The Purification and Properties of Fragments of Trypsinogen Obtained by Cyanogen Bromide Cleavage*

T. HOFMANN†

From the Department of Biochemistry, University of Washington, Seattle

Received October 15, 1963

The peptide chain of trypsinogen was cleaved at the carboxyl side of the two methionine residues by treatment with cyanogen bromide. The products were converted to *S*-β-aminoethyl derivatives. The three fragments resulting from the specific cleavage were purified and analyzed for their amino acid composition and N-terminal residues. Going from the N-terminal to the C-terminal residue the fragments consist of 92, 74, and 63 amino acid residues. On the basis of the composition and knowledge of partial sequences it has been possible to allocate all the known peptides to one of the three fragments. The most interesting fact which emerges is that, as in chymotrypsin, the histidine residues are located in the N-terminal region and the active-center serine is in the C-terminal region. There is at least one peptide bond involving a leucine or isoleucine and an unknown residue which is hydrolyzed nonspecifically under the acidic conditions used for the cyanogen bromide reaction.

Bovine pancreatic trypsinogen is composed of a single polypeptide chain of 229 amino acid residues cross-linked by six disulfide bonds. Walsh *et al.* (1962a,b) have described amino acid sequences found in tryptic and chymotryptic digests of *S*-sulfotrypsinogen, while reports from Mikes *et al.* (1961, 1962) and Tomasek *et al.* (1963) deal with the structure of peptides of peptic and tryptic hydrolysates of diisopropylphosphoryl trypsin and of a tryptic digest of *S*-sulfotrypsinogen.

However, difficulties were encountered both in finding sufficient overlapping peptides to piece the whole sequence together and in accounting for all the residues of the molecule because of the presence of insoluble fractions in the digests. In order to overcome at least some of these difficulties and to simplify the elucidation of the unknown regions of the molecule, we applied the method for the specific cleavage of peptide bonds involving the carboxyl group of methionine residues which had been introduced by Gross and Witkop (1961) and was applied by these authors to ribonuclease (Gross and Witkop, 1962), and more recently by Edmundson (1963) to sperm whale myoglobin. The reagent, cyanogen bromide, converts methionine residues into C-terminal homoserine and its lactone. Trypsinogen contains two methionine residues and accordingly should give rise to three fragments after cyanogen bromide cleavage and fission of the six disulfide bridges. In what follows these fragments will be defined, going from the amino-terminal residue to the carboxy-terminal residue, as the A, B, and C fragments. According to expectation the A and the B fragments will have homoserine or its lactone or both as C-terminal residues.

This paper describes the isolation, purification, and chemical analysis of the three fragments, which are formed by cyanogen bromide cleavage of trypsinogen, reduction of the disulfide bridges, and their conversion of *S*-β-aminoethylcysteine residues (Lindley, 1956). As a consequence of this work it has been possible to allocate all peptides of known sequence to their respective fragments and to obtain at least the backbone of the linear amino acid sequence of the trypsinogen molecule.

* This investigation was supported by a U. S. Public Health Service grant (GM 04617) and a Research Travel Grant from the Wellcome Trust, London. A preliminary report was given (Hofmann *et al.*, 1963).

† Permanent address: Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada.

EXPERIMENTAL

Materials.—Bovine pancreatic trypsinogen was obtained from Worthington Biochemical Corp. as a 1-× crystallized preparation containing 50% $MgSO_4$ (lots 619 and 707). It contained less than 0.5% active trypsin. After autocatalytic activation it had a molecular activity of 1300 μmoles BAEE[1]/min × μmole enzyme at 25°. This corresponds to the highest published specific activity of trypsin k_0 = 0.36 μmole/min × mg protein-N (Green and Neurath, 1953). Nevertheless, as Table I shows, the protein was, from the point of view of end groups, not homogeneous, since it contained, in addition to the expected N-terminal valine, significant amounts of aspartic acid, serine, and phenylalanine. (The same was also true of other trypsinogen samples of full potential activity.) It is probable that the extra N-terminal amino acids are not due to contaminants but arise from splits in the peptide chain. As will be shown, these splits have considerably complicated the problem of purification of the fragments resulting from treatment with cyanogen bromide. Before use the trypsinogen was dissolved in 0.01 M HCl and dialyzed against 0.001 M HCl until free of sulfate.

TABLE I
N-TERMINAL AMINO ACIDS OF TRYPSINOGEN, HCl-TREATED TRYPSINOGEN, AND CYANOGEN BROMIDE–TREATED TRYPSINOGEN (VALINE = 1.0)

DNP-Amino Acids	Trypsinogen	HCl Treated	CNBr Treated
DNP-aspartic acid[a]	0.07	0.06	(0.26)[b]
DNP-serine	0.05	0.08	(0.32)[b]
DNP-leucine and/or DNP-isoleucine	0	0.23	0.3
DNP-phenylalanine	0.11	0.09	0.92
DNP-valine	1.0	1.0	1.0
DNP-threonine	—	—	(0.18)[b]

[a] The values are uncorrected for hydrolytic and chromatographic losses. [b] These high values were obtained only when the dinitrophenylation reaction was carried out for prolonged periods (up to 18 hours).

Trypsinogen concentrations were determined by

[1] The following abbreviations are used here: FDNB = 1-fluoro-2,4-dinitrobenzene; DNP = 2,4-dinitrophenyl-; DFP = diisopropylphosphofluoridate; BAEE = benzoyl-L-arginine ethyl ester; DEAE = diethylaminoethyl-; CM = carboxymethyl-.

measuring the optical densities at 280 mμ, using a value $E^{1\%}_{1cm}$ = 13.9 (Davie and Neurath, 1955).

Cyanogen bromide and β-aminoethylbromide·HBr were products of Eastman Organic Chemicals Div. and were used without further purification.

Sephadex dextran gels, grades 25 and 75, were obtained from Pharmacia, Uppsala, Sweden.

CM-cellulose (0.8 meq/g) and *DEAE-cellulose* (0.9 meq/g) were purchased from the Brown Co. Before use they were washed with 0.5 M NaOH–0.1 M NaCl followed by water, 0.1 N HCl, and water. Finally they were equilibrated with the appropriate buffers.

Starch, hydrolyzed for starch-gel electrophoresis, was obtained from the Connaught Medical Research Laboratories, University of Toronto, Canada.

Amino Acid Analysis.—The hydrolysis of protein fractions was carried out for 20–90 hours at 106° with glass-distilled constant-boiling HCl in tubes sealed under vacuum. The acid was removed in a rotary evaporator and the residue dissolved in the *p*H 2.2 citrate buffer of Moore *et al.* (1958). Amino acid analyses were performed with a Beckman/Spinco automatic amino acid analyzer, Model 120, according to Spackman *et al.* (1958) and Moore *et al.* (1958).

The integration constants for homoserine and homoserine lactone were determined on purified samples. With the analyzer for which the factors for leucine and lysine are 21.8 and 24.9, respectively, the following values were obtained: homoserine, 18.5; lactone, 14.2. These values are in agreement with those of Neumann *et al.*, (1962). The position on the columns of Spackman *et al.* (1958) agreed with those reported by Gundlach *et al.* (1959). The integration constant of *S*-β-aminoethylcysteine was determined in a hydrolysate of a chromatographically pure sample of *S*-β-aminoethylglutathione. The *S*-β-aminoethylcysteine content of the hydrolysate was assumed to be the same as the glutamic acid and glycine content. The constant calculated was 22.8 (lysine = 24.9); this constant includes a factor correcting for any destruction which may have occurred during hydrolysis. As reported (Raftery and Cole, 1963), the position of the peak on the 15-cm column is between lysine and histidine. In peptide fractions where one of the three amino acids was present in small amounts, it was found advantageous to increase the column height to 19 cm in order to obtain complete separation of the three amino acids.

Tryptophan was determined by the method of Spies and Chambers (1949), as modified by Harrison and Hofmann (1961).

Terminal Amino Acids.—N-terminal amino acids were determined by the FDNB method of Sanger (1945) as described by Porter (1950). In some experiments more vigorous conditions were used; for some samples the time of coupling was increased from 2 hours to 18 hours, while others were coupled in 8 M urea.

For those samples which still possessed potential tryptic activity which might be revealed during the dinitrophenylation reaction, DFP or benzamidine (Mares-Guia and Shaw, 1963) were used as inhibitors.

C-terminal amino acids were determined with carboxypeptidase A (Harris, 1955). The author is most grateful to Dr. K. A. Walsh for these analyses.

Starch-Gel Electrophoresis.—This was done by the method of Smithies (1955). All gels were made up in 6.5 M urea buffers of *p*H 3.0 and 8.6, as the fragments resulting from cyanogen bromide treatment did not migrate in the absence of urea. The gels were prepared by the standard technique and had the following composition: *p*H *3.0*, 10 g starch, 60 ml 0.05 M formic acid which had been adjusted to *p*H 3.0 with 8 N NaOH, 32 g urea; *p*H *8.6*, 10 g starch, 60 ml 0.076 M Tris–0.005 M citric acid, 32 g urea. They were allowed to solidify overnight at 25° before use. The samples, usually the dry proteins, were dissolved in the urea-containing buffers at a concentration of about 10–20 mg/ml, taken up on a small filter paper strip, and inserted into the gel. Experiments were made at approximately 10 v/cm gel, for 2–3 hours.

TABLE II
AMINO ACID COMPOSITION OF TRYPSINOGEN AND CYANOGEN BROMIDE-TREATED TRYPSINOGEN, BEFORE AND AFTER CONVERSION TO *S*-β-AMINOETHYL DERIVATIVES

Amino Acid	Number of Residues/Mole Trypsinogen (MW 24,000)		
	Trypsinogen[a]	CN-Br Trypsinogen	Aminoethyl Derivative
Aspartic acid	26	26.4	25.9
Threonine[b]	10	9.8	9.8
Serine[b]	33	33.	33.
Glutamic acid	14	14.2	14.2
Proline	9	8.4	8.3
Glycine	25	26.	25.
Alanine	14	14.	14.
Half-cystine	12	10.8	0
Valine[c]	18	18.	18.
Methionine	2	0.17	0.09
Isoleucine[c]	15	15.	15.
Leucine	14	14.3	13.5
Tyrosine	10	9.2	9.1
Phenylalanine	3	3.04	2.65
Homoserine lactone + homoserine		1.75	1.78
Lysine	15	15.	15.4
Histidine	3	2.9	2.4
Arginine	2	2.1	2.2
Aminoethylcysteine			11.4
Tryptophan	4	Not determined	

[a] These values were determined on the trypsinogen samples used for this work and agree with those determined by K. A. Walsh (unpublished). All values are calculated on the assumption that there are 14 alanine residues in trypsinogen (K. A. Walsh). [b] Extrapolated to zero time of hydrolysis. [c] Value after 90 hours' hydrolysis.

Reaction with Cyanogen Bromide.—Gross and Witkop (1962) and Edmundson (1963) carried out the cleavage with CNBr in 0.1 N HCl for 24 hours at room temperature. When these conditions were used with trypsinogen only about 50% of the methionine residues reacted. When, however, the temperature was raised to 30° and the time was increased to 30 hours, methionine was converted to homoserine and its lactone to the extent of 92–95%, as measured by the residual methionine (see Table II).

In a typical experiment, trypsinogen (2.4 g = 100 μmoles) in 230 ml 0.2 M HCl was mixed with 800 mg CNBr (7.5 mmoles) in 230 ml H_2O. The solution, which remained clear during the whole reaction period, was kept at 30° for 30 hours. After freeze-drying the material was obtained quantitatively as a white powder. In order to remove the excess reagents completely, this was dissolved in about 200 ml H_2O and freeze-dried again.

Preliminary results indicated that during the reaction there was a progressive loss in the activatability of the trypsinogen which was approximately parallel to the cleavage of the methionine bonds.

N-Terminal Analysis of the Reaction Mixture.—Trypsinogen contains two methionine residues which are followed in the amino acid sequence by leucine (K. A. Walsh, personal communication) and phenyl-

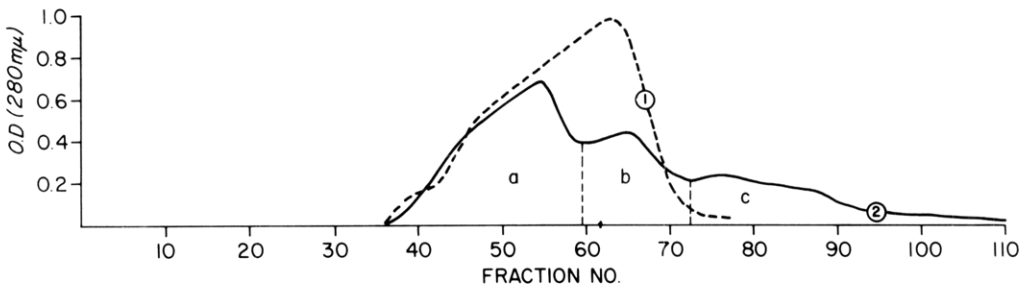

FIG. 1.—Chromatography of *S*-β-aminoethyl derivatives of ① trypsinogen and ② trypsinogen after treatment in 0.1 N HCl at 30° for 30 hours on Sephadex G-75, 60–100 mesh, 5 × 100 cm in 0.2 M acetic acid, flow rate 35 ml/hour, fraction size 17.5 ml.

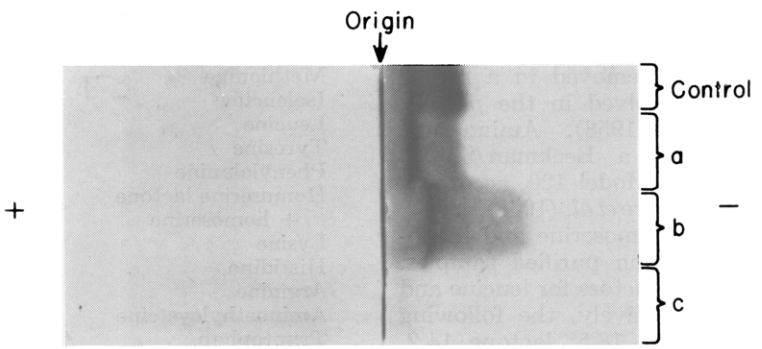

FIG. 2.—Photograph of starch-gel electrophoresis patterns of *S*-β-aminoethyl derivatives of trypsinogen before (control) and after (a,b,c) HCl treatment at *p*H 8.6 in 6.5 M urea; a, b, and c correspond to the peaks of Fig. 1.

alanine, respectively (Walsh *et al.*, 1962b). These two residues should appear as new N-terminal amino acids during the CNBr cleavage. When the reaction mixture was analyzed for N-terminal residues by the Sanger technique (Table I), only phenylalanine was found in amounts equivalent to the valine which is N-terminal in trypsinogen; aspartic acid, serine, threonine, and leucine were found in smaller but significant quantities. The first three of these appeared only after prolonged dinitrophenylation and were present in variable amounts; they probably arose from peptide bonds which were split during the dinitrophenylation reaction. However, the N-terminal leucine or isoleucine residues appeared to have been liberated as a consequence of the CNBr reaction, although for unknown reasons only in low yield. Since, however, subsequent control experiments from which CNBr was omitted also yielded N-terminal leucine and/or isoleucine residues in similar amounts they must have arisen mostly from a nonspecific hydrolytic split of the peptide chain under the acidic reaction conditions. This is also borne out by the fact that the purified B-fragment yields DNP-leucine only in very poor yield (see below). The possibility remained, therefore, that the cleavage of one of the methionine bonds led to the appearance of an N-terminal residue which would not be easily detected by the FDNB technique. Such residues could be glycine, proline, half-cystine, and tryptophan. All of these residues can be ruled out on experimental grounds. The mild conditions of hydrolysis recommended by Porter (1950) should have revealed DNP-glycine and DNP-proline. Dinitrophenylation of oxidized, cyanogen bromide–treated trypsinogen should have given rise to DNP-cysteic acid arising from N-terminal half-cystine. Tryptophan can be ruled out because all four residues are accounted for in known amino acid sequences (Walsh *et al.*, 1962b; Tomasek *et al.*, 1963). None of these contains methionine.

The FDNB reaction was also carried out under conditions more vigorous than those recommended by Porter (1950). The reaction time was increased from 2 hours to 16 hours; the reaction was also carried out in 8 M urea where the DNP-protein remained soluble, but no additional N-terminal residues were found and no increase in the yield of DNP-leucine was obtained. It must be concluded that although leucine was probably liberated as new N-terminal residue during the CNBr reaction it is nonreactive toward the FDNB reagent.

Nonspecific Peptide Bond Cleavage During CNBr Reaction.—It was mentioned earlier that during the CNBr treatment approximately 0.3 residue of N-terminal leucine or isoleucine per N-terminal valine was liberated. A control experiment in which trypsinogen was incubated in 0.1 N HCl for 30 hours at 30° in the absence of CNBr revealed the appearance of about 0.23 residue of N-terminal leucine or isoleucine (Table I), indicating that the acidic conditions are responsible for the hydrolytic breakage of a particularly labile peptide bond. The breaking of this bond leads to a breakdown in the molecule as is shown in Figure 1, which represents the result of an experiment in which trypsinogen which had been treated with 0.1 N HCl as described above was converted to *S*-β-aminoethyltrypsinogen (see below) and then chromatographed on a Sephadex G-75 column (curve 2). A sample of trypsinogen which had not been treated with 0.1 N HCl but had been converted to *S*-aminoethyltrypsinogen was chromatographed on the same column for comparison (curve 1). Figure 1 clearly demonstrates the appearance of at least two and possibly three new peaks which move more slowly than the main peak. The appearance of new fragments can also be demonstrated by starch-gel electrophoresis (Figure 2). The three peaks, (a), (b), and (c) (Fig. 1, curve 2) were compared with *S*-β-aminoethyltrypsinogen which had not been exposed to acidic conditions (Fig. 1, curve 1).

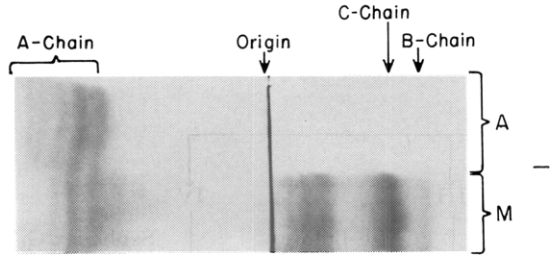

FIG. 3.—Photograph of starch-gel electrophoresis patterns of mixture of fragments of trypsinogen obtained by cyanogen bromide treatment (*S*-β-aminoethyl derivatives), M, and purified A fragment, A, at *p*H 8.6 in 6.5 M urea.

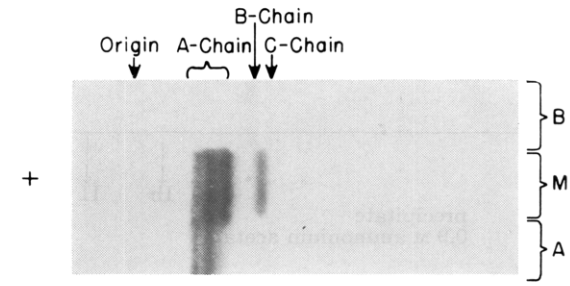

FIG. 4.—Photograph of starch-gel electrophoresis patterns of mixture of fragments of trypsinogen obtained by cyanogen bromide treatment (*S*-β-aminoethyl derivatives), M, and purified A and B fragments, A, B at *p*H 3,0 in 6.5 M urea.

Comparison indicates that peak (a) is identical with the untreated trypsinogen while peak (b) is different and moves further toward the cathode because it is more basic and of lower molecular weight as is also shown by the Sephadex chromatography. Peak (c) shows only a small amount of material moving toward the cathode and a weak streaking component moving toward the anode. The tentative sequence of trypsinogen which has been proposed (Walsh *et al.*, 1964) shows that if the half-cystine residues are converted to the positively charged *S*-β-aminoethylcysteines, the N-terminal region of the molecule through to the first methionine residue is relatively rich in free carboxyl side chains and deficient in positively charged side chains, while the remainder of the molecule contains only five free carboxyl groups and twenty-two positively charged side chains. The only region which can give rise to sizable peptide fragments with a net negative charge at *p*H 8.6 is thus the N-terminal region. It can be concluded, therefore, that peak (c) originates from this region and that the nonspecific split probably occurs in the N-terminal part of the molecule. This is also borne out by the amino acid analysis of the three peaks: peak (a) has the composition of trypsinogen, while peak (b), although similar, gives low values for aspartic acid and histidine, the two amino acids which predominate in the A fragment; peak (c), on the other hand, is rich in these two amino acids, relative to the others.

Reduction and Substitution of Disulfide Bridges.—Since trypsinogen contains six disulfide bridges, the cleavage of the peptide chain at the two methionine residues does not necessarily lead to the formation of separable fragments. Thus, chromatography on a Sephadex G-75 column indicated that the material was still one component, and that the three fragments were held together by the disulfide bridges. In order to separate the fragments, the cleavage of these linkages and the conversion of the cysteine side chains into stable, nonreactive side chains became essential. The method of Cavallini *et al.* (1955), which converts cysteine into *S*-β-aminoethylcysteine, was chosen in preference to other methods, such as oxidation to cysteic acid, or conversion to *S*-sulfo or *S*-carboxymethyl derivatives, because the resulting fragments were readily soluble in slightly acid media, especially 0.2 M acetic acid, probably because of their strongly cationic character. It was hoped that this would reduce aggregation because of electrostatic repulsion.

Contrary to published reports on this method (Lindley, 1956; Tietze *et al.*, 1957) which indicate that a high *p*H is required for this reaction, it was found (K. A. Walsh, to be published) that virtually quantitative substitution could be obtained under the following conditions: The freeze-dried material obtained after the CNBr reaction (2.4 g, 100 μmoles) was dissolved in 250 ml of 8 M urea in 0.05 M Tris-HCl buffer, *p*H 8.0, and the solution was cooled to 10°. Ten ml of β-mercaptoethanol (150 mmoles, 120-fold excess) was added in order to reduce the disulfide bonds. The flask was flushed with N_2, sealed, and kept at 0° for 24 hours.

β-aminoethylbromide·HBr (160 g = 780 mmoles which is a 5-fold excess over sulfhydryl groups) was added, followed by sufficient 10 N NaOH to keep the *p*H between 7.8 and 8.0. The mixture was allowed to warm to room temperature, with periodic adjustment of the *p*H for about 1 hour. Thereafter the mixture was left at 37° for 5 hours and then in the cold room overnight. The total volume was about 400 ml. The protein material was then separated from the reagents by gel filtration on Sephadex G-25. Two portions of 200 ml each were applied to a column 5 × 100 cm and the material was eluted with 0.2 M acetic acid. The protein fractions were free of urea and inorganic ions and were freeze-dried. Yield was 2.24 g (93%). Amino acid analysis of this material showed complete conversion of the 12 half-cystine residues to *S*-β-aminoethylcysteine. The composition agrees closely with that of trypsinogen except that the values for histidine and phenylalanine are somewhat low (Table II).

If the material had been free from nonspecific peptide-bond splits, it should have contained only the three fragments produced by the CNBr cleavage. However, starch-gel electrophoresis in urea (Figs. 3 and 4) showed that the reaction mixture was considerably more complex. At *p*H 8.6 (Fig. 3), up to twelve bands can be clearly seen, about half of them migrating toward the anode, the other half toward the cathode. At *p*H 3.0 (Fig. 4), where the bands are more diffuse, there are at least five clearly discernible bands all moving cathodically. The difference in number of bands between the two *p*H values indicates that at least some of the bands in the *p*H 8.6 runs arose from partial deamidation owing to the acidic conditions of the cyanogen bromide reaction, and thus the same fragment could be represented by more than one band. The five bands in the *p*H 3.0 run would then represent the three expected fragments due to the methionine bond cleavage and two additional bands arising from the nonspecific hydrolytic split discussed above. However, the possibility cannot be excluded that the additional N-terminal groups present in the original trypsinogen preparation represent additional splits in the peptide chain which could give rise to small additional bands.

RESULTS

Purification of Fragments Obtained by Cyanogen Bromide Treatment.—Because of the presence of more

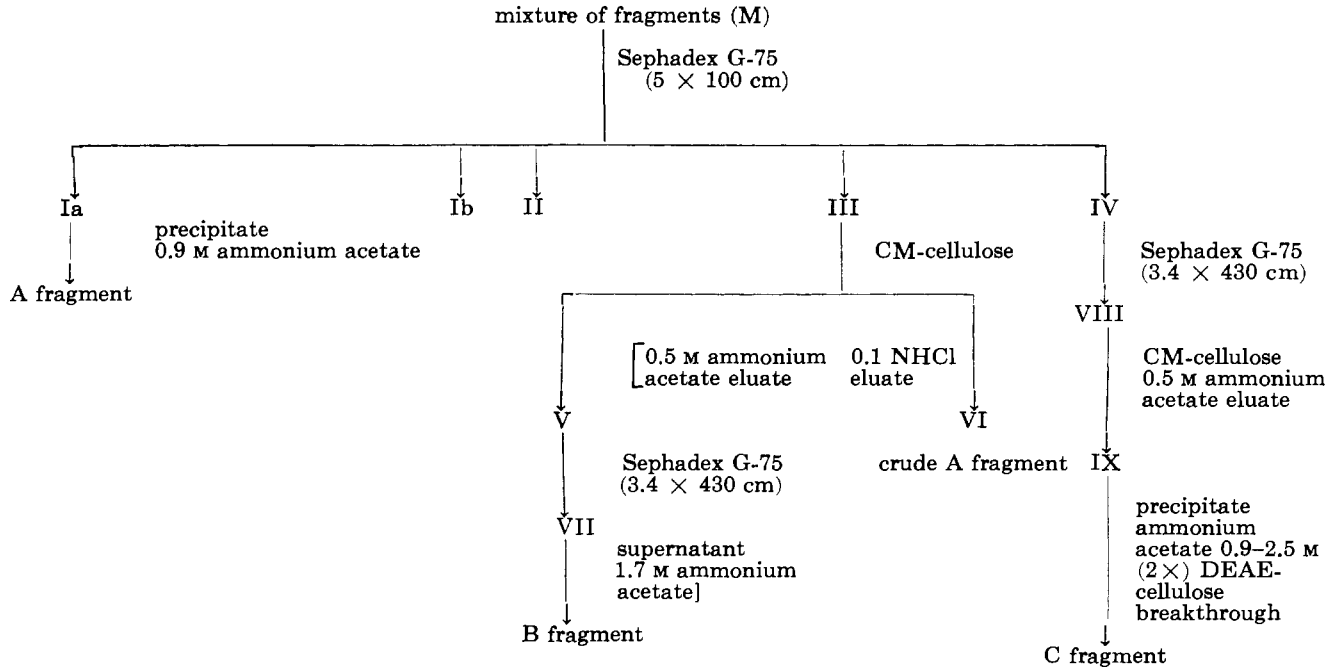

FIG. 5.—Flow sheet for the purification of the *S*-β-aminoethyl derivatives of cyanogen bromide–treated trypsinogen.

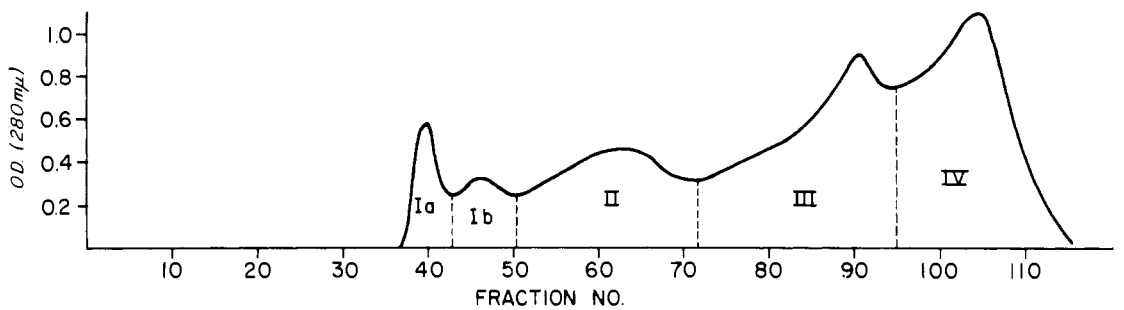

FIG. 6.—Chromatography of the mixture of *S*-β-aminoethyl derivatives of fragments obtained after cyanogen bromide treatment of trypsinogen on Sephadex G-75, 120–200 mesh, 5 × 100 cm, in 0.2 M acetic acid, flow rate 30 ml/hour, fraction size 15 ml.

than the three expected fragments, the purification problem was more complex than was originally expected. The method used is summarized on the flow sheet in Figure 5. Unfortunately, the yields of the individual fragments, while sufficient for analysis of their composition, terminal residues, and homogeneity, are not adequate for sequence analyses. Another method is therefore being developed for the preparation, in good yield, of fragments of lower purity, but sufficient for sequence determinations, and will be described elsewhere.

The reaction mixture after the cyanogen bromide treatment and conversion to the *S*-β-aminoethyl derivative was applied in five 400-mg quantities to a Sephadex G-75 column (5 × 100 cm). Before the column was poured, the Sephadex (120–200 mesh) was allowed to swell in 0.5 M NaCl for 24 hours and was then washed extensively with 0.2 M acetic acid. The material was eluted with 0.2 M acetic acid and emerged in five distinct peaks (Fig. 6). The fractions under each peak were pooled as indicated by the dotted lines and freeze-dried.

The following approximate yields were obtained: Fraction Ia, 80 mg; Ib, 800 mg; II, 300 mg; III, 800 mg; IV, 750 mg.

Determination of the amino acid composition and of the N-terminal residues was used to identify the various peaks. Peak fraction Ia appeared to consist of enriched A fragment and was used for its purification. Peak fraction Ib contained residual methionine and also some free cystine and predominantly valine as N-terminal residue, and was considered to consist mainly of intact and only partly degraded trypsinogen. The N-terminal residues of peak fraction II were mainly valine and leucine and thus must have come at least in part from the fragments arising from the nonspecific split discussed earlier. Peak fraction III was used for the preparation of the B fragment and peak fraction IV for the preparation of the C fragment.

A Fragment.—The material from peak fraction Ia

TABLE III
N-TERMINAL AMINO ACID RESIDUES IN PURIFIED FRAGMENTS OF TRYPSINOGEN OBTAINED AFTER CYANOGEN BROMIDE TREATMENT AND CONVERSION TO *S*-β-AMINOETHYL DERIVATIVES[a]

DNP-Amino Acid	A Fragment	B Fragment[b]	C Fragment
DNP-aspartic acid	0.04		
DNP-serine	Trace	0.03	0.04
DNP-leucine and/or DNP-isoleucine		0.09	0.03
DNP-phenylalanine	0.07		0.7
DNP-valine	0.7		0.03

[a] The values given are in moles/mole fragment. They are uncorrected for mechanical and hydrolytic losses.
[b] Dinitrophenylation carried out in 8 M urea for 16 hours.

TABLE IV
AMINO ACID COMPOSITION OF PURIFIED FRAGMENTS OF TRYPSINOGEN OBTAINED AFTER CYANOGEN BROMIDE TREATMENT AND CONVERSION TO S-β-AMINOETHYL DERIVATIVES

Amino Acid	Fragment A	Fragment B	Fragment C	A + B + C	Trypsinogen
	(Number of residues/mole)				
Aspartic acid	15.3	5.1	5.1	25.5	26
Threonine[a]	3.0	5.0	1.85	9.85	10
Serine[a]	12.	15.8	7.5	35.3	33
Glutamic acid	7.7	2.2	5.4	15.3	14
Proline	2.2	4.	1.95	8.15	9
Glycine	10.1	5.2	11.	26.3	25
Alanine	4.4	6.8	2.7	13.5	14
Valine[b]	9.5	2.2	6.1	17.8	18
Isoleucine[b]	6.4	5.0	3.0	14.4	15
Leucine	4.3	7.5	1.7	13.5	14
Tyrosine	4.8	2.0	2.85	9.65	10
Phenylalanine	2.	Trace	0.84	2.84	3
Homoserine lactone + homoserine	0.75	0.98	Trace	1.73	
Lysine	3.4	6.0	5.7	15.1	15
Histidine	2.5	0.03	Trace	2.53	3
Arginine	1.25	1.04	0.15	2.34	2
Aminoethylcysteine	2.8	3.4	4.3	10.5	12
Tryptophan	1.2	0.87	2.1	4.17	4
Integral number of residues	92.	73.	63.	228.	229

[a] Extrapolated to zero-time hydrolysis. [b] Value after 90 hours' hydrolysis.

(Fig. 6, about 80 mg) was dissolved in 7 ml 0.2 M acetic acid and brought to 0.9 M ammonium acetate by the addition of 0.7 ml of a 10 M solution. The precipitate which formed was centrifuged off, suspended in a few ml water, and freeze-dried. It was finally dried over NaOH pellets and concentrated H_2SO_4 to remove ammonium acetate. Yield: 65 mg of white fluffy material, soluble in 0.2 M acetic acid. This material was not quite homogeneous, as judged by starch-gel electrophoresis (Figs. 4 and 8) and by N-terminal analysis (Table III), valine contributing some 86% of the N-terminal residues. However, attempts to obtain further purification by chromatography on Sephadex gels and on cellulose columns, or by salt fractionations, failed to reduce the contaminating N-terminal groups. The analysis for C-terminal residues with carboxypeptidase showed the presence of homoserine and some other residues. Unfortunately, insufficient material was available for a time study of the liberation of C-terminal residues. The amino acid composition of this fragment is given in Table IV.

The presence of nearly stoichiometric amounts of homoserine and the fact that valine is N-terminal suggests that this fraction is the A fragment. This conclusion was supported by the isolation, in good yield, of the N-terminal peptide Val-Asp_4-Lys.Ileu.Val.Gly.-Gly.Tyr. (Walsh *et al.*, 1962a), from a chymotryptic digest of this fragment (K. A. Walsh and co-workers, to be published).

B Fragment.—The starting material for the isolation of this fragment was peak fraction III (Fig. 6). Amino acid analysis indicated that this material contained less than stoichiometric amounts of histidine and phenylalanine, but stoichiometric amounts of homoserine. As there was only a small amount of N-terminal valine, the material should contain mainly the B fragment. Its purification was judged by following the loss in histidine and phenylalanine. A considerable part of the impurities could be removed by absorbing the material on CM-cellulose and eluting with ammonium acetate.

The freeze-dried material from peak fraction III was applied in 100-mg portions, dissolved in about 5 ml 0.2 M acetic acid, to a column (3.4 × 18 cm) of CM-cellulose in 0.2 M acetic acid, at a rate of about 100 ml/hour. All the material was adsorbed and about 60–70% of the ultraviolet-absorbing material could be eluted by changing the eluent to 0.5 M ammonium acetate, *p*H 5.2 (peak V, Fig. 7). This material contained only small amounts of histidine and phenylalanine and was used for the further purification of the B fragment. The remaining 30–40% could be eluted with 0.1 N HCl after the column had been washed with water to remove all ammonium acetate (peak VI, Fig. 7). This fraction was found to be relatively rich in histidine, phenylalanine, and aspartic acid, and thus probably represented material from the A fragment. However, it was very heterogeneous and was therefore not used further. The fractions of peak V (Fig. 7) were freeze-dried and freed of ammonium acetate *in vacuo* over NaOH and H_2SO_4. The white powder (about 350 mg) was dissolved in 0.2 M acetic acid and chromatographed on a long Sephadex G-75 column (3.4 × 430 cm) in an effort to obtain a separation of the B and C fragments. The pattern obtained by elution with 0.2 M acetic acid is shown in Figure 7. Amino acid analyses showed that fraction VII contained the lowest amounts of histidine and phenylalanine, and a 1:1 ratio of arginine and homoserine. The slower fractions were mixtures of the B and C fragments with other contaminants as judged by amino acid analysis and N-terminal groups.

Peak VII, after freeze-drying, about 150 mg, was dissolved in 12 ml 0.2 M acetic acid and brought to 1.7 M ammonium acetate by adding 2.4 ml 10 M ammonium acetate. The precipitate was centrifuged off. To the supernatant were added 5.5 ml 10 M ammonium acetate, and after standing for 2 hours the precipitate was centrifuged off and dissolved in 0.2 M acetic acid. Freeze-drying, followed by drying over NaOH and H_2SO_4, gave a yellowish white powder. Yield: about 80 mg. This material represented the final preparation of B fragment. The amino acid composition (Table IV) shows that the contamination with histidine and phenylalanine-containing fragments is small. The N-terminal amino acids are given in

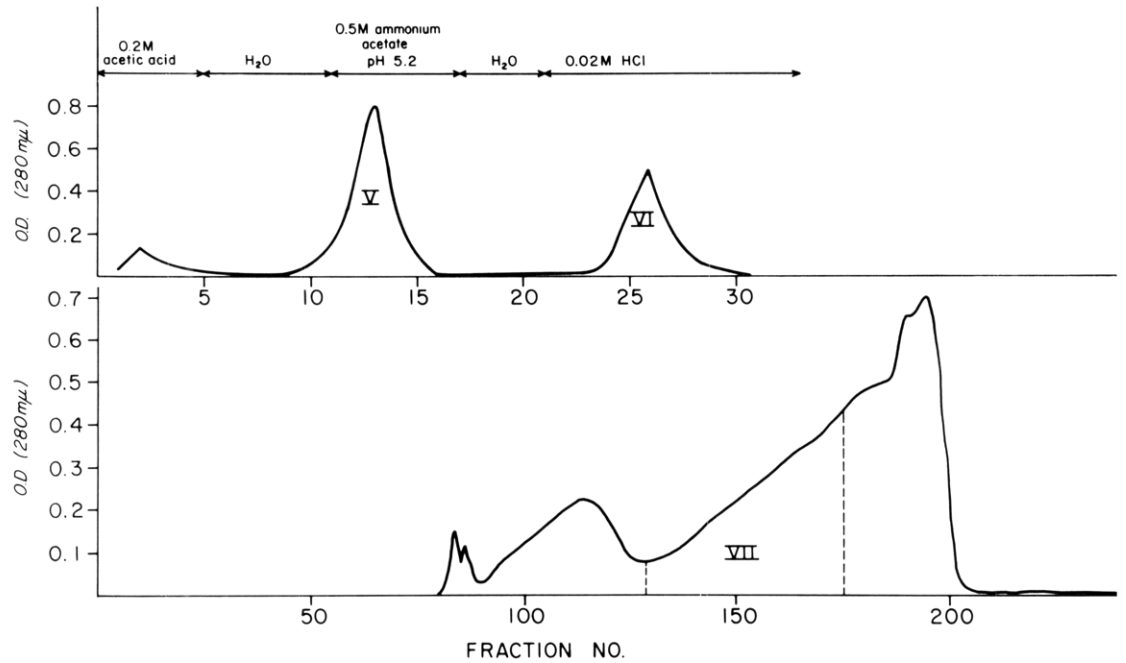

FIG. 7.—Top: Chromatography of peak III (Figure 5) on CM-cellulose (3.4 × 18 cm), flow rate 100 ml per hour. Bottom: Rechromatography of peak V on Sephadex G-75, 60–100 mesh, 3.4 × 430 cm, in 0.2 M acetic acid, flow rate 30 ml/hour, fraction size 15 ml.

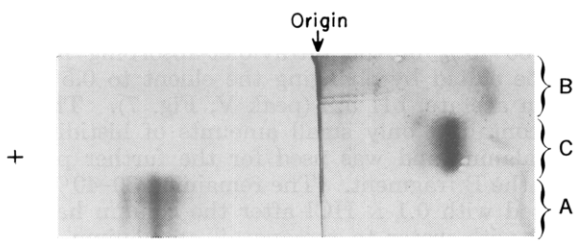

FIG. 8.—Photograph of starch-gel electrophoresis patterns of purified fragments of trypsinogen obtained by cyanogen bromide treatment (*S*-β-aminoethyl derivatives) at *p*H 8.6 in 6.5 M urea, fragments A, B, and C.

Table III. In spite of vigorous conditions of the dinitrophenylation reaction and a careful search for labile DNP-amino acids (such as DNP-glycine and DNP-proline) only 0.09 residue of N-terminal leucine or isoleucine was found. Leucine aminopeptidase did not liberate free amino acids. Starch-gel electrophoresis at *p*H 8.6 (Fig. 8) showed that the material is apparently homogeneous within the limits of detectability. In addition, carboxypeptidase A liberated only homoserine, lesser amounts of asparagine, serine, and a trace of threonine, amino acids which are expected from the known sequence predicted for the C-terminal region of the B fragment: .... Thr.Ser.$AspNH_2$.-HoSer. (Walsh *et al.*, 1962b).

C Fragment.—When the material of peak IV (Fig. 6) (650 mg) was chromatographed on a Sephadex G-75 column (3.4 × 430 cm) in 0.2 M acetic acid (Fig. 9), it became apparent that there had been contamination with some small material which appeared as a small separate peak on the 4.3-m-long column. Examination showed this fraction to consist of a number of peptides of unknown origin. The main peak (VIII) was collected, freeze-dried, and chromatographed on a CM-cellulose column as described for the B fragment. The peak emerging with 0.5 M ammonium acetate buffer *p*H 5.2 (corresponding to peak V in Fig. 7) was analyzed for its amino acid content. Arginine, histidine, and homoserine were present in amounts of less than one residue and therefore it was concluded that the main part of this fraction came from the C-terminal part of the trypsinogen molecule. Elution of the column with 0.1 M HCl gave a small peak which, as in the purification of the B fragment, was relatively rich in histidine and aspartic acid, and thus appears to originate from the A fragment.

Further purification of the main peak was achieved by fractionation with ammonium acetate. The material was dissolved in 0.2 M acetic acid (10 mg/ml). Ammonium acetate (10 M) was added to make the solution 0.9 M. The precipitate, which was high in histidine, was centrifuged off and discarded. A heavy precipitate which appeared on adjusting the supernatant to 2.5 M was collected, dissolved in water, and freeze-dried. The material was reprecipitated between 0.9 M and 2.7 M ammonium acetate and freed of ammonium acetate by freeze-drying.

Amino acid and N-terminal analyses showed that there was still considerable contamination with histidine, arginine, and homoserine. Most of the histidine and some of the arginine could be removed on a DEAE-cellulose column (2 × 60 cm) which was equilibrated with 0.005 M ammonium acetate, *p*H 8.0. A solution of the freeze-dried material from the preceding step in this same buffer was passed slowly into the column, and washed through with the same solvent. About 75% of the ultraviolet-absorbing material passed through. It was collected by freeze-drying and, as Table IV shows, it was virtually free of histidine and homoserine, but still contained some arginine. Its purity was estimated at about 85%. Starch-gel electrophoresis at *p*H 8.6 (Fig. 8) presented essentially one diffuse band. Carboxypeptidase A liberated in good yield asparagine, serine, isoleucine, alanine, threonine, and glutamine. The amounts liberated were compatible with the sequence $GluNH_2$.Thr.Ala.Ileu.Ser.$AspNH_2$ proposed by Walsh *et al.* (1964) for the C-terminal part of the whole trypsinogen molecule and hence for the C-terminal of the C fragment. Insufficient material was available for a more detailed study.

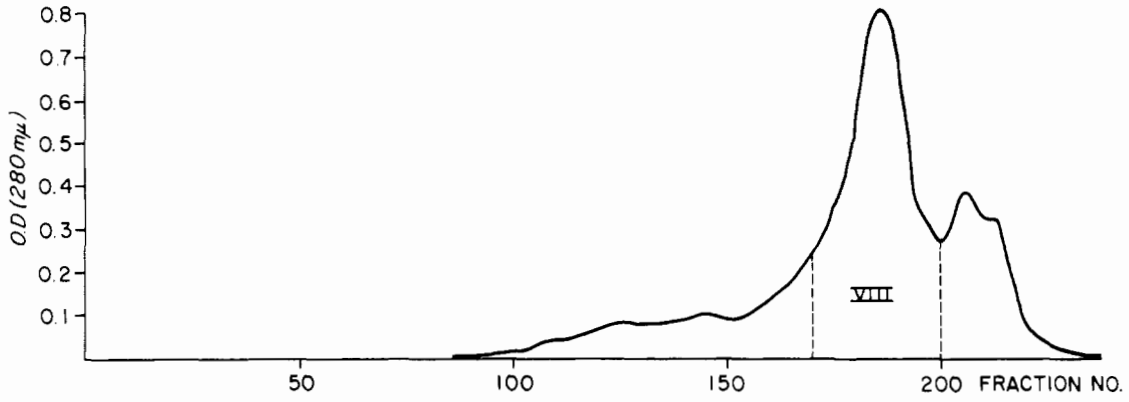

FIG. 9.—Rechromatography of peak IV (Figure 5) on Sephadex G-75 (3.4 × 430 cm) in 0.2 M acetic acid, flow rate 30 ml/hour.

DISCUSSION

The peptide chain of trypsinogen, like that of ribonuclease (Gross and Witkop, 1962) and sperm whale myoglobin (Edmundson, 1963), has been successfully cleaved by cyanogen bromide. Quantitative cleavage was obtained at *p*H 1 in all three cases. However, in trypsinogen there appears to be a peptide bond which is especially labile under these conditions and gives rise to N-terminal leucine or isoleucine. An attempt was therefore made to find conditions for the cyanogen bromide reaction under which this nonspecific hydrolysis would be prevented.

When trypsinogen was treated with cyanogen bromide for 30 hours at 30° in 0.01 N hydrochloric acid (*p*H about 2.0), 1 M acetic acid (*p*H about 2.4), or 0.1 M acetic acid (*p*H about 2.9) only 10%, 20%, and 0%, respectively, of the methionines were converted to homoserine while in 50% (v/v) formic acid (*p*H about 0.4–0.6) and 50% (v/v) acetic acid (*p*H about 1.5–1.7) 95% and 35%, respectively, of the methionines reacted. Unfortunately the nonspecific split was not prevented. These experiments give, however, some interesting information about the conditions under which cyanogen bromide cleaves methionine peptide bonds in proteins. The reaction went nearly to completion in 30 hours only at the low *p*H values. In the region of *p*H 2 some reaction occurred while at higher *p*H values no reaction could be detected. Ribonuclease behaves in a similar way while in model compounds cleavage can be obtained at *p*H values as high as 6 (Gross and Witkop, 1962). These authors ascribe the lack of reactivity of the methionines in ribonuclease at the higher *p*H values to the steric unavailability in the native protein and suggest that denaturation is a prerequisite for reaction. It is difficult to accept this explanation for trypsinogen, because the protein is not denatured in 0.1 N hydrochloric acid where cleavage is essentially quantitative. In 50% acetic acid where denaturation is extensive the reactivity of the methionine is much lower, while in 50% formic acid, which has a lower *p*H than the acetic acid, the reaction is again nearly quantitative. Thus it would appear that the *p*H is an important factor controlling the reaction, but other solvent effects cannot be ruled out at present.

In view of the complexity of the reaction mixture obtained after the cyanogen bromide reaction and the conversion of the reduced half-cystine residues into *S*-β-aminoethylcysteine residues, the purification and identification of the three theoretical fragments presented considerably greater difficulties than was the case in ribonuclease (Gross and Witkop, 1962) and myoglobin (Edmundson, 1963). In the early stages the identification of the crude fragments was greatly aided by the knowledge of large parts of the amino acid sequence (Walsh *et al.*, 1962b, and personal communication; Tomasek *et al.*, 1963). A number of factors added to the difficulties, however. Thus, the A fragment could not be obtained in a homogeneous state probably because it was accompanied by a fragment arising from the nonspecific split which appears to comprise a large section of the A fragment of similar properties. Furthermore, there is evidence of interactions between fragments leading to aggregations. Evidence for this assumption is provided by the fact that peak Ia, which consists largely of A fragment, emerged in front of peak Ib, which contains intact methionine and cystine, and thus represents incompletely degraded material and hence should be larger than the A fragment. In addition, histidine residues which are located solely in the A fragment were found in relatively large quantities in all the fractions of Figure 6 and thus provide evidence that the A fragment interacts with other fragments.

The separation of the B and C fragments also presented some difficulties. These two chains are not only similar in size (74 and 63 residues, respectively), but they also resemble each other in composition. Each has nearly the same number of positive charges, contributed by the lysine, *S*-β-aminoethylcysteine, and arginine residues, and each also has the same number of free carboxyl groups. In addition, the sum of the residues with hydrophobic side chains (aromatic residues, leucine, isoleucine, and valine) is similar, i.e., 18 and 17 for the B and C fragments, respectively. It is interesting to find that in spite of the similarity in size, gel filtration on long columns of Sephadex G-75 brings about a useful separation.

In spite of these difficulties, it has been possible to obtain the three fragments in 10–20% yield and about 85–90% purity. This purity is considered sufficient to give unambiguous compositions. As Table IV shows, the sum of the different residues of the individual fragments is equal to the sum of the corresponding residues in trypsinogen with only minor deviations.

An interesting point which will require further investigation is the unusually low reactivity of the N-terminal leucine residue in the B fragment. As has been shown, even under the most vigorous reaction conditions of dinitrophenylation, DNP-leucine was obtained only in 10% yield, and leucine aminopeptidase did not liberate any leucine at all. The N-terminal sequence predicted for the B fragment is Leu.Ileu.Lys.LeuLys . . . (Walsh *et al.*, 1964). It seems possible that once the ε-amino groups of the

two lysine residues have been substituted with the dinitrophenyl group, this region of the molecule becomes so strongly hydrophobic that the α-amino group of the terminal leucine becomes unavailable.

On the basis of the results obtained in this work and the known sequences of a considerable number of peptides it has been possible to allocate unambiguously all these peptides to one of the three fragments. The preliminary structure of the trypsinogen resulting from this and other studies has been presented (Walsh *et al.*, 1964) and will not be repeated here.

One important aspect, however, may be mentioned. Tomasek *et al.* (1963) propose the following sequence of 84 amino acids:

Ser.Ala.Tyr.Pro.Gly.Glu(NH_2).Ileu.Thr.Ser.Asp(NH_2).
←
Met.Phe.Cys.Ala.Gly.Tyr.Leu.Glu.Gly.Gly.Lys.
——T-29——
Asp(NH_2).Ser.Cys.Glu(NH_2).Gly.Gly.Asp.SER.Gly.

Pro.Val.Val.Cys.Ser.Gly.Lys|Leu.Lys.Ser.Ala.Ala.Ser.
——→|←——T-64——
Leu.Asp.(NH_2).Ser.Arg|Val.Ala.Ser.(Leu,Ser,Pro,Ileu,
——→|←——
Thr,Ser,Cys,Ala,Gly,Ala,Ser,Thr,Glu,Cys,Leu,)Ileu.

Ser.Gly.Try.Gly.Asp(NH_2).Thr.Lys|Ser.Ser.Gly.Thr.
——T-24——→|←——
Ser.Tyr.Pro.Asp.Val.Leu.Lys.
——T-47——→

As this sequence includes the dipeptide sequence . . .-Met.Phe . . . it must represent the C-terminal region of the B fragment and the N-terminal region of the C fragment. If this sequence is correct then the C fragment should contain a minimum of 7 alanines, 13 serines, and 1 arginine. As Table IV shows, however, there are only 3 alanines, 7 serines, and no arginine in the C fragment, and thus the work presented here does not support the conclusion of Tomasek *et al.* (1963). However, peptides T-64 and T-24 can be placed into the B fragment on the basis of the following argument: A peptic peptide, Asp(NH_2).Ser.Arg.Val.-Ala.Ser (P-49, Mikes *et al.*, 1961) indicates that peptides T-64 and T-24 overlap. They contain 5 alanines and 3 threonines and therefore cannot be allocated to the A fragment, because the A fragment contains only 4.4 alanines and 3 threonines (Table IV); of these 1 alanine and 2 threonines belong to the N-terminal peptide of trypsinogen, Val.Asp_4.Lys.Ileu.Val.Gly.-Gly.Tyr.Thr.Cys.Gly.Ala.Asp(NH_2).Thr.(Val,Pro).Tyr (Walsh *et al.*, 1962a).

A most interesting result which emerges from this study is that all of the three histidine residues are in the A fragment and must be located between residues number 20, the end of the N-terminal peptide, and 92, the end of the A fragment. The active-center serine residue, indicated in the above sequence by capital letters, is the 18th residue from the N terminal of the C fragment and therefore occupies approximately position 183. This linear separation of the serine from the histidines supports the model proposed by Dixon and Neurath (1957) for trypsinogen. Recently, Brown and Hartley (1963) showed that in chymotrypsinogen the two histidine residues are at position 39 and 55 and the active-center serine residue is at position 193, thus in both "serine proteases" two of the components of the active center are widely separated in the linear structure of the molecule.

Acknowledgments

I wish to express my deep gratitude to Dr. H. Neurath for his most generous hospitality and his continuous interest in this work. My sincere thanks are also due to Dr. K. A. Walsh for many fruitful suggestions, for carrying out the carboxypeptidase experiments, and for making available the data on the amino acid composition of trypsinogen and on amino acid sequences. I am indebted to Mrs. H. Froste and Miss D. L. Kauffman for their valuable assistance and to Mr. P. Schneider for carrying out the numerous amino acid analyses. I am also grateful to the University of Sheffield, England, for granting leave of absence.

References

Brown, J. R., and Hartley, B. S. (1963), *Biochem. J. 89*, 59P.
Cavallini, D., de Marco, D., Mondovi, B., and Azzone, G. F. (1955), *Experientia 11*, 62.
Davie, E. W., and Neurath, H. (1955), *J. Biol. Chem. 212*, 515.
Dixon, G., and Neurath, H. (1957), *Fed. Proc. 16*, 791.
Edmundson, A. B. (1963), *Nature 198*, 354.
Green, N. M., and Neurath, H. (1953), *J. Biol. Chem. 204*, 379.
Gross, E., and Witkop, B. (1961), *J. Am. Chem. Soc. 83*, 1510.
Gross, E., and Witkop, B. (1962), *J. Biol. Chem. 237*, 1856.
Gundlach, H. G., Stein, W. H., and Moore, S. (1959), *J. Biol. Chem. 234*, 1754.
Harris, J. I. (1955), *Methods Biochem. Analy. 2*, 383.
Harrison, P. M., and Hofmann, T. (1961), *Biochem. J. 80*, 38P.
Hofmann, T., Walsh, K. A., Kauffman, D. L., and Neurath, H. (1963), *Fed. Proc. 22*, 528.
Lindley, H. (1956), *Nature 178*, 647.
Mares-Guia, M., and Shaw, E. (1963), *Fed. Proc. 22*, 528.
Mikes, O., Holeysovsky, V., Tomasek, V., Keil, B., and Sorm, F. (1962), *Collection Czech. Chem. Commun. 27*, 1964.
Mikes, O., Holeysovsky, V., Tomasek, V., and Sorm, F. (1961), *Collection Czech. Chem. Commun. 26*, 1048.
Moore, S., Spackman, D. H., and Stein, W. H. (1958), *Anal. Chem. 30*, 1185.
Neumann, N. P., Moore, S., and Stein, W. H. (1962), *Biochemistry 1*, 68.
Porter, R. R. (1950), *Methods Med. Res. 3*, 256.
Raftery, M. A., and Cole, R. D. (1963), *Biochem. Biophys. Res. Commun. 10*, 467.
Sanger, F. (1945), *Biochem. J. 39*, 507.
Smithies, O. (1955), *Biochem. J. 61*, 629.
Spackman, D. H., Stein, W. H., and Moore, S. (1958), *Anal. Chem. 30*, 1190.
Spies, J. R., and Chambers, D. C. (1949), *Anal. Chem. 21*, 1249.
Tietze, F., Gladner, J. A., and Folk, J. E. (1957), *Biochim. Biophys. Acta 26*, 659.
Tomasek, V., Mikes, O., Holeysovsky, V., Keil, B., and Sorm F. (1963), *Biochim. Biophys. Acta 69*, 186.
Walsh, K. A., Kauffman, D. L., and Neurath, H. (1962a), *Biochemistry 1*, 893.
Walsh, K. A., Kauffman, D. L., and Neurath, H. (1962b), *Biochim. Biophys. Acta 65*, 540.
Walsh, K. A., Kauffman, D. L., Sampath Kumar, K. S. V., and Neurath, H. (1964), *Proc. Nat. Acad. Sci. U. S. 51* (in press).