Amino-Acid Sequence of Porcine Pepsin

(protein/cyanogen bromide/active site)

J. TANG, P. SEPULVEDA, J. MARCINISZYN, JR., K. C. S. CHEN*, W-Y. HUANG†, N. TAO*, D. LIU, AND J. P. LANIER

Laboratory of Protein Studies, Oklahoma Medical Research Foundation, and the Department of Biochemistry and Molecular Biology, University of Oklahoma School of Medicine, Oklahoma City, Okla. 73104

ABSTRACT   As the culmination of several years of experiments, we propose a complete amino-acid sequence for porcine pepsin, an enzyme containing 327 amino-acid residues in a single polypeptide chain. In the sequence determination, the enzyme was treated with cyanogen bromide. Five resulting fragments were purified. The amino-acid sequence of four of the fragments accounted for 290 residues. Because the structure of a 37-residue carboxyl-terminal fragment was already known, it was not studied. The alignment of these fragments was determined from the sequence of methionyl-peptides we had previously reported. We also discovered the locations of active-site aspartyl residues, as well as the pairing of the three disulfide bridges. A minor component of commercial crystalline pepsin was found to contain two extra amino-acid residues, Ala-Leu-, at the amino-terminus of the molecule. This minor component was apparently derived from a different site of cleavage during the activation of porcine pepsinogen.

Although pepsin (EC 4.3.3.1) was one of the first enzymes to be discovered and purified in crystalline form, its structure-function relationship is understood only superficially. In recent years, workers have realized that the complete elucidation of the mechanism of action of an enzyme depends upon the detailed knowledge of the chemical and three-dimensional structure of its molecule. For this reason, we undertook a long-term study of the complete amino-acid sequence of porcine pepsin.

A number of studies have contributed to our knowledge of the partial sequence of porcine pepsin. They include investigations of the amino-acid sequences at the regions near the carboxyl-terminus (1-4), the amino-terminus (5, 6), the three disulfide bonds (7), the tryptophanyl residues (8, 9), the phosphoseryl residue (7, 10), two separate active-site aspartyl residues (11, 12), and a number of small peptides (13, 14). We have confirmed most of these previous findings and now present the complete amino-acid sequence of this enzyme.

METHODS

Porcine pepsin (three-times crystallized) was obtained from Worthington. The enzyme was either alkali-denatured (7) or reduced and aminoethylated (15) and subjected to cyanogen-bromide cleavage (16). Preliminary experiments indicated that at room temperature one of the methionyl bonds (Met-Thr at residues 80-81) was cleaved less than 5% by cyanogen

* Present address: The Rockefeller University, New York, N.Y. 10021.

† Present address: The Veterans Administration Hospital, Kansas City, Mo. 64128.

bromide. The extent of cleavage was increased to 30-35% when the reaction was carried out at 37°. This reaction temperature, 37°, was used for most of the later experiments. The fragments were separated by column chromatography on Sephadex G-75 and DEAE-cellulose. The detailed procedures will be published separately (Chen, K. C. S., Huang, W.-Y., Tao, N. & Tang J., manuscript in preparation). Proteolytic digests were carried out using α-chymotrypsin, trypsin, thermolysin, subtilisin, staphylococcal protease, β-lytic protease, and papain. Some of the cyanogen bromide fragments were chemically cleaved at the tryptophanyl bonds with 2-(2-nitrophenylsulfenyl)-3-methyl-3-bromoindolenine (17). The resulting peptides were separated on Sephadex columns or/and on high-voltage electrophoresis. The amino-acid sequences of the peptides were determined by dansyl-Edman degradation, hydrolysis with carboxypeptidases, and hydrazinolysis. The conditions for these procedures were similar to those described (7, 12).

RESULTS AND DISCUSSION

The complete amino-acid sequence of porcine pepsin is shown in Fig. 1. The enzyme molecule contains 327 residues in a single polypeptide chain and has a molecular weight of 34,644. The final amino-acid composition based on this sequence is shown in Table 1.

Before we could establish the complete primary structure, the amino-acid sequences of four cyanogen bromide fragments had to be determined. These fragments are located between residue numbers 1 and 80, 81 and 199, 200 and 246, and 247 and 290. To determine their sequences, we performed from three to six different enzymatic or chemical cleavages on each cyanogen bromide fragment. Automatic Edman degradations were carried out for two cyanogen bromide fragments starting at residues 1 and 81. In each case, about 30 cycles were carried out. These results were useful in placing or confirming the peptides in these regions. The complete sequence contains six one-residue overlaps at residue numbers 114, 123, 154, 238, 262, and 275. Although the correctness of these overlaps is supported by the amino-acid compositions of the respective fragments, investigations are being conducted to provide stronger evidence for these overlaps. The fragment containing the carboxyl-terminal 37 residues (residues 291-327) was not completely determined in this laboratory because the sequence of this region was already known (1-4). However, we did obtain from the tryptophan-cleavage of this fragment a decapeptide (residue 291-300) with a sequence identical to that previously reported (4). The placement of the cyanogen bromide

fragments in the pepsin molecule was accomplished with the amino-acid sequences near the four methionyl residues. These sequences we had either previously reported (7, 12) or determined (Chen *et al.*, in preparation).

Based on our previous study (7), the three disulfide bridges are paired between residues 45 and 50, 206 and 210, and 250 and 283. The only phosphoseryl residue is at position 68, and the histidyl residue is located at position 53. The two arginyl residues and the single lysyl residue, which have been shown to be near the carboxyl-terminus, are located at residue numbers 308, 316, and 320, respectively. Altogether, 44 acidic residues (including phosphoserine) are present throughout the entire polypeptide chain, in contrast to the localization of the basic residues described above.

The epoxide-reactive, active-site aspartyl residue, which has been shown to participate in the peptic catalysis as a carboxylate ion (12, 18, 19), is at residue number 32. Another active-site aspartyl residue, which reacts with diazo inactivators (11, 20), is located at residue number 215. These aspartyl residues, and the arginyl residue at position 316, which we previously showed to be in the vicinity of the active center (21), must be near each other in the tertiary structure of the enzyme.

The sequences around the two active-center aspartyl residues are homologous, an interesting feature pointed out pre-

viously (12); in addition we have found that the sequences around the first four of the five tryptophans are also homologous. These phenomena and the evidence derived from comparison of segments within the pepsin molecule have led us to believe that this porcine enzyme has evolved from a small polypeptide chain by means of gene doubling and translocation. The evidence and significance of these aspects will be included in a separate paper (Sepulveda, P. & Tang, J., manuscript in preparation).

In our studies on the amino-acid sequence of the α-chymotryptic peptides derived from the amino-terminus of pepsin, we isolated two different peptides. The major peptide contained the amino-terminal 14 residues that have the sequence shown in Fig. 1. A minor peptide contained two additional residues at the amino-terminus but was otherwise identical in sequence. These two sequences are shown in the following:

Major yield:
Ile-Gly-Asp-Glu-Pro-Leu-Glu-Asn-Tyr-Leu-Asp-Thr-Glu-Tyr

Minor yield:
Ala-Leu-Ile-Gly-Asp-Glu-Pro-Leu-Glu-Asn-Tyr-Leu-Asp-Thr-Glu-Tyr

Pepsin from the commercial preparation apparently contains a minor component, Ala-pepsin. These two species of pepsins are obviously derived from different sites of cleavage during
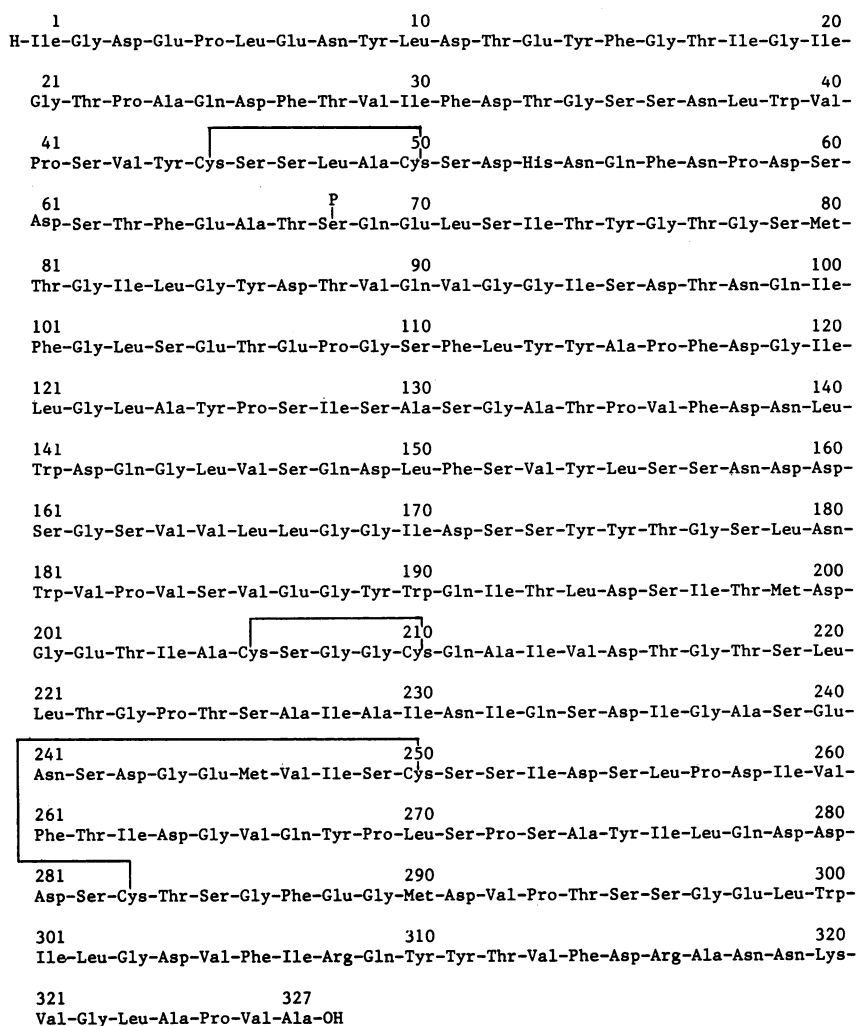
```
  1                          10                          20
H-Ile-Gly-Asp-Glu-Pro-Leu-Glu-Asn-Tyr-Leu-Asp-Thr-Glu-Tyr-Phe-Gly-Thr-Ile-Gly-Ile-

 21                          30                          40
Gly-Thr-Pro-Ala-Gln-Asp-Phe-Thr-Val-Ile-Phe-Asp-Thr-Gly-Ser-Ser-Asn-Leu-Trp-Val-

 41                          50                          60
Pro-Ser-Val-Tyr-Cys-Ser-Ser-Leu-Ala-Cys-Ser-Asp-His-Asn-Gln-Phe-Asn-Pro-Asp-Ser-

 61                    P     70                          80
Asp-Ser-Thr-Phe-Glu-Ala-Thr-Ser-Gln-Glu-Leu-Ser-Ile-Thr-Tyr-Gly-Thr-Gly-Ser-Met-

 81                          90                         100
Thr-Gly-Ile-Leu-Gly-Tyr-Asp-Thr-Val-Gln-Val-Gly-Gly-Ile-Ser-Asp-Thr-Asn-Gln-Ile-

101                         110                         120
Phe-Gly-Leu-Ser-Glu-Thr-Glu-Pro-Gly-Ser-Phe-Leu-Tyr-Tyr-Ala-Pro-Phe-Asp-Gly-Ile-

121                         130                         140
Leu-Gly-Leu-Ala-Tyr-Pro-Ser-Ile-Ser-Ala-Ser-Gly-Ala-Thr-Pro-Val-Phe-Asp-Asn-Leu-

141                         150                         160
Trp-Asp-Gln-Gly-Leu-Val-Ser-Gln-Asp-Leu-Phe-Ser-Val-Tyr-Leu-Ser-Ser-Asn-Asp-Asp-

161                         170                         180
Ser-Gly-Ser-Val-Val-Leu-Leu-Gly-Gly-Ile-Asp-Ser-Ser-Tyr-Tyr-Thr-Gly-Ser-Leu-Asn-

181                         190                         200
Trp-Val-Pro-Val-Ser-Val-Glu-Gly-Tyr-Trp-Gln-Ile-Thr-Leu-Asp-Ser-Ile-Thr-Met-Asp-

201                         210                         220
Gly-Glu-Thr-Ile-Ala-Cys-Ser-Gly-Gly-Cys-Gln-Ala-Ile-Val-Asp-Thr-Gly-Thr-Ser-Leu-

221                         230                         240
Leu-Thr-Gly-Pro-Thr-Ser-Ala-Ile-Ala-Ile-Asn-Ile-Gln-Ser-Asp-Ile-Gly-Ala-Ser-Glu-

241                         250                         260
Asn-Ser-Asp-Gly-Glu-Met-Val-Ile-Ser-Cys-Ser-Ser-Ile-Asp-Ser-Leu-Pro-Asp-Ile-Val-

261                         270                         280
Phe-Thr-Ile-Asp-Gly-Val-Gln-Tyr-Pro-Leu-Ser-Pro-Ser-Ala-Tyr-Ile-Leu-Gln-Asp-Asp-

281                         290                         300
Asp-Ser-Cys-Thr-Ser-Gly-Phe-Glu-Gly-Met-Asp-Val-Pro-Thr-Ser-Ser-Gly-Glu-Leu-Trp-

301                         310                         320
Ile-Leu-Gly-Asp-Val-Phe-Ile-Arg-Gln-Tyr-Tyr-Thr-Val-Phe-Asp-Arg-Ala-Asn-Asn-Lys-

321            327
Val-Gly-Leu-Ala-Pro-Val-Ala-OH
```

Fig. 1.   Amino-acid sequence of porcine pepsin.

TABLE 1. *Amino-acid composition of porcine pepsin based on amino-acid sequence*

Amino acid	Residues per molecule
Lysine	1
Histidine	1
Arginine	2
Aspartic Acid	30
Asparagine	12
Threonine	27
Serine	43
Phosphoserine	1
Glutamic Acid	13
Glutamine	13
Proline	15
Glycine	35
Alanine	16
Half-cystine	6
Valine	22
Methionine	4
Isoleucine	25
Leucine	27
Tyrosine	15
Phenylalanine	14
Tryptophan	5
Total	327
Molecular weight	34,644

the activation of pepsinogen. Pepsinogen has been shown to be activated by two different mechanisms, an intramolecular and a pepsin-catalyzed activation mechanism (22). It has also been suggested that the former mechanism produces only Ile-pepsin, while the latter produces pepsin molecules with different amino-termini (22). Evidence suggests that Ala-pepsin is one of the products in the pepsin-catalyzed activation of pepsinogen.

NOTE ADDED IN PROOF

We have found that the isoleucyl residue at position 230 is present in some molecules of pepsin and absent in others. The one-residue overlap at position 238 has been substantiated.

Preliminary reports of this work were presented in refs. 23 and 24. We thank Dr. Robert Heinrikson of University of Chicago for the sequenator analysis of cyanogen bromide fragments.

1. Dopheide, T. A. A., Moore, S. & Stein, W. H. (1967) *J. Biol. Chem.* **242**, 1833–1837.
2. Perham, R. N. & Jones, G. M. T. (1967) *Eur. J. Biochem.* **2**, 84–89.
3. Matveeva, R. A., Krivtsov, V. E. & Stepanov, V. M. (1968) *Biokhimiya* **33**, 167–174.
4. Kostka, V., Morávek, L. & Sorm, F. (1970) *Eur. J. Biochem.* **13**, 447–454.
5. Tang, J. (1970) *Biochem. Biophys. Res. Commun.* **41**, 697–703.
6. Stepanov, V. M., Timokhina, E. A., Baratova, L. A., Belyanova, L. P., Korzhenko, V. P. & Zhukova, I. G. (1971) *Biochem. Biophys. Res. Commun.* **45**, 1482–1487.
7. Tang, J. & Hartley, B. S. (1970) *Biochem. J.* **118**, 611–623.
8. Dopheide, T. A. A. & Jones, W. M. (1968) *J. Biol. Chem.* **243**, 3906–3911.
9. Vesenev, V. I., Kuznetsov, S., Jr. & Stepanov, V. M. (1970) *Biokhimiya* **35**, 787–796.
10. Vakhitova, É. A., Amirkhanyan, M. M. & Stepanov, V. M. (1970) *Biokhimiya* **35**, 1164–1169.
11. Bayliss, R. S., Knowles, J. R. & Wybrandt, G. B. (1969) *Biochem. J.* **113**, 377–386.
12. Chen, K. S. C. & Tang, J. (1972) *J. Biol. Chem.* **247**, 2566–2574.
13. Stepanov, V. M., Amirkhanyan, M. M., Belenkii, B. G., Valyulis, R. A., Vaknitova, É. A., Pugacheva, I. B. & Senyutenkova, L. G. (1970) *Biokhimiya* **35**, 283–290.
14. Balyulis, R. A. & Stepanov, V. M. (1971) *Biokhimiya* **36**, 358–362.
15. Slobin, L. I. & Singer, S. J. (1968) *J. Biol. Chem.* **243**, 1777–1786.
16. Steers, E., Jr., Craven, Z. R., Anfinsen, C. B. & Bethune, J. L. (1965) *J. Biol. Chem.* **240**, 2478–2484.
17. Fontana, A. (1972) *Methods in Enzymology*, eds. Hirs, C. H. W. & Timasheff, S. N. (Academic Press, New York), Vol. XXV, pp. 419–423.
18. Tang, J. (1971) *J. Biol. Chem.* **246**, 4510–4517.
19. Hartsuck, J. A. & Tang, J. (1972) *J. Biol. Chem.* **247**, 2575–2580.
20. Rajagopalan, T. G., Moore, S. & Stein, W. H. (1966) *J. Biol. Chem.* **241**, 4940–4950.
21. Huang, W-Y. & Tang, J. (1972) *J. Biol. Chem.* **247**, 2704–2710.
22. Al-Janabi, J., Hartsuck, J. A. & Tang, J. (1972) *J. Biol. Chem.* **247**, 4628–4632.
23. Tang, J., Huang, W-Y., Sepulveda, P., Chen, K. C. S., Marciniszyn, J., Jr., Tao, N. & Liu, D. (1973) *Fed. Proc.* **32**, 2035.
24. Tang, J., Sepulveda, P., Marciniszyn, J., Jr. & Liu, D. (1973) *Abstract of the Ninth International Congress of Biochemistry*, Stockholm, July 1–7, p. 47.