine and thus no O-glycosidic linkage to serine or threonine. Sequence analysis identified two potential glycosylation sites, Asn-Xaa-Ser/Thr, in protein S at positions 458 and 468. The amino acid compositions and sequence analysis of a peptide (residues 464–481), obtained by digestion of the carbohydrate-containing cyanogen bromide fragment with the Glu-specific enzyme from *S. aureus* V8, showed that position 468 was not glycosylated. However, an acid hydrolysate of the peptide Ile-Asn-Tyr-Lys-Asn-Glu (residues 454–459), obtained from a similar digest, contained glucosamine, and sequence analysis gave no asparagine in position 5, indicating that Asn-458 is glycosylated.

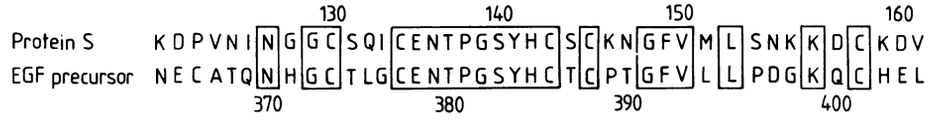

FIG. 4. Sequence comparison of part of the second EGF-homology region (residues 121–161) in protein S with part of EGF-homology region C (residues 364–404) in the mouse EGF precursor (25). Standard one-letter amino acid abbreviations are used.

Homologies to Other Proteins and Tentative Arrangement of Disulfide Bonds. Alignment of the leader sequence in protein S with the corresponding sequences in the other vitamin K-dependent plasma proteins shows that it consists of two parts (22–24). The amino-terminal part (residues −41 to −18) resembles a typical hydrophobic signal peptide. Residues −17 to −1 show a pronounced conservation of sequence among the vitamin K-dependent proteins and end with four consecutive arginine residues in protein S. Although the cleavage site of the signal peptidase has not been identified, it has been proposed that the carboxyl-terminal part of the leader sequence is cleaved off later than the signal peptide (22, 23). This region may be a common structural feature that is recognized by the vitamin K-dependent carboxylase. We recently published the sequence of the 100 amino-terminal amino acid residues in bovine protein S (14). The Gla-containing region, residues 1–44, has a high degree of homology to the other vitamin K-dependent clotting factors and has one intrachain disulfide bond (Cys-17 to Cys-22; ref. 14). The sequence containing residues 45–75 has no counterpart among the other vitamin K-dependent proteins and is inserted between the Gla region and the region that is homologous to the epidermal growth factor (EGF) precursor. Thrombin cleaves protein S at the arginine residues in positions 70 and 52, leaving the Gla-rich domain linked to the large carboxyl-terminal fragment by a disulfide bond involving Cys-47.

A diagonal plot (not shown) of the cysteine-rich region in protein S (residues 80–244) shows it to consist of a structural element that occurs in four copies, each one containing six cysteine residues and being homologous to internal-homology regions in the EGF precursor (25). There is a particularly strong homology between residues 125–161 in protein S and residues 370–401 in the EGF precursor (Fig. 4). Presumably there is independent folding of the four structural units, leading to a disulfide bond arrangement that from the homology is inferred to be identical to that in the EGF precursor. On this assumption, the half-cystine residues from positions 80 to 245 were paired in disulfide bonds (Fig. 5). The aspartic acid in position 95, which is in the amino-terminal EGF domain, is hydroxylated to *erythro*-β-hydroxyaspartic acid (14). Preliminary evidence (not shown) indicates that the asparagine residues in positions 136, 178, and 217 are modified.

Unlike the other vitamin K-dependent plasma proteins (prothrombin factors IX and X and protein C; ref. 26), protein S has no resemblance to the serine proteases. This was corroborated by a sequence comparison using the Dayhoff data base, which did not show significant homology to any other sequenced protein either (27). The carboxyl-terminal part has six cysteine residues. Their arrangement in disulfide bonds was investigated by cyanogen bromide cleavage of the intact molecule (see *Materials and Methods*). The material from the Sephadex G-75 column had an amino acid composition identical to the sum of the two cyanogen bromide fragments containing residues 570–598 and 599–634. Sequence analysis confirmed that the two carboxyl-terminal fragments are linked by a disulfide bond. That the amino acid composition of the purified material from the Sephacryl S-200 column was identical to the sum of the two cyanogen bromide fragments containing residues 394–415 and 416–482 was corroborated by sequence analysis of the same material. Thus the pairing of cysteine residues in positions 17, 22, 408, 434, 597, and 614 has been experimentally determined, whereas the cysteine residues in the EGF-homology region (residues 80–244) have been paired by homology. The pairing of the four remaining cysteine residues—i.e., position 47 with position 71 and position 247 with 526—is tentative.

The vitamin K-dependent plasma proteins have amino-terminal Gla-containing regions and are with one exception [protein Z (28), whose function is unknown] proenzymes of serine proteases (26). Prothrombin is unique insofar as it has two so-called "kringle" structures inserted between the Gla-rich domain and the serine protease part (29). Factors IX and X and protein C have two EGF-homology regions between the Gla region and the serine protease part but have

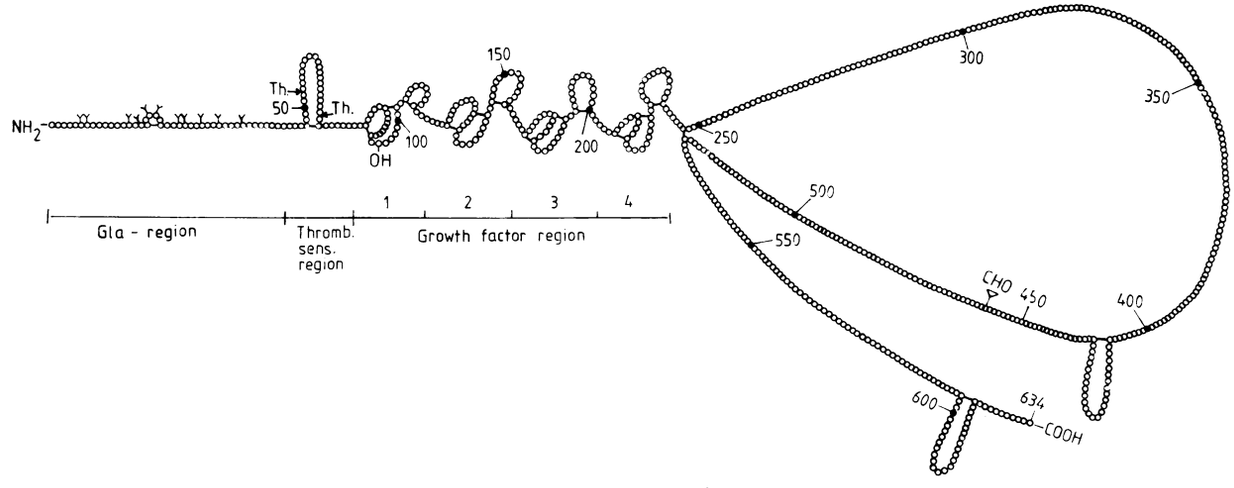

FIG. 5. Domain structure of protein S. See text for the pairing of cysteine residues by disulfide bonds. Th, thrombin-sensitive bonds; CHO, carbohydrate side chain; OH, hydroxyl group of β-hydroxyaspartic acid (residue 95); ○, γ-carboxyglutamic acid.

no kringle structures (22–24). Protein Z is homologous to the latter three proteases but lacks two of the residues in the charge-relay system (28). The amino acid sequence of protein S clearly shows that it must be placed in a third group, independent of the prothrombin group and the group comprising factors IX and X and proteins C and Z. Carboxyl-terminal to the Gla residues, protein S has a region with two thrombin-sensitive bonds, which is followed by an EGF-homology region that is twice as long as in factors IX and X, protein C, and protein Z. However, the most prominent structural feature of protein S is the large carboxyl-terminal fragment (residues 245–634), which is apparently unrelated to the serine proteases. The unique structure of protein S is reflected in its function. In addition to being a cofactor to activated protein C in the degradation of factors Va and VIIIa (5, 6), protein S can form a complex with C4b-binding protein (9, 10). The complexed form of protein S is not a cofactor to activated protein C (7), nor is free protein S that has been cleaved by thrombin. Whether protein S complexed with C4b-binding protein is important in the regulation of the complement system is not known. Knowledge of the amino acid sequence of protein S will make functional analysis of the various parts of the molecule possible.

Dr. Robert Wydro (Integrated Genetics, Framingham, MA) generously provided the cDNA library and the synthetic oligonucleotides used in this investigation and provided invaluable support and advice. The expert technical assistance of Ingrid Dahlqvist, Bergisa Hildebrand, Monica Jönsson, Margareta Kjellberg, and Ulla Peterson is gratefully acknowledged. The gas-phase sequencer was procured through a grant from the Wallenberg Foundation. The investigation was supported by grants from the Swedish Medical Research Council (Project B85-03X-04487-11C and Project B85-13X-7143-01A) and from the Magnus Bergvall Trust, the Alfred Österlund Trust, and the Albert Påhlsson Trust.

1. DiScipio, R. G., Hermodson, M. A., Yates, S. G. & Davie, E. W. (1977) *Biochemistry* **16,** 698–706.
2. DiScipio, R. G. & Davie, E. W. (1979) *Biochemistry* **18,** 899–904.
3. Stenflo, J. & Jönsson, M. (1979) *FEBS Lett.* **101,** 377–381.
4. Dahlbäck, B. (1983) *Biochem. J.* **209,** 837–846.
5. Walker, F. J. (1981) *J. Biol. Chem.* **256,** 11128–11131.
6. Gardiner, J. E., McGann, M. A., Berridge, C. N., Fulcher, C. A., Zimmerman, T. S. & Griffin, J. H. (1984) *Circulation* **70,** 820.
7. Comp, P. C., Nixon, R. R., Cooper, M. R. & Esmon, C. T. (1984) *J. Clin. Invest.* **74,** 2082–2088.
8. Schwarz, H. P., Fischer, M., Hopmeister, P., Batchard, M. A. & Griffin, J. H. (1984) *Blood* **64,** 1297–1300.
9. Dahlbäck, B. & Stenflo, J. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 2512–2516.
10. Dahlbäck, B. (1983) *Biochem. J.* **209,** 847–856.
11. Dahlbäck, B. & Hildebrand, B. (1983) *Biochem. J.* **209,** 857–863.
12. Drakenberg, T., Fernlund, P., Roepstorff, P. & Stenflo, J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1802–1806.
13. Fernlund, P. & Stenflo, J. (1983) *J. Biol. Chem.* **258,** 12509–12512.
14. Dahlbäck, B., Lundwall, Å. & Stenflo, J. (1986) *J. Biol. Chem.* **261,** 5111–5115.
15. Sugo, T., Dahlbäck, B., Holmgren, A. & Stenflo, J. (1986) *J. Biol. Chem.* **261,** 5116–5120.
16. Walker, F. J. (1984) *J. Biol. Chem.* **259,** 10335–10339.
17. Woods, D. E., Markham, A. F., Richer, A. T., Goldberger, G. & Colten, H. R. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 5661–5665.
18. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3963–3965.
19. Bankier, A. T. & Barrell, B. G. (1983) in *Techniques in Nucleic Acid Biochemistry*, ed. Flavell, R. A. (Elsevier, Limerick, Ireland), Vol. 85, pp. 1–73.
20. Staden, R. (1982) *Nucleic Acids Res.* **10,** 4731–4751.
21. Staden, R. (1982) *Nucleic Acids Res.* **10,** 2951–2961.
22. Foster, D. C., Yoshitake, S. & Davie, E. W. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4673–4677.
23. Long, G. L., Belagaje, R. & MacGillivray, R. T. A. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 5653–5656.
24. Fung, M. R., Hay, C. W. & MacGillivray, R. T. A. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 3591–3595.
25. Doolittle, R. F., Feng, D. F. & Johnsson, M. S. (1984) *Nature (London)* **307,** 558–560.
26. Jackson, C. M. & Nemerson, Y. (1980) *Annu. Rev. Biochem.* **49,** 765–811.
27. Dayhoff, M. O., Barkier, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91,** 524–545.
28. Hojrup, P., Jensen, M. S. & Petersen, T. E. (1985) *FEBS Lett.* **184,** 333–338.
29. Magnusson, S. (1975) in *Proteases and Biological Control*, eds. Reich, E., Rifkin, D. B. & Shaw, E. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 123–149.