Amino Acid Sequence of *Streptomyces griseus* Trypsin. Cyanogen Bromide Fragments and Complete Sequence†

Robert W. Olafson,‡ Lubomir Jurášek, Michael R. Carpenter, and Lawrence B. Smillie*

ABSTRACT: Information compiled by automatic Edman degradation of *Streptomyces griseus* trypsin coupled with previous data has permitted the assignment of the first 36 residues at the NH_2 terminus of the protein. Cyanogen bromide cleavage at the three methionine residues followed or preceded by reduction and aminoethylation resulted in the production of four fragments, Cn1 to Cn4, which were separated by gel filtration on Sephadex G-50 or G-75. Fragments Cn4 (15 residues) and Cn3 (5 residues) were shown to be derived from the NH_2 terminus of the protein while Cn2 (47 residues and devoid of homoserine) was from the COOH terminus. The arrangement of the fragments was thus Cn4–Cn3–Cn1–Cn2. Automatic Edman degradation in the sequenator coupled with peptides derived from α-lytic protease and chymotryptic digestion and from the peptic and tryptic peptides previously elucidated have permitted the sequence determination of fragments Cn1 and Cn2 and therefore of the whole protein. These studies show that extensive regions of identity or similarity exist between *Streptomyces griseus* trypsin and bovine trypsin. These include the NH_2-terminal four residues, the sequences near histidine-57 (chymotrypsinogen A numbering system), aspartic acid-102, aspartic acid-189, and serine-195, the regions of the three disulfide bridges, and the COOH-terminal end (residues 225–229) of the proteins. When aligned to maximize homology the identity of residues is 34%. This identity is increased to 54% when only those residues classified as internal by Stroud *et al.* (Stroud, R. M., Kay, L. M., and Dickerson, R. E. (1971), *Cold Spring Harbor Symp. Quant. Biol. 36*, 125) are considered. These results indicate that the folding of the polypeptide chains of the two enzymes is very similar and are in agreement with the very similar enzymic, chemical, and physical properties of the two enzymes.

Trypsin, isolated from Pronase, a commercial product derived from culture filtrates of *Streptomyces griseus* has been shown to be remarkably similar in its enzymic, physical, and chemical properties to bovine trypsin (Olafson and Smillie, 1975). To determine the degree of similarity of its primary structure with the bovine enzyme and with other proteases from mammalian and bacterial sources, studies directed toward the elucidation of its complete amino acid sequence were initiated (Jurášek *et al.*, 1969). Toward this end, we have previously described the isolation and sequence determination of peptic peptides accounting for some 95% of the molecule (Jurášek and Smillie, 1973). These peptides included the three disulfide bridges where some 50% identity with bovine trypsin was observed. This identity included not only the active serine-195 and histidine-57 residues but also aspartic acid-189 known to play an important role in binding of positively charged trypsin substrates. A subsequent report has described the extension of this sequence information by the isolation and elucidation of the soluble tryptic peptides (Jurášek and Smillie, 1974), accounting for some 65% of the molecule. From chymotryptic and α-lytic protease digests of the insoluble tryptic peptides an additional 20% was recovered in unique sequences. In the present paper we describe the automatic Edman degradation of the intact protein and the isolation and characterization of the four cyanogen bromide fragments produced by cleavage at its three methionine residues. Sequence analysis of these by automatic and manual methods coupled with previous sequence information has allowed the complete primary structure of the enzyme to be assigned.

Materials and Methods

Streptomyces griseus trypsin (S.G. trypsin[1]) was isolated from Pronase (B grade; Calbiochem, Los Angeles, Calif.) by ion-exchange chromatography on CM-Sephadex C-50 as previously described (Jurášek *et al.*, 1971). A further purification was usually performed by rechromatography on either Bio-Rex 70 (Jurášek *et al.*, 1969) or on SE-Sephadex C-50 (Olafson and Smillie, 1975).

Cleavage of methionyl bonds with cyanogen bromide was carried out either on the native enzyme or after reduction and aminoethylation of the disulfide bridges. In the latter case the enzyme was initially inactivated with diisopropyl phosphorofluoridate, reduced with mercaptoethanol, and aminoethylated with ethylenimine as previously described (Jurášek and Smillie, 1974). Cleavage with cyanogen bromide (Gross, 1967) was effected in 70% formic acid (20 mg of protein/ml) by adding 3.3 mg of cyanogen bromide per mg of protein. After incubation at room temperature (in the dark) for 20 hr the solution was diluted with 10 vol of water and lyophilized. Protein which had not been reduced and aminoethylated before treatment with cyanogen bromide was then converted to the S-β-aminoethylated derivative (Raftery and Cole, 1966). After lyophilization the reduced and aminoethylated fragments were fractionated on either

† From the Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada. *Received May 22, 1974.* Taken, in part, from a dissertation submitted by R.W.O. to the Faculty of Graduate Studies of the University of Alberta in partial fulfillment of the requirements for the degree of Doctor of Philosophy. This investigation was supported by a grant from the Medical Research Council of Canada. R.W.O. was supported by a Medical Research Council of Canada Studentship (1969–1972).

‡Present address: Institute of Oceanography, University of British Columbia, Vancouver 8, B.C.

[1] Abbreviations used are: S.G. trypsin, *Streptomyces griseus* trypsin.

Table I: Automatic Sequence Analysis of S.G. Trypsin.

Edman Cycle	Residue[a] (Yield in nmol)	Edman Cycle	Residue[a] (Yield in nmol)	Edman Cycle	Residue[a] (Yield in nmol)
1	Val (99)	14	Phe (36)	27	Tyr (2)
2	Val (100)	15	Met (50)[b]	28	Ala (8)
3	Gly (87)	16	Val (45)	29	Gln (12)
4	Gly (142)	17	Arg (21)	30	Asp (5)
5	Thr (74)	18	Leu (47)	31	Ile (3)
6	Arg (50)	19	Ser (25)[b]	32	Val (4)
7	Ala (95)	20	Met (35)[b]	33	Leu (4)
8	Ala (56)	21	Gly (20)	34	Thr (3)
9	Gln (79)	22		35	Ala (3)
10	Gly (98)	23	Gly (23)	36	Ala (2)
11	Glu (54)	24	Gly (26)	37	His (1)
12	Phe (51)	25	Ala (16)		
13	Pro (22)	26	Leu (14)		

[a] Residues analyzed by amino acid analysis after HI hydrolysis (Smithies *et al.*, 1971) of thiazolinones derived from 8 mg (350 nmol) of S.G. trypsin. [b] Determined by gas chromatography (see Materials and Methods).

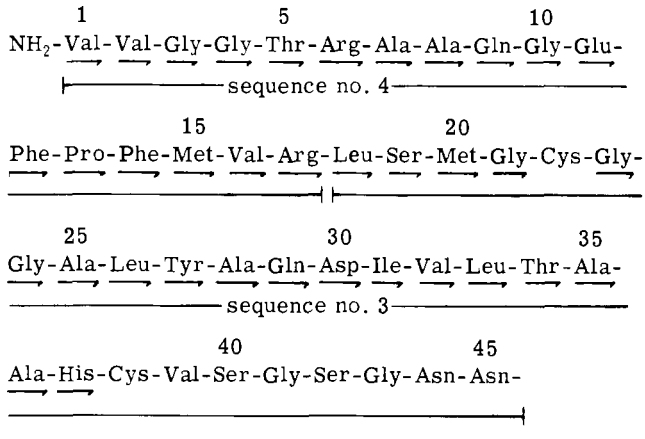

FIGURE 1: NH_2-terminal sequence of S.G. trypsin as elucidated by automatic sequencer analysis and from structures of unique sequences 4 and 3 previously elucidated (Jurášek and Smillie, 1974); → indicates indentification on sequenator.

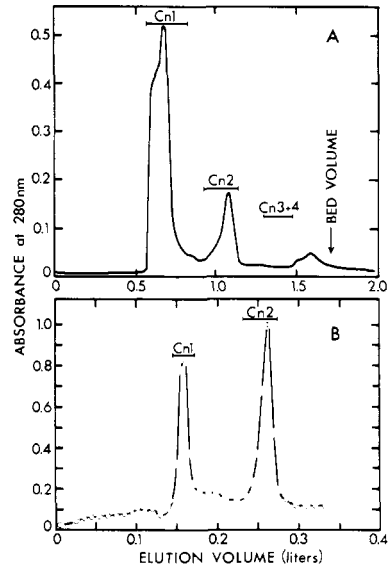

FIGURE 2: (A) Separation of cyanogen bromide fragments of reduced and S-β-aminoethylated S.G. trypsin (100 mg) on a column (5 × 88 cm) of Sephadex G-50 fine in 0.5 M acetic acid. (B) Separation of fragments of S.G. trypsin (25 mg) produced by cleavage with cyanogen bromide followed by reduction and alkylation with ethylenimine. The column (2.0 × 100 cm) of Sephadex G-75 superfine was equilibrated with 0.65 M acetic acid containing 8 M urea.

Sephadex G-50 in 0.5 M acetic acid or on Sephadex G-75 in 0.65 M acetic acid–8 M urea (Figure 2A and B). Fractions Cn1 and Cn2 were located from their absorbance at 280 nm and fractions Cn3 and Cn4 by spotting small aliquots (50 μl) on paper and subjecting these to high-voltage electrophoresis at pH 6.5. Fractions were pooled as indicated and the peptides of fractions Cn3 and Cn4 further purified by high-voltage electrophoresis at pH 6.5 and 1.8.

Maleylation was performed by the method of Butler *et al.* (1969). Performic acid oxidized S.G. trypsin (2.7 μmol) prepared by the method of Hirs (1967) was dispersed in 0.2 M borate buffer (pH 9.0) (0.65 mg of protein/ml) and four 1.6-ml aliquots of 1 M maleic anhydride in 1,4-dioxane added at 5-min intervals. The pH was maintained constant in the pH-Stat by titration with 1 N NaOH. Maleylation of fragment Cn1 (12.5 μmol) was carried out in 320 ml of 25 mM sodium borate buffer (pH 9.0)–50 mM mercaptoethanol–8 M urea with the addition of 1.6-ml quantities of 1 M maleic anhydride at 5-min intervals for 50 min. The maleylated products were then exhaustively dialyzed against 50 mM *N*-ethylmorpholine acetate buffer (pH 7.5) at room temperature.

α-Lytic protease digestion was performed on fragments Cn1 and Cn2. To Cn1 (18.8 μmol), dissolved in 12 ml of 50% acetic acid, was added 60 ml of water and 3 ml of *N*-ethylmorpholine. The pH was quickly adjusted to 8.2 with 6 N NH_4OH and digestion initiated immediately by the addition of 3.6 mg of the enzyme (prepared in the laboratory of Dr. D. R. Whitaker). After 40 min at 31° an additional 3.6 mg of enzyme was added. The digestion mixture, which was opalescent, became clear after 128 min when the reaction was terminated by lyophilization. Fragment Cn2 (0.6 μmol) was digested as a suspension with 0.4 mg of α-lytic protease in 2 ml of 0.05 M *N*-ethylmorpholine acetate buffer (pH 8.0). After 132 min at 25° the suspension clarified and the digestion was terminated by lyophilization.

Chymotryptic Digestions. Maleylated fragment Cn1 (12.5 μmol) in 500 ml of 50 mM *N*-ethylmorpholine acetate

Table II: Amino Acid Compositions[a] of Peptides of Cyanogen Bromide Fractions Cn3 and Cn4.

	Cn3A	Cn3B	Cn4A1	Cn4A2	Cn4B1	Cn4B2
Arginine	0.99 (1)	0.75 (1)	0.86 (1)	1.10 (1)	0.61 (1)	0.72 (1)
Threonine			0.94 (1)	0.90 (1)	1.13 (1)	0.98 (1)
Serine	0.92 (1)	0.99 (1)				
Glutamic acid			1.94 (2)	1.97 (2)	2.57[d] (2)	1.96 (2)
Proline			0.93 (1)	0.84 (1)	1.41 (1)	1.00 (1)
Glycine			2.97 (3)	2.94 (3)	3.00 (3)	2.93 (3)
Alanine			2.00 (2)	2.03 (2)	1.85 (2)	1.95 (2)
Valine	0.99 (1)	1.00 (1)	0.43[c] (1)	1.68[c] (2)	0.59[c] (1)	1.20[c] (2)
Leucine	1.00 (1)	1.08 (1)				
Phenylalanine			1.94 (2)	2.00 (2)	1.47 (2)	2.00 (2)
Homoserine[b]	0.95 (1)	0.66 (1)	0.86 (1)	0.95 (1)	*d* (1)	0.77 (1)
N-Aminoethylvaline			0.44 (1)		0.55 (1)	
NH_2 terminus	Val	Val	AEV[e]	Val		Val
pH 6.5 mobility[f]	+0.60	+0.32	+0.19	+0.19	0	0
pH 6.5 charge[g]	+2	+1	+1	+1	0	0
pH 1.8 mobility[h]	1.22	1.15	0.88	0.67	0.86	0.66
% yield	3.0	1.3	1.8	4.0	0.2	1.4

[a] Values are given in residues per molecule. The assumed integral values are given in parentheses. [b] Sum of homoserine and its lactone. [c] These values were consistently low due to the Val-Val bond in the Cn4 peptides. [d] Homoserine did not resolve from glutamic acid in this analysis. [e] AEV is aminoethylated valine. [f] Relative to aspartic acid ($m = -1.00$). [g] Calculated by the method of Offord (1966). [h] Relative to serine ($m = 1.00$).

Table III: Automatic Sequence Analysis of Fragment Cn2.

Edman Cycle	Residue (Yield in nmol[a])	Edman Cycle	Residue (Yield in nmol[a])	Edman Cycle	Residue (Yield in nmol[a])
1	Phe (100)	10	Ile (80)	19	Tyr (35)
2	Arg[b]	11	Gln (15)	20	Gly (15)
3	Lys (75)	12	Val (60)	21	
4	Asp (110)	13	Gly (25)	22	Ala (40)
5	Asn (120)	14	Ile (60)	23	
6	Ala (105)	15	Val (50)	24	Pro (5)
7	Asp (100)	16	Ser (35)	25	Gly (15)
8	Glu (70)	17	Trp (40)		
9	Trp (75)	18	Gly (15)		

[a] Analysis was by gas chromatography of the phenylthiohydantoins or their trimethylsilylated derivatives derived from 5 mg (0.9 μmol) of fragment. [b] Determined qualitatively by the method of Yamada and Itano (1966).

Table IV: Amino Acid Compositions[a] and Properties of Selected α-Lytic Protease Peptides of Fragment Cn2.

	Cn2α7b	Cn2α5d
Threonine	0.92 (1)	0.64 (1)
Serine		1.09 (1)
Glutamic acid	1.00 (1)	
Alanine		1.23 (1)
Valine	1.21 (1)	
Tyrosine	0.67 (1)	
Phenylalanine		1.00 (1)
NH_2 terminus	Val	Thr
pH 6.5 mobility	−0.30	0
pH 6.5 charge	−1	0

[a] Values are given in residues per molecule. The assumed integral values are given in parentheses. Calculation of mobility and charge are as in Table II. The dansyl-Edman method established the sequence of α7b to be Val-Tyr-Thr-Glu and of α5d to be Thr-Phe-Ala-Ser.

buffer (pH 7.5) was digested with 10 mg of α-chymotrypsin (Worthington, 3× crystallized) at 28° for 37 hr. The digestion was terminated by adjusting the pH to 3.5 with formic acid and the peptides demaleylated by incubation at 60° for 6 hr. Some precipitation occurred during the demaleylation and was removed by centrifugation. The supernatant was lyophilized.

Performic acid oxidized and maleylated S.G. trypsin (2.7 μmol) was digested at room temperature (21°) with 0.33 mg of α-chymotrypsin in 5 ml of 1 mM *N*-ethylmorpholine acetate buffer (pH 8.0) in the pH-Stat. The pH was maintained with 0.2 N NH_4OH and the reaction terminated by lyophilization after 5 hr.

Fragment Cn2 (0.4 μmol) was suspended in 5 ml of 0.05 M *N*-ethylmorpholine acetate buffer (pH 8.0) and incubated with 0.2 mg of α-chymotrypsin at 30° for 7 hr. The suspension was solubilized during the digestion.

Peptide Fractionation. The peptides resulting from α-lytic protease and chymotryptic digestion of fragment Cn1 were initially separated into three fractions (I, II, and III) on the anion exchanger AG1-X2 (Bio-Rad Laboratories, Richmond, Calif.) by the method of Landon (1964) and as previously described (Johnson and Smillie, 1971b). Each of these fractions was then subjected to ion-exchange chromatography on Chromobead P resin (Technicon Corp., Chauncey, N.Y.) using gradient system IV as described by Welinder and Smillie (1972). Eluates were analyzed with a

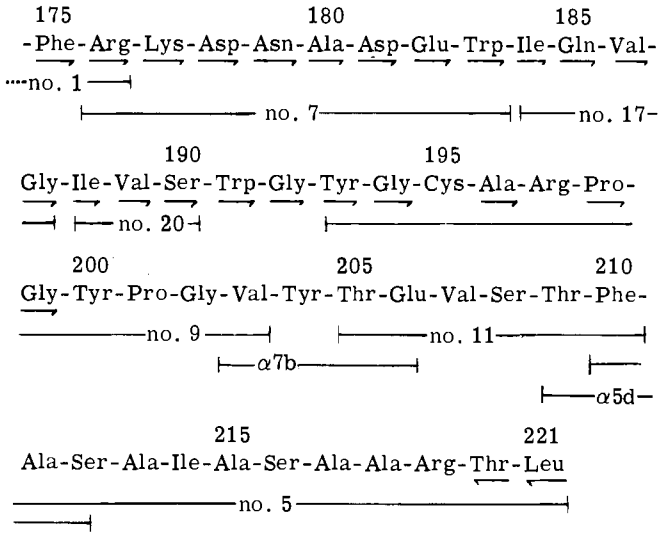

FIGURE 3: Amino acid sequence of fragment Cn2. Unique sequences 17 and 20 had been previously elucidated from peptic digests of S.G. trypsin (Jurášek and Smillie, 1973) and 1, 5, 7, 9, and 11 from both peptic and tryptic digests of reduced and aminoethylated S.G. trypsin (Jurášek and Smillie, 1974); → refers to identification of residues by automatic sequenator analyses and ← to COOH-terminal analysis of S-β-aminoethylated S.G. trypsin with carboxypeptidase A. Peptides α5d and α7b are selected peptides isolated from α-lytic protease digests of Cn2 (see Table IV). The numbering of residues is according to the final sequence of S.G. trypsin as shown in Figure 5.

Technicon peptide analyzer with the modifications described previously (Welinder and Smillie, 1972). Further purifications were by high-voltage electrophoresis on paper at pH 6.5, 1.8, and 3.5.

Sequence Methods. Amino acid analyses, manual sequence methods (dansyl-Edman), calculation of peptide mobilities, net charge, and assignment of Asx or Glx residues as Asp, Asn, Glu, or Gln were as before (Johnson and Smillie, 1971a,b; Hodges and Smillie, 1972). COOH-terminal analysis with carboxypeptidase A of inactivated, reduced, and aminoethylated S.G. trypsin was performed according to Johnson and Smillie (1972). Automatic sequencer analyses were done with a Beckman Sequencer, Model 890B. Samples of protein or fragment (5–10 mg) were dissolved in 0.4 ml of anhydrous trifluoroacetic acid and dried on the wall of the spinning reaction cup before starting the standard protein sequencing program. All other procedures were as described by Edman and Begg (1967) except that the chlorobutane was made 0.005% with respect to 1,4-butanedithiol. All reagents were of "Sequanal" grade (Pierce Chemical Co.). The amino acid residues were identified on a Beckman gas chromatograph, Model GC-50, as phenylthiohydantoins and their trimethylsilylated derivatives according to the procedures described by Pisano and Bronzert (1969) and Hermodson *et al.* (1972). Arginine residues were detected on paper by the method of Yamada and Itano (1966) and histidine according to Heilmann *et al.* (1957). In some cases the residues have been identified as their free amino acids on the amino acid analyzer after treatment with 6 N HCl, 0.05% β-mercaptoethanol at 110° for 24 hr under vacuum, or after hydrolysis with HI as described by Smithies *et al.* (1971).

Results

Automatic Edman Degradation of S.G. Trypsin. The results of these analyses as performed on native S.G. trypsin are reported in Table I. These data are in perfect agreement with the NH_2-terminal sequence of 17 residues previously reported for the enzyme (see unique sequence no. 4 of Jurášek and Smillie, 1974) and extend the sequence to residue

Table V: Amino Acid Compositions[a] of Cyanogen Bromide Fragments of S.G. Trypsin.

	Cn1[b]	Cn2[b]	Cn3[c]	Cn4[b]	Total	S.G. Trypsin[b,d]
Lysine	5.8 (6)	1.0 (1)			6.8	7.0 (7)
Histidine	0.8 (1)				0.8	1.1 (1)
Arginine	5.2 (4)	3.3 (3)	1.0 (1)	0.9 (1)	10.4	9.0 (9)
Aspartic acid	15.9 (16)	3.5 (3)			19.4	19.5 (19)
Threonine	13.1 (14)	3.1 (3)		0.9 (1)	17.1	17.7 (18)
Serine	11.2 (9)	3.5 (4)	0.9 (1)		15.6	15.3 (14)
Glutamic acid	13.4 (13)	3.3 (3)		1.9 (2)	18.6	18.8 (18)
Proline	6.1 (6)	1.9 (2)		0.9 (1)	8.9	8.7 (9)
Glycine	21.0 (22)	5.5 (5)		3.0 (3)	29.5	30.6 (30)
Alanine	19.0 (19)	7.0 (7)		2.0 (2)	28.0	28.0 (28)
Half-cystine	4.1[e] (5)	0.7[e] (1)			4.8[e]	6.6[f] (6)
Valine	12.3 (12)	3.9 (4)	1.0 (1)	0.9[g] (2)	18.1	19.2 (18)
Methionine	0.9[h] (1)		1.0 (1)[h]	0.9 (1)[h]	2.8[h]	3.0[i] (3)
Isoleucine	6.2 (6)	2.6 (3)			8.8	8.6 (9)
Leucine	10.0 (10)	1.2 (1)	1.0 (1)		12.2	11.9 (12)
Tyrosine	6.7 (6)	2.9 (3)			9.6	8.8 (9)
Phenylalanine	3.0 (2)	2.0 (2)		1.9 (2)	6.9	6.2 (6)
Tryptophan[j]	(2)	(2)				(4)

[a] Values are given in residues per molecule. Values in parentheses are based on final sequence for each fragment and for whole molecule. [b] Values calculated relative to alanine assuming 19.0, 7.0, 2.0, and 28.0 alanines for Cn1, Cn2, Cn4, and S.G. trypsin, respectively. [c] Values calculated relative to leucine equal to 1.0 residue. [d] Calculated from the amino acid analyses of Jurášek *et al.* (1969). [e] Determined as *S*-β-aminoethylcysteine. [f] Determined as cysteic acid. [g] Determined as the sum of valine and *N*-aminoethylvaline. [h] Determined as the sum of homoserine and homoserine lactone. [i] Determined as methionine sulfone. [j] Determined from sequence only.

Table VI: Automatic Sequence Analysis of Cn1.

Edman Cycle (Residue No.)	Residue[a] (Yield in nmol)	Edman Cycle (Residue No.)	Residue[a] (Yield in nmol)	Edman Cycle (Residue No.)	Residue[a] (Yield in nmol)
1 (21)	Gly (17)	14 (34)	Thr (24)	27 (47)	
2 (22)		15 (35)	Ala (25)	28 (48)	Ile/Leu (3.0)
3 (23)	Gly (50)	16 (36)	Ala (26)	29 (49)	Thr (1.3)
4 (24)	Gly (30)	17 (37)	His[b] ( )	30 (50)	Ala (6)
5 (25)	Ala (50)	18 (38)		31 (51)	Thr (3)
6 (26)	Leu (57)	19 (39)	Val (20)	32 (52)	
7 (27)	Tyr (50)	20 (40)	Ser (12)	33 (53)	
8 (28)	Ala (44)	21 (41)	Gly (12)	34 (54)	Val (2)
9 (29)	Gln (16)	22 (42)	Ser (9)	35 (55)	Val (2)
10 (30)	Asp (45)	23 (43)	Gly (11)	36 (56)	
11 (31)	Ile (30)	24 (44)	Asn (5)	37 (57)	Ile/Leu (2)
12 (32)	Val (34)	25 (45)	Asx[c] (13)		
13 (33)	Leu (28)	26 (46)	Thr (6)		

[a] Residues derived from 5 mg (0.3 μmol) of fragment analyzed as their phenylthiohydantoins or trimethylsilylated derivatives by gas chromatography. [b] Determined qualitatively by the method of Heilmann *et al.* (1957). [c] Found only as aspartic acid but from independent evidence known to be asparagine (see unique sequence no. 3 of Jurášek and Smillie, 1974).

37. This latter extension gave an overlap for unique sequences no. 4 and no. 3 previously reported and established the sequence of the five residues at positions 29 to 33 in the latter unique sequence. These data thus provided the sequence of the NH_2 terminus of S.G. trypsin from residues 1 to 45 which is summarized in Figure 1.

Cyanogen Bromide Cleavage and Fractionation of Fragments. The Sephadex G-50 elution pattern of 100 mg of protein which had been reduced and aminoethylated before cyanogen bromide cleavage is shown in Figure 2A. Fractions Cn1 and Cn2 were well resolved from one another but the former appeared to be contaminated with high molecular weight material, presumably uncleaved polypeptide chain, as judged by amino acid and NH_2-terminal analyses. Fraction Cn2 was homogeneous by these criteria and was taken to dryness by lyophilization. Fractions Cn3 and Cn4 did not absorb at 280 nm and were detected by high-voltage electrophoresis of fraction aliquots at pH 6.5. The peptides in these two fractions were not well resolved and were pooled together and lyophilized. The material absorbing at 280 nm following these fractions was devoid of peptides and was discarded.

The presence of high molecular weight material in fraction Cn1 could be attributed to either incomplete cleavage because of the partial alkylation of methionines during the reaction with ethylenimine prior to the cyanogen bromide reaction or to the aggregation of Cn1 with the other fragments in the eluting solvent employed. To avoid these possibilities, cyanogen bromide cleavage was carried out on S.G. trypsin whose disulfide bridges were intact. The resulting fragments were then reduced and alkylated with ethylenimine as described under Materials and Methods. Fractionation of the resulting S-β-aminoethylated fragments was then performed on a column of Sephadex G-75 in 0.65 M acetic acid–8 M urea as shown in Figure 2B. Amino acid and NH_2-terminal analyses indicated that fraction Cn1 was homogeneous.

Purification and Characterization of Fractions Cn3 and Cn4. Six peptides were purified by high-voltage electrophoresis at pH 6.5 and 1.8. Their electrophoretic mobilities, amino acid compositions, and NH_2-terminal analyses are presented in Table II. From an inspection of these data it is clear that they represent different chemical entities of two peptide sequences which from the automatic Edman degradation data of Table I are derived from the NH_2-terminal end of the protein. Thus, the two peptides Cn3A and Cn3B are clearly two varieties of fragment Cn3 (residues 16–20) differing only in the presence of homoserine lactone or homoserine at their COOH-teminal ends. Similarly, different varieties of fragment Cn4 (residues 1–15) were observed depending on whether homoserine (Cn4B1 and Cn4B2) or homoserine lactone (Cn4A1 and Cn4A2) was present. This fraction was further complicated by the presence of N-aminoethylated valine at the NH_2 terminus of two varieties (Cn4A1 and Cn4B1) of this peptide. We have previously presented evidence (Jurášek and Smillie, 1974) that during the preparation of the reduced and aminoethylated protein a certain proportion of the NH_2-terminal valine of the protein is alkylated with ethylenimine. The properties of the dansylated NH_2-terminal residues of these two peptide varieties were completely consistent with those of dansyl-aminoethylvaline as previously described.

Sequence Analysis of Fraction Cn2. Amino acid analyses of this fragment showed it to be devoid of homoserine and allowed its assignment to the COOH-terminal end of the protein. Its sequence has been elucidated both by automatic Edman degradation in the sequencer and by the manual dansyl-Edman procedure applied to peptides derived from α-lytic protease and chymotryptic digestion of the fragment. The automatic sequencer results (Table III) permitted the assignment of the first 25 residues at the NH_2 terminus of the fragment with the exception of the residues at cycles 21 and 23. Digestion with α-lytic protease and chymotrypsin resulted in the production of 23 and 12 peptides, respectively, which were separated and purified by high-voltage electrophoresis on paper at pH 6.5, 1.8, and 3.5. Of these only two were required as overlaps of unique sequences previously elucidated. The amino acid compositions and sequence data of these two α-lytic protease peptides (Cn2α7b and Cn2α5d) are given in Table IV. The compositions and sequence data of the remaining α-lytic and chymotryptic peptides were entirely consistent with the se-

Table VII: Amino Acid Compositions, Purification Procedures, Mobilities, Net Charges, and Sequence Data of Selected α-Lytic Protease and Chymotryptic Peptides of Cyanogen Bromide Fragment Cn1.

Peptide[a]	Purification Methods[b]	pH 6.5 Mobility[c]	Calcd Net Charge[d]	Analyses (Molar Ratios) and Sequence Data	Residue No.[e]	Yield (%)
α-I-2-2	3.5	0	0	Gly - Ser - Gly - Asn - Asn - Thr - Ser (Ile, Thr) 1.04 0.95 1.04 1.06 1.06 0.95 0.95 1.00 0.95	41–49	0.28
α-I-5-4	1.8; 3.5	0	0	Gly - Ser - Gly - Asn - Asn - Thr - Ser 1.00 0.94 1.00 1.09 1.09 0.92 0.94	41–47	0.21
α-I-7-3	1.8; 3.5	0	0	Ala - Thr - Gly - Gly - Val - Val 1.07 1.02 1.10 1.10 0.85 0.85	50–55	0.31
α-I-24-6	6.5; 1.8	+0.41	+1	Leu - Lys - Ile - Ala 1.15 0.94 1.00 0.90	95–98	0.54
α-II-5-3	1.8	0	0	Ala (Tyr, Asn, Gln, Gly, Thr, Phe, Thr, Val, Ala) 0.89 1.09 0.94 1.11 1.37 0.90 1.13 0.90 0.78 0.89	102–111	0.48
α-III-6-3	6.5; 1.8	−0.45	−1	Asx - Leu - Glx - Ser - Ala 1.00 0.72 0.85 0.92 0.87	56–60	0.11
α-III-14-2b	6.5; 1.8	−0.06	0	Ser - Ala - Val 0.74 0.99 1.02	59–61	0.22
α-III-11-1b	6.5; 1.8	0	0	Ala (Tyr, Asn, Glu, Gly, Thr, Phe, Thr) 0.99 0.94 1.04 1.01 1.07 0.96 1.00 0.95	102–109	0.34
α-III-11-2	6.5; 1.8	−0.04	0	Tyr - Asn - Gln - Gly - Thr - Phe - Thr - Val 0.81 1.00 1.03 1.07 0.96 0.98 0.96 0.84	103–110	0.05
α-III-26-3	6.5; 1.8	+0.18	+1	Gly - Trp - Gly - Ala - Asx - Arg - Glx (Gly, Gly, Ser, Glx) Glx 0.92 + 0.92 1.19 1.09 0.97 1.06 0.92 0.92 0.89 1.06 1.06 - Arg - Tyr - Leu - Leu 0.97 1.00 1.01 1.01	112–127	0.50
Ch-I-15-9	6.5; 1.8	−0.03	0	Thr - Val - Ala - Gly - Trp 0.95 0.95 1.01 1.06 +	109–113	0.44
Ch-LD[f]	6.5; 6.5	+0.25	+1	Leu - Leu - Lys - Ala - Asn - Val - Pro - Phe 1.01 1.01 0.90 1.13 1.04 1.03 0.84 1.01	126–133	

[a] Peptide nomenclature: α and Ch refer to peptides isolated from α-lytic protease and chymotryptic digests, respectively. Roman numerals refer to fraction recovered from AG1-X2. The first arabic number refers to fraction number recovered from Chromobead Type P ion-exchange chromatography. The second arabic number refers to order of migration of peptide in first paper electrophoretic system where the most basic peptide was given the lowest number. [b] Numbers refer to high-voltage electrophoretic purifications at pH 6.5, 3.5, or 1.8, [c] Mobilities are expressed relative to aspartic acid ($m = -1$). [d] Calculated according to Offord (1966). [e] Residue numbers in completed sequence (see Figures 4 and 5). [f] This peptide was isolated by the lysine diagonal method (LD) of Butler *et al.* (1969); see details in text.

quence of fragment Cn2 as detailed in Figure 3 and further information concerning them may be obtained from the authors. The assignment of the amino acid sequence of Cn2 (Figure 3) was made possible by combining the unique sequences previously elucidated from peptic and tryptic digests of the whole S.G. trypsin molecule (Jurášek and Smillie, 1973, 1974) with the sequencer results and the two α-lytic peptides described in the present work. The assignments of -Thr-Leu to the COOH terminus of fragment Cn2 and of this fragment to the COOH terminus of the whole S.G. trypsin molecule were consistent with the finding that leucine followed by threonine were released most rapidly from inactivated, reduced, and aminoethylated S.G. trypsin by treatment with carboxypeptidase A.

Sequence Analysis of Fraction Cn1. The assignment of fragments Cn3 and Cn4 to the NH_2-terminal end of the S.G. trypsin and of Cn2 to the COOH terminus indicated that the cyanogen bromide fragments could be ordered in the molecule as Cn4–Cn3–Cn1–Cn2. Amino acid analyses of Cn1 coupled with the sequences of Cn3 and Cn4 and the amino acid analysis of Cn2 and its sequence indicated that the total amino acid composition of the protein was accounted for by these four fragments (see Table V). These data further indicated that fragment Cn1 contained approximately 154 residues. The automatic sequenator results on intact S.G. trypsin (Table I) provided an overlap for fragments Cn3 and Cn1 and the sequence of the first 17 residues at the NH_2 terminus of fragment Cn1. These residues were clearly part of the sequence no. 3 previously elucidated (Jurášek and Smillie, 1974) which extended the sequence by a further eight residues to position 45 of the S.G. trypsin sequence.

Application of the automatic sequenator to fragment Cn1 (Table VI) confirmed these results and permitted the

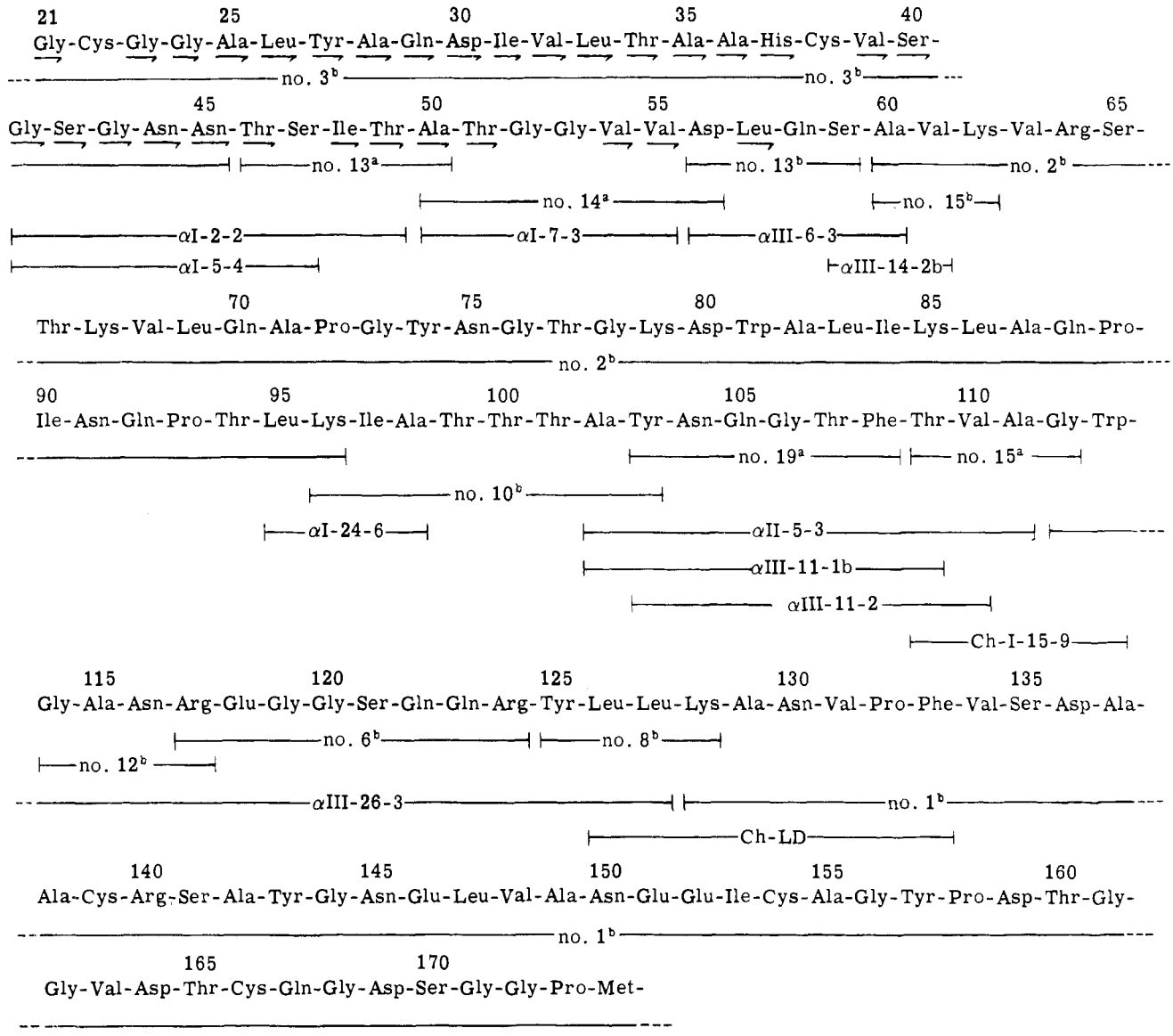

FIGURE 4: Amino acid sequence of fragment Cn1. Superscripts a and b refer to unique sequences previously elucidated from peptic (Jurášek and Smillie, 1973) and from tryptic (Jurášek and Smillie, 1974) digests, respectively; → denotes residues identified by automatic sequenator analyses of Cn1. Numbering of residues is according to the final sequence elucidated for the whole S.G. trypsin molecule (see Figure 5). Nomenclature of peptikdes is as described in Table VII.

1 5 10 15 20 25
Val-Val-Gly-Gly-Thr-Arg-Ala-Ala-Gln-Gly-Glu-Phe-Pro-Phe-Met-Val-Arg-Leu-Ser-Met-Gly-Cys-Gly-Gly-Ala-
30 35 40 45 50
Leu-Tyr-Ala-Gln-Asp-Ile-Val-Leu-Thr-Ala-Ala-His-Cys-Val-Ser-Gly-Ser-Gly-Asn-Asn-Thr-Ser-Ile-Thr-Ala-
55 60 65 70 75
Thr-Gly-Gly-Val-Val-Asp-Leu-Gln-Ser-Ala-Val-Lys-Val-Arg-Ser-Thr-Lys-Val-Leu-Gln-Ala-Pro-Gly-Tyr-Asn-
80 85 90 95 100
Gly-Thr-Gly-Lys-Asp-Trp-Ala-Leu-Ile-Lys-Leu-Ala-Gln-Pro-Ile-Asn-Gln-Pro-Thr-Leu-Lys-Ile-Ala-Thr-Thr-
105 110 115 120 125
Thr-Ala-Tyr-Asn-Gln-Gly-Thr-Phe-Thr-Val-Ala-Gly-Trp-Gly-Ala-Asn-Arg-Glu-Gly-Gly-Ser-Gln-Gln-Arg-Tyr-
130 135 140 145 150
Leu-Leu-Lys-Ala-Asn-Val-Pro-Phe-Val-Ser-Asp-Ala-Ala-Cys-Arg-Ser-Ala-Tyr-Gly-Asn-Glu-Leu-Val-Ala-Asn-
151 155 160 165 170 175
Glu-Glu-Ile-Cys-Ala-Gly-Tyr-Pro-Asp-Thr-Gly-Gly-Val-Asp-Thr-Cys-Gln-Gly-Asp-Ser-Gly-Gly-Pro-Met-Phe-
180 185 190 195 200
Arg-Lys-Asp-Asn-Ala-Asp-Glu-Trp-Ile-Gln-Val-Gly-Ile-Val-Ser-Trp-Gly-Tyr-Gly-Cys-Ala-Arg-Pro-Gly-Tyr-
205 210 215 220
Pro-Gly-Val-Tyr-Thr-Glu-Val-Ser-Thr-Phe-Ala-Ser-Ala-Ile-Ala-Ser-Ala-Ala-Arg-Thr-Leu

FIGURE 5: Complete amino acid sequence of S.G. trypsin. Disulfide bridges are between half-cystine residues 22–38, 139–154, and 166–195.

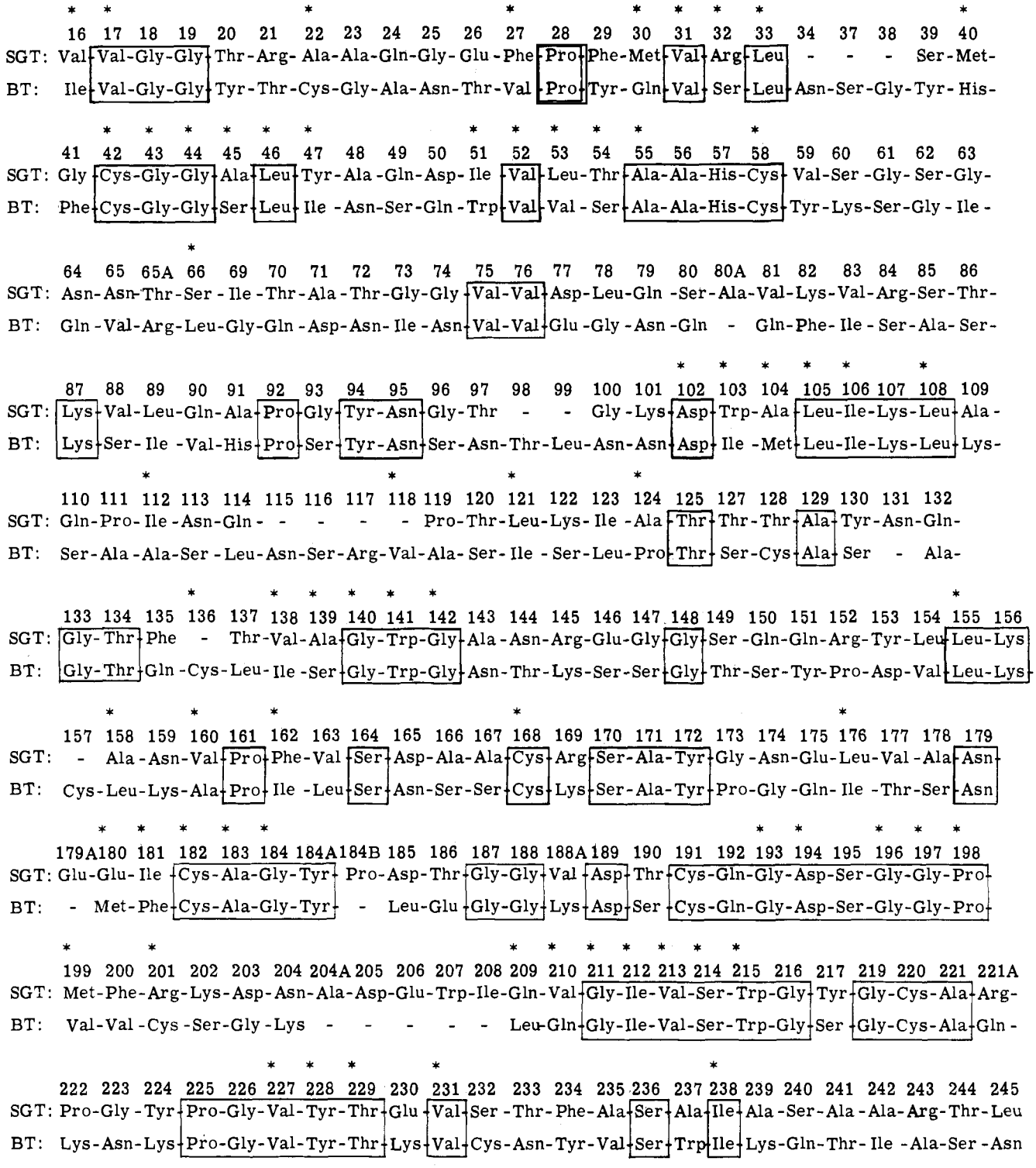

FIGURE 6: Comparison of amino acid sequences of S.G. trypsin (SGT) and bovine trypsin (BT). Identical residues are framed. The asterisk denotes internal residues inaccessible to water in BT (Stroud *et al.*, 1971). The numbering system is that of chymotrypsinogen A (Hartley and Kauffman, 1966). The sequence of BT was taken from Walsh and Neurath (1964) and Holeyšovský *et al.* (1967).

additional assignment of sequence through to residue 57 although the residues at positions 47, 52, 53, and 56 were not identified nor could isoleucine and leucine be differentiated at positions 48 and 57.

It was clear from the sequence information accumulated at this stage of the work that many of the unique sequences previously elucidated from peptic and tryptic digests of the intact S.G. trypsin molecule (Jurášek and Smillie, 1973, 1974) and not already assigned to the other fragments were derived from the Cn1 fragment. In an attempt to provide suitable overlaps for these sequences, the Cn1 fragment was digested with α-lytic protease and chymotrypsin (see Materials and Methods).

The lyophilized peptides from each digest were separated into three fractions, I, II, and III, by chromatography on AG1-X2, and each of these then applied separately to a Chromobead P column. The eluted peptides were pooled, lyophilized, and purified further, if necessary, by high-volt-

age electrophoresis on paper at pH 6.5, 3.5, and 1.8. By these procedures, over 70 peptides, ranging in size from free amino acids to 19 residues, were isolated in adequate purity for amino acid sequence analysis. We report here only those which have provided new information about the structure and which have provided overlaps for the previously elucidated sequences. Further information concerning the analysis of those peptides not reported in this paper may be obtained from the authors or by consulting Olafson (1973). In no case were the analytical data for these peptides inconsistent with the elucidated final sequence of Cn1 (see Figure 4).

Details of the amino acid composition, purification procedure, electrophoretic mobility, net charge, and sequence information for each of the selected α-lytic protease peptides and the single chymotryptic peptide are assembled in Table VII. Coupled with previous information derived from peptic and tryptic peptides and from the automatic sequenator results (Table VI), these analyses permitted the alignment of the complete sequence for the Cn1 fragment as shown in Figure 4.

Although this sequence was consistent with all of the analytical data, the available information provided only a single residue overlap in the region of the lysine residue at position 128. Further evidence was required for the verification of the sequence in this region. It was anticipated that by chymotryptic digestion of the protein under conditions of low enzyme to substrate ratios it would be possible to produce a peptide derived from residues 126 to 133 of the sequence shown in Figure 4. Selective isolation of this peptide should be possible by the diagonal lysine technique (Butler *et al.*, 1969) and would provide an excellent overlap for this region. Toward this end, 2.7 μmol of performic acid oxidized S.G. trypsin was maleylated, dialyzed, and finally digested at a 1:200 molar ratio of chymotrypsin to protein at pH 8.0 for 5 hr. After performing the diagonal procedure at pH 6.5, a peptide was isolated off the diagonal at a position corresponding to that anticipated for a peptide with the composition of residues 126–133 (*i.e.*, with a mobility of +0.25 relative to aspartic acid). Isolation of this peptide by the usual procedures allowed its composition and sequence to be determined (peptide Ch-LD of Table VII) and provided the necessary overlap in this region of the sequence.

Discussion

The determination of the amino acid sequences of each of the four cyanogen bromide fragments of S.G. trypsin and their overlapping from sequences elucidated around each of the three methionine residues have allowed the completion of the complete primary structure of the molecule (Figure 5). S.G. trypsin consists of a single polypeptide chain of 221 residues with three disulfide bridges at positions 22–38, 139–154, and 166–195. The amino acid composition of the protein as determined earlier (Jurášek *et al.*, 1969) is accounted for, within experimental error, by the final sequence (see Table V). The validity of the deduced primary structure is strengthened by the fact that in no case were peptides derived from several proteolytic digests which could not be accounted for by the elucidated structure.

On the basis of partial sequences about the disulfide bridges of S.G. trypsin we have previously described the remarkable homology between this protein and bovine trypsin. The completion of the sequence of S.G. trypsin now permits a more complete comparison of this homology. For the purposes of this comparison the sequences of S.G. trypsin and bovine trypsin have been aligned in Figure 6 to maximize their homology. It is clear that the overall lengths of the two polypeptide chains are very similar (221 residues in S.G. trypsin and 223 in bovine trypsin) and that although the number of disulfide bridges in S.G. trypsin is only 3 compared to 6 in bovine trypsin, the positions of these common bridges are identical in the two proteins. In addition there are many regions of the structure where the two sequences show extensive regions of identity or similarity. These include the NH_2-terminal four residues, the sequences near histidine-57 (chymotrypsinogen-A numbering scheme), aspartic acid-102, aspartic acid-189, serine-195, and regions near the COOH-terminal end of the protein (residues 216 and 226) known to be involved in the structure of the substrate-binding pocket of bovine trypsin. In this alignment the proportion of residues which are identical in the two proteins is 34%. When only those residues classified as internal and inaccessible to solvent are considered this identity is increased to 54%. This degree of homology is remarkable in view of the widely different phylogenetic sources of the two enzymes and indicates that the family of tryptic enzymes has been highly conserved throughout evolution. It also indicates that the folding of the polypeptide chains of the *Streptomyces griseus* enzyme and bovine trypsin is very similar and that the geometries of the active sites and substrate binding sites are very closely related, a conclusion in agreement with the very similar enzymic, chemical, and physical properties of the two enzymes (Olafson and Smillie, 1975). These predictions are being currently explored in this laboratory by comparative model building of *Streptomyces griseus* trypsin based on its amino acid sequence and the three-dimensional structure of bovine trypsin (Stroud *et al.*, 1971, 1974; Krieger *et al.*, 1974).

Acknowledgment

The authors are grateful to Mr. M. Nattriss for the many amino acid analyses performed in this work.

References

Butler, P. J. G., Harris, J. I., Hartley, B. S., and Leberman, R. (1969), *Biochem. J. 112*, 679.

Edman, P., and Begg, G. (1967), *Eur. J. Biochem. 1*, 80.

Gross, E. (1967), *Methods Enzymol. 11*, 461.

Hartley, B. S., and Kauffman, D. L. (1966), *Biochem. J. 101*, 229.

Heilmann, J., Barollier, J., and Watzke, E. (1957), *Hoppe-Seyler's Z. Physiol. Chem. 309*, 219.

Hermodson, M. A., Ericsson, L. H., Titani, K., Neurath, H., and Walsh, K. A. (1972), *Biochemistry 11*, 4493.

Hirs, C. H. W. (1967), *Methods Enzymol. 11*, 197.

Holeyšovský, V., Tomášek, V., Mikeš, O., Danilová, A., and Šorm, F. (1967), *Collect. Czech. Chem. Commun. 32*, 655.

Johnson, P., and Smillie, L. B. (1971a), *Can. J. Biochem. 49*, 548.

Johnson, P., and Smillie, L. B. (1971b), *Can. J. Biochem. 49*, 1083.

Johnson, P., and Smillie, L. B. (1972), *Can. J. Biochem. 50*, 589.

Jurášek, L., Fackre, D., and Smillie, L. B. (1969), *Biochem. Biophys. Res. Commun. 37*, 99.

Jurášek, L., Johnson, P., Olafson, R. W., and Smillie, L. B. (1971), *Can. J. Biochem. 49*, 1195.

Jurášek, L., and Smillie, L. B. (1973), *Can. J. Biochem. 51,* 1077.
Jurášek, L., and Smillie, L. B. (1974), *Can. J. Biochem. 52,* 382.
Krieger, M., Kay, L. M., and Stroud, R. M. (1974), *J. Mol. Biol. 83,* 209.
Landon, M. (1964), Ph.D. Thesis, University of Birmingham, England.
Offord, R. E. (1966), *Nature (London) 211,* 591.
Olafson, R. W. (1973), Ph.D. Thesis, University of Alberta, Edmonton, Alberta, Canada.
Olafson, R. W., and Smillie, L. B. (1975), *Biochemistry,* preceding paper.
Pisano, J. J., and Bronzert, T. J. (1969), *J. Biol. Chem. 224,* 5597.
Raftery, M. A., and Cole, R. D. (1966), *J. Biol. Chem. 241,* 3457.
Stroud, R. M., Kay, L. M., and Dickerson, R. E. (1971), *Cold Spring Harbor Symp. Quant. Biol. 36,* 125.
Stroud, R. M., Kay, L. M., and Dickerson, R. E. (1974), *J. Mol. Biol. 83,* 185.
Smithies, O., Gibson, D. Fanning, E. M., Goodfliesh, R. M., Gilman, J. G., and Ballantyne, D. L. (1971), *Biochemistry 10,* 4912.
Walsh, K. A., and Neurath, H. (1964), *Proc. Nat. Acad. Sci. U.S. 52,* 884.
Welinder, K. G., and Smillie, L. B. (1972), *Can. J. Biochem. 50,* 63.
Yamada, S., and Itano, H. A. (1966), *Biochim. Biophys. Acta 130,* 538.

Nuclear Magnetic Resonance Studies of Histone IV Solution Conformation†

A. Eugene Pekary,‡ Hsueh-Jei Li,§ Sunney I. Chan,* Chen-Jung Hsu, and Thomas E. Wagner#

ABSTRACT: The 220-MHz high-resolution proton magnetic resonance (PMR) spectrum of histone IV has been examined as a function of histone concentration, salt concentration, and pD. The hydrophobic C-terminal portion of the histone IV monomer appears to be largely PMR "invisible" indicating that this region of the polypeptide contains rigid secondary structure. Further loss of PMR resonance areas with increased histone IV concentration in neat D_2O has been attributed to self-aggregation involving a monomer–dimer equilibrium. An equilibrium between the monomer and large aggregates, on the other hand, appears to dominate at NaCl concentrations above 0.01 *M*. pD studies reveal an abrupt increase in histone IV aggregation at pD <0.8 and precipitation of histone IV at pD values in the neighborhood of its isoelectric point, pD $\simeq$ 11.

Cationic histones participate in nuclear chromatin function by forming electrostatic–hydrophobic complexes with anionic DNA (Lewin, 1969; 1970; Tanford, 1961; Bradbury et al., 1967; Smart and Bonner, 1971). In the presence of high ionic strength, however, such complex formation is suppressed (Hoare and Johns, 1971), presumably by counterion stabilization of the components. A similar complex destabilizing effect occurs when the pH lies outside the interval between the component isoelectric points. Nucleohistones may dissociate in this latter case because the sign of the component net charges will then be the same. For these reasons salt and acid extraction are the most commonly employed techniques for dissociating histones from purified nuclear chromatin (Elgin et al., 1971).

Upon dissociation from DNA, however, the polybasic histones experience increased aggregation, particularly at high ionic strengths (Tanford, 1961; Bradbury et al., 1967). Among the well-characterized histones, histone IV self-associates in electrolyte-containing media to the greatest extent (Diggle and Peacocke, 1971; Edwards and Shooter, 1969; Jirgensons et al., 1966). Investigations by optical rotatory dispersion (ORD) (Bradbury et al., 1965; Jirgensons and Hnilica, 1965; Tuan and Bonner, 1969), circular dichroism (CD) (Wagner, 1970; Shih and Fasman, 1971; Li et al., 1971), and CD and fluorescence (Li et al., 1972) have revealed that ionic strength changes affect not only the histone IV self-association tendency but also its conformation. According to these studies, the histone IV "extended coil" may acquire up to 20% α-helical content in uni-univalent salt solutions exceeding 0.1 *M*. Additional secondary and tertiary structure may also be present in the solution conformation of this histone since previous interpretations of ORD and CD spectra for "unordered" proteins based on a combination of random and helical segments may not have been entirely adequate (Dearborn and Wetlaufer, 1970; Li et al., 1971). The time course of changes in the his-

† From the Arthur Amos Noyes Laboratory of Chemical Physics, California Institute of Technology, Pasadena, California 91125 (Contribution No. 4565), and Cornell University Graduate School of Medical Sciences, New York, New York 10021. *Received July 18, 1974.* This work was supported by U.S. Public Health Service Grant GM-14523 from the National Institute of General Medical Sciences (to S.I.C.) and Grant CA-08748 from the National Cancer Institute (to T.E.W.).

‡ Population Council Postdoctoral Fellow, 1970–1972. Present address: Endocrinology Laboratory, Wadsworth VA Hospital, Wilshire and Sawtelle Boulevards, Los Angeles, California 90073.
§ Present address: Department of Chemistry, Brooklyn College of the City University of New York, Brooklyn, New York 11210.
# Present address: Department of Chemistry, Ohio University, Athens, Ohio 45701.