Amino acid sequence of bovine protein Z: a vitamin K-dependent serine protease homolog

Peter Højrup, Margit S. Jensen and Torben E. Petersen*

Department of Molecular Biology and Plant Physiology, University of Aarhus, DK-8000 Aarhus C, Denmark

Received 2 April 1985

The amino acid sequence of protein Z has been determined from sequence analysis performed on fragments obtained by chemical and enzymatic degradations. The polypeptide consists of a single chain containing 396 amino acid residues (M_r 43 677). Comparison with the vitamin K-dependent plasma proteins reveals an extensive homology. The N-terminal part, containing 13 γ-carboxyglutamic acid and one β-hydroxyaspartic acid residue, is extensively homologous to and of similar length to the light chain of factor X. The remainder of protein Z is homologous to the serine proteases and of similar size to the heavy chain of factor Xa, but of the active site residues only aspartic acid-102 is present. Histidine-57 and serine-195 are replaced in protein Z by threonine and alanine, respectively. The physiological function of protein Z is still uncertain.

γ-Carboxyglumatic acid *β-Hydroxyaspartic acid* *Haptoglobin* *Blood coagulation*

1. INTRODUCTION

Protein Z is a single chain glycoprotein which has been purified from bovine [1,2], and human plasma [3]. It has a molecular mass of 50 kDa and contains 13 γ-carboxyglutamic acid (Gla) residues within the amino-terminal 40 residues [4]. This amino-terminal sequence of protein Z is highly homologous with those of the well known vitamin K-dependent coagulation factors including protein C and protein S, but sufficiently different from these factors to identify protein Z as a distinct vitamin K-dependent protein. No functional assay is available for protein Z, and despite the strong similarity with the coagulation factors, antibodies raised against protein Z show no effect in standard clotting assays ([3]; unpublished). Furthermore, it has not been possible to activate protein Z to a serine protease with proteolytic enzymes such as thrombin, factor Xa and trypsin (unpublished). Here we present the amino acid sequence of bovine protein Z, showing that protein Z is homologous to the serine proteases but due to mutation of essential amino acids, protein Z cannot function as a proteolytic enzyme.

* To whom correspondence should be addressed at: Department of Biochemistry, SJ-70, University of Washington, Seattle, WA 98195, USA

Abbreviations: Gla, γ-carboxyglutamic acid; Pth, phenylthiohydantoin

2. MATERIALS AND METHODS

Bovine protein Z was purified by adsorption to barium citrate followed by ion-exchange chromatography as described [2]. A sample of protein Z was also obtained from M.P. Esnouf.

Trypsin, chymotrypsin and pepsin were purchased from Worthington, thermolysin and soybean trypsin inhibitor from Sigma and staphylococcal protease from Miles Laboratories. Sephacryl S-200 and Sephadex G-50, fine were from Pharmacia.

The sequences of the smaller peptides were mainly determined manually, while longer sequences were determined using a Beckman se-

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30 © 1985 Federation of European Biochemical Societies

quencer 890C. Detailed technical sequence methods, including peptide separation by HPLC have been described [4,5]. β-Hydroxyaspartic acid was determined by amino acid analysis on a Beckman 121 MB analyzer as described by Drakenberg et al. [6], and a sample of β-hydroxyaspartic acid used as a standard was obtained from J. Stenflo.

2.1. *Degradation of protein Z*

Protein Z (41 mg) was degraded for 20 h with cyanogen bromide (80 mg in 70% formic acid) followed by reduction and carboxymethylation. The fragments were separated on a column of Sephacryl S-200 (2.5 × 80 cm) in 8 M urea, 50 mM ammonium acetate, pH 5.0. Reduced and carboxymethylated protein Z (150 mg) was digested with 1.8 mg trypsin for 3 h at 37°C in 0.1 M ammonium bicarbonate, pH 8.3, and the reaction stopped by adding 2 mg soybean trypsin inhibitor. The resulting peptides were separated on a column of Sephadex G-50 (2.5 × 100 cm) in 0.1 M ammonium bicarbonate, pH 8.3, and further purified by high-voltage paper electrophoresis at pH 6.5 and 2.1 or by HPLC. A similar digest using 130 mg protein Z was performed with chymotrypsin and the reaction stopped by adding 2 mg phenylmethanesulfonyl fluoride dissolved in 1 ml ethanol. The last digest of reduced and carboxymethylated protein Z (80 mg) was with 1 mg staphylococcal protease for 12 h at 37°C in 0.1 M ammonium bicarbonate. The peptides from this digest were also separated on a column of Sephadex G-50 (2.5 × 100 cm) in 0.1 M ammonium bicarbonate, pH 8.3. The smaller peptides were further purified by HPLC, while the larger peptides were subdigested with pepsin. Subdigests with thermolysin were performed in 50 mM pyridinium acetate, 10 mM $CaCl_2$, pH 6.5, at 37°C while the pepsin subdigest was in 5% formic acid.

3. RESULTS

The amino acid sequence of bovine protein Z was determined from peptides obtained by digestion of reduced and carboxymethylated protein Z with trypsin, chymotrypsin or staphylococcal protease. The amino-terminal sequences of the cyanogen bromide fragments were also determined. The sequence of protein Z is shown in fig.1 together with the fragments used for elucidation of the structure. It contains 396 amino acid residues including 13 γ-carboxyglutamic acids, one β-hydroxyaspartic acid, 3 asparagine-linked oligosaccharide groups and one threonine-linked oligosaccharide group. The M_r value of the peptide part of protein Z is 43 677, and the amino acid composition is shown in table 1.

Thirty-three tryptic peptides were isolated and characterized corresponding to the entire sequence of protein Z except for residues 321–347. This stretch of the sequence is particularly rich in hydrophobic residues, and the peptide was probably lost because of low solubility. No special effort was taken to isolate this peptide as the sequence was unambiguously deduced from the chymotryptic and staphylococcal protease peptides. Thirty-two chymotryptic peptides were isolated and characterized. Their composition accounts for 96% of the sequence. All the tryptic and chymotryptic peptides are indicated in fig.1, but only 3 of the staphylococcal protease peptides are shown because only minimal evidence is presented.

In the evidence for the amino acid sequence, the overlap between T14 and T15 is based on the 3 residue chymotryptic peptide C18. This results in a single residue overlap, but it is most unlikely that there are unknown peptides located between T14 and T15 as the homology with factor X and haptoglobin is good in this region (fig.2). Also, the composition of peptide Sp4Pe1 supports this conclusion.

3.1. *Postsynthetic modifications*

Thirteen γ-carboxyglutamic acids have been found in protein Z in positions 7, 8, 11, 15, 17, 20, 21, 26, 27, 30, 33, 36 and 40 as published elsewhere [4]. The positions of the Gla residues are the same as the 10 in prothrombin plus the two in factor X and IX and protein C in the clotting factors. Protein Z has an extra Gla-11 which is unique for this protein.

One β-hydroxyaspartic acid has been assigned to position 64 in protein Z. We have not been able to obtain full sequence evidence for the exact location of the β-hydroxyaspartic acid mainly due to the carbohydrate group attached to the nearby Asn-59. As judged by amino acid analysis, the peptide C3Th1 (residues 57–65) contained one β-hydroxyaspartic acid. This peptide was purified by

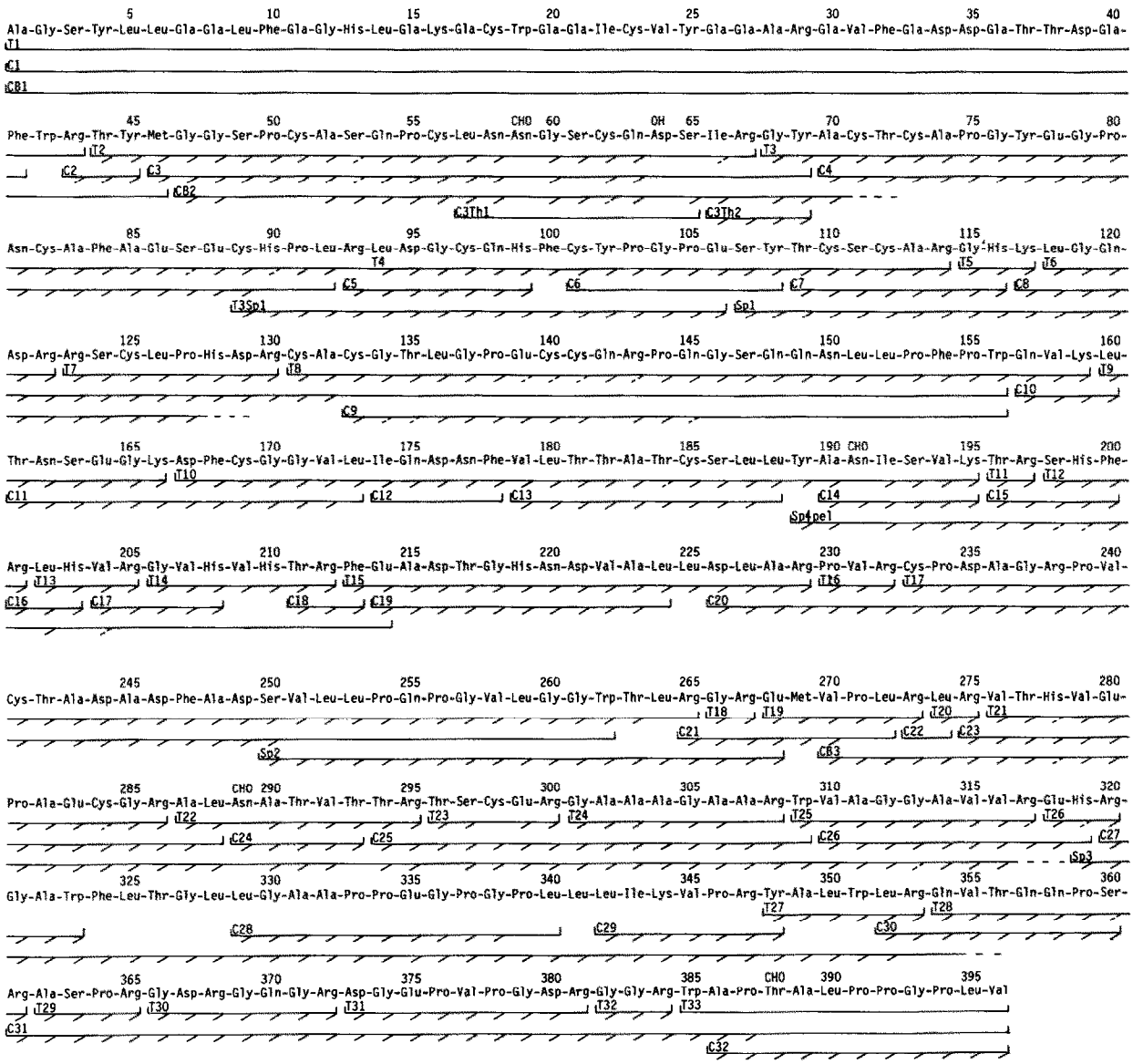

Fig.1. Amino acid sequence of bovine protein Z. The figure summarizes the peptides used in deducing the sequence. Gla, γ-carboxyglutamic acid; CHO, carbohydrate; OH, hydroxy group attached to aspartic acid (β-hydroxyaspartic acid); T1–T33, fragments isolated from a tryptic digest; C1–C32, chymotryptic peptides; CB1–CB3, cyanogen bromide peptides; Sp, staphylococcal protease peptides; C3Th1, thermolytic digest of peptide C3; Sp4Pe1, peptic peptide from Sp4. Sequence evidence for residues 1–42 has been published in [4].

HPLC after digestion of peptide C3 with thermolysin but the yield obtained was too low for an amino acid sequence determination. In none of the other sequence runs of this area was a clear Pth derivative of β-hydroxyaspartic acid seen. Nevertheless, the unambiguous finding by amino acid analysis of one β-hydroxyaspartic acid in peptide C3Th1 and in all the other peptides containing position 64, the consistent absence of a Pth derivative in this position, and the high degree of homology with protein C [6], factor IX [9] and factor X [9,11], all of which contain one β-

Table 1

Amino acid composition of bovine protein Z

	Sequence	Amino acid analysis
Cys	24	21
Asp	18	–
Asn	9	–
Asp-OH	1	–
Asx	–	30
Thr	23	21
Ser	18	22
Glu	14	–
Gln	15	–
Gla	13	–
Glx	–	46
Pro	32	30
Gly	43	38
Ala	35	32
Val	25	29
Met	2	4
Ile	5	8
Leu	38	36
Tyr	9	10
Phe	12	12
Lys	6	11
His	12	11
Trp	8	7
Arg	34	29

Asp-OH, β-hydroxyaspartic acid; Gla, γ-carboxyglutamic acid. The amino acid analysis is taken from [2]

hydroxyaspartic acid in the corresponding position, strongly indicate that a β-hydroxyaspartic acid is located in position 64. The finding of one β-hydroxyaspartic acid is in agreement with the amino acid analysis results of Fernlund and Stenflo [12], who found 0.85 mol β-hydroxyaspartic acid per mol protein in bovine protein Z.

Three oligosaccharide groups containing glucosamine were found attached to Asn-59, Asn-191 and Asn-289 all obeying the Asn-X-Ser/Thr glycosylation rule. A fourth oligosaccharide group containing galactosamine was found attached to Thr-388 in agreement with the Thr/Ser-X-X-Pro glycosylation rule. An oligosaccharide is found in the position equivalent to Thr-388 in factor X. Direct sequence evidence was not found for Asn-59, but the position was deduced from the consistent presence of Asx and carbohydrate in all amino acid analysis of peptides covering this area and in the absence of a Pth derivative during sequencing of the carbohydrate-containing step.

4. DISCUSSION

Protein Z is a single chain protein whose sequence shows homology with the factor X-like coagulation factors. This is illustrated in fig.2, where protein Z is aligned with factor X and haptoglobin. The Gla region consists of the amino-terminal 42 residues which are found in all the vitamin K-dependent coagulation factors. Then follows a region (residues 43–134) which shows 45% identity with the light chain of factor X including a β-hydroxyaspartic acid in position 64. A short stretch of amino acid residues leads to the part of protein Z which is homologous with the serine proteases. The latter region shows 23% identity with the heavy chain of factor X. This is only slightly less than the identity of haptoglobin to factor X (25%). It is clear that like haptoglobin protein Z cannot show any proteolytic activity because the residue that would correspond to the active serine (195 in chymotrypsin) is Ala-311 in protein Z and also the active site histidine (57 in chymotrypsin) is changed to a threonine (Thr-184). As in haptoglobin, Asp-102 (chymotrypsin numbering) is however present in protein Z (Asp-221).

Most serine proteases are activated by cleavage of a single peptide bond on the carboxy-terminal side of an arginine residue resulting in a new amino-terminal valine or isoleucine (Ile-16 in chymotrypsin). The arrow in fig.2 points at the activation bond in factor X. A similar activation of protein Z seems to be impossible as the area around the corresponding bond is quite different in protein Z compared to the serine proteases, and would not be a good substrate for an activating Arg-specific protease.

The 17 'extra' amino acid residues found in haptoglobin are not seen in protein Z (inserted between position 295 and 296 using protein Z numbers), but a carboxy-terminal tail similar to that found in factor X is present with the insertion of an extra 12 residues in protein Z. Also, the *O*-linked carbohydrate is present in a homologous position (Thr-388).

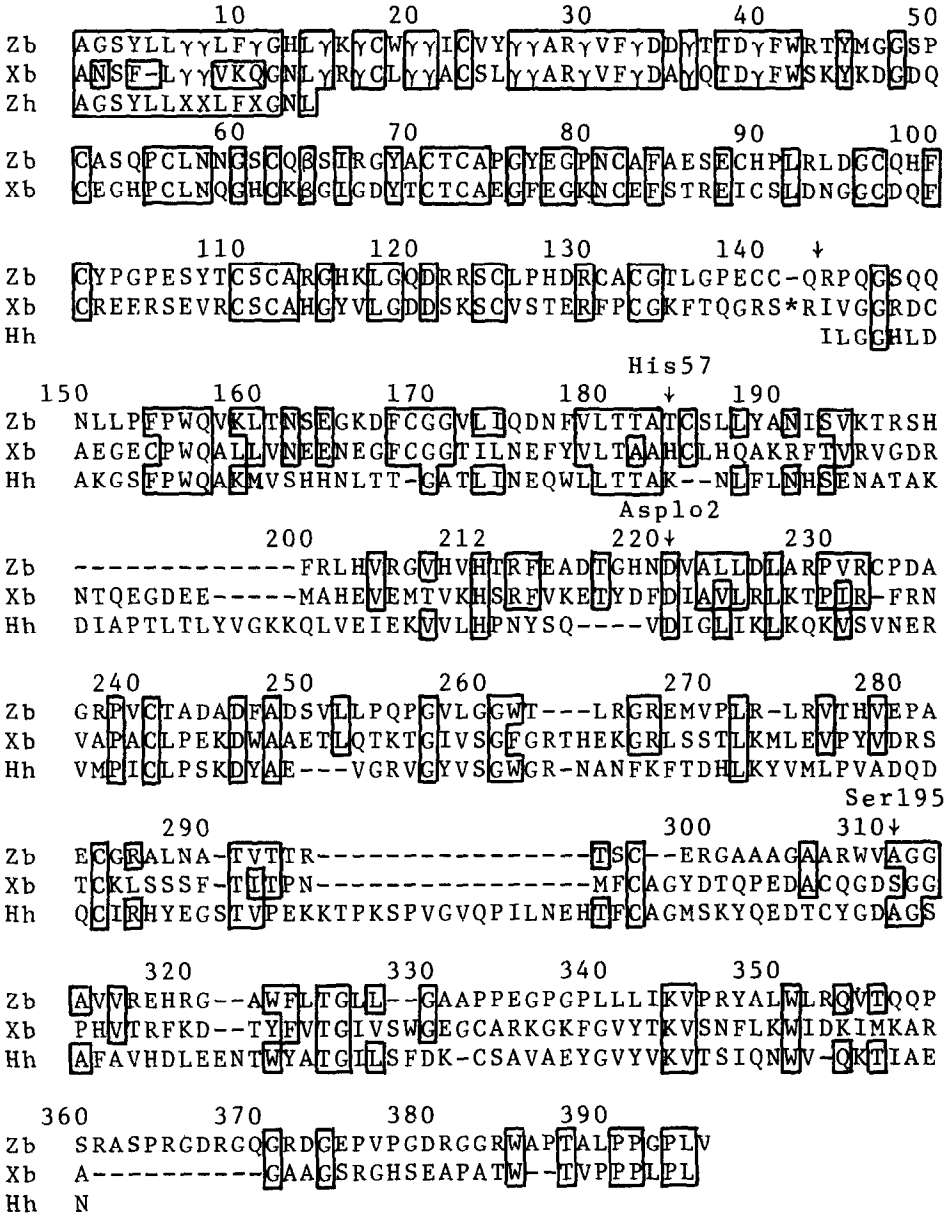

Fig.2. Sequence comparison of bovine protein Z (Zb) with bovine factor X (Xb) [7–9], human protein Z (Zh) [3] and human haptoglobin (Hh) [10]. Boxed residues are identical to corresponding residues in protein Z. The star in the sequence of factor X indicates the omission of 50 residues of the activation peptide. The arrow points to the activation bond in factor X. The standard single-letter code is used except for γ, γ-carboxyglutamic acid; β, β-hydroxyaspartic acid; X, unidentified residues presumed to be γ-carboxyglutamic acid.

All the half-cystines found in the light chain of factor X are also present in protein Z and of the 3 disulfide loops common to most serine proteases (histidine 42–58, methionine 168–182 and serine 191–220, using chymotrypsin numbering) only the histidine and methionine loops are present in protein Z. The two half-cystines forming the disulfide bond in the serine loop are probably replaced by Ala-307 and Pro-333 in protein Z.

Four half-cystines in protein Z (residues 131,

140, 141 and 233) have not been found in the other proteins, and are therefore likely to constitute two 'additional' disulfide bridges. As two of these half-cystines are adjacent to each other in the sequence they cannot form a mutual disulfide bond and therefore either Cys-140 or Cys-141 must be bonded to Cys-131 and the other one to Cys-233.

In protein Z Cys-241 is found in a position corresponding to one of the half-cystines (Cys-122 in chymotrypsin) which form an inter-chain bridge common to all the two-chain serine proteases. This is in agreement with the number of half-cystines found in the area homologous with the light chain of factor X.

This work shows that protein Z is a serine protease homolog which on structural grounds should be unable to function as a serine protease. The physiological function of protein Z is still not known, but the homology with the other vitamin K-dependent plasma proteins indicates an affinity for the same surfaces and like haptoglobin it may serve as a binding protein.

ACKNOWLEDGEMENTS

We thank N. Axelsen, Statens Seruminstitut, København, for raising the antibodies against protein Z. This work was supported by the US National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD (grant HL 16238) and the Danish Science Research Council (grant 11-2301).

REFERENCES

[1] Prowse, C.V. and Esnouf, M.P. (1977) Biochem. Soc. Trans. 5, 255–256.
[2] Petersen, T.E., Thøgersen, H.C., Sottrup-Jensen, L., Magnusson, S. and Jörnvall, H. (1980) FEBS Lett. 114, 278–282.
[3] Broze, G.J. and Miletich, J.P. (1984) J. Clin. Invest. 73, 933–938.
[4] Højrup, P., Roepstorff, P. and Petersen, T.E. (1982) Eur. J. Biochem. 126, 343–348.
[5] Skorstensgaard, K., Thøgersen, H.C., Vibe-Pedersen, K., Petersen, T.E. and Magnusson, S. (1982) Eur. J. Biochem. 128, 605–623.
[6] Drakenberg, T., Fernlund, P., Roepstorff, P. and Stenflo, J. (1983) Proc. Natl. Acad. Sci. USA 80, 1802–1806.
[7] Enfield, D.L., Ericsson, L.H., Walsh, K.A., Neurath, H. and Titani, K. (1975) Proc. Natl. Acad. Sci. USA 72, 16–19.
[8] Titani, K., Fujikawa, K., Enfield, D.L., Ericsson, L.H., Walsh, K.A. and Neurath, H. (1975) Proc. Natl. Acad. Sci. USA 72, 3082–3086.
[9] Sugo, T., Fernlund, P. and Stenflo, J. (1984) FEBS Lett. 165, 102–106.
[10] Kurosky, A., Barnett, D.R., Lee, T.-H., Touchstone, B., Hay, R.E., Arnott, M.S., Bowman, B.H. and Fitch, W.M. (1980) Proc. Natl. Acad. Sci. USA 77, 3388–3392.
[11] McMullen, B.A., Fujikawa, K., Kisiel, W., Sasagawa, T., Howald, W.N., Kwa, E.Y. and Weinstein, B. (1983) Biochemistry 22, 2875–2884.
[12] Fernlund, P. and Stenflo, J. (1983) J. Biol. Chem. 258, 12509–12512.