N-TERMINAL AMINO ACID SEQUENCE OF 55 RESIDUES OF HOG PEPSIN

L. MORÁVEK

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

Received January 16th, 1974

Peptides from the chymotrypsin and thermolysin digests of the N-terminal tryptic fragment of pepsin were isolated and characterized. The information from these digests, combined with other sequence data, provided the N-terminal sequence of 55 residues of hog pepsin.

The cyanogen bromide cleavage of hog pepsin[1,2] gives rise to a large fragment which involves the N-terminal half of the chain. In a previous paper[3] the N-terminal cyanogen bromide fragment was aminoethylated and cleaved by trypsin. The resulting tryptic fragments were isolated and their sequence in the chain determined. In this study peptides from the chymotrypsin and thermolysin digests of the N-terminal tryptic fragment CB2-RAE-T12 (ref.[3]) were isolated and characterized. The information from these digests, combined with selected peptides from the chymotrypsin and thermolysin digests of the parent cyanogen bromide fragment CB2 (ref.[1]), gave the N-terminal sequence of hog pepsin comprising 55 amino-acid residues.

EXPERIMENTAL

The N-terminal tryptic fragment CB2-RAE-T12 (ref.[3]) and the N-terminal cyanogen bromide fragment CB2 (ref.[1]) were isolated in earlier studies. The chymotryptic digest of fragment CB2-RAE-T12 was prepared from 100 mg of the peptide. A 1% solution of this material was made alkaline (phenol red) with ammonium carbonate and digested with chymotrypsin (prepared by trypsin activation of a five-times crystallized preparation of chymotrypsinogen, Léčiva, Prague, Czechoslovakia) at a weight ratio 1 : 50, 4h, at 37°C. The digest was evaporated, dissolved in 2 ml of distilled water and fractionated by discontinuous recycling chromatography[4] on a standard 150 . 1·8 cm column of Sephadex G-15 in dilute ammonium hydroxide (pH 9). Descending elution was used. All fractions (6 ml/20 min) were assayed by paper chromatography of aliquots (0·1 ml). The part of the effluent containing the peptides was manually reapplied to the top of the column in the described manner[4]. After three cycles an effective bed height of 450 cm was achieved. Using this relatively laborious procedure, based on gel filtration, a better recovery of peptides was obtained than that, observed in experiments with other digests, using fractionation on Dowex 1-X2 columns. The cleavage of 100 mg of fragment CB2-RAE-T12 with thermolysin (B grade, Calbiochem, U.S.A.) was done under the same conditions as used with chymotrypsin. The pH was adjusted with 0·1M ammonium carbonate, containing 0·001M calcium chloride. The fraction-

ation of the thermolysin digest was performed by a single gel filtration step on Sephadex G-15, under the conditions described above. The digestion of two 900 mg-portions of the cyanogen bromide fragment CB2 with chymotrypsin and thermolysin was done in the same manner, as described with fragment CB2-RAE-T12. The digests were fractionated on a 60 . 1·8 cm column of Dowex 1-X2, using the procedure, described by Guest and coworkers[5]. Fractions were evaluated by paper chromatography of aliquots (0·2 ml). The purification of peptides from all digests was completed by paper chromatography and electrophoresis. These techniques and also methods of characterization have been described elsewhere[1]. The amino-acid composition of the peptides is presented in Table I. All peptides from the digests of fragment CB2-RAE-T12 were included, whereas only several peptides were selected from the digests of the larger fragment CB2. All peptides from the latter digests will be included in another paper[6]. Tryptophan, tyrosine, or phenylalanine was assigned as C-terminal amino acid in chymotryptic peptides, in accordance with the specificity of chymotrypsin. In the absence of these amino acids leucine was assigned as the C-terminal amino acid. The N-terminal amino acids were determined by the dansylation technique (as the 1-dimethylamino-naphthalene-5-sulfonyl derivatives[7]), using the procedure described by Novotný and Franěk[8]. All chemicals used were of G.R. purity grade. The N-terminal sequences were determined by Edman degradation[9] as described in a previous paper[10]. Sequences of peptides thus obtained are presented in Fig. 1.

RESULTS AND DISCUSSION

In a preceding study[3], three fragments (CB2-RAE-Tl2, CB2-RAE-T2121 and CB2-RAE-Tl1b) were obtained by tryptic cleavage of the N-terminal cyanogen bromide fragment CB2 (ref.[1]). The fragments were partly characterized and their sequence in the chain was determined, as presented in the second line of Fig. 1. The information, obtained from the overlapping peptides, isolated from the chymotrypsin and thermolysin digests of the tryptic fragment CB2-RAE-Tl2 and of the parent cyanogen bromide fragment CB2 has made it possible to arrange these peptides in two sequence regions. One of them, contained in peptide M17-41, can be assigned to the N-terminus of fragment CB2-RAE-T12, whereas the other one, containing the S-(β-aminoethyl)cysteine residue (Aec), represents the remaining C-terminal part of the fragment. There is just one site in the chain of the fragment CB2-RAE-T12, between residues 14(Tyr) and 15(Phe), without an overlap. However, the existence of a phenylalanine residue at the N-terminus of the chymotryptic peptide M12-4 would be explained only, if the amino acid preceding this phenylalanine participated in a bond, very susceptible to chymotryptic cleavage. Since the only tryptophan residue and all phenylalanine residues of this fragment (Table II) have been assigned to positions inside these two sequence regions, as have all the tyrosines with the exception of residue 14, the only residue providing this susceptible bond must be Tyr(14). Supporting evidence for this conclusion is obtained by an examination of the amino acid analysis of peptide M17-54. This peptide, which is closely related to the N-terminal peptide M17-41, contains a residue of phenylalanine. The values for serine, glycine and valine indicate that the peptide had not been completely purified and

for this reason it was not included in Fig. 1. However, the presence of 0·9 residues of phenylalanine is in agreement with the conclusion and indicates, that in this case chymotryptic cleavage most likely occurred behind phenylalanine(15).

The number of the individual amino-acid residues in the sequence of fragment CB2-RAE-T12 (Fig. 1) is in a relatively good agreement with the amino-acid composi-

TABLE I

Amino Acid Composition of Peptides

Peptides from the chymotrypsin digest of CB2-RAE-T12 are symbolized by the letter "C", from the thermolysin digest by the letters "TH". Peptides derived from the fragment CB2 are marked by the letter "M" for the chymotrypsin peptides and by the letter "V" for the thermolysin peptides. The peptides were analyzed after 20-h hydrolysis; none of the samples contained arginine, lysine, or methionine.

Peptide	His	Cys[a]	Asp	Thr	Ser	Glu	Pro	Gly	Ala	Val	Hse[b]	Ile	Leu	Tyr	PheTrp[c]
C 112	—	—	1·2	2·1	—	1·0	1·2	3·0	1·0	—	—	1·7	—	—	1·0 —
C 1221	—	—	1·9	1·0	1·9	—	—	1·1	—	—	—	—	0·9	—	— —
C 1242	—	—	1·1	1·0	—	1·0	—	—	—	—	—	—	1·0	0·7	— —
C 24	—	—	2·0	1·2	2·0	—	—	1·1	—	—	—	—	1·0	—	— +
C 441	—	—	—	—	1·0	—	1·2	—	—	2·0	—	—	—	0·9	— —
C 451	—	—	—	1·0	—	—	—	—	—	1·0	—	0·8	—	—	1·1 —
C 51	—	1·0	—	—	—	—	—	—	—	—	—	—	—	—	— —
TH 1-11	—	—	1·0	1·0	—	1·0	1·0	1·0	1·0	—	—	0·9	—	—	— —
TH 1-21	—	—	1·2	1·1	—	1·1	1·0	2·0	1·0	—	—	1·8	—	—	— —
TH 2-1	—	—	2·1	1·0	1·8	—	—	1·0	—	0·9[d]	—	0·9[d]	—	—	0·9 —
TH 3-32	—	—	1·1	1·2	—	1·0	—	—	—	—	—	—	1·0	1·6	— —
TH 6-2	—	—	—	—	—	—	—	1·1	—	—	—	0·9	—	—	— —
TH 8-1	—	—	—	1·0	—	—	—	1·1	—	—	—	—	—	—	0·7 —
TH 9-1	—	1·0	—	—	—	—	—	—	—	1·0	—	—	—	—	1·1 —
TH 9-3	—	—	—	1·1	—	—	—	—	—	—	—	—	—	—	0·8 —
TH 10-31	—	—	—	—	0·8	—	1·1	—	—	1·0	—	—	1·0	—	— +
TH 10-32	—	—	—	—	—	—	—	—	—	1·0	—	—	—	0·8	— —
M 12-4	—	—	1·1	2·0	—	1·1	1·1	3·0	1·0	—	—	1·9	—	—	1·7 —
M 16-12	—	—	1·9	0·9	1·9	—	—	1·1	—	—	—	—	1·0	—	— +
M 17-41	—	—	3·0	1·0	—	2·7	1·2	1·2	—	—	—	0·8	1·9	1·4	— —
M 17-54	—	—	2·8	1·1	1·0	2·4	0·9	1·5	—	0·3	—	1·1	1·9	1·6	0·9 —
V 17-11	0·9	0·8	2·2	1·1	—	1·0	—	—	1·0	—	—	0·9	—	—	— —
V 19-11	—	—	2·3	—	—	1·9	0·9	1·1	—	—	—	1·0	1·2	0·7	— —

[a] The half-cysteine residues present in the C- and TH- category peptides was determined as the S-(β-aminoethyl)cysteine derivative. Those present in the M and V categories were determined as cysteic acid. [b] Hse stands for the homoserine. [c] Tryptophan was determined only qualitatively. [d] Values for valine and isoleucine were obtained after 70 h hydrolysis.

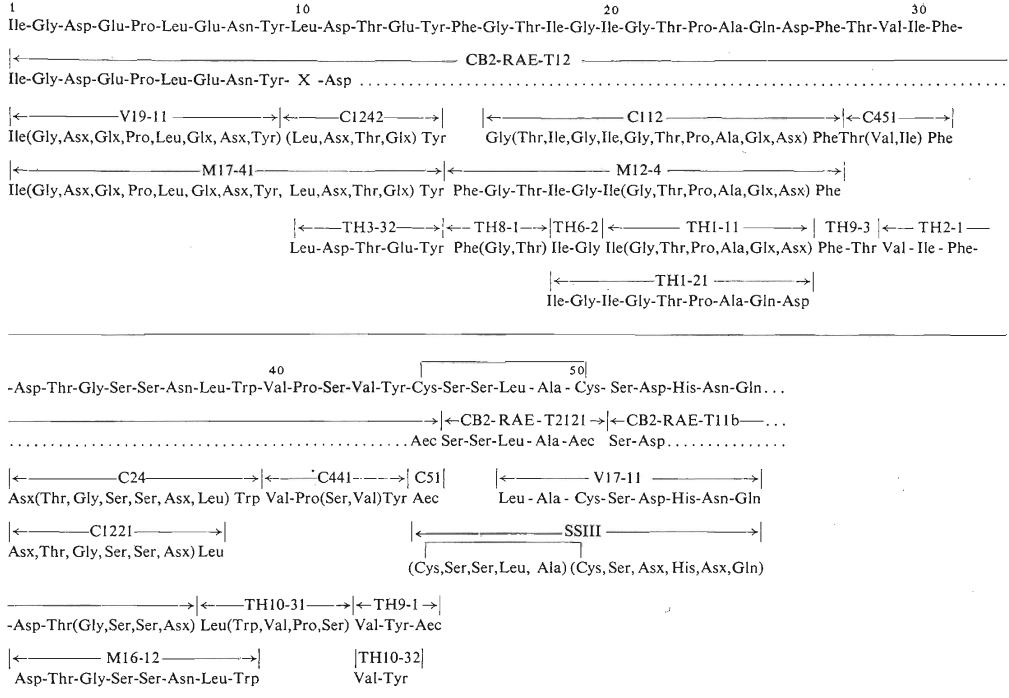
1
10
20
30
Ile-Gly-Asp-Glu-Pro-Leu-Glu-Asn-Tyr-Leu-Asp-Thr-Glu-Tyr-Phe-Gly-Thr-Ile-Gly-Ile-Gly-Thr-Pro-Ala-Gln-Asp-Phe-Thr-Val-Ile-Phe-
CB2-RAE-T12
Ile-Gly-Asp-Glu-Pro-Leu-Glu-Asn-Tyr- X -Asp
V19-11
C1242
Ile(Gly,Asx,Glx,Pro,Leu,Glx, Asx,Tyr) (Leu,Asx,Thr,Glx) Tyr
C112
C451
Gly(Thr,Ile,Gly,Ile,Gly,Thr,Pro,Ala,Glx,Asx) PheThr(Val,Ile) Phe
M17-41
M12-4
Ile(Gly,Asx,Glx, Pro,Leu, Glx,Asx,Tyr, Leu,Asx,Thr,Glx) Tyr Phe-Gly-Thr-Ile-Gly-Ile(Gly,Thr,Pro,Ala,Glx,Asx) Phe
TH3-32
TH8-1
TH6-2
TH1-11
TH9-3
TH2-1
Leu-Asp-Thr-Glu-Tyr Phe(Gly,Thr) Ile-Gly Ile(Gly,Thr,Pro,Ala,Glx,Asx) Phe-Thr Val - Ile - Phe-
TH1-21
Ile-Gly-Ile-Gly-Thr-Pro-Ala-Gln-Asp
40
50
-Asp-Thr-Gly-Ser-Ser-Asn-Leu-Trp-Val-Pro-Ser-Val-Tyr-Cys-Ser-Ser-Leu - Ala - Cys- Ser-Asp-His-Asn-Gln . . .
CB2- RAE - T2121
CB2-RAE-T11b
. . . . . Aec Ser-Ser-Leu - Ala-Aec Ser-Asp . . . . . .
C24
C441
C51
V17-11
Asx(Thr, Gly, Ser, Ser, Asx, Leu) Trp Val-Pro(Ser, Val)Tyr Aec
Leu - Ala - Cys-Ser-Asp-His-Asn-Gln
C1221
SSIII
Asx,Thr, Gly, Ser, Ser, Asx) Leu
(Cys,Ser,Ser,Leu, Ala) (Cys, Ser, Asx, His,Asx,Gln)
TH10-31
TH9-1
-Asp-Thr(Gly,Ser,Ser,Asx) Leu(Trp,Val,Pro,Ser) Val-Tyr-Aec
M16-12
TH10-32
Asp-Thr-Gly-Ser-Ser-Asn-Leu-Trp
Val-Tyr

FIG. 1

Alignment of Peptides in the N-Terminal Sequence of Pepsin

The nomenclature of peptides is that as described in Table I. The disulfide peptide SS III was published in a previous paper[12]. Aec stands for S-(β-aminoethyl)cysteine. Symbol X in the N-terminal sequence of fragment CB2-RAE-T12 signifies an ambiguous result in the 10th step of the Edman degradation.

TABLE II

Amino Acid Composition of Fragment CB2-RAE-T12

No arginine, lysine, histidine or methionine was found in the sample; Hse stands for homoserine, Aec for S-(β-aminoethyl)cysteine.

Amino Acid	Hydrolysis time		Found in
	20 h	70 h	peptides
Asp	6·2	6·1	6
Thr	5·0	4·2	5
Ser	4·5	3·4	3
Glu	3·4	3·7	4
Pro	2·8	2·9	3
Gly	5·2	4·9	5
Ala	1·2	1·3	1
Val	3·0	3·2	3
Ile	3·3	3·4	4
Leu	3·7	3·5	3
Tyr	3·0	2·7	3
Phe	2·6	2·5	3
Hse	0·1	0·1	—
Aec	0·5	0·4	1
Trp[a]	—	—	1
		Total	*45*

[a] Tryptophan was not determined quantitatively. However, there was only one tryptophan-containing peptide in the chymotryptic map of fragment CB2-RAE-T12.

tion of this fragment (Table II). The lower value for aminoethylcysteine is due to the fact, that during the tryptic hydrolysis of the aminoethylated fragment CB2 (CB2-RAE) a part of the material was cleaved behind Tyr(44), as indicated by the existence of the thermolysin peptide TH10-32 Val-Tyr(43-44). In studies with carboxypeptidase A approximately 50% of the material was found to be shorter by the aminoethylcysteine residue[11]. The higher values for serine and leucine are to be ascribed to

residual impurities in the starting material because the sequence data do not allow for the incorporation of these residues into the fragment. The information from peptides CB2-RAE-T2121 and CB2-RAE-T11b, obtained in a previous paper[3] and from the histidine-containing peptide V17-11 extends the sequence of fragment CB2-RAE-T12 to 55 residues from the N-terminus of pepsin (Fig. 1). The disulfide peptide SS III (ref.[12]) indicates that the two half-cystine residues present in this sequence form a disulfide bond in the native enzyme, as published also by other authors[13].

The here presented N-terminal sequence has been submitted to the Atlas of Protein Sequence and Structure in 1971 (ref.[14]). The recently published N-terminal sequence by Stepanov and coworkers[15] is in complete agreement with results from this laboratory. The N-terminal sequence including 9 residues by Tang[16] and the sequence Asp-Thr-Gly-Ser-Ser-Asn (Fig. 1, res. 32–37) by Chen and Tang[17] is in good agreement with our results. However, its N-terminal extension by a tripeptide with one overlapping residue of aspartic acid in the sequence Ile-Val-Asp-Thr-Gly-Ser-Ser-Asn (ref.[17]) is not reconcilable with our results. With respect to the present stage of knowledge of pepsin in our laboratory[18], the possibility of repetition of the above mentioned hexapeptide sequence in a different site of the chain is almost excluded. Also the two sequences, published by Pugacheva and coworkers[19] can be assigned to positions in the N-terminal sequence, one of them to res. 16–27 and the other beginning at res. 45. The tryptophan sequence Leu-Trp-Val-Pro-Ser, published by Dopheide and Jones[20], can be assigned to res. 38–42 of our sequence.

As already stated by other authors[15,21], a high degree of sequential homology between N-terminal sequences of pepsin and chymosin (rennin) can be observed. This homology seems to be extended almost to the whole chain of the two proteins[18,22].

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. Morávek L., Kostka V.: This Journal *38*, 304 (1973).
2. 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).
3. Morávek L.: FEBS Letters *23*, 337 (1972).
4. Morávek L.: J. Chromatog. *59*, 343 (1971).
5. Guest J. R., Carlton B. C., Yanofsky C.: J. Biol. Chem. *242*, 5397 (1967).
6. Morávek L.: Unpublished results.
7. Gray W. R., Hartley B. S.: Biochem. J. *89*, 59 P (1963).
8. Novotný J., Franěk F.: Chem. listy *62*, 995 (1968).
9. Edman P.: Acta Chem. Scand. *10*, 761 (1956).

10. Kostka V., Morávek L., Šorm F.: European J. Biochem. *13*, 447 (1970).
11. Kostka V.: Unpublished results.
12. Keil B., Morávek L., Šorm F.: This Journal *32*, 1968 (1967).
13. Tang J., Hartley B. S.: Biochem. J. *118*, 611 (1970).
14. Morávek L. in the book: *Atlas of Protein Sequence and Structure* (M. O. Dayhoff, Ed.), National Biochemical Research Foundation, Silver Spring, Md., submitted September 1971.
15. 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).
16. Tang J.: Biochem. Biophys. Res. Commun. *41*, 697 (1970).
17. Chen K. C. S., Tang J.: J. Biol. Chem. *247*, 2566 (1972).
18. Morávek L., Kostka V.: FEBS Letters *35*, 276 (1973).
19. Pugačeva I. B., Stepanov V. M., Amirchanjan M. M.: Biochimija *36*, 993 (1971).
20. Dopheide T. A. A., Jones W. M.: J. Biol. Chem. *243*, 3906 (1968).
21. Pedersen V. B., Foltmann B.: FEBS Letters *35*, 250 (1973).
22. Foltmann B., Kauffman D., Parl M., Andersen P. M.: Neth. Milk Dairy J. *27*, 288 (1973).

Translated by the author.