Primary Structure of the Hydrophobic Plant Protein Crambin†

Martha M. Teeter,* Jonathan A. Mazer, and James J. L'Italien‡

ABSTRACT: Crambin, a hydrophobic plant seed protein, consists of a single chain of 46 amino acids with a calculated molecular weight of 4720. The primary structure was determined by using solid-phase sequencing techniques and was confirmed through X-ray crystallographic analysis of the protein at 1.5-Å resolution [Hendrickson, W. A., & Teeter, M. M. (1981) *Nature (London) 290*, 107–112]. High-performance liquid chromatographic separation of the proteolytic fragments from crambin led to the identification of two sites of microheterogeneity. The three disulfide bonds were located at positions 3–40, 4–32, and 16–26 from the crystallographic data. Comparison of the primary structure with known sequences revealed that crambin is homologous with the plant toxins purothionin and viscotoxin. Methods to estimate protein secondary structure were applied and found to predict all of crambin's structure except its amphiphilic helix.

Crambin was first isolated by Van Etten et al. (1965) from the aqueous acetone seed extract of *Crambe abyssinica*, a plant in the same family as mustard and rape (the Crucifer family). Single crystals of this water-insoluble protein were grown from aqueous ethanol solutions. These crystals were observed by one of us (M.M.T.) to diffract X-rays to a minimum interplanar distance of 0.88 Å. This is the highest "resolution" which has been seen in the crystallographic analysis of any protein to date. An interest in solving the crystal structure (Teeter & Hendrickson, 1979) of crambin prompted us to undertake its sequence analysis. In the later stages, the sequence analysis and the crystal structure determination proceeded simultaneously. This unusual situation enabled us to corroborate the primary structure by independent high-resolution X-ray diffraction analysis (Hendrickson & Teeter, 1981).

Crambin is a hydrophobic protein as seen by its solubility in organic solvent–water mixtures. Its low polarity (31%), calculated by the method of Vanderkooi & Capaldi (1972), is characteristic of intrinsic membranes and may be related to its function.

When the present work was initiated, we were aware of no relationship between crambin and any other protein. However, after the sequence was completed, a computer search (Dayhoff, 1978) revealed that crambin was homologous with the plant toxins viscotoxin (Samuelsson et al., 1968) from mistletoe and purothionin (Mak & Jones, 1976) from wheat endosperm.

Materials and Methods

Crambin was obtained from Van Etten et al. (1965) as a recrystallized aqueous acetone extract from the defatted seed meal of the plant *Crambe abyssinica*. Iodo[1-^{14}C]acetic acid was purchased from New England Nuclear, TPCK-trypsin[1] was from Worthington, and TLCK-treated chymotrypsin and carboxypeptidase A were from Sigma. Glass-distilled methanol and acetonitrile for high-performance liquid chromatography (HPLC) were obtained from either Burdick and Jackson or Baker, while all other chemicals and solvents were reagent grade or better.

Chemical Modification of Cystine. The disulfide bonds of crambin were reduced and carboxymethylated with iodo[^{14}C]acetic acid according to the method of Crestfield et al. (1963). After carboxymethylation, the protein was desalted on a Sephadex G-10 column in 0.1 M NH_4HCO_3 and lyophilized. Alternatively, some crambin was oxidized with performic acid by the method of Hirs (1956).

Enzymatic Digestion of Crambin. Lyophilized crambin, which had previously been carboxymethylated or oxidized, was digested in 0.1 M NH_4HCO_3 with TPCK-trypsin using a 1/50 enzyme/substrate weight ratio. The trypsin was added in two aliquots, and the digestion was performed at 37 °C for a total of 4 h. The reaction was stopped by quick-freezing and subsequent lyophilization. Chymotryptic digestions were performed on carboxymethylated crambin tryptic C fragment. These digestions were carried out in the same manner as the tryptic digests except that a 1/15 enzyme/substrate weight ratio was used. The peptides which resulted from the various enzymatic digestions were separated either on a column (2.0 × 6.0 cm) of DEAE-cellulose (Whatman DE-52) equilibrated with 0.1 M NH_4HCO_3 using a linear NH_4HCO_3 gradient (50 mL of 0.4 M into 50 mL of 0.01 M) of by reverse-phase HPLC.

High-Performance Liquid Chromatography of Peptides. Peptide separations were performed by using a Waters Associates system consisting of two Model 6000 A pumps, a Model 720 system controller, a Model 730 data module, and two UV detectors connected in series. Crambin–peptide mixtures (approximately 30 nmol) were dissolved in 50% acetic acid and applied directly to a Waters Associates μBondapak C_{18} column (0.39 × 30 cm). The solvent system used, a gradient of acetonitrile into 20 mM KH_2PO_4 at pH 2.2 (Ozols et al., 1980), permitted simultaneous UV detection of the peptide–amide bond (at 206–214 nm) and of the aromatic residues (at 254 or 280 nm) in the column eluant. Entire peaks were collected in individual test tubes in a fraction collector which was activated by a time-delay switch. These fractions were then neutralized with triethylamine and evaporated before amino acid analysis.

Amino Acid Analysis. The amino acid composition of purified peptides was routinely determined by acid hydrolysis in 5.7 N HCl under vacuum at 130 °C for 6 h. The hydro-

† From the Department of Chemistry, Boston University, Boston, Massachusetts 02215. *Received January 19, 1981.* This work was supported in part by National Science Foundation Grants PCM80-03929 to M. M. Teeter and PCM79-04910 to R. A. Laursen.

‡ Present address: Department of Internal Medicine, Yale Medical School, New Haven, CT.

[1] Abbreviations used: HPLC, high-performance liquid chromatography; TLC, thin-layer chromatography; TPCK-trypsin, trypsin treated with L-1-chloro-(3-(toluenesulfonamido)-4-phenyl-2-butanone; TLCK-chymotrypsin, chymotrypsin treated with L-1-chloro-(3-(toluenesulfonamido)-7-amino-2-heptanone; DEAE-cellulose, diethylaminoethylcellulose; PTH, phenylthiohydantoin; CM-Cys, carboxymethylcysteine; Δ-Ser, dehydroserine.

0006-2960/81/0420-5437$01.25/0

Table I: Amino Acid Composition of Crambin

amino acid	amino acid composition previously reported[a]	determined from amino acid analysis[b]	determined from sequence
cysteic acid	5.6	5.8	6
aspartic acid	4.6	4.0	1
asparagine			3
threonine	5.2	5.8	6
serine	2.4	2.5	2-3[c]
glutamic acid	1.0	1.1	1
glutamine			0
proline	4.3	4.4	4-5[c]
glycine	3.9	4.0	4
alanine	4.8	5.1	5
valine	1.9	1.9	2
methionine	0	0	0
isoleucine	4.3	4.0	4-5[d]
leucine	1.0	1.3	1-2[d]
tyrosine	1.9	1.4	2
phenylalanine	1.0	1.0	1
histidine	0	0	0
lysine	0	0	0
arginine	1.9	2.0	2
total no. of residues	43.8	44.3	46

[a] Van Etten et al. (1965). [b] Values shown were determined from a time course for oxidized crambin. Values for Ser and Thr are from 24-h hydrolysis, and values for Val and Ile are from 120-h hydrolysis. [c] These values include one site of serine–proline microheterogeneity. [d] These values include one site of isoleucine–leucine microheterogeneity.

lysates were dried under vacuum and then analyzed on a Beckman 119 CL amino acid analyzer using a 97-min sodium citrate program (Beckman Application Notes, 1977a). The presence of phosphate salts in those peptides which were isolated by HPLC did not adversely affect amino acid analysis. The composition of oxidized crambin was determined by acid hydrolysis at 110 °C for 24 and 120 h.

Sequencing by Solid-Phase Edman Degradation. Purified peptides (100 nmol) from crambin were routinely attached to aminopolystyrene by carbodiimide activation of the carboxyl groups, essentially by the method of Wittmann-Liebold & Lehmann (1975). Attachment yields were 30–40%. Solid-phase Edman degradation (Laursen, 1966) was performed on a Sequemat Mini-15 solid-phase sequencer using the standard programs and solvents. Conversion from thiazolones to phenylthiohydantoin (PTH) amino acids was done either manually using 1 M HCl for 10 min at 80 °C (Edman & Begg, 1967) or automatically with a Sequemat P-6 autoconverter in line with the sequencer (Horn & Bonner, 1976). Identification of phenylthiohydantoins was performed by TLC on silica gel plates (Laursen, 1971), by back-hydrolysis (Méndez & Lai, 1975), or by HPLC (Bhown et al., 1978). In the latter case, the phenylthiohydantoins were separated on an Altex Ultrasphere RP-18 column (0.39 × 15 cm) with 80 mM sodium acetate (pH 4.1) as the aqueous buffer component and methanol as the eluting solvent. Phenylthiohydantoins (1–2 nmol) were generally identified in a single chromatographic run of less than 22 min. For HPLC results, repetitive yields were 90–95%.

Confirmation of the C-Terminal Sequence by Carboxypeptidase A. The C-terminal sequence of crambin was determined in a time-course study by carboxypeptidase A digestion of crambin tryptic C fragment. The reaction was performed in NH_4HCO_3 buffer (pH 7) using a 1/10 (w/w) enzyme/substrate ratio (Potts, 1967). Aliquots were removed and quenched in pH 2.2 sodium citrate as the digestion proceeded. Samples were injected directly onto the amino acid analyzer using either our standard system for protein hydrolysates or a modified physiological fluids program in lithium citrate buffer (Beckman Application Notes, 1977b) for Asn.

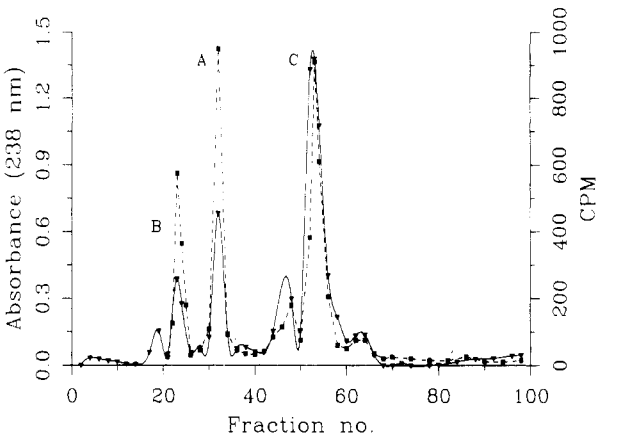

FIGURE 1: Separation of the tryptic peptides on DEAE-cellulose. The chromatographic separation was monitored by the absorbance at 238 nm (—) and by scintillation counting of [^{14}C]CM-cysteine (---).

Table II: Amino Acid Composition of the Tryptic Peptides

amino acid	tryptic A peptide	tryptic B peptide	tryptic C peptide
cysteic acid	2.1 (2)[a]	1.4 (1)	3.0 (3)
aspartic acid		1.8 (2)	2.1 (2)
threonine	1.9 (2)		4.0 (4)
serine	1.0 (1)	1.0 (1)	0.9 (1-2)[c]
glutamic acid			1.0 (1)
proline	1.1 (1)		3.4 (3-4)[c]
glycine			4.0 (4)
alanine	1.0 (1)		4.0 (4)
valine	0.8 (1)[b]	1.0 (1)	
isoleucine	0.7 (1)[b]		2.8 (4-5)[b,c]
leucine			1.0 (1-2)[c]
tyrosine			1.7 (2)
phenylalanine		0.9 (1)	
arginine	1.1 (1)	1.2 (1)	
total no. of residues	10	7	29

[a] Values in parentheses refer to the number of residues determined by sequence analysis. [b] The cleavage of Ile-7–Val-8, Ile-33–Ile-34, and Ile-34–Ile-35 is expected to be only 50% complete in the 6-h, 130 °C hydrolysis. [c] Microheterogeneity occurs here.

Results

Amino Acid Composition. The amino acid composition for oxidized crambin is summarized in Table I. The agreement with a previous report (Van Etten et al., 1965) and with the determined sequence is quite good. The presence of L-allo-isoleucine was detected by Van Etten at the level 0.11 mol of L-allo-Ile/mol of crambin for an 18-h hydrolysis and 0.29 mol of L-allo-Ile/mol of crambin for a 120-h hydrolysis. We have found at least 0.16 mol of L-alloisoleucine/mol of crambin in a 96-h hydrolysis and less at shorter hydrolysis times. Van Etten et al. (1965) suggested L-allo-Ile was produced by an acid-catalyzed rearrangement during a 120-h hydrolysis. When we heat Ile alone under similar conditions for 120 h, approximately 0.02 mol of L-allo-Ile/mol of Ile is produced or 0.1 mol of L-allo-Ile/4.5 mol of Ile for comparison with crambin. Thus, Van Etten's hypothesis about the origin of L-allo-Ile is supported but not proven by our experiments.

Isolation of Peptides. Digestion of reduced and carboxymethylated crambin with trypsin produced three fragments (A, B, and C) which were separated on DEAE-cellulose (Figure 1). Alternatively, oxidized crambin (1 μmol) was digested, and the fragments were separated by HPLC (~40

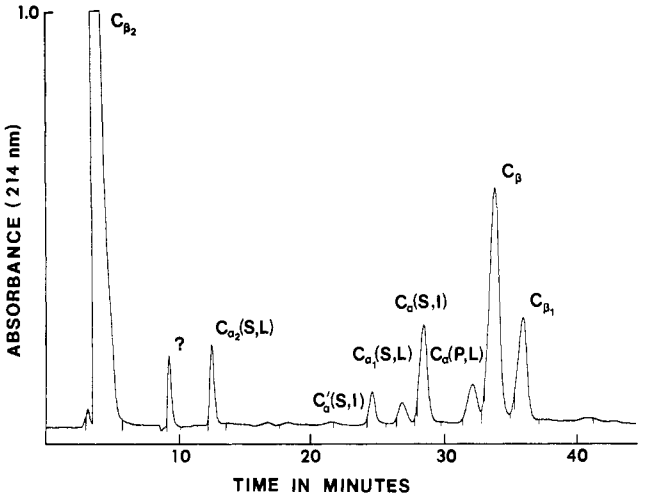

FIGURE 2: Separation of the chymotryptic fragments of the tryptic C peptide on reverse-phase HPLC. The single-letter codes S, L, I, and P represent the amino acids Ser, Leu, Ile, and Pro, respectively. S and P are at residue 22, and I and L are at residue 25.

Table III: Amino Acid Composition of the DEAE-cellulose-Separated Chymotryptic Peptides of the Tryptic C Peptide

amino acid	chymotryptic C_α	chymotryptic C_β
CM-cysteine	*a* (1)[b]	2.0 (2)
asparatic acid		1.3 (2)
threonine	1.7 (2)	1.6 (2)
serine	*c* (0–1)[d]	
glutamic acid	1.0 (1)	
proline	1.5 (1–2)[d]	2.0 (2)
glycine	1.4 (1)	3.0 (3)
alanine	2.0 (2)	2.1 (2)
valine		
isoleucine	*c* (0–1)[d]	2.5 (3)
leucine	1.6 (1–2)[d]	
tyrosine	0.4 (1)	0.8 (1)
phenylalanine		
arginine		
total no. of residues	12	17

[a] Not determined. [b] Values in parentheses refer to the number of residues determined by sequence analysis. [c] Below detection limits. [d] Microheterogeneity occurs here.

nmol). The amino acid compositions of the purified oxidized peptides are presented in Table II. Yields of the oxidized fragments were 70–80%. The largest fragment, C, was further digested with chymotrypsin, and the resulting peptides were separated either on DEAE-cellulose or by reverse-phase HPLC (Figure 2). Despite the presence of two Tyr residues in the C peptide, only two fragments (C_α and C_β) were obtained on DEAE-cellulose. The amino acid compositions of these are given in Table III. Note that the analysis for C_α is quite poor. The HPLC separation revealed microheterogeneity in the C_α peptide as described below.

Sequencing. The phenylthiohydantoins resulting from solid-phase Edman degradation were identified by TLC (for tryptic peptides A, B, and C and for intact crambin) or HPLC (for tryptic peptides B and C, intact crambin, and chymotryptic fragment C_β). The location of [^{14}C]carboxymethyl-Cys-PTH was confirmed by scintillation counting. PTH-Arg was identified by back-hydrolysis with HI. The order of the tryptic peptides was established by sequencing intact crambin through the first 27 residues. The sequence data are summarized in Figure 3.

The C-terminal sequence of the C tryptic peptide was determined by carboxypeptidase digestion to be Tyr-Ala$_{45}$-Asn (Figure 4). The digestion stopped after Tyr, providing further evidence that the next residue was Asp since carboxypeptidase A cleaves Asp residues slowly. Final crystallographic refinement at 1.5 Å revealed geometry and hydrogen-bonding patterns consistent with the assignments of Asp to residue 43 and Asn to residue 46.

Microheterogeneity. Although DEAE-cellulose chromatography of the chymotryptic digest of C showed only two fragments, fractionation by reverse-phase HPLC revealed a much more complex picture (Figure 2). An analysis of amino

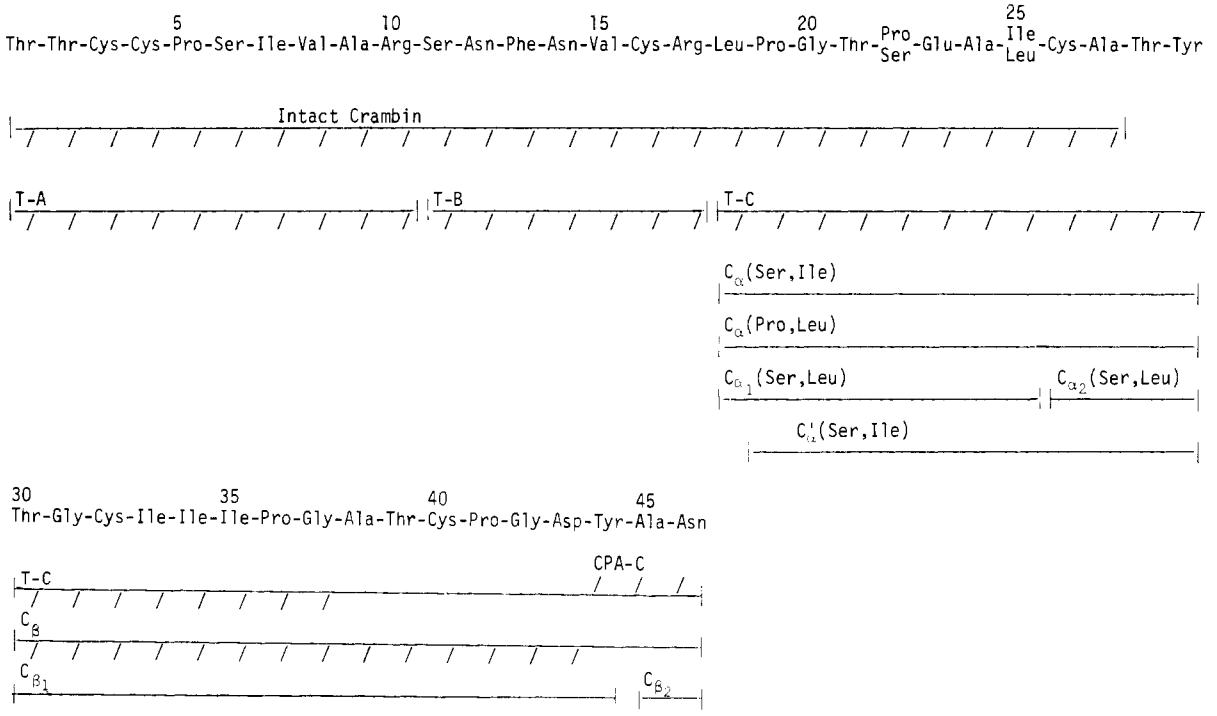

FIGURE 3: Summary of the sequence of crambin. T-A, T-B, and T-C are the tryptic peptides; C_α and C_β are the chymotryptic peptides of the tryptic C peptide; and CPA-C is the carboxypeptidase A digest of the C tryptic peptide. For C_α, the residues in parentheses indicate the observed heterogeneity at sites 22 and 25, respectively.

Table IV: Amino Acid Composition of the Reverse-Phase HPLC Separated Chymotryptic Peptides of the Tryptic C Peptide

amino acid	C_α(S,I)[a]	C_α'(S,I)	C_α(P,L)	C_{α_1}(S,L)	C_{α_2}(S,L)	C_β	C_{β_1}	C_{β_2}
CM-cysteine	0.3 (1)[b]	0.4 (1)	*c* (1)	0.5 (1)	~0.4[d] (1)	2.0 (2)	0.9[d] (2)	
aspartic acid						1.3 (2)	1.3 (1)	1.1 (1)
threonine	2.1 (2)	1.5 (2)	1.8 (2)	1.1 (1)	1.2 (1)	1.6 (2)	1.7 (2)	
serine	1.0 (1)	1.0 (1)		0.6 (1)				
glutamic acid	1.0 (1)	1.0 (1)	1.0 (1)	0.8 (1)				
proline	0.8 (1)	*e* (1)	1.5 (2)	0.5 (1)		2.0 (2)	2.0 (2)	
glycine	1.1 (1)	1.3 (1)	1.4 (1)	0.9 (1)		3.0 (3)	3.3 (3)	
alanine	2.1 (2)	1.9 (2)	2.1 (2)	0.8 (1)	1.1 (1)	2.1 (2)	0.7 (1)	1.0 (1)
valine								
isoleucine	0.8 (1)	0.9 (1)				2.5 (3)	2.6 (3)	
leucine	1.0 (1)		1.6 (2)	1.0 (1)				
tyrosine	0.7 (1)	0.7 (1)	0.4 (1)	0.4	1.0 (1)	0.8 (1)	0.7 (1)	
phenylalanine								
arginine								
total no. of residues	12	11	12	8	4	17	15	2

[a] Single-letter designations are S (Ser), P (Pro), L (Leu), and I (Ile). Ser and Pro are at residue 22, and Ile and Leu are at residue 25. [b] Values in parentheses refer to the number of residues determined from sequence analysis. [c] Not determined. [d] Oxidized forms of CM-Cys were observed here. [e] Estimated visually to be ~1.

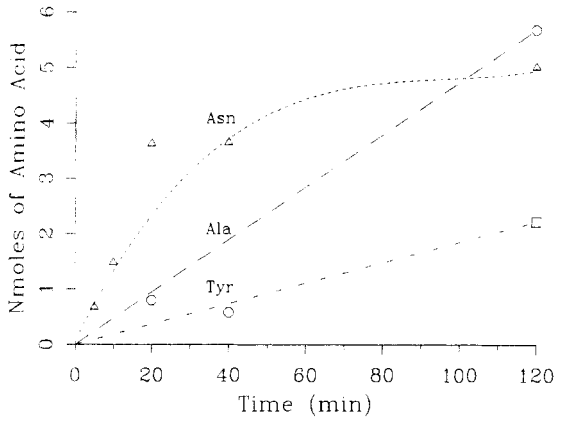

FIGURE 4: Carboxypeptidase A digestion of the C peptide. The digestion released first Asn (···), then Ala (- · -), and finally Tyr (- - -). Although this experiment does not permit an accurate estimate of the rates of release of the amino acids (i.e., the slopes are not well determined), it does qualitatively establish the C-terminal sequence.

acid compositions of these fragments (Table IV) showed that many of the peptides had similar compositions but differed in the relative amounts of Pro, Ser, Leu, and Ile. This led us to conclude that there are two sites of heterogeneity, Pro/Ser-22 and Ile/Leu-25. The major chymotryptic cleavage is a Tyr-29 which gives rise to C_α and C_β. A minor cleavage at Tyr-44 gives C_{β_1} and C_{β_2}. Two 12-residue peptides were isolated which correspond to C_α with Ser-22 and Ile-25 [C_α(Ser,Ile)] and to C_α with Pro-22 and Leu-25 [C_α(Pro,Leu)]. Chymotryptic cleavage at Leu-25 of C_α(Ser,Leu) produced C_{α_1}(Ser,Leu) and C_{α_2}(Ser,Leu). C_α'(Ser,Ile), which occurs in low yield, appears to be derived from the most abundant peptide, C_α(Ser,Ile), by cleavage of the Leu-18–Pro-19 bond. The fourth possible variant (Pro-22, Ile-25) was not detected here but was observed to predominate in the X-ray investigation (Hendrickson & Teeter, 1981). The estimated crystallographic ratios of Pro:Ser and Ile:Leu were 60:40 for both sites. However, solubility differences among these forms might have led to selection of different variants in the material analyzed in the crystallographic and sequence studies. Thus, although the naturally occurring ratios of heterogeneous species have not been determined by our results, the presence of microheterogeneity is firmly established.

Independent evidence for this heterogeneity came from low PTH yields at residues 22 and 25 in the sequence determination of C. HPLC analysis of the PTH for residue 22 showed two peaks. The major one was PTH-Pro (4.8 nmol), and a minor one could have been PTH-Ala or PTH-Δ-Ser (3.6 nmol). Residue 25 appeared to be Ile with a smaller amount of Leu (8.0 and 3.6 nmol, respectively).

DEAE-cellulose separation, which is based on anionic charge difference, could not distinguish the various forms of C_α, but reverse-phase HPLC, which is based on retention increasing with hydrophobicity, could. It is possible to predict the order of elution of the peptides from reverse-phase HPLC by applying the empirical hydrophobic constants of Meek (1980) or of Rekker (1977). The relative hydrophobicity accurately predicts the order of elution for most of the peptides shown in Figure 2 except for C'(Ser,Ile) and C_{α_1}(Ser,Leu) which are inverted (Meek) or very close (Rekker). Constants are not available for CM-Cys for cysteic acid.

Disulfide Bonds. The disulfide linkages were determined by analysis of the X-ray structure (Hendrickson & Teeter, 1981) to be Cys-3–Cys-40, Cys-4–Cys-32, and Cys-16–Cys-26. It is interesting to note that the same disulfide linkages have been demonstrated chemically for purothionin (Hase et al., 1978) and are proposed for viscotoxin (Samuelsson & Pettersson, 1971a).

Discussion

Crambin's hydrophobic character, i.e., its insolubility in water, can be better understood by examining the distribution of hydrophobic and hydrophilic residues over the surface of crambin. The backbone of crambin, as determined by its crystal structure (Hendrickson & Teeter, 1981), assumes the shape of the Greek letter Γ (Figure 5). The vertical stem of the Γ is composed of two antiparallel α helices, while the horizontal arm has a β sheet and a β turn. Disulfide bonds hold subdomains within these arms rigidly in place. As viewed in Figure 5, the Ile_{33}-Ile_{34}-Ile_{35}-Pro_{36} sequence falls at the bottom of the β sheet, while the more polar residues Thr_1-Thr_2-Cys_3 lie at the top. Hydrophobic residues (Ile-7, Val-8, Val-15, Leu-18, Ile/Leu-25, and Tyr-29) line the left side of the α helices, but three of the four charged side groups (Arg-10, Arg-17, and Glu-23) as well as the carboxy and amino termini cluster along the right side. Thus, the hydrophobic residues are found primarily along the left side of the vertical arm and on the underside of the molecule, and most hydrophilic residues line the inner bend of the Γ. This separation of hydrophobic and hydrophilic residues in the tertiary structure of a protein has been called amphiphilic character (Fukushima et al., 1979). If self-association occurs between hydrophobic regions in crambin, nonpolar solvents would be

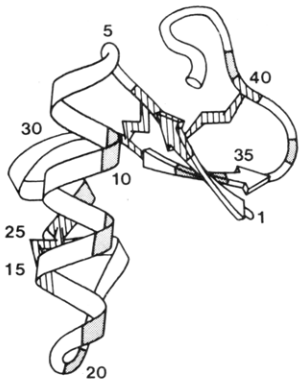

FIGURE 5: Schematic drawing of the backbone of crambin. This representation of crambin was drawn by Jane S. Richardson to be faithful to computer drawings of the skeleton. Arrows depict β strands. The disulfides are drawn as "lightning flashes". The crambin residues which are homologous with the plant toxins viscotoxin and purothionin are indicated on the backbone. Cys residues are striped, and all other homologous residues are shaded.

needed to dissolve it. Such an intermolecular, hydrophobic interaction is found in crambin crystals. The lower left edge of the α helices (Figure 5) makes van der Waals contact with a similar region in an adjacent, symmetry-related molecule, and water is completely excluded. The fact that such associations exist suggests why water alone cannot solubilize crambin.

Crambin is nearly 45% α helical. The helices are amphiphilic, similar to those predicted for several membrane-associated proteins. In the apoprotein apo-C-III (Segrest et al., 1974), the protein may be attached to the surface of the lipid bilayer by means of the apolar face of the predicted amphiphilic helices. In the case of the transmembrane protein bacteriorhodopsin, one model (Engelman et al., 1980) suggests the hydrophobic side chains of the α helices are on the exterior of the protein in contact with the phospholipid alkyl chains. Polar side chains on the other side of the helices form intramolecular associations among the helices. This interaction protects the polar residues from the hydrophobic lipid environment and holds the helices in place.

In light of the hydrophobic character of crambin, it is of interest to test the schemes devised from water-soluble protein structures for prediction of the secondary structure from the primary sequence (Chou & Fasman, 1974; Leavitt, 1978). A comparison between the predicted and observed secondary structure (Figure 6) shows an overall agreement of 70%. This is an average performance for these methods (Chou & Fasman, 1977). However, it is significant that the prediction is poorest for the long, amphiphilic top α helix (residues 7–18). This region, as noted above, may be most characteristic of many membrane-bound proteins. Thus, these models are only partially successful in their predictive ability for this hydrophobic protein.

The two sites of microheterogeneity in crambin are four residues apart in the linear sequence but adjacent in the three-dimensional structure. Pro/Ser-22 and Ile/Leu-25 are located in the lower left edge of the lower α helix (Figure 5). In the crystal structure, this section of the helix participates in strong intermolecular hydrophobic interactions (see above). Residue 25 lies in the center of this hydrophobic region, and either Ile or Leu would preserve the hydrophobic interaction. Residue 22 is located at the edge of this water-excluding region, where it could hydrogen bond to water. One of the heterogeneous residues here (Ser) would permit such bonding.

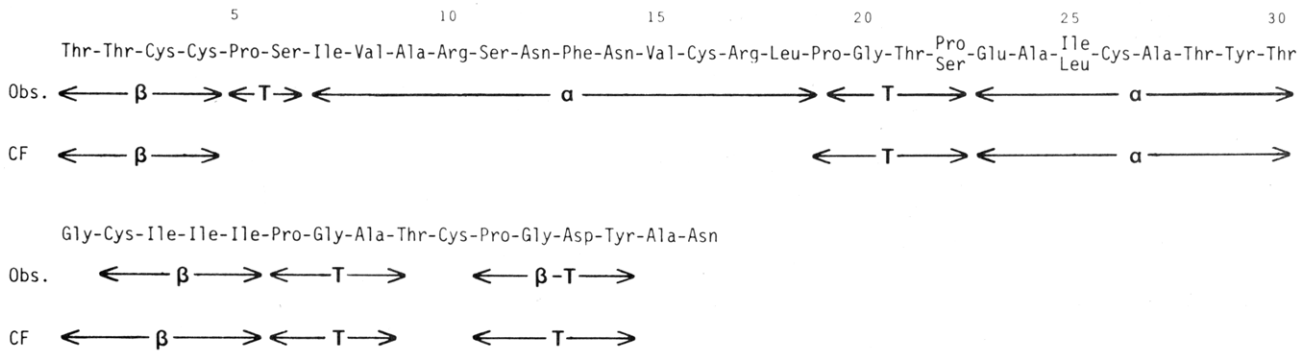

FIGURE 6: Comparison of the secondary structure prediction with the actual crystallographic secondary structure. α indicates the α helix, β is the β sheet, T is a turn region, and β-T is a β turn. The method of Chou & Fasman (1974), designated CF, was used for this comparison. The frequencies of Leavitt (1978) were in 90% agreement with those of Chou and Fasman and so are not shown here. The observed structure is from the crystallographic analysis (Hendrickson & Teeter, 1981).

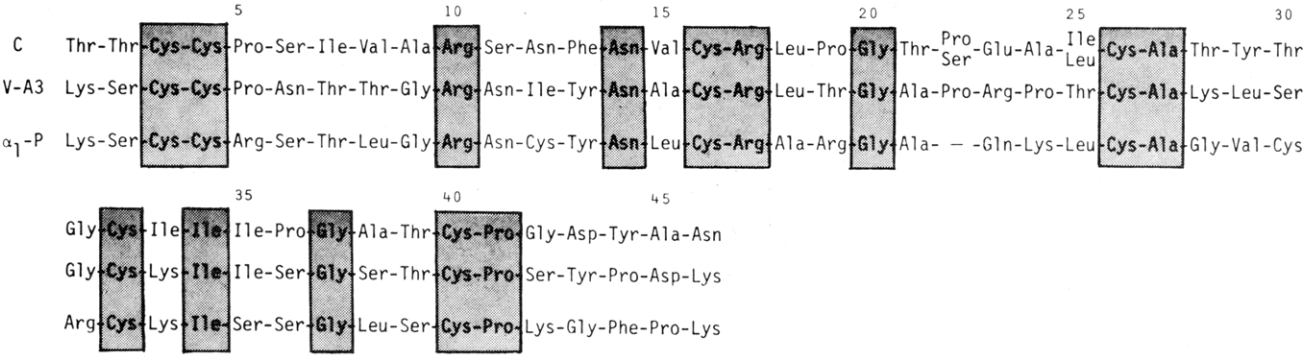

FIGURE 7: Comparison of the amino acid sequences of crambin and the plant toxins. C is the sequence of crambin (this work), V-A3 is viscotoxin A3 (Samuelsson et al., 1968), and α_1-P is α_1-purothionin (Jones & Mak, 1977; Ohtani et al., 1977). A total of 30% of the residues (shaded) are identical in all three proteins, and 48% are homologous using a mutation criterion (Dayhoff, 1978). A total of 43% are identical in crambin and viscotoxin A3 (50% are homologous); 33% are identical in crambin and α_1-purothionin (46% are homologous). A total of 55% are identical in viscotoxin A3 and α_1-purothionin (56% are homologous).

Finally, both sites of microheterogeneity represent a minimum of one base mutation in the nucleic acid which codes for this protein.

A search of the available protein sequence data (Dayhoff, 1978) revealed that crambin has a strong homology with the plant toxins purothionin and viscotoxin (Figure 7). All are 45 or 46 residues long and exhibit conservation of the three disulfide linkages (Mak & Jones, 1976; Hase et al., 1978; Samuelsson et al., 1968) which form the rigid core of the crambin molecule. A fourth disulfide bond found in purothionin (Cys-12–Cys-30) could be made in crambin with the appropriate amino acid substitutions. The significance of the homologous regions can be better understood when these regions are placed on the backbone of crambin (Figure 5). Many conserved residues fall in the inner bend of the Γ-shaped crambin molecule. This is the most hydrophilic region in crambin and contains many important intramolecular interactions which hold the two arms of the molecule together. For example, Arg-10 forms a salt bridge with the carboxy-terminal residue 46, hydrogen bonds to the carbonyl and hydroxyl of residue 2, and interacts with the amide oxygen of Asn-14 through a tightly bound water molecule. This complex association around Arg-10 could exist in all of the proteins. The hydrophobic residues are generally conserved functionally; i.e., Tyr-29 is replaced by Leu or Val and Ile/Leu-25 by Thr or Leu. However, the specific hydrophobic residues are less strongly retained than the specific hydrophilic amino acids at the inner bend. It may be that the hydrogen-bonding ability and/or the polarity of the inner bend must be conserved in this family of proteins.

The principal differences between crambin and the plant toxins lie in the number and type of charged residues in each. Crambin has only four charged residues while the plant toxins have eight to ten charged amino acids at neutral pH. Crambin has a p*I* of ~7, but purothionin and viscotoxin are very basic proteins. The low number of charged residues contributes to the insolubility of crambin in water and the high number to the solubility of the plant toxins.

Preliminary experiments (M.M. Teeter and J. R. Petithory, unpublished experiments) have shown that crambin is not toxic to 3T3 cells in tissue culture, whereas purothionin is toxic at the microgram per milliliter level. However, crambin's insolubility may prevent it from reaching the site of toxicity. It is also conceivable that the high positive charge, which is absent in crambin, plays a role in the toxicity of purothionin and viscotoxin.

The plant toxins as a group show considerable sequence heterogeneity. The three forms of purothionin (α_1, α_2, and β) each differ from the others at five to six sites (Jones & Mak, 1977). In viscotoxin, two pairs of the three forms (A3/B and A3/A2) differ at eight and nine residues, respectively (Samuelsson et al., 1968; Samuelsson & Pettersson, 1971b; Olson & Samuelsson, 1972). A2 and B differ at only two sites. Thus, except for the A2 and B forms of viscotoxin, considerably more heterogeneity is found in these plant toxins than in crambin. Most of the changes are conservative of the functional character of the amino acids. However, a few of the changes alter the acidity or basicity of a residue. These variations account for the fact that the toxicities of α_1- and β-purothionins are different toward different types of phytopathogenic bacteria (Fernández de Caleya et al., 1972). Such variation may provide an advantage to the plant in attacking invasive bacteria, if this is the biological function of these plant toxins, as some have suggested (Fernández de Caleya et al., 1972; Jones & Mak, 1977).

Although the function of crambin remains unknown, it may serve as an important model for membrane-bound proteins and for plant toxins. The primary and tertiary structures of crambin provide an example for the presence of amphiphilic helices in a hydrophobic protein. The strong sequence homology of crambin with plant toxins suggests they have similar three-dimensional structures, similar origins, and perhaps similar functions as well.

Acknowledgments

We thank Dr. Richard A. Laursen for advice and the use of his facilities, David Katz for technical assistance, and Dr. Susan Hochschwender and Dr. G. J. Quigley for helpful discussions.

References

Beckman Application Notes (1977a) 118/119CL-AN-001 (April) Beckman Instruments, Fullerton, CA.

Beckman Application Notes (1977b) 118/119CH-AN-002 (April) Beckman Instruments, Fullerton, CA.

Bhown, A. S.; Wissenger, A. K., Mole, J. E., & Bennett, J. C. (1978) *J. Chromatogr. 148*, 532–535.

Chou, P. Y., & Fasman, G. D. (1974) *Biochemistry 13*, 222–245.

Chou, P. Y., & Fasman, G. D. (1977) *Trends Biochem. Sci. (Pers. Ed.) 2*, 128–131.

Crestfield, A. M., Moore, S., & Stein, W. H. (1963) *J. Biol. Chem. 238*, 622–627.

Dayhoff, M. O. (1978) *Atlas of Protein Sequence and Structure 5*, Suppl. 3, National Biomedical Research Foundation, Washington, DC.

Edman, P., & Begg, G. (1967) *Eur. J. Biochem. 1*, 80–92.

Engelman, D. M., Henderson, R., McLachlan, A. D., & Wallace, B. A. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 2023–2027.

Fernández de Caleya, R., González-Pascual, B., García-Olmedo, F., & Carbonero, P. (1972) *Appl. Microbiol. 23*, 998–1000.

Fukushima, D., Kupferberg, J. P., Yokoyama, S., Droom, D. J., Kaiser, E. T., & Kezdy, F. J. (1979) *J. Am. Chem. Soc. 101*, 3703–3704.

Hase, T., Matsubara, H., & Yoshizumi, H. (1978) *J. Biochem. (Tokyo) 83*, 1671–1678.

Hendrickson, W. A., & Teeter, M. M. (1981) *Nature (London) 290*, 107–112.

Hirs, C. H. W. (1956) *J. Biol. Chem. 219*, 611–621.

Horn, M. J., & Bonner, A. G. (1976 in *Solid-Phase Methods in Protein Sequence Analysis* (Previero, A., Coletti-Previero, M. A., Eds.) pp 163–176, Elsevier/North-Holland, Amsterdam.

Jones, B. L., & Mak, A. S. (1977) *Cereal Chem. 54*, 511–523.

Laursen, R. A. (1966) *J. Am. Chem. Soc. 88*, 5344–5346.

Laursen, R. A. (1971) *Eur. J. Biochem. 20*, 89–102.

Leavitt, M. (1978) *Biochemistry 17*, 4277–4285.

Mak, A. S., & Jones, B. L. (1976) *Can. J. Biochem. 54*, 835–842.

Meek, J. L. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1632–1636.

Méndez, E., & Lai, C. Y. (1975) *Anal. Biochem. 68*, 47–53.

Ohtani, S., Okada, T., Yoshizumi, H., & Kogamiyama, H. (1977) *J. Biochem. (Tokyo) 82*, 753–767.

Olson, T., & Samuelsson, G. (1972) *Acta Chem. Scand. 26*, 585–595.

Ozols, J., Heinemann, F. S., & Gerard, C. (1980) in *Methods in Peptide and Protein Sequence Analysis* (Birr, C., Ed.)

pp 417–432, Elsevier/North-Holland Biomedical Press, New York.
Potts, J. T. (1967) *Methods Enzymol. 11*, 648–664.
Rekker, R. F. (1977) *The Hydrophobic Fragmental Constant*, p 301, Elsevier Scientific, New York.
Samuelsson, G., & Pettersson, B. M. (1971a) *Eur. J. Biochem. 21*, 86–89.
Samuelsson, G., & Pettersson, B. M. (1971b) *Acta Chem. Scand. 25*, 2048–2055.
Samuelsson, G., Seger, L., & Olson, T. (1968) *Acta Chem. Scand. 22*, 2624–2642.
Segrest, J. P., Jackson, R. L., Morrisett, J. D., & Gotto, A. M. (1974) *FEBS Lett. 38*, 247–253.
Teeter, M. M., & Hendrickson, W. A. (1979) *J. Mol. Biol. 127*, 219–223.
Vanderkooi, G., & Capaldi, R. A. (1972) *Ann. N.Y. Acad. Sci. 195*, 135–138.
Van Etten, C. H., Nielsen, H. C., & Peters, J. E. (1965) *Phytochemistry 4*, 467–473.
Wittmann-Liebold, B., & Lehmann, A. (1975) in *Solid-Phase Methods in Protein Sequence Analysis* (Laursen, R. A., Ed.) pp 81–90, Pierce Chemical Co., Rockford, IL.

Minor Collagens of Chicken Hyaline Cartilage†

Charles A. Reese and Richard Mayne*

ABSTRACT: Analysis has been made of the minor collagens which are solubilized by limited pepsin digestion of chicken sterna and which remain in solution after precipitation of type II collagen at 0.9 M NaCl–0.5 M HOAc. The precipitate obtained by further dialysis to 1.2 M NaCl–0.5 M HOAc was shown to contain the 1α, 2α, and 3α chains previously isolated from human cartilages [Burgeson, R. E., & Hollister, D. W. (1979) *Biochem. Biophys. Res. Commun. 87*, 1124–1131]. Mapping by polyacrylamide gel electrophoresis of the CNBr and *Staphylococcus aureus* V8 protease peptides derived from the 1α, 2α, and 3α chains strongly suggested that (i) 1α and 2α are different from the α1(V) and α2(V) chains of type V collagen and (ii) 3α is closely related to the α1(II) chain. An additional collagenous molecule of higher molecular weight (called HMW) was present in the 1.2 M NaCl precipitate, and considerably more HMW could be precipitated by dialysis to 2.0 M NaCl–0.5 M HOAc. HMW contained disulfide bonds and, after reduction, gave rise to three components, C-1, C-2, and C-3, of apparent M_r 87 500, 51 000, and 36 400, respectively. The precipitate obtained at 2.0 M NaCl–0.5 M HOAc also contained a lower molecular weight collagenous molecule (called LMW), which contained disulfide bonds and had an apparent molecular weight before reduction of 30 000. Cultured chick chondrocytes isolated from embryonic sterna were shown to synthesize 1α, 2α, and 3α by radioactive labeling and fluorography of polyacrylamide gels. The amino acid analyses and solubility properites of 1α, 2α, and HMW demonstrate that these components are closely related to the type V collagen isolated from other tissues, and it is suggested that 1α, 2α, and HMW should be regarded as members of a type V family of collagens.

Several studies have shown that the major collagen present in all hyaline cartilages is type II collagen of chain composition [α1(II)]$_3$ [reviewed by Miller (1976)]. However, several groups have recently reported the isolation of small amounts of other collagenous molecules from hyaline cartilages. Burgeson & Hollister (1979), working with human and bovine hyaline cartilages, isolated three collagenous chains which were designated 1α, 2α, and 3α. These authors proposed that 1α and 2α were previously undescribed collagen chains but were nevertheless closely related to the α1(V) and α2(V) chains of type V collagen.[1] The chain designated 3α appeared closely related to the α1(II) chain of type II collagen and may be an overglycosylated form of this chain. These results have been confirmed for neonatal pig hyaline cartilage (Shimokomaki et al., 1980) and for bovine nasal cartilage (Ayad et al., 1981), with both of these groups also reporting the isolation of an additional collagenous component of apparent M_r 110 000, which contained disulfide bonds and was reducible to a single component of apparent M_r 33 000.

In this paper, we report our analyses of the minor collagens present in chicken hyaline cartilage. The 1α, 2α, and 3α chains have been isolated from this tissue, and in addition, two disulfide-bonded collagenous molecules have been obtained which are similar to, but nevertheless are different from, the disulfide-bonded molecule previously isolated from mammalian cartilages (Shimokomaki et al., 1980; Ayad et al., 1981).

Materials and Methods

Isolation and Fractionation of Collagens. Sterna (200 g) from adult chickens were stripped of perichondrium, diced into cold, distilled water, and homogenized by using a Polytron homogenizer (Brinkman Instruments). After centrifugation, the sterna were extracted with 4 M guanidine, 50 mM Tris-HCl,[2] pH 7.4 (two extractions, 12 h each, 4 °C), followed by extensive washing of the sterna with cold, distilled water. The sterna were resuspended in 0.5 M acetic acid and 0.2 M NaCl containing pepsin (1 mg/mL, Worthington) as described previously (Burgeson & Hollister, 1979) and extracted with gentle stirring (4 °C, 20 h). After centrifugation (30000*g*, 30 min), the supernatant was brought to pH 8.0 by addition of 5 M NaOH. After standing overnight, the solution was

† From the Departments of Anatomy and Biochemistry, University of Alabama Medical Center, Birmingham, Alabama 35294. *Received April 13, 1981.* Supported in part by a Faculty Research Grant to R.M.
* Address correspondence to this author. R.M. is an Established Investigator of the American Heart Association.

[1] The nomenclature used for the type V collagen chains is that proposed by Bornstein & Sage (1980), so that the previously described αB and αA chains (Burgeson et al., 1976) are now designated the α1(V) and α2(V) chains.

[2] Abbreviations used: CNBr, cyanogen bromide; CM, carboxymethyl; $NaDodSO_4$, sodium dodecyl sulfate; Tris, tris(hydroxymethyl)aminomethane.

0006-2960/81/0420-5443$01.25/0