BOVINE PEPSIN: THE SEQUENCE OF THE FIRST 65 AMINO ACID RESIDUES (COMPLETING THE SEQUENCE OF THE FIRST 110 RESIDUES OF BOVINE PEPSINOGEN)

Marianne K. HARBOE and B. FOLTMANN,
Institute of Biochemical Genetics, University of Copenhagen, Ø.Farimagsgade 2A, DK-1353 Copenhagen K, Denmark

Received 17 August 1975

1. Introduction

In a previous communication [1] we reported the ultimate N-terminal amino acid sequence of bovine pepsin* together with the sequence overlapping the activation segment. The latter sequence was obtained from a fragment of 60 residues isolated after tryptic digestion of maleylated pepsinogen. The total sequence of this fragment has now been established and we have been able to determine the sequence of the succeeding 10 residues in the peptide chain of pepsin. Together with the N-terminal amino acid sequence of pepsinogen [2] this completes the sequence of the first 110 residues of the zymogen. The predicted homology with chymosin [3] has been confirmed.

2. Materials and methods

2.1. *Enzymes and reagents*

Pepsinogen and pepsin were from the same preparations as used in [1], chymotrypsin was purchased from Worthington Biochemicals Inc. and elastase from Whatman Biochemicals Ltd. Reagents and chromatographic equipment were the same as used in [2] and [4].

*Where nothing else is stated 'pepsinogen' and 'pepsin' will in the following text refer to bovine pepsinogen A and pepsin A (EC 3.4.23.1), the predominant protease of adult bovine gastric juice.

Abbreviation: Dns(dansyl), 5-dimethylaminonapthalene-1-sulfonyl.

2.2. *Digestions*

The tryptic and the thermolytic digests were the same as described in [1]. Thermolytic peptides are marked by Th. The tryptic peptides from maleylated pepsinogen are designated TM-1 to TM-7, numbered from the N-terminus of the zymogen. TM-1 to TM-3 are accounted for in [2]. TM-4 is described in this communication. Chymotryptic digestion was carried out in 0.04 M *N*-ethylmorpholine acetate, pH 8.0, for 3 h at 37°C, peptide/enzyme ratio 250/1 (mol/mol). Elastase digestion: 0.2 M *N*-ethylmorpholine acetate, pH 8.0, for 30 min at 37°C, peptide/enzyme ratio 200/1 (mol/mol). The peptides from these digestions are marked by C and E respectively.

2.3. *Purification of peptides*

The large tryptic fragments were purified by gel filtration on Sephadex G-100 [1]. Low molecular weight peptides were purified by high voltage paper electrophoresis/paper chromatography [4]. The net charge of the peptides were evaluated from the electrophoretic mobilities at pH 6.5 [5], these are expressed relative to DnsOH (−1.0) [6]. Cystine peptides were purified by the cysteic acid diagonal method [7].

2.4. *Amino acid analyses and sequencing*

Amino acid analyses were performed with a Durrum D-500 analyzer after hydrolysis for 20 h with constant boiling HCl in sealed, evacuated tubes. To condense this paper we have omitted tables on the amino acid composition of the individual peptides. However, the results were of grade 'a' in the classification suggested by Ambler [8], i.e. the analyses show-

ed stoichiometric ratios of amino acid residues of a given peptide which agreed with integers better than ± 10%; no impurities were present in amounts as great as 0.2 mol/mol.

The peptides were sequenced by manual Edman degradation followed by dansylation and chromatography of the Dns-amino acids on polyamide layer sheets [6] and/or conversion of the liberated thiazolinone amino acids into free amino acids [9]. Conversion with HI was carried out in sealed evacuated tubes [2] and the analyses took place as ordinary amino acid analyses.

3. Results and discussion

The results are illustrated in fig.1. To facilitate alignment of the amino acid sequences of the gastric proteases and their zymogens the residues are numbered according to the numbering used previously starting from the N-terminus of prochymosin [4]. Chymotrypsin and elastase digestions were performed on TM-4. The location of most amino acid residues has been obtained independently in three or four peptides, but to compress the results only such peptides are shown or mentioned that are necessary for overlaps or contribute to location of amides.

The sequence from Glu_{42} to Tyr_{55} has been published previously [1], but our present results contain a correction to the location of an amide group in Th-1. After two and three Edman degradations the remaining peptide was purified by paper electrophoresis at pH 6.5. After two Edman degradations an acidic peptide with the composition 2.0 Glx and 0.9 Pro was isolated while three Edman degradations gave a neutral peptide consisting of 1.0 Glx and 0.9 Pro. This shows that residue no. 50 must be Gln.

Peptide C-3 had a mobility of −0.95, indicating a net charge of −2. The overlaps around Phe_{61} represent a single residue overlap. However, TM-4 contains only three phenylalanines and since the overlaps around the two others are firmly established the sequence shown is the only one possible. The amide in Gln_{71} and Asp_{72} was determined by the electrophoretic mobility of C-4 after ten and eleven Edman degradations (−0.85, −1.00 respectively). Th-3 had a mobility of −0.5, but

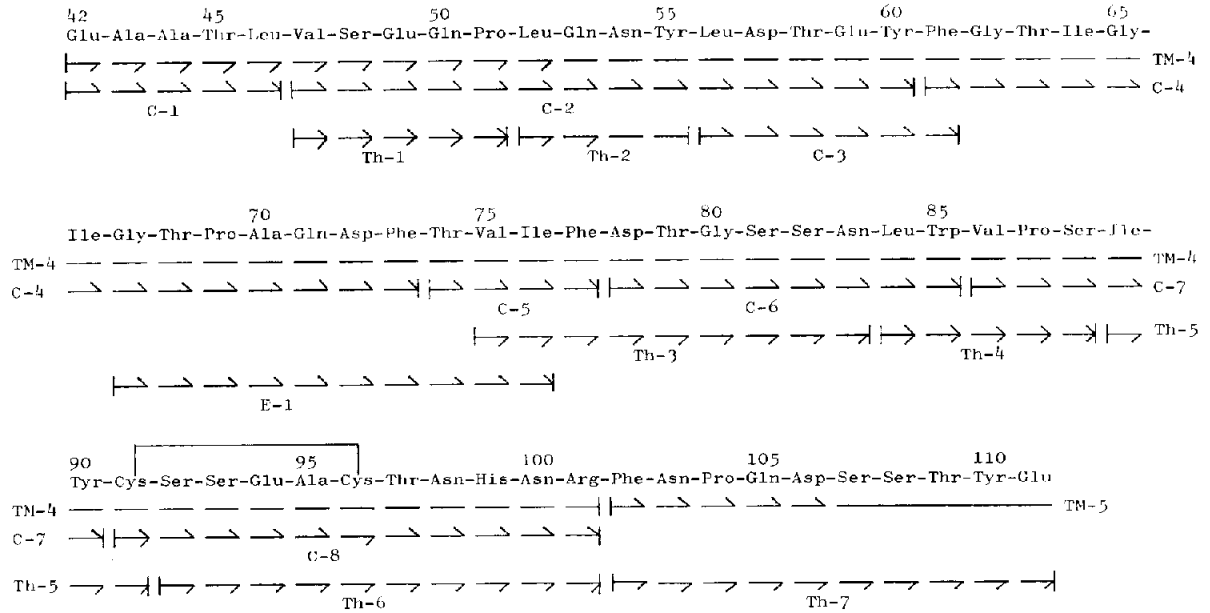

Fig.1. The amino acid sequence from residue no. 42 to residue no. 111 of bovine pepsinogen. The N-terminus of active pepsin is Val_{47}. Notation: (——→) sequential Edman degradation/dansylation, (——►) sequential Edman degradation with identification of the thiazolinone derivatives after conversion to free amino acids. (———) present from amino acid analyses.

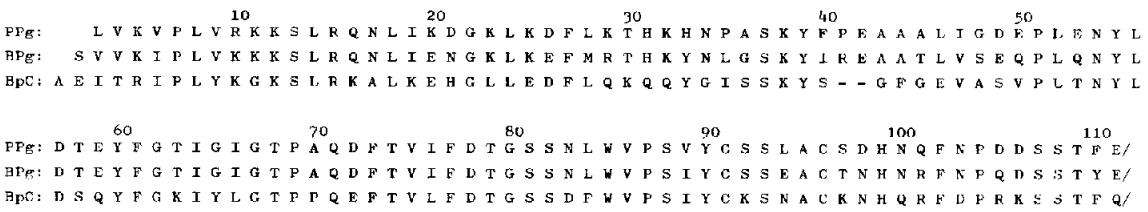

Fig.2. Comparison between the N-terminal amino acid sequences of porcine pepsinogen (PPg) [12–15], bovine pepsinogen (BPg) ([2] and this paper) and prochymosin (BpC) [4,10]. The results are expressed in the single letter code: A (Ala), C (Cys), D (Asp), E (Glu), F (Phe), G (Gly), H (His), I (Ile), K (Lys), L (Leu), M (Met), N (Asn), P (Pro), Q (Gln), R (Arg), S (Ser), T (Thr), V (Val), W (Trp), Y (Tyr). The N-termini of the pepsins are I_{47} and V_{47}, while chymosin starts at G_{45}.

after four Edman degradations the peptide became neutral. The S-S loop Cys_{91} to Cys_{96} was isolated as two off diagonal cycsteic acid peptides in the thermolytic digest and one off diagonal peptide after digestion with chymotrypsin. The electrophoretic mobility of Th-6 changed from −0.20 to +0.25 after the third Edman degradation.

The connection between Arg_{101} and Phe_{102} is obtained from indirect but solid evidence. Bovine pepsinogen contains 6 arginine residues, three are found in the activation segment [2] and two are accounted for in the last 20 residues [10,11]; thus Arg_{101} must precede the large tryptic fragment (TM-5, peak II [1]) and the N-terminal sequence of this is identical to that of Th-7. The electrophoretic mobility of Th-7 was −0.9; after three, four and five Edman degradations the mobilities were −1.0, −1.1 and −0.7 respectively.

The peptide bonds cleaved by the enzymatic degradations used in this study correspond to the expected except that it was surprising that elastase cleaved after Ile_{66} and Ile_{76}. 30 nmol of the corresponding peptide was recovered after paper electrophoresis at pH 6.5 and pH 2 from a digest of 100 nmol of TM-4.

The homology among the three gastric zymogens analysed so far is summarized in fig.2. The obvious homology between the pepsins substantiates the deduced connection between Arg_{101} and Phe_{102}. Out of 111 positions, 51 are identical in all three structures, while 84 are identical in the two pepsinogens. It is further observed that a major part of the substitutions is conservative in the sense that polarity of the side chains is maintained.

It has been reported that in porcine pepsin [16] and in chymosin [17] Asp_{78} has a special reactivity for labelling with epoxy compounds; it is noteworthy that this residue is located in highly conservative surroundings corresponding to what one would expect for an active center residue.

Acknowledgements

This investigation has been supported by The Carlsberg Foundation and by The Danish Natural Science Research Council.

References

[1] Harboe, M. and Foltmann, B. (1973) FEBS Lett. 34, 311–314.
[2] Harboe, M., Andersen, P. M., Foltmann, B., Kay, J. and Kassell, B. (1974) J. Biol. Chem. 249, 4487–4494.
[3] Pedersen, V. B. and Foltmann, B. (1973) FEBS Lett. 35, 250–254.
[4] Pedersen, V. B. and Foltmann, B. (1975) Eur. J. Biochem. 55, 95–103.
[5] Offord, R. E. (1966) Nature (London) 211, 591–593.
[6] Hartley, B. S. (1970) Biochem. J. 119, 805–822.
[7] Brown, J. R. and Hartley, B. S. (1966) Biochem. J. 101, 214–228.
[8] Ambler, R. P. (1974) Biochem. J. 137, 3–14.
[9] Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G. and Ballantyne, D. L. (1971) Biochemistry 10, 4912–4921.
[10] Foltmann, B., Kauffman, D., Parl, M. and Andersen, P. M. (1973) Neth. Milk Dairy J. 27, 288–297.
[11] Parl, M., Andersen, P. M. and Foltmann, B. (1973) Abstr. 9th Int. Congr. Biochem. (Stockholm), 95.

[12] Ong, E. B. and Perlmann, G. E. (1968) J. Biol. Chem. 243, 6104–6109.
[13] Pedersen, V. B. and Foltmann, B. (1973) FEBS Lett. 35, 255–256.
[14] Moravek, L. and Kostka, V. (1974) FEBS Lett. 43, 207–211.
[15] Tang, J., Sepulveda, P., Marciniszyn, J., Chen, K. C. S., Huang, W-Y., Tao, N., Liu, D. and Lainer, J. P. (1973) Proc. Nat. Acad. Sci. U.S.A. 70, 3437–3439.
[16] Chen, K. C. S. and Tang, J. (1972) J. Biol. Chem. 247, 2566–2574.
[17] Chang, W-J. and Takahashi, K. (1974) Biochem. J. (Tokyo) 76, 467–474.