TENTATIVE AMINO ACID SEQUENCE OF HOG PEPSIN

L. MORÁVEK and V. KOSTKA
Institute of Organic Chemistry and Biochemistry, Czechoslovak Academy of Sciences, Prague, Czechoslovakia

Received 2 July 1973

The collection of sequential data in the 'Atlas of Protein Sequence and Structure' [1] is lacking information on the complete amino acid sequence of pepsin (EC 3.4.4.1) even though this problem has been studied in several laboratories for a number of years. The relatively slow progress of sequential work on pepsin can be ascribed to a certain degree to peculiarities of its amino acid composition. The low arginine and lysine content of pepsin [2] considerably reduces the applicability of its cleavage by trypsin.

Important sequential studies on hog pepsin have been carried out in the laboratories of S. Moore and W.H. Stein and of G.E. Perlmann at the Rockefeller University, in the laboratory of V.M. Stepanov, and in the laboratories of J. Tang and B.S. Hartley. The most recent summary of these results is presented in the 1972 issue of the 'Atlas' [1].

The progress of independent sequential work on hog pepsin carried out in our laboratory has enabled us to derive a tentative structure which accounts for 324 amino acid residues and is reported in this short communication. The hitherto unpublished experimental details of our sequential work as well as a comparison of our results with the data of other authors will be reported in a series of forthcoming papers.

The tentative amino acid sequence of hog pepsin presented here (fig. 1) is essentially based on the results of studies on the cyanogen bromide fragments of S-sulfo-pepsin [3], complemented by data obtained in experiments with the tryptic digestion of aminoethylated pepsin [4]. Smaller peptides were isolated in relatively large quantities from the thermolysin [5] and chymotryptic [6] digest of S-sulfo-pepsin. The links between the individual half-cystine residues were determined in our laboratory [7] in the early stage of work.

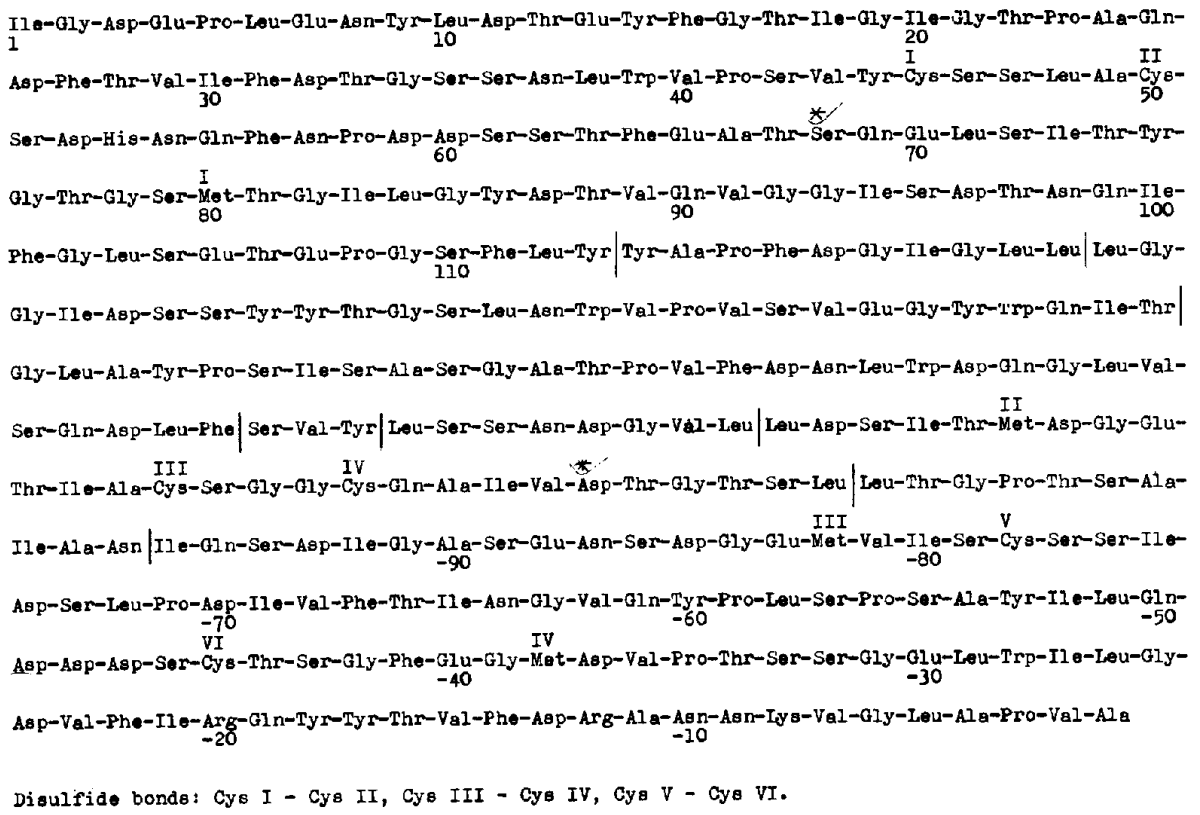
In the process of the arrangement of sequential data certain more specific problems were also studied. Sequences around the methionine residues [4, 8] have helped us to determine the order of the cyanogen bromide fragments. The presence or absence of amides was checked by analysis of peptides from various types of digests [4].

The cyanogen bromide cleavage of S-sulfo-pepsin at its four methionine residues gives rise to five specific fragments. Characteristics of these fragments were published in a previous paper [3]. In addition to the above five cyanogen bromide fragments, one more fragment was isolated; the latter includes the *N*-terminal region of pepsin up to methionine residue No. II (fig. 1). The existence of this fragment indicates incomplete cyanogen bromide cleavage of the bond Met-Thr involving methionine residue No. I. A similar observation was made in experiments with the cyanogen bromide cleavage of a large tryptic fragment of aminoethylated pepsin [4] derived from the same region of the pepsin molecule.

The systematic sequential studies of the individual cyanogen bromide fragments were begun with the *C*-terminal fragment. The complete sequence of its 37 residues, involving the two arginines and the only lysine of pepsin, was determined [9].

In subsequent work our attention was focused on the *N*-terminal region of pepsin containing two half-cystine residues [3]. After a preliminary study [10], the cyanogen bromide fragment involving the *N*-terminal region of pepsin was aminoethylated and subjected to tryptic digestion [11]. All three fragments resulting from this digestion were isolated and characterized. One of the two larger tryptic fragments which has Ser No. 51 as the *N*-terminus (fig. 1), was found

North-Holland Publishing Company – Amsterdam



Disulfide bonds: Cys I - Cys II, Cys III - Cys IV, Cys V - Cys VI.

Fig. 1. Tentative amino acid sequence of hog pepsin. Amino acid residues included in the completely determined terminal regions are numbered. Negative numbers are used to mark the residues in the carboxyl region, starting at the *C*-terminus. Roman numerals denote individual residues of half-cystine and methionine. Vertical bars mark sites of missing overlaps. The positions of the sequences in the region between Met I and Met II are not unambiguous; the order of the sequences between Met II and Met III is final. The serine residue (No. 68) which is marked by an asterisk, was found to be phosphorylated (cf. ref. [12]). An asterisk is also used to mark the aspartic acid residue in the sequence Ile–Val–Asp–Thr–Gly–Thr–Ser, reported to be part of the active site of pepsin [16].

to contain the phosphoserine residue of pepsin [12]. In subsequent studies the complete amino acid sequence of the other high molecular weight tryptic fragment was determined [13]. This 45-residue fragment represents the *N*-terminus of the pepsin molecule [13]. The half-cystine residues localized in this region occupy positions Nos. 45 and 50 (fig. 1). The smallest one of the three tryptic peptides was found to include residues 46–50. In the later work the amino acid sequence up to methionine No. I (residue No. 80) was determined [8]; the phosphoserine residue occupies position No. 68.

The determination of the complete sequence of 43 residues between Met III and Met IV/14/ permitted Met III to be positioned at No. −82 (i.e. 82 from the *C*-terminus, see legend to fig. 1), and half-cystine residues Cys V and Cys VI at Nos. −78 and −45. The determined amino acid sequence was confirmed by a number of identical results obtained with the tryptic digest of aminoethylated pepsin [4]; the aspartyl residue No. −64 originally reported [14] as Asp was found to be an Asn [4].

Peptides from the region between Met II and Met III [8] afforded three as yet unlinked sequences of unambiguous order (fig. 1).

The region between Met I and Met II represents the largest specific cyanogen bromide fragment [3]. The tentative structure of this part of the molecule includes sequences around these two methionine residues as well as five sequences whose order is not yet final.

Rajagopalan et al. [2] have determined the total

number of amino acid residues in pepsin (excluding tryptophan) as being 315. The analysis of the tryptophan content in pepsin carried out in our laboratory [15] has revealed the presence of five residues of this amino acid. All the five tryptophan residues have been included in the tentative structure. Using the number of 315 residues by Rajagopalan et al. [2] we arrive at the total of 320. Our tentative structure accounts for 324 residues. Considering the possible overlapping of the unlinked sequences, the two numbers are in relatively good agreement. Our final conclusions in this respect will be made with the knowledge of the complete amino acid sequence.

References

[1] Atlas of Protein Sequence and Structure (Dayhoff, M.O., ed), National Biomedical Research Foundation, Silver Spring, Md.
[2] Rajagopalan, T.G., Moore, S. and Stein, W.H. (1966) J. Biol. Chem. 241, 4940.
[3] Morávek, L. and Kostka, V. (1973) Collection Czech. Commun. 38, 304.
[4] Kostka, V., unpublished results.
[5] Morávek, L. and Kysilka, Č., Collection Czech. Chem. Commun., in the press.
[6] Kostka, V. and Salvetová, A. unpublished results.
[7] Keil, B., Morávek, L. and Šorm, F. (1967) Collection Czech. Chem. Commun. 32, 1968.
[8] Morávek, L., unpublished results.
[9] Kostka, V., Morávek, L. and Šorm, F. (1970) European J. Biochem. 13, 447.
[10] Trufanov, V.A., Kostka, V., Keil, B. and Šorm, F. (1969) European J. Biochem. 7, 544.
[11] Morávek, L. (1972) FEBS Letters 23, 337.
[12] Verdier, C.H. (1954) Acta Chem. Scand. 8, 1302.
[13] Morávek, L., Submitted to the 'Atlas', September 1971.
[14] Morávek, L., submitted to the 'Atlas', May 1972.
[15] Neradová, V. and Kostka, V., unpublished results.
[16] Bayliss, R.S., Knowles, J.R. and Wybrandt, G.B. (1969) Biochem. J. 113, 377.