

COMPLETE AMINO ACID SEQUENCE OF HOG PEPSIN

L. MORÁVEK and V. KOSTKA

*Institute of Organic Chemistry and Biochemistry, Czechoslovak Academy of Sciences,
166 10 Prague 6, Czechoslovakia*

Received 9 April 1974

1. Introduction

The amino acid sequence of hog pepsin or of some parts of its polypeptide chain has been studied in a number of laboratories. The most recent summary of the results obtained is presented in the 1972 issue of the Atlas of Protein Sequence and Structure [1]. Our aim has been to determine the complete covalent structure of this enzyme and thus to provide, together with the complete structures of bovine trypsinogen [2] and chymotrypsinogen [3] elucidated in this laboratory earlier, complete and independent sequence information on the fundamental proteolytic enzymes of the digestive tract.

The determination of the complete structure of the whole molecule was based on the results of sequential studies on large fragments designated CB [4] obtained by cyanogen bromide cleavage of pepsin at its four methionine residues [5] (fig. 1). The studies on the individual cyanogen bromide fragments were begun after the determination of the disulfide bonds of the 6 half-cystine residues of pepsin [6]. The complete amino acid sequence of fragment CB1 [7], representing the 37-residue C-terminal part of the molecule, was determined. Later, the 55-residue N-terminal amino acid sequence of pepsin [8,9] was

established by sequential analysis of fragments resulting from tryptic digestion of aminoethylated fragment CB2 [10]. The information on the disulfide bonds [6] was extended by the determination of longer amino acid sequences around the half-cystine residues [11]. The amino acid sequence of fragment CB5 was determined next [12]. Simultaneously the order of all cyanogen bromide fragments in the terminal sequences of these fragments [4] and of the sequence around the methionine residues [13]. The methionine sequences were not reconcilable completely with the earlier data of other authors [14]. In addition to the main line of approach — the study of cyanogen bromide fragments large tryptic fragments of aminoethylated pepsin [15] and the thermolysin [16] and chymotryptic [17] peptides derived from the whole *S*-sulfo-pepsin were investigated.

The sequential information on pepsin and other, unpublished, data was summarized in the form of a tentative structure, involving 324 amino acid residues, which appeared in September 1973 [18]. This structure comprised the 113-residue N-terminal amino acid sequence and the 96-residue C-terminal sequence. These regions account for complete amino acid sequences of fragments CB4, CB5 and CB1, and for a part of fragment CB3. (Compared to the amino acid

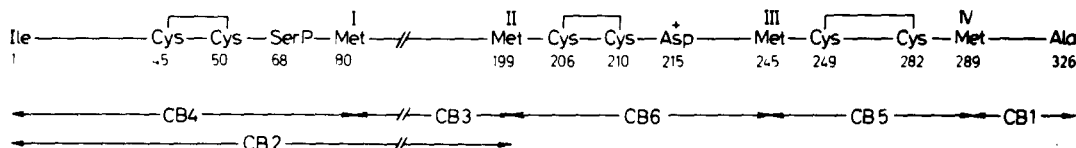


Fig. 1. Location of certain important amino acid residues and arrangement of cyanogen bromide fragments (marked CB) [4] in pepsin chain. Serine residue No. 68 was found to be phosphorylated [29]. Aspartic acid residue No. 215 (marked by a cross) in sequence . . Ile-Val-Asp-Thr-Gly-Thr-Ser . . is a part of the active site of pepsin [30].

sequence of fragment CB5 submitted to the 'Atlas' [12], a serine residue was placed to position No. 76 [18]. Likewise, residue No. 64, originally reported to be aspartic acid [12], was found to be an asparagine). The tentative amino acid sequence of fragment CB6 [19] was presented in the form of three sections of noninterchangeable order. The completion of the amino acid sequence of hog pepsin required therefore the arrangement of shorter sequences within fragment CB3 and the determination of the two missing links of fragment CB6.

Important progress in the sequential studies of hog pepsin has been made also by Tang and coworkers who, in July 1973, proposed a 355-residue structure [20]. This structure was based on the amino acid sequences of the first four cyanogen bromide fragments from the N-terminus, as determined by the authors [20], and on the amino acid sequence of the remaining cyanogen bromide fragment CB1 (fig. 1). This 37-residue sequence was elucidated completely in our laboratory [7] and partly also by other authors [21–23]. The assumed number of 355 residues in the pepsin chain was considerably higher than would be expected having regard to the amino acid composition [5] and molecular weight [24] of hog pepsin. In December 1973, Tang and coworkers reported another amino acid sequence of hog pepsin comprising 327 residues [25]; this sequence, similarly to that, published earlier [20], uses data of other authors on the C-terminal region [7,21–23].

Sequences established by Stepanov and coworkers were summarized by the author [26] and in the 'Atlas' [1] and are subject of some more recent reports, cf., e.g. [27,28].

In this paper, the complete amino acid sequence of hog pepsin is presented, based entirely on the results obtained in this laboratory.

2. Results and discussion

The information necessary for the completion of the amino acid sequence of hog pepsin was obtained by the analysis of peptides derived from the region of the chain between Met I and Met II (fig. 1), prepared by digestion with subtilisin [15], thermolysin [19], and *N*-bromosuccinimide [19], and by the amino acid analysis of purified cyanogen bromide frag-

ments CB3 and CB6 [19]. Detailed data on sequential analysis of the individual sections of the pepsin chain will be given in a series of experimental papers.

The arrangement of the interchangeable shorter sequences falling into fragment CB3 [18] is shown in the scheme in fig. 2. Two alterations were made in the amino acid sequences of the individual regions [18]. Thermolysin peptide (Phe, Asx, Gly, Ile, Leu) Gly was found to be C-terminated with glycine and therefore the sequence of a chymotryptic peptide (residues No. 118–123, fig. 2) was reexamined and shown to be Asp–Gly–Ile–Leu–Gly–Leu. From these results the order of residues No. 121 and 122 was changed to . . . Leu–Gly . . . The analysis of a chymotryptic peptide derived from the sequential region originally reported as Leu–Ser–Ser–Asn–Asp–Gly–Val–Leu [18], permitted us to extend this region to Leu–Ser–Ser–Asn–Asp–Asp–Ser–Gly–Ser–Val–Val–Leu (fig. 2, residues No. 155–166). The complete amino acid sequence of pepsin resulted from the arrangement of the known, unlinked sequential regions, separated in the tentative sequence [18] by vertical bars.

The link between tyrosine residues No. 113 and 114 provides the subtilisin peptide Leu–Tyr–Tyr–Ala–Pro–Phe–Asp–Gly (Ile, Leu, Gly). The next link follows from the 2-residue overlap in positions No. 122 and 123. By using the thermolysin peptide Phe–Ser and the subtilisin peptide Asp–Leu–Phe–Ser–Val–Tyr, an extension of the continuous sequence at phenylalanine No. 151 is obtained. The order of the remaining two sections between Tyr No. 154 and the C-terminal sequence Leu–Asp–Ser–Ile–Thr–Met (residues No. 194–199) was established from the results of carboxypeptidase A cleavage of fragments CB2 and CB3. The links at positions No. 154/155 and 166/167, and the overlap involving positions No. 122–123 are derived from the agreement between the amino acid sequence proposed and the amino acid composition of purified fragment CB3 [19]. As supporting evidence serves the finding that the bond at the amino terminus of Leu No. 155 is susceptible to cleavage both by chymotrypsin and *N*-bromosuccinimide. The latter observation requires the presence of a tyrosine residue at the amino-terminal side of Leu No. 155, because the complete knowledge of sequences around tryptophans excludes tryptophan as an alternative. Similarly, the suscepti-



Fig. 2. Arrangement of interchangeable sequence regions [18] falling into cyanogen bromide fragment CB3. Individual regions are marked by lines.

bility of the bond at the amino terminus of Leu No. 167 to chymotrypsin is reconcilable with the proposed bond Leu (166)–Leu (167).

The linkage of the three noninterchangeable regions corresponding to the part of the molecule between Met II and Met III follows from the agreement of the total sum of amino acid residues in these three regions with the amino acid composition of purified fragment CB6 [19]. The finding of the chymotryptic peptide (Ala, Ile, Val, Asp, Thr₂, Gly, Ser, Leu) Leu (res. No. 212–221) shows that the leucine residues in positions 220–221 do not overlap. The complete amino acid sequence of hog pepsin contains 326 residues, i.e. 2 residues more than the tentative sequence [18]; the extension of the latter by 4 residues at positions No. 160–166 was compensated partly by the residue overlap at positions No. 122–123.

When our amino acid sequence of pepsin is compared with that of the 4 cyanogen bromide fragments sequenced by Tang and coworkers [25], certain differences in this region of the pepsin chain emerge. These authors report that positions No. 60–61 are

occupied by the sequence . . Ser–Asp . . , whereas the information obtained from two independent experiments in our laboratory shows the sequence . . Asp–Ser . . . A comparison with the corresponding region [26] of calf chymosin (rennin, EC 3.4.4.3), a protein showing a high degree of sequential homology with hog pepsin, favors the version reported by us:

pepsin (this paper)	. . PRO–Asp–Asp–SER–SER–THR–PHE . .
cymosin [26]	. . PRO–Arg–Lys–SER–SER–THR–PHE . .
pepsin [25]	. . PRO–Asp–Ser–Asp–SER–THR–PHE . .

Tang and coworkers also report [25] that 'the isoleucyl residue at position No. 230 is present in some molecules of pepsin and absent in others'. We analyzed 4 peptides derived from this region in this laboratory and did not find this isoleucine residue in any of the peptides. This observation leads us to conclude that this isoleucine residue cannot be includ-

Ile - Gly - Asp-Glu-Pro-Leu-Glu-Asn-Tyr¹⁰-Leu-Asp-Thr-Glu-Tyr-Phe-Gly-Thr-Ile-Gly²⁰-Ile-Gly-Thr-Pro-Ala-Gln-
 Asp-Phe-Thr-Val³⁰-Ile-Phe-Asp-Thr-Gly-Ser-Ser-Asn-Leu-Trp-Val⁴⁰-Pro-Ser-Val-Tyr-Cys-Ser-Ser-Leu-Ala-Cys⁵⁰-
 Ser-Asp-His-Asn-Gln-Phe-Asn-Pro-Asp-Asp-Ser-Ser-Thr-Phe-Glu-Ala-Thr-Ser-Gln-Glu-Leu-Ser-Ile-Thr-Tyr-
 Gly-Thr-Gly-Ser-Met⁸⁰-Thr-Gly-Ile-Leu-Gly-Tyr-Asp-Thr-Val-Gln-Val-Gly-Gly-Ile-Ser-Asp-Thr-Asn-Gln-Ile¹⁰⁰-
 Phe-Gly-Leu-Ser-Glu-Thr-Glu-Pro-Gly-Ser¹¹⁰-Phe-Leu-Tyr-Tyr-Ala-Pro-Phe-Asp-Gly-Ile¹²⁰-Leu-Gly-Leu-Ala-Tyr-
 Pro-Ser-Ile-Ser¹³⁰-Ala-Ser-Gly-Ala-Thr-Pro-Val-Phe-Asp-Asn-Leu-Trp-Asp-Gln-Gly-Leu-Val-Ser-Gln-Asp-Leu¹⁵⁰-
 Phe-Ser-Val-Tyr-Leu-Ser-Ser-Asn-Asp-Asp-Ser-Gly-Ser-Val-Val-Leu-Leu-Gly-Gly¹⁷⁰-Ile-Asp-Ser-Ser-Tyr-Tyr-
 Thr-Gly-Ser-Leu-Asn-Trp-Val-Pro-Val-Ser-Val-Glu-Gly-Tyr-Trp-Gln-Ile-Thr-Leu-Asp-Ser-Ile-Thr-Met-Asp²⁰⁰-
 Gly-Glu-Thr-Ile-Ala-Cys-Ser-Gly-Gly²¹⁰-Cys-Gln-Ala-Ile-Val-Asp-Thr-Gly-Thr-Ser-Leu-Leu-Thr-Gly-Pro-Thr-
 Ser-Ala-Ile-Ala-Asn-Ile-Gln-Ser-Asp-Ile-Gly-Ala-Ser-Glu-Asn-Ser-Asp-Gly-Glu-Met-Val-Ile-Ser-Cys-Ser²⁵⁰-
 Ser-Ile-Asp-Ser-Leu-Pro-Asp-Ile²⁶⁰-Val-Phe-Thr-Ile-Asn-Gly-Val-Gln-Tyr-Pro-Leu-Ser-Pro-Ser-Ala-Tyr-Ile-
 Leu-Gln-Asp-Asp-Asp-Ser-Cys-Thr-Ser²⁸⁰-Gly-Phe-Glu-Gly-Met-Asp-Val-Pro-Thr-Ser-Ser-Gly-Glu-Leu-Trp-Ile³⁰⁰-
 Leu-Gly-Asp-Val-Phe-Ile-Arg-Gln-Tyr³¹⁰-Tyr-Thr-Val-Phe-Asp-Arg-Ala-Asn-Asn-Lys-Val³²⁰-Gly-Leu-Ala-Pro-Val-
 Ala³²⁶

Disulfide bonds: Cys(45)-Cys(50), Cys(206)-Cys(210), Cys(249)-Cys(282)

Fig. 3. Amino acid sequence of hog pepsin.

ed in the amino acid sequence of the major pepsin component. The last difference pertains to residue No. 263. This residue was found to be an aspartic acid [12] in sequential work on fragment CB5, in agreement with the data of Tang and coworkers [25]. However, peptide Thr-Ile-Asn-Gly-Val-Gln-Tyr (Pro,Leu,Ser,Pro,Ser,Ala)Tyr was isolated later from the chymotryptic digest [17] of *S*-sulfo-pepsin and residue No. 263 was shown to be an asparagine from the determination of the net charge of the peptide and independently from sequential degradation experiments.

The existence of Ala-pepsin [25], reported to constitute a minor component of hog pepsin, was confirmed by the isolation of peptide Ala-Leu-Ile-Gly-Asp-Glu-Pro-Leu-Glu-Asn-Tyr (Leu,Asx,Thr,Glx,Tyr,Phe) from the chymotryptic digest [17] of *S*-sulfo-pepsin.

References

- [1] Atlas of Protein Sequence and Structure (Dayhoff, M. O., ed.), National Biomedical Research Foundation, Silver Spring, Md.
- [2] Mikeš, O., Tomášek, V., Holeyšovský, V. and Šorm, F. (1967) Collection Czech. Chem. Commun. 32, 655.
- [3] Meloun, B., Kluh, I., Kostka, V., Morávek, L., Prusik, Z., Vaněček, J., Keil, B. and Šorm F. (1966) Biochim. Biophys. Acta 130, 543.
- [4] Morávek, L. and Kostka, V. (1973) Collection Czech. Chem. Commun. 38, 304.
- [5] Rajagopalan, T. G., Moore, S. and Stein, W. H. (1966) J. Biol. Chem. 241, 4940.
- [6] Keil, B., Morávek, L. and Šorm F. (1967) Collection Czech. Chem. Commun. 32, 1968.
- [7] Kostka, V., Morávek, L. and Šorm, F. (1970) Eur. J. Biochem. 13, 447.
- [8] Morávek, L., Submitted to the 'Atlas', September 1971.
- [9] Morávek, L., Collection Czech. Chem. Commun., in the press.
- [10] Morávek, L. (1972) FEBS Letters 23, 337.
- [11] Morávek, L., Collection Czech. Chem. Commun., in the press.
- [12] Morávek, L., Submitted to the 'Atlas', May 1972.
- [13] Morávek, L. and Kostka, V., Collection Czech. Chem. Commun., in the press.
- [14] Tang, J. and Hartley, B. S. (1970) Biochem. J. 118, 611.
- [15] Kostka, V., unpublished results.
- [16] Morávek, L. and Kysilka, Č., Collection Czech. Chem. Commun., in the press.

- [17] Kostka, V. and Salvetová, A., Collection Czech. Chem. Commun., in the press.
- [18] Morávek, L. and Kostka, V. (1973) FEBS Letters 35, 276.
- [19] Morávek, L., unpublished results.
- [20] Tang, J., Sepulveda, P., Marciniszyń Jr., J. and Liu, D. (1973) 9th International Congress of Biochemistry, Stockholm, Abstract 2all.
- [21] Dopheide, T. A. A., Moore, S. and Stein, W. H. (1967) J. Biol. Chem. 242, 1833.
- [22] Perham, R. N. and Jones, G. M. T. (1967) Eur. J. Biochem. 284.
- [23] Matveeva, R. A., Krivtsov, V. E. and Stepanov, V. M. (1968) Biokhimiya 33, 167.
- [24] Williams, R. C., Jr. and Rajagopalan, T. G. (1966) J. Biol. Chem. 241, 4951.
- [25] Tang, J., Sepulveda, P., Marciniszyń Jr., J., Chen, K. C. S., Huang, W.-Y., Tao, N., Liu, D. and Lanier, J. P. (1973). Proc. Natl. Acad. Sci. U.S. 70, 3437.
- [26] Ostoslavskaya, V. I., Amirkhanyan, M. M., Valulis, R. A., Vasenev, V. I., Vakhitova, E. I., Katrukha, S. P., Kuznetsov, Yu. S., Lapuk, Ya. I., Levin, E. D. Matveeva, R. A., Muratova, G. L., Pugacheva, I. B., Surova, I. A. and Stepanov, V. M. (1970) IUPAC 7th International Symposium on the Chemistry of Natural Products, Riga, Abstract A35.
- [27] Katrukha, S. P. and Stepanov, V. M. (1972) Biokhimiya 37, 329.
- [28] Stepanov, V. M., Baratova, L. A., Pugacheva, I. B., Belyanova, L. P., Revina, L. P. and Timokhina, E. A. (1973) Biochem. Biophys. Res. Commun. 54, 1134.
- [29] Verdier, C. H. (1954) Acta Chem. Scand. 8, 1302.
- [30] Bayliss, R. S., Knowles, J. R. and Wybrandt, G. B. (1969) Biochem. J. 113, 377.