THE AMINO ACID SEQUENCE OF THE FIRST 61 RESIDUES OF CHYMOSIN (RENNIN EC 3.4.4.3)

Vibeke Barkholt PEDERSEN and Bent FOLTMANN
Institute of Biochemical Genetics, University of Copenhagen, Ø. Farimagsgade 2A, DK-1353 Copenhagen K, Denmark

Received 3 July 1973

1. Introduction

In the first two communications [1, 2] on the primary structure of chymosin (rennin EC 3.4.4.3) analyses of a series of low molecular weight tryptic and chymotryptic peptides have been reported, among these the *N*-terminal sequence, the S–S bridges and the *C*-terminal sequence. It was shown that of the five arginine groups two were located near an S–S loop and two were found within the last 20 residues of the peptide chain [2].

If the peptide chain of chymosin is cleaved only after the arginyl residues the uneven distribution of these and the resulting difference in size of fragments should facilitate fractionation of the digest by gel filtration. Such experiments have now been carried out, and this paper describes the primary structure up to the first arginine (no 57). This residue has previously [2] been found to be followed by a tetrapeptide which extends the sequence to arginine no 61.

2. Materials and methods

2.1. *Enzymes*

Chymosin was prepared and purified by chromatography as previously described [3]. The main chromatographic fraction (chymosin B) was maleylated according to Butler et al. [4].

The following enzymes were used in the experiments described (all ratios are expressed in moles per mole): Trypsin, tocylphenylalanylchloromethylketone (TPCK)-treated (Worthington, Freehold, N.J., USA). a) 1/150 for 15 min at 12°C, pH maintained at 7.8 by addition of 0.05 M NaOH [5], the reaction was terminated by addition of Trasylol® (pancreatic trypsin inhibitor, Bayer, Germany); b) 1/100 for 3 hr at 24°C otherwise as a). Chymotrypsin (Novo, Copenhagen, Denmark) 1/100 in *N*-ethyl-morpholine acetate (NEMAc) 0.05 M pH 8.0 for 20 hr at 22°C. Elastase (Whatman Biochemicals, Maidstone, UK) 1/150 in 0.05 M NEMAc pH 8.0 for 4 hr at 37°C. Thermolysin (Daiwa Kasei, Osaka, Japan) 1/100 in 0.05 M NEMAc pH 8.0 containing 0.01 M $CaCl_2$ for 16 hr at 37°C. Papain (Sigma, St. Louis, Mo., USA) 1/75 in 0.2 M pyridine acetate pH 6.5 containing 0.5% β-mercaptoethanol for 20 hr at 22°C.

2.2. *Purification and analysis of peptides*

Peptides from the tryptic digest containing more than 50 residues were purified by gel filtration on Sephadex G-100, as described in fig. 1. Low molecular weight peptides were purified by high voltage paper electrophoresis in liquid cooled tanks as in [1], some peptides being further purified by paper chromatography c.f. table 1.

Amino acid analyses were performed with a 'Durrum 500' or a 'BioCal 201' analyzer after hydrolysis witj redistilled HCl in vacuum sealed tubes for 16–20 hr at 110°C. The results are expressed in stoichiometric ratios without corrections. An approximate value of the tryptophan contents in TM-1 was obtained from the spectrophotometric measurements [6]. In the small peptides tryptophan was identified by Ehrlich staining [7] on filter paper. Sequencing was carried out by sequential Edman degradation–dansylation [8] and the 1-dimethylaminonapthalene-5-sulphonyl (Dns)-amino acids were identified by thin

 North-Holland Publishing Company – Amsterdam

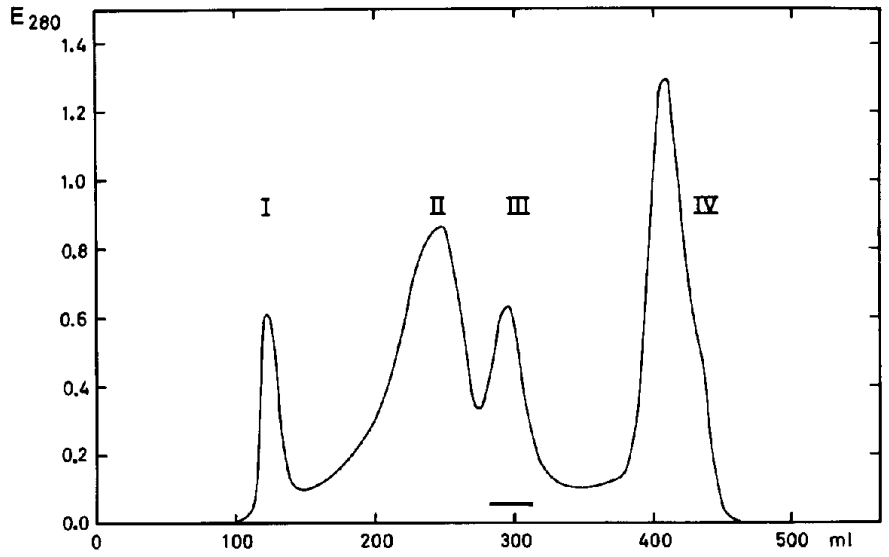

Fig. 1. Gel filtration of tryptic fragments from maleylated chymosin. The digest was remaleylated before it was applied to a column of Sephadex G-100 (2.5 X 100 cm). Eluent NH_4HCO_3 0.05 M pH 8.0. Flow rate 19 ml/hr. III represents TM-1 (the eluate indicated by the solid bar was subjected to re-gelfiltration before analysis).

layer chromatography on polyamide layer sheets [9, 10]. Occasionally *C*-terminal sequences were established by hydrolysis with carboxypeptidase-A [11].

The numbers of the tryptic and the chymotryptic peptides are rationalized according to their position in the peptide chain. The other peptides are just characterized with numbers from our protocols. Capital letters indicate the enzyme used in the digestion, T: Trypsin, C: Chymotrypsin, El: Elastase, Th: Thermolysin, TM: Tryptic digest of maleylated chymosin.

3. Results

Maleylated chymosin was subjected to digestion with trypsin under restricted conditions (a). In order to increase electrostatic repulsion between the fragments, the digest was remaleylated, and the products of the digestion were separated by gel filtration on Sephadex G-100. The results are illustrated in fig. 1. Peak I is an aggregation product, II contains two large fragments, III represents TM-1 and IV is a mixture of three smaller tryptic peptides (TM-2 and the two *C*-terminal peptides, TM-5 and TM-6) together with excess maleic acid.

Amino acid analyses of TM-1 gave the following composition:

Asp (6.0) Thr (4.0) Ser (5.3) Glu (5.1) Pro (3.9) Gly (4.1)
Ala (2.3) Val (3.8) 1/2 Cys (0.7) Met (0.3) Ile (2.2) Leu (4.1)
Tyr (3.7) Phe (3.9) His (0.8) Lys (2.9) Arg (1.0) Trp (ca. 1).

A chymotryptic digest of TM-1 was subjected to high voltage paper electrophoresis at pH 6.5. A guide strip was demaleylated and examined by the diagonal electrophoretic technique [4]. From the diagonal map it was evident that three off-diagonal spots were identical with those previously found in the *N*-terminal sequence [2]. TM-1 was thus identified as the *N*-terminal tryptic fragment. The chymotryptic peptides were purified and sequenced as summarized in table 1. The sequences of the peptides C-1 to C-4 have been published [2] and will not be further discussed. The electrophoretic mobility of C-5 indicated a single negative charge at pH 6.5 [12]. This peptide was further digested with papain from which two neutral peptides Leu–Gly and Thr–Pro–Pro–Gln were obtained together with one acidic Glu–Phe. This establishes the location of the amide. C-7 was obtained

Table 1
Analyses of peptides from which the amino acid sequence of TM-1 is established.

Residue no in the peptide chain	Peptide no	Electrophoretic mobility pH 6.5	Further purification or notes	Amino acid sequence
1– 8	C-1		a)	Gly–Glu–Val–Ala–Ser–Val–Pro–Leu
9–11	C-2		a)	Thr–Asn–Tyr
12–17	C-3		a)	Leu–Asp–Ser–Gln–Tyr–Phe
18–21	C-4		a)	Gly–Lys–Ile–Tyr
22–29	C-5	–0.45	BAWP	Leu–Gly–Thr–Pro–Pro–Gln–Glu–Phe
			+pH 3.5	0.9 1.0 0.9 2.0 2.1 1.0
30–33	C-6	0	pH 1.9	Thr–Val–Leu–Phe
				1.1 1.0 1.0 0.9
34–40	C-7a	–1.05	pH 3.5	Asp–Thr–Gly–Ser–Ser–Asp–Phe
				2.0 1.0 1.1 2.0 0.9
34–41	C-7b	–0.9	BAWP	Asp–Thr–Gly–Ser–Ser–Asp–Phe–Trp
			+pH 3.5	2.0 0.9 1.0 1.8 1.0 +
42–46	C-8		b)	Val,Pro,Ser,Ile,Tyr
47–54	C-9		c)	Cys–Lys–Ser–Asn–Ala–Cys–Lys–Asn
55–57	C-10a		c)	His–Gln–Arg
23–30	$El_{(876)}$	–0.40	BAWP	Gly–Thr–Pro–Pro–Gln–Glu–Phe–Thr
				1.0 2.0 1.8 2.2 1.0
23–31	$El_{(877)}$	–0.35	BAWP	Gly–Thr–Pro–Pro–Gln–Glu–Phe–Thr–Val
				0.9 1.7 2.2 2.2 1.0 1.2
33–37	$El_{(882)}$	–0.6	BAWP	Phe–Asp–Thr–Gly–Ser–Ser
				0.8 1.0 1.0 1.1 1.9
39–44	$El_{(875)}$	–0.5	BAWP	Asp–Phe–Trp–Val–Pro–Ser
				0.8 0.9 + 1.2 1.0 0.9
45–57	$Th_{(587)}$	+0.8	oxidized pH 6.5	Ile–Tyr–Cys–Lys–Ser–Asn–Ala–Cys–Lys–Asn–His–Gln–Arg
				0.8 0.7 1.8 1.9 1.0 2.3 1.0 1.2 1.0 1.2
20–29	$T_{(458)}$	–0.3	pH 3.5	Ile–Tyr–Leu–Gly–Thr–Pro–Pro–Gln–Glu–Phe
				0.9 0.8 1.0 1.2 0.9 2.2 2.0 0.9
30–46	$T_{(462)}$	–0.6	pH 3.5	Thr–Val–Leu–Phe–Asp–Thr–Gly–Ser–Ser–Asp–Phe–Trp–Val–Pro–Ser–Ile–Tyr
				1.8 1.8 1.0 1.8 2.1 1.0 2.7 + 1.0 0.9 0.8

Notation of the peptides is described in the text. (C-10a is the chymotryptic peptide from TM-1, the corresponding chymotryptic peptide of the entire chymosin overlaps TM-2 with Phe no. 58 [2]). The electrophoretic mobility at pH 6.5 is calculated relative to an internal marker of 1-dimethylaminonaphthalene-5-sulfonic acid (–1.0). Further purification indicates subsequent paper electrophoresis at the pH shown, or descending paper chromatography in butan-1-ol–acetic acid–water–pyridine (BAWP) (15:3:10:12, by vol) [22]. – Quantitative amino acid composition with stoichiometric ratios shown below the bars. + Positive staining for tryptophan. → Location of the residue by Edman degradation–dansylation. ← Residues liberated by carboxypeptidase-A. a) and c) the sequences of these peptides have been published in [1] and [2]. b) This peptide was never analysed (cf. the text), the expected amino acids of the peptide are however included in the table to complete the sequence of the chymotryptic peptides.

Table 2
Comparison between the *N*-terminal amino acid sequences of gastric proteases.

	10																			20
Chymosin	Gly	Glu	Val	Ala	Ser	Val	Pro	Leu	Thr	Asn	Tyr	Leu	Asp	Ser	Gln	Tyr	Phe	Gly	Lys	Ile
Bovine pepsin			Val	Ser	Gln	Glu	Pro	Leu	Gln	Asn	Tyr	(Leu	. Asx	. Thr	. Glx	. Tyr	. Phe	. Gly	. Thr	. Ile .
Porcine pepsin			Ile	Gly	Asp	Glu	Pro	Leu	Glu	Asn	Tyr	Leu	Asn	Thr	Glu	Tyr	Phe	. Gly	Thr	Ile
Common							Pro	Leu		Asn	Tyr	Leu				Tyr	Phe	Gly		Ile
	30																			40
Chymosin	Tyr	Leu	Gly	Thr	Pro	Pro	Gln	Glu	Phe	Thr	Val	Leu	Phe	Asp	Thr	Gly	Ser	Ser	Asp	Phe
Bovine pepsin	Tyr	. Ile	. Gly	. Thr	. Pro	. Ala	. Glx	. Asx	. Phe	. Thr)	Val	Ile	Phe	Asp	Thr	Gly	Ser	Ser	Asn	. Leu
Porcine pepsin	Gly	Ile	Gly	Thr	Pro	Ala	Gln	Asp	Phe	. Thr	Val	Ile	Phe	. Asp	Thr	Gly	Ser	Ser	Asn	. Leu
Common			Gly	Thr	Pro				Phe	Thr	Val		Phe	Asp	Thr	Gly	Ser	Ser		
	50																			60
Chymosin	Trp	Val	Pro	Ser	Ile	Tyr	Cys	Lys	Ser	Asn	Ala	Cys	Lys	Asn	His	Gln	Arg	Phe	Asp	Pro	Arg/
Bovine pepsin	Trp	Val	Pro	Ser	. Ile	Tyr	Cys	. Ser	Ser	Glu	Ala	Cys	Thr	Asn	His	Asn	Arg/				
Porcine pepsin	Trp	Val	Pro	Ser	. Val	Tyr	Cys	Ser	Ser	Leu	Ala	Cys	Ser	Asp	His	Asn	Gln	Phe	Asn	Pro	Asp/
Common	Trp	Val	Pro	Ser		Tyr	Cys		Ser		Ala	Cys			His						

Ch: Chymosin; Bp: Bovine pepsin; Pp: Porcine pepsin; ( ): The residues between brackets are only known from amino acid composition. . : The sequence or overlap is inferred from homology; / : End of peptide. The results are expressed in the single letter code according to the recommendations of IUB–IUPAC. ((1969) Biochem. J. 113, 1). The numbered bars below the structure correspond to the peptides shown in table 1.

in two forms with and without tryptophan, and the electrophoretic mobility indicated two negative charges, indicating absence of amides. C-8 was never prepared in amount sufficient for a reliable analysis. The reason for this apparent loss of C-8 may be that this rather apolar peptide was dissolved in the cooling liquid during electrophoresis. The sequences of C-9 and C-10 have previously been published [1]. The peptides were observed both separately and joined together due to incomplete cleavage of the Asn–His bond.

Some of the overlapping peptides were obtained after digestion with elastase, which in our experiments gave preferential cleavage after Leu and Ser, but cleavage after Ala, Asn, Thr and Val was also observed. Peptides representing most of TM-1 have been purified, but for the sake of clarity only those which contribute additional information will be considered. $El_{(876)}$ and $El_{(877)}$ provide together with $El_{(882)}$ the overlaps between C-5, C-6 and C-7, while $El_{(875)}$ extends the sequence of C-7 into the unidentified C-8. The last part of C-8 was found in a peptide from a thermolytic digest after the S–S bridge had been oxidized with performic acid ($Th_{(587)}$; only the two first residues were directly identified, but its characteristic amino acid composition clearly places this peptide as the C-terminal part of TM-1. Unspecific hydrolysis with trypsin in conventional digestion for 3 hr has caused many problems during the course of this work, but on some occasions the peptides produced by such unspecific activity have provided valuable overlaps. The overlap between C-4 and C-5 was obtained by $T_{(458)}$, and $T_{(463)}$ includes C-6, C-7 and C-8.

The structure which arises from the peptides in table 1 accounts for the amino acid composition of TM-1 except for an apparent methionine containing impurity.

4. Discussion

The results are compiled in table 2, which also includes published sequences of porcine pepsin together with preliminary results from bovine pepsin at present under investigation in our laboratory [13, 14]. The three fragments of porcine pepsin up to residue no 33 have been reported by Stepanov et al. (15–17). The sequence from Asp 34 to Asn 39 is from Chen and Tang [18], these authors did not find Phe 33, but considering the homology we have preferred to use the results of Stepanov et al. The sequence of porcine pepsin around Trp no 41 is from Dopheide and Jones [19], the sequence surrounding the S–S loop reported by Tang and Hartley [20] has been extended by Revina et al. [21]. In bovine pepsin the residues from no 12 to no 30 have not yet been sequenced, but the amino acid composition of the fragment fits very well into a homologous sequence.

The present results indicate that in the *N*-terminal part of the peptide chain approximately half of the amino acid residues in chymosin, bovine and porcine pepsin are common, the two pepsins being slightly more related to each other than to chymosin.

Acknowledgements

We thank Mr. E. Geertsen and Mrs P. Warnich-Hansen for valuable technical assistance. The starting material of rennet powder was a gift from Chr. Hansen's Laboratory, Copenhagen. The work was supported by the Danish Natural Science Research Council.

References

[1] Foltmann, B. and Hartley, B.S. (1967) Biochem. J. 104, 1064.
[2] Foltmann, B. (1970) Phil. Trans. Roy. Soc. London, Ser. B 247, 147.
[3] Foltmann, B. (1966) Compt. Rend. Trav. Lab. Carlsberg 35, 143.
[4] Butler, P.J.G., Harris, J.I., Hartley, B.S. and Leberman, R. (1969) Biochem. J. 112, 679.
[5] Hapner, K.D. and Wilcox, Ph.E. (1970) Biochemistry 9, 4470.
[6] Beaven, G.H. and Holiday, E.R. (1952) Advan. Protein Chem. 7, 320.
[7] Leggett Bailey, J. (1967) Techniques in Protein Chemistry p. 21, Elsevier, Amsterdam, London, New York.
[8] Gray, W.R. (1967) in: Methods in Enzymology, Vol. 11, p. 469 (Hirs, C.H.W., ed.), Academic Press, New York, London.
[9] Woods, K.R. and Wang, K.-T. (1967) Biochim. Biophys. Acta 133, 369.
[10] Hartley, B.S. (1970) Biochem. J. 119, 805.
[11] Ambler, R.P. (1967) in: Methods in Enzymology, Vol. 11, p. 436 (Hirs, C.H.W., ed.), Academic Press, New York, London.
[12] Offord, R.E. (1966) Nature 211, 591.
[13] Harboe, M. and Foltmann, B. (1973) FEBS Letters 34, 311.
[14] Harboe, M. and Foltmann, B., unpublished results.
[15] Stepanov, V.M., Timokhina, E.A., Baratova, L.A., Belyanova, L.P., Korzhenko, V.P. and Zhukova, I.G. (1971) Biochem. Biophys. Res. Commun. 45, 1482.
[16] Pugacheva, I.B., Stepanov, V.M. and Amirkhanyan, M.M. (1971) Biokhimiya 36, 993.
[17] Stepanov, V.M. (1972) submitted to 'Atlas of Protein Sequence and Structure' Vol. 5, (Dayhoff, M.O., ed.), Natl. Biomed. Res. Foundation, Washington, DC, USA.
[18] Chen, K.C.S. and Tang, J. (1972) J. Biol. Chem. 247, 2566.
[19] Dopheide, T.A.A. and Jones, W.M. (1968) J. Biol. Chem. 243, 3906.
[20] Tang, J. and Hartley, B.S. (1970) Biochem. J. 118, 611.
[21] Revina, L.P., Yakhitova, E.A., Pugacheva, I.B., Lapuk, Ya.I. and Stepanov, V.M. (1972) Biokhimya 37, 1074.
[22] Waley, S.G. and Watson, J. (1953) Biochem. J. 55, 328.