Perutz, M. F., & Mathews, F. S. (1966) *J. Mol. Biol. 21*, 199–202.
Peticolas, W. L., Nafie, L., Stein, P., & Fanconi, B. (1970) *J. Chem. Phys. 52*, 1576–1584.
Scheidt, W. R. (1978) *Porphyrins 3*, 463–511.
Shelnutt, J. A., O'Shea, D. C., Yu, N.-T., Cheung, L. D., & Felton, R. H. (1976) *J. Chem. Phys. 64*, 1156–1165.
Shelnutt, J. A., Cheung, L. D., Chang, R. C. C., Yu, N.-T., & Felton, R. H. (1977) *J. Chem. Phys. 66*, 3387–3398.
Sono, M., & Asakura, T. (1975) *J. Biol. Chem. 250*, 5227–5232.
Spaulding, L. D., Chang, R. C. C., Yu, N.-T., & Felton, R. H. (1975) *J. Am. Chem. Soc. 97*, 2517–2525.
Spiro, T. G., & Strekas, T. C. (1974) *J. Am. Chem. Soc. 96*, 338–345.
Spiro, T. G., Stong, J. D., & Stein, P. (1979) *J. Am. Chem. Soc. 101*, 2648–2655.
Stryer, L., Kendrew, J. C., & Watson, H. C. (1964) *J. Mol. Biol. 8*, 96–104.
Teale, F. W. J. (1959) *Biochim. Biophys. Acta 35*, 543.
Tsubaki, M., Nagai, K., & Kitagawa, T. (1980) *Biochemistry 19*, 379–385.
Wright, P. G., Stein, P., Burke, J. M., & Spiro, T. G. (1979) *J. Am. Chem. Soc. 101*, 3531–3535.
Yonetani, T., & Asakura, T. (1969) *J. Biol. Chem. 244*, 4580–4588.
Yu, N.-T. (1977) *CRC Crit. Rev. Biochem. 4*, 229–280.
Yu, N.-T., & Srivastava, R. B. (1980) *J. Raman Spectrosc. 9*, 166–171.

Amino Acid Sequence of a Collagenolytic Protease from the Hepatopancreas of the Fiddler Crab, *Uca pugilator*†

Gregory A. Grant,* Kenneth O. Henderson, Arthur Z. Eisen, and Ralph A. Bradshaw

ABSTRACT: The amino acid sequence of a collagenolytic protease from the hepatopancreas of the fiddler crab, *Uca pugilator*, was determined from the structures of overlapping tryptic, chymotryptic, thermolytic, staphylococcal protease, and cyanogen bromide peptides together with automated sequencer analysis of the intact protein. Crab collagenase is a serine protease composed of 226 residues which is capable of degrading the native triple helix of collagen under physiological conditions. When aligned for optimal homology, crab collagenase displays 35% identity with bovine trypsin, 38% with bovine chymotrypsin B, and 32% with porcine elastase. The six half-cystinyl residues in crab collagenase correspond to those forming three of the five disulfide bonds in chymotrypsin. The residues forming the charge relay system of the active site of chymotrypsin (His-57, Asp-102, and Ser-195) are found in corresponding regions in crab collagenase, and the sequences around these residues are well conserved. The primary structure of crab collagenase is the first reported for a serine protease from crustacean hepatopancreas and the first reported for a serine protease possessing the unusual property of being able to degrade native helical collagen.

The proteolytic enzyme isolated from the hepatopancreas of the fiddler crab, *Uca pugilator* (Eisen & Jeffrey, 1969; Eisen et al., 1970, 1973), is capable of degrading collagen under conditions that do not denature the protein and is characterized by the cleavage of the native collagen helix at several sites in the area of the TC^{A}_{75} locus of the molecule (Eisen & Jeffrey, 1969). Unlike vertebrate collagenases, however, the crab hepatopancreas collagenase exhibits trypsin- and chymotrypsin-like activities, determined with synthetic substrates, as inherent properties of the same molecule (Eisen et al., 1973). In addition, this collagenolytic protease resembles the vertebrate serine proteases with respect to inhibition by diisopropyl fluorophosphate and soybean trypsin inhibitor, pH optimum, and approximate molecular weight (ca. 25000) (Eisen et al., 1973).

The complete amino acid sequence of this enzyme has been determined from overlapping cyanogen bromide, tryptic, chymotryptic, thermolytic, and *Staphylococcus aureus* V8 protease peptides. The sequence unequivocally demonstrates that this collagenolytic enzyme is a serine protease of the trypsin family. In this respect, it differs from vertebrate collagenases which are zinc metalloenzymes (Seltzer et al., 1977) and is unique in that no other known serine protease is capable of degrading the native triple helix of collagen. This sequence is the first reported for a serine protease from crustacean hepatopancreas.

Materials and Methods

The enzyme was obtained from live fiddler crabs as previously described (Eisen et al., 1973). L-1-(Tosylamido)-2-phenylethyl chloromethyl ketone treated trypsin, chymotrypsin, and carboxypeptidases A and B were purchased from Worthington Biochemical Corp. Thermolysin was purchased from Calbiochem and *S. aureus* V8 protease from Miles. Carboxypeptidase Y and phenyl isothiocyanate for manual sequencing were obtained from Pierce Chemical Co. *o*-Iodosobenzoate was purchased from Chemical Dynamics and β-mercaptoethanol from Eastman. Iodoacetic acid, trifluoroacetic acid, and dansyl chloride were purchased from Sigma Chemical Co., and cyanogen bromide and diisopropyl fluorophosphate were from Aldrich Chemical Co. Iodo-[^{14}C]acetic acid was obtained from New England Nuclear,

† From the Division of Dermatology, Department of Medicine (G. A.G., K.O.H., and A.Z.E.), and the Department of Biological Chemistry (G.A.G. and R.A.B.), Washington University School of Medicine, St. Louis, Missouri 63110. *Received April 3, 1980.* This investigation was supported by U.S. Public Health Service Grants AM 13362, AM 12129, and AM 07284.

0006-2960/80/0419-4653$01.00/0

and diisopropyl fluoro[^{32}P]phosphate was from Amersham/Searle. Polyamide thin-layer plates were from Pierce, and silica gel plates were from Analtech Inc. Dowex 50-X8 was from Beckman (Type AA15), and Dowex 1-X2 was from Sigma (200–400 mesh). All other sequencing reagents were purchased from Beckman, and all other materials were of the highest grade commercially available.

Pyridine was redistilled after the addition of solid ninhydrin, and iodoacetate was recrystallized from petroleum ether before use.

Cyanogen Bromide Digestion and Carboxymethylation. Lyophilized enzyme pretreated with diisopropyl fluorophosphate was dissolved in 75% trifluoroacetic acid (1 μmol/mL), and a 30-fold excess (by weight) of solid cyanogen bromide was added. After incubation in the dark at room temperature for 24 h, the reaction mixture was diluted fivefold with water, flushed with nitrogen, and lyophilized. The protein was dissolved in 8–10 mL of 0.1 M ammonium bicarbonate and 6 M guanidine hydrochloride and the pH adjusted to 9.5 with 6 N NaOH. A fivefold molar excess of dithiothreitol to protein thiol groups was added, and the mixture was incubated at room temperature for 2 h. After the pH was adjusted to 8.0 with 6 N HCl, a mixture of iodoaceic acid (5:1 molar excess of iodoacetic acid to total thiol groups) and 50 μCi of iodo[^{14}C]acetic acid was added and the pH maintained at 8.0 with 6 N NaOH. The alkylation was stopped after 5 min by the addition of a fivefold molar excess of β-mercaptoethanol to iodoacetic acid and the pH was maintained at 7.0 for 10 min. The S-[^{14}C]carboxymethylated cyanogen bromide fragments were then purified by chromatography on a 2.6 × 300 cm column of Sephadex G-75 in 0.1 M ammonium bicarbonate, pH 8.0.

Proteolytic Digests. Enzymatic digestion of the carboxymethylated cyanogen bromide fragments with trypsin and chymotrypsin was performed in 0.1 M ammonium bicarbonate, pH 8.8, at 37 °C. The appropriate enzyme (1% w/w) was added at zero time and again at 2 h, and the reaction was allowed to proceed for 18–20 h. The reaction was stopped either by acidification to pH 2.0 or by direct gel filtration chromatography.

Following repeated treatment with diisopropyl fluorophosphate, carboxymethylation of the whole protein was performed as described for the cyanogen bromide peptides. Tryptic digestion of the whole carboxymethylated enzyme was performed as described for the cyanogen bromide peptides.

Thermolytic digestion of the carboxymethylated cyanogen bromide fragments was performed in 0.1 M ammonium bicarbonate, pH 8.0, in the presence of 2 mM $CaCl_2$ at 37 °C. Thermolysin (2% w/w) was added at zero time, and the incubation was stopped after 4 h.

The *o*-iodosobenzoate digestion of cyanogen bromide peptide C3 was carried out in 3 mL of 70% acetic acid and 4 M guanidine hydrochloride by using a 2:1 w/w ratio of reagent to peptide in the dark at room temperature for 24 h. The reaction was stopped by adding a fivefold molar excess of β-mercaptoethanol over *o*-iodosobenzoate. Insoluble peptide material which resulted was redissolved in 8 M urea, the pH was adjusted to 8.8 with NaOH, 1% w/w trypsin was added, and the reaction was allowed to proceed at room temperature. After 30 min another 1% w/w trypsin was added and the reaction was stopped 30 min later by acidification to pH 4.0 with 1 N acetic acid. *S. aureus* V8 protease (3% w/w) was then added and the mixture incubated for 18 h at 37 °C. After the pH was adjusted to 8.0 with NaOH the sample was chromatographed on a 2.6 × 200 cm column of Sephadex G-75 in 0.1 M ammonium bicarbonate, pH 8.0.

Peptide Separation. Gel filtration was performed on columns of Sephadex G-25, G-50, or G-75 equilibrated and developed with either 0.1 M ammonium bicarbonate, pH 8.0, or 0.1 N acetic acid. Eluants were monitored by absorption at 230 nm. Ion-exchange chromatography was performed on columns of Dowex 50-X8, Dowex 1-X2, or sulfoethyl-Sephadex developed with pyridine acetate buffers as described previously (Bradshaw et al., 1969; Grant & Bradshaw, 1978). Peptides were detected by the ninhydrin reaction following alkaline hydrolysis on an automated peptide analyzer (Hill & Delaney, 1969). A portion of the column effluent (5–10%) was diverted to the peptide analyzer, and the remainder was collected. Ion-exchange chromatography was also performed on columns of DEAE-cellulose eluted with linear gradients of ammonium bicarbonate. Eluants were monitored by absorption at 230 nm. Preparative high-voltage paper electrophoresis was performed with a Gilson Model D electrophorator at 2500 V for $1^3/_4$ h in pyridine acetate buffer, pH 3.75. Peptides were spotted horizontally across the width of the paper, and migration distance was detected by ninhydrin treatment of a strip removed from each side of the paper. Unreacted peptides were eluted from the remaining paper with 0.1 N acetic acid.

Amino Acid Analysis. Samples of peptide or protein were hydrolyzed with 6 N HCl in evacuated sealed glass tubes at 110 °C for 24–96 h. Analyses were performed on Beckman 119, 120, or 121 amino acid analyzers. The presence of tryptophan in peptides was determined with Ehrlich's reagent (Glazer et al., 1975).

Sequencing Procedures. The amino-terminal sequences of the protein and the S-carboxymethylated peptides were determined by automated Edman degradation in a Beckman 890C sequencer using 0.1 or 0.33 M Quadrol buffer. Some peptides were sequenced in the same manner with the addition of polybrene to the sample (Klapper et al., 1978) or by the dansyl–Edman procedure (Gray, 1972). Phenylthiohydantoins were identified, after conversion from the anilinothiazolinones, by gas–liquid chromatography (Pisano & Bronzert, 1969) or thin-layer chromatography (Jacobs & Niall, 1975). Amides were identified directly from thin-layer chromatography. Histidine and arginine were identified by spot tests or by high-pressure liquid chromatography. Analyses were performed on a Du Pont C-18 column run at 80 mL/h. The column was equilibrated in a solution of 20% acetonitrile and 80% 0.1 M sodium phosphate, pH 2.1, and elution was achieved by increasing the acetonitrile concentration of the eluant (3%/min) to a final concentration of 50%. The column was monitored at 269 nm. Dansyl amino acids were identified by two-dimensional thin-layer chromatography on polyamide sheets (Hartley, 1970). Carboxy-terminal residues were determined by hydrolysis with carboxypeptidases A or B at pH 8.5 (Shearer et al., 1967) or carboxypeptidase Y at pH 5.5. (Hayashi et al., 1973).

Results

The proposed structure for the crab collagenolytic protease is shown in Figure 1. The structure was constructed from peptides derived from cyanogen bromide (C), tryptic (T), chymotryptic (Ch), thermolytic (Th), and *S. aureus* V8 protease (S) hydrolyses. Large peptides produced by digestion of the whole enzyme are designated by a single letter (C or T). Peptides produced by subdigestion of the cyanogen bromide or tryptic fragments are designated by the letter corresponding to the parent peptide followed by additional letters indicating the enzyme(s) used for the subdigestion. The

Table I: Amino Acid Composition of Crab Collagenase

	amino acid analysis[a]			integral	
	24 h	48 h	96 h	value	sequence
Asp	28.0	28.0	28.0	28	28
Thr	19.9	21.5	19.5	22	22
Ser	13.3	12.8	10.8	14	14
Glu	12.7	12.6	12.9	13	12
Pro	12.2	12.5	12.6	12	12
Gly	25.9	26.2	26.1	26	27
Ala	18.8	18.8	18.8	19	20
Val	16.7	19.4	20.5	21	24
$^1/_2$-Cys	5.6	6.1	5.9	6	6
Met	2.8	3.1	2.9	3	3
Ile	13.4	15.2	16.1	16	17
Leu	12.1	12.5	12.5	13	13
Tyr	6.4	7.7	6.8	7	7
Phe	7.6	6.9	7.5	8	8
His	3.2	3.5	3.6	4	4
Lys	0.9	1.1	1.0	1	1
Arg	3.0	3.6	3.9	4	4
Trp				4	4
					226

[a] Amino acid analyses were determined from duplicate 24-, 48-, and 96-h hydrolyses. Serine and threonine were extrapolated to zero time hydrolysis, $^1/_2$-cystine was determined as *S*-(carboxymethyl)cysteine, methionine was determined as the sulfone after performic acid oxidation, and tryptophan was estimated spectrophotometrically (Edelhoch, 1967).

peptides produced by a particular digestion are listed with arabic numerals, following the letter designating the type of digestion.

The proposed structure contains 226 amino acids and is in good agreement with the composition determined by amino acid analysis listed in Table I. The calculated molecular weight from the sequence is 23 505.

Sequencer Analysis of the Intact Enzyme. Sequencer analysis of the intact protein identified 41 residues. This analysis, which is summarized in Table II, yielded the location of one of the methionyl residues which allowed the alignment of two of the four cyanogen bromide fragments (vide infra) and placed them at the amino-terminal segment of the protein.

Isolation of Fragments. Cyanogen bromide digestion of the protein yielded four peptides, C1–C4. These were completely separated and obtained in homogeneous form by chromatography on Sephadex G-75 (Figure 2). The radioactivity profile of the S-[^{14}C]carboxymethylated peptides indicates that all of the half-cystinyl residues reside in C2 and C4. Amino acid analyses indicate that four of the six half-cystinyl residues are contained in C4 and the remaining two are in C2. When the protein was treated with diisopropyl fluoro[^{32}P]phosphate prior to cyanogen bromide digestion, all of the ^{32}P radioactivity was found in C4, indicating that this peptide contained the seryl active-site residue. Amino acid analysis also revealed that only C4 lacked homoserine and thus indicated that it was the carboxy-terminal peptide. This observation, along with the automated sequencer analysis of the whole protein, allowed the alignment of the cyanogen bromide peptides and provided the basis upon which the entire sequence is constructed.

Trypsin digestion of S-[^{14}C]carboxymethylated protein yielded several peptides after chromatography on Sephadex G-75.[1] The first major peak contained a single peptide that corresponded to the amino-terminal segment of the protein. Although this peptide did not contain any radioactive label, it did contain half-cystinyl residues 26 and 42. These residues

[1] See paragraph at end of paper regarding supplementary material.

Table II: Quantitation of Automated Sequencer Analyses

	crab collagenase[a]		C3[a]		C4	
position	residue	yield[e]	residue	yield[e]	residue	yield[e]
1	Ile	144	Asp	393	Asn	*b*
2	Val	143	Gly	201	Ser	*b*
3	Gly	26	Ala	334	Ala	360
4	Gly	24	Gly	257	Asp	140
5	Val	105	Phe	316	Cys	*c*
6	Glu	108	Val	307	Asp	197
7	Ala	158	Asp	343	Ala	330
8	Val	97	Val	262	Val	315
9	Pro	46	Val	301	Tyr	*b*
10	Asn	*b*	Leu	328	Gly	170
11	Ser	*b*	Gly	244	Ile	360
12	Trp	*b*	Ala	264	Val	210
13	Pro	18	His	*d*	Thr	*b*
14	His	*d*	Asn	*b*	Asp	165
15	Gln	*b*	Ile	306	Gly	75
16	Ala	92	Arg	*d*	Asn	*b*
17	Ala	98	Glu	301	Ile	245
18	Leu	58	Asp	267	Cys	*c*
19	Phe	48	Glu	275	Ile	120
20	Ile	41	Ala	222	Asp	27
21	Asp	70	Thr	*b*		
22	Asp	75	Gln	*b*		
23	Met	44	Val	152		
24	Tyr	*b*	Thr	*b*		
25	Phe	31	Ile	154		
26	Cys	*c*	Gln	*b*		
27	Gly	24	Ser	*b*		
28	Gly	25	Thr	*b*		
29	Ser	*b*	Asp	107		
30	Leu	32	Phe	128		
31	Ile	51	Thr	*b*		
32	Ser	*b*	Val	115		
33	Pro	11	His	*d*		
34	Glu	30	Gln	*b*		
35	Trp	*b*	Asn	*b*		
36	Ile	39	Tyr	*b*		
37	Leu	46	Asn	*b*		
38	Thr	*b*	Ser	*b*		
39	Ala	23	Phe	37		
40	Ala	28	Val	65		
41	His	*d*	Ile	43		
42			Ser	*b*		
43			Asn	*b*		
44			Asp	40		
45			Ile	*b*		
46			Ala	33		
47			Val	14		
48			Val	*b*		
49			Arg	*e*		

[a] Sequenced in the presence of polybrene. [b] Not quantitated. Identified by thin-layer chromatography. [c] Not quantitated. Identified by scintillation counting as *S*-([^{14}C]carboxymethyl)cysteine. [d] Not quantitated. Identified by chemical spot test. [e] Values are expressed as nanomoles. Approximately 200 nmol of collagenase, 450 nmol of C3, and 400 nmol of C4 were applied to the sequencer.

are apparently resistant to reduction and alkylation while the polypeptide chain of the protein is still intact. Cleavage of the protein with CNBr prior to alkylation, however, renders these residues susceptible to attack (Figure 2, C2). Peptide T1 was also obtained from this digest. Sequencer analysis of this peptide revealed the overlap of C3 into C4 and thus confirmed the alignment of the cyanogen bromide peptides. The remaining tryptic fragments, which could be only partially resolved by DEAE-cellulose chromatography, failed to yield any additional information.

As seen in Figure 1, the automated sequencer analysis of the intact protein gave the complete sequence of the amino-terminal cyanogen bromide fragment C1 and supplied the overlap into C2. The complete sequence of C2 was determined

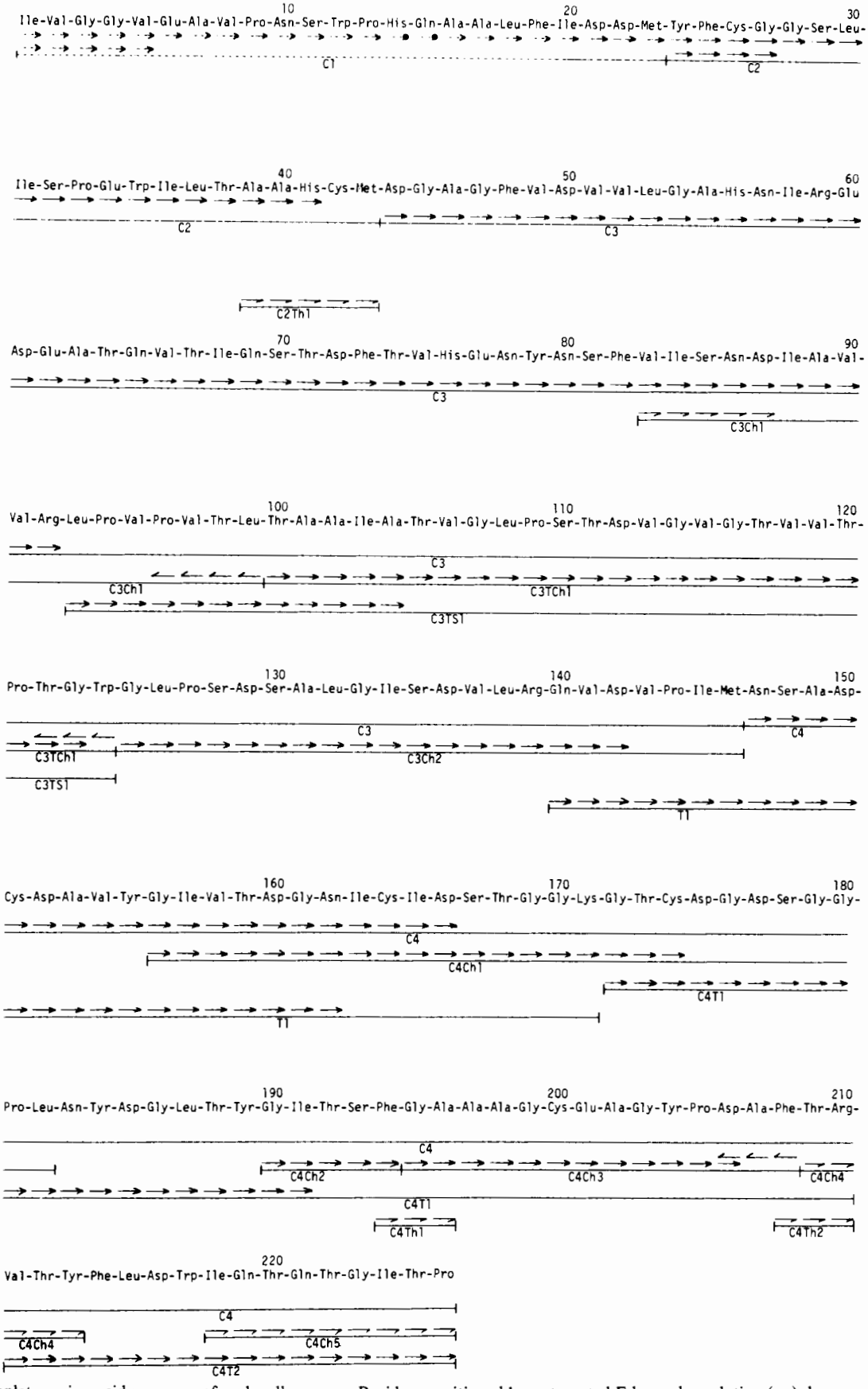

FIGURE 1: Complete amino acid sequence of crab collagenase. Residues positioned by automated Edman degradation (→), by manual Edman degradation (→), and with carboxypeptidase (—) are indicated. Peptides are designated as described in the text.

from this analysis and from peptides obtained from a thermolytic digestion of C2. These peptides were isolated by chromatography on Dowex 50-X8 and by high-voltage paper electrophoresis. Peptide C2Th1 was the carboxy-terminal peptide, contained the second half-cystinyl residue, and gave the necessary overlap to complete the sequence of C2.

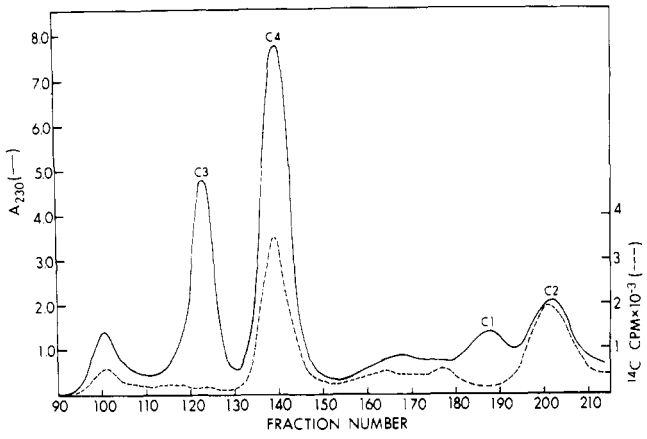

FIGURE 2: Purification of cyanogen bromide peptides. Cyanogen bromide cleavage of 5 μmol of crab collagenase was performed as described in the text. The elution profile shown is from chromatography of the digest on a column (2.5 × 300 cm) of Sephadex G-75 in 0.1 M ammonium bicarbonate, pH 8.0. The column flow rate was 30 mL/h and 15-min fractions were collected. Peptides were detected by absorbance at 230 nm (—) and by liquid scintillation counting (---) of radiolabeled *S*-([^{14}C]carboxymethyl)cysteine-containing peptides.

The structure of cyanogen bromide fragment C3 was constructed from automated sequencer analysis of C3 (summarized in Table II) and from overlapping peptides obtained by tryptic and chymotryptic digestion of C3, *o*-iodosobenzoate digestion followed by tryptic and *S. aureus* protease digestion of C3, and tryptic digestion of the whole protein. Tryptic digestion of C3 produced three peptides which were partially separated on Sephadex G-75. Two large peptides corresponding to residues 44–92 and 93–139 coeluted on Sephadex G-75 and were not further resolved. The third peptide corresponded to residues 140–146. Under the conditions used for the tryptic digestion, cleavage at arginyl-59 did not occur.

The peptides produced by chymotryptic digestion of the large tryptic peptides of C3 were separated on Sephadex G-50 followed by sulfoethyl-Sephadex chromatography. This yielded several peptides, only one of which, C3TCh1, was needed to complete the sequence of C3 (Figure 1).

Chromatography of the chymotryptic digest of C3 on Dowex 50-X8 followed by chromatography on Dowex 1-X2 yielded several peptides. Two of these peptides, C3Ch1 and C3Ch2, were useful in completing the sequence of this fragment. C3Ch1 overlapped the amino-terminal sequence of C3, and C3Ch2 overlapped into T1 to gain the remaining carboxy-terminal sequence of C3.

Sephadex G-75 chromatography of the tryptic and *S. aureus* protease digest of the urea-solubilized precipitate obtained from the *o*-iodosobenzoate digestion of C3 produced a pool which contained two major peptides. One of these peptides originated from the amino terminal of C3 and consisted of residues 44–62. Knowledge of the sequence of this peptide from sequencer analysis of C3 allowed the sequence of the other peptide, C3TS1, to be deduced and revealed the overlap between C3Ch1 and C3TCh1. In addition, C3TS1, along with C3Ch1, completed the overlap of the amino-terminal sequence of C3 with the rest of the peptide.

The structure of cyanogen bromide fragment C4 was constructed from overlapping peptides obtained by tryptic, chymotryptic, and thermolytic digestion of C4. Sephadex G-50 chromatography of the chymotryptic digestion of C4 produced two major pools. Chromatography of the first pool on QAE-Sephadex yielded a single clean peptide, C4Ch1, that contained the seryl active-site residue and overlapped with the amino-terminal sequence of C4. Chromatography of the second pool on Dowex 50-X8 produced two peptides, C4Ch3 and C4Ch5, which coeluted at the beginning of the column and were cleanly separated by chromatography on Dowex 1-X2. C4Ch2 and C4Ch8 also coeluted and were separated on Dowex 1-X2.[1] Eight peptides suitable for sequence analysis were obtained from this digest. As seen in Figure 1, five of these, C4Ch1–C4Ch5 were needed to complete the sequence of C4.

The tryptic digest of C4 was chromatographed on Sephadex G-75 and yielded two major pools. Pool II contained a single peptide, C4T2, which originated from the carboxy-terminal end of C4 and provided the overlap between C4Ch4 and C4Ch5. Pool I was subsequently chromatographed on DEAE-cellulose and yielded at least two peptides. One of these, C4T1, contained the seryl active-site residue, completed the sequence around this residue, and overlapped C4Ch1 and C4Ch2.

The thermolytic digest of C4, chromatographed on Dowex 50-X8, followed by Dowex 1-X2 chromatography of heterogeneous pools, yielded at least 12 peptides, 2 of which, C4Th1 and C4Th2, gave the necessary overlaps to complete the sequence of C4.

Carboxypeptidase Digestion. Carboxypeptidase Y digestions of C3Ch1, C3TCh1, and C4Ch3 confirmed the carboxy-terminal sequences of these peptides and were consistent with their compositions. Carboxypeptidase Y digestion of the whole protein or of C4, however, failed to confirm the carboxy-terminal sequence of the whole protein as deduced from the peptide alignments in C4. Although proline was identified as a major residue released in these digestions, several other residues that are not present at the carboxyl end of the deduced sequence were also released in similar yield. This may have been due to endopeptidase activity present in commercial preparations of this enzyme that would manifest itself to a greater degree on large peptide substrates or to the possibility that there are additional residues present at the carboxy terminal of this protein. However, the observations that trypsin, chymotrypsin, and thermolysin all produced peptides (C4T2, C4Ch5, and C4Th10,[1] respectively) corresponding to this region that terminate with a prolyl residue support the conclusion that the protein terminates with the sequence presented in Figure 1. This is also consistent with the composition of the protein (Table I).

Amide Assignments. All amide assignments were made by thin-layer chromatography of the phenylthiohydantoins during the sequence analysis of the respective peptides (Jacobs & Niall, 1975).

All aspartic acid and glutamic acid residues encountered in the dansyl–Edman procedure were subsequently encountered during automated sequence analysis which allowed the proper amide assignments to be made.

Discussion

An unambiguous structure for the 226 residues comprising crab collagenase can be constructed from the peptides described. The four large peptides produced by cyanogen bromide cleavage at the three methionine residues have been used as the basis upon which the entire sequence of the enzyme was constructed. Automated sequencer analysis provides identification of the first 41 residues and the overlap of the first two cyanogen bromide peptides. Thermolytic peptides derived from C2 complete the structure of C2 and extend the sequence to Met-43. Tryptic peptide T1 provides the overlap of the remaining two cyanogen bromide peptides and confirms the alignment of these peptides. Although chemical overlap of C2 to C3 is lacking, the overlap of C1 to C2 and C3 to C4

```
           16      20                                            40                                      60
           1                                     20                                              40
Coll       I V G G V E A V P N S W P H Q A A L F I D D M Y - - - - F C G G S L I S P E W I L T A A H C M D G A G F V D V V L G A
CB bovine  I V N G E D A V P G S W P W Q V S L Q D S T G F - - - H F C G G S L I S E D W V V T A A H C G V T T S D V - V V A G E
Tr bovine  I V G G Y T C G A N T V P Y Q V S L - - N S G Y - - - H F C G G S L I N S Q W V V S A A H C Y K S G I Q V R L - - G Q
E porcine  V V G G T E A Q R N S W P S Q I S L Q Y R S G S S W A H T C G G T L I R Q N W V M T A A H C V D R E L T F R V V V G E

                             80                                        100                                         120
                     60                                      80                                          100
Coll       H N I R - E D E A T Q V T I Q S T D F T V H E N Y N S F V I S - - N D I A V V R L P V P V T L T A A I A T V G L P S T
CB bovine  F D Q G L E T E D T Q V L K - I G K V F K N P K F S I L T V R - - N D I T L L K L A T P A Q F S E T V S A V C L P S -
Tr bovine  D N - N V V E G N E Q F I S - A S K S I V H P S Y N S N T L N - - N D I M L I K L K S A A S L N S R V A S T S L P T S
E porcine  H N L N Q N N G T E Q Y V G - V Q K I V V H P Y W N T D D V A A G Y D I A L L R L A Q S V T L N S Y V Q L G V L P R -

                                         140                                       160
                                 120                                       140                                               160
Coll       D V G V - - - - G T V V T P T G W G L P S - D S A L G I S D V L R Q V D V P I M - N S A D C D - A V Y - G I - V T D G
CB bovine  - A D E D F P A G M L C A T T G W G K - T K Y N A L K T P D K L Q Q A T L P I V S N T - D C R K - - Y W G S R V T D V
Tr bovine  C A S - - - - A G T Q C L I S G W G N - T K S S G T S Y P D V L K C L K A P I L S N S - S C K - S A Y P G - Q I T S N
E porcine  - A G T I L A N G S P C Y I T G W G L - T R T N G Q L - A Q T L Q Q A Y L P T V D Y A - I C S S S S Y W G S T V K N S

          180                                          200                                       220
                                                 180                                             200
Coll       N I C I D - S T G G K G T C D G D S G G P L N Y - - D G - - L T Y G I T S F G A A A G C - E A G Y P D A F T R V T Y F
CB bovine  M I C A G - A S G V - S S C M G D S G G P L V C Q K N G A W T L A G I V S W G S S - T C S - T S T P A V Y A R V T A L
Tr bovine  M F C A G Y L E G G K D S C Q G D S G G P V V C - - S G K - - L Q G I V S W G S - - G C A Q K N K P G V Y T K V C N Y
E porcine  M V C A G - G N G V R S G C Q G D S G G P L H C L V N G Q Y A V H G V T S F V S R L G C N V T R K P T V F T R V S A Y

                       240
                     220
Coll       L D W I Q T Q T G I - T P
CB bovine  M P W V Q - E T - L A A N
Tr bovine  V S W I K - Q T - I A S N
E porcine  I S W I N - N V - I A S N
```

FIGURE 3: Comparison of the sequence of crab collagenase (Coll) with those of bovine trypsin (Tr bovine), bovine chymotrypsin B (CB bovine), and porcine elastase (E porcine). The sequences of these serine proteases, other than that of crab collagenase, were taken from Woodbury et al. (1978) which also gives an explanation of the one letter code. Residues are numbered according to the chymotrypsin numbering system (top) and as they occur in crab collagenase as presented in Figure 1 (bottom).

leaves this as the only possibility. Furthermore, this is consistent with the amino acid composition of the enzyme and the homology to the pancreatic serine proteases. Automated sequencer analysis provides identification of the first 49 residues of C3. This region is extended by compositional and sequence overlap of C3Ch1 which is in turn overlapped by C3TS1. The overlap at residues 92 and 93 is further strengthened by the three tryptic peptides generated from C3. These peptides correspond to residues 44–92, 93–139, and 140–146. As already stated in the text, cleavage of Arg-59 does not occur under the conditions used. Sequence analysis shows the amino-terminal sequence of peptide 93–139 to be Leu-Pro-Val-Pro-Val. The C3 tryptic peptides account for all of the residues in C3, are consistent with the trypsin-sensitive residues in this fragment, and confirm the placement of Arg-92 contiguous to Leu-93. The overlap of C3TCh1 with C3TS1 extends the sequence to Trp-124. The overlap of C3Ch2 with T1 completes the sequence of C3 and extends the sequence of Met-146. The placement of C3Ch2 contiguous with C3TCh1 is supported by the invariance of the glycyltryptophanylglycyl sequence in other serine proteases and by additional homology on either side. The first 20 residues of C4 are identified by automated sequencer analysis, and the chymotryptic, tryptic, and thermolytic residues derived from this region are consistent with this structure. This region is overlapped by C4Ch1 which is in turn overlapped by C4T1 and extends the sequence to Ile-191. The final overlaps are provided by C4Ch2, C4Ch3, C4Ch4, C4Th1, C4Th2, and C4T2 and complete the structure of C4 and the enzyme.

When the protein sequences are optimally aligned, the degree of sequence identity of crab collagenase with bovine trypsin (35%), chymotrypsin B (38%), and porcine elastase (32%) leaves little doubt that this enzyme is a serine protease of the trypsin class (Figure 3). This is supported by the selective inhibition and modification of the enzyme at seryl-178 by diisopropyl fluorophosphate.

The residues forming the charge relay system of the active site in serine proteases (His-57, Asp-102, and Ser-195 in bovine chymotrypsin A) are found in corresponding regions in crab collagenase (His-41, Asp-87, and Ser-178), and the sequences around these residues are highly conserved.

The six half-cystinyl residues in crab collagenase correspond to 6 of the 10 half-cystinyl residues in chymotrypsin. Since none of the half-cystinyl residues in crab collagenase are present as free thiols, these residues probably form disulfide bonds identical with three of the five disulfide bonds present in chymotrypsin. In crab collagenase, these bonds would link Cys-26 to Cys-42, Cys-151 to Cys-164, and Cys-174 to Cys-200 and would correspond to Cys-42 to Cys-58, Cys-168 to Cys-182, and Cys-191 to Cys-220, respectively, in chymotrypsin. These disulfide bonds are present in all other known serine proteases except the group-specific protease from rat small intestine (Woodbury et al., 1978) which does not contain Cys-191 to Cys-220 (chymotrypsin numbering) corresponding to Cys-174 to Cys-200 in crab collagenase (Figure 1).

In addition to these similarities, there are also some obvious differences. In chymotrypsin and trypsin, serine-189 and asparatic acid-177, respectively, form the bottom of the sub-

strate binding site. In crab collagenase this position is occupied by Gly-172. At the top of the binding pocket, where trypsin and chymotrypsin contain glycyl residues (Gly-226), crab collagenase contains a corresponding aspartic acid residue (Asp-206). In elastase, a threonine residue is found in this position and is thought to participate, at least in part, in blocking the binding pocket to all but the smallest side chains on the substrate. It is not known what role these differences may play in the substrate binding interactions of crab collagenase, but they may be significant with respect to the collagenolytic activity of this enzyme.

The low isoelectric point of crab collagenase (p*I* 3.0) is also reflected in its primary structure by the low content of basic residues compared to acidic residues. Crab collagenase contains only 9 basic residues (4 histidyl, 4 arginyl, and a single lysyl at position 171) in contrast to 25 acidic residues. This is considerably lower than that of other serine proteases (p*I* 8–11) and would impart a substantial negative charge on crab collagenase at its catalytic pH optimum (pH 8.0–8.5). What significance this may have in regard to its action on collagen or its physiological role as a digestive protease is not known.

The ability of crab collagenase to attack and specifically cleave the native triple helix of collagen is a unique feature of this serine protease. In this regard, the conformational nature of the active site and the regions involved in substrate binding are of particular interest. The structural features of the enzyme may give considerable insight into the determinants of substrate specificity of this class of enzymes as well as the mammalian metalloenzyme collagenases. In addition, investigations into the nature of the interaction of crab collagenase with collagen may lead to a better understanding of the structural features of collagen that participate in the initial cleavage of that molecule at a specific site by tissue collagenase (Eisen & Jeffrey, 1969; Eisen et al., 1970). Crystallographic and chemical modification studies designed to elucidate these points are now in progress.

Acknowledgments

We thank Dat T. Phan for his excellent technical assistance.

Supplementary Material Available

Isolation of peptides by various chromatographic methods (Figures 4–17) and amino acid composition of isolated peptides and location of peptides not listed in Figure 1 (Tables III–VII) (17 pages). Ordering information is given on any current masthead page.

References

Bradshaw, R. A., Garner, W. H., & Gurd, F. R. N. (1969) *J. Biol. Chem. 244*, 2149–2158.

Edelhoch, J. (1967) *Biochemistry 6*, 1948–1954.

Eisen, A. Z., & Jeffrey, J. J. (1969) *Biochim. Biophys. Acta 191*, 517–526.

Eisen, A. Z., Bauer, E. A., & Jeffrey, J. J. (1970) *J. Invest. Dermatol. 55*, 359–373.

Eisen, A. Z., Henderson, K. O., Jeffrey, J. J., & Bradshaw, R. A. (1973) *Biochemistry 12*, 1814–1822.

Glazer, A. N., DeLange, R. J., & Sigman, D. S. (1975) in *Chemical Modification of Proteins* (Work, T. S., & Work, E., Eds.) p 183, American Elsevier, New York.

Grant, G. A., & Bradshaw, R. A. (1978) *J. Biol. Chem. 253*, 2727–2731.

Gray, W. H. (1972) *Methods Enzymol. 25*, 333–344.

Hartley, B. S. (1970) *Biochem. J. 119*, 805–822.

Hayashi, R., Moore, S., & Stein, W. H. (1973) *J. Biol. Chem. 248*, 2296–2302.

Hill, R. L., & Delaney, R. (1969) *Methods Enzymol. 11*, 339–351.

Jacobs, J. W., & Niall, H. D. (1975) *J. Biol. Chem. 250*, 3629–3636.

Klapper, D. G., Wilde, C. E., & Capra, J. D. (1978) *Anal. Biochem. 85*, 126–131.

Pisano, J. J., & Bronzert, T. J. (1969) *J. Biol. Chem. 244*, 5597–5607.

Seltzer, J. L., Jeffrey, J. J., & Eisen, A. Z. (1977) *Biochim. Biophys. Acta 485*, 179–187.

Shearer, W. T., Bradshaw, R. A., Gurd, F. R. N., & Peters, T. (1967) *J. Biol. Chem. 242*, 5451–5459.

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