Biochemistry 13, 2974.
Tanford, C. (1962), *Adv. Protein Chem. 17,* 69.
Tanford, C., and Kirkwood, J. G. (1957), *J. Am. Chem. Soc. 79,* 5333.
Tanford, C., and Roxby, R. (1972), *Biochemistry 11,* 2192.
Vigna, R. A. (1973), Ph.D. Thesis, Indiana University.
Vigna, R. A., Gurd, L. J., and Gurd, F. R. N. (1974), *J. Biol. Chem. 249,* 4144.
Watson, H. C. (1969), *Prog. Stereochem. 4,* 299.
Ysern, M. E. (1973), M.S. Thesis, Indiana University.

Amino Acid Sequence of Dogfish Trypsin†

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

ABSTRACT: The amino acid sequence of pancreatic trypsin from the spiny Pacific dogfish (*Squalus acanthias*) has been determined and compared with the sequences of bovine and porcine trypsin. Dogfish trypsin contains one less amino acid residue (222) than the other two enzymes. Two-thirds of the residues in corresponding positions in dogfish and bovine trypsin are identical and the sequences of all three enzymes are homologous. Of the 223 amino acid residues of bovine trypsin, 77 are replaced without significant changes in function. Seven replacements, all conservative, occur in the interior of the protein; the remainder are on the surface. All residues known to be components of the active site of bovine trypsin are present in corresponding positions in dogfish trypsin. Comparison of the three enzymes suggests calcium binding sites in dogfish trypsin. A corrected sequence of bovine trypsin identifies residue 67 as Asn and residues 84–87 as Ser-Asn-Thr-Leu.

The recognition that the homologous serine proteases, bovine chymotrypsin (Matthews et al., 1967), elastase (Shotton and Watson, 1970), and trypsin (Stroud et al., 1971), have similar three-dimensional conformations has provided an experimental basis for relating the details of their structure both to the specificity of their function and to their evolution (de Haen et al., 1975). Although no crystallographic data are yet available for pancreatic serine proteases of other species, knowledge of their amino acid sequences in conjunction with the model of the analogous bovine enzymes can yield useful information about phylogenetic evolution and about the tolerance of function to structural change.

We have previously reported a partial amino acid sequence of trypsin from the Spiny Pacific Dogfish *Squalus acanthias* (Bradshaw et al., 1970) and the complete amino acid sequence of porcine trypsin (Hermodson et al., 1973). Both enzymes are homologous with the bovine enzyme but a detailed comparison had to await the completion of the amino acid sequence of dogfish trypsin which is presented herein. The functional analogy of bovine and dogfish trypsins was previously demonstrated by Tye (1971).

Materials and Methods

Sephadex of various grades was obtained from Pharmacia Fine Chemicals. 4-Vinylpyridine (Baker) was distilled under reduced pressure and stored at −20°. Pepsin and α-chymotrypsin were products of Worthington Biochemical Corp. Before use, chymotrypsin was treated with α-*N*-tosyllysine chloromethyl ketone to inactivate trypsin.

Dogfish trypsin was isolated from frozen pancreas glands which had been freshly excised from dogfish caught in Puget Sound. Within 3 days the frozen glands (400 g containing approximately 250 mg of trypsinogen) were minced and blended in a Waring Blendor for 40 sec with 2 vol of cold water. The slurry was agitated with paddles for 1 hr at 4° in an electric ice cream freezer to complete the extraction and then centrifuged at 12,500 *g* for 1 hr. After filtration through a pad of glass wool, the supernatant was pumped at a rate of 1800 ml/hr onto a 9 × 75 cm column of DEAE-cellulose at 4°, previously equilibrated with 0.01 *M* NaCl–0.005 *M* Tris-HCl (pH 8.0). The column was washed overnight with the same buffer and then with 3 l. of 0.25 *M* NaCl–0.005 *M* Tris-HCl (pH 8.0). At this point the salt concentration was raised to 0.50 *M* NaCl (in 0.005 *M* Tris, pH 7.0) and 400 fractions of 25 ml each were collected. Fractions containing trypsinogen were pooled (yielding 160 mg of trypsinogen) and acidified to pH 2.7 with concentrated HCl and a precipitate was discarded after centrifugation at 15,800*g*. Solid ammonium sulfate was added (313.5 g/l.) to 50% saturation and the trypsinogen collected as a precipitate by centrifuging at 15,800*g* for 30 min. The precipitate (110 mg of trypsinogen) was dissolved in 100 ml of 1 m*M* HCl, dialyzed overnight against 1 m*M* HCl, and lyophilized. The dry powder was dissolved in 50 ml of 0.01 *M* Tris-HCl–0.05 *M* $CaCl_2$–0.5 *M* KCl (pH 8), and the solution readjusted to pH 8. The trypsinogen was converted to trypsin by treatment with 3 mg of dogfish trypsin for 1–2 hr at 36°, pH 8.0. A small precipitate was removed by centrifugation at 4° and the supernatant dialyzed for 24 hr against 1 m*M* HCl to remove salts.

The lyophilized product (114 mg of trypsin) was dissolved in 60 ml of 0.1 *M* Tris–0.05 *M* $CaCl_2$–0.5 *M* KCl (pH 8), the pH readjusted to pH 8, and a small precipitate discarded after centrifugation. A 1.5 × 84 cm column of chicken ovomucoid immobilized on Sepharose (Robinson et al., 1971) was equilibrated with 0.02 *M* Mes buffer[1] (pH

† From the Department of Biochemistry, University of Washington, Seattle, Washington 98195. *Received November 20, 1974.* This work has been supported by research grants from the National Institutes of Health (GM-15731) and the American Cancer Society (BC 91-P).

[1] Abbreviations used are: PE-, *S*-pyridylethyl-; Mes, 2-(*N*-morpholino)ethanesulfonic acid; BzArgOEt, benzyl-L-arginine methyl ester; Pth derivative, phenylthiohydantoin derivative.

6.0) containing 0.05 *M* $CaCl_2$ and 0.5 *M* KCl. After application of 10 ml of pH 8 buffer (described above), followed by the trypsin solution, the column was washed with pH 6 buffer until no protein emerged. Trypsin was then eluted by application of 0.1 *M* formic acid–0.5 *M* KCl (pH 1.9). Fractions containing tryptic activity were pooled, dialyzed against 1 m*M* HCl, and lyophilized. The overall yield was 25–30% of that in the crude extract (approximately 70 mg).

Trypsin activity was measured in a pH-Stat using Bz-ArgOEt[1] as substrate (Walsh and Wilcox, 1970). Before assay, trypsinogen was converted to trypsin by 3-hr incubation at 36° of 0.1-ml aliquots with 0.1 ml of 0.2 *M* Tris-HCl–0.1 *M* $CaCl_2$–1.0 *M* KCl (pH 8) containing 3.3 μg of bovine trypsin.

Amino acid analyses were performed with Beckman amino acid analyzers (Spackman et al., 1958) after acid hydrolysis by the method of Moore and Stein (1963). Homoserine lactone was converted to homoserine prior to analyses by the method of Ambler (1965). Cystine was determined after reduction and alkylation as PE-cysteine.

Automated Edman degradations were performed with a Beckman Sequencer, Model 890, by the method of Edman and Begg (1967) as modified by Hermodson et al. (1972). Repetitive yields of 95–97% were routinely observed. Pth[1] amino acids were identified by gas chromatography of the trimethylsilyl derivatives. The sequenator reagents were of Sequenal Grade (Pierce Chemical Co.).

Manual Edman degradations were performed by a modification of the method of Peterson et al. (1972) using reagents from Pierce Chemical Co. as follows. The peptide (0.5–1.0 μmol) was dissolved in a glass-stoppered 12-ml centrifuge tube by addition of 200 μl of 5% *N,N*-dimethylallylamine in 50% pyridine (pH 9.5). The tube was flushed with purified N_2 for 30 sec and 10 μl (5 μl for the second degradation) of phenyl isothiocyanate was then added. The contents were mixed and incubated at 50° for 20 min with occasional shaking. The reaction mixture was then extracted twice with 1-ml aliquots of benzene (stored under N_2). The benzene layer was separated by centrifugation and discarded. The aqueous layer was freeze-dried, first at room temperature for 60 min and then at 50–60° for 30 min.

Trifluoroacetic acid (200 μl, 100 μl for the second degradation) and 5 μl of ethanethiol were added. The tube was flushed with purified N_2 for 5–10 sec and then incubated at 50° for 7 min. Trifluoroacetic acid and ethanethiol were then evaporated under a rapid stream of N_2 so that the peptide would spread into a thin film at the bottom of the tube. The thiazolinone derivative was extracted twice with 1-ml aliquots of peroxide-free ethyl ether (Matheson Coleman and Bell, Analyzed Grade) containing 0.5% ethanethiol freshly prepared each day. After drying the ether extract, the thiazolinones were converted to the Pth derivatives and identified in the same manner as in the automated Edman degradations. The residual peptide was dried under a gentle stream of N_2 and then subjected to the next cycle of the degradation.

With this method, 10–20 consecutive amino acid residues could be usually identified and recovered with a repetitive yield of 70–90%, except for the following.

(a) SERINE. Extensive destruction was observed and the yield of Pth-δ-Ser was very poor (less than 5%). Thus, even starting with 0.5 μmol of the peptide, serine residues could be identified only within the first five degradations.

(b) ASPARAGINE AND GLUTAMINE. The yields were poor (5–10%) and derivatives of both the acid and the amide forms were obtained.

(c) THREONINE. The yield was poor (5–10%) and several unidentified peaks in addition to that of Pth-Thr were observed on gas chromatograms.

(d) PE-CYSTEINE, ARGININE, AND HISTIDINE. Identifications were only qualitative.

(e) SHORT PEPTIDES. When the peptide contained fewer than five residues, yields dropped to 5–20%, particularly with nonpolar peptides.

Cleavage of methionyl bonds with cyanogen bromide was effected in 70% formic acid (10 mg of protein and 20 mg of reagent/ml) for 15–20 hr at room temperature in the dark (Gross, 1967). The reaction product was diluted with 5 vol of water and lyophilized.

Enzymatic digests were performed for 15–20 hr at 37° using 10 mg of peptide/ml and 0.2 mg of enzyme/ml. Chymotryptic digests were performed in 1% NH_4HCO_3 (pH 8); peptic digests were performed in 5% formic acid (pH 1.5–2.0). In each case, the products were lyophilized and then dissolved in 2 ml of 8% pyridine and the pH adjusted to approximately 10 with triethylamine. The mixture was fractionated on a 0.9 × 50 cm column of Dowex 1-X2 using a four-chamber gradient of pyridine–acetate buffers (Wikler et al., 1970).

Results

Properties of Purified Dogfish Trypsin. Using the procedures described herein, approximately 25% of the tryptic activity of crude extracts of dogfish pancreas was recovered in the final purified product. The specific activity toward BzArgOEt and toward protein substrates was the same as that of bovine trypsin. After gel electrophoresis in sodium dodecyl sulfate, two bands were seen, a major one corresponding to a molecular weight of 23,000 and a minor broader band corresponding to a molecular weight of about 11,000. Sequenator analysis revealed the presence of two simultaneously degrading polypeptide chains, a major sequence of Ile-Val-Gly-Gly-Tyr and a minor sequence (about 30% of the major one) of Asn-Val-Asp-Leu-Ile. Since both bovine and porcine trypsins occur as mixtures of a single-chain form and a two-chain form (Schroeder and Shaw, 1968; Hermodson et al., 1973), the possible presence of such a mixture in the dogfish preparations was examined.

After reduction with dithioerythritol and alkylation with 4-vinylpyridine in 6 *M* guanidine · HCl (Hermodson et al., 1973), dogfish trypsin could be separated into two fractions by gel filtration on Sephadex G-50 (Figure 1). These two fractions had identical amino acid compositions. One fraction (designated as β in Figure 1) yielded a single band (mol wt 23,000) during electrophoresis in sodium dodecyl sulfate and a single amino-terminal sequence (Ile-Val-Gly-Gly-Tyr) by sequenator analysis. The other fraction (designated δ in Figure 1) yielded during electrophoresis a broader band (mol wt approximately 11,000) and equal amounts of two amino-terminal sequences (Ile-Val-Gly-Gly-Tyr and Asn-Val-Asp-Leu-Ile).

The two chains in fraction δ were separated from each other by chromatography on SE-sephadex in 8 *M* urea (Figure 2). The sum of the amino acid compositions of the two fractions (δ-C and δ-N) was equal to the composition of the single-chain β form (Table I). Since the amino-terminal sequence of δ-N was identical with that of the β form, it was concluded that the δ form of trypsin consists of two chains, δ-N and δ-C, which have resulted from autolytic

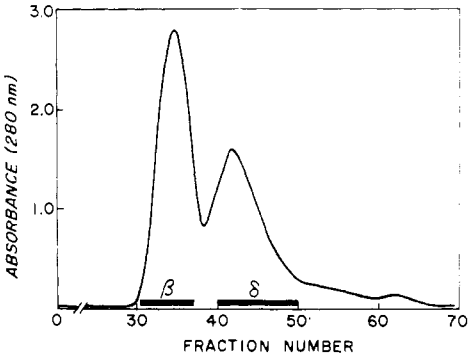

FIGURE 1: Separation of PE-trypsin (dogfish) into two species by gel filtration. The protein (100 mg) was applied in a volume of 5 ml to a column (2.5 × 110 cm) of Sephadex G-50 equilibrated and eluted with 9% formic acid. Fractions (5 ml) were collected at 40 ml/hr and pooled as indicated.

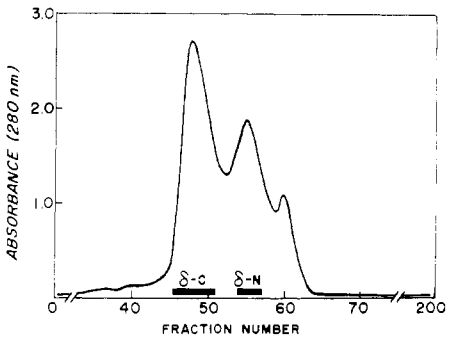

FIGURE 2: Separation of dogfish PE-δ-trypsin into amino-terminal (δ-N) and carboxy-terminal (δ-C) fragments by chromatography on a column of SE-Sephadex C-25 (2.5 × 30 cm). The protein (160 mg) was dissolved in 5 ml of 0.05 *M* sodium formate-8 *M* urea (pH 3.1) and applied to the column previously equilibrated with the same buffer. After elution with 50 ml of the buffer (not shown), a linear gradient was prepared from 500 ml of 0.05 *M* sodium formate-8 *M* urea (pH 3.1) to 500 ml of 1 *M* sodium acetate-8 *M* urea (pH 5.0). Fractions (5 ml) were collected at 50 ml/hr and pooled as indicated.

cleavage near the center of the single chain of β-trypsin. The alignment of these fragments of δ-trypsin is illustrated diagrammatically in Figure 3, where the residue numbers refer to the final sequence (Figure 4).

Extended sequenator analysis of PE-β-trypsin identified the amino-terminal sequence up to residue 48 (Figure 4). Similar analysis of fragment δ-C identified 31 residues, ending with Met-Asp-?-Ala-Val. From the compositional data of Table I, these 31 residues could be tentatively placed at positions 105–135 (Figure 4). The fourth degradation cycle of δ-C yielded leucine and proline in a ratio of 2:1, indicating polymorphism.

Fractionation of Cyanogen Bromide Fragments. The mixture of fragments derived by treatment of PE-β-trypsin (120 mg) with CNBr was fractionated by gel filtration (Figure 5). Of the six major pooled fractions, fractions IV and V were homogeneous as judged by amino acid analysis and by sequence analysis. Fraction I contained several fragments which were not completely cleaved by CNBr. Chromatography of fractions II, III, and VI on SE-Sephadex or Dowex 1-X2 shown in Figure 6 yielded ten pure fragments (CB1 to CB10) whose amino acid compositions are listed in Table II. The sum of the amino acid contents of these ten fragments is in close agreement with that of β-trypsin, indicating that the fragments indeed originated from cleavage of the nine methionyl bonds.

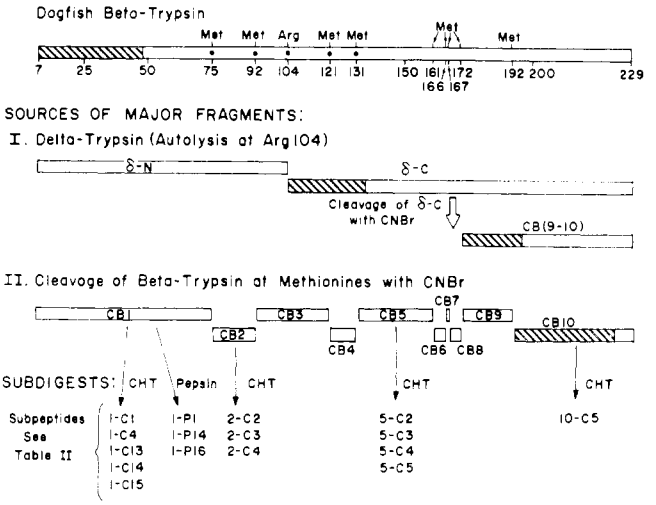

FIGURE 3: Diagrammatic summary of fragments generated from dogfish trypsin for sequence analysis. The top bar represents the polypeptide chain of PE-β-trypsin. The shaded areas of the horizontal bars represent portions of fragments examined by sequenator analysis. The length of each horizontal bar is directly proportional to the length of the protein or fragment. The nomenclature of chymotryptic (CHT) and peptic peptides corresponds to that summarized in Table III.

Table I: Amino Acid Compositions of Dogfish β-Trypsin and Its Autolytic Fragments δ-N and δ-C.

	Observed Amino Acid Composition[a]			
	β	δ-N[b]	δ-C[b]	Σ(δ-N + δ-C)
Lys	5.2 (5)	3.1 (3)	2.1 (2)	5.2
His	7.9 (8)	4.5 (5)	2.6 (3)	7.1
Arg	7.2 (7)	5.0 (5)	2.3 (2)	7.3
Asp	24.6 (24)	10.7 (11)	12.2 (13)	22.9
Thr	7.2 (7)	1.7 (2)	4.6 (5)	7.3
Ser	16.4 (17)	7.3 (8)	8.2 (9)	15.5
Glu	14.9 (15)	6.3 (6)	8.9 (9)	15.2
Pro	10.0 (10)	4.9 (5)	5.3 (5)	10.2
Gly	28.4 (28)	9.8 (9)	19.3 (19)	29.1
Ala	16.0 (16)	8.0 (8)	8.0 (8)	16.0
Cys[c]	11.4 (12)	3.1 (3)	8.5 (9)	11.6
Val	17.5 (18)	7.1 (7)	10.1 (11)	17.2
Met	8.9 (9)	2.3 (2)	5.9 (7)	8.2
Ile	12.6 (14)	6.4 (8)	5.0 (6)	11.4
Leu	14.0 (14)	6.9 (7)	6.9 (7)	13.8
Tyr	11.7 (12)	5.7 (6)	5.9 (6)	11.6
Phe	1.0 (1)	1.1 (1)	0 (0)	1.1
Trp	5.4 (5)	*d* (2)	*d* (3)	*d*
Total	(222)	(98)	(134)	

[a] The numbers in parentheses indicate those found by sequence analysis (Figure 4). [a] Isolated as illustrated in Figure 2. Their alignment is indicated in Figure 3. [c] Determined as *S*-pyridylethylcysteine. [d] Not determined.

In a separate experiment, a fragment was derived by cleavage of the δ-C portion of PE-δ-trypsin with CNBr (Figure 3), followed by gel filtration (Figure 7). The amino acid composition of this fragment agreed with that of the sum of fragments CB9 and CB10. However, the fragment contained a single methionyl residue whereas CB9 contained a homoseryl residue. Subsequent sequenator analysis (Figure 3) established that this fragment (CB9–10) possessed the same amino-terminal sequence as CB9. Thus, fragment CB9–10 appears to overlap CB9 and CB10 indicating that the methionyl bond (residue 192 in Figure 4) had not been cleaved by CNBr.

```
PORCINE:                                        Thr     Ala Ala
BOVINE:                                  10     Thr     Gly Ala
DOGFISH:                  ILE-VAL-GLY-GLY-TYR-GLU-CYS-PRO-LYS-

Asn Ser Val[a]    Tyr Gln                   Ser     Ser
Asn Thr Val       Tyr Gln              25   Ser             30
HIS-ALA-ALA-PRO-TRP-THR-VAL-SER-LEU-ASN-VAL-GLY-TYR-HIS-PHE-

                          Asn Ser Gln
                  35      Asn Ser Gln  40                 45
CYS-GLY-GLY-SER-LEU-ILE-ALA-PRO-GLY-TRP-VAL-VAL-SER-ALA-ALA-

            Lys Ser                                       Asn
            Lys Ser Gly                55                 Asp Asn
HIS-CYS-TYR-GLN-ARG-ARG-ILE-GLN-VAL-ARG-LEU-GLY-GLU-HIS-ASP-

    Asp Val Leu             Asn     Gln Phe     Asn Ala Ala Lys
    Asn Val Val  65         Asn     Gln Phe     Ser Ala     Lys
ILE-SER-ALA-ASN-GLU-GLY-ASP-GLU-THR-TYR-ILE-ASP-SER-SER-MET-

Ile     Thr                 Phe Asn     Asn Thr     Asp
Ser     Val         80 Ser      Asn Ser Asn Thr     Asn      90
VAL-ILE-ARG-HIS-PRO-ASN-TYR-SER-GLY-TYR-ASP-LEU-ASP-ASN-ASP-

                                Ser         Thr         Ser Arg
                  95        Lys Ser Ala 100 Ser         Ser Arg
ILE-MET-LEU-ILE-LYS-LEU-SER-LYS-PRO-ALA-ALA-LEU-ASN-ARG-ASN-

    Ala Thr Val                 Arg Ser         Ala         Thr
    Ala Ser       110               Ser 115     Ser         Thr
VAL-ASP-LEU[b]-ILE-SER-LEU-PRO-THR-GLY-CYS-ALA-TYR-ALA-GLY-GLU-

Glu                                         Lys Ser Ser Gly Ser
Gln               125                   130 Lys Ser Ser Gly Thr
MET-CYS-LEU-ILE-SER-GLY-TRP-GLY-ASN-THR-MET-ASP-GLY-ALA-VAL-

    Tyr Pro Ser Leu                 Lys
    Tyr Pro     Val     Lys         Lys         Ile     150
SER-[c]-GLY-ASP-GLN-LEU-GLN-CYS-LEU-ASP-ALA-PRO-VAL-LEU-SER-

    Ser Ser         Ser Ser             Gln         Gly
    Ser Ser     155 Ser             160 Gln         Ser 165
ASP-ALA-GLU-CYS-LYS-GLY-ALA-TYR-PRO-GLY-MET-ILE-THR-ASN-ASN-

    Ile             Phe Leu
    Phe     Ala 170     Leu         175                 180
MET-MET-CYS-VAL-GLY-TYR-MET-GLU-GLY-GLY-LYS-ASP-SER-CYS-GLN-

                                            Gln
                  185                   Ser Lys         195
GLY-ASP-SER-GLY-GLY-PRO-VAL-VAL-CYS-ASN-GLY-MET-LEU-GLN-GLY-

                                    Gln Lys Asn Lys
                  200 Ser           Gln Lys Asn Lys     210
ILE-VAL-SER-TRP-GLY-TYR-GLY-CYS-ALA-GLU-ARG-ASP-HIS-PRO-GLY-

            Lys         Asn         Asn         Gln Gln
            Lys 215     Asn         220         Lys Gln 225
VAL-TYR-THR-ARG-VAL-CYS-HIS-TYR-VAL-SER-TRP-ILE-HIS-GLU-THR-

       (Ala,Asn)
           Asn
ILE-ALA-SER-VAL
```

FIGURE 4: The amino acid sequence of dogfish β-trypsin. The residue numbers correspond to those of bovine trypsinogen (Walsh and Neurath, 1964). The sequences of the bovine enzyme (Walsh, 1970; Mikes et al., 1966) and porcine enzyme (Hermodson et al., 1973) are identical except for the residues which are set one (bovine) or two lines (porcine) above the dogfish sequence. The bovine sequence has been corrected at positions 67 (Asn) and 84–87 (Ser-Asn-Thr-Leu) by sequenator analysis of a tryptic fragment (Leu_{56}-Arg_{105}) of succinylated trypsin. The fragment was isolated as described by Hermodson et al. (1973): superscript a, isoleucine was also seen at this position in porcine trypsin; superscript b, proline was also observed at position 108 of dogfish trypsin, indicating polymorphism; superscript c, a gap is inserted in dogfish trypsin to optimize the alignment. Homologous residues whose side chains are internal in porcine elastase and bovine chymotrypsin are underlined (Shotton and Watson, 1970).

Amino Acid Sequence of the Individual CNBr Fragments. The relative lengths and the compositions of the ten fragments (and the overlap fragment) are summarized in Figure 3 and Table II, respectively. Although the alignment of these fragments in the whole molecule will not be presented until later (vide infra), for purposes of clarity the residue numbers of the complete structure (Figure 4) will be used in presenting the proof of structure of each fragment.

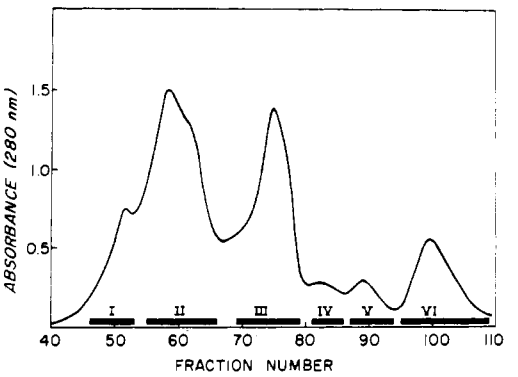

FIGURE 5: Fractionation of a cyanogen bromide digest of PE-β-trypsin (120 mg) on a column (2.5 × 110 cm) of Sephadex G-50, equilibrated and eluted with 9% formic acid. Fractions were pooled as indicated by the horizontal bars.

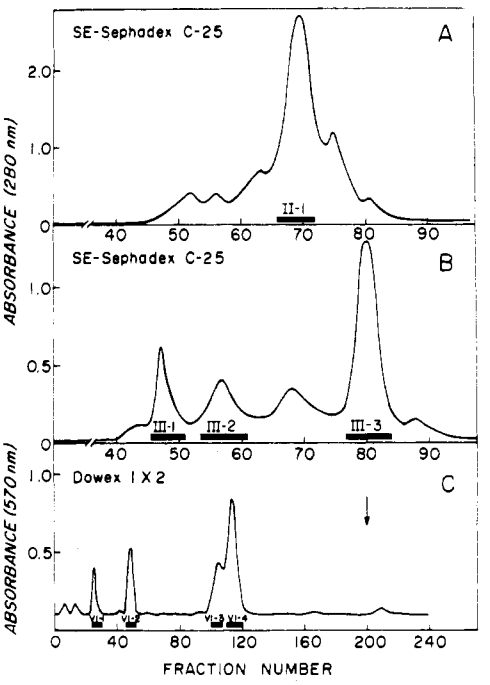

FIGURE 6: Purification of fragments of PE-β-trypsin described in Figure 5: (A) fraction II was applied to a column (2.5 × 30 cm) of SE-Sephadex C-25 and the chromatogram developed as described in Figure 2; (B) fraction III was applied and fractionated as described for fraction II; (C) fraction VI was dissolved in 2 ml of 3% pyridine and applied to a column (0.9 × 50 cm) of Dowex 1-X2 equilibrated at 37° with 3% pyridine. The column was developed at 37° with a four-chamber gradient of pyridine–acetate buffers (Wikler et al., 1970) and finally stripped with 50% acetic acid (arrow). Fractions (2 ml) were collected at a rate of 20 ml/hr. Peptides were detected by their absorbance at 280 nm or with ninhydrin (absorbance at 570 nm).

CB1 (RESIDUES 7–75). A major portion of the sequence of CB1 was derived by sequenator analysis of PE-β-trypsin which ended with Ala-His-Cys-Tyr_{48}. The remainder was elucidated by conventional analysis of eight peptides derived by enzymatic digestion of CB1 (Figure 8). Of 15 chymotryptic peptides and 16 peptic peptides isolated, most contained sequences which were redundant with those obtained by sequenator analysis. Eight peptides provided information necessary to complete the sequence of the frag-

Table II: Amino Acid Compositions[a] of Fragments of Dogfish Trypsin Derived by Cleavage with CNBr.

Peptide Residues (Fig 4) Source Amino Acid	CB1 7–75 II-1[b]	CB2 76–92 V[c]	CB3 93–121 III-2[b]	CB4 122–131 VI-4[b]	CB5 132–161 III-1[b]	CB6 162–166 VI-2[b]	CB7 167 VI-1[b]	CB8 168–172 VI-3[b]	CB9 173–192 IV[c]	CB10 193–229 III-3[b]	ΣCB1 through CB10 inc.[d]	β-Trypsin[e] 7–229	CB(9–10) 173–229 II[f]
Lys	1.2 (1)		2.0 (2)		1.1 (1)				1.2 (1)		5	5.2	1.4 (1)
His	3.6 (4)	0.9 (1)								2.8 (3)	8	7.9	2.6 (3)
Arg	3.1 (3)	1.0 (1)	1.0 (1)		0.1 (0)					1.9 (2)	7	7.2	1.9 (2)
Asp	5.1 (5)	4.7 (5)	3.1 (3)	1.1 (1)	4.2 (4)	1.9 (2)			3.1 (3)	1.3 (1)	24	24.6	4.3 (4)
Thr	2.3 (2)		0.9 (1)	1.0 (1)	0.2 (0)	1.0 (1)				1.9 (2)	7	7.2	2.4 (2)
Ser	5.3 (6)	1.1 (1)	2.0 (2)	1.0 (1)	1.9 (2)				2.2 (2)	2.7 (3)	17	16.4	4.5 (5)
Glu	5.9 (6)	0.2 (0)	1.2 (1)		3.2 (3)				2.2 (2)	3.1 (3)	15	14.9	4.9 (5)
Pro	3.0 (3)	0.9 (1)	2.4 (2)		2.4 (2)				1.1 (1)	1.3 (1)	10	10.0	1.8 (2)
Gly	8.3 (8)	1.3 (1)	2.2 (2)	2.0 (2)	4.1 (4)			1.3 (1)	6.0 (6)	4.5 (4)	28	28.4	10.4 (10)
Ala	6.0 (6)	0.4 (0)	4.0 (4)		4.0 (4)					2.3 (2)	16	16.0	1.9 (2)
Cys[g]	3.2 (3)		0.6 (1)	0.8 (1)	1.6 (2)			0.8 (1)	2.0 (2)	1.7 (2)	12	11.4	4.0 (4)
Val	4.7 (6)	0.5 (1)	1.2 (1)		2.1 (2)			1.0 (1)	1.7 (2)	4.5 (5)	18	17.5	6.5 (7)
Met[h]	0.8 (1)	1.2 (1)	0.9 (1)	1.1 (1)	0.9 (1)	1.0 (1)	1.0 (1)	0.7 (1)	0.8 (1)		9	8.9	1.1 (1)
Ile	4.6 (5)	1.5 (2)	1.9 (2)	1.0 (1)	0.4 (0)	1.0 (1)				2.5 (3)	14	12.6	2.9 (3)
Leu	2.9 (3)	1.0 (1)	4.8 (5)	1.0 (1)	3.3 (3)					1.0 (1)	14	14.0	1.0 (1)
Tyr	4.2 (4)	1.9 (2)	0.9 (1)		1.0 (1)			1.0 (1)		2.8 (3)	12	11.7	2.9 (3)
Phe	1.0 (1)										1	1.0	0
Trp	N.D. (2)			N.D. (1)						N.D. (2)	5	5.4	N.D. (2)
Total	(69)	(17)	(29)	(10)	(29)	(5)	(1)	(5)	(20)	(37)	(222)		(57)
Yield (%)	30	26	31	18	34	19	8	16	48	23			10

[a] Residues/peptide. Figures in parentheses indicate those calculated from sequence analysis. [b] From Figure 6. [c] From Figure 5. [d] Sum of compositions derived by sequence analysis. [e] Measured composition (Table I). [f] A peptide overlapping CB9 and CB10, derived from δ-C (Figure 7). [g] Determined as *S*-pyridylethylcysteine. [h] Determined as homoserine except where underlined.

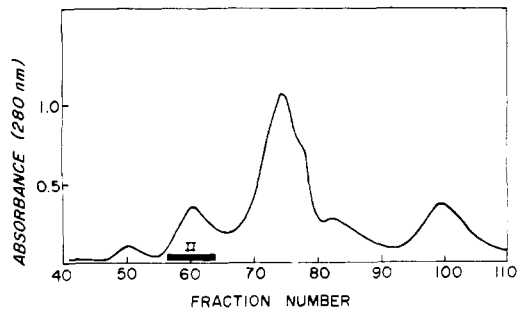

FIGURE 7: Fractionation of a cyanogen bromide digest of fragment δ-C (60 mg) on Sephadex G-50 under conditions identical with those of Figure 5.

ment. Their compositions are listed in Table III and the results of manual Edman degradations are summarized in Figure 8. The derived sequence is in accord with the amino acid composition of CB1 (Table II) if the slow hydrolysis of Ile_7-Val_8 and Val_{41}-Val_{42} is taken into account.

CB2 (RESIDUES 76-92). This 17-residue peptide was subjected to manual Edman degradation. Except for residue 83 (serine), 12 amino-terminal residues were identified. The remainder was elucidated from three peptides generated by chymotryptic digestion (Table III). The sum of compositions of the three peptides corresponded to that of fragment CB2. Since only peptide 2-C4 contained homoserine it was positioned at the carboxyl terminus of CB2. The compositions of peptides 2-C2 and 2-C3 agreed with the amino-terminal sequence of fragment CB2. Manual Edman degradations of peptides 2-C3 and 2-C4 completed the sequence of fragment CB2 (Figure 8).

CB3 (RESIDUES 93-121). Manual Edman degradation established the sequence of residues 93-114 except for residue 110 (serine). Beginning with position 104, the following sequence was observed: Arg-Asn-Val-Asp-Leu-Ile-?-Leu-Pro-Thr-Gly. The amino-terminal sequence of fragment δ-C was identical with this sequence beginning with Asn_{105}, and since serine was identified at position 110 the sequence could be extended to Met_{121}. The amino acid composition calculated from the sequence of fragment CB3 (Figure 4) is in agreement with that of fragment CB3 (Table II).

CB4 (RESIDUES 122-131). The sequence of this decapeptide was determined by nine cycles of manual Edman degradation. The results placed amino acids at all positions except 125 (Ser) and 130 (Thr). The final identifications were made from the sequence of fragment δ-C.

CB5 (RESIDUES 132-161). Manual Edman degradation established the sequence of 17 residues of this 29-residue peptide. The remainder was elucidated from overlapping chymotryptic peptides as indicated in Figure 8. The residues at positions 150 and 154 could not be identified by Edman degradation but the amino acid composition of chymotryptic peptide 5-C5 (Table III) indicated that only serine and PE-cysteine residues could occupy these two positions. Since PE-cysteinyl residues, but not seryl residues, are susceptible to chymotryptic cleavage,[2] it was concluded that PE-cysteine occupies position 154 and serine position 150. This sequence is confirmed by the sequence Leu-Asp-Ala-Pro-Val-(Leu,Ser,Asp,Ala,Glu)-Cys-Lys reported by Bradshaw et al. (1970) which is consistent with the sequence of residues 144-155.

[2] D. L. Enfield and K. Titani, unpublished experiments on chymotryptic digestion of fragments of bovine blood coagulation factor X_1.

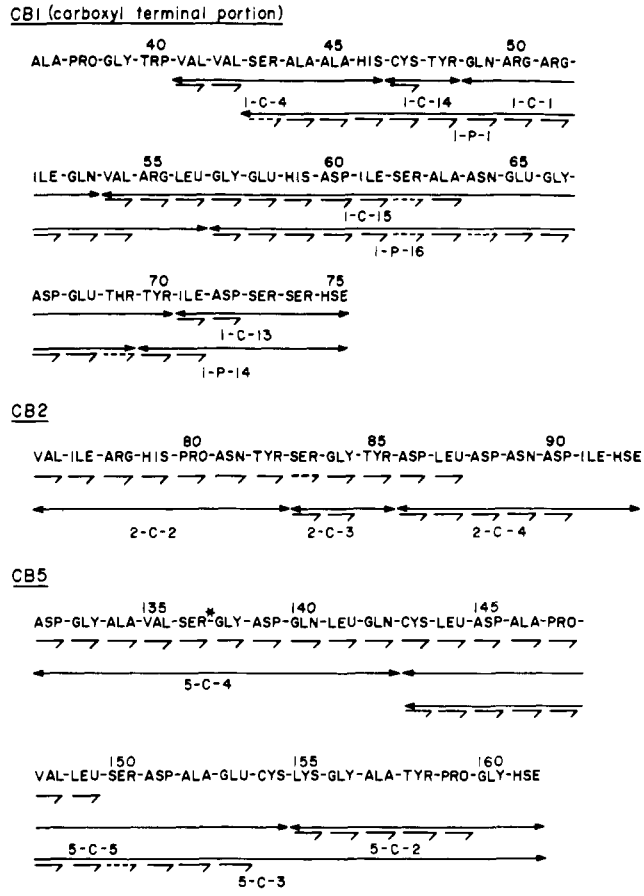

FIGURE 8: Summaries of the proof of sequence of the carboxyl-terminal portion of fragments CB1, CB2, and CB5. The numbering corresponds to that in Figure 4. The asterisk indicates the position at which a gap is placed in Figure 4. Peptides described in Table III were subjected to Edman degradation (→). A dashed arrow (⇠) indicates a postion at which a positive identification was not made.

CB6 (RESIDUES 162-166). Two cycles of the manual Edman degradation identified the amino terminus. Since homoserine must be carboxyl terminal, the peptide must have the sequence Ile-Thr-Asx-Asx-Hse. During paper electrophoresis at pH 6.5, two peaks were observed, a neutral peptide and a basic one. Homoseryl peptides are known to migrate as pairs, one peptide containing homoserine and the other homoserine lactone. Peptide CB6 must contain two asparagine residues to account for the observed charge.

CB7 (RESIDUE 167). Amino acid analysis revealed only homoserine (Table II).

CB8 (RESIDUES 168-172). Three cycles of the manual Edman degradation were sufficient to establish the sequence Cys-Val-Gly-Tyr-Hse.

CB9 (RESIDUES 173-192). Nine cycles of the manual Edman degradation established the sequence Glu-Gly-Gly-Lys-Asp-Ser-Cys-Gln-Gly. This sequence is identical with that of the amino terminus of the larger fragment (CB9-10). Sequenator analysis of the latter fragment established the sequence of CB9 (Figure 3).

CB10 (RESIDUES 193-229). This fragment was relatively insoluble in aqueous solution at neutral pH. Sequenator analysis identified 30 residues (Figure 3), ending with the sequence Val-Ser-Trp-Ile_{222}. Chymotryptic digestion of fragment CB10 yielded a peptide (10-C5) whose composition agreed with that predicted from the difference between those of fragment CB10 and residues 193-221. Manual

Table III: Amino Acid Compositions[a] of Peptides Derived by Enzymatic Digestion and Chromatography on Dowex 1-X2.

	Chymotryptic Cleavage of CB1					Peptic Cleavage of CB1			Chymotryptic Cleavage of CB2			Chymotryptic Cleavage of CB5				Chymotryptic Cleavage of CB10	Tryptic Cleavage of *S*-Aminoethyltrypsinogen[b]
Peptide Amino Acid	1-C1	1-C4	1-C13	1-C14	1-C15	1-P1	1-P14	1-P16	2-C2	2-C3	2-C4	5-C2	5-C3	5-C4	5-C5	10-C5	
Lys												1.4 (1)	1.2 (1)				
His		1.2 (1)			1.0 (1)	0.7 (1)		1.0 (1)	0.9 (1)							0.6 (1)	
Arg	2.3 (2)				0.9 (1)	2.8 (3)			1.0 (1)								
Asp			1.1 (1)		2.9 (3)		1.1 (1)	3.1 (3)	0.9 (1)		4.1 (4)		2.0 (2)	1.8 (2)	2.1 (2)		2.0 (2)
Thr					0.9 (1)			0.9 (1)								0.9 (1)	0.9 (1)
Ser		0.9 (1)	1.8 (2)		1.0 (1)	0.8 (1)	1.9 (2)	1.1 (1)		0.9 (1)			1.2 (1)	0.8 (1)	1.0 (1)	1.0 (1)	
Glu	1.7 (2)				2.9 (3)	2.1 (2)		3.0 (3)					1.4 (1)	1.8 (2)	1.2 (1)	1.1 (1)	
Pro									1.2 (1)			1.2 (1)	1.8 (2)		0.9 (1)		1.1 (1)
Gly					2.1 (2)			2.0 (2)		1.0 (1)		1.9 (2)	1.7 (2)	1.8 (2)			2.0 (2)
Ala		2.0 (2)			1.0 (1)	2.0 (2)		1.1 (1)				1.0 (1)	3.0 (3)	1.0 (1)	2.0 (2)	1.0 (1)	1.0 (1)
Cys[c]				1.0 (1)		1.0 (1)							1.8 (2)		1.7 (2)		1.1[d] (1)
Val		1.8 (2)			1.0 (1)	1.0 (1)			0.4[e] (1)				1.3 (1)	1.0 (1)	1.1 (1)	1.0 (1)	
Met[f]			1.0 (1)				1.1 (1)				0.8 (1)	0.6 (1)	0.8 (1)				3.1 (3)
Ile	1.0 (1)		0.9 (1)		1.0 (1)	1.1 (1)	0.9 (1)	0.7 (1)	0.4[e] (1)		0.9 (1)					1.7 (2)	0.9 (1)
Leu					1.0 (1)	0.7 (1)					1.0 (1)		2.3 (2)	1.0 (1)	1.9 (2)		
Tyr				0.9 (1)	0.8 (1)	0.8 (1)	0.9 (1)		0.8 (1)	0.8 (1)		0.9 (1)	0.5 (1)				0.8 (1)
Total	(5)	(6)	(5)	(2)	(17)	(14)	(6)	(13)	(7)	(3)	(7)	(7)	(19)	(10)	(12)	(8)	(13)
Yield (%)	13	27	40	17	25	15	37	16	81	78	85	30	16	53	19	30	15

[a] Residues per peptide. Figures in parentheses indicate those calculated from sequence analysis. [b] Isolated by Bradshaw et al. (1970) by fractionation on Dowex 50-X8 and Dowex 1-X2 of a tryptic digest of *S*-aminoethyltrypsinogen. [c] Determined as *S*-pyridylethylcysteine. [d] Determined as *S*-aminoethylcysteine. [e] Fifteen-hour hydrolysate. [f] Determined as homoserine except where underlined.

Edman degradation of peptide 10-C5 established its sequence as Ile-His-Glu-Thr-Ile-Ala-Ser-Val (residues 222–229).

Alignment of the Fragments. The ten CNBr fragments were aligned in the following manner. β-Trypsin, CB1, and δ-N all have the same amino-terminal sequence indicating that CB1 and δ-N are fragments derived from the amino-terminal portion of β-trypsin. As already discussed, fragment δ-C corresponds to the 125 carboxyl-terminal residues. Since the sequence of the ten amino-terminal residues of δ-C is identical with the sequence Asn-Val-Asp-Leu-Ile-Ser-Leu-Pro-Thr-Gly within fragment CB3, fragment CB3 must overlap δ-N and δ-C. Thus, the first 12 residues of CB3 represent the carboxyl terminus of δ-N (residues 93–104) and the other 17 residues represent the amino terminus of δ-C (residues 105–121).

The sum of the amino acids of fragment CB1 and of the first 12 residues of fragment CB3 accounts for 81 residues of the 98-residue fragment δ-N. The composition of the remaining 17 residues is consistent with the inclusion of fragment CB2 in δ-N. Thus, the three CNBr fragments must be in the order CB1–CB2–CB3. This conclusion is confirmed by the agreement between the composition of fragment δ-C and those of CB4 through CB10 plus that portion of CB3 not included in fragment δ-N.

Fragment CB10 must be the carboxyl terminus of trypsin since this is the only fragment lacking homoserine. Sequenator analysis of fragment CB9–10 placed 26 residues in sequence (Figure 3). The first 20 residues correspond to CB9 and the last 7 residues (Met-Leu-Gln-Gly-Ile-Val-Ser) provided an overlap with the amino-terminal region of CB10.

At this point, there remained only the alignment of fragments CB4 through CB8 in the region of residues 122–172 (Figure 3). Sequenator analysis of the fragment δ-C established that CB3 was followed by CB4. The analysis terminated with Met_{131}-Asp-?-Ala-Val. Since only CB5 has an amino-terminal aspartyl residue (in the sequence Asp-Gly-Ala-Val-), CB5 must follow CB4. This extends the sequence to Met_{161}, terminating with a sequence Gly-Ala-Tyr-Pro-Gly-Met_{161}.

Proof of the alignment of the remaining fragments was derived from a tryptic peptide (Table III) isolated by Bradshaw et al. (1970) which was shown to have the following sequence: Gly-Ala-Tyr-Pro-Gly-Met-Ile-(Thr,Asn)-Asn-Met-Met-Cys. The first six residues correspond to the carboxyl end of CB5, the next six residues dictate the placement of CB6 and CB7, and the terminal cysteine residue must be the amino-terminal residue of the adjacent CNBr fragment. Since only CB8 has an amino-terminal cysteine residue, it must follow CB7.

Of the nine overlapping sequences required to establish the order of the ten CNBr fragments in Figure 3, only four long overlapping segments were demonstrated, viz. CB3 to CB4, CB5 to CB6 to CB7, and CB9 to CB10. In addition, a short overlap linked CB4 to CB5, whereas only a single residue linked CB7 to CB8. Three placements were made by compositional arguments, i.e. the location of CB2 between CB1 and CB3 and the linkage of CB8 to CB9. The data of Bradshaw et al. (1970) confirm the placement of CB2 by providing overlapping sequences to CB1 and CB3. The juxtaposition of CB8 and CB9 is confirmed by a tryptic peptide isolated by Bradshaw et al. (1970) having the sequence Val-Gly-Tyr-Met-Glx-Gly-Gly-Lys. This corresponds to residues 169–176 and unambiguously places CB8 next to CB9.

Discussion

The amino acid sequence of dogfish trypsin has been established by a combination of sequenator analysis and conventional procedures. Over 50% of the 222 amino acids could be placed by sequenator analysis of PE-β-trypsin and of three major fragments; the remaining structure was resolved by conventional sequence analysis of ten cyanogen bromide fragments and of various peptides generated by enzymatic digestions. The pairing of the half-cystinyl residues yet needs to be determined but since all 12 half-cystinyl residues are in the same linear positions in dogfish, bovine, and porcine trypsins, it seems reasonable to assume that they are paired in identical manner and that the three-dimensional structures of these three trypsins are very similar. Since the activation peptide of dogfish trypsinogen has previously been reported to contain seven amino acid residues, in the sequence Ala-Pro-Asp-Asp-Asp-Asp-Lys (Hermodson et al., 1971), the zymogen contains a total of 229 amino acid residues.

When compared to bovine and porcine trypsin, the dogfish enzyme lacks an amino acid residue in position 137 (tyrosine). This residue is on the surface of the bovine enzyme, well removed from the active site and, according to Kenner et al. (1968), is rapidly nitrated by reaction with tetranitromethane. In common with porcine trypsin, the dogfish enzyme appears to have two allotypic variants. In the porcine trypsin, the replacements occur in position 18, whereas in the dogfish enzyme they occur in position 108 (Leu/Pro). The evidence is tentative since only the peptide containing leucine was isolated but not that containing proline.

Comparison of dogfish and bovine trypsins reveals 77 replacements among 223 residues, equivalent to 65% identity. Assuming that the species have diverged from a common ancestor approximately 400 million years ago, and applying the equation of Dickerson (1971), it follows that m, the number of replacements per 100 residues, is 42.3 and the unit evolutionary period (UEP) is 9.4 million years. This compares with a UEP of 3.0 million years obtained by comparison of bovine and porcine trypsins (Hermodson et al., 1973).

Bovine and dogfish trypsins differ at 77 loci; porcine and dogfish trypsins differ at 74 loci. By comparing bovine and porcine trypsins, 40 mutable sites have been identified; of these, 5 are in the interior of the molecule and the remainder on the surface (Hermodson et al., 1973). By comparing dogfish with bovine and porcine trypsins, 46 additional mutable sites can be added, making a total of 86 such sites in the bovine enzyme. Assuming a similar folding of bovine and dogfish trypsins, two of these new sites are in the interior and involve in each instance conservative substitutions (Tyr_{20}/Trp and Gln_{21}/Thr). The remaining 44 new sites are on the exterior of the molecule and differ from bovine or porcine trypsin in a conservative (17 sites) and in a radical (27 sites) manner.

Hermodson et al. (1973) have listed 19 components of the active site which are preserved in bovine and porcine trypsin. The same sites are preserved in the dogfish enzyme, consistent with the functional requirements of these three trypsins.

Bovine and dogfish trypsin, unlike the porcine enzyme, bind calcium ions. The site of binding has not been proven in any instance, but Stroud et al. (1971) have suggested that a combination of Asp_{59}, Asp_{139}, and Glu_{65} may be effective in chelating calcium ions in the bovine enzyme.

Since Asp_{59} in the bovine enzyme corresponds to His_{59} in both dogfish and porcine trypsins, residue 59 cannot be a ligand in dogfish trypsin. It is possible, however, that the adjacent residue Asp_{60} serves the same role as Asp_{59} in the bovine enzyme, whereas in the porcine enzyme the corresponding site (His_{59}, Asn_{60}) obviously precludes calcium binding.

Acknowledgments

We thank Mrs. L. Black for her assistance in the preparation of dogfish trypsin, Dr. M. A. Hermodson, Mrs. B. Moody, Mrs. S. Balian, and Mrs. M. Jarvis for their assistance in the sequence analyses, and Mr. R. R. Granberg for his assistance in the amino acid analyses.

References

Ambler, R. P. (1965), *Biochem. J. 96,* 32P.
Bradshaw, R. A., Neurath, H., Tye, R. W., Walsh, K. A., and Winter, W. P. (1970), *Nature (London) 226,* 237.
de Haen, C., Neurath, H., and Teller, D. C. (1975), *Biochemistry* (in press).
Dickerson, R. E. (1971), *J. Mol. Evol. 1,* 26.
Edman, P., and Begg, G. (1967), *Eur. J. Biochem. 1,* 80.
Gross, E. (1967), *Methods Enzymol. 11,* 238.
Hermodson, M. A., Ericsson, L. H., Neurath, H., and Walsh, K. A. (1973), *Biochemistry 12,* 3146.
Hermodson, M. A., Ericsson, L. H., Titani, K., Neurath, H., and Walsh, K. A. (1972), *Biochemsitry 11,* 4493.
Hermodson, M. A., Tye, R. W., Reeck, G. R., Neurath, H., and Walsh, K. A. (1971), *FEBS Lett. 14,* 222.
Kenner, R. A., Walsh, K. A., and Neurath, H. (1968), *Biochem. Biophys. Res. Commun. 33,* 353.
Matthews, B. W., Sigler, P. B., Henderson, R., and Blow, D. M. (1967), *Nature (London) 214,* 652.
Mikes, O., Holeysovsky, V., Tomasek, V., and Sorm, F. (1966), *Biochem. Biophys. Res. Commun. 24,* 346.
Moore, S., and Stein, W. H. (1963), *Methods Enzymol. 6,* 819.
Peterson, J. D., Nehrlich, S., Oyer, P. E., and Steiner, D. F. (1972), *J. Biol. Chem. 247,* 4866.
Robinson, N. C., Tye, R. W., Neurath, H., and Walsh, K. A. (1971), *Biochemistry 10,* 2743.
Schroeder, D. D., and Shaw, E. (1968), *J. Biol. Chem. 243,* 2943.
Shotton, D. M., and Watson, H. C. (1970), *Nature (London) 225,* 811.
Spackman, D. H., Stein, W. H., and Moore, S. (1958), *Anal. Chem. 30,* 1190.
Stroud, R. M., Kay, L. M., and Dickerson, R. E. (1971), *Cold Spring Harbor Symp. Quant. Biol. 36,* 125.
Tye, R. W. (1971), Ph.D. Dissertation, University of Washington, Seattle.
Walsh, K. A. (1970), *Methods Enzymol. 19,* 41.
Walsh, K. A., and Neurath, H. (1964), *Proc. Natl. Acad. Sci. U.S.A. 52,* 884.
Walsh, K. A., and Wilcox, P. E. (1970), *Methods Enzymol. 19,* 31.
Wikler, M., Titani, K., and Putnam, F. W. (1970), *J. Biol. Chem. 245,* 2158.

Protein Content of Chromatin Fractions Separated by Sucrose Gradient Centrifugation†

D. Doenecke‡ and B. J. McCarthy*

ABSTRACT: When sheared chromatin is centrifuged in a steep sucrose gradient, two broad peaks are resolved. DNA extracted from both fractions has approximately the same molecular weight. The basis for this fractionation seems to be differential aggregation. The slowly sedimenting material shows a lower protein/DNA ratio than the rapidly sedimenting chromatin as judged by equilibrium density centrifugation in CsCl after formaldehyde fixation or under nonionic conditions. After selective removal of histone f1 and further shear, most of the slowly sedimenting chromatin material appears as free DNA in steep cesium chloride gradients. The data are consistent with several recent reports concerning the subunit structure of chromatin.

The earliest method to be developed for chromatin fractionation involved fragmentation and differential centrifugation (Frenster *et al.,* 1963). Although methods based upon other principles have since been developed, this same basic methodology has proved to be most popular and convenient for obtaining two chromatin fractions exhibiting different biochemical and physical properties. The centrifugation approach was later refined through the use of sucrose gradients (Chalkley and Jensen, 1968; Yunis and Yasmineh, 1971; Duerksen and McCarthy, 1971; McCarthy et al., 1973).

In the present communication we present further characterization of the two fractions obtained after shearing chromatin of *Drosophila* cells, Schneider's line 2 (Schneider, 1972). The results demonstrate a major difference in the protein/DNA ratio of the two peaks.

† From the Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143. *Received October 31, 1974.* Supported by a grant from the National Science Foundation (GB 35704).

‡ Recipient of a research fellowship from the Deutsche Forschungsgemeinschaft. Present address: Physiologisch-chemisches Institut Lahnberge, D 355 Marburg Lahn, Federal Republic of Germany.