SEQUENTIAL STUDIES ON PEPTIDES PREPARED BY CHYMOTRYPTIC DIGESTION OF CHICKEN PEPSINOGEN AND PARTIAL COVALENT STRUCTURE OF THE PROTEIN

Miroslav BAUDYŠ and Vladimír KOSTKA

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

Received January 27th, 1982

Dedicated to Prof. Dr S. M. Rapoport on the occasion of his 70th birthday.

The peptides obtained by chymotryptic digestion of the S-sulfo derivative of chicken pepsinogen and of the large tryptic fragments of the same derivative were subjected to sequential analysis. The results obtained afforded information on nonoverlapping sequences of 329 amino acid residues in the molecule of the zymogen and yielded data on links between individual fragments in the middle part of the chain obtained in earlier work. A partial covalent structure of the protein is presented which contains 356 amino acid residues in eight segments of uninterchangeable order.

In earlier papers from this Laboratory experiments were reported designed to elucidate the amino acid sequence of different parts of the polypeptide chain of chicken pepsin and its zymogen. In these experiments we determined the primary structure of the N- and C-terminal part of the molecule[1,2], the neighborhood of the half-cystine residues and the way in which these residues are linked one to another by disulfide bonds[3], the amino acid sequence around the aspartic acid residue reactive to diazo inhibitors[4], the vicinity of the cysteine residue[5], and the site of attachment of the sugar moiety to the polypeptide chain[6]; the amino acid sequence of a number of residues derived from the middle part of the chain has also been elucidated[7]. The data obtained enabled us to propose a partial primary structure of chicken pepsinogen[8] containing 322 amino acid residues arranged in 19 segments of various lengths. The arrangement of these segments required the knowledge of their overlaps. Since the majority of our sequential data had been derived from the investigation of tryptic[2] and thermolytic[7] peptides we assumed that these overlaps might yield the chymotryptic peptides.

This paper describes the results of studies on peptides prepared by chymotryptic digestion of chicken pepsinogen and of its large tryptic fragments.

EXPERIMENTAL

Material

The preparation of chicken pepsinogen, its S-sulfo derivative, and of its large tryptic fragments have been described before[1,2,6]. Chymotrypsin was a twice crystallized commercial preparation of Worthington Biochemical Corp. (Freehold, NJ., USA). Sephadex G-25 was from Pharmacia (Uppsala, Sweden), Dowex 1—X2 from Fluka (Buchs, Switzerland). The source of the chemicals for automatic and manual sequential degradation of peptides and for the identification of the degradation products has been described elsewhere[7].

Methods

Chymotryptic digestion of S-sulfo-pepsinogen: The protein was dissolved to a 1% solution in water (final pH 8·2, adjusted by ammonia). Chymotrypsin (enzyme to substrate ratio 1 : 100, w/w) was added and the solution was incubated at 37°C for 2 h with occasional adjustment of pH to 8·0 by ammonia. The digestion was discontinued afterwards by the addition of glacial acetic acid (to pH 4·0). A small quantity of precipitate which had formed was centrifuged off and washed twice with water at pH 4·0. The washings were pooled with the first supernatant and the solution was lyophilized. The chymotryptic digestion of the large tryptic peptides has been described before[6].

Ion exchange chromatography of the chymotryptic digest: The chymotryptic digest was dissolved to a 0·3% solution in 1% *sym*-collidine acetate (pH 8·7). The pH of the sample was adjusted to 9·0 by 2M-NaOH. The sample solution was subsequently placed onto a column of Dowex 1—X2 (2·4 × 50 cm), equilibrated with 1% *sym*-collidine acetate (pH 8·7). The column was eluted by an exponential concentration gradient of acetic acid[9] as described elsewhere[10]. The effluent fractions were pooled according to paper chromatography[11] of aliquots. The pooled material was rotary evaporated to dryness and then fractionated by paper techniques.

Amino acid analysis: The samples of peptides were hydrolyzed 20 or 70 h in 6M-HCl at 110°C *in vacuo*. The amino acid analysis[12] was carried out in a Durrum D-500 analyzer. The tryptophan content was determined according to Penke and coworkers[13].

N-*Terminal end group analysis of peptides*: The dansyl technique[14] and thin layer chromatography identification of the products on polyamide layers[15] were used.

Sequential analysis of peptides: The phenyl isothiocyanate method[16] was employed. The amino acid phenylthiohydantoins resulting from the conversion of the thiazolinones obtained either by manual or automatic (Beckman 890 C Amino Acid Sequenator) degradation were identified by thin layer chromatography[17] on silica gel or by gas chromatography (Beckman GC-65 Gas Chromatograph), either as such or after silylation[18].

RESULTS AND DISCUSSION

In earlier sequential studies on chicken pepsinogen[2-4] we used tryptic[2,6] and thermolytic[7] digestion for the fragmentation of the chain of the protein. Chymotryptic digestion was employed in one special experiment[3] only and the digest was not systematically investigated. Such an investigation was carried out in this study in an effort to obtain overlaps which would enable us to link together the tryptic fragments

TABLE I

Amino acid composition of peptides obtained by chymotryptic digestion of chicken S-sulfo-pepsinogen (C-peptides) and by chymotryptic digestion of large tryptic fragments of the same derivative (TC-peptides)

The values given were determined by methods described in the text and are not corrected

Designation of peptide	Number of amino acid residues mol/mol of peptide																	
	Lys	His	Arg	Cys	Asp	Thr	Ser	Glu	Pro	Gly	Ala	Val	Met	Ile	Leu	Tyr	Phe	Trp
C-1			0·9						1·0			0·8			0·9			
C-2	1·7									1·0								
C-3							0·9								1·1			
C-4	1·1		0·8					0·9							0 9			
C-5	0·8	0·6			2·1			1·1		1·1					3·3		1·0	
C-6	1·1														0·9			
C-7	0·9				1·1		1·0		1·0		1·0					0·4		
C-8						1·7	1·1	2·2	0·8		1·0		0·9			0·4		
C-9					1·1	0·9										0·7		
C-10					1·1		1·0				1·0		0·6			1·8		
C-11					1·3	2·0	1·3	1·8	1·0	1·9				1·7			1·0	
C-12	1·0	0·8		1·0[a]	1·3		3·1				1·2							
C-13	1·2		0·8		1·4	1·1	3·0		1·0							0·7	0·8	
C-14											1·1			0·9		1·0		
C-15						1·4	2·1			4·0			0·9	1·1	1·4	0·7		
C-16							1·0			1·7				0·9	1·0	0·7		
C-17						0·9	1·7	1·9	0·9	1·8					1·0		1·8	
C-18						1·0	1·6	1·9	1·0	1·6					0·9		0·9	
C-19		1·2					0·9	1·0							0·9			
C-20					0·9			0·8			0·6	0·9			1·0		0·9	
C-21	1·0				1·2	1·0	1·8	1·4		2·3					1·0		0·9	
C-22												1·1			1·0		1·0	
C-23	1·0				1·6	1·7			0·9	2·0				1·0		0·8		
C-24	1·2					1·7				1·1				1·0		0·8		
C-25						1·0	0·9	1·1			1·1					0·9		0·8
C-26	1·0		1·0		2·0	1·9		1·0		1·2		2·0	0·4	0·9		0·7		
C-27	0·8		1·0		2·1	1·1				1·3		2·0				0·7		
C-28				0·6[a]	1·6	2·5	1·3	1·4		1·4	1·1	0·8		0·8	1·6		0·9	
C-29								1·0	1·0	1·0	1·0	1·0	0·3			0·7		
C-30		1·2			1·0					1·0	1·0			0·9			1·0	
C-31			1·0					1·2						0·9		0·9		
C-32			1·0		2·4						1·1	0·7		0·7			1·1	
C-33			1·0					1·0						0·6				
C-34												0·8		0·6			1·0	
C-35	1·0		0·7		2·6		1·8		1·0	1·0	0·9	1·0			2·0			
C-36	0·9						1·9		0 9	1·0		1·1			2·3			
C-37	2·0	0·7							1·0							0·5		
C-38	3·0						0·8			1·0					1·0			
C-39		0 6					1·0							0·9				
C-40	1·8		1·0	0·9	4·5	1·3	2·5	1·4	1·5	2·5		2·3		3·5	2·4		1·0	

TABLE I
(*Continued*)

Designation of peptide	Number of amino acid residues mol/mol of peptide																	
	Lys	His	Arg	Cys	Asp	Thr	Ser	Glu	Pro	Gly	Ala	Val	Met	Ile	Leu	Tyr	Phe	Trp
TC-1		0·8							1·0			1·1			1·0	0·8		
TC-2					1·2	2·6	1·1	2·0	0·8		1·0		0·7			1·7		
TC-3					1·0		0·9				1·0		0·8			2·0		
TC-4					1·0	1·8	1·0	2·0	1·0	1·9				1·9			0·9	
TC-5					2·0	1·0	1·9			1·1					1·0			0·8
TC-6							0·9		1·0			0·9		1·0		0·9		
TC-7					1·0	1·8	0·9	1·0				1·9				0·9		
TC-8											1·0			0·9				
TC-9						1·0	1·9			4·0			0·9	1·0	1·0	0·9		
TC-10					1·0	1·0	1·0				1·0	1·8						
TC-11						1·0	1·8	2·0	0·9	1·9					1·0		2·0	
TC-12				1·0[a]	2·0					2·1				1·0	1·8	1·8	1·0	
TC-13					2·0	1·1	3·6		2·0	1·2	1·9	1·0	1·0	1·0			1·9	
TC-14		0·9					1·0	1·1					0·8		1·0			
TC-15					1·0			1·0			1·0	0·9			1·0		0·9	
TC-16							0·9					1·0				1·0		
TC-17	1·0			1·0[a]														
TC-18												1·0			1·0		1·0	
TC-19	1·1	0·8		0·9[a]	5·0	1·0	3·8	1·2	0·9	2·1		1·8		2·0	2·0		0·9	
TC-20		1·5			1·0					1·0	1·0			0·9		1·0		
TC-21						1·0	1·0		1·0		2·0				1·0	1·0		
TC-22				1·0[a]	3·1	1·9	1·1	4·7	0·9	3·7		0·9	1·7		2·9		1·1	0·8
TC-23					1·0					1·0		0·9		0·9	0·9		1·0	
TC-24			1·0											1·0				
TC-25						1·0	1·9								1·0			
TC-26	1·0						0·8								1·0			

[a] Determined as cysteic acid after oxidation by performic acid[19].

and to arrange the small thermolytic peptides. S-Sulfo-pepsinogen and also certain large tryptic peptides, obtained in earlier work[2,6], were used to start with.

The chymotryptic digestion of S-sulfo-pepsinogen was allowed to proceed at an enzyme to substrate ratio of 1 : 100 for 2 h at 37°C. We assumed that a shorter hydrolysis time might lead to preferential cleavage of the bonds at the carboxyl side of tryptophan, tyrosine, phenylalanine, and leucine. We prepared, therefore, the digest under these conditions and were able to isolate a total number of 40 peptides (*cf.* Table I, peptides C-1 to C-40), ranging in size from a dipeptide (C-3) to a 28-resi-

TABLE II

Amino acid sequence of peptides from chymotryptic digest of chicken S-sulfo-pepsinogen (C-peptides) and from the chymotryptic digest of large tryptic fragments (TC-peptides) of the same pepsinogen derivative. The sequences were determined by methods described under Experimental. In certain cases the peptide was ascribed the C-terminal amino acid with respect to the known specificity of chymotrypsin. The amino acids in peptides incompletely sequenced are arranged according to the expected order

Designation	Structure (composition)
C-1	(Arg,Val,Pro)Leu
C-2	Lys-Gly-Lys
C-3	Ser-Leu
C-4	(Arg,Lys,Glx)Leu
C-5	Lys-Asp-His-Gly-Leu-Leu-Glu(Asx,Phe,Leu)
C-6	Leu-Lys
C-7	(Asn,Pro,Ala,Ser,Lys)Tyr
C-8	Thr-Ala-Thr-Glu-Ser(Tyr,Glx,Pro,Met)
C-9	(Thr,Asn)Tyr
C-10	(Met,Asp,Ala,Ser,Tyr)Tyr
C-11	Gly(Thr,Ile,Ser,Ile,Gly,Thr,Pro,Glx,Glx,Asx)Phe
C-12	(Ser,Ser,Ala,Cys,Ser,Asx,His,Lys)
C-13	Arg-Phe-Asp(Pro,Ser,Lys,Ser,Ser,Thr,Tyr)
C-14	Ile-Ala-Tyr
C-15	Gly(Thr,Gly,Ser,Met,Ser,Gly,Ile,Leu,Gly,Tyr)
C-16	(Ser,Gly,Ile,Leu,Gly,Tyr)
C-17	Gly-Leu(Ser,Glx,Thr,Glx,Pro,Gly,Ser,Phe)Phe
C-18	Gly(Leu,Ser,Glx,Thr,Glx,Pro,Gly,Ser)Phe
C-19	(Ser,His,Gln)Leu
C-20	Val-Ala-Gln-Asp(Leu,Phe)
C-21	Leu-Ser-Lys(Asx,Gly,Glx,Thr,Gly,Ser)Phe
C-22	Val-Leu-Phe
C-23	Gly-Gly-Ile-Asp-Pro-Asn-Tyr-Thr-Thr-Lys
C-24	Thr-Thr-Lys-Gly-Ile-Tyr
C-25	Ser-Ala-Glu-Thr-Tyr-Trp
C-26	(Glx,Ile,Thr,Met,Asx,Arg,Val,Thr,Val,Gly,Asx,Lys)Tyr
C-27	Asp-Arg-Val-Thr-Val-Gly-Asn-Lys-Tyr
C-28	(Phe,Thr,Cys,Glx,Ala,Ile,Val,Asx,Thr,Gly,Thr,Ser,Leu,Leu)
C-29	Val-Met-Pro-Gln-Gly-Ala-Tyr
C-30	His-Ile-Asn-Gly(His, Ala)Phe
C-31	Ile-Arg-Glu-Tyr
C-32	(Val,Ile,Phe,Asx,Arg,Ala,Asx)
C-33	(Ile,Arg,Glx)
C-34	(Val,Ile)Phe
C-35	(Asx,Arg,Ala,Asx,Asx,Lys,Val,Gly,Leu,Ser,Pro,Leu,Ser)

TABLE II
(*Continued*)

Designation	Structure (composition)
C-36	(Lys,Val,Gly,Leu,Ser,Pro,Leu,Ser)
C-37	Lys-Lys-His-Pro-Tyr
C-38	Lys-Lys-Gly-Lys-Ser-Leu
C-39	Ser-Ile-His
C-40	Asn-Arg-Ile-Ile-Lys-Asp-Leu-Gly-Val (Ser, Ser, Asp, Gly, Glu, Ile, Ser, Cys, Asp, Asp, Ile, Ser, Lys, Leu, Pro, Asp, Val, Thr) Phe
TC-1	Tyr-His-Pro-Val-Leu
TC-2	Thr-Ala-Thr-Glu-Ser-Tyr-Glu-Pro-Met-Thr-Asn-Tyr
TC-3	Met-Asp-Ala-Ser-Tyr-Tyr
TC-4	Gly-Thr-Ile-Ser-Ile-Gly-Thr-Pro-Gln-Gln-Asp-Phe
TC-5	Asp-Thr-Gly-Ser-Ser-Asn-Leu-Trp
TC-6	Val-Pro-Ser-Ile-Tyr
TC-7	Val-Ser-Thr-Asn-Glu-Thr-Val-Tyr
TC-8	Ile-Ala-Tyr
TC-9	Gly-Thr-Gly-Ser-Met-Ser-Gly-Ile-Leu-Gly-Tyr
TC-10	Asp-Thr-Val-Ala-Val-Ser
TC-11	Gly-Leu-Ser-Glu-Thr-Glu-Pro-Gly-Ser-Phe-Phe
TC-12	Tyr-Tyr-X-Asn-Phe-Asp-Gly-Ile-Leu-Gly-Leu
TC-13	Ala-Phe-Pro-Ser-Ile-Ser-Ser-Ser-Gly-Ala-Thr-Pro-Val-Phe-Asp-Asn-Met
TC-14	Met-Ser-Gln-His-Leu
TC-15	Val-Ala-Gln-Asp-Leu-Phe
TC-16	Ser-Val-Tyr
TC-17	Cys-Lys
TC-18	Val-Leu-Phe
TC-19	Asp-Leu-Gly-Val-Ser-Ser-Asp-Gly-Glu-Ile-Ser-Cys-Asp-Asp-Ile-Ser-Lys-Leu-Pro-Asp--Val-Thr-Phe-His
TC-20	His-Ile-Asn-Gly-His-Ala-Phe
TC-21	Thr-Leu-Pro-Ala-Ser-Ala-Tyr
TC-22	Val-Leu-Asn-Glu-Asp-Gly-Ser-Cys-Met-Leu-Gly-Phe-Glu-Asn-Met-Gly-Thr-Pro-Thr--Glu-Leu-Gly-Glu-Gln-Trp
TC-23	Ile-Leu-Gly-Asp-Val-Phe
TC-24	Ile-Arg
TC-25	Ser-Ser-Thr-Tyr
TC-26	Leu-Ser-Lys

X in peptide TC-12 stands for a phenylthiohydantoin not determined unambiguously.

due peptide (C-40). Some peptides overlap each other because of chymotryptic cleavage of the same part of the chain at different residues (C-2, C-3 and C-38, C-6

and C-37, C-15 and C-16, C-17 and C-18, C-23 and C-24, C-26 and C-27, C-31 and C-33, C-32 and C-34, C-35 and C-36). Of the 40 peptides isolated (*cf.* Table II) 4 are C-terminated with lysine and were therefore formed by cleavage with trypsin obviously present as a contaminant in the chymotrypsin preparation used. Of additional 36 peptides 30 are C-termined with an amino acid corresponding to the main site of chymotryptic attack, *i.e.* with tyrosine, phenylalanine, and leucine. Of the

remaining 6 peptides two are C-terminated with serine (the C-terminal peptides of the molecule, C-35and C-36); in other cases the chain was cleaved after an asparagine or a glutamic acid residue (peptides C-32 and C-33, resp.), after methionine (C-8) and histidine (C-39).

The peptides prepared by the cleavage of large tryptic fragments[6] (TC-peptides) range in size from 2 to 25 amino acid residues. They are exclusively C-terminated by amino acids corresponding to the main specificity of chymotrypsin (and trypsin) except for two peptides, TC-13 and TC-10; the existence of the latter is especially striking. The peptides C-terminated with basic residues are derived from the carboxyl termini of the corresponding tryptic fragments.

All TC-peptides and certain selected C-peptides were subjected to detailed sequential analysis. The data obtained provide information on nonoverlapping sequences of 329 residues in the chain of pepsinogen.

The results of earlier sequential studies on chicken pepsinogen and pepsin[1-7] permit us to unambiguously determine the positions of all peptides in the chain. Figure 1 shows the partial amino acid sequence of the polypeptide chain of chicken pepsinogen containing 356 amino acid residues and summarizing the results of all our sequential experiments (this study included). The chymotryptic peptides obtained either by direct digestion of S-sulfo-pepsinogen or from the large tryptic fragments are set in capital letters and marked by horizontal bars and the number of the peptide. As can be seen in the figure the C- and TC-peptides practically cover the entire chain of the protein. In spite of that some individual sites of the chain have not been obtained either from the direct digest or from the large tryptic fragments. Some of these sites represent the overlaps sought. We have not been able, *e.g.*, to obtain the peptide preceding the first active aspartic acid residue (32). It is reasonable to assume that the strongly hydrophobic peptide Ser-Val-Ile-Phe, which most likely

←

FIG. 1

Peptides arising from chymotryptic digestion of chicken pepsinogen and partial structure of the protein. The peptides studied are marked (with the exception of overlapping peptides C-16, C-18, and TC-24) by horizontal bars and the corresponding symbol (C, TC). The sequences elucidated in the present study are set in capital letters, the sequences determined in earlier work[1-7] are set in small letters. The interruption of the chain is marked by a slanting line (/). The segment cleaved off (at the site marked →) in the process of zymogen activation (propart) is continuously numbered (p1 through p42). The numbers in brackets above certain important residues correspond to the pepsin numbering system proposed by Foltmann[21], these numbers are shown to facilitate a comparison of our structure with the covalent and three-dimensional structures of other carboxyl proteinases. The half-cystine residues (45) and (50), (206) and (210), and 250) and (283) are linked by disulfide bonds[3]

is formed by chymotryptic cleavage of this site, could have been lost during isolation procedure. Chymotyptric cleavage after the serine residue at the C-terminus of peptide TC-10 is particularly striking. As yet, we cannot make any final conclusions on the total number of amino acid residues picked up in the individual digests and therefore we do not know the exact length of the segment between the C-terminal serine of peptide TC-10 and the sequence Ile-Phe-Gly-... following in the chain.

In our earlier study[8] an effort has been made to summarize our data in a structure containing 322 amino acid residues arranged in 19 segments of different degree of interchangeability in the chain. The results of the present analysis of the chymotryptic digest of S-sulfo-pepsinogen and of its large tryptic fragments permit us to extend this structure to 356 residues arranged in 7 segments as shown in Fig. 1. Even though we are still missing 6 overlaps, these segments are not interchangeable. This is evident from their positions in the chain (*e.g.* in the case of the two terminal segments) and from their homology with the covalent structures of the corresponding parts of molecules of other carboxyl proteinases and their zymogens. We have pointed to these homologies in detail in some of our preceding experimental papers[1-7] and also in two review articles[8,20]. The determination of the missing overlaps and of the complete amino acid sequence of chicken pepsinogen will result from the final sequential experiments now under progress in our Laboratory.

We are indebted to Mrs E. Bulantová and Mrs J. Zaoralová for skillful technical assistance in this study. We thank Mr K. Grüner and Mrs A. Čermáková for the sequential degradation experiments. The amino acid analyses were carried out by Mr J. Zbrožek, Mrs V. Himrová and Mrs E. Dršková.

REFERENCES

1. Kostka V., Keilová H., Grüner K., Zbrožek J.: This Journal *42*, 3691 (1977).
2. Baudyš M., Keilová H., Kostka V.: This Journal *45*, 1144 (1980).
3. Keilová H., Baudyš M., Kostka V.: This Journal *44*, 2284 (1979).
4. Keilová H., Kostka V., Baudyš M.: This Journal *45*, 2131 (1980).
5. Pohl J., Baudyš M., Kostka V.: This Journal, in press.
6. Baudyš M., Kostka V., Grüner K., Pohl J.: This Journal *47*, 709 (1982).
7. Baudyš M., Kostka V., Keilová H.: This Journal *46*, 1994 (1981).
8. Kostka V., Keilová H., Baudyš M. in the book: *Proteinases and their Inhibitors* (V. Turk, L. Vitale, Eds.), p. 125. Mladinska knjiga — Pergamon Press, Ljubljana — Oxford 1981.
9. Guest J. R., Carlton B. C., Yanofsky C.: J. Biol. Chem. *242*, 5397 (1967).
10. Kostka V.: This Journal *46*, 640 (1981).
11. Walley S. G., Watson J.: Biochem. J. *55*, 328 (1953).
12. Spackman D. H., Stein W. H., Moore S.: Anal. Chem. *37*, 1108 (1965).
13. Penke B., Ferenczi R., Kovacs K.: Anal. Biochem. *60*, 45 (1974).
14. Gray W. R.: Methods Enzymol. *11*, 139 (1967).
15. Hartley B. S.: Biochem. J. *119*, 805 (1970).
16. Edman P.: Acta Chem. Scand. *4*, 277 (1950).
17. Edman P., Henschen A. in the book: *Protein Sequence Determination* (S. B. Needleman, Ed.), Vol. 8, p. 232, Springer, Berlin 1975.

18. Klebe J. F., Finbeiner H., White D. M.: J. Amer. Chem. Soc. *88*, 3390 (1966).
19. Moore S.: J. Biol. Chem. *238*, 235 (1963).
20. Kostka V.: *Thesis*, Institute of Organic Chemistry and Biochemistry, Czechoslovak Academy of Sciences, Prague 1981.
21. Foltmann B.: Assays in Biochemistry *17*, 52 (1982).

Translated by the author (V. K.).