Ray, W. J., & Koshland, D. E., Jr. (1961) *J. Biol. Chem. 236*, 1973–1979.
Roy, S., & Colman, R. F. (1979) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 38*, 313.
Rubin, C. S., & Rosen, O. M. (1975) *Annu. Rev. Biochem. 44*, 831–887.
Sandoval, I. V., & Cuatrecasas, P. (1976) *Biochemistry 15*, 3424–3432.
Segel, I. H. (1975) in *Enzyme Kinetics*, p 108, Wiley, New York.
Thornburg, W., & Lindell, T. J. (1977) *J. Biol. Chem. 252*, 6660–6665.
Traugh, J. A., Ashby, C. D., & Walsh, D. A. (1974) *Methods Enzymol. 38*, 290–299.
Walinder, O. (1973) *Biochim. Biophys. Acta 293*, 140–149.
Weber, K., & Osborn, M. (1969) *J. Biol. Chem. 244*, 4406–4412.
Weller, M. (1979) in *Protein Phosphorylation* (Lagnado, J. R., Ed.) Academic Press, London and New York.
Wyatt, J. L., & Colman, R. F. (1977) *Biochemistry 16*, 1333–1342.
Zoller, M. J., & Taylor, S. S. (1979) *J. Biol. Chem. 254*, 8363–8368.

Amino Acid Sequence of Crayfish (*Astacus fluviatilis*) Trypsin I_f[†]

Koiti Titani, Tatsuru Sasagawa, Richard G. Woodbury, Lowell H. Ericsson, Herbert Dörsam, Marinette Kraemer, Hans Neurath, and Robert Zwilling*

ABSTRACT: The complete amino acid sequence of trypsin from the crayfish *Astacus fluviatilis* has been determined. The protein was fragmented with cyanogen bromide after S-carboxymethylation of the reduced disulfide bonds and by trypsin after S-carboxymethylation as well as after succinylation of lysine residues and aminoethylation of the reduced disulfide bonds. Peptides were purified by gel filtration and by reversed-phase high-performance liquid chromatography. Stepwise degradation was performed in a spinning cup sequencer. The enzyme contains 237 amino acid residues and has a molecular weight of 25 030. In contrast to bovine trypsin, it contains three rather than six disulfide bonds which are paired in the same fashion as those in trypsin from *Streptomyces griseus*. The constituents of the active site of bovine trypsin are present in corresponding positions in the crayfish enzyme. Crayfish trypsin shows 43.6% sequence identity with the bovine enzyme as compared to 40.0% identity with the *S. griseus* enzyme. The present analysis affords the first detailed view into the evolution of trypsins at the invertebrate level.

Ever since the isolation of crystalline trypsin by Kunitz & Northrop (1936), this enzyme has served as a prototype in the study of the structure and function of proteolytic enzymes. Trypsin-related proteases fulfill a variety of physiological functions (Reich et al., 1975; de Haën et al., 1975), and their phylogeny has been traced back to the prokaryotic level (Olafson et al., 1975).

Despite the current detailed knowledge of the structure and function of mammalian serine proteases, of which trypsin is a prominent member, information of the early evolutionary history of this family of proteases is relatively fragmentary. Indeed, it is remarkable how much less is known about the invertebrate proteases than about their vertebrate or bacterial counterparts (Zwilling & Neurath, 1981). In order to fill this gap of the phylogenetic development of trypsins, and identify those amino acid residues which have been conserved during the evolution of the enzyme, we have undertaken a determination of the amino acid sequence of crayfish trypsin. This "old" serine protease lies on the evolutionary pathway from which decapode crustacea and mammals have diverged some 700 million years ago.

In contrast to its mammalian counterpart, crayfish trypsin is a rather acidic protein. It is resistant to autodigestion, is irreversibly denatured below pH 3, and has no known zymogen form (Zwilling et al., 1969). Some of its structural features are more closely akin to bacterial than to bovine trypsin (Zwilling et al., 1975). The present study describes the complete covalent structure of this invertebrate trypsin which, for the reasons given above, might be considered a "missing link" between prokaryotic and vertebrate serine proteases (Zwilling et al., 1980).

Materials and Methods

Crayfish trypsin is elaborated by the hepatopancreas and is secreted into the stomachlike cardia, which is the best source for this enzyme. The dark brown digestive fluid (containing approximately 3 mg of trypsin/mL) was collected from the cardia by the method previously described (Zwilling & Neurath, 1981). Initially, the enzyme was purified by affinity chromatography on soybean trypsin inhibitor (SBTI)[1] covalently linked to Sepharose (Zwilling & Neurath, 1981). This procedure provided a highly purified preparation in good yield but failed to separate the five multiple forms known to coexist in trypsin of *Astacus fluviatilis* (Pfleiderer & Zwilling, 1972). In order to effect their separation, we subjected the crude extract to anion-exchange chromatography on DEAE-Sephacel (pH 8.0, NaCl gradient, 0.2–2.2 M) and to gel filtration on Sephadex G50, fine (3 × 95 cm column, 0.01 M Tris-HCl buffer, pH 8.0, containing 0.4 M NaCl), repeating each step

[†] From the Howard Hughes Medical Institute and the Department of Biochemistry, University of Washington, Seattle, Washington 98195 (K.T., T.S., R.G.W., L.H.E., and H.N.), and the Department of Physiology, Institute of Zoology, University of Heidelberg, D-6900 Heidelberg, FRG (H.D., M.K., and R.Z.). *Received October 13, 1982.* This work was supported by the National Institutes of Health (GM-15731) and by research grants from the Deutsche Forschungsgemeinschaft, Bonn-Bad Godesberg, FRG (Zw 17/5 and Zw 17/7).

[1] Abbreviations: I_f, band I of crayfish (*Astacus fluviatilis*) trypsin; AE, aminoethyl; CM, carboxymethyl; HPLC, high-performance liquid chromatography; Pth, phenylthiohydantoin; TPCK, tosylphenylalanyl chloromethyl ketone; SBTI, soybean trypsin inhibitor; DEAE, diethylaminoethyl; Tris, tris(hydroxymethyl)aminomethane.

0006-2960/83/0422-1459$01.50/0 © 1983 American Chemical Society

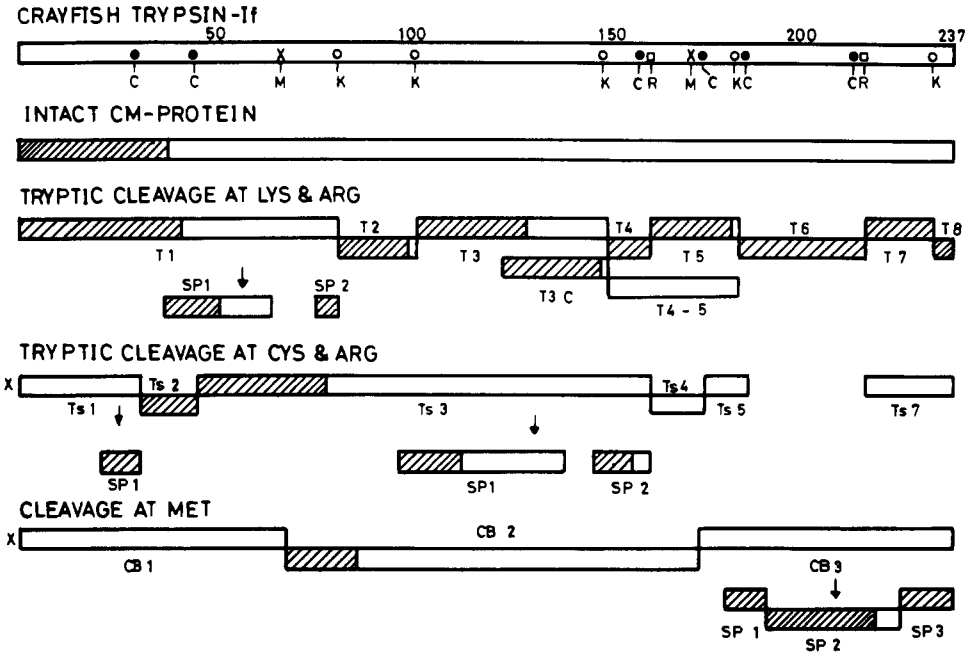

FIGURE 1: Diagrammatic summary of the major fragments generated for sequence analysis of crayfish trypsin I_f. The top bar represents the whole protein and indicates the location of cleavage sites, i.e., lysine, K (O), arginine, R (□), cysteine, C (●), and methionine, M (×), residues. The hatched section of each bar represents the portion of the sequence determined by automated Edman degradation. X denotes that the amino terminus is blocked by succinylation. For acronyms of the fragments, see the text.

twice (Zwilling et al., 1981). The material prepared in this manner migrated in disc gel electrophoresis (15% gel) as a single band, I_f, with the lowest cationic mobility among the multiple forms at pH 8.6. This material, after lyophilization, was used for the completion of the sequence analysis.

TPCK–trypsin (bovine) was purchased from Worthington. *Staphylococcus aureus* V8 protease was obtained from Miles. Cyanogen bromide, ethylenimine, and succinic anhydride were products of Eastman, Pierce, and MCB, respectively. Iodo-[^{14}C]acetic acid (12.56 mCi/mmol) was purchased from New England Nuclear. Acetonitrile for HPLC was obtained from Burdick and Jackson. Trifluoroacetic acid (Pierce) was distilled after refluxing with CrO_3. Reagents and solvents for the sequencer were obtained from Beckman. Polybrene was purchased from Aldrich.

Sephadex columns of various grades were products of Pharmacia. Columns for reversed-phase HPLC were μBondapak C18 and CN (Waters), SynChropak RP-P (SynChrom), and Zorbax ODS (Du Pont).

Amino acid analyses were performed with a Dionex D-500 amino acid analyzer following manufacturer's instructions. Sequence analyses were performed with a Beckman 890C sequencer according to the method of Edman & Begg (1967) as modified by Brauer et al. (1975), in the presence of polybrene (Tarr et al., 1978). Pth derivatives of amino acids were identified in a semiquantitative manner by two complementary systems of reversed-phase HPLC (Hermann et al., 1978; Ericsson et al., 1977).

Protein was reduced with dithiothreitol and carboxymethylated with iodo[^{14}C]acetic acid by a modification of the method of Crestfield et al. (1963). Protein was succinylated by the method of Yaoi et al. (1964) and then reduced and aminoethylated by the method of Cole (1967).

Peptide separation by reversed-phase HPLC was performed with a Varian 5000 liquid chromatograph. The column was equilibrated with 0.1% trifluoroacetic acid. The peptide mixture was injected into the column in either 0.1% trifluoroacetic acid or 6 M guanidine hydrochloride and eluted by increasing concentrations of acetonitrile containing 0.08% trifluoroacetic acid (Dunlap et al., 1978; Mahoney & Hermodson, 1980).

Results

General Strategy. The strategy employed for the sequence determination of crayfish trypsin I_f is shown in Figure 1.

First, Edman degradation of intact CM-protein (20 nmol) yielded the amino-terminal sequence of 35 residues (Table I) except for several tentative identifications. Although the analysis was repeated several times, the sequence determination could not be further extended because of an unexpectedly rapid increase of background Pth's which presumably was caused by the extremely acidic nature of the protein (Table II).

Second, the CM-protein was digested with TPCK–trypsin (bovine). Sequence analysis of 10 isolated peptides yielded nearly 90% of the total sequence (213 out of 237 residues). In this analysis, it was realized that the first residue to be cleaved by trypsin is located too far from the amino terminus. Hence, the protein was subjected to tryptic cleavage at cysteinyl and arginyl bonds after succinylation of lysyl residues followed by S-aminoethylation of cysteinyl residues as described under Materials and Methods. Analysis of peptides isolated from this and of peptides obtained by cleavage with cyanogen bromide provided the remaining sequence and the overlaps of the primary tryptic peptides.

Products of Lysyl and Arginyl Cleavage. A tryptic digest (at 37 °C, pH 8, for 15 h) of 400 nmol of CM-protein was resolved by gel filtration into five fractions (Figure 2). Fractions 1 and 2 mainly contained one peptide, T1 (residues 1–81). Fraction 1 was assumed to be an aggregated product. Fraction 3 contained two peptides, T3 and T4–5 (minor overlapping peptide), which were further separated by reversed-phase HPLC (Figure 3A). Five peptides, T2, T3C, T4, T5, and T6, in fraction 4 and two peptides, T7 and T8 in fraction 5 were also separated by HPLC as shown in panels B and C, respectively, of Figure 3. The amino acid compositions of these peptides are listed in Table III. In all, 10 tryptic peptides were isolated, of which one (T4–5) generated by incomplete cleavage of the Arg–Asp bond (residues

Table I: Summary of Automated Edman Degradations of Peptides Derived from Crayfish Trypsin I_f

fragment analyzed (residue no.)	approximate amount applied (nmol)	repetitive yield (%)	continuous sequence proven	identification	
				tentative	poor
intact CM-protein (1-237)	20	97	1-36	20, 21, 26, 29, 30, 32, 33	24, 28
T1 (1-81)	27	92	1-41	41	28, 36
Ts1-SP1 (21-30)	10	89	21-30	30	
Ts2 (31-46)	50	80	31-46	46	
T1-SP1 (38-64)	9	89	38-50		
Ts3 (47-160)	20	92	47-77		75, 76
CB2 (68-172)	20	89	68-88	81	78, 85, 86
T1-SP2 (75-81)	10	89	75-81		
T2 (82-101)	20	82	32-101	98, 99, 100, 101	
Ts3-SP1 (97-138?)	9	86	97-109		107
T3 (102-148)	15	87	102-128		
T3C (123-148)	13	86	123-148	146, 147, 148	
Ts3-SP2 (145-158?)	4	80	145-154		
T4 (149-160)	15	95	149-160		
T4-5 (149-182)	35	89	149-164		
T5 (161-182)	20	87	161-182	181, 182	
CB3-SP1 (180-188)	9	[a]	180-184		
T6 (183-215)	20	92	183-214		
CB3-SP2 (189-224)	12	86	189-218	203, 209, 216, 217, 218	208, 212, 213, 215
T7 (216-233)	10	96	216-232	231	
CB3-SP3 (225-237)	12	87	225-235		
T8 (234-237)	15	[a]	234-237		

[a] Values were not estimated.

Table II: Amino Acid Composition (Residues per Molecule) of Crayfish Trypsin I_f[a]

amino acids[b]	I	II	III
Asp (D)	28.2	30.2	18
Asn (N)			12
Thr (T)	14.7[c]	14.0	16
Ser (S)	17.5[c]	21.1	19
Glu (E)	21.6	21.5	13
Gln (Q)			7
Pro (P)	9.3	9.3	10
Gly (G)	29.9	29.6	30
Ala (A)	18.0	18.2	18
1/2-Cys (C)	6.3[d]	4.1[d]	6
Val (V)	17.3[e]	18.6	19
Met (M)	2.3	1.4	2
Ile (I)	13.5[e]	11.2	14
Leu (L)	15.4	18.0	17
Tyr (Y)	10.2	11.2	12
Phe (F)	7.6	9.3	9
His (H)	4.3	4.9	5
Lys (K)	3.9	5.3	5
Arg (R)	1.7	2.0	2
Trp (W)	ND[f]	ND[f]	3
	221.6[g]	229.9[g]	237[g]

[a] I, by amino acid analysis of the whole protein; II, the sum of compositions of eight major tryptic peptides by amino acid analysis (Table I); III, composition calculated from the sequence (Figure 4). [b] One-letter codes of amino acids in parentheses. [c] Extrapolated to zero-time hydrolysis. [d] Analyzed as CM-Cys. [e] Taken from values of 96-h acid hydrolysates. [f] Not determined. [g] Total.

160–161) was an overlapping peptide of T4 and T5 and another (T3C) was derived from the carboxyl-terminal half of T3 by a chymotrypsin-like cleavage. The sum of compositions of eight major peptides, excluding the two minor peptides T3C and T4–5, is in good agreement with that of the whole protein and also with the composition calculated from the complete sequence (Table II). Peptide T8 must have been derived from the carboxyl terminus of the whole protein because it lacked both lysine and arginine.

All of the residues of four isolated peptides, T4, T6, T7, and T8, were placed by sequence analysis (Table I and Figure 4), except that the carboxyl-terminal lysine or arginine residues

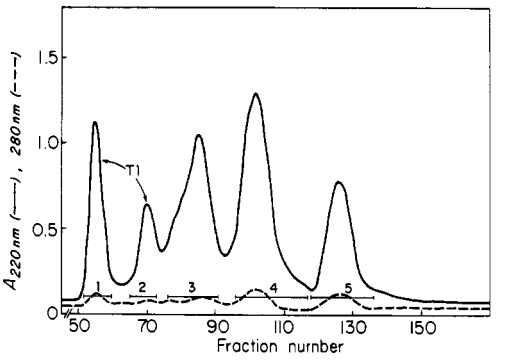

FIGURE 2: Fractionation of a tryptic digest of S-CM-protein (400 nmol) on a column (1.5 × 200 cm) of Sephadex G-50 superfine in 0.1 M NH_4HCO_3, pH 8. Fractions of 1 mL were collected at a flow rate of 9 mL/h and pooled as indicated by horizontal bars.

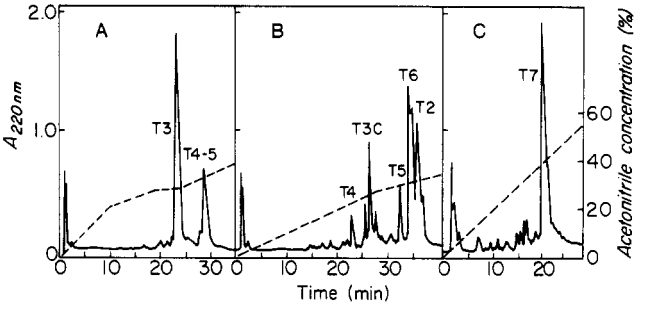

FIGURE 3: Purification of tryptic peptides by reversed-phase HPLC. (A) Pool 3 in Figure 2 was applied to a SynChropak RP-P column (0.41 × 25 cm) equilibrated with 0.1% trifluoroacetic acid and eluted at a flow rate of 1 mL/min by a gradient of 0.1% trifluoroacetic acid to 0.08% trifluoroacetic acid in acetonitrile as shown by the broken line. Peptides were recovered by lyophilization. (B) Pool 4 in Figure 2 was separated in a similar manner as in (A). (C) Pool 5 in Figure 2 was separated in a similar manner as in (A) on a Waters μBondapak C18 column (0.4 × 30 cm).

in T6 and T7 were assigned on the basis of their compositions. The sequence analysis of T1 (Table I) indicates that the peptide is derived from the amino-terminal portion of the whole protein. The analysis of T1 confirmed and extended the amino-terminal sequence of the protein up to Ile-41 except for

Table III: Amino Acid Compositions[a] of Peptides Obtained from Tryptic Digest of S-Carboxymethylated Crayfish Trypsin I_f (Residues per Molecule)

peptide:	T1	T2	T3	T3C	T4	T4-5	T5	T6	T7	T8	T1-SP1[c]	T1-SP2[c]
residues:[b]	1-81	82-101	102-148	123-148	149-160	149-182	161-182	183-215	216-233	234-237	38-64	75-81
amino acids												
CM-Cys	1.6 (2)				0.7 (1)	1.5 (2)	0.8 (1)	0.9 (2)			0.6 (1)	
Asx	7.1 (8)	6.1 (6)	5.8 (6)	3.6 (3)	1.4 (1)	4.2 (5)	3.9 (4)	4.1 (3)	0.8 (1)	1.1 (1)	4.2 (4)	
Thr	4.0 (5)		6.0 (7)	4.6 (6)	1.1 (1)	0.9 (1)		2.1 (2)	0.9 (1)		1.2 (1)	1.8 (2)
Ser	6.2 (7)	2.0 (1)	3.2 (3)	2.1 (1)	1.8 (1)	2.0 (2)	1.7 (1)	5.2 (5)	1.0 (1)		1.6 (1)	1.5 (1)
Glx	9.0 (11)	1.6 (1)	3.6 (3)	2.7 (2)	1.5 (1)	3.1 (3)	2.7 (2)	1.8 (1)	1.3 (1)		3.9 (3)	1.4 (1)
Pro	2.3 (2)		2.6 (3)	1.1 (1)	1.0 (1)	1.6 (2)	1.0 (1)	0.6 (1)	1.8 (1)		1.1 (1)	
Gly	9.4 (10)		6.8 (7)	5.0 (5)		3.6 (4)	3.6 (4)	7.7 (7)	2.0 (2)		4.5 (4)	
Ala	5.0 (5)		3.9 (4)	1.4 (1)	1.3 (1)	2.8 (3)	2.0 (2)	4.0 (4)		2.0 (2)	3.0 (3)	
Val	5.3 (6)		4.6 (4)	4.1 (3)	3.3 (3)	3.8 (4)	1.2 (1)	0.5 (1)	2.7 (3)	1.0 (1)	1.6 (2)	0.9 (1)
Met	1.0 (1)					0.5 (1)	0.4 (1)					
Ile	3.8 (5)	2.1 (3)	1.6 (2)	0.9 (1)		1.5 (2)	1.6 (2)	1.2 (1)	1.0 (1)		2.0 (2)	1.2 (1)
Leu	5.2 (5)	5.0 (5)	4.0 (4)	1.1 (1)	1.0 (1)	1.0 (1)		2.8 (2)			1.4 (1)	
Tyr	4.4 (5)	1.1 (1)				0.9 (1)	1.0 (1)	2.0 (2)	2.7 (2)		2.8 (3)	
Phe	5.6 (6)	1.1 (1)	1.3 (1)			1.1 (1)	1.3 (1)					
His	1.9 (2)	1.0 (1)	1.0 (1)						0.9 (1)		0.7 (1)	
Lys	1.4 (1)	1.0 (1)	1.3 (1)	1.3 (1)		1.1 (1)	0.8 (1)		0.9 (1)			0.7 (1)
Arg					1.0 (1)	0.9 (1)		1.0 (1)				
Trp[d]			(1)	(1)				(1)	(1)			
no. of residues	81	20	47	26	12	34	22	33	18	4	27	7
% yield	36	22	22	9	14	28	8	15	31	27	27	54

[a] By amino acid analysis or (in parentheses) calculated from the sequence (Figure 4). [b] For residue numbering, see Figure 4. [c] Subpeptides of peptide T1 generated by staphylococcal protease. [d] Not determined.

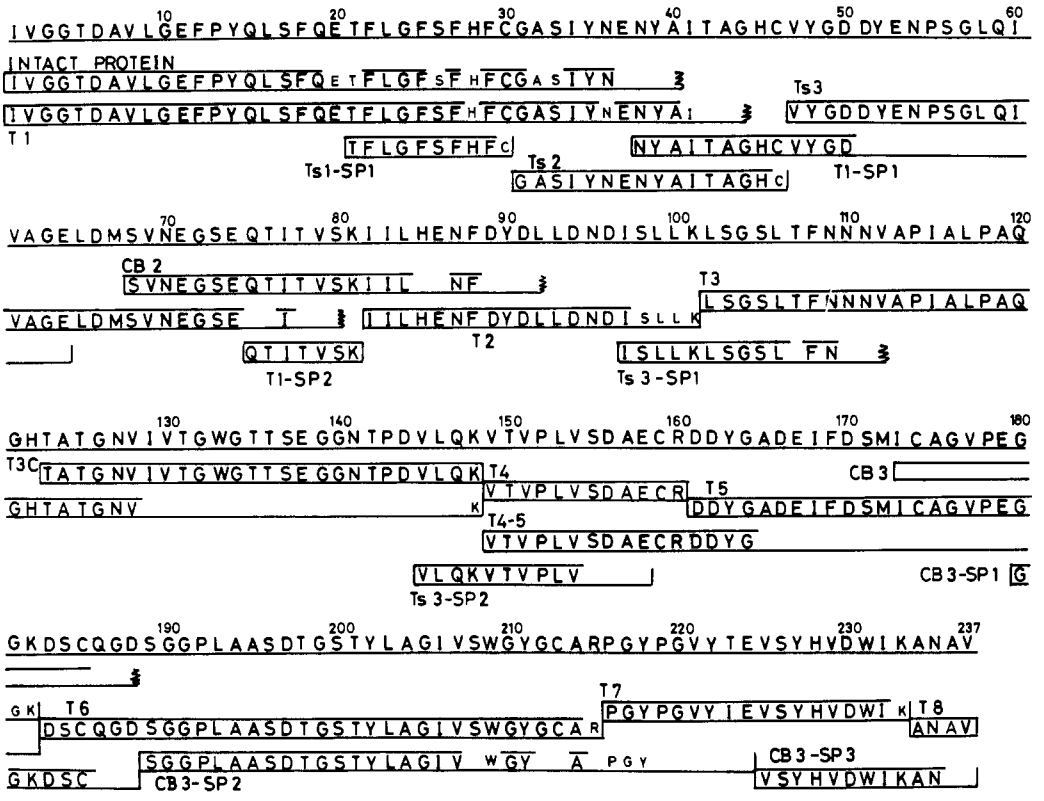

FIGURE 4: Summary proof of the sequence of crayfish trypsin I_f. The one-letter code (see Table II) within bars designates amino acids unambiguously identified by Edman degradation (capital letters) or those tentatively identified, placed by composition and cleavage specificity (small letters). The length of each bar is proportional to the length of that peptide. Enclosures of the top of a bar indicate a proven sequence; gaps in the upper enclosure signify portions of that sequence that were not identified. Tryptic peptides and cyanogen bromide fragments are designated by the prefix T and CB, respectively. Tryptic peptides after succinylation are designated by the prefix Ts. The peptides generated by staphylococcal protease are designated by the prefix SP.

His-28, which was later confirmed by the analysis of Ts1-SP1, a subpeptide of Ts1 generated by staphylococcal protease.

Peptide T1 (50 nmol) was further digested with staphylococcal protease and the digest separated by HPLC as shown in Figure 10A of the supplementary material (see paragraph at end of paper regarding supplementary material). The sequence analysis of T1-SP1 (Table I) extended the amino-terminal sequence of the whole protein up to Asp-50, confirming Ile-41 tentatively identified in the analysis of T1 (Figure 4). The carboxyl-terminal sequence Gln-Thr-Ile-Thr-Val-Ser-Lys (residues 75–81) of T1 was obtained by the analysis of T1-SP2 (Table I), another subpeptide of T1 gen-

erated by staphylococcal protease (Figure 10A, supplementary material, and Table III).

Peptide T2 was sequenced through the carboxyl terminus but identifications of the last four residues, Ser-Leu-Leu-Lys (residues 98–101), were tentative (Table I). Peptide T3 was partially sequenced through Val-128 (Table I). The sequence of T3C, a minor peptide generated by a chymotrypsin-like split, overlapped that of the last six residues of T3 and extended it to Lys-148, which appeared to be the carboxyl terminus of T3, though identifications of the last three residues were tentative (Table I).

Peptide T5 was sequenced to the carboxyl-terminal Lys-182 (Table I). The partial sequence of T4–5 (Table I) provided the overlap of T4 and T5.

Products of Cysteinyl and Arginyl Cleavage. A tryptic digest of the S-AE-succinyl protein (200 nmol) was primarily separated by gel filtration as shown in Figure 6 (supplementary material). Fraction 1 contained a homogeneous large peptide, Ts3 (residues 47–160), derived from the middle portion of the protein. Five more peptides were isolated from fractions 2 and 3 by HPLC (Figure 9A and 9B, supplementary material). The amino acid compositions of six isolated peptides, Ts1–Ts6, are listed in Table V (supplementary material) together with those of subpeptides of Ts1 and Ts3 generated by staphylococcal protease. Peptide T7 was not entirely homogeneous by composition. Presumably, a hypothetical peptide, Ts6 (residues 186–215), which was missed in the present experiment, contaminated this fraction.

As described above, the sequence analysis of subpeptide Ts1-SP1 (residues 21–30) (Table I) confirmed His-28 (Figure 4). The analysis of Ts3 (residues 47–160) (Table I) extended the amino-terminal sequence of the protein up to Ile-77, which was assigned to the third residue in T1-SP2 derived from the carboxyl-terminal portion of T1. The peptide (25 nmol) was further digested with staphylococcal protease to yield two overlapping peptides, Ts3-SP1 and Ts3-SP2, which were separated by a combination of gel filtration and HPLC as shown in Figure 9 (supplementary material). Sequence analysis of Ts3-SP1 (Table I) provided the overlap of tryptic peptides T2 and T3 and also confirmed the tentative sequence of the carboxyl-terminal portion of T2 (Figure 4). The analysis of Ts3-SP2 also provided the overlap of trytpic peptides T3 and T4 and confirmed the sequence of residues 146–148 which had been tentatively placed in T3C.

Products of Methionyl Cleavage. The last set of the sequence proof was obtained with fragments generated by cleavage of the protein with cyanogen bromide. S-AE-succinyl protein (150 nmol) was treated with cyanogen bromide in 72% formic acid for 15 h. The digest was separated by gel filtration as shown in Figure 7 (supplementary material). Although the mixture was clearly resolved into two symmetrical peaks, the first peak appeared by composition to be a mixture of two fragments, CB1 and CB2 (Table VI, supplementary material). Sequence analysis of this fraction (Table I), however, yielded only the sequence of CB2 starting with Ser-68, because the amino terminus of CB1 was blocked by succinylation prior to cleavage. This analysis provided the overlap of the primary tryptic peptides, T1 and T2, and confirmed the carboxyl-terminal sequence of T1 (Figure 4).

Fragment CB3 was homogeneous and consisted of residues 173 through 237 (the carboxyl terminus of the protein) (Table V, supplementary material). Sequence analysis of subpeptide CB3-SP1 (Table I) provided the overlap of tryptic peptides T5 and T6, confirming the carboxyl-terminal sequence of T5 (Figure 4). Two more subpeptides of fragment CB3, CB3-SP2 and CB3-SP3, were isolated by HPLC from a digest of the fragment with staphylococcal protease (Figure 10C, supplementary material) and subjected to sequence analysis (Table I), which provided the last two overlaps to complete the total sequence as shown in Figure 4.

Discussion

Trypsins from bacterial, invertebrate, and vertebrate sources are members of the same protein family. Although separated from each other by long time spans of evolution, they have nevertheless retained characteristics of their common origin. Confirmatory evidence of this conclusion is furnished by the complete amino acid sequence of crayfish (*Astacus fluviatilis*) trypsin presented in this paper, which provides the first detailed insight into the evolution of trypsins at the invertebrate level.

An earlier analysis of the partial amino-terminal sequence of a related species of crayfish trypsin (*Astacus leptodactylus*) has already revealed indications of a homologous relationship to bacterial as well as mammalian trypsins (Zwilling et al., 1975). The analysis of homology can now be extended to the entire sequence of the *A. fluviatilis* enzyme.

The sequence analysis was complicated by the fact that this enzyme occurs in multiple forms, a phenomenon which is generally characteristic of decapode crustacea. The multiple forms differ in electrophoretic and immunological properties (Linke et al., 1969). Trypsin from *A. leptodactylus* occurs in but two closely related forms, whereas in *A. fluviatilis* five different forms (I_f–V_f) can be distinguished (Pfleiderer & Zwilling, 1972). This multiplicity proved to be a major impediment in the sequence analysis since the highly specific affinity chromatographic procedure used for the isolation of this enzyme failed to resolve the isozymes and thus precluded the ordering of the fragments into a unique sequence. The problem was finally solved by subjecting the crude extract to anion-exchange chromatography, followed by gel filtration (Zwilling et al., 1981). Material prepared in this manner displayed a single band upon disc gel electrophoresis (having the lowest mobility among all multiple forms at pH 8.6) and provided the basis for the completion of the sequence. The determination of the amino-terminal sequence presented the greatest difficulties in the present study, since, in contrast to bovine trypsin, the crayfish enzyme is an acidic protein and contains only five lysyl and two arginyl groups and none among the first 80 amino-terminal residues that would be susceptible to tryptic cleavage (Figure 4). The sequence of this portion of the molecule was eventually resolved by the use of two peptides, Ts-2 and Ts-3, which were isolated by cleavage of the polypeptide chain at cysteinyl residues 30 and 46 (see Results).

Most of the sequence of *A. fluviatilis* trypsin reported in this paper is unambiguous and is based on replicate or overlapping sequence analyses. Some of the data are derived from single analyses only, but in each such case, the sequence data are consistent with the composition of small peptides derived from corresponding regions. The weakest point in the present proof of structure appears to be the Arg-Pro overlap between peptides T6 and T7 (residues 215–216) which unexpectedly was cleaved by trypsin. Analysis of peptide CB3-SP2 failed to provide an unambiguous proof for the Arg-Pro overlap. Identification of Arg-215 is based on a combination of a tentative Pth-Arg in the sequence analysis and the composition of peptide T6. However, this portion of the sequence (residues 205–226, Figure 4) is identical with that of the *Streptomyces griseus* enzyme.

Bovine and crayfish trypsins differ in chain length (223 and 237 residues, respectively). This difference is accounted for

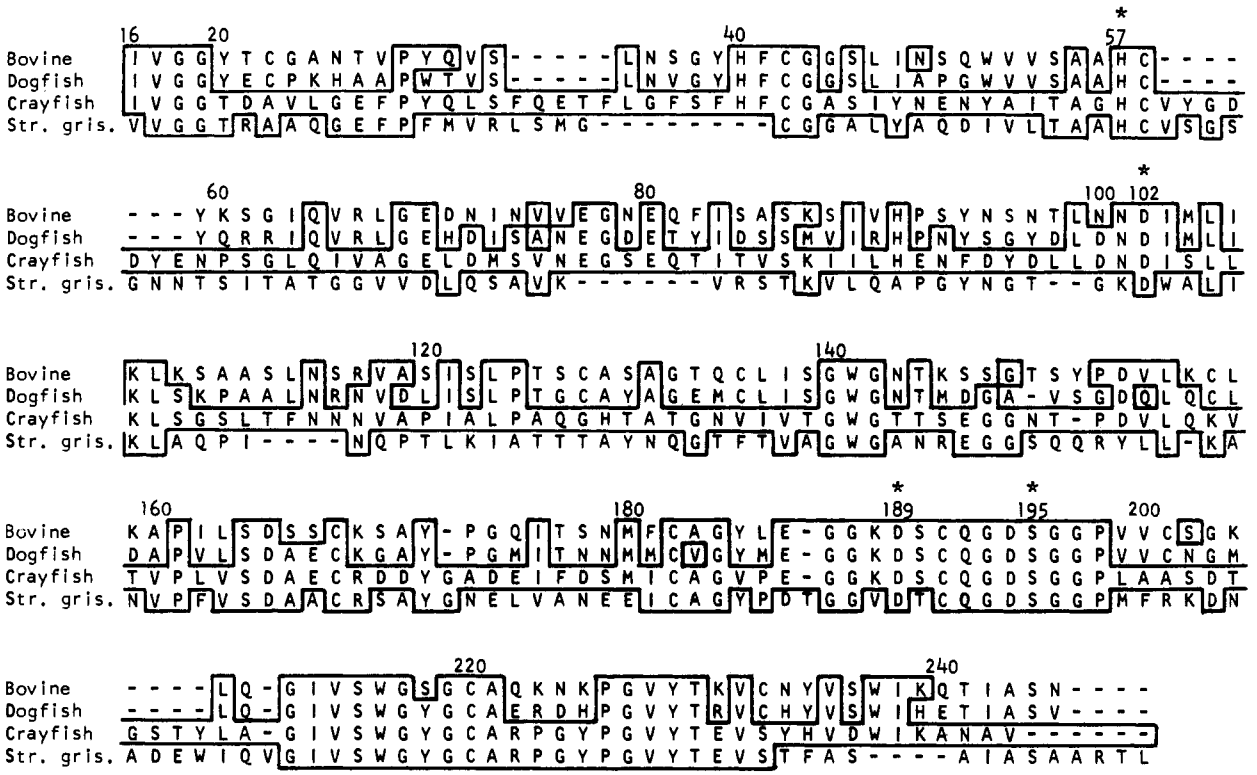

FIGURE 5: Comparison of sequences of prokaryotic, invertebrate, and vertebrate trypsins. The alignment is based on de Haën et al. (1975), and the scoring scheme is that of Dayhoff (1972). Residues identical with those in crayfish trypsin I_f are enclosed. Members of the catalytic triad and the primary substrate specificity determining Asp-189 are indicated by an asterisk. The bovine chymotrypsinogen numbering system is used (Titani et al., 1975).

Table IV: PAM[a] Computed from Percent Differences

trypsin	bovine	dogfish	crayfish	*S. griseus*
bovine	0.0	48.2	111.2	151.2
dogfish	48.2	0.0	103.1	141.5
crayfish	111.2	103.1	0.0	136.8
S. griseus	151.2	141.5	136.8	0.0

[a] PAM = accepted point mutation (rearranged acronym), considered a measure of the amount of evolutionary change (Dayhoff, 1972).

by four insertions and three deletions in the crayfish enzyme when compared to the bovine enzyme (see Figure 5). When the sequences of the four trypsins are optimally aligned, as shown in Figure 5, and allowance is made for insertions and deletions (de Haën et al., 1975), the following identities are calculated: bovine/crayfish, 43.6%; dogfish/crayfish, 44.7%; *S. griseus*/crayfish, 40.1%. This comparison clearly indicates structural homology of crayfish trypsin with both vertebrate and bacterial trypsins and also shows that there is a greater sequence coincidence of crayfish trypsin with the vertebrate than with the bacterial enzymes. Viewed in another way, the relationship among the primary structures of the four trypsins can be analyzed by the Dayhoff method (Dayhoff, 1972) which expresses the degree of evolutionary change in terms of "accepted point mutations" (PAMs). The results of such calculations are given in Table IV which also shows a greater sequence coincidence of crayfish trypsin with the bovine than with the *S. griseus* enzyme (111.2 PAMs as compared to 136.8 PAMs).

The number of half-cystine residues and their positions in the sequence of crayfish trypsin coincide with those of the *S. griseus* enzyme, justifying the prediction that they are paired in the same fashion. These include Cys-42–Cys-58 (His loop), Cys-168–Cys-182 (Met loop), and Cys-191–Cys-220 (Ser loop). The additional three cystine bridges in bovine trypsin (positions 22–157, 127–232, and 136–201) evidently have been acquired after the lines of development leading to decapode crustacea and to mammals, respectively, have separated from one another. On the basis of this criterion, the importance of which was pointed out by de Haën et al. (1975), crayfish trypsin must be placed more closely to bacterial than to vertebrate trypsins and must be inserted before the point of divergence of the branches leading to pancreatic elastase, chymotrypsin, and trypsin (Zwilling et al., 1980). Since the disulfide bridge Cys-136–Cys-201 is also present in rat mast cell protease II (Woodbury et al., 1978) but absent in crayfish trypsin, the invertebrate enzyme must have evolved separately from all known eukaryotic serine proteases.

As could be predicted from functional considerations, the constituents of the active site found in vertebrate trypsins are also present in corresponding positions in the crayfish enzyme. These include the components of the catalytic triad, His-57, Asp-102, and Ser-195 (bovine chymotrypsinogen numbering system), and residues Ile-16 and Asp-194 which in the other serine proteases are known to form a salt bridge stabilizing the catalytic site. Crayfish trypsin also shares with the bovine enzyme the residues which determine the primary and secondary binding sites, i.e., Asp-189, Gly-216, and Gly-226.

Comparison of crayfish and *S. griseus* trypsins shows a remarkably long stretch of coincidence in the carboxyl-terminal region (residues 212–233) which in the vertebrate enzymes is interrupted by six radical replacements (positions 222–225, 231, and 233). The earlier observations that, in contrast to bovine trypsin, crayfish trypsin is resistant to autocatalytic degradation (Zwilling et al., 1969) can now be fully understood on the basis of the present sequence analysis which shows that in position 145, which in the bovine enzyme is occupied by lysine (a point of autocatalytic cleavage), the crayfish enzyme contains a trypsin-resistant seryl residue.

All observations recorded in this paper are consistent with the idea that crayfish trypsin reflects the state of serine proteases in an important phase of its evolutionary history and that it has preserved many of the early properties characteristic of this protein family.

Acknowledgments

We thank Santosh Kumar, Roger D. Wade, and Elizabeth Blackwood for their excellent technical assistance in HPLC, amino acid analyses, and Edman degradations. We are also indebted to Dr. K. A. Walsh for encouragement and helpful discussions and to Dr. D. C. Teller for computer analysis of sequence homologies.

Supplementary Material Available

Two tables and five figures describing the further purification of TS subpeptides and subpeptides of T, Ts, and CB peptides and their amino acid compositions (5 pages). Ordering information is given on any current masthead page.

Registry No. Trypsin, 9002-07-7; trypsin (*Astacus fluviatilis* I_f reduced), 84582-73-0.

References

Brauer, A. W., Margolies, M. N, & Haber, E. (1975) *Biochemistry 14*, 3029–3035.

Cole, R. D. (1967) *Methods Enzymol. 11*, 315–316.

Crestfield, A. M., Moore, S., & Stein, W. H. (1963) *J. Biol. Chem. 245*, 1895–1902.

Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure*, Vol. 5.

De Haën, C., Neurath, H., & Teller, D. C. (1975) *J. Mol. Biol. 92*, 225–259.

Dunlap, C. E., III, Gentleman, S., & Lowney, L. I. (1978) *J. Chromatogr. 160*, 191–198.

Edman, P., & Begg, G. (1967) *Eur. J. Biochem. 1*, 80–91.

Ericsson, L. H., Wade, R. D., Gagnon, J., McDonald, R. M., Granberg, R. R., & Walsh, K. A. (1977) *Solid Phase Methods in Protein Sequence Analysis* (Previero, A., & Coletti-Previero, M. A., Eds.) pp 137–142, Elsevier/North-Holland, Amsterdam.

Hermann, J., Titani, K., Ericsson, L. H., Wade, R. D., Neurath, H., & Walsh, K. A. (1978) *Biochemistry 17*, 5672–5679.

Kunitz, M., & Northrop, J. H. (1936) *J. Gen. Physiol. 19*, 991.

Linke, R., Zwilling, R., Herbold, D., & Pfleiderer, G. (1969) *Hoppe-Seyler's Z. Physiol. Chem. 350*, 877–885.

Mahoney, W. C., & Hermodson, M. A. (1980) *J. Biol. Chem. 255*, 11199–11203.

Olafson, R. W., Jurásek, L., Carpenter, M. R., & Smillie, L. B. (1975) *Biochemistry 14*, 1168–1177.

Pfleiderer, G., & Zwilling, R. (1972) *Naturwissenschaften 59*, 396–405.

Reich, E., Rifkin, D. B., & Shaw, E., Eds. (1975) in *Proteases and Biological Control*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Tarr, G. E., Beecher, J. F., Bell, M., & McKean, D. J. (1978) *Anal. Biochem. 84*, 622–627.

Titani, K., Ericsson, L. H., Neurath, H., & Walsh, K. A. (1975) *Biochemistry 14*, 1358–1366.

Woodbury, R. G., Katunuma, N., Kobayashi, K., Titani, K., & Neurath, H. (1978) *Biochemistry 17*, 811–819.

Yaoi, Y., Titani, K., & Narita, K. (1964) *J. Biochem. (Tokyo) 56*, 222–229.

Zwilling, R., & Neurath, H. (1981) *Methods Enzymol. 80*, 633–643.

Zwilling, R., Pfleiderer, G., Sonneborn, H. H., Kraft, V., & Stucky, J. (1969) *Comp. Biochem. Physiol. 28*, 1275–1287.

Zwilling, R., Neurath, H., Ericsson, L. H., & Enfield, D. L. (1975) *FEBS Lett. 60*, 247–249.

Zwilling, R., Neurath, H., & Woodbury, R. G. (1980) in *Protides of the Biological Fluids* (Peeters, H., Ed.) pp 115–118, Pergamon Press, Oxford.

Zwilling, R., Dörsam, H., Torff, H.-J., & Rödl, J. (1981) *FEBS Lett. 127*, 75–78.