THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 262, No. 29, Issue of October 15, pp. 13916–13919, 1987
Printed in U.S.A.
© 1987 by The American Society for Biochemistry and Molecular Biology, Inc.

Primary Structure of a Protease Isoinhibitor from Bovine Spleen

A POSSIBLE INTERMEDIATE IN THE PROCESSING OF THE PRIMARY GENE PRODUCT*

(Received for publication, May 21, 1987)

Donatella Barra‡§, Maurizio Simmaco‡, Francesco Bossa‡, Evandro Fioretti¶, Mauro Angeletti¶, and Franca Ascoli‖

From the ‡Dipartimento di Scienze Biochimiche e Centro di Biologia Molecolare del Consiglio Nazionale delle Ricerche, Università di Roma La Sapienza, 00185 Roma, Italy, the ¶Dipartimento di Biologia Cellulare, Università di Camerino, 62032 Camerino, Italy, and the ‖Dipartimento di Medicina Sperimentale e Scienze Biochimiche, Università di Roma Tor Vergata, 00173 Roma, Italy

Sequence studies on the protease isoinhibitor I isolated from bovine spleen have revealed that it consists of two molecular variants which differ only in the presence of an additional COOH-terminal residue, asparagine, in the less abundant form. The complete amino acid sequence shows that they are composed of 65 or 66 residues and predicts M_r of 7223 or 7338, respectively. The sequences correspond exactly to the 58-residue polypeptide chain of spleen isoinhibitor II plus NH_2- and COOH-terminal extensions of 2 and 5 or 6 amino acid residues, respectively. Moreover the entire sequences are located within the 100-residue structure deduced from the mRNA and DNA sequences of the putative precursor. These data support the idea that the molecular variants of isoinhibitor I are either mature proteins with distinct functional roles, or intermediates in the multistage processing of the primary product of gene expression, which eventually leads to the mature protein, *i.e.* inhibitor II.

Analysis of eukaryotic genes and of the corresponding mRNA sequences on the one hand, and of the mature protein product on the other, has often shown that the latter is the result of post-translational proteolytic processing of the primary product of expression. The NH_2- and COOH-terminal extensions of these larger precursor molecules, whose structures may be deduced from the gene sequence, are thought to be responsible for transport across membranes or, in general, for the final targetting of the mature protein (1–3). However, direct proof of the occurrence of such precursors and information about their maturation are difficult to obtain *in vivo*.

In this paper we report the primary structure of a serine protease inhibitor (inhibitor I) isolated from bovine spleen (4). Comparison of its sequence with that of the corresponding translation product predicted from the gene sequence (5) shows inhibitor I is either a mature protein with a distinct functional role, or is an intermediate generated in the course of proteolytic processing of the primary gene product. The final product of this process should be bovine spleen inhibitor II, another isoinhibitor already isolated and sequenced in our laboratory (6).

* This work was supported in part by a grant from the Ministero Pubblica Istruzione. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom all correspondence should be addressed: Dipartimento di Scienze Biochimiche, Università La Sapienza, Piazzale Aldo Moro, 5, 00185 Roma, Italy.

EXPERIMENTAL PROCEDURES

Materials—Bovine spleen inhibitor I was isolated as previously described by affinity chromatography on immobilized trypsin, followed by SP-Sephadex C-50 chromatography at pH 6.5 (4). After this step, the fraction containing inhibitor I was further purified on a Mono S column in the Pharmacia fast protein liquid chromatography system, as described (6), followed by high performance liquid chromatography (HPLC)[1] on a reverse-phase column under the conditions reported below for peptide purification. *Staphylococcus aureus* V8 protease was from Miles; CNBr was from Fluka. All other chemicals were of analytical grade.

Amino Acid Sequence Determination—0.8 mg of HPLC-purified inhibitor I was reduced and carboxymethylated as reported (6).

50 nmol of carboxymethylated protein was treated with a few crystals of CNBr overnight at room temperature in 0.1 ml of 70% formic acid and then lyophilized. The resulting peptide mixture was purified by HPLC using a Beckman model 332 instrument, on a reverse-phase column (Brownlee, Aquapore RP-300, 10 μm, 4.6 × 250 mm) with a linear gradient of the following solvents: solvent A, 0.2% trifluoroacetic acid; solvent B, 0.1% trifluoroacetic acid in acetonitrile/isopropyl alcohol 4:1 (v/v), at a flow rate of 1 ml/min. Elution of the peptides was monitored on a Beckman 165 spectrophotometer at 220 nm.

A second aliquot of carboxymethylated protein (20 nmol) was digested with *S. aureus* protease at 37 °C for 3 h at an enzyme/substrate ratio of 1:30. The peptide mixture was fractionated by HPLC as reported above.

NH_2-terminal residues of purified peptides were determined according to Gray (7). Amino acid analysis was carried out with an LKB 4400 instrument equipped with a Spectra-Physics System I Computing Integrator after hydrolysis of samples in 6 N HCl at 110 °C for 24 h.

Sequence analysis was performed on an Applied Biosystems model 470A gas-phase protein sequencer equipped with an Applied Biosystems model 120A PTH Analyzer for the on-line detection of phenylthiohydantoins. Samples were dissolved in 10% acetic acid and loaded onto trifluoroacetic acid-treated glass fiber filters coated with Polybrene and prewashed according to the manufacturer's instructions.

RESULTS

Gas-phase sequencing of carboxymethylated spleen inhibitor I gave the sequence to residue 41 (CM-protein, Table I).

The elution profile of the sample treated with CNBr is shown in Fig. 1. Amino acid analyses of the purified peptides are reported in Table II, as well as those of peptides obtained after digestion with *S. aureus* protease. Results of sequencing of peptides B1, B2, B3, and B4 are reported in Table I; peptide B3a gave the same sequence as B3. Only peptide S3 in the *S. aureus* digest was sequenced to provide the overlap of peptide B2 with B3 (Table I).

Analytical data from peaks B3, B3a, and B4 of the CNBr

[1] The abbreviations used are: HPLC, high performance liquid chromatography; BPTI, bovine pancreatic trypsin inhibitor.

This is an Open Access article under the CC BY license.

TABLE I

Amino acid sequence data for bovine spleen inhibitor I

Position	Residue	Peptide, (cycle no.), yield (pmol)					
		CM-protein (500 pmol)	B 1 (1200 pmol)				
-2	Ala	(1) 178	(1) 520				
-1	Gln	(2) 78	(2) 320				
+1	Arg	(3) 12	(3) 195				
2	Pro	(4) 78	(4) 268				
3	Asp	(5) 83	(5) 155				
.	Phe	(6) 50	(6) 190				
5	CmCys	(7) 45	(7) 130				
6	Leu	(8) 40	(8) 125				
7	Glu	(9) 35	(9) 83				
8	Pro	(10) 50	(10) 108				
9	Pro	(11) 71	(11) 123				
10	Tyr	(12) 35	(12) 68				
11	Thr	(13) 22	(13) 35				
12	Gly	(14) 38	(14) 58				
13	Pro	(15) 35	(15) 45				
14	CmCys	(16) 19	(16) 28				
15	Lys	(17) 23	(17) 15				
16	[illegible]	(18) 25	(18) 35				
17	Lys	(19) 19	(19) 13	B 2			
18	Met	(20) 15	(20) X[a]	(1500 pmol)			
19	Ile	(21) 18	RY[a] 88%	(1) 1150			
20	Arg	(22) 8		(2) 1098			
21	Tyr	(23) 15		(3) 1170			
22	Phe	(24) 18		(4) 953			
23	Tyr	(25) 15		(5) 948			
24	Asn	(26) 20		(6) 710			
25	Ala	(27) 22		(7) 825			
26	Lys	(28) 9		(8) 820			
27	Ala	(29) 21		(9) 833			
28	Gly	(30) 18		(10) 518			
29	Phe	(31) 13		(11) 640			
30	CmCys	(32) X[a]		(12) 403			
31	Glu	(33) X[a]		(13) 422			
32	Thr	(34) X[a]		(14) 290			
33	Phe	(35) 10		(15) 370			
34	Val	(36) 8		(16) 333			
35	Tyr	(37) 5		(17) 348			
36	Gly	(38) 11		(18) 310			
37	Gly	(39) 13		(19) 343			
38	CmCys	(40) X[a]		(20) 188			
39	Lys	(41) 4		(21) 165			
40	Ala	RY[a] 95%		(22) 198			
41	Lys			(23) 188			
42	Ser			(24) 158			
43	Asn			(25) 123			
44	Asn			(26) 155			
45	Phe			(27) 120			
46	Arg			(28) 113			
47	Ser			(29) 53			
48	Ala			(30) 88			S 3
49	Glu			(31) 70			(1000 pmol)
50	Asp			(32) 55			(1) 470
51	CmCys			(33) 38	B 3	B 4	(2) 179
52	Met			(34) X[a]	(1200 pmol)	(1200 pmol)	(3) 385
53	Arg			RY[a] 92%	(1) 893	(1) 1070	(4) 290
54	Thr				(2) 740	(2) 973	(5) 340
55	CmCys				(3) 613	(3) 615	(6) 303
56	Gly				(4) 635	(4) 588	(7) 306
57	Gly				(5) 593	(5) 550	(8) 310
58	Ala				(6) 743	(6) 675	(9) 287
59	Ile				(7) 463	(7) 430	(10) 269
60	Gly				(8) 458	(8) 443	(11) 210
61	Pro				(9) 420	(9) 353	(12) 168
62	Arg				(10) 263	(10) 208	(13) 130
63	Glu				(11) 106	(11) 185	(14) 42
64	Asn				RY[a] 93%	(12) 93	RY[a] 91%
						RY[a] 93%	

[a] X, not identified; RY, repetitive yield.

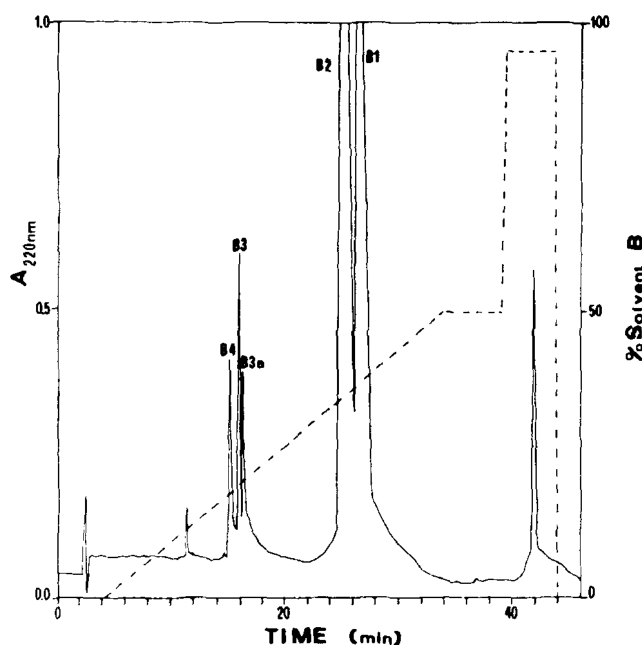

FIG. 1. **Reverse-phase high performance liquid chromatography of the CNBr digest of carboxymethylated protein.** Flow rate: 1.0 ml/min. Column: Aquapore RP-300, 10 μm, 4.6 × 250 mm. Solvent systems: *A*, 0.2% trifluoroacetic acid; *B*, 0.1% trifluoroacetic acid in acetonitrile/isopropyl alcohol 4:1 (v/v). ——, absorbance at 220 nm. - - -, percent of solvent B in the gradient elution. The *numbers* above the peaks refer to the peptides subsequently found in these peaks.

digest clearly show the existence of two variants of the COOH-terminal fragment, one ending with Glu-63 and the other with Asn-64, according to the numbering adopted for isoinhibitor II, as reported in Fig. 2. The occurrence of peaks B3 and B3a, with identical amino acid compositions and sequences, could be explained on the basis of differences in the oxidation state of methionine and/or carboxymethylcysteine residues. The presence of these COOH-terminal peptides indicates the existence of two different polypeptide chains in the sample analyzed. The difference of 1 amino acid residue was not detected in the original inhibitor I preparation, whereas variants of the small COOH-terminal fragment could be separated easily by HPLC. Their yields (see Table II) indicate a predominance of the shorter version of the polypeptide chain.

The complete sequence reported in Fig. 2 shows that inhibitor I is composed either of 65- or 66-amino acid residues, corresponding to molecular weights of 7223 or 7338, respectively, and contains NH_2- and COOH-terminal extensions with respect to inhibitor II, reported for comparison in the same figure. The two sequences are aligned under the sequence of the primary translation product deduced from the sequence of the spleen inhibitor gene, as reported by Creighton and Charles (5).

DISCUSSION

Isoinhibitors from bovine tissues, in particular from bovine spleen (4), belong to the BPTI (bovine pancreatic trypsin inhibitor) family (8). The name BPTI was initially given to a well-known protein of 58 amino acids with a tertiary structure stabilized by three disulfide bonds, which inhibits a number of serine proteases (9). Several studies revealed the presence of four isoinhibitors (I–IV) in bovine spleen (4, 6, 10–12). BPTI, itself, has been identified together with isoinhibitor IV; the other three isoinhibitors, which are present in different amounts in various bovine tissues, have structural and functional properties very similar to BPTI. They all have basic isoelectric points, low molecular weights, and show inhibitory activity toward trypsin, chymotrypsin, and kallikrein. BPTI is found in mast cells (13), whereas the other three isoinhibitors have been localized immunocytochemically in smooth muscle cells of some bovine spleen blood vessels (12). Spleen inhibitor II has been purified and sequenced: it has the same number of amino acids as BPTI, but with seven substitutions. These substitutions are such that the inhibitory activity and the overall tertiary structure are not significantly affected. Only a small decrease in the affinity constant for porcine kallikrein was observed in comparison with BPTI (6).

Anderson and Kingston (14, 15) have shown that two different but highly homologous genes are present in the bovine genome, one encoding BPTI and the other spleen inhibitor II. More recently, Creighton and Charles (5) were able to characterize the two genes further, showing that they are constructed of three exons and two introns. From the

TABLE II

Amino acid composition of fragments obtained after digestion of carboxymethylated inhibitor I (CM-protein) with CNBr (B) and S. aureus protease (S)

The composition from sequence analysis of each peptide is indicated by the numbers in parentheses.

Peptide residue nos.	CM-protein	B1 −2–18	B2 19–52	B3 53–63	B3a 53–63	B4 53–64	S1 −2–31	S2 32–49	S3 50–63
CmCys	5.3 (6)	1.6 (2)	2.5 (3)	0.8 (1)	0.6 (1)	0.6 (1)	2.5 (3)	1.0 (1)	1.8 (2)
Asp	5.2 (5)	0.9 (1)	4.0 (4)			0.8 (1)	2.0 (2)	1.9 (2)	1.0 (1)
Thr	2.7 (3)	1.0 (1)	0.9 (1)	0.7 (1)	0.6 (1)	0.8 (1)	0.8 (1)	0.9 (1)	1.0 (1)
Ser	1.7 (2)		2.0 (2)					1.7 (2)	
Glu	5.0 (5)	2.0 (2)	2.0 (2)	1.1 (1)	0.7 (1)	1.0 (1)	2.9 (3)	1.0 (1)	1.0 (1)
Pro	4.8 (5)	3.5 (4)		1.0 (1)	1.1 (1)	1.0 (1)	3.7 (4)		1.0 (1)
Gly	6.7 (7)	1.2 (1)	3.1 (3)	3.0 (3)	3.1 (3)	3.0 (3)	2.0 (2)	2.0 (2)	2.9 (3)
Ala	6.8 (7)	2.0 (2)	3.8 (4)	1.1 (1)	1.0 (1)	1.1 (1)	3.9 (4)	1.8 (2)	1.0 (1)
Val	1.0 (1)		1.0 (1)					0.9 (1)	
Met	1.4 (2)						0.8 (1)		0.7 (1)
Ile	1.8 (2)		0.9 (1)	1.0 (1)	1.0 (1)	0.8 (1)	0.9 (1)		0.9 (1)
Leu	1.0 (1)	0.9 (1)					1.0 (1)		
Tyr	3.4 (4)	0.7 (1)	2.5 (3)				2.7 (3)	0.6 (1)	
Phe	4.8 (5)	0.8 (1)	3.6 (4)				2.9 (3)	1.7 (2)	
Lys	4.7 (5)	1.7 (2)	2.9 (3)				3.0 (3)	2.0 (2)	
Arg	4.8 (5)	0.7 (1)	1.8 (2)	1.8 (2)	1.5 (2)	1.7 (2)	1.9 (2)	0.8 (1)	1.6 (2)
Hse/Hsl		+ (1)	+ (1)						
NH_2-terminal residue	Ala	Ala	Ile	Arg	Arg	Arg	Ala	Thr	Asx
Total yield %		36	34	19	20	11	32	41	35

```
                                          -30                                              -20
Gene-derived sequence   Met-Lys-Met-Ser-Arg-Leu-Cys-Leu-Ser-Ile-Ala-Leu-Leu-Val-Leu-Leu-Gly-Thr-Leu-Ala-Ala-
Inhibitor II
Inhibitor I

         -10                                         -1  +1                                     10
Ser-Thr-Pro-Gly-Cys-Asp-Thr-Ser-Asn-Gln-Ala-Lys-Ala-Gln-Arg-Pro-Asp-Phe-Cys-Leu-Glu-Pro-Pro-Tyr-Thr-Gly-Pro-
                                                        Arg-Pro-Asp-Phe-Cys-Leu-Glu-Pro-Pro-Tyr-Thr-Gly-Pro-
                                                Ala-Gln-Arg-Pro-Asp-Phe-Cys-Leu-Glu-Pro-Pro-Tyr-Thr-Gly-Pro-

                             20                                      30                                       40
Cys-Lys-Ala-Lys-Met-Ile-Arg-Tyr-Phe-Tyr-Asn-Ala-Lys-Ala-Gly-Phe-Cys-Glu-Thr-Phe-Val-Tyr-Gly-Gly-Cys-Lys-Ala-
Cys-Lys-Ala-Lys-Met-Ile-Arg-Tyr-Phe-Tyr-Asn-Ala-Lys-Ala-Gly-Phe-Cys-Glu-Thr-Phe-Val-Tyr-Gly-Gly-Cys-Lys-Ala-
Cys-Lys-Ala-Lys-Met-Ile-Arg-Tyr-Phe-Tyr-Asn-Ala-Lys-Ala-Gly-Phe-Cys-Glu-Thr-Phe-Val-Tyr-Gly-Gly-Cys-Lys-Ala-

                                  50                                    60
Lys-Ser-Asn-Asn-Phe-Arg-Ser-Ala-Glu-Asp-Cys-Met-Arg-Thr-Cys-Gly-Gly-Ala-Ile-Gly-Pro-Arg-Glu-Asn-Leu
Lys-Ser-Asn-Asn-Phe-Arg-Ser-Ala-Glu-Asp-Cys-Met-Arg-Thr-Cys-Gly-Gly-Ala
Lys-Ser-Asn-Asn-Phe-Arg-Ser-Ala-Glu-Asp-Cys-Met-Arg-Thr-Cys-Gly-Gly-Ala-Ile-Gly-Pro-Arg-Glu-Asn
                                                                                           ↑
```

FIG. 2. **The amino acid sequence of isoinhibitor I from bovine spleen is compared with the sequences of isoinhibitor II (6) and of the primary translation product inferred from the sequence of the corresponding gene (5).** The *numbering* refers to the sequence of isoinhibitor II. The two variants found in the preparation of isoinhibitor I differ in the presence (the less abundant variant) or the absence of a COOH-terminal asparaginyl residue; this further site of processing is indicated by the *vertical arrow*.

virtually identical sequence, these authors suggested that the two genes arose by gene duplication or by gene conversion. In addition they demonstrated that the middle exons of both genes primarily encode the mature proteins, while the other two exons clearly indicate that the two proteins are biosynthesized as larger precursors, having 33 or 35 residues as NH_2-terminal extensions and 7 residues as COOH-terminal extensions. The question remained whether isoinhibitors I and III are the products of different genes or originate from one of the two genes encoding isoinhibitor II or BPTI (isoinhibitor IV), respectively, through some kind of post-translational modification. The present studies indicate the following: (i) the protein has 65 residues, has the same sequence as spleen inhibitor II with an NH_2-terminal extension of two amino acids and a COOH-terminal extension of five amino acids; (ii) a small amount (20%) of a variant containing an additional amino acid (Asn) at the COOH-terminal extension is also present; (iii) the sequences of these extensions are in agreement with that deduced from the gene which codes for isoinhibitor II (5).

In conclusion, sequence studies on both variants of isoinhibitor I help unravel the complex picture of multistage processing of the primary product of expression of the gene homologous to that coding for BPTI, processing which eventually leads to the final mature protein (isoinhibitor II). It seems that the long NH_2-terminal sequence (33 or 35 amino acids) of the gene product is the first to be removed by limited proteolysis, while the COOH-terminal extension is necessary for the final targetting of the protein, before the last processing step takes place. This hypothesis can now be tested through *in vitro* processing studies of bovine spleen inhibitor I. Obviously the hypothesis is not incompatible with the possibility that inhibitor I is a mature expression product, which displays in certain bovine tissues or subcellular locations a distinct antiproteolytic function. Additional interest in this system derives from the recent finding of minute amounts of BPTI-like isoinhibitors in human serum (16).

REFERENCES

1. Blobel, G. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 1496–1500
2. Perlman, D., and Halvorson, H. O. (1983) *J. Mol. Biol.* **167,** 391–409
3. Von Hejne, G. (1983) *Eur. J. Biochem.* **133,** 17–21
4. Fioretti, E., Binotti, I., Barra, D., Citro, G., Ascoli, F., and Antonini, E. (1983) *Eur. J. Biochem.* **130,** 13–18
5. Creighton, T. E., and Charles, I. G. (1987) *J. Mol. Biol.* **194,** 11–21
6. Fioretti, E., Iacopino, G., Angeletti, M., Barra, D., Bossa, F., and Ascoli, F. (1985) *J. Biol. Chem.* **260,** 11451–11455

7. Gray, W. R. (1972) *Methods Enzymol.* **25,** 121–138
8. Laskowski, M., Jr., and Kato, I. (1980) *Annu. Rev. Biochem.* **49,** 593–626
9. Fritz, H., and Wunderer, G. (1983) *Drug. Res.* **33,** 479–494
10. Fioretti, E., Citro, G., and Ascoli, F. (1984) *Prep. Biochem.* **14,** 389–401
11. Fioretti, E., Angeletti, M., Citro, G., and Ascoli, F. (1985) *Prep. Biochem.* **15,** 211–219
12. Businaro, R., Fioretti, E., Fumagalli, L., Citro, G., De Renzis, G., and Ascoli, F. (1987) *Eur. J. Biochem.* **165,** 79–82
13. Fritz, H., Kruck, H., Russe, J., and Liebich, H. G. (1979) *Hoppe-Seyler's Z. Physiol. Chem.* **360,** 437–444
14. Anderson, S., and Kingston, I. B. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 6838–6842
15. Kingston, I. B., and Anderson, S. (1986) *Biochem. J.* **233,** 443–450
16. Fioretti, E., Angeletti, M., Citro, G., Barra, D., and Ascoli, F. (1987) *J. Biol. Chem.* **262,** 3586–3589