CYSTINE SEQUENCES OF HOG PEPSIN

L. MORÁVEK

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

Received February 4th, 1974

The amino-acid sequences around the six half-cystine residues of pepsin were determined. These results complete the information on the cystine sequences and provide sequential overlaps useful for the arrangement of the tryptic fragments of aminoethylated pepsin.

The disulfide bonds were the subject of one of our earliest papers[1] on the structure of hog pepsin. It was found that two of the cystines form short loops, whereas the third disulfide bond links two half-cystine residues farther apart in the chain of pepsin. A more recent study on the same subject by Tang and Hartley[2] corroborated these results and furthermore revealed the amino-acid sequences of the cystine peptides. However, the N-terminal extensions of the two short loops N-terminated by half-cystine residues, were not discovered. Since the half-cystine residues could function as potential sites of tryptic cleavage after their conversion to S-(β-aminoethyl)-cysteine[3,4], the amino acid sequences in their vicinity provide information necessary for the arrangement of such tryptic fragments in the chain. In the preceding paper[5], the N-terminal sequence of pepsin H_2N-Ile-Gly-Asp-Glu-Pro-Leu-Glu-Asn-
-Tyr-Leu-Asp-Thr-Glu-Tyr-Phe-Gly-Thr-Ile-Gly-Ile-Gly-Thr-Pro-Ala-Gln-Asp-Phe-
-Thr-Val-Ile-Phe-Asp-Thr-Gly-Ser-Ser-Asn-Leu-Trp-Val-Pro-Ser-Val-Tyr-Cys-Ser-
-Ser-Leu-Ala-Cys-Ser-Asp-His-Asn-Gln.. was determined. It contains two half-cystine residues (positions 45 and 50), forming a disulfide loop, and because it is assigned to the N-terminus of pepsin, it will be named cystine sequence A. The cystine sequence A provides the missing N-terminal extension of one of the loops. The sequence, published by Pugacheva and coworkers[6], beginning at residue 45, also can be assigned to the region of this loop. However, the N-terminal link of half-cystine(45) is missing.

In this paper, the N-terminal extension of the other short disulfide loop is presented, as well as sequential information on the remaining third disulfide. To obtain this information, peptides from the thermolysin digest of S-sulfopepsin[7] or peptides from various enzymic digests of fraction CB5 obtained by cyanogen bromide cleavage of pepsin[8], were used.

EXPERIMENTAL

The disulfide peptides isolated in the previous paper[1] are marked by symbol "SS" (Fig. 1 and Fig. 2). Their N-terminal groups were determined by the dansylation technique[9] (as the 1-dimethylaminonaphthalene-5-sulfonyl derivatives), their C-terminal amino acids were also determined by dansylation of the amino acids liberated by hydrazinolysis[10]. The amino-acid composition of peptides presented in this paper is included in Table I. Peptides marked by the symbol "Th" were isolated from the thermolysin digest of S-sulfopepsin[7]. All other digests were prepared by cleavage of fraction CB5 of the cyanogen bromide hydrolysate of S-sulfopepsin, isolated in our previous study[8]. Characterization of fraction CB5 has revealed that it contains one major component, designated fragment CB5, and characterized by the N-terminal sequence Val-Ile-Ser... (extended in more recent experiments with Edman degradation[11] of oxidized material to Val--Ile-Ser-Cys... (ref.[12])). In addition to fragment CB5 it contains a smaller amount of another component, characterized by the N-terminal sequence Asp-Gly-Glu-Thr and designated fragment CB6 (ref.[8]). Peptides isolated from the chymotryptic digest of the aminoethylated[3,4] fraction CB5 are marked by symbol FCB5-RAE-C. Peptides from digests of the same material by trypsin or thermolysin are marked by symbols, ending in a —T, or Th, respectively. The short symbol "U" was used for peptides from the thermolysin digest of the original fraction CB5, in which the half-cystine residues were present in the S-sulfonated form. The two remaining digests, marked by FCB5-C(ox) and FCB5-RAE-T(b), were produced for the specific isolation of some half-cystine-containing peptides, based on the charges, due to the presence of the residues of cysteic acid, or S-(β-aminoethyl)cysteine. The chymotryptic digest of fraction FCB5 was oxidized by the performic acid oxidation mixture (1 volume of 30% hydrogen peroxide and 9 volumes of 99% formic acid) and peptides, bearing a negative charge on paper electrophoresis (pH 1·9), were isolated. These peptides were marked by the symbol FCB5-C(ox). Peptides from the tryptic digest of the aminoethylated fraction CB5, which were positively charged on paper electrophoresis at pH 5·6, were marked by the symbol FCB5-RAE-T(b). The preparation of all digests mentioned above, as well as their fractionation will be described in papers on the sequence of fragment CB5 (ref.[12]). or on fragment CB6 (ref.[13]). Methods for the purification and characterization of the peptides have been described in our previous papers[8,14]. Homoserine was assigned as C-terminal in peptides containing this amino acid. The residue of S-(β-aminoethyl)cysteine was assigned to the C-terminus when present in peptides resulting from cleavage by trypsin. The N-terminal amino acids were determined by the dansylation technique[9] using the procedure by Novotný and Franěk[15]. The N-terminal sequences were determined by Edman degradation[11] as described in a previous paper[14].

RESULTS AND DISCUSSION

In the preceding paper[5] the cystine sequence A, assigned to the N-terminal region of pepsin was determined. The remaining four half-cystine residues are present in the cystine sequences B and C.

Cystine sequence B. The alignment of the peptides in sequence B is presented in Fig. 1. This sequence contains the two half-cystine residues of the disulfide peptide SS I (ref.[1]), which is comparable with the peptide Cys-Ser-Gly-Gly-Cys-Gln (ref.[2]). Our peptide Thl provides the N-terminal and C-terminal extension of this peptide, and allows for the other peptides in Fig. 1 to be assigned to the region preceding the

third half-cystine residue of pepsin. The N-terminal sequence Asp-Gly-Glu-Thr is identical with the sequence ascribed to the N-terminus of our cyanogen bromide fragment CB6 (ref.[8]) and with the N-terminus of fragment B-5, described by Ostoslavskaja and coworkers.[16] Results of our study on the cyanogen bromide fragments[8] and of our study on the methionine sequences of pepsin[17] permit us to assign the cystine sequence B to the N-terminus of fragment CB6, occupying the region between methionine residues II and III. The preliminary results of the study of fragment CB6 (ref.[13]) have revealed that the cystine sequence B is followed in the chain by the sequence Ile-Val-Asp-Thr-Gly..., which involves the aspartic acid residue of the active site of pepsin[18].

Cystine sequence C. This sequence containing two half-cystine residues farther apart in the chain, can be subdivided to sequence C1 (containing the fifth half-cystine) and sequence C2 (containing the sixth half-cystine residue of pepsin). As found in a study on the sequence of fragment CB5 (ref.[12]), both half-cystines are present in this fragment, between the methionine residues III and IV (ref.[8,17]).

(Cys,Ser,Gly,Gly,Cys,Gln)
SS I

Ile-Ala-Cys-Ser-Gly-Gly-Cys-Gln-Ala
Th1

Asp-Gly-Glu-Thr-Ile-Ala-Aec
FCB5-RAE-C20-22

(Ser,Gly,Gly)Aec
FCB5-RAE-T4-41

(Asx,Gly,Glx,Thr,Ile,Ala,Aec,Ser,Gly,Gly)Aec
FCB5-RAE-T(b)31

(Asx,Gly,Glx,Thr,Ile,Ala,Cys,Ser,Gly,Gly,Cys,Glx)
FCB5-C(ox)11

Asp-Gly-Glu-Thr...
fragment CB6

Asp-Gly-Glu-Thr-Ile-Ala-Cys-Ser-Gly-Gly-Cys-Gln-Ala

FIG. 1

Cystine Sequence B

The nomenclature of peptides is given in the experimental section. Aec stands for S-(β-aminoethyl)cysteine.

Sequence C1. This sequence contains peptide SS II-a, described in our previous study[1]. It can be unambiguously included in the N-terminal sequence obtained by Edman degradation of oxidized fragment CB5 (ref.[12], *cf.* Fig. 2). As will be discussed in a paper on the sequence of fragment CB5 (ref.[12]), the remaining part of the disulfide sequence Cys-Ser-Ser-Ile-Asp-Gln of Tang and Hartley[2] was not corroborated in full. A sequence including the same half-cystine residue was published by Katrukha and Stepanov[19].

Sequence C2. This sequence represents the C-terminal part of fragment CB5 (ref.[12]), including the methionine residue IV (ref.[17]). The alignment of peptides in sequence C2 is presented in Fig. 2. The sequence of the half-cystine peptide U4-11 was obtained by Edman degradation. However, the product obtained in the 8th degradation step was not identified. When peptide FCB5-RAE-Th14-34, containing S-(β-aminoethyl)cysteine, was subjected to tryptic cleavage, two fragments were liberated. Because of the specificity of trypsin, the S-(β-aminoethyl)cysteine residue present in peptide FCB5-RAE-Th14-34-T6 should be assigned to the C-terminal position. The amino-acid composition of this peptide when compared with the N-terminal part

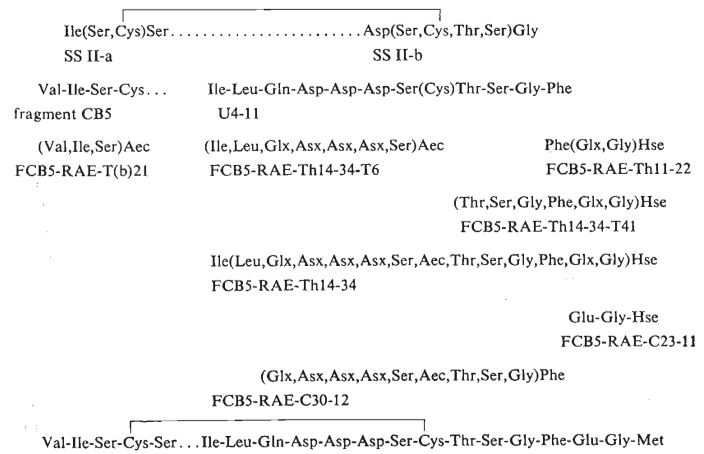

FIG. 2
Cystine Sequence C

TABLE I

Amino-Acid Composition of Peptides

The peptides were analyzed after 20 h hydrolysis; none of the samples contained lysine, histidine, arginine, proline, methionine, tyrosine or tryptophan.

	Cys	Asp	Thr	Ser	Glu	Gly	Gly	Ala	Hse[b]	Ile	Leu	Phe
Th1	1·8[a]	—	—	1·1	1·1	2·0	1·9	—	—	1·0	—	—
FCB5-RAE-C20-22	0·8	1·0	0·9	—	1·0	1·0	1·0	—	—	1·0	—	—
FCB5-RAE-C23-11	—	—	—	—	0·9	0·8	—	—	1·0	—	—	—
FCB5-RAE-C30-12	0·6	3·1	1·0	1·9	1·2	1·2	—	—	—	—	—	0·8
FCB5-RAE-T4-41	0·8	—	—	1·0	—	2·1	—	—	—	—	—	—
FCB5-RAE-Th11-22	—	—	—	—	1·0	1·0	—	—	1·1	—	—	0·9
FCB5-RAE-Th14-34	0·5	3·1	1·0	1·9	2·0	1·9	—	—	1·1	1·0	1·0	1·0
FCB5-RAE-Th14-34-T41	—	—	0·8	1·0	1·0	2·0	—	—	1·0	—	—	0·8
FCB5-RAE-Th14-34-T6	0·8	3·0	—	1·2	1·1	—	—	—	—	0·9	1·0	—
FCB5-C(ox)11	2·0[a]	1·1	1·0	1·1	2·0	3·0	1·1	—	—	1·0	—	—
FCB5-RAE-T(b)21	0·9	—	—	1·0	—	—	—	0·7	—	0·7	—	—
FCB5-RAE-T(b)31	1·2	0·9	0·9	1·1	1·1	2·9	0·8	—	—	1·0	—	—
U4-11	0·8[a]	3·0	1·0	1·9	1·2	1·2	—	—	—	0·9	0·9	0·9

[a] Cys determined as cysteic acid; in all other peptides Cys was determined as S-(β-aminoethyl) cysteine. [b] Symbol Hse stands for homoserine.

of peptide U4-11 places the half-cystine in position 8 of that peptide. The amino-acid composition of peptide FCB5-RAE-Th14–34-T41, also liberated by tryptic cleavage, completed the total amino-acid composition of peptide FCB5-RAE-Th14–34. These results only partially corroborated the sequence of the cystine peptide S 1b of Tang and Hartley[2]. Our sequence C2 gives a greater degree of homology when compared with that of chymosin (rennin)[20]:

pepsin Ile-Leu-GLN-ASP-Asp-Asp-Ser-CYS-THR-SER-GLY-PHE-Glu-Gly-Met
chymosin Thr-Ser-GLN-ASP-Gln-Gly-Phe-CYS-THR-SER-GLY-PHE
than the other sequence Glu-Asx-Asx-Ser-Cys-Thr-Ser-Asp-Ser-Asp-Ser (ref.[2]).

The here presented cystine sequences have been submitted to the Atlas of Protein Sequence and Structure in 1972 (ref.[21]) and are included in our tentative sequence of pepsin[22]. The recently published N-terminal sequence of pepsin by Stepanov and coworkers[23] is in complete agreement with our cystine sequence A.

I wish to thank Mrs J. Viková for careful technical assistance. I am indebted to Mr K. Grüner and Mrs A. Čermáková, who carried out skillfully the Edman degradation experiments, as well as to Mr J. Zbrožek, Miss V. Himrová and Mrs E. Držková for the amino acid analyses.

REFERENCES

1. Keil B., Morávek L., Šorm F.: This Journal *32*, 1968 (1967).
2. Tang J., Hartley B. S.: Biochem. J. *118*, 611 (1970).
3. Lindley H.: Nature *178*, 647 (1956).
4. Raftery M. A., Cole R. D.: Biochem. Biophys. Res. Commun. *10*, 467 (1963).
5. Morávek L.: This Journal, in press.
6. Pugačeva I. B., Stepanov V. M., Amirchanjan M. M.: Biochimija *36*, 993 (1971).
7. Morávek L., Kysilka Č.: This Journal, in press.
8. Morávek L., Kostka V.: This Journal *38*, 304 (1973).
9. Gray W. R., Hartley B. S.: Biochem. J. *89*, 59P (1963).
10. Akabori S., Ohno K., Narita K.: Bull. Chem. Soc. Japan *25*, 214 (1952).
11. Edman P.: Acta Chem. Scand. *10*, 761 (1956).
12. Morávek L.: Unpublished results.
13. Morávek L.: Unpublished results.
14. Kostka V., Morávek L., Šorm F.: European J. Biochem. *13*, 447 (1970).
15. Novotný J., Franěk F.: Chem. listy *62*, 995 (1968).
16. Ostoslavskaja V. I., Pugačeva I. B., Vachitova E. A., Krivcov V. F., Muratova G. L., Levin E. D., Stepanov V. M.: Biochimija *33*, 331 (1968).
17. Morávek L., Kostka V.: This Journal, in press.
18. Bayliss R. S., Knowles J. R., Wybrandt G. B.: Biochem. J. *113*, 377 (1969).
19. Katrucha S. P., Stepanov V. M.: Biochimija *37*, 329 (1972).
20. Foltmann B., Hartley B. S.: Biochem. J. *104*, 1064 (1967).
21. Morávek L. in the book: *Atlas of Protein Sequence and Structure* (M. O. Dayhoff, Ed.), National Biochemical Research Foundation, Silver Spring, Md., submitted May 1972.
22. Morávek L., Kostka V.: FEBS Letters *35*, 276 (1973).
23. 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.