AMINO-ACID SEQUENCES AROUND METHIONINE AND TRYPTOPHAN RESIDUES OF HOG PEPSIN

L.Morávek and V.Kostka

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

Received February 12th, 1974

The amino-acid sequences around the four methionine residues and the five tryptophan residues of hog pepsin were determined in studies on selected peptides from enzymatic digests of the protein and with the use of data on terminal amino-acid sequences of the cyanogen bromide fragments. The sequences around the methionine residues determine the order of the cyanogen bromide fragments in the polypeptide chain of pepsin.

The relatively low methionine content of hog pepsin (4 residues, ref.[1,2]) led us to use in our sequential studies cyanogen bromide cleavage[2,3] as a fundamental fragmentation procedure proceeding specifically[4]. Since this procedure brings about peptide bond cleavage at the carboxyl side of methionine residues, it is necessary to know the amino-acid sequence around methionine in order to arrange the arising cyanogen bromide fragments in the polypeptide chain. Methionine sequences of hog pepsin were studied by Tang and Hartley[5] and some partial data were also reported by Vasenev and coworkers[6]. Some of our results obtained in the process of characterization of the cyanogen bromide fragments[2] of S-sulfo-pepsin and of the peptides from the thermolysin[7] and chymotryptic[8] digest of this protein were in disagreement with the data published[5,6]. For these reasons, methionine peptides selected from enzymatic digests were subjected to detailed examination in order that we might arrange the cyanogen bromide fragments on the basis of our own information only.

Tryptophan is another important amino acid of minor occurrence in pepsin. The problem of the number of tryptophan residues of hog pepsin has been discussed in our previous paper[9]. All the experimental data obtained by us indicate the presence of five tryptophan residues in the molecule of this enzyme. Similarly to methionine, the amino acid sequences around tryptophan residues in hog pepsin were the subject of special studies by other authors[10–12]. From sequential studies on selected tryptophan peptides, carried out in this Laboratory, we were able to determine the neighbourhood of all five tryptophan residues and thus to complement the data published so far.

EXPERIMENTAL

The subject of sequential investigation were peptides chosen from several enzymatic digests; these peptides characterize the neighborhood of all methionine residues (Fig. 1) and of tryptophan residues tentatively marked by Roman numerals II—IV (Fig. 2). The amino-acid sequence around tryptophan I (Fig. 3) is known from the N-terminal sequence of pepsin published earlier[13]; the neighborhood of tryptophan V is involved in the complete amino-acid sequence of cyanogen bromide fragment CBI (ref.[14], Fig. 3). The N- and C-terminal amino-acid sequences of the cyanogen bromide fragments[2], designated by the symbol "CB", as well as some already published cystine sequences[15], which fall into the regions around methionine, were used to characterize the neighborhood of the methionine residues.

Peptides for studies on the methionines were chosen on the basis of their content of methionine or homoserine, certain other peptides, permitting a broader neighborhood of this residue to be characterized, were also used. Tryptophan – containing peptides were sought for by a qualitative test, *i.e.* by dipping paper chromatograms in a 1% solution of *p*-dimethylaminobenzaldehyde in a mixture of acetone and concentrated hydrochloric acid (9 : 1, *v/v*); the tryptophan-containing peptides were stained violet[16].

The isolation and characterization of peptides marked by the symbol "Th", obtained from the thermolysin digest of S-sulfo-pepsin, as well as of peptides marked "C", prepared by chymotryptic digestion of the same substrate, have been described before[7,8]. Peptides marked "M" were obtained from the chymotryptic digest[17] of cyanogen bromide fragment CB2 (ref.[2]), peptides marked "V" from the thermolysin digest of the same fragment. Peptides marked "S" were chosen from the subtilisin digest of tryptic fragment RAEP-tA22 (ref.[18]) of aminoethylated pepsin; peptide RAEP-tA42-C353 was obtained by chymotryptic hydrolysis of another tryptic fragment, RAEP-tA42 from the same digest[18].

The amino-acid analysis of peptides was effected by the method of Spackman, Stein and Moore[19], as described in detail elsewhere[14]. The results of the analyses are summarized in Table I. The N-terminal amino acids were determined by the dansylation technique[20,21]. The stepwise degradation of peptides according to Edman[22] was carried out by the technique described before[14]. The C-terminal amino acid of chymotryptic peptides was derived in most cases from the known specificity of this enzyme (Trp, Tyr,Phe); in the remaining cases it was determined from experiments with carboxypeptidase A cleavage of the peptides *e.g.*[23]). The presence of amides was determined either directly in the course of stepwise degradation of the peptides, in some other cases from the net charge of the peptide on electrophoresis (pH 5·6, ref.[24]).

RESULTS AND DISCUSSION

In studies on the terminal amino-acid sequences of the cyanogen bromide fragments[2] we determined preliminarily their order as CB4-CB3-CB6-CB5-CB1. The results obtained, however, did not confirm in full any of the methionine sequences reported by other authors[5]. Therefore we undertook a detailed investigation of peptides characterizing the neighborhood of the methionine residues. The results of studies on the individual methionine sequences are summarized in Fig. 1.

Methionine sequence I. This sequence involves the C-terminal region of fragment CB4 and the N-terminus of fragment CB3. The bond ...Met-Thr... was cleaved only partly by cyanogen bromide; therefore in addition to fragments CB4 and CB3,

TABLE I

Amino-Acid Analyses of Peptides

The designation of the peptides is described in the experimental section. The analyses were made on 20-h hydrolysates. None of the peptides contained lysine, histidine, arginine, or cysteine.

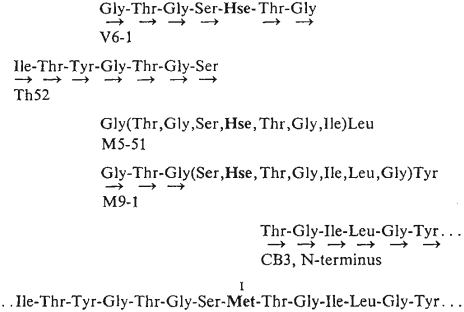
Designation of peptide	mol/mol of Peptide														
	Asp	Thr	Ser	Glu	Pro	Gly	Ala	Val	Met	Ile	Leu	Tyr	Phe[a]	Trp[a]	Aec[b]
C-II-5	2·1		1·3	1·7		1·1		1·1			1·9		0·8		
C-IX-4			0·8	1·0		1·2		1·2				0·9		+	
M 5-51		2·0	1·0			2·9			1·1[c]	1·0	1·0				
M 9-1		1·9	1·0			3·6			1·2[c]	1·0	1·0	0·4			
M 10-11	2·0		1·1	1·1			1·0				1·0			+	
M 12-2	1·0	1·0	1·0						1·2[c]	0·9					
RAEP-tA 42-C 353	1·1	0·9	1·8	1·8	0·9	1·8		1·0	0·8		1·0			+	
S 1042	1·1	1·1	1·1			1·1					1·1	1·7			
S 1262			1·1	1·0	1·1			2·9						+	
S 146	0·8		1·0	1·1	1·2			3·2			0·7			+	
S 15524		1·0		1·1		1·2				1·0		0·8		+	
S 1611	1·9	1·9	1·0	1·1		1·1	1·0		0·9	1·9	1·0				0·8
S 2032	3·0	1·0	0·9	1·1	0·9	1·1	0·8	1·1			1·3		0·8	+	
Th 9	2·3		2·0	2·0		1·1	0·8		0·8						
Th 30	1·2	1·9		1·1		1·1			0·8	1·0					
Th 34	2·1	1·0			1·2		0·8	1·0					0·9		
Th 52		2·0	1·1			2·0				1·0		0·9			
Th 83	1·1			1·0		1·1					1·0			+	
Th 84	1·0				1·1			1·0			0·9			+	
V 6-1		1·9	1·2			2·7			1·1[c]						
V 9-1		1·0	1·0			1·0						0·8			
V 14-7	1·2		1·0	1·0				1·1			1·8				
V 19-21				1·9		1·1		1·0				0·7		+	

[a] The tryptophan content was determined qualitatively only. [b] Aec stands for S-(β-aminoethyl)cysteine. [c] Determined as homoserine, formed by cyanogen bromide cleavage.

FIG. 1

Methionine Sequences of Pepsin

The methionine residues are numbered according to their order in the polypeptide chain of pepsin. The designation of the peptides is explained in the experimental section. $\rightharpoonup$ stands for the individual steps of Edman degradation, $\leftharpoondown$ symbolizes the order of C-terminal amino acids determined by cleavage of peptides with carboxypeptidase A. Hse stands for homoserine.



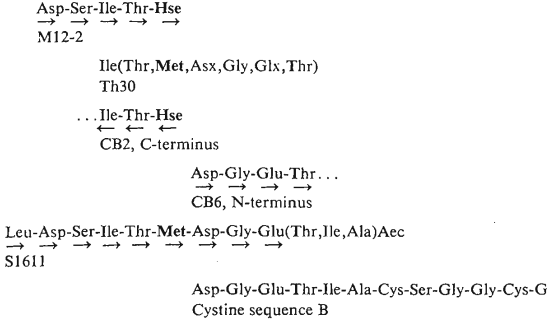


FIG. 1
(*Continued*)

Ala-Ser-Glu-Asn-Ser-Asp-Gly-Glu-**Met**
→ → → → → → → →
Th9

Val-Ile-Ser-Cys
→ → → →
CB6, N-terminus

Val-Ile-Ser-Cys-Ser
Cystine sequence C1

III
...Ala-Ser-Glu-Asn-Ser-Asp-Gly-Glu-**Met**-Val-Ile-Ser-Cys-Ser...

Methionine sequence III

Ile-Leu-Gln-Asp-Asp-Asp-Ser-Cys-Thr-Ser-Gly-Phe-Glu-Gly-**Met**
Cystine sequence C2

Asp-Val-Pro-Thr-Ser-Ser-Gly-Glu-Leu-Trp...
→ → → → → → → → → →
CB1, N-terminus

Glu-Gly-**Met**-Asp-Val-Pro-Thr(Ser,Ser,Gly,Glu,Leu)Trp
→ → → → → → →
RAEP-tA42-C353
IV

...Ile-Leu-Gln-Asp-Asp-Asp-Ser-Cys-Thr-Ser-Gly-Phe-Glu-Gly-**Met**-Asp-Val-Pro-Thr-Ser-Ser-Gly-Glu-Leu-Trp...

Methionine sequence IV

fragment CB2 was also isolated. The latter, in which fragments CB4 and CB3 are linked together, involves Met I and Met II. The N-terminal part of the sequence, *i.e.* ...Ile-Thr-Tyr-Gly-Thr-Gly-Ser-Met... is reconcilable with sequence I of Tang and Hartley[5]. However, the other part of their sequence, ...Asp-Val-Pro-Thr-Ser from our data represents the amino acid sequence at the carboxyl side of Met IV (Fig. 1); it is identical with the N-terminus of fragment CB1 which involves the C-terminal region of pepsin[2]. Neither did we confirm in our studies the sequence Tyr-Gly-Thr-Ser-Gly-Thr-Met, determined by Vasenev and coworkers[6].

Methionine sequence II. This sequence involves the identical C-terminal regions of fragments CB3 and CB2 and the N-terminal region of fragment CB6, comprising cystine sequence B (ref.[15]). The part of the sequence at the amino side of the methionine residue is identical with sequence III of Tang and Hartley[5], *i.e.* Asp-Ser-Ile--Thr-Met(Asp,Gly,Glu)(Ala,Tyr), whereas the rest of their sequence is reconcilable with our results only partly.

Methionine sequence III. The sequence around Met III involves the C-terminus of fragment CB6 (ref.[17]) and the N-terminus of fragment CB5, whose amino acid sequence has preliminarily been published[25,26]. We did not find any peptide in this study corresponding to the uninterrupted amino acid sequence around Met No III. Since, however, all the remaining methionine links and also all terminal amino-acid sequences of the cyanogen bromide fragments have been determined unambiguously, there is only one pair of unlinked terminal sequences left which permits methionine sequence III to be derived unambiguously. The sequence thus obtained is partly comparable with sequence II of Tang and Hartley[5]: Ala-Ser-Glu-Asx-Ser-Asx-Gly--Glu-Met-Ile(Val,Tyr).

Methionine sequence IV. This sequence overlaps the determined sequences of fragments CB5 (ref.[25,26]) and CB1 (ref.[14]). A part of our methionine sequence IV, ...Glu-Gly-Asp-Met..., is reconcilable with a part of sequence IV of Tang and Hartley[5], Glx-Gly-Asx-Ser(Thr,Ser,Val)Glu-Gly-Asp-Met-Asx(Thr,Ser,Ser,Glu,Gly,Val,Leu). However, the other part of our sequence, *i.e.* ...Asp-Val-Pro-Thr-Ser-... is identical with a part of sequence I of Tang and Hartley[5]. The revised version of the sequence, reported by Cheng and Tang[27], *i.e.* Phe-Glu-Gly-Met-Asp-Val-Pro-Thr-Ser- is in full agreement with our Methionine sequence IV.

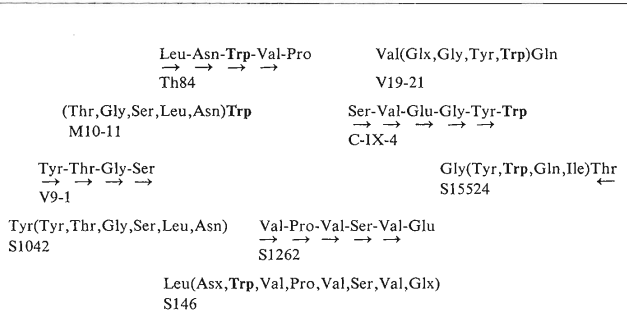
The results obtained in this study fully confirm the order of cyanogen bromide fragments of hog pepsin, CB4-CB3-CB6-CB5-CB1, proposed by us earlier[2].

The results obtained in our earlier study[9] show that hog pepsin contains five tryptophan residues. The amino acid sequences around tryptophan were studied by Dopheide and Jones[10]. These authors determined the neighbourhood of four tryptophan residues. Similar sequential studies, focused on tryptophan residues, were carried out by Vasenev and coworkers[11,12]. The N-terminal 55-residue amino acid sequence[13] of hog pepsin and the amino acid sequence of the 37-residue C-ter-

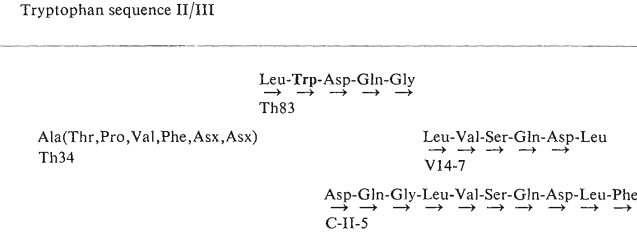
FIG. 2

Tryptophan Sequences of Pepsin

Tryptophan residues II through IV are numbered tentatively. The order of tryptophan sequences II/III and Trp IV in the region between Met I and Met II (ref.[36]) has not been determined as yet. The designation of the peptides is described in the experimental section. The meaning of symbols → and ← is given in the legend to Fig. 1.



Tryptophan sequence II/III



Tryptophan sequence IV

minal fragment CB1 (ref.[14]), both determined in this laboratory, include tryptophan residues I and V.

Tryptophan sequence I (Fig. 3). This amino acid sequence involves Trp I whose close neighborhood is characterized by sequence II of Dopheide and Jones[10].

A comparable sequence, Asp-Glu-Ser-Gly-Leu-Trp-Val-Pro-Ser-Val-Tyr(Ser,Tyr) of Vasenev and coworkers[12], was reported[28] later from the laboratory of these authors in a form which is identical with our data.

Tryptophan sequence II/III. Our studies on the region between Met II and Met III (Fig. 3, ref.[17,18]) have shown that this part of the pepsin molecule, identical with fragment CB3, contains 3 tryptophan residues. Tryptophan sequence II/III involves two of these three residues. Their designation is tentative because the position of this pair within fragment CB3 is interchangeable with tryptophan IV. Sequence IV and sequence III, reported by Dopheide and Jones[10], can be aligned with our tryptophan sequence II/III. From our knowledge, however, a tyrosine residue is present at the amino side of Trp III (Fig. 2). The sequence around Trp II reported by Vasenev and coworkers[12] is in agreement with our data; their sequence around Trp III (ref.[12]) also lacks the tyrosine residue.

FIG. 3

Amino-Acid Sequences around Residues of Minor Occurrence in Pepsin

The Figure summarizes sequences around the residues of lysine, histidine, arginine, methionine, tryptophan, and half-cystine. The completely determined amino-acid sequences of the terminal regions are numbered. Negative numbers were used to mark the residues in the C-terminal region, starting from the C-terminus.

H_2N-Ile-Gly-Asp-Glu-Pro-Leu-Glu-Asn-Tyr-Leu-Asp-Thr-Glu-Tyr-Phe-Gly-Thr-Ile-Gly-Ile-
(10, 20)
-Gly-Thr-Pro-Ala-Gln-Asp-Phe-Thr-Val-Ile-Phe-Asp-Thr-Gly-Ser-Ser-Asn-Leu-**Trp**-Val-
(I; 30, 40)
-Pro-Ser-Val-Tyr-**Cys**-Ser-Ser-Leu-Ala-**Cys**-Ser-Asp-**His**-Asn-Gln-. . .-Ile-Thr-Tyr-Gly-
(I, II; 50)
-Ser-**Met**-Thr-Gly-Ile-Leu-Gly-Tyr-. . .-Tyr-Tyr-Thr-Gly-Ser-Leu-Asn-**Trp**-Val-Pro-Val-
(II; 80)
-Ser-Val-Glu-Gly-Tyr-**Trp**-Gln-Ile-Thr-. . ./. . .-Gly-Ala-Thr-Pro-Val-Phe-Asp-Asn-Leu-
(III)
-**Trp**-Asp-Gln-Gly-Leu-Val-Ser-Gln-Asp-Leu-Phe-. . .-Leu-Asp-Ser-Ile-Thr-**Met**-Asp-Gly-
(IV; II)
-Glu-Thr-Ile-Ala-**Cys**-Ser-Gly-Gly-**Cys**-Gln-Ala-. . .-Ala-Ser-Glu-Asn-Ser-Asp-Gly-Glu-
(III, IV; −90)
-**Met**-Val-Ile-Ser-**Cys**-Ser-. . .-Ile-Leu-Gln-Asp-Asp-Asp-Ser-**Cys**-Thr-Ser-Gly-Phe-Glu-
(III, V; VI; −80, −50, −40)
-Gly-**Met**-Asp-Val-Pro-Thr-Ser-Ser-Gly-Glu-Leu-**Trp**-Ile-Leu-Gly-Asp-Val-Phe-Ile-**Arg**-
(IV, V; −20)
-Gln-Tyr-Tyr-Thr-Val-Phe-Asp-**Arg**-Ala-Asn-Asn-Lys-Val-Gly-Leu-Ala-Pro-Val-Ala-
(−10)
-CO_2H

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

Tryptophan sequence IV. This sequence is reconcilable with the data of other authors[10,12]; it defines, however, a larger region of 20 amino-acid residues.

Tryptophan sequence V (Fig. 3). Tryptophan residue V is contained in fragment CB1 (ref.[14]); its neighborhood was not characterized in the studies by other authors[10,12].

We thank Mrs J. Viková and Mrs E. Bulantová for skillful technical assistance. We are indebted to Mr K. Grüner and to Mrs A. Čermáková for the stepwise degradation of the peptides, and to Mr J. Zbrožek, Miss V. Himrová and Mrs E. Držková for the amino-acid analyses.

REFERENCES

1. Rajagopalan T. G., Moore S., Stein W. H.: J. Biol. Chem. *241*, 4940 (1966).
2. Morávek L., Kostka V.: This Journal *38*, 304 (1973).
3. Ostoslavskaja V. I., Pugačeva I. B., Vachitova E. A., Krivcov V. F., Muratova G. L., Levin E. D., Stepanov V. M.: Biochimija *33*, 331 (1966).
4. Gross, E., Witkop B.: J. Biol. Chem. *237*, 1856 (1962).
5. Tang J., Hartley B. S.: Biochem. J. *118*, 611 (1970).
6. Vasenev V. I., Kuznecov J. S., Stepanov V. M.: Chim. Prir. Sojedin. *5*, 609 (1969).
7. Morávek L., Kysilka Č.: This Journal, in press.
8. Kostka V., Salvetová A.: This Journal, in press.
9. Kostka V., Neradová V.: This Journal, in press.
10. Dopheide T. A. A., Jones W.: J. Biol. Chem. *243*, 3906 (1968).
11. Vasenev V. I., Kuznecov J. S., Stepanov V. M.: Biochimija *35*, 787 (1970).
12. Vasenev V. I., Kuznecov J. S., Stepanov V. M.: Biochimija *37*, 898 (1972).
13. Morávek L.: This Journal, in press.
14. Kostka V., Morávek L., Šorm F.: Eur. J. Biochem. *13*, 447 (1970).
15. Morávek L.: This Journal, in press.
16. Smith I.: Nature *171*, 43 (1953).
17. Morávek L.: Unpublished results.
18. Kostka V.: Unpublished results.
19. Spackman D. H., Stein W. H., Moore S.: Anal. Chem. *30*, 1190 (1958).
20. Gray W. R., Hartley B. S.: Biochem. J. *89*, 59 P (1963).
21. Novotný J., Franěk F.: Chem. listy *62*, 995 (1968).
22. Edman P.: Acta Chem. Scand. *10*, 761 (1956).
23. Kostka V., Morávek L., Kluh I., Keil B.: Biochim. Biophys. Acta *175*, 459 (1969).
24. Mikeš O.: This Journal *22*, 831 (1957).
25. Morávek L. in the book: *Atlas of Protein Sequence and Structure* (M. O. Dayhoff, Ed.), National Biomedical Research Foundation, Silver Spring, Md., submitted May 1972.
26. Morávek L., Kostka V.: FEBS Letters *35*, 276 (1973).
27. Chen K. C. S., Tang J.: J. Biol. Chem. *247*, 2566 (1972).
28. Stepanov V. M., Baratova L. A., Pugačeva I. B., Beljanova L. P., Revina L. P., Timochina E. A.: Biochem. Biophys. Res. Commun. *54*, 1164 (1973).

Translated by the author (V. K.).