Schrode, J., and Laskowski, M., Jr. (1971), *Fed. Proc., Fed. Amer. Soc. Exp. Biol. 30*, 1320 Abstr.
Schroeder, D. D., and Shaw, E. (1968), *J. Biol. Chem. 243*, 1968.
Sealock, R. W. (1972), Ph.D. Thesis, Purdue University, West Lafayette, Ind.
Tschesche, H. (1967), *Z. Physiol. Chem. 348*, 1216.
Tschesche, H., and Klein, H. (1968), *Z. Physiol. Chem. 349*, 1645.
Tschesche, H., Klein, H., and Reidel, G. (1971), *Proc. Int. Conf. Proteinase Inhibitors, 1st, 1970*, 299.
Tschesche, H., and Obermeier, R. (1971), *Proc. Conf. Proteinase Inhibitors, 1st, 1970*, 135.
Tschesche, H., and Wachter, E. (1970), *Eur. J. Biochem. 16*, 187.
Wyckoff, H. W., Tsernoglou, D., Hanson, A. W., Knox, J. R., Lee, B., and Richards, F. M. (1970), *J. Biol. Chem. 245*, 305.

Determination of the Amino Acid Sequence of Porcine Trypsin by Sequenator Analysis†

Mark A. Hermodson,*‡ Lowell H. Ericsson, Hans Neurath, and Kenneth A. Walsh

ABSTRACT: The amino acid sequence of porcine trypsin has been determined by sequenator analysis of the reduced and S-pyridylethylated protein and of eight suitably chosen peptide fragments. The fragments were the products of cleavage by autolysis, cyanogen bromide, 2-(2-nitrophenylsulfenyl)-3-methyl-3-bromoindolenine, hydroxylamine, and trypsin, respectively. All but the last 2 of the 223 amino residues were uniquely placed by these analyses. Comparison of this sequence with that of bovine trypsin indicated 82% identity, corresponding to a unit evolutionary period of approximately 3 million years. Of the 41 amino acid substitutions, 36 are on the surface of the bovine enzyme and 5 in the interior. The latter are of the conservative type. All residues known to be components of the active site of bovine trypsin are present in identical positions in porcine trypsin, but the porcine enzyme does not possess the calcium binding site identified in the bovine enzyme.

The automated method of Edman and Begg (1967) for determining amino acid sequences of proteins and large peptides offers several major advantages over previous procedures. The method is more rapid; it is applicable to longer sequences; and hence the elucidation of the entire sequence requires a smaller number of fragments than do conventional methods. Ideally, it might be possible to determine the amino acid sequence of a protein entirely by sequenator analysis. Practical restrictions are imposed less by the sequenator method *per se* (Smithies *et al.*, 1971; Hermodson *et al.*, 1972a) than the paucity of efficient procedures for the cleavage of polypeptide chains at specific sites. As a test of the feasibility of deriving a total protein sequence solely by automated techniques, we have analyzed porcine trypsin. This enzyme resembles bovine trypsin in molecular weight and amino acid composition (Walsh, 1970) and like the latter exists in a single-chain (β) form, and a two-chain (α) form which is a product of autolysis (Schroeder and Shaw, 1968; Hermodson *et al.*, 1972a). Elucidation of the amino acid sequence of porcine trypsin can also contribute to an understanding of the homology and phylogenetic variations within the class of trypsin-like enzymes.

Materials and Methods

Porcine trypsin (Novo Industri, Copenhagen) was purified by affinity chromatography on columns of chicken ovomucoid covalently linked to Sepharose (Robinson *et al.*, 1971). The purified enzyme was a 10:1 mixture of the β (single-chain) and α (two-chains) forms (Hermodson *et al.*, 1972a). The protein was reduced and S-pyridylethylated (Cavins and Friedman, 1970) in solutions containing 20–30 mg of protein/ml, 6 M guanidine hydrochloride, 0.13 M Tris, and 0.1 mg of ethylenedinitrilotetraacetate/ml (pH 7.6). Dithioerythritol was added to yield a 20-fold molar excess over the concentration of protein disulfide groups. After 3 hr at room temperature, the mixture was treated for 90 min with 3.0 mol of 4-vinylpyridine (Baker)/mol of dithioerythritol. The solution was then acidified to pH 2.0 with 88% formic acid and the alkylated protein separated from reagents and salts on a column of Sephadex G-75 equilibrated and eluted with 9% aqueous formic acid. Three fractions were obtained (Figure 1) corresponding to the single chain of β-trypsin and to the two fragments of α-trypsin. These fractions are designated β-trypsin and fragments α-N (amino terminal) and α-C (carboxyl terminal), respectively. Fragments α-N and α-C were further purified by chromatography on the same column.

Sequenator analyses were performed with a Beckman Sequencer, Model 890, by the method of Edman and Begg (1967) as modified by Hermodson *et al.* (1972a). The sequenator reagents were of "Sequenal" grade (Pierce Chemical Co.).

Amino acid analyses were performed with a Beckman amino acid analyzer, Model 120C. *S*-Pyridylethylcysteine is resistant to acid hydrolysis (Cavins and Friedman, 1970) and elutes as a single discrete peak between ammonia and arginine. A chromatographically pure standard of PE-cysteine[1] was obtained from Pierce Chemical Co. and prior to use was dried

† From the Department of Biochemistry, University of Washington, Seattle, Washington 98195. *Received March 2, 1973*. This work has been supported by research grants from the National Institutes of Health (GM 15731) and the American Cancer Society (NP-18N).

‡ Present address: Division of Medical Genetics, University of Washington, Seattle, Wash. 98195.

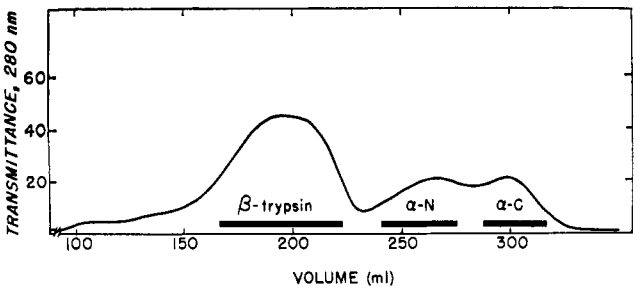

FIGURE 1: Separation of PE-β-trypsin and fragments α-N and α-C by gel filtration on Sephadex G-75. Reduced and S-pyridylethylated porcine trypsin (125 mg) was applied in a volume of 6 ml to a column (115 × 2.5 cm) of Sephadex G-75, equilibrated and eluted with 9% formic acid at 25 ml/hr. Fractions were pooled as indicated by the horizontal bars.

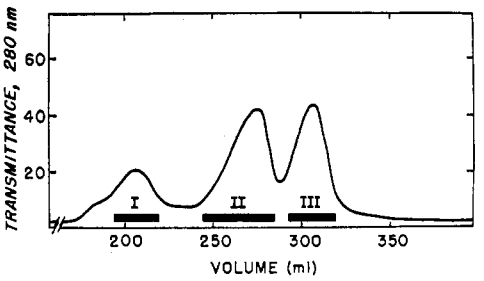

FIGURE 2: Fractionation of a cyanogen bromide digest of PE-β-trypsin (100 mg) on a column (115 × 2.5 cm) of Sephadex G-50, equilibrated and eluted with 9% formic acid at 30 ml/hr. Fractions were pooled as indicated by the horizontal bars.

for 3 days at $<10^{-2}$ Torr. Reaction with ninhydrin gave a color yield 0.82 of that of lysine.

Cleavage of methionyl bonds in β-trypsin (Gross, 1967) was effected in 70% formic acid (20–30 mg of protein/ml) by adding solid cyanogen bromide (2 mg of reagent/mg of protein). After 18–20 hr of reaction at room temperature (in the dark), the solution was diluted with 5–10 volumes of water and lyophilized. The products were separated by gel filtration on a column of Sephadex G-50 which was equilibrated and eluted with 9% formic acid. Three fractions were obtained and designated fractions I, II, and III (Figure 2).

Cleavage of Asn–Gly peptide bonds (Bornstein, 1969) was effected with hydroxylamine in a solution containing 5–10 mg of protein/ml, 2 M NH_2OH, 0.2 M K_2CO_3, and 6 M guanidine-HCl (pH 9.5). After 4 hr of reaction at 45°, the mixture was acidified by dropwise addition of 88% formic acid while the solution was agitated vigorously on a Vortex mixer (one or two drops of octanol were added to prevent foaming). The products were fractionated by gel filtration.

Cleavage of tryptophanyl bonds (Omenn *et al.*, 1970) was effected by dissolving the protein in 80% formic acid (20 mg protein/ml) and adding 50 equiv of BNPS-Skatole[1] (Pierce Chemical Co.). The mixture was allowed to react for 30 min at room temperature, then diluted with water to lower the formic acid concentration to 70%. Mercaptoethanol (0.1-ml/ml solution) was added to destroy the excess reagent and to reduce methionine sulfoxide residues to methionine residues. The mixture was allowed to react overnight at room temperature. Following twofold dilution with water, it was extracted with 1-ml portions of ethyl acetate until the extracts were colorless. The mixture of peptides in the aqueous layer was fractionated by gel filtration.

Cleavage of arginyl bonds was effected with trypsin after acylation of the ε-amino groups of lysyl residues. The protein, or peptide fragment, was dissolved in 6 M guanidine-HCl (20–30 mg of protein/ml), the pH was adjusted to 8.0 on a pH-Stat at 20° by the addition of 12 N NaOH and citraconic anhydride (liquid) or succinic anhydride (finely divided powder) was added in 25-mg portions up to ten times the weight of the protein with vigorous stirring. The rate of additions was slow enough so that the pH of the solution remained between pH 7 and 8. The acylated product was then desalted either by dialysis at pH 8.0 or by gel filtration on Sephadex G-25 equilibrated and eluted with 0.1 N NH_4HCO_3. The modified protein, or peptide, was treated with trypsin (1 mg of Worthington "TPCK-trypsin"/100 mg of substrate) at pH 8.0 for 5 min at 37°. Following tryptic digestion, the citraconyl groups were removed by lowering the pH to 2.5 by the addition of formic acid and incubation at room temperature for 3 hr. The product was lyophilized and the peptides were separated by gel filtration on Sephadex G-50 in 9% formic acid. ε-Aminosuccinyl groups are stable. The tryptic digest of the succinylated peptide was lyophilized and fractionated by gel filtration on Sephadex G-50 in 0.1 M NH_4HCO_3.

Peptide mixtures were subfractionated by gel filtration. The effluent was examined at 280 nm in a flow-through cell in an LKB Uvicord II column monitor.

Results

The amino acid sequence of reduced and S-pyridylethylated porcine trypsin was determined by sequenator analyses of the intact polypeptide chain and of eight peptide fragments. One of these fragments was found in the starting material, apparently generated by autolysis. Two more were generated by cleavage with cyanogen bromide. The remaining five fragments were generated by chemical or enzymatic cleavage at tryptophanyl, arginyl, or asparaginyl bonds. The sources of the fragments and their relation to the total sequence are illustrated diagrammatically in Figure 3, where the residue numbers refer to the final sequence (Figure 4) and the nine sequenator analyses are identified by the shaded segments, A–I. The amino acid residues are numbered according to those of bovine trypsinogen (Walsh and Neurath, 1964).

Although seven different digests or subdigests were examined in this study, in two cases only (α-trypsin and CNBr cleavage) were all the cleavage products recovered and characterized. These will be considered first since several of these fragments have served in turn as starting materials for subsequent fragmentation.

Identification of the Fragments α-N and α-C of α-Trypsin. Sequenator analysis of the active enzyme gave two sequences in a ratio of approximately 10:1 (Hermodson *et al.*, 1972a). The predominant sequence (Ile-Val-Gly) was identical with that of the single-chain form of bovine β-trypsin whereas the minor sequence (Ser-Ser-Gly) was identical with that of the internal sequence of the α form of bovine trypsin beginning with residue 132 (Schroeder and Shaw, 1968). Thus the presence of both a β form and an α form of porcine trypsin was indicated.

Gel filtration on Sephadex G-75 separated porcine PE-trypsin into three components (Figure 1). The first two components emerging from the column had identical amino-

[1] Abbreviations used are: BNPS-skatole, the compound formed by the reaction of *N*-bromosuccinimide with 2-(2-nitrophenylsulfenyl)-3-methylindole. The product is probably 2-(2-nitrophenylsulfenyl)-3-methyl-3-bromoindolenine; PE, *S*-pyridylethyl.

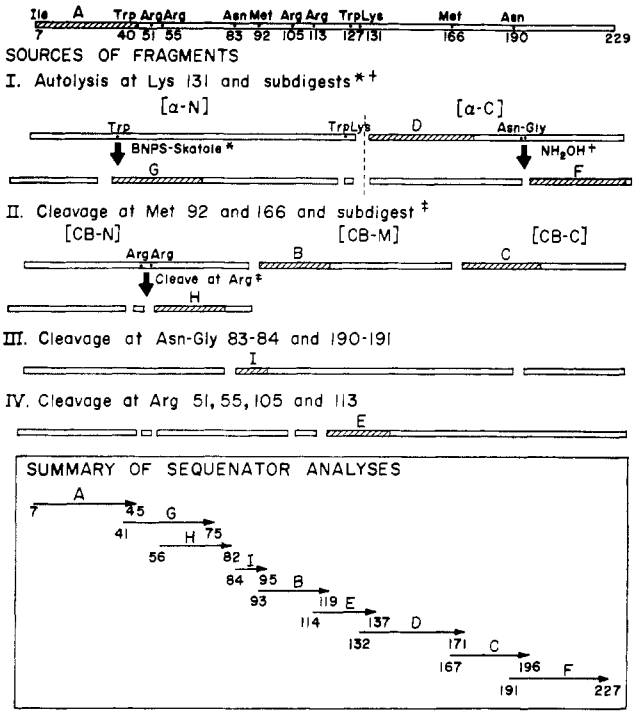

FIGURE 3: Diagrammatic summary of fragments generated from PE-β-trypsin for sequenator analysis. The top bar represents the polypeptide chain of PE-β-trypsin and the residues which are important for its fragmentation. The capital letters, A–I, identify the sequenator analyses in the order in which they were performed and described in the text. The shaded section of each horizontal bar indicates the segment of structure determined by that analysis. The overlaps among these segments are illustrated in the lower inset.

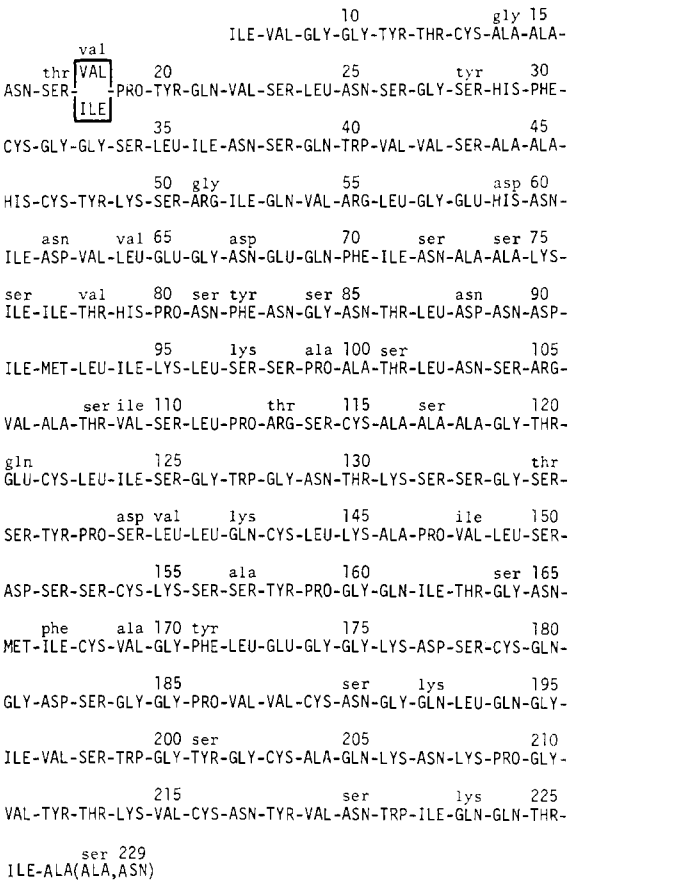

FIGURE 4: The amino acid sequence of porcine β-trypsin. The residue numbers correspond to those of bovine β-trypsin (Walsh and Neurath, 1964). The sequence of the bovine enzyme is that of Walsh and Neurath (1964) and Walsh (1970) except for residues 84–87 which were not sequenced and for which the data of Mikes *et al.* (1966) were used. The bovine sequence is identical except for those residues which are set above the porcine sequence. Polymorphism in the porcine sequence is evident in position 18 (Hermodson *et al.*, 1972a).

terminal sequences, Ile-Val-Gly, which corresponded to the predominant sequence of the intact enzyme. The amino terminal sequence of the third component, Ser-Ser-Gly, was the same as that of the minor component of the original mixture. It was concluded that the first component was the single-chain β form, and that the other two components corresponded to the two fragments of the α form, α-N, and α-C.

Identification of Cyanogen Bromide Fragments. Since amino acid analysis of β-trypsin yielded two methionyl residues (Table I), three unique fragments were expected from cleavage with cyanogen bromide. Gel filtration on Sephadex G-50 yielded three fractions (Figure 2), designated I, II, and III. For clarity, their characterization will be described in reverse order.

Sequenator analysis of fraction III gave a single amino acid sequence, Ile-Cys-Val. Since this fraction was devoid of homoserine (Table I), it must have originated from the carboxyl-terminal region of β-trypsin. Hence this fraction was designated fragment CB-C.

Analysis of fraction II revealed a mixture of two peptides. In order to separate them, the mixture was lyophilized, dissolved in a minimum volume of 0.05 M acetic acid, and adjusted to pH 5.5 by the addition of pyridine. One of the peptides precipitated at this pH and was removed by centrifugation, redissolved in 0.05 M acetic acid, and precipitated again with pyridine. The two supernatant solutions were combined and lyophilized. The precipitate was dissolved in 10% acetic acid and lyophilized. The precipitate and the supernatant fractions were then examined separately in the sequenator. The precipitate yielded a single sequence, Ile-Val-Gly, identical with the amino-terminal sequence of β-trypsin. This fragment must have originated from the amino-terminal region of porcine trypsin and was therefore designated fragment CB-N. The supernatant fraction yielded a single sequence Leu-Ile-Lys. Since the amino- and carboxyl-terminal regions of β-trypsin had already been accounted for by CB-N and CB-C, respectively, the fraction recovered from the supernatant solution must represent the middle portion of the sequence and was, therefore, designated fragment CB-M. These assignments were confirmed by amino acid analysis which showed that the amino acid composition of β-trypsin agreed with the sum of the residues of fragments CB-N, CB-C, and CB-M.

Fraction I represented approximately 10% of the weight of the original protein. Sequenator analysis revealed a single sequence, Ile-Val-Gly, but the amino acid composition of fraction I was identical with that of fraction II which contained equal amounts of CB-N and CB-M. Fraction I was thus assumed to be an "overlap" fragment consisting of CB-N and CB-M joined by an uncleaved homoseryl peptide bond.

Derivation of Partial Amino Acid Sequences (Analyses A–D). In addition to β-trypsin, five fragments were available for extended sequenator analyses. These included fragments α-N, α-C, and three cyanogen bromide fragments, CB-N, CB-C, and CB-M.

Since fragments α-N and CB-N were both derived from the amino-terminal region of β-trypsin, extended sequence information was obtained only from three fragments (α-C,

TABLE I: Amino Acid Composition[a] of Porcine Trypsin and of Fragments (CB) Generated by Cleavage with Cyanogen Bromide.

	PE-β-Trypsin		CB-N		CB-M		CB-C	
Amino Acid	Anal.	Sequence	Anal.[b]	Sequence	Anal.[b]	Sequence	Anal.[b]	Sequence
Asp	23.8	6	13.1	3	4.2	1	6.7	2
Asn		18		10		3		5
Thr	10	10	3.0	3	4.4	5	1.9	2
Ser	27	27	7.6	8	14.7	16	3.3	3
Glu	17.3	5	7.4	3	3.6	1	7.3	1
Gln		12		4		2		6
Pro	9.0	9	1.4	2	4.1	5	1.6	2
Gly	25.2	25	8.4	8	6.1	6	11.1	11
Ala	15.3	15	6.3	6	6.0	6	3.4	3
Val	16.2	16.4[c]	5.1	6.4[c]	2.9	3	5.5	7
Met	2.2	2	0.8[e]	1	0.8[e]	1		
Ile	15.5	15.6[c]	6.4	8.6[c]	2.7	3	3.0	4
Leu	16.6	16	5.0	5	8.6	9	2.2	2
Tyr	8.2	8	2.7	3	1.8	2	2.6	3
Phe	4.0	4	2.8	3			0.9	1
Lys	9.7	10	2.1	2	3.4	4	3.6	4
His	3.9	4	4.1	4				
Arg	3.9	4	1.8	2	1.8	2		
Cys/2[f]	11.7	12	3.3	3	3.8	4	4.7	5
Trp	4.0[d]	4	ND	1	ND	1	ND	2
Total		223		86		74		63

[a] Results expressed as residues per molecule. Triplicate samples were analyzed after 24 hr of hydrolysis in 6 N HCl and duplicate samples after 48, 72, and 96 hr. Values for threonine and serine are extrapolated to zero time. Values for valine and isoleucine are the average of 96-hr hydrolysates. All other values are averages of analyses. [b] Single 24-hr hydrolyses in 6 N HCl. Valine and isoleucine values are low since Val–Val, Ile–Ile, Ile–Val, and Val–Ile bonds are not completely cleaved in this time period. [c] See text for discussion of the polymorphism at residue 18. [d] Determined by hydrolysis in 4.2 N NaOH (Hugli and Moore, 1972). [e] Determined as homoserine and homoserine lactone. [f] Determined as *S*-pyridylethylcysteine.

CB-C, and CB-M) and β-trypsin itself. These analyses, A–D, are identified in Figure 3.

ANALYSIS A of β-trypsin identified the first 39 residues of the entire polypeptide chain. Two amino acids, valine and isoleucine, were obtained at position 18. The sum of their yields was equivalent to the yield expected for a single residue at that cycle, indicating the presence of polymorphism (Hermodson *et al.*, 1972a).

ANALYSIS B of fragment CB-M identified the first 27 residues of this fragment. From the compositional data of Table I and the knowledge that fragment CB-M follows fragment CB-N, these 27 residues could be tentatively identified as residues 93–119 in the sequence of β-trypsin. Analogously, the first 30 residues of fragment CB-C (analysis C) could be placed beginning with residue 167.

ANALYSIS D of fragment α-C yielded a sequence of 40 amino acid residues ending with the sequence Met-Ile-Cys-Val-Gly-Phe which overlapped the amino-terminal region of fragment CB-C (Ile-Cys-Val-Gly-Phe). Thus a continuous 65-residue sequence was established from Ser_{132} to Ile_{196} (Figure 4). Two residues (Gln_{161} and Thr_{163}) were only tentatively identified in analysis D. Their identities were confirmed by comparing the amino acid composition of a peptide obtained by cyanogen bromide cleavage of fragment α-C (comprising residues 132–166) with the composition calculated from the sequence of these residues (Table II).

Analyses A–D placed 131 residues of β-trypsin (59%) in three segments of the polypeptide chain. These correspond to residues 7–45, 93–119, and 132–196 (Figure 4). This information, together with the knowledge of the composition of the missing segments (Table I), determined the selection of additional cleavage sites to generate suitable fragments for the completion of the analysis (Figure 3). This experimental approach is most easily illustrated by the fragments required for the completion of the carboxyl-terminal sequence of the molecule (residues 119–229) which will be described first.

ANALYSIS E (RESIDUES 114–137). Amino acid analyses of the cyanogen bromide fragments indicated that fragments CB-N and CB-M each contained two arginyl residues whereas fragment CB-C contained none (Table I). Since the amino-terminal sequence of the protein did not include arginine up to residue 45 (analysis A), and the two arginine residues of fragment CB-M were tentatively placed at positions 105 and 113 (analysis B), the four arginine residues of the molecule must be distributed within a segment of not greater than 68 residues (residues 46–113). Since analysis B identified the sequence beyond the fourth arginine (113) as Arg_{113}-Ser-Cys-Ala-Ala-Ala-Gly (Figure 4), cleavage of β-trypsin at the arginyl residues should produce four relatively small fragments from the amino-terminal half of the molecule and one large carboxyl-terminal fragment of approximately 120 residues, beginning with the sequence Ser-Cys-Ala-Ala-Ala-Gly. Accordingly, PE-β-trypsin was citraconylated, digested with trypsin, and after removal of the citraconyl groups, fractionated on Sephadex G-50. Fractions of the first peak were pooled as indicated in Figure 5A and subjected to sequenator analysis. The first 6 of 24 identified residues agreed with the expected amino-terminal sequence of the carboxyl-terminal fragment. The last seven of the 24 residues (Lys-Ser-Ser-Gly-Ser-Ser-Tyr) provided an overlap of 6 residues with

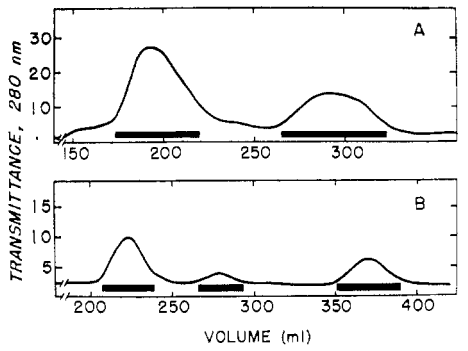

FIGURE 5: (A) Fractionation of a tryptic digest of citraconylated β-trypsin (50 mg) on a column (115 × 2.5 cm) of Sephadex G-50, equilibrated and eluted with 9% formic acid at 30 ml/hr. Prior to fractionation, the citraconyl groups were removed by incubation at pH 2.5 for 3 hr at room temperature. The salt peak (not shown) also contained two small peptides. Fractions were pooled as shown by the solid bars. (B) Fractionation of the products of cleavage of fragment α-C (50 mg) with hydroxylamine on a column (115 × 2.5 cm) of Sephadex G-50, equilibrated and eluted with 9% formic acid at 30 ml/hr. Fractions were pooled as indicated by the solid bars.

the amino terminus of fragment α-C (analysis D) and allowed the identification of the site of autolytic cleavage as a Lys–Ser bond. This analysis completed the sequence of fragment CB-M.

The second peak from the G-50 column (Figure 5A) contained approximately equimolar amounts of two peptides, and since attempts to separate them proved unsuccessful, they were not further characterized.

ANALYSIS F (RESIDUES 191–227). Sequenator analysis of fragment CB-C (analysis C, Figure 3) extended the sequence five residues beyond an Asn–Gly bond (Asn_{190}-Gly-Gln-Leu-Gln-Gly-Ile). Cleavage of fragment CB-C by hydroxylamine would have yielded two such fragments of similar size, which would have been difficult to separate. Thus fragment α-C was chosen to generate a fragment comprising residues 191–229. The products were separated by gel filtration on Sephadex G-50 (Figure 5B). The first peak fraction contained intact fragment α-C. The second peak fraction contained the amino-terminal cleavage product, and the third the desired carboxyl-terminal fragment beginning with the sequence Gly-Gln-Leu-Gln-Gly-Ile. Sequenator analysis of this fragment identified 37 residues ending with Trp-Ile-Gln-Gln-Thr-Ile-Ala_{227}.

Determination of the Carboxyl-Terminal Sequence of β-Trypsin. The carboxyl-terminal fragment generated from fragment α-C by treatment with hydroxylamine (analysis F) served as starting material for the identification of the carboxyl-terminal sequence of β-trypsin. This fragment (6 mg) was digested with 0.3 mg of "TLCK-chymotrypsin" (Worthington) for 14 hr at 37° and fractionated by chromatography on Dowex 1-X2 under conditions identical with those published previously (Hermodson *et al.*, 1972b). One of the fractions was an octapeptide containing the six carboxyl-terminal residues of analysis F plus one residue each of Ala and Asx (Table II). Since the sequence of the first six residues has already been established, the last two amino acid residues must be (Ala, Asx).

The placement of this octapeptide at the carboxyl terminus of the protein was confirmed by digestion of β-trypsin with yeast carboxypeptidase C as previously described (Hermodson *et al.*, 1972b). This exopeptidase released sequentially

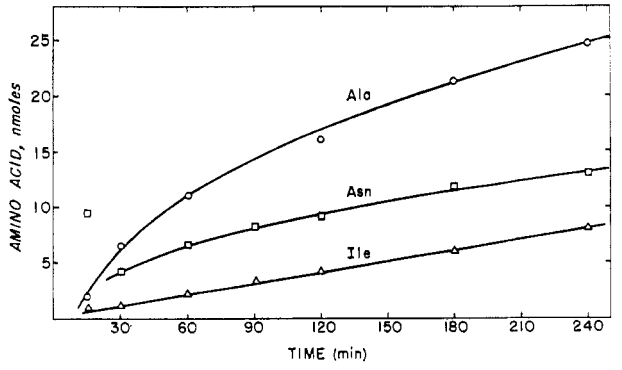

FIGURE 6: Progressive liberation of carboxyl-terminal amino acids from PE-β-trypsin by yeast carboxypeptidase C: alanine (○), asparagine (□), isoleucine (△). The 15-min value for asparagine was anomalously high.

TABLE II: Composition of Peptides Derived from Fragments of PE-β-Trypsin.[a]

Amino Acid	Tryptic Cleavage of CB-N: Residues 56–75[c]	Tryptic Cleavage of CB-N: Residues 76–82[c]	CNBr Cleavage of α-C: Residues 132–166[d]	Chymotryptic Peptide[b]: Residues 222–229[e]
Asx	4.3 (4)	5.9 (6)	2.0 (2)	1.2 (1)
Thr	0.2 (0)	1.8 (2)	1.0 (1)	1.0 (1)
Ser	0.2 (0)	0	8.7 (10)	0.1 (0)
Glx	3.7 (4)	0.2 (0)	2.0 (2)	1.9 (2)
Pro	0	1.0 (1)	2.5 (3)	0
Gly	2.1 (2)	1.1 (1)	2.9 (3)	0.3 (0)
Ala	2.0 (2)	0.1 (0)	1.0 (1)	2.0 (2)
Val	1.0 (1)	0.1 (0)	0.9 (1)	0
Met	0	1.0 (1)	(1)[f]	0
Ile	2.0 (2)	1.8 (3)	0.9 (1)	1.8 (2)
Leu	1.9 (2)	1.0 (1)	3.7 (4)	0
Tyr	0	0	1.7 (2)	0
Phe	1.0 (1)	1.1 (1)	0	0
Lys	1.0 (1)	0	2.0 (2)	0.2 (0)
His	0.9 (1)	0.9 (1)	0	0
Arg	0	0	0	0
Cys/2	0	0	2.0 (2)	0
Trp	0	0	ND (0)	ND (0)

[a] Amino acid analyses were performed after 24 hr of hydrolysis in 6 N HCl at 110°. Cysteine was determined as *S*-pyridylethylcysteine. The data are expressed as residues per molecule. Numbers in parentheses are the residues found in the sequence. [b] A fragment comprising residues 191–229 was isolated as described in analysis F (see the text and Figure 3). [c] These peptides were isolated by paper electrophoresis (Whatman paper No. 3, pH 6.5, 2000 V, 90 min) of a tryptic digest of fragment CB-N. $Methionine_{92}$ was determined as homoserine + homoserine lactone. The peptide bond Ile_{76}–Ile_{77} was not completely hydrolyzed and hence the yield of isoleucine was low. [d] This peptide was isolated by treatment of fragment α-C with cyanogen bromide, followed by gel filtration on Sephadex G-50. [e] This peptide (comprising residues 222–229) was isolated by chromatography on Dowex 1-X2 of a chymotryptic digest of the carboxyl-terminal fragment generated by treatment of PE-β-trypsin with hydroxylamine. [f] Homoserine and homoserine lactone were present in the hydrolysate but were not quantitated since they did not fully separate from glutamic acid and lysine, respectively.

alanine, asparagine, and isoleucine (Figure 6). Although the yield of alanine exceeded that of asparagine throughout the analysis, the known position of Ala_{227} three residues from the end allows for either of the following sequences: Ile_{226}-Ala-Asn-Ala-OH or Ile_{226}-Ala-Ala-Asn-OH. The choice between these alternatives cannot be made on the basis of the present data.

ANALYSIS G (RESIDUES 41–75). The completion of the structural analysis of β-trypsin required the identification of the amino acid sequence of the carboxyl-terminal half of fragment CB-N (residues 46–92, of Figure 4). Analysis A of β-trypsin defined the amino-terminal sequence through residue 45 (Trp-Val-Val-Ser-Ala-Ala_{45}). Cleavage of the Trp–Val bond with BNPS-skatole appeared to provide a suitable initiation point for sequenator analysis. The next tryptophanyl residue in the sequence was known to be five residues from the carboxyl terminus of fragment α-N (residue 127). Cleavage of the tryptophanyl bonds at residues 40 and 127 would yield a large fragment (residues 41–127) which should be readily separated from the smaller peptides, residues 7–40 and 128–131. Hence fragment α-N was chosen for fragmentation with BNPS-skatole. After treatment with this reagent, the reaction mixture was fractionated on Sephadex G-50 (Figure 7A). Since cleavage at tryptophan was incomplete, the major peak fraction of Figure 7A contained peptides having the amino-terminal sequence of fragment α-N. This fraction was discarded. Sequenator analysis of the fraction corresponding to the shoulder of the main peak (solid bar in Figure 7A) indicated the presence of a predominant sequence of Val-Val-Ser-Ala-Ala. Although a minor sequence, Ile-Val-Gly-Gly-Tyr, was obtained in a 10% yield compared to the major one, it could be distinguished from the former by its identity with the amino-terminal sequence of α-N. This analysis provided the sequence of residues 41–75 and extended the amino-terminal sequence of β-trypsin through the arginine residues 51 and 55.

ANALYSIS H (RESIDUES 56–82). Tryptic cleavage of arginyl bonds in succinylated fragment CB-N followed by gel filtration on Sephadex G-50 (Figure 7B) generated a fraction with the expected amino-terminal sequence of a subfragment comprising residues 56–92 (Leu-Gly-Glu-His-Asn, Figure 4). Extended analysis terminated abruptly with Phe_{82}. Lys_{75}, which was tentatively identified in analysis G, did not give a recognizable product in this analysis because the peptide was succinylated. It is probable that this fragment (56–92) was contaminated with a peptide containing residues 7–51 which could not be detected in the sequenator because its α-amino group (of isoleucine 7) was blocked by succinylation. A small contamination by the tetrapeptide Ile-Gln-Val-Arg (residues 52–55) was detected but caused no ambiguity of interpretation since its sequence (proven in analysis G) could be subtracted from the major sequence.

ANALYSIS I (RESIDUES 84–95). The only unidentified sequence was a segment between Phe_{82} and Met_{92}. This sequence was elucidated by analysis of a fragment obtained by cleavage of an Asn–Gly bond in β-trypsin with hydroxylamine (Bornstein, 1969). Evidence for the existence of two Asn–Gly bonds was obtained from gel filtration of hydroxylamine treated PE-β-trypsin which indicated the presence of at least five components.[2] One of these, which emerged last, was pure

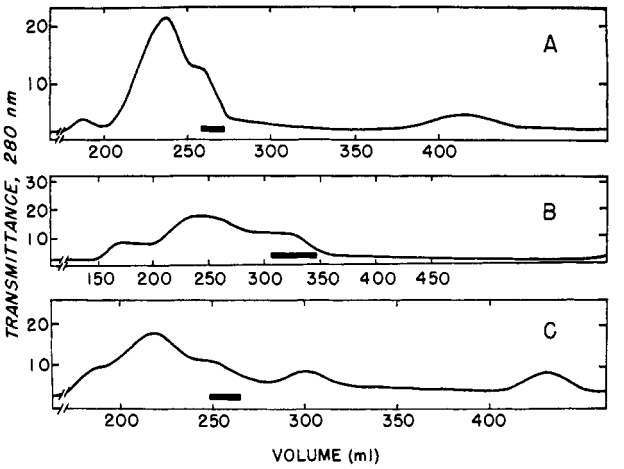

FIGURE 7: (A) Fractionation of the products of cleavage of fragment α-C (50 mg) with BNPS-skatole on a column (115 × 2.5 cm) of Sephadex G-50, equilibrated and eluted with 9% formic acid at 30 ml/hr. (B) Fractionation of a tryptic digest of succinylated fragment CB-N (50 mg) on a column (90 × 2.5 cm) of Sephadex G-50, equilibrated and eluted with 0.1 M NH_4HCO_3 at 40 ml/hr. (C) Fractionation of the products of cleavage of PE-β-trypsin (100 mg) with hydroxylamine, on a column (115 × 2.5 cm) of Sephadex G-50, equilibrated and eluted with 9% formic acid at 30 ml/hr. Each figure is a trace of the chart from the column monitor. The fractions were pooled as indicated by the horizontal bars.

and identical with the carboxyl-terminal subfragment (residues 191–229) obtained by cleavage of fragment α-C with hydroxylamine (see Figure 5B and analysis F, Figure 3). Sequenator analysis of the fraction indicated by the bar in Figure 7C gave the sequence Gly-Asn-Thr-Leu-Asp-Asn-Asp-Ile-Met_{92}-Leu-Ile-Lys which provided eight residues leading to Met_{92} as well as a tripeptide sequence which overlaps fragment CB-M. The amino-terminal sequence of porcine trypsin was again observed in this analysis as a contaminant; the desired fragment predominated by a ratio of 5:2 over the contaminant, and since the sequence of the contaminant was known, it did not hinder the analysis. This analysis completed the sequence of fragment CB-N except for the residue(s) between Phe_{82} and the amino-terminal glycine of analysis I.

Confirmation of Lys_{75} and of Residues 82–86. Since the identification of residue 75 as lysine was equivocal (see above), additional evidence was sought. Fragment CB-N contained two lysyl and two arginyl residues (Table I). When CB-N (30 mg) was digested with 0.5 mg of "TPCK-trypsin" (Worthington) for 3 hr at 37° (pH 8) and fractionated on a 2.5 × 115 cm column of Sephadex G-50 in 9% formic acid, only one peptide could be detected by monitoring the effluent at 280 nm. This peptide corresponded to a fragment comprising residues 7–49. Reaction with ninhydrin after alkaline hydrolysis gave evidence of another broad peak fraction, just prior to the salt peak, which did not absorb at 280 nm and hence must have been devoid of tryptophan and tyrosine. This fraction was pooled, lyophilized, and dissolved in 10% acetic acid. Paper electrophoresis at pH 6.5 indicated the presence of two acidic peptides which were eluted and analyzed. The composition of the more acidic of the two peptides corresponded to that of the sequence of residues 56–75 (Table II) and thus confirmed Lys_{75}.

The composition of the less acidic peptide corresponded to that of the sequence of residues 76–82 plus residues 84–92 plus Asx. Since the specificity of hydroxylamine cleavage required that the residue preceding Gly_{84} be asparagine, the

[2] Since cleavage by hydroxylamine is rarely complete, cleavage of a single Asn–Gly bond would yield 3 components (intact chain plus 2 fragments) and cleavage of two Asn–Gly bonds would yield 6 components (intact chain, 3 fragments and 2 overlapping fragments).

composition of this tryptic peptide fully accounted for the sequence of residues 82–86 (Phe-Asn-Gly-Asn-Thr) and completed the sequence of fragment CB-N. The sequence deduced from these analyses is in full accord with the composition of fragment CB-N determined by amino acid analysis (Table I).

Discussion

The present work constitutes a determination of the amino acid sequence of a protein primarily by sequenator analysis. With the exceptions of the carboxyl-terminal dipeptide sequence, which remains to be resolved, and Asn_{83}, all 223 amino acid residues of S-pyridylethylated porcine trypsin could be placed by sequenator analysis of suitably chosen peptide fragments. A total of eight fragments was required for sequenator analysis, and four more peptides for confirming the identification of six residues.

In contrast, sequence analysis of 316 residues of thermolysin (which employed only in part sequenator analysis) required 173 fragments (Titani *et al.*, 1972). The determination of the sequence of bovine carboxypeptidase A (307 residues) involved 199 fragments (Bradshaw *et al.*, 1969), whereas the original sequence analysis of 51 residues of insulin by Sanger and coworkers required 152 peptide fragments (Sanger and Thompson, 1953). These statistics testify to the enormous savings in time and effort of the sequenator methods as compared to conventional methods of sequence analysis.

The reductive cleavage of cystinyl residues and their conversion to *S*-pyridylethylcysteine residues have circumvented problems previously encountered with other cysteinyl derivatives. Whereas *S*-aminoethylcysteine is not readily identifiable by gas–liquid chromatography or by thin-layer chromatography, and *S*-carboxymethylcysteine cannot be differentiated from serine by gas–liquid chromatography, *S*-pyridylethylcysteine is readily identified by thin-layer chromatography (Hermodson *et al.*, 1972a).

The methods employed herein were not free of problems, however. For instance, identification of the Asn_{83}–Gly_{84} sequence involved subfractionation of a tryptic digest of a cyanogen bromide fragment to confirm the evidence based on the specificity of cleavage by hydroxylamine. The identification of the Ala_{227}–Ala_{228} sequence is circumstantial at best and based on homology with bovine trypsin. Placement of disulfide bonds was not attempted in this work. All twelve half-cystinyl residues are in the same positions as in the bovine enzyme and hence were presumed to form identical pairs of disulfide bonds. The actual pairing should be resolved by current investigations of the crystal structure of porcine enzyme by Dr. David Blow (personal communication).

Partial sequences of porcine trypsin have been previously published by Travis and Liener (1965) and by Smith and Liener (1967). These sequences agree with the present data with the exception of the following residues (those of the previous authors are given in parentheses): Ile_{61}(Asx); Asp_{62}(Ile); Asp_{177}(Asn); Asp-Ser-Gly_{184}(Gly-Asp-Ser); Asn-Gly-Gln_{192}(Gly-Gln-Gln).).

The sequence of the activation peptide of porcine trypsinogen has been previously reported as Phe-Pro-Thr-$[Asp]_4$-Lys (Charles *et al.*, 1963). Thus the zymogen comprises 231 amino acid residues as compard to 229 in bovine trypsinogen. Charles *et al.* (1963) proposed the carboxyl-terminal sequence Thr-(Gln,Ile)-Ala-Asn on the basis of digestion with carboxypeptidase A. Our sequence, Ile_{222}-Gln-Gln-Thr-Ile-Ala-(Ala,Asn), is quite consistent with their data, given the difficuly of placing dipeptides such as Ala-Ala and Gln-Gln by analysis of carboxypeptidase digests.

Comparison of the sequences of the porcine and bovine enzymes reveals no gaps and 41 replacements among 223 residues, equivalent to 82% identity. Assuming that these two species have diverged from each other 60 million years ago and applying the equation of Dickerson (1971), it follows that m, the number of replacements per 100 residues, is 20.3 and the unit evolutionary period (UEP) is 2.96 million years.

Of the 41 amino acid substitutions, 36 are on the surface of the molecule, as defined for serine proteases by Shotton and Hartley (1970) and by Stroud *et al.* (1971), and 5 in the interior. These interior residues are as follows (for each residue number, the residues are given in the order of bovine trypsin, porcine trypsin, and bovine chymotrypsin; chymotrypsin residue numbers are in parentheses): residue 18—Val, Val/Ile, Trp (27); residue 109—Ile, Val, Val (121); residue 148—Ile, Val, Leu (162); residue 167—Phe, Ile, Ile (181); residue 169—Ala, Val, Ala (183). All these substitutions are of the conservative type and involve hydrophobic amino acid residues.

The preservation of function requires that the amino acid residues of the active site remain invariant (Dickerson and Geis, 1969). This is in fact the case when bovine and porcine trypsin are compared to each other. The components of the active site include the following residues (Stroud *et al.*, 1971): His_{46}, Asp_{90}, Ser_{183}, and Ser_{198} (the "charge-relay" system); Ile_7 (forming an ion pair with Asp_{182}); Ser_{23}, His_{29}, Gly_{181}, and Gly_{184}. The following residues probably make direct contacts with substrates of trypsin and chymotrypsin (Hartley, 1968; Segal *et al.*, 1971) and are likewise invariant in bovine and porcine trypsins: Asp_{177}, Ser_{178}, Cys_{179}, Gln_{180}, Ser_{198}, Trp_{199}, Gly_{200}, Gly_{210}, and Leu_{87}. The first four of these residues, together with Gly_{200} and Gly_{210}, are part of the substrate binding pocket; residues 198–200 form a hydrogen-bond structure which is antiparallel to that of the substrate. Leu_{87} is conceivably a P_2 site in chymotrypsin and trypsin as well. Recently Robertus *et al.* (1972) have proposed that the transition state in chymotryptic catalysis is stabilized by hydrogen bonds between the substrate and Gly_{193}, Ser_{195}, and Ser_{214} in chymotrypsin. The corresponding residues in bovine trypsin (Gly_{181}, Ser_{183}, and Ser_{198}) are also found in porcine trypsin.

On the basis of the three-dimensional structure of bovine trypsin Stroud *et al.* (1971) have proposed two alternative calcium binding sites: either Asp_{59} (residue 71 according to the chymotrypsinogen sequence) and Asp_{139} (153), or Asp_{59} and Glu_{65} (78). The latter chelate structure might stabilize the top loop of the enzyme. Since in porcine trypsin neither residue 59 (histidine) nor residue 139 (serine) contains a carboxyl group, neither calcium binding site could exist. Indeed, there is little evidence that porcine trypsin, which is more resistant to autolysis than the bovine enzyme, binds calcium (Vithayathil *et al.*, 1961; Buck *et al.*, 1962; Lazdunski and Delaage, 1965). A strict comparison of the two enzymes with regard to their interaction with calcium or other cations may therefore be illusory.

Acknowledgments

The authors are indebted to Dr. K. Titani for many helpful suggestions, to R. R. Granberg for the amino acid analyses, and to Leyardia Black for the purification of the porcine trypsin by affinity chromatography.

References

Bornstein, P. (1969), *Biochem. Biophys. Res. Commun. 36*, 957.
Bradshaw, R. A., Ericsson, L. H., Walsh, K. A., and Neurath, H. (1969), *Proc. Nat. Acad. Sci. U. S. 63*, 1389.
Buck, F. F., Vithayathil, A. J., Bier, M., and Nord, F. F. (1962), *Arch. Biochem. Biophys. 97*, 417.
Cavins, J. F., and Friedman, M. (1970), *Anal. Biochem. 35*, 489.
Charles, M., Rovery, M., Guidoni, A., and Desnuelle, P. (1963), *Biochim. Biophys. Acta 69*, 115.
Dickerson, R. E. (1971), *J. Mol. Evol. 1*, 26.
Dickerson, R. E., and Geis, I. (1969), The Structure and Action of Proteins, New York, N. Y., Harper & Row.
Edman, P., and Begg, G. (1967), *Eur. J. Biochem. 1*, 80.
Gross, E. (1967), *Methods Enzymol. 11*, 238.
Hartley, B. S. (1968), *Phil. Trans. Roy. Soc. London, Ser. B, 257*, 77.
Hermodson, M. A., Ericsson, L. H., Titani, K., Neurath, H., and Walsh, K. A. (1972a), *Biochemistry 11*, 4493.
Hermodson, M. A., Kuhn, R. W., Walsh, K. A., Neurath, H., Eriksen, N., and Benditt, E. P. (1972b), *Biochemistry 11*, 2934.
Hugli, T. E., and Moore, S. (1972), *J. Biol. Chem. 247*, 2828.
Lazdunski, M., and Delaage, M. (1965), *Biochim. Biophys. Acta 105*, 541.
Mikes, O., Holeysovsky, V., Tomasek, V., and Sorm, F. (1966), *Biochem. Biophys. Res. Commun. 24*, 346.
Omenn, G. S., Fontana, A., and Anfinsen, C. B., Jr. (1970), *J. Biol. Chem. 245*, 1895.
Robertus, J. D., Kraut, J., Alden, R. A., and Birktoft, J. H. (1972), *Biochemistry 11*, 4293.
Robinson, N. C., Tye, R. W., Neurath, H., and Walsh, K. A. (1971), *Biochemistry 10*, 2743.
Sanger, F., and Thompson, E. O. P. (1953), *Biochem. J. 53*, 353.
Schroeder, D. D., and Shaw, E. (1968), *J. Biol. Chem. 243*, 2943.
Segal, D. M., Powers, J. C., Cohen, G. H., Davies, D. R., and Wilcox, P. E. (1971), *Biochemistry 10*, 3728.
Shotton, D. M., and Hartley, B. S. (1970), *Nature (London) 225*, 802.
Smith, R. A., and Liener, I. E. (1967), *J. Biol. Chem. 242*, 4033.
Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G., and Ballantyne, D. L. (1971), *Biochemistry 10*, 4912.
Stroud, R. M., Kay, L. M., and Dickerson, R. E. (1971), *Cold Spring Harbor Symp. Quant. Biol. 36*, 125.
Titani, K., Hermodson, M. A., Ericcson, L. H., Walsh, K. A., and Neurath, H. (1972), *Nature (London), New Biol. 238*, 35.
Travis, J., and Liener, I. E. (1965), *J. Biol. Chem. 240*, 1967.
Vithayathil, A. J., Buck, F., Bier, M., and Nord, F. F. (1961), *Arch. Biochem. Biophys. 92*, 532.
Walsh, K. A. (1970), *Methods Enzymol. 19*, 41.
Walsh, K. A., and Neurath, H. (1964), *Proc. Nat. Acad. Sci. U. S. 52*, 884.

Isolation and Characterization of the Cyanogen Bromide Peptides from the α1(II) Chain of Bovine and Human Cartilage Collagen†

Edward J. Miller* and Linda G. Lunde

ABSTRACT: Insoluble collagen prepared from three mammalian cartilages (bovine nasal septum, bovine articular cartilage, and human articular cartilage) was cleaved with cyanogen bromide (CNBr). The solubilized peptides which in each instance accounted for over 95% of the tissue collagen were isolated by molecular sieve and ion-exchange chromatography and characterized with respect to molecular weight and amino acid composition. Twelve and ten peptides were isolated as CNBr cleavage products of the bovine and human cartilage collagens, respectively. For each species, the isolated peptides account for virtually the entire length of an α1(II) chain. The data indicate that identical peptides are derived from both bovine cartilages although peptides from articular cartilage collagen exhibit slightly less hydroxylation of lysyl residues. The peptides from bovine cartilage collagen are very similar to those from human cartilage collagen and exhibit only small differences in amino acid composition. Furthermore, all of the peptides characterized in the present study are clearly recognizable as homologs of those previously isolated from chick α1(II), demonstrating a close structural homology between the α1(II) chains in the cartilages of various species.

Previous studies on the collagen in several cartilaginous tissues have indicated that the cartilage collagen molecule is comprised of three identical α chains. The latter have been designed α1(II) chains to indicate their derivation from a genetic locus distinct from the loci directing the synthesis of α1(I) and α2 chains found in the collagen of bone, skin, tendon, and other tissues (Miller and Matukas, 1969; Trelstad *et al.*, 1970; Miller, 1971a; Miller *et al.*, 1971; Strawich and Nimni, 1971; Toole *et al.*, 1972). Thus, the cartilage collagen molecule with the chain composition $[\alpha 1(II)]_3$ represents a striking example of tissue specificity in the biosynthesis of this macromolecule.

† From the Department of Biochemistry and Institute of Dental Research, University of Alabama Medical Center, Birmingham, Alabama 35294. *Received March 29, 1973.* This work was supported by U. S. Public Health Service Grant DE-02670 from the National Institute of Dental Research.