Blaha, K., Frick, I., Bezpalova, Z., and Kaurov, A. (1970), *Collect. Czech. Chem. Commun. 35*, 3557.
Blaha, K., Frick, I., and Rudinger, J. (1969), *Collect. Czech. Chem. Commun. 34*, 3497.
Bush, C. A. (1971), *in* Physical Techniques in Biological Research, Vol. IA, Oster, G., Ed., New York, N. Y., Academic Press, p 439.
Bush, C. A., and Tinoco, I., Jr. (1967), *J. Mol. Biol. 23*, 601.
Bystrov, V. F., Portnova, S. L., Tsetlin, V. I., Ivanov, V. T., and Ovchinnikov, Yu. A. (1969), *Tetrahedron 25*, 493.
Chen, A. K., and Woody, R. W. (1971), *J. Amer. Chem. Soc. 93*, 29.
Deber, C. M., Scatturin, A., Vaidya, V. M., and Blout, E. R. (1970), *Peptides: Chem. Biochem., Proc. Amer. Peptide Symp., 1st 1968*, 163.
Hooker, T., and Schellman, J. A. (1970), *Biopolymers 9*, 1319.
IUPAC–IUB (1970), *Biochemistry 9*, 3471.
Kopple, K. D. (1971), *Biopolymers 10*, 1139.
Kopple, K. D., Go, A., Logan, R. H., and Savrda, J. (1972), *J. Amer. Chem. Soc. 94*, 973.
Kopple, K. D., Ohnishi, M., and Go, A. (1969a), *J. Amer. Chem. Soc. 91*, 4264.
Maigret, B., Perahia, D., and Pullman, B. (1971), *Biopolymers 10*, 107.
Pysh, E. S. (1970), *Science 167*, 290.
Ramachandran, G. N., Chandrasekaran, R., and Kopple, K. D. (1971), *Biopolymers 10*, 2113.
Ramachandran, G. N., and Sasisekharan, V. (1968), *Advan. Protein Chem. 23*, 2083.
Schywzer, R. (1959), *Rec. Chem. Progr. 20*, 147.
Stern, A., Gibbons, W. A., and Craig, L. C. (1968), *Proc. Nat. Acad. Sci. U. S. 61*, 734.
Tinoco, I., Jr. (1962), *Advan. Chem. Phys. 4*, 113.
Tinoco, I., Jr. (1963), *Rad. Res. 20*, 133.
Urry, D. W., Ohnishi, M., and Walter, R. (1970), *Proc. Nat. Acad. Sci. U. S. 6*, 111.
Urry, D. W., Ruiter, A. L., Starcher, B. C., and Hinners, T. A. (1969), *Antimicrob. Ag. Chemother. 1968*, 87.
Venkatachalam, C. M. (1968), *Biopolymers 6*, 1425.
Ziaudin, Kopple, K. D., and Bush, C. A. (1972), *Tetrahedron Lett.*, 483.
Ziegler, S. M., and Bush, C. A. (1971), *Biochemistry 10*, 1330.

Amino Acid Sequence of Thermolysin. Isolation and Characterization of the Fragments Obtained by Cleavage with Cyanogen Bromide†

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

ABSTRACT: Commercial preparations of crystalline thermolysin were purified by gel filtration under conditions which minimized autolysis. Gel electrophoresis in the presence of sodium dodecyl sulfate and gel filtration in 6 M guanidine hydrochloride indicated the presence of a single polypeptide chain with a molecular weight of approximately 37,000–38,000. On the basis of the amino acid composition and metal content the minimum molecular weight was calculated to be 34,800. End-group analyses yielded amino-terminal isoleucine and carboxyl-terminal lysine. All attempts to detect carbohydrates gave negative results. As predicted from the presence of two methionine residues, three major fragments were obtained after treatment with cyanogen bromide. The molecular weights and amino acid compositions of these fragments were determined. End-group analyses of these fragments, together with characterization of a small amount of a larger overlap fragment, established the linear order within the polypeptide chain of thermolysin.

Thermolysin is an endopeptidase isolated in crystalline form from culture filtrates of the thermostable microorganism *Bacillus thermoproteolyticus* (Endo, 1962). The enzyme is representative of a group of neutral metalloproteases (Matsubara and Feder, 1971) which are largely bacterial in origin, operate at neutral pH, and are inhibited by chelating agents but not by inhibitors of "sulfhydryl" or "serine" proteases. The substrate specificity of neutral metalloproteases is preferentially directed toward internal peptide bonds wherein the imino nitrogen is contributed by hydrophobic residues (Matsubara and Feder, 1971). In this respect these enzymes resemble pancreatic carboxypeptidase A, although they are endopeptidases while the latter enzyme is an exopeptidase (Pétra, 1970).

Ohta *et al.* (1966) reported for thermolysin a preliminary molecular weight of 37,500 and an amino acid composition which revealed an abundance of hydrophobic residues. Latt *et al.* (1969) reported that the enzyme contains 1 g-atom of zinc per 37,500 g which is essential for the enzyme activity. In these features thermolysin resembles carboxypeptidase A (Neurath and Bradshaw, 1970). In contrast to carboxypeptidase A, thermolysin also contains 3–4 g-atoms of calcium per 37,500 g (Latt *et al.*, 1969) and maintains activity at higher temperatures than bovine carboxypeptidase A (Endo, 1962; Pétra, 1970). The presence of calcium ions appears to be

† From the Department of Biochemistry, University of Washington, Seattle, Washington 98195. *Received March 6, 1972.* This work has been supported by research grants from the National Institutes of Health (GM 15731) and the American Cancer Society (NP-18N).

important for the thermal stability of the enzyme (Feder *et al.*, 1971).

Because of our interest in its possible evolutionary relationship to carboxypeptidase A (Neurath and Bradshaw, 1970), we have undertaken a study of the amino acid sequence of thermolysin. Also, thermolysin appears to be representative of a group of neutral metalloproteases whose structure–function relationship is relatively unexplored (Matsubara and Feder, 1971). Simultaneously and independently Matthews and coworkers have carried out X-ray diffraction analysis of the enzyme (Matthews *et al.*, 1972b). Exchange and integration of data from both of these analyses have facilitated the construction of a model of the fine structure of the enzyme (Matthews *et al.*, 1972a). In a series of four papers the details of the chemical analyses are presented. This first paper describes the purification and chemical characterization of the enzyme and of fragments derived by cleavage with cyanogen bromide.

Experimental Procedure

Materials. Crystalline thermolysin (from *Bacillus thermoproteolyticus*, Rokko) was purchased from Daiwa Kasei K. K., Osaka, Japan. The preparation contained 65% protein, 21% calcium acetate, and 11% sodium acetate. Among several preparations of the enzyme examined, one (Lot No. TOBBO1 with a specific activity of 9080 PU/mg) was selected and used throughout the present study because of its high quality as determined by end-group analyses and disc gel electrophoresis. Proteins used as standards in the various analyses were products of Worthington (ovalbumin, carboxypeptidase A, α-chymotrypsin, lysozyme, trypsin, trypsinogen, and catalase), of Mann (myoglobin), and of Sigma (insulin). Prothrombin and α_1-acid glycoprotein were donated by Dr. K. Fujikawa (Department of Biochemistry, University of Washington) and Dr. S. Hakomori (Department of Pathobiology, Universtiy of Washington), respectively. These proteins were used without further purification.

Sephadex G-75, medium, and Sepharose 6B were purchased from Pharmacia Fine Chemicals, and Bio-Gel P-60 (100–200 mesh) and P-100 (100–200 mesh) from Bio-Rad Laboratories.

Cyanogen bromide was obtained from Eastman Organic Chemicals and stored in a sealed desiccator at 4° when not in use. Trifluoroacetic acid, obtained from J. T. Baker Chemical, was refluxed with solid CrO_3 (2 g/100 ml) for 6 hr and the fraction which boiled between 71 and 73° collected. Guanidine hydrochloride was purchased from Heico Inc. and furylacryloylglycyl-L-leucinamide (FAGLA)[1] from Cyclo Chemical Co. Other reagents used were of analytical grade.

Methods. Acid hydrolysis was carried out by the method of Moore and Stein (1963). Amino acid analyses were performed with Spinco 120 amino acid analyzers (Spackman *et al.*, 1958). Homoserine lactone in acid hydrolysates was converted to homoserine by the method of Ambler (1965).

Tryptophan and tyrosine were determined spectrophotometrically by the method of Goodwin and Morton (1946). The tryptophan content was also measured after base hydrolysis (Noltmann *et al.*, 1962), after acid hydrolysis with *p*-toluenesulfonic acid (Liu and Chang, 1971), and by counting the chymotryptic peptides stained with *p*-dimethylaminobenzaldehyde on a peptide map (Katz *et al.*, 1959).

Carbohydrate Analysis. Total hexose was determined by the orcinol method of Winzler (1955) and by the phenol–sulfuric acid method of Hirs (1967). The results were expressed in terms of a 1:1 weight ratio of galactose–mannose standard. Total amino sugar was determined by the Elson–Morgan method as modified by Gatt and Berman (1966). Glucosamine·HCl was used as standard. Neuraminic acid (sialic acid) was determined by the Warren procedure (1959) standardized with *N*-acetylneuraminic acid. Carbohydrate analysis was also performed by the gas chromatographic procedure of Yang and Hakomori (1971) using α_1-acid glycoprotein as standard. Attempts to detect carbohydrate in samples were also made qualitatively by the procedure of Zacharius *et al.* (1969) using prothrombin as standard.

End-Group Analysis. Amino-terminal residues were determined by the fluorodinitrobenzene method of Sanger (1945) and by the cyanate method of Stark (1967). In both procedures the reaction was performed in 8 M urea. In the former method, DNP-amino acids were identified by one-dimensional paper chromatography using *tert*-amyl alcohol saturated with potassium phthalate buffer (pH 6.0) or 1.5 M phosphate buffer (pH 6.0) (Fraenkel-Conrat *et al.*, 1955). A Beckman Sequencer (Model 890A) was employed for determining amino-terminal sequences according to the method of Edman and Begg (1967) as modified by Hermodson *et al.* (1970). The products of degradation were identified by gas–liquid chromatography. Carboxyl-terminal sequences were determined using carboxypeptidases A and B (Ambler, 1967).

Molecular Weight Determination. Molecular weights were determined by gel electrophoresis in the presence of sodium dodecyl sulfate (Weber and Osborn, 1969) using standard proteins with known molecular weights. Molecular weights were also determined by gel filtration on a 1.5 × 80 cm column of Sepharose 6B in 6 M guanidine hydrochloride (pH 5.0) at a flow rate of approximately 3 ml/hr (Fish *et al.*, 1969). The column was standardized with ovalbumin (43,000), trypsinogen (24,000), lysozyme (14,300), and insulin (6,400 without reduction of disulfide bonds).

Activity Assay. Thermolysin activity was assayed spectrophotometrically by an adaptation of the methods of Feder (1968) and Latt *et al.* (1969). Aliquots (20 μl) of 0–0.1 mM enzyme solutions were added to each 2 ml of 1 mM of FAGLA in 0.05 M Tris–0.01 M $CaCl_2$–0.1 M NaCl (pH 7.5). The reaction was followed at room temperature by measuring absorption at 345 nm for 10 min. Activity was arbitrarily expressed as the decrease in the absorbance (ΔA/min) calculated from a linear portion of the reaction in the first few minutes.

Purification of Thermolysin. One gram of commercial enzyme was dissolved in 50 ml of cold 1 M NaCl–0.01 M sodium acetate–0.002 M calcium acetate, adjusted to pH 6.0 with 0.01 N acetic acid. The solution was adjusted to approximately pH 11 with 0.2 N NaOH, then immediately neutralized to pH 6.0 with 0.2 N acetic acid. After clarification by centrifugation at 10,000 rpm at 0° for 10 min, the supernatant solution was placed on a 5 × 100 cm column of Sephadex G-75 equilibrated in the cold room (0–5°) with 1 M NaCl–0.01 M sodium acetate–0.002 M calcium acetate (pH 6.0). Elution was carried out with the same buffer and the effluent was collected in 20-ml aliquots at a flow rate of 120 ml/hr. The column was monitored both by absorbance at 280 nm and in a Technicon Autoanalyzer with ninhydrin after alkaline hydrolysis of 100-μl aliquots. Enzyme activity was assayed in 20-μl aliquots. Fractions containing active enzyme were

[1] The following abbreviations are used: FAGLA, furylacryloylglycyl-L-leucinamide; FDNB, 1-fluoro-2,4-dinitrobenzene; DNP, 2,4-dinitrophenyl.

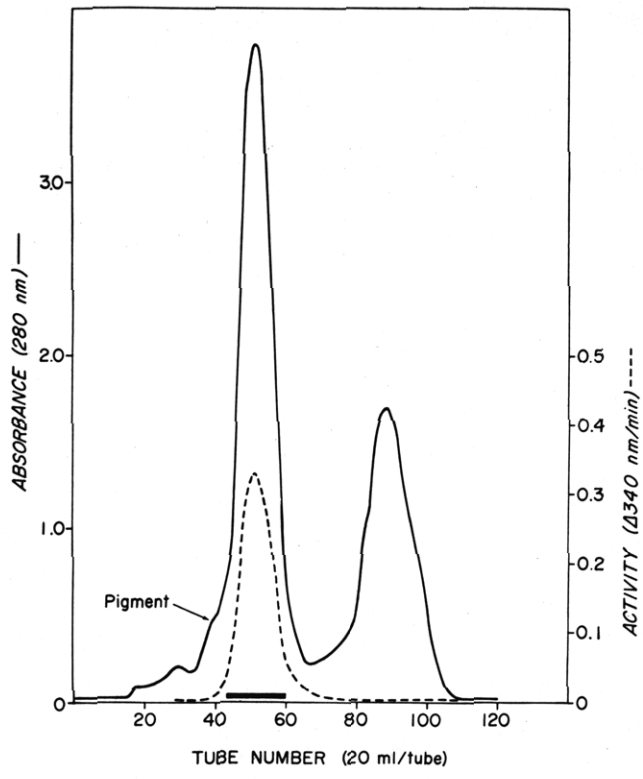

FIGURE 1: Purification of thermolysin on a column of Sephadex G-75 at pH 6 (see text for details).

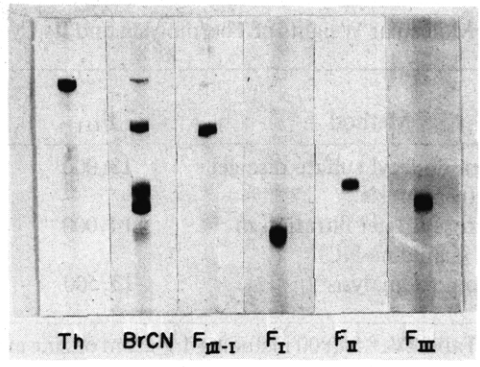

FIGURE 2: Gel electrophoresis in the presence of sodium dodecyl sulfate of purified thermolysin (Th), the whole cyanogen bromide digest (CNBr), and four isolated cyanogen bromide fragments. Samples of 0.02 mg were applied to the gels. The concentration of cross-links was 1.2%.

combined and denatured by precipitation with trichloroacetic acid at a final concentration of 1%. A precipitate was collected by centrifugation at 10,000 rpm at 0° for 10 min, washed once with 100 ml of 1% trichloroacetic acid, twice with 100 ml of 50% methanol, and finally suspended in 100 ml of water and lyophilized.

Cyanogen Bromide Cleavage. Cleavage with cyanogen bromide was carried out by the method of Gross and Witkop (1962) as modified by Bargetzi *et al.* (1964) and Sampath Kumar *et al.* (1964). Approximately 500 mg of purified and denatured thermolysin were dissolved in a mixture of 30 ml of purified trifluoroacetic acid and 10 ml of water. Cyanogen bromide (1 g) was added and the mixture allowed to stand at room temperature for 15–20 hr in the dark. The reaction was terminated by slowly pouring the mixture into 200 ml of chilled water. After gentle stirring at 0–5° for 1 hr, the precipitate was collected by refrigerated centrifugation, suspended in 50 ml of water, and lyophilized. The recovery of peptides in this precipitate was practically 100% by weight.

Results

Purification and Characterization of Enzyme Protein. A typical elution profile of the commercial enzyme on a column of Sephadex G-75 is shown in Figure 1. Only the first peak showed activity against FAGLA. Some yellowish pigment, which eluted with the leading edge of this peak, was discarded. The enzyme was recovered by precipitation with trichloroacetic acid immediately after chromatography to avoid further autolysis. It was concluded that recovery of polypeptide in the precipitate was complete since the supernatant showed no absorbance at 280 nm and did not react with ninhydrin before or after alkaline hydrolysis. The average isolation yield of the enzyme protein was 525 mg from 1 g of the starting material. This corresponds to 80% recovery since the starting material contained approximately 65% protein by weight.

The second peak in Figure 1 appeared to be a mixture of small peptides. This fraction did not precipitate with trichloroacetic acid but reacted with ninhydrin, more so after alkaline hydrolysis. The size of the peak was much smaller when the enzyme was recrystallized prior to gel filtration. These peptides may well have been derived from autodigestion during storage or shipment. If so, digestion appears to have been of the "all or none" type since no fragments of an intermediate size are apparent in Figure 1.

After precipitation with trichloroacetic acid the purified enzyme protein irreversibly lost activity and solubility at neutral pH. According to disc gel electrophoresis in the presence of sodium dodecyl sulfate (Figure 2), the protein contained a single polypeptide chain. By comparison with standard proteins (Figure 3) the molecular weight was estimated to be approximately 37,000. Gel filtration in 6 M guanidine

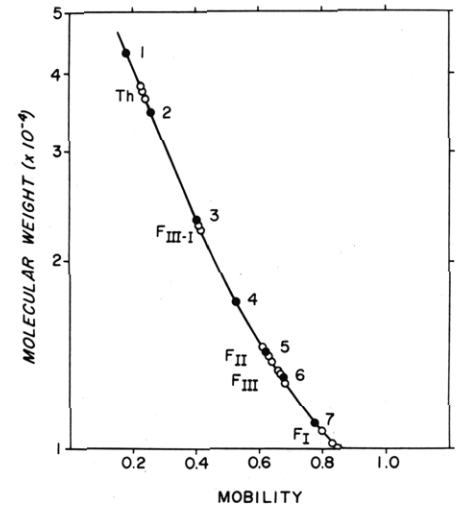

FIGURE 3: Determination of the molecular weights of thermolysin and the cyanogen bromide fragments by the method of Weber and Osborn (1969) (1, ovalbumin; 2, carboxypeptidase A; 3, trypsin; 4, myoglobin; 5, lysozyme; 6 and 7, α-chymotrypsin). Solid circles, standard proteins; open circles, unknown samples. Samples of 0.02 mg each were applied to the gels. The concentration of cross-links was 1.2%.

TABLE I: Molecular Weights of Thermolysin and Its Cyanogen Bromide Fragments.

Method	F_{III}	F_I	F_{II}	$F_{III} + F_I + F_{II}$	Thermolysin	F_{III-I}	$F_{III} + F_I$
Sodium dodecyl sulfate disc gel electrophoresis	13,000	10,000	14,000	37,000	37,000	23,000	23,000
Sepharose 6B gel filtration in 6 M guanidine-HCl	14,000	10,000	14,000	38,000	38,000	23,000	24,000
Amino acid analysis[a]	13,500	9,100	11,700	34,300	34,600[b]	22,100	22,600

[a] See Table IV. [b] 34,800 including 1 g-atom of zinc and 3 g-atoms of calcium per mole of enzyme.

hydrochloride also indicated the presence of a single polypeptide with a molecular weight of approximately 38,000. On the other hand, a minimum molecular weight of 34,600 was calculated from the amino acid composition (assuming that the enzyme is composed only of amino acid residues). The results of molecular weight determinations are summarized in Table I.

End-group analyses also provided evidence of only one polypeptide chain. The sole end group detected with FDNB was 0.80 mole of DNP-Leu (or Ile) per 34,600 g of protein. Analysis by the cyanate method established the amino-terminal residue as isoleucine (Table II). Sequenator analysis revealed a unique sequence of 15 residues at the amino terminus of the protein as shown in Figure 4.

By use of carboxypeptidases A and B the carboxyl-terminal sequence was shown to be -Val-Lys-OH (Table III). Carboxypeptidase A only released traces of serine, glycine, alanine, and valine during 6-hr digestion at 37° (each less than 0.05 mole per 34,600 g). Carboxypeptidase B established lysine at the carboxyl terminus (0.90 mole per 34,600 g in 6 hr at 37°). Digestion with a mixture of carboxypeptidases A and B indicated that glycine, alanine, and valine were released after lysine and valine, but the order of these residues is not certain.

TABLE II: Amino-Terminal Residues of Thermolysin and Its Cyanogen Bromide Fragments by the Cyanate Method.[a]

Amino Acid	Thermolysin	F_{III}	F_I	F_{II}	F_{III-I}
Asx				0.08	
Thr	0.06		0.05	0.06	
Ser			0.06	*0.83*	
Glx			0.09		
Pro				*0.15*	
Gly		0.10			
Ala		0.05			
Val			*0.87*		
Ile[b]	*0.66*	*0.88*			*0.81*
Leu		0.06			0.05

[a] Values are corrected by recovery factors given by Stark (1967) and shown as residues per molecule using molecular weights obtained by amino acid analysis (Table I). Values less than 0.05 after correction are omitted. [b] Value includes that for alloisoleucine.

The amino acid composition of the purified protein obtained from duplicate analyses of hydrolysates for four different times is shown in Table IV. There appears to be good agreement between the present results and those of Ohta *et al.* (1966) except for values for tryptophan. In our experiments the tryptophan content was determined by four different methods (Table V). In each case, approximately 3 moles of tryptophan were found per 34,600 g. The spectrophotometric method also yielded 28 residues of tyrosine, in agreement with the value obtained by acid hydrolysis.

From the present results (Table IV) it appears that the protein is composed of 318 amino acid residues (316 residues from the complete amino acid sequence, Titani *et al.*, 1971; Titani *et al.*, 1972) and on this basis a molecular weight of 34,587 is calculated, assuming that the protein consists only of amino acid residues. Small differences between molecular weights estimated by physical methods and by amino acid analysis (Table II) allowed for the possibility that thermolysin might be a glycoprotein. Therefore, carbohydrate analyses were carried out by various methods but no trace of carbohydrate could be detected. Disc gels of 0.02–0.2 mg of purified (or crude) enzyme treated with the periodic acid–Schiff base reagent (Zacharius *et al.*, 1969) were completely negative whereas control gels containing 0.02 mg of prothrombin showed carbohydrate. Gas chromatographic analyses by the method of Yang and Hakomori (1971) indicated a small contaminant that did not correspond to any known sugar deriva-

TABLE III: Digestion of Thermolysin with Carboxypeptidases A and B.[a]

	Time of Digestion (hr)			
	1	3	6	24
Lys	0.66	0.84	0.89	0.97
Val	0.43	0.81	1.20	1.62
Gly	0.16	0.32	0.52	0.92
Ala	0.11	0.27	0.39	0.86
Ser	0.12	0.17	0.15	0.20

[a] Digestion was carried out at 37°. Values are expressed in moles of amino acid released from 34,600 g of protein. Amino acids of less than 0.10 mole were omitted. The following sequence is indicated:

(Ala, Val, Gly)-Val-Lys-OH

```
                1                   5                  10                  15                 316
Thermolysin     Ile - Thr- Gly-Thr-Ser -Thr-Val- Gly- Val- Gly- Arg-Gly- Val- Leu-Gly----------Val- Lys

                1                   5                  10                  15                  20
Fragment FIII   Ile - Thr- Gly-Thr- Ser -Thr- Val- Gly-Val- Gly- Arg-Gly-Val- Leu-Gly-Asp-Gln-Lys- Asn-Ile-

                                   25                  30                  35                 120
                Asn-Thr- Thr-Tyr- Ser -Thr- Tyr-Tyr- Tyr-Leu-Gln-Asp-Asn-Thr-Arg-Gly---------- Met-

               121                 125                 130                 135                 140
Fragment FI     Val -Tyr- Gly-Asp-Gly- Asp-Gly-Gln-Thr-Phe-Ile - Pro- Leu-Ser -Gly- Gly-Ile - Asp-Val- Val-

                                   145                 150                 155                 160
                Ala-His - Glu-Leu-Thr- His -Ala-Val- Thr-Asp-Tyr- Thr- Ala-Gly-Leu-Ile - Tyr-Gln- Asn-Glu-

                                   165                                                         205
                Ser -Gly- Ala- Ile - Asn-Glu-Ala-Ile ----------------------------------------------Met-

               206                 210                 215                 220                 225
Fragment FII    Ser -Asp-Pro- Ala-Lys- Tyr- Gly-Asp-Pro-Asp-His- Tyr- Ser -Lys- Arg-Tyr-Thr-Gly- Thr- Gln-

                                   230                 235                 240                 245
                Asp-Asn-Gly-Gly- Val- His- Ile - Asn-Ser- Gly- Ile - Ile - Asn-Lys- Ala- Ala- Tyr-Leu-Ile - Ser-

                                  316
                ----------------Lys
```

FIGURE 4: Partial amino acid sequence of thermolysin. The amino-terminal sequences of thermolysin and its cyanogen bromide fragments were determined by duplicate analyses with the sequenator. (Identification of Ser_{234} in F_{II} was marginal.) In the case of F_{II} about 5–10% of the chain began with Pro_{208} and 5–10% with Pro_{214} (see text). The linear arrangement of the three fragments and the carboxyl-terminal analyses are discussed in the text. Residue numbers are those established in the complete sequence (Titani *et al.*, 1972).

tive. Other analyses for carbohydrate were essentially negative.

Isolation of the Cyanogen Bromide Fragments. When after cyanogen bromide treatment the reaction mixture was diluted with water, all of the detectable peptide fragments precipitated quantitatively, indicating that the fragments were all fairly large in size. Amino acid analysis of the whole mixture of precipitated fragments showed that at least 98% of the methionine residues were converted to homoserine. As to be expected from the presence of two methionine residues per mole of protein (Table IV), sequenator analysis of the whole digest showed the presence of three components in similar yields: 1st turn: Val, Ser, and Ile; 2nd turn: Tyr, Asp, and Thr; 3rd turn: Pro and Gly (approximately twice as much as Pro). On the other hand, the sodium dodecyl sulfate gel electrophoresis of the whole digest indicated the presence of four fragments having molecular weights of approximately 23,000, 14,000, 13,000, and 10,000, respectively (Figure 2 and Table I). This discrepancy in the number of fragments is explained below by the isolation of a fragment (F_{III-I}) containing fragments F_{III} and F_I joined by an uncleaved homoseryl residue.

The primary fractionation of the cyanogen bromide digest by gel filtration is shown in Figure 5. The lyophilized mixture of the fragments (approximately 500 mg) was dis-

TABLE IV: The Amino Acid Composition of Thermolysin.[a]

Amino Acid	Hydrolysis Time (hr) 24	48	72	96	Aver-age	Integral Value: Present Analysis	Ohta *et al.* (1966)
Lys	11.18	10.69	11.16	10.80	10.96	11	12
His	8.20	8.15	8.15	8.35	8.21	8	9
Arg	10.48	9.50	10.17	9.98	10.03	10	10
Asx	44.39	43.85	45.10	43.62	44.24	44	43
Thr	23.60	22.91	22.49	21.77	24.72[b]	25	23
Ser	24.67	23.51	22.18	20.34	26.58[b]	27	23
Glx	22.41	21.49	21.91	21.69	21.87	22	20
Pro	8.14	8.20	8.12	8.22	8.17	8	8
Gly	35.53	35.00	35.67	35.89	35.52	36	36
Ala	28.00	28.00	28.00	28.00	28.00	28	28
Cys/2	0.00	0.00	0.00	0.00	0.00	0	0
Val	20.50	20.91	21.69	21.78	21.78[c]	22	24
Met	2.03	2.04	2.01	2.22	2.07	2	2
Ile	16.82	17.04	17.53	17.52	17.52[c]	18	18
Leu	16.04	15.88	16.04	15.85	15.95	16	17
Tyr	28.54	28.02	28.42	28.43	28.35	28	29
Phe	10.12	10.13	10.20	10.33	10.19	10	10
Trp						3[d]	5
Total						318	317

[a] The values for each hydrolysis time are the average of two analyses. Calculations were based on 28.00 residues of Ala per molecule. [b] Extrapolated to zero time. [c] Equivalent to 96-hr value. [d] Taken from the data in Table V.

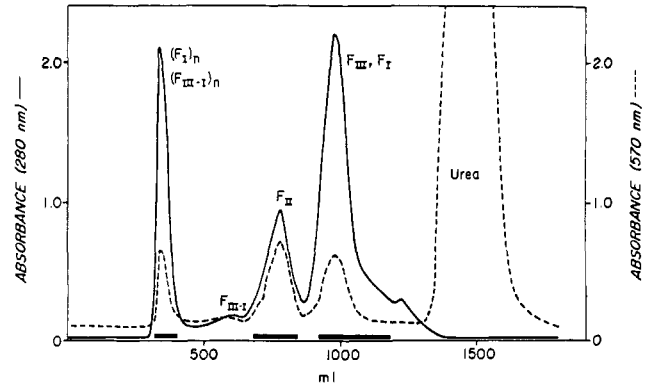

FIGURE 5: Separation of the cyanogen bromide fragments of thermolysin on a 5.0 × 100 cm column of Bio-Gel P-100 equilibrated with 1.0 M formic acid. The effluent was monitored both at 280 nm (solid line) and by ninhydrin analysis after alkaline hydrolysis of 50-μl aliquots (broken line). For further details, see the text.

TABLE V: Tryptophan and Tyrosine Contents of Thermolysin and Its Cyanogen Bromide Fragments.[a]

Method[b]		CNBr Fragment[c] F_{III}	F_I	F_{II}	$F_{III} + F_I + F_{II}$	Thermolysin[c]
Spectrophotometric method	Trp	1.55 (2)	1.02 (1)	0.05 (0)	3	2.67 (3)
	Tyr	14.72 (15)	5.00 (5)	7.78 (8)	28	28.19 (28)
Hydrolysis with $Ba(OH)_2$ (48 hr)		1.55 (2)	0.71 (1)	0.18 (0)	3	2.55 (3)
Hydrolysis with *p*-toluenesulfonic acid	24 hr	1.61 (2)	0.57 (1)	0.00 (0)	3	2.62 (3)
	72 hr	1.54 (2)	0.62 (1)	0.00 (0)	3	2.67 (3)
Peptide map		2	1	0	3	3

[a] The data are averages of duplicate analyses. Calculations were made on the basis of molecular weights obtained from amino acid compositions (Table IV). [b] See the text for details. [c] Integral values are shown in parentheses.

solved in 20 ml of 1.0 M formic acid containing 8 M urea, and applied to a 5 × 100 cm column of Bio-Gel P-100 equilibrated with the acid only. The fragments were eluted at room temperature with 1 M formic acid, and the effluent collected in 10-ml aliquots at a flow rate of 60 ml/hr. The column was monitored by absorbance at 280 nm and by ninhydrin analysis (Technicon Autoanalyzer) after alkaline hydrolysis of 50-μl aliquots. The peptide fragments were recovered by lyophilization.

Estimation of molecular weights by sodium dodecyl sulfate gel electrophoresis and examination of the elution volumes on the P-100 column indicated that the first peak contained aggregates of both fragment F_I and of fragment F_{III-I}, designated in Figure 5 as $(F_I)_n$ and $(F_{III-I})_n$, respectively. The separation of these two fragments was accomplished by chromatography either on a 2.5 × 100 cm column of Bio-Gel P-100 in 1 M formic acid–8 M urea or on a 2.5 × 100 cm column of Bio-Gel P-60 in 0.1 M ammonium bicarbonate (pH 8.0) (Figure 6). In both columns the flow rate was 30 ml/hr

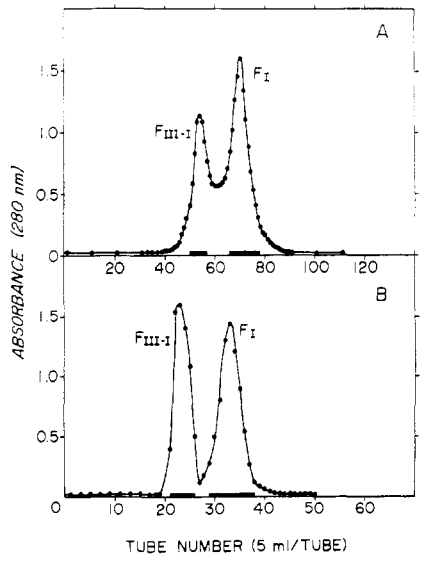

FIGURE 6: The separaton of fragment F_{III-I} on a 2.5 × 100 cm column of (a) Bio-Gel P-100 in 8 M urea–M formic acid or (b) Bio-Gel P-60 in 0.1 M ammonium bicarbonate, pH 8.0. Aliquots of 5 ml were collected and combined as shown by the solid bars.

and the columns were monitored by absorbance at 280 nm. In the P-100 column, urea was removed by dialysis against water. It appeared that fragment F_I dissociated while F_{III-I} remained polymeric at the neutral pH which was employed for elution of the P-60 column.

The second peak in the primary separation (Figure 5) contained the homogeneous monomeric fragment F_{II} and occasionally trace amounts of monomeric fragment F_{III-I} at the leading edge. In the latter case, the fragment was purified by rechromatography on the same column as that used for the primary separation.

The third peak was a mixture of fragments F_{III} and F_I, both of which appeared to be monomeric. These two fragments were separated by differential precipitation in 2 M urea at neutral pH. The lyophilized mixture of the third fraction was dissolved in a small volume of 8 M urea–0.1 M ammonium bicarbonate (pH 8.0) (100 mg per 5 ml), and then diluted with water to 2 M urea. The resulting precipitate (F_{III}) was collected by centrifugation. The precipitate and the supernatant were separately dialyzed against water and lyophilized. Average isolation yields of the four fragments are shown in Table VI.

Molecular Weight of the Cyanogen Bromide Fragments. Each isolated fragment was found to be almost homogeneous by several criteria. Sodium dodecyl sulfate disc gel electrophoresis indicated the presence of a single component in

TABLE VI: Yields of the Cyanogen Bromide Fragments.[a]

	Amount Isolated (mg)	Amount Expected (mg)[b]	% Yield
F_I[c]	60	133	45
F_{II}	170	170	100
F_{III}	140	197	71
F_{III-I}	20	330	6.1

[a] From 500 mg of protein. [b] Calculated on the basis of molecular weights of fragments obtained from amino acid compositions (Table I). [c] Sum of the fragments isolated from the first and third peaks in Figure 5.

TABLE VII: The Amino Acid Compositions of the Cyanogen Bromide Fragments of Thermolysin.

Amino Acid	F_{III} Average[a]	F_{III} Integral Value	F_I Average[a]	F_I Integral Value	F_{II} Average[a]	F_{II} Integral Value	$F_{III} + F_I + F_{II}$[b]	Thermolysin[c]	F_{III-I} Average[a]	F_{III-I} Integral Value	$F_{III} + F_I$[b]
Lys	2.83	3	1.17	1	6.78	7	11	11	4.05	4	4
His	2.70	3	2.20	2	2.88	3	8	8	4.62	5	5
Arg	5.40	5	0.98	1	4.00	4	10	10	6.17	6	6
Asx	21.12	21	11.82	12	11.17	11	44	44	32.09	32	33
Thr	10.89[d]	11	5.85[d]	6	7.82[d]	8	25	25	16.68[d]	17	17
Ser	8.98[d]	9	5.59[d]	6	10.52[d]	11	26	27	15.03[d]	15	15
Hse	0.78[d]	1	0.83[d]	1	0.00[d]	0	2	2(Met)	1.59[d]	2	2
Glx	5.14	5	8.06	8	8.01	8	21	22	14.35	14	13
Pro	2.01	2	2.62	3	3.09	3	8	8	4.58	5	5
Gly	11.90	12	10.52	11	13.39	13	36	36	23.41	23	23
Ala	11.00	11	6.00	6	11.00	11	28	28	17.00	17	17
Val	7.36[e]	7	5.66[e]	6	8.04[e]	8	21	22	12.96[e]	13	13
Ile	3.95[e]	4	7.84[e]	8	5.66[e]	6	18	18	11.85[e]	12	12
Leu	4.88	5	4.62	5	5.66	6	16	16	9.52	10	10
Tyr	14.81	15	5.04	5	7.57	8	28	28	16.91	17	20
Phe	3.96	4	3.03	3	3.05	3	10	10	6.92	7	7
Trp		2[f]		1[f]		0[f]	3	3		3[f]	3
Total		120		85		110	315	318		202	205

[a] The data are averages of duplicate analyses of 24-, 48-, and 72-hr acid hydrolysates. Calculations were made on the basis of 11.00, 6.00, 11.00, and 17.0 residues of Ala per molecule of fragment F_{III}, F_I, F_{II}, and F_{III-I}, respectively. [b] Total of integral values. [c] Taken from Table IV. [d] Extrapolated to zero time. [e] Equivalent to 72-hr value. [f] Taken from Table V.

each fragment (Figure 2). During gel filtration in 6 M guanidine hydrochloride, each fragment eluted as a single and symmetrical peak. Molecular weights of fragments F_I, F_{II}, and F_{III} obtained by these physical methods are summarized in Table I, together with minimum molecular weights calculated from the amino acid compositions (Table VII). There appears to be good agreement between values obtained by physical and chemical methods except for fragment F_{II}. As shown in Figure 2, this peptide appeared to be slightly larger in size than F_{III} whereas on the basis of the amino acid composition its minimum molecular weight was smaller than that of F_{III}. Sequence analysis later indicated that there was no significant error in the amino acid composition, and the molecular weight of fragment F_{II} calculated from the final sequence was 11,830. Hence this inversion in electrophoretic mobilities observed for fragments F_{II} and F_{III} may suggest that in the presence of sodium dodecyl sulfate, their mobilities depend not only on molecular size but also on amino acid composition.

Amino-Terminal Residues. End-group analyses provided final evidence for homogeneity of each fragment and facilitated the alignment of these fragments in the original protein molecule. By FDNB analysis, the only detectable products from fragments F_I, F_{II}, and F_{III} were DNP-valine (0.79 mole/9100 g), DNP-serine (0.70 mole/11,700 g), and DNP-leucine (or isoleucine) (0.85 mole/13,500 g), respectively. Essentially the same results were obtained by the cyanate method (Table II) which established the amino-terminal residue of F_{III} as isoleucine. In the case of fragment F_{II}, the cyanate method yielded a small but significant amount of amino-terminal proline (0.15 mole/11,700 g) (which is difficult to identify by the FDNB method), together with the major amino-terminal serine. Sequenator analysis (Figure 4) revealed that this amino-terminal proline residue was derived by cleavage at the third and the ninth residues of the major sequence of the fragment. Subsequent structural data (Figure 4) indicated that an aspartic acid–proline bond was cleaved in each case, probably due to the 75% trifluoroacetic acid employed in the cyanogen bromide reaction (Piszkiewicz *et al.*, 1970).

Sequenator analyses of each fragment are summarized in Figure 4. F_{III} corresponds to the amino terminus of the whole protein. Extended degradations were facilitated by the the optimal size and physical properties of the fragments. The amino-terminal sequence of F_{III-I} is identical with that of F_{III} (and the whole protein). Thus F_I is at the carboxyl terminus of F_{III-I}.

Carboxyl-Terminal Residues. The carboxyl end group of fragments F_I and F_{III} was assumed to be homoserine since each contained a single residue of homoserine per molecule (Table VII). By digestion with carboxypeptidase B at 37° for 6 hr (0.86 mole/11,700 g), the carboxyl terminus of fragment F_{II} was found to be lysine. This is identical with the carboxyl-terminal residue of the whole molecule.

From the results of these end-group analyses, it follows that the order of the three fragments in the original molecule is F_{III}-F_I-F_{II}.

Amino Acid Composition. Amino acid compositions of the three fragments, F_I, F_{II}, and F_{III}, are shown in Table VII. Tryptophan contents are taken from Table V which includes the results obtained by four different determinations. The sum of amino acid compositions of the three fragments approximates the composition of the whole protein, indicating that the amino acid residues in the original molecule are

well recovered in these fragments. From each amino acid composition, molecular weights were calculated to be 9100, 11,700, and 13,500 for fragments F_I, F_{II}, and F_{III}, respectively. The sum of these values also agrees closely with the molecular weight of the whole protein (Table I).

Fragment F_I consisted of approximately 85 amino acid residues. It was characterized by a low content of basic residues and high content of potentially acidic residues. In fact, fragment F_I was highly acidic and quite soluble above pH 5 but it tended to aggregate and precipitate at high concentrations below pH 5. This property explains its mobility on Bio-Gel P-100 (Figure 5).

Fragment F_{II} consisted of approximately 110 amino acid residues and was characterized by a relatively high content of basic residues. Low absorbancy (280 nm) of the fragment (Figure 5) is explained by the absence of tryptophan.

Fragment F_{III} consisted of approximately 120 amino acid residues with a high content of tyrosine. Two of the three tryptophan residues were found in this fragment. The presence of these aromatic residues resulted in the highest absorbancy and the lowest solubility of the three principal fragments. However, this property facilitated the separation of the fragment from fragment F_I by differential precipitation in 2 M urea.

Fragment F_{III-I} was isolated in only 6% yield, suggesting that it was a minor component. Its identity as a fragment that overlapped F_{III} and F_I was established by molecular weight comparisons (Table I), end-group data (Figure 4), and amino acid compositions (Table VII). The fragment contained two homoserine residues per molecule, indicating that one methionine residue cleaved normally whereas the other remained as an internal homoseryl residue. The alignment of fragment F_{III} at the amino terminus of the protein indicates that the carboxyl-terminal methionine in F_{III} is not completely cleaved and that about 6% occurs as a homoseryl residue joining F_{III} to F_I.

Discussion

Contrary to expectation, none of several commercial preparations of thermolysin so far tested showed satisfactory homogeneity by end-group analysis or disc gel electrophoresis in the presence of sodium dodecyl sulfate. Recrystallization, as employed by other workers (Endo, 1962; Mätsubara, 1970; Ohta *et al.*, 1966; Latt *et al.*, 1969), yielded a heterogeneous product after salt removal by dialysis. Recrystallization of the enzyme as the apo form in the presence of 1,10-phenanthroline (Latt *et al.*, 1969) failed to yield a protein of sufficient homogeneity for sequence analysis. It was finally found that a homogeneous preparation with a single amino-terminal isoleucine could be obtained by denaturation with trichloroacetic acid immediately after recrystallization (50% yield) or after gel filtration (80% yield). Yellowish pigment and autolysis products were present in most commercial preparations; the former was difficult to remove by recrystallization but was completely removed by gel filtration. The use of trichloroacetic acid had some disadvantages in that the enzyme was irreversibly inactivated and was insoluble at neutral pH. Nonetheless, the molecular weight and amino acid composition of the recovered protein were almost identical with those obtained by Ohta *et al.* (1966) after recrystallization.

Studies on the purified protein revealed a single polypeptide chain with a molecular weight of approximately 34,600. This value should be corrected to 34,800 for native enzyme by including 1 g-atom of zinc and 3 g-atoms of calcium per molecule (Latt *et al.*, 1969). Apparent differences between molecular weights estimated by physical and chemical methods are assumed to result from inherent errors in the physical methods since all attempts to detect nonpeptide components gave negative results. The ultraviolet absorption spectrum (Ohta *et al.*, 1966) indicates that thermolysin has no nucleotide component and the present study shows that it is not a glycoprotein.

Amino acid analysis of the purified protein disclosed the presence of two methionine residues and the complete absence of cysteine or cystine, in agreement with the data of Ohta *et al.* (1966). However, in contrast to their data, we found only three tryptophan residues instead of five using four different methods of analysis. This value is consistent with the data from the cyanogen bromide fragments. Thermolysin is characterized by an abundance of hydrophobic amino acid residues (99 of the 316 residues are Val, Met, Ile, Leu, Tyr, Phe or Trp). Sequence analysis is facilitated by the presence of but two methionine residues which should yield three fragments by cyanogen bromide cleavage.

In attempting cleavage with cyanogen bromide it was found that the purified enzyme was insoluble in the usual solvent, 75% formic acid. The protein did dissolve in 75% trifluoroacetic acid but some destruction of tryptophan may occur in this solvent (Nomoto *et al.*, 1969). These strongly acidic conditions may be responsible for the minor cleavage of aspartic acid–proline bonds near the amino terminus of fragment F_{II} (Figure 4). In any event, as predicted, three major fragments were obtained by cleavage with cyanogen bromide. The sum of the molecular weights of these fragments and the sum of their separate amino acid compositions agree closely with those of the whole protein.

Both methionine residues were completely converted to homoserine but one homoseryl bond (residue 120) was not completely cleaved. An "overlap peptide," fragment F_{III-I}, was found (in 7% yield) having an amino-terminal sequence, molecular weight, and amino acid composition consistent with a homoseryl residue (120) at the carboxyl terminus of F_{III} remaining attached to the amino-terminal serine (121) of F_I. This "overlap peptide" was useful in aligning the three fragments. Taken together with the absence of homoserine in F_{II} and the demonstration of carboxyl-terminal lysine in both F_{II} and the whole protein, the linear sequence of the three fragments was F_{III}-F_I-F_{II} as shown in Figure 4.

Thus it is possible at this level of sequence analysis to place 127 of the 316 residues in sequence and to align the cyanogen bromide fragments unambiguously in the thermolysin molecule. The remaining structure has been reported in a preliminary communication (Titani *et al.*, 1972) and its proof is detailed in accompanying papers. The orientation of this structure in the three-dimensional architecture of the enzyme is the subject of a separate communication (Matthews *et al.*, 1972a) but it is relevant to the present discussion that the alignment of the fragments was clearly confirmed by the X-ray data.

Acknowledgments

The authors are indebted to Mr. R. R. Granberg for his assistance in the amino acid analyses and to Dr. S. Hakomori for his assistance in the carbohydrate analyses.

References

Ambler, R. P. (1965), *Biochem. J. 96*, 32P.
Ambler, R. P. (1967), *Methods Enzymol. 11*, 436.

Bargetzi, J.-P., Thompson, E. O. P., Sampath Kumar, K. S. V., Walsh, K. A., and Neurath, H. (1964), *J. Biol. Chem. 239*, 3767.
Edman, P., and Begg, G. (1967), *Eur. J. Biochem. 1*, 80.
Endo, S. (1962), *Hakko Kogaku Zasshi 40*, 346.
Feder, J. (1968), *Biochem. Biophys. Res. Commun. 32*, 326.
Feder, J., Garrett, L. R., and Wildi, B. S. (1971), *Biochemistry 10*, 4552.
Fish, W. W., Mann, K. G., and Tanford, C. (1969), *J. Biol. Chem. 244*, 4989.
Fraenkel-Conrat, H., Harris, J. I., and Levy, A. L. (1955), *Methods Biochem. Anal. 2*, 359.
Gatt, R., and Berman, E. R. (1966), *Anal. Biochem. 15*, 167.
Goodwin, T. W., and Morton, R. A. (1946), *Biochem. J. 40*, 628.
Gross, E., and Witkop, B. (1962), *J. Biol. Chem. 237*, 1856.
Hermodson, M. A., Ericsson, L. H., and Walsh, K. A. (1970), *Fed. Proc., Fed. Amer. Soc. Exp. Biol. 29*, 728.
Hirs, C. H. W. (1967), *Methods Enzymol. 11*, 411.
Katz, A. M., Dreyer, W. J., and Anfinsen, C. B. (1959), *J. Biol. Chem. 234*, 2897.
Latt, S. A., Holmquist, B., and Vallee, B. L. (1969), *Biochem. Biophys. Res. Commun. 37*, 333.
Liu, T.-Y., and Chang, Y. H. (1971), *J. Biol. Chem. 246*, 2842.
Matsubara, H. (1970), *Methods Enzymol. 19*, 642.
Matsubara, H., and Feder, J. (1971), The Enzymes, Vol. 3, 3rd ed, Boyer, P. D., Ed., New York, N. Y., Academic, 721.
Matthews, B. W., Colman, P. M., Jansonius, J. N., Titani, K., Walsh, K. A., Neurath, H., and Schoenborn, B. P. (1972a), *Nature (London)* (in press).
Matthews, B. W., Jansonius, J. N., Colman, P. M., and Schoenborn, B. P. (1972b), *Nature (London)* (in press).
Moore, S., and Stein, W. H. (1963), *Methods Enzymol. 6*, 819.
Neurath, H., and Bradshaw, R. A. (1970), *Accounts Chem. Res. 3*, 249.
Noltmann, E. A., Mahowald, T. A., and Kuby, S. A. (1962), *J. Biol. Chem. 237*, 1146.
Nomoto, M., Srinivasan, N. G., Bradshaw, R. A., Wade, R. D., and Neurath, H. (1969), *Biochemistry 8*, 2755.
Ohta, Y., Ogura, Y., and Wada, A. (1966), *J. Biol. Chem. 241*, 5919.
Pétra, P. H. (1970), *Methods Enzymol. 19*, 460.
Piszkiewicz, D., Landon, M., and Smith, E. L. (1970), *Biochem. Biophys. Res. Commun. 40*, 1173.
Sampath Kumar, K. S. V., Clegg, J. B., and Walsh, K. A. (1964), *Biochemistry 3*, 1728.
Sanger, F. (1945), *Biochem. J. 39*, 507.
Spackman, D. H., Stein, W. H., and Moore, S. (1958), *Anal. Chem. 30*, 1190.
Stark, G. R. (1967), *Methods Enzymol. 11*, 125.
Titani, K., Hermodson, M. A., Ericsson, L. H., Walsh, K. A., and Neurath, H. (1972), *Nature (London)* (in press).
Titani, K., Hermodson, M. A., Walsh, K. A., and Neurath, H. (1971), *Fed. Proc., Fed. Amer. Soc. Exp. Biol. 30*, 1241.
Warren, L. (1959), *J. Biol. Chem. 234*, 1971.
Weber, K., and Osborn, M. (1969), *J. Biol. Chem. 244*, 4406.
Winzler, R. J. (1955), *Methods Biochem. Anal. 2*, 279.
Yang, H., and Hakomori, S. (1971), *J. Biol. Chem. 246*, 1192.
Zacharius, R. M., Zell, T. E., Morrison, J. H., and Woodlock, J. J. (1969), *Anal. Biochem. 30*, 148.

Cross-β Protein Structures. I. Insulin Fibrils†

Michael J. Burke‡ and Malcolm A. Rougvie*

ABSTRACT: Structural studies have been carried out upon the fibrils which are produced when insulin is heated in acid. Electron microscopy and low-angle X-ray diffraction data suggest that the fibrils have a uniform cross-section with dimensions of about 29 by 47 Å. Infrared dichroism and wide-angle X-ray diffraction clearly demonstrate the cross-β structure of the fibrils. Ultraviolet circular dichroism and optical rotatory dispersion spectra are consistent with a β conformation. A detailed structure is suggested in which individual insulin molecules are stacked in layers 4.7 Å thick in the direction of the fibril axis, with one molecule per layer. Interaction of the extended chains in adjacent layers is believed to be of the parallel pleated-sheet type.

Under appropriate solvent conditions, the globular protein insulin polymerizes to form submicroscopic fibrils. The chemistry of the polymerization process has been studied extensively by Waugh (*e.g.*, 1957) who found that it could be reversed yielding biologically active, crystallizable insulin. Because of the reversibility of the process he suggested as early as 1944 (Waugh, 1944) that the fibrils are arrays of only slightly distorted globular insulin molecules. Koltun *et al.* (1954) supported this proposal on the basis of X-ray diffraction analysis. Their analysis has been accepted by Reithel (1963) in a review of protein association, and more recently by other authors (Beaven *et al.*, 1969). However, primarily on the basis of infrared dichroism measurements, Ambrose and Elliott (1951) concluded that the fibrils are composed of polypeptide chains which are in the β conformation and which extend transverse to the fibril axis (*i.e.*, cross-β structure).

Other evidence related to the question is discussed briefly in the review by Klostermeyer and Humbel (1966). In this paper we provide evidence which affirms the essential correct-

† From the Department of Biochemistry and Biophysics, Iowa State University, Ames, Iowa 50010. *Received December 30, 1971.* This work was supported by Grant-in-Aid GM-11127 from the National Institutes of Health of the Department of Health, Education, and Welfare.

‡ Present address: Department of Chemistry, University of Minnesota, Minneapolis, Minn. 55455.