

The Amino-Acid Sequence and Carbohydrate Content of Phospholipase A₂ from Bee Venom

Rudolf A. SHIPOLINI, George L. CALLEWAERT, Richard C. COTTRELL, and Charles A. VERNON

The Christopher Ingold Laboratories, Department of Chemistry, University College London

(Received June 12, 1974)

The complete amino acid sequence of phospholipase A₂ from the venom of the common European honey-bee (*Apis mellifica*) has been determined. The sequence of amino acid residues at the N-terminus was obtained by direct application of the Edman degradation technique and that at the C-terminus by digestion with carboxypeptidases A and B. Digestion of the reduced and carboxymethylated enzyme with trypsin yielded a completely soluble peptide mixture, all the components of which were isolated and their structures determined. Overlaps of the tryptic peptides were deduced from the structures of peptides isolated after digestion of the reduced and carboxymethylated enzyme with chymotrypsin and of the reduced and aminoethylated enzyme with a protease specific for cleavage at lysine residues. Two remaining ambiguities were resolved by peptides obtained from a digest of the reduced, carboxymethylated and maleylated enzyme with trypsin. The phospholipase A₂ consists of a single chain of 128 amino acid residues and contains attached carbohydrate residues. The composition and point of attachment of the carbohydrate moiety have been established.

Phospholipase A₂ is the name given to a family of enzymes which catalyse the hydrolysis of fatty acid ester bonds at the 2-carbon position of 3-*sn*-phosphoglycerides. Enzymes with this specificity have been isolated from pancreatic tissue [1], bee venom [2] and a variety of snake venoms [3]. The enzymes have certain characteristics in common, for example relatively small size, high disulphide-bridge content, heat stability and requirement for Ca²⁺ ions for activity. There are, however, important differences. In particular, the pancreatic enzyme appears to be active on the aggregated (micellar) form of phospholipids [4], whereas the substrate for the bee-venom enzyme has been shown to be the dispersed (monomolecular) phospholipid [2]. It is also of interest that the enzyme from pancreas is secreted as an inactive zymogen which may be activated by limited proteolysis with trypsin [5]. The complete amino acid

sequence of the zymogen and the positions of the disulphide bridges in porcine pancreatic phospholipase A₂ have been reported [6, 7].

Our interest in phospholipase A₂ from bee venom arose from the observation [2] that the enzyme crystallises readily which, combined with its relatively small size, suggests that it may be amenable to total structure determination. As a preliminary to crystallographic analysis we have determined the primary structure of the enzyme; this work has been the subject of a preliminary report [8]. A missing overlap in the earlier work has now been established with certainty and three corrections have been made to the previously published sequence. In addition, the composition and point of attachment of the carbohydrate moiety have been established. An interesting aspects of the present work has been the use of a recently characterized protease isolated from the basidiomycete *Armillaria mellea* [9]; the enzyme has been shown to have specificity for cleavage of peptide bonds N-terminal to lysine residues [10] and the results presented here confirm and extend the previous observations. The positions of the four disulphide bridges of phospholipase A₂

Abbreviation. Dansyl, 5-dimethylaminonaphthalene-1-sulphonyl.

Enzymes. Phospholipase A₂ (EC 3.1.1.4); trypsin (EC 3.4.21.4); α -chymotrypsin (EC 3.4.21.1) carboxypeptidase A (EC 3.4.12.2); carboxypeptidase B (EC 3.4.12.3); *A. mellea* protease (EC 3.4.21.-).

have been determined and details are presented in the accompanying paper.

MATERIALS AND METHODS

Materials

Phospholipase A₂ from bee venom was purified as previously described [2]. Trypsin (treated with diphenylcarbamyl chloride) and carboxypeptidases A and B (treated with diisopropylphosphofluoridate) were obtained from Sigma (London) Chemical Co. (London SW6, U. K.). α -Chymotrypsin (BDH Chemicals Ltd, Poole, Dorset, U. K.) was purified from residual tryptic activity by ion-exchange chromatography [11]. *A. mellea* protease was a generous gift from the Pharmaceuticals Division, ICI Ltd, U. K. Phenylisothiocyanate (BDH) was purified by distillation at reduced pressure, and stored under nitrogen at -20°C . Pyridine (BDH) and *N*-ethylmorpholine (Eastman-Kodak Co., Rochester, N.Y., U.S.A.) were redistilled at atmospheric pressure through columns of glass. Iodoacetic acid (BDH) was recrystallised twice from toluene and maleic anhydride (BDH) was purified by sublimation. Dimethylallylamine (Pierce Chemical Co., Rockford, Ill., U.S.A.) was used without further purification. All other reagents were of the highest purity available.

Preparation of Derivatives

a) Reduction of Disulphide Bonds. Phospholipase A₂ (160 mg) was dissolved in Tris-HCl buffer (0.2 M, pH 8.0, 16 ml) containing urea (8 M) and EDTA (0.02 M). A 10 M excess of dithioerythritol was added and the solution stored at room temperature under nitrogen for 4 h.

b) Carboxymethylation. To a solution of the reduced enzyme prepared as above was added a small excess (calculated on the basis of the amount of dithioerythritol used) of iodoacetic acid dissolved in NaOH (0.2 N). The mixture was stored under nitrogen in the dark at room temperature for 30 min. The reduced carboxymethylated enzyme was isolated by passage through a column of Sephadex-G50 (superfine) equilibrated in acetic acid (5% v/v). The protein was concentrated by rotary evaporation and lyophilised. Quantitative amino acid analysis showed that 95% of the cystine content had been converted to the carboxymethyl derivative.

c) Aminoethylation [12]. To a solution of the reduced enzyme were added four equal portions of ethylene-imine at 10-min intervals. The final amount added was a 50-fold excess over the dithioerythritol used in the reduction stage. Reduced 2-aminoethylated

enzyme was isolated as in (b) above, and amino acid analysis showed that modification of cystine residues was 98% complete.

d) Maleylation. The free amino groups of the reduced carboxymethylated enzyme were maleylated using the conditions described by Butler *et al.* [13]. Excess reagents were removed by gel filtration through Sephadex G-50 (fine) equilibrated in NH₄OH (0.1 M). The protein fraction was concentrated by rotary evaporation and dried by lyophilisation. Dansylation of the reduced carboxymethylated maleylated enzyme showed that modification of amino groups was extensive, but the extent of blocking was not quantitated.

Proteolytic Digests of Modified Phospholipase A₂

a) Tryptic Digest of Reduced Carboxymethylated Maleylated Enzyme. The modified enzyme (150 mg) was dissolved in *N*-ethylmorpholine-acetic acid buffer (0.2 M, pH 8.0, 50 ml) containing CaCl₂ (5 mM) and trypsin (2% w/w) was added. The solution was incubated for 4 h at 35°C and then lyophilised. Maleyl groups were removed by incubation of the digest in acetic acid (pH 3.5, 10 ml) for 6 h at 50°C , after which the peptide mixture was again lyophilised.

b) Tryptic Digest of Reduced Carboxymethylated Enzyme. The modified enzyme (100 mg) was dissolved in NH₄HCO₃ (0.1 M, 10 ml) containing CaCl₂ (5 mM) and trypsin (1% w/w) was added. After 4 h at 35°C the reaction was stopped by acidification with glacial acetic acid, and the digest lyophilised.

c) Chymotryptic Digest of Reduced Carboxymethylated Enzyme. The modified enzyme (100 mg) was dissolved in water (10 ml) and 1 N NaOH added to pH 8.0. Chymotrypsin (2% w/w) was added and the solution was incubated for 8 h at 37°C in the cell of a Radiometer pH-Stat (type TTT1). The pH was maintained by automatic addition of NaOH (0.2 M). After incubation, the pH was lowered to 3.5 by addition of 1 N HCl and the peptide mixture lyophilised.

*d) Digestion of Reduced 2-Aminoethylated Enzyme with *A. mellea* Protease.* The modified enzyme (100 mg) was dissolved in *N*-ethylmorpholine-acetic acid buffer (0.2 M, pH 8.0, 10 ml) and digested with *A. mellea* protease (1% w/w) for 20 h at 35°C . The digest was lyophilised.

e) Digestion of Reduced Carboxymethylated Enzyme with Carboxypeptidases A and B. The modified enzyme (15 mg) was dissolved in *N*-ethylmorpholine-acetic acid buffer (0.2 M, pH 8.5, 3 ml) and digested with carboxypeptidase A (2.5% w/w) at 37°C . Aliquots (0.4 ml) were taken after 0.5, 1.2 and 24 h, acidified with glacial acetic acid (0.1 ml) and the precipitated protein was removed by centrifugation. The samples

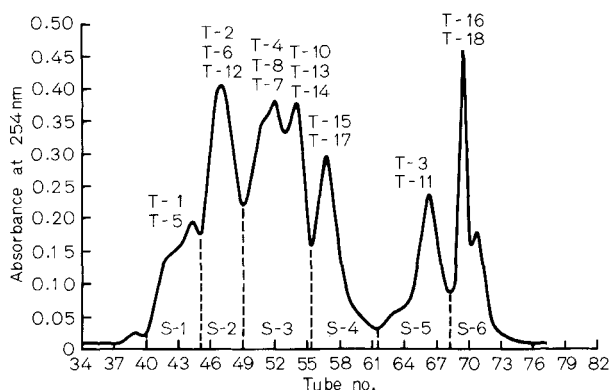


Fig. 1. Elution profile of the tryptic digest of reduced carboxymethylated enzyme after passage through Sephadex G-25. The column (1.5 × 150 cm) was equilibrated with acetic acid (5% v/v). Vertical bars indicate pooled fractions which were taken for further purification. Also shown are the identifying symbols for the peptides contained in each pooled fraction

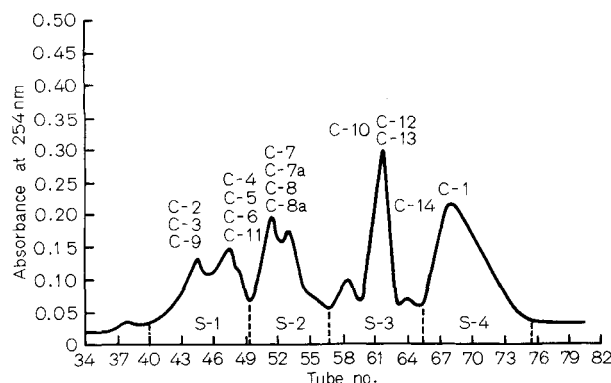


Fig. 2. Elution profile of the chymotryptic digest of reduced carboxymethylated enzyme after passage through Sephadex G-25. The conditions were as stated in Fig. 1

were dried by evaporation and finally subjected to quantitative amino acid analysis. Digestion with carboxypeptidase B was carried out similarly, aliquots being taken after 2, 4, 6 and 24 h.

Peptide Purification

In each case the initial purification procedure was gel filtration through Sephadex. For the tryptic digest of reduced carboxymethylated maleylated enzyme, a column (3.4 × 150 cm) of Sephadex G-50 (superfine) equilibrated in NH_4OH (0.1 M) was used, the sample being applied in 10 ml of the same solvent. In the other three cases columns (1.5 × 150 cm) of Sephadex G-25 (superfine) equilibrated in acetic acid (5% v/v) were used, the samples being applied in 3 ml of the same solvent. Elution profiles for the digests are shown in Fig. 1–3. Fractions were pooled as shown in Fig. 1–3 and further purified by high-voltage paper electrophoresis at pH 6.5 (pyridine–acetic acid–

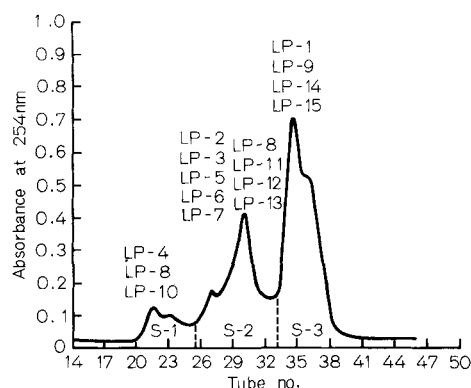


Fig. 3. Elution profile of the digest of reduced 2-aminoethylated enzyme with *A. mellea* protease after passage through Sephadex G-25. The conditions were as stated in Fig. 1

water, 25:1:224 by vol.) on Whatman no. 3MM paper. Voltages of 60 V/cm were maintained for 60–80 min. After electrophoresis, peptides were localised on guide strips by spraying with ninhydrin (0.2% w/v in acetone). Separated components were eluted from the paper (acetic acid, 5% v/v) and their purity assessed by thin-layer chromatography on cellulose plates using the solvent system *n*-butanol–acetic acid–pyridine–water (90:18:60:72 by vol. [14]). Samples requiring further purification were subjected to paper chromatography (Whatman 3MM) using the same solvent system.

In addition to spraying with ninhydrin, guide strips were also examined for the presence of peptides containing tyrosine and histidine using the Pauly reagent [15], for tryptophan-containing peptides using the Ehrlich reagent [16] and for arginine-containing peptides using the Sakaguchi reagent [17].

N-terminal Amino-Acid Analysis

Except in the case of peptides produced by digestion with *A. mellea* protease, N-terminal amino acid analysis by the dansylation technique was used as the final criterion of peptide purity [18]. Dansyl amino acids were identified by chromatography on polyamide thin-layer sheets (7.5 × 7.5 cm) [19]. Additional solvent systems were used to distinguish between the dansyl derivatives of hydroxy amino acids and of acidic amino acids [20] and between the monodansyl derivatives of histidine, lysine and arginine [21].

Quantitative and Qualitative Amino-Acid Analysis

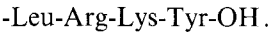
Hydrolysis of samples was carried out as described previously [2] and quantitative analyses were performed using a Technicon Autoanalyzer operated as a single-column system. For small peptides, qualitative

Table 1. *Amino acids liberated from carboxymethylated enzyme by treatment with carboxypeptidases A and B*
Carboxymethylated enzyme was treated with carboxypeptidases A and B for the times shown, after which aliquots originally containing approximately 130 nmol of protein were analysed for amino acid content

Enzyme	Time of hydrolysis	Amino acids liberated
	h	(nmol)
Carboxypeptidase A	0.5	Tyr (45) Lys (39)
	1	Tyr (120) Lys (110)
	24	unchanged
Carboxypeptidase B	2	Tyr (45) Lys (37)
		Arg (26) Leu (24)
	24	Tyr (125) Lys (122)
		Arg (90) Leu (90)

The C-Terminal Amino-Acid Sequence

Samples of carboxymethylated enzyme were digested with carboxypeptidase A and with carboxypeptidase B; the results are shown in Table 1. Treatment with carboxypeptidase A liberated only tyrosine and lysine, tyrosine being in larger quantity at short times of hydrolysis. Carboxypeptidase B also liberated tyrosine and lysine and in addition arginine and leucine in approximately equal amounts. The failure of carboxypeptidase A to liberate the antipenultimate residue strongly suggested that it was arginine rather than leucine and hence the following partial sequence was deduced for the C-terminal region:



Peptide	Sequence
T-1	Ile-Ile-Tyr-Pro-Gly-Thr-Leu-Trp-CmCys-Gly-His-Gly-Asx-Lys
T-2	Ser-Ser-Gly-Pro-Asn-Glu-Leu-Gly-Arg → →
T-3	Phe-Lys
T-4	His-Thr-Asp-Ala-CmCys-CmCys-Arg → → → → → → →
T-5	Thr-His-Asx-Met-CmCys-Pro-Asx-Val-Met-Ser-Ala-Gly-Glx-Ser-Lys
T-6	His-Gly-Leu-Thr-Asp-Thr-Ala-Ser-Arg →
T-7	Leu-Ser-CmCys-Asn-Asx-Asx-Asx-Phe-Tyr-Lys →
T-8	Asn-Ser-Ala-Asx-Thr-Ile-Ser-Ser-Tyr →
T-9	Phe-Val-Gly-Lys
T-10	Met-Tyr-Phe-Asn-Leu-Ile-Asn-Thr-Lys → → → → → → →
T-11	CmCys-Tyr-Lys
T-12	Leu-Glu-His-Pro-Val-Thr-Gly-CmCys-Gly-Glu-Arg → → → → → → → → →
T-13	Thr-Glu-Gly-Arg → → →
T-14	CmCys-Leu-His-Tyr-Thr-Val-Asp-Lys → → → → →
T-15	Ser-Lys-Pro-Lys
T-16	Val-Tyr-Gln-Trp-Phe → → → →
T-17	Asp-Leu-Arg →
T-18	Lys-Tyr

Fig. 4. *Amino-acid sequences of tryptic peptides from reduced carboxymethylated enzyme.* Residues not underlined were determined by the dansyl-Edman method. Residues marked → were determined by the direct Edman method. Amide

assignments were carried out either by isolation of the phenylthiohydantoin derivative (→) or by measurements of electrophoretic mobility at pH 6.5 (→) as described in the text

Table 2. Amino-acid compositions of the peptides from the tryptic digest of reduced carboxymethylated enzyme

Amino acid	T-1 ^c	T-2	T-3	T-4	T-5	T-6	T-7	T-8
Cm-cysteine ^a	0.87 (1)	—	—	1.83 (2)	0.87 (1)	—	0.74 (1)	—
Aspartic acid	0.84 (1)	0.93 (1)	—	0.91 (1)	1.83 (2)	0.87 (1)	3.57 (4)	1.82 (2)
Threonine	0.91 (1)	—	—	0.90 (1)	0.91 (1)	1.83 (2)	—	0.91 (1)
Serine	—	1.84 (2)	—	—	1.74 (2)	0.87 (1)	0.87 (1)	2.61 (3)
Glutamic acid	—	0.89 (1)	—	—	0.87 (1)	—	—	—
Proline	1.10 (1)	1.00 (1)	—	—	1.10 (1)	—	—	—
Glycine	2.91 (3)	1.97 (2)	—	—	1.01 (1)	1.01 (1)	—	—
Alanine	—	—	—	1.00 (1)	0.97 (1)	0.97 (1)	—	0.98 (1)
Valine	—	—	—	—	0.91 (1)	—	—	—
Methionine	—	—	—	—	1.71 (2)	—	—	—
Isoleucine	0.87 (1)	—	—	—	—	—	—	1.03 (1)
Leucine	1.00 (1)	1.03 (1)	—	—	—	1.02 (1)	1.40 (1)	—
Tyrosine	0.78 (1)	—	—	—	—	—	0.97 (1)	1.01 (1)
Phenylalanine	—	—	0.98 (1)	—	—	—	1.05 (1)	—
Lysine	0.97 (1)	—	1.13 (1)	—	1.13 (1)	—	1.13 (1)	—
Histidine	1.00 (1)	—	—	1.10 (1)	1.07 (1)	1.05 (1)	—	—
Arginine	—	0.97 (1)	—	0.94 (1)	—	0.94 (1)	—	—
Tryptophan ^b	+ (1)	—	—	—	—	—	—	—

^a The value given is the sum of carboxymethyl-cysteine and cysteic acid.^b Determined spectrophotometrically.

Table 3. Amino-acid compositions of the peptides from the chymotryptic digest of reduced carboxymethylated enzyme

Amino acid	C-1	C-2 ^c	C-3	C-4	C-5	C-6	C-7
Cm-cysteine ^a	—	0.67 (1)	2.60 (3)	—	0.65 (1)	—	—
Aspartic acid	—	2.37 (2)	2.90 (3)	—	3.47 (4)	1.90 (2)	—
Threonine	0.90 (1)	—	2.10 (2)	—	—	0.90 (1)	—
Serine	—	2.20 (2)	—	2.20 (2)	0.83 (1)	2.63 (3)	—
Glutamic acid	—	0.60 (1)	—	1.00 (1)	—	—	—
Proline	1.10 (1)	1.20 (1)	1.40 (1)	—	—	—	—
Glycine	1.00 (1)	3.90 (4)	—	1.90 (2)	—	—	1.20 (1)
Alanine	—	—	1.00 (1)	1.00 (1)	—	1.00 (1)	—
Valine	—	—	1.10 (1)	—	—	—	1.00 (1)
Methionine	—	—	1.50 (2)	—	—	—	0.70 (1)
Isoleucine	0.81 (1)	—	—	—	—	1.00 (1)	—
Leucine	1.00 (1)	0.70 (1)	—	1.00 (1)	—	—	—
Tyrosine	1.00 (1)	—	—	—	0.91 (1)	0.80 (1)	1.10 (1)
Phenylalanine	—	0.57 (1)	—	—	1.02 (1)	—	0.90 (1)
Lysine	—	1.36 (1)	1.20 (1)	1.20 (1)	—	1.40 (1)	1.10 (1)
Histidine	—	1.30 (1)	2.20 (2)	1.00 (1)	—	—	—
Arginine	—	0.70 (1)	1.20 (1)	—	—	—	—
Tryptophan ^b	+ (1)	—	—	—	—	—	—

^a The value given is the sum of carboxymethyl-cysteine and cysteic acid.^b Determined spectrophotometrically.^c The analysis of peptide C-2 also showed approximately 2 residues of glucosamine.

Cleavage of tyrosine and leucine by the carboxypeptidase B was presumably due to contamination with carboxypeptidase A.

Tryptic Peptides from Reduced Carboxymethylated Enzyme

Eighteen peptides were obtained from the tryptic digest of reduced carboxymethylated enzyme; their

amino acid compositions are shown in Table 2. In the case of peptide T-1, a peak was eluted on amino acid analysis immediately before valine. This was shown to be due to glucosamine and comparison with standards indicated a recovery of about 2 moles of glucosamine/mole of peptide. The recovery of isoleucine from peptide T-1 was very low (approximately one residue instead of two); this was due to the

T-9	T-10	T-11	T-12	T-13	T-14	T-15	T-16	T-17	T-18	Total
—	—	0.87 (1)	0.91 (1)	—	0.84 (1)	—	—	—	—	8
—	1.78 (2)	—	—	—	0.87 (1)	—	—	0.90 (1)	—	16
—	0.87 (1)	—	0.89 (1)	0.93 (1)	0.91 (1)	—	—	—	—	10
—	—	—	—	—	—	0.90 (1)	—	—	—	10
—	—	—	1.73 (2)	0.87 (1)	—	—	0.86 (1)	—	—	6
—	—	—	1.01 (1)	—	—	1.05 (1)	—	—	—	5
1.00 (1)	—	—	2.00 (2)	1.00 (1)	—	—	—	—	—	11
—	—	—	—	—	—	—	—	—	—	4
1.00 (1)	—	—	1.00 (1)	—	1.13 (1)	—	0.97 (1)	—	—	5
—	0.83 (1)	—	—	—	—	—	—	—	—	3
—	0.73 (1)	—	—	—	—	—	—	—	—	3 ^d
—	0.81 (1)	—	1.02 (1)	—	1.12 (1)	—	—	1.10 (1)	—	8
—	0.96 (1)	1.03 (1)	—	—	0.97 (1)	—	0.92 (1)	—	1.01 (1)	8
1.12 (1)	1.02 (1)	—	—	—	—	—	1.02 (1)	—	—	5
1.13 (1)	1.00 (1)	1.13 (1)	—	—	1.12 (1)	2.28 (2)	—	—	1.12 (1)	11
—	—	—	1.13 (1)	—	1.06 (1)	—	—	—	—	6
—	—	—	0.97 (1)	0.91 (1)	—	—	—	0.92 (1)	—	6
—	—	—	—	—	—	—	+	(1)	—	2

^c The analysis of peptide T-1 also showed approximately 2 residues of glucosamine.

^d The total content of isoleucine is 4 residues. The yield of isoleucine from peptide T-1 was low due to the occurrence of the sequence Ile-Ile.

C-7a	C-8	C-8a	C-9	C-10	C-11	C-12	C-13	C-14
—	0.71 (1)	0.65 (1)	1.53 (2)	—	—	—	—	—
—	1.90 (2)	1.72 (2)	—	—	0.81 (1)	—	1.00 (1)	—
—	0.90 (1)	0.87 (1)	2.10 (2)	—	0.90 (1)	—	—	—
—	—	—	—	—	0.80 (1)	—	—	—
—	—	—	2.60 (3)	—	—	1.00 (1)	—	—
—	—	—	1.50 (1)	—	1.30 (1)	—	—	—
1.10 (1)	—	—	2.80 (3)	—	—	—	—	—
—	—	—	—	—	—	—	—	—
1.00 (1)	—	—	1.00 (1)	—	2.00 (2)	—	—	—
0.67 (1)	—	—	—	—	—	—	—	—
—	0.70 (1)	0.60 (1)	—	—	—	—	—	—
—	0.80 (1)	0.70 (1)	1.90 (2)	—	—	—	1.10 (1)	—
0.90 (1)	0.90 (1)	0.80 (1)	—	1.00 (1)	0.90 (1)	—	0.80 (1)	—
—	0.90 (1)	—	—	—	—	—	—	1.82
1.20 (1)	1.20 (1)	1.30 (1)	1.00 (1)	—	1.60 (3)	—	1.10 (1)	—
—	—	—	1.00 (1)	1.00 (1)	—	—	—	—
—	—	—	1.90 (2)	—	—	—	1.10 (1)	—
—	—	—	—	—	—	+	(1)	—

occurrence of the isoleucine residues in the sequence Ile-Ile which is very resistant to acid hydrolysis.

All of the peptides were subjected to complete structure determination either by the dansyl-Edman technique or, in the case of peptides T-4, T-10, T-12, T-13, T-14 and T-16, by the direct Edman method. The amino acid sequences obtained are shown in Fig. 4, and the positions of the peptides in the total

sequence are shown in Fig. 7. It can be seen from Fig. 7 that the tryptic peptides contained all of the amino acid residues present in phospholipase A₂.

Chymotryptic Peptides from Reduced Carboxymethylated Enzyme

The amino acid compositions of peptides isolated from the chymotryptic digest of reduced carboxy-

Peptide	Sequence
C-1	Ile-Ile-Tyr-Pro-Gly-Thr-Leu-Trp
C-2	CmCys-Gly-His-Gly-Asx-Lys-Ser-Ser-Gly-Pro-Asx-Glx-Leu-Gly-Arg-Phe
C-3	Lys-His-Thr-Asx-Ala-CmCys-CmCys-Arg-Thr-His-Asp-Met-CmCys-Pro-Asn-Val-Met
C-4	Ser-Ala-Gly-Glu-Ser-Lys-His-Gly-Leu
C-5	Ser-CmCys-Asx-Asx-Asx-Asx-Phe-Tyr
C-6	Lys-Asn-Ser-Ala-Asp-Thr-Ile-Ser-Ser-Tyr
C-7	Phe-Val-Gly-Lys-Met-Tyr
C-7a	Val-Gly-Lys-Met-Tyr
C-8	Phe-Asx-Leu-Ile-Asx-Thr-Lys-CmCys-Tyr
C-8a	Asn-Leu-Ile-Asn-Thr-Lys-CmCys-Tyr
C-9	Lys-Leu-Glu-His-Pro-Val-Thr-Gly-CmCys-Gly-Glu-Arg-Thr-Glx-Gly-Arg-CmCys-Leu
C-10	His-Tyr
C-11	Thr-Val-Asp-Lys-Ser-Lys-Pro-Lys-Val-Tyr
C-12	Gln-Trp
C-13	Asp-Leu-Arg-Lys-Tyr
C-14	Free Phe

Fig. 5. Amino-acid sequences of chymotryptic peptides from reduced carboxymethylated enzyme. Residues not underlined were determined by the dansyl-Edman method. Residues

marked → were determined by the direct Edman method. Amide assignments were carried out by isolation of the phenylthiohydantoin derivative for those residues marked →

methyated enzyme are shown in Table 3 and the amino acid sequences are given in Fig. 5. The sequences are generally consistent with the amino acid compositions except in the case of peptide C-1, where the occurrence of the sequence Ile-Ile resulted in a low yield of isoleucine. In addition to the expected amino acids, peptide C-2 contained approximately two residues of glucosamine. Free phenylalanine was found in the digest to an extent of about 1.8 moles/mole of reduced carboxymethylated enzyme. This arose from complete hydrolysis of the bonds to residue Phe¹²³ and partial cleavage of the bonds C-terminal to Phe⁷⁶ and Phe⁸². Consistently, the chymotryptic peptides originating from the region of residues 76 to 90 were found both in forms with and without the N-terminal phenylalanine residues (peptides C-7 and C-7a, and peptides C-8 and C-8a).

The positions of the chymotryptic peptides in the structure of phospholipase A₂ are shown in Fig. 7. No peptide was found corresponding to residues 51–53.

Table 4. Amino-acid compositions of some of the peptides from the digest of reduced 2-aminoethylated enzyme with *A. mellea* protease

Amino acid	L-5	L-6	L-7
Aspartic acid	0.87 (1)	3.51 (4)	1.78 (2)
Threonine	1.78 (2)	—	0.87 (1)
Serine	1.67 (2)	—	2.47 (3)
Glutamic acid	—	—	—
Proline	—	—	—
Glycine	1.00 (1)	—	1.00 (1)
Alanine	0.96 (1)	—	0.96 (1)
Valine	—	—	1.00 (1)
Methionine	—	—	—
Isoleucine	—	—	1.00 (1)
Leucine	2.21 (2)	—	—
Tyrosine	—	0.87 (1)	0.93 (1)
Phenylalanine	—	0.96 (1)	0.97 (1)
Aminoethylcysteine ^a	—	0.78 (1)	—
Lysine	1.10 (1)	—	1.12 (1)
Histidine	1.10 (1)	—	—
Arginine	0.94 (1)	—	—

^a The value given is the sum of aminoethylcysteine and cysteic acid.

Peptide	Sequence
L-1	Ile-Ile-Tyr-Pro-Gly-Thr-Leu-Trp
L-2	AetCys-Gly-His-Gly-Asx-Lys-Ser-Ser-Gly-Pro-Asx-Glx-Leu-Gly-Arg-Phe
L-3	Lys-His-Thr-Asx-Ala-AetCys
L-4	AetCys-Arg-Thr-His-Asx-Met-AetCys-Pro-Asx-Val-Met-Ser-Ala-Gly-Glx-Ser
L-5	Lys-His-Gly-Leu-Thr-Asx-Thr-Ala-Ser-Arg-Leu-Ser
L-6	AetCys-Asn-Asn-Asn-Asp-Phe-Tyr → → → →
L-7	Lys-Asn-Ser-Ala-Asp-Thr-Ile-Ser-Ser-Tyr-Phe-Val-Gly → →
L-8	Lys-Met-Tyr-Phe-Asn-Leu-Ile-Asx-Thr →
L-9	Lys-AetCys-Tyr
L-10	Lys-Leu-Glx-His-Pro-Val-Thr-Gly
L-11	AetCys-Gly-Glx-Arg-Thr-Glu-Gly-Arg →
L-12	AetCys-Leu-His-Tyr-Thr-Val-Asp →
L-13	Lys-Ser-Lys-Pro
L-14	Lys-Val-Tyr-Gln-Trp-Phe-Asp-Leu-Arg → →
L-15	Lys-Tyr

Fig. 6. Amino-acid sequences of peptides from the digest of reduced 2-aminoethylated enzyme with *A. mellea* protease. Residues not underlined were determined by the dansyl-

Edman method. The states of amidation of residues marked → were determined by measurements of electrophoretic mobility at pH 6.5

Peptides Obtained by Digestion of Reduced 2-Aminoethylated Enzyme with *A. mellea* Protease

Fifteen major peptides were isolated from the digest of reduced 2-aminoethylated enzyme with *A. mellea* protease. Three of these (L-5, L-6 and L-7) were of particular importance and were subjected to quantitative amino acid analysis (Table 4); qualitative compositions by the method of complete dansylation of hydrolysates were determined for the remaining twelve peptides and were consistent with the amino acid sequences in all cases. The amino acid sequences of all the peptides were established by the dansyl-Edman method and are shown in Fig. 6. States of amidation of side-chains were assigned by studies of electrophoretic mobility at pH 6.5 as indicated in Fig. 6.

Peptides from the Digest of Reduced Carboxymethylated Maleylated Enzyme with Trypsin

The majority of the material from this digest was insoluble in the solvents used for peptide separation;

consequently only a small number of peptides were isolated from the digest. Of these, two were particularly useful in establishing overlaps and will be described here.

Peptide TM-1. The amino acid composition (with recoveries in brackets) was: Asp (1.1), Thr (1.1), Ala (1.1), CmCys + cysteic acid (1.7), Phe (1.0), Lys (1.0), His (1.1), Arg (1.0). The amino acid sequence, determined by the direct Edman method was:

Phe-Lys-His-Thr-Asp-Ala-CmCys-CmCys-Arg.

Peptide TM-2. The amino acid composition was: Asp (0.39), Thr (1.0), Glu (1.66), Pro (1.1), Gly (2.0), Val (1.0), CmCys + cysteic acid (1.7), Leu (1.0), Tyr (0.8), Lys (1.0), His (1.0), Arg (0.9). The amino acid sequence, determined by the direct Edman method was:

CmCys-Tyr-Lys-Leu-Glu-His-Pro-Val-Thr-Gly-CmCys-Gly-Glu-Arg.

It can be seen from Fig. 7 that peptide TM-2 arose from cleavage at Lys⁸⁸ indicating that maleylation of this residue either did not occur or was not complete.

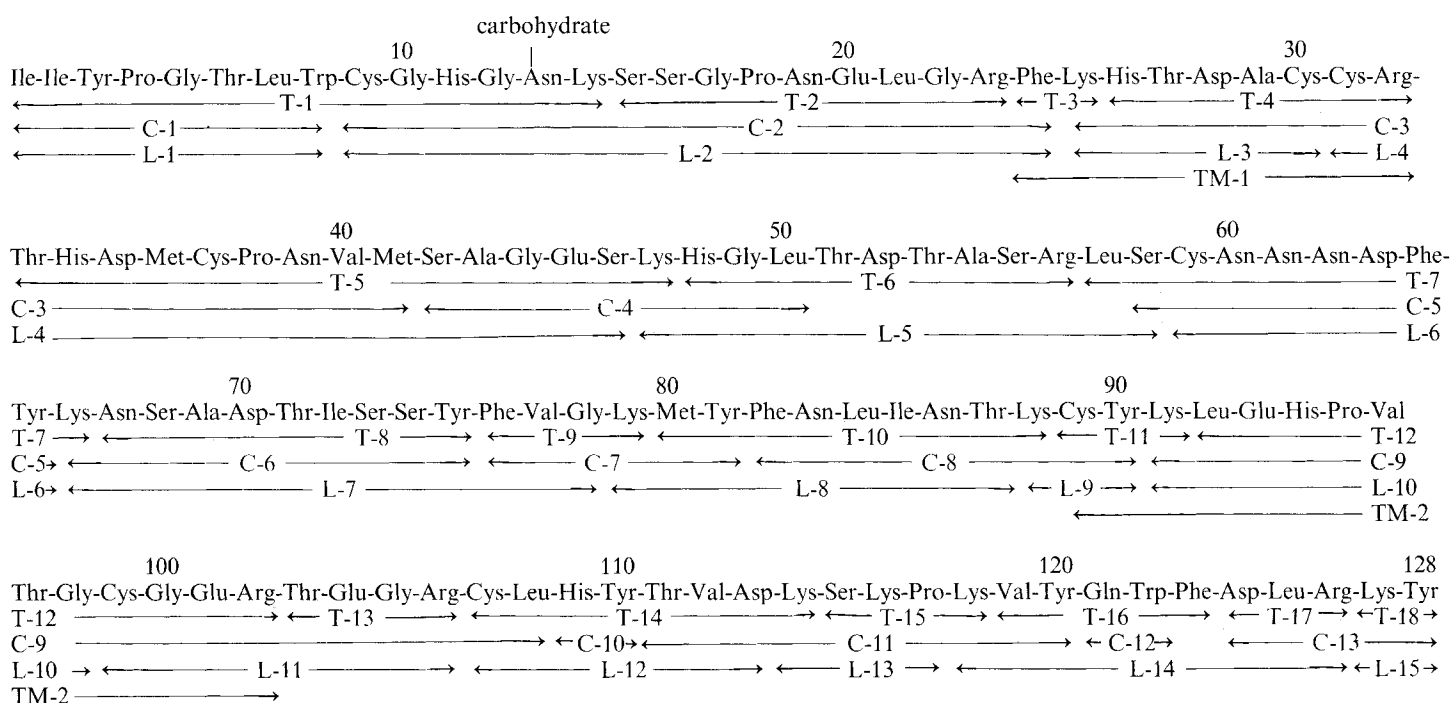


Fig. 7. The amino-acid sequence of phospholipase A₂ from bee venom. The symbols T and C refer to peptides isolated after cleavage of reduced carboxymethylated enzyme with trypsin and chymotrypsin respectively; L refers to peptides from the

digest of reduced 2-aminoethylated enzyme with *A. mellea* protease and TM to peptides from digestion of reduced carboxymethylated maleylated enzyme with trypsin

Table 5. Carbohydrate content of phospholipase A₂ and of peptide C-2

Carbohydrate residue	Phospholipase A ₂		Peptide C-2	
	recovery	molar ratios ^a	recovery	molar ratios
	nmol		nmol	
Fucose	0.49	1.0	0.070	1.0
Mannose	4.01	8.2	0.557	8.0
Galactose	0.43	0.9	0.034	0.5
N-Acetylglucosamine ^b	1.94	4.0	0.245	3.5

^a Based on fucose = 1.

^b The method of analysis does not distinguish between the presence of glucosamine or N-acetylglucosamine in the original sample.

Composition of the Carbohydrate Moiety

The carbohydrate contents of native phospholipase A₂ and of peptide C-2 were determined by the method of Bolton *et al.* [24] as described in Methods. In the case of the native enzyme the weight of protein taken was 2.9 mg (approximately 0.18 μmol) and after hydrolysis and conversion of the carbohydrate residues, aliquots containing 1/200 of the original material (*i.e.* 0.90 nmol) were taken for analysis. The results

obtained (Table 5) suggest that the native enzyme contained approximately half a residue each of fucose and galactose, 4 residues of mannose and 2 residues of N-acetylglucosamine. Peptide C-2 contained carbohydrate residues in similar molar ratios to those of the native enzyme. The value for galactose could not be obtained accurately since the peak was incompletely resolved from that of mannose. No attempt was made to weigh the sample of peptide C-2 taken for analysis because of the very small amount available.

DISCUSSION

The peptides isolated from the four proteolytic digests gave sufficient information to establish the structure of phospholipase A₂ unambiguously (Fig. 7). There were two main reasons why results from all four digests were required to establish the structure. First, treatment of the reduced carboxymethylated enzyme with trypsin resulted in cleavage at Tyr⁷⁵ and Phe¹²³ in addition to the normal cleavage at lysine and arginine. These non-specific cleavage, which still occurred when the time of tryptic digestion was drastically reduced, meant that overlaps between peptides T-8 and T-9 and between T-16 and T-17 could not be established by digestion with chymotrypsin. These overlaps were provided by peptides

L-7 and L-14 from the digestion with *A. mellea* protease. The second major problem was the occurrence in three positions of sequences of the type X-Lys where X is a point of chymotryptic cleavage (*i.e.* at positions 24–25, 65–66 and 90–91). The overlaps in these three regions were not, therefore, obtainable unambiguously from either the digest with chymotrypsin or with *A. mellea* protease. Treatment of the reduced carboxymethylated maleylated enzyme with trypsin yielded two peptides (TM-1 and TM-2), the first of which overlapped tryptic peptides T-3 and T-4, and the second overlapped T-11 and T-12; the overlap between peptides T-7 and T-8 is then established by elimination. The structure shown in Fig. 7 differs from that previously reported in three respects [8]. Two residues originally given as aspartic acid (61 and 67) are in fact asparagine, and a leucine residue previously positioned between residues 64 and 65 has been deleted. In the case of residue 67, the error was typographical. The other two arose as a result of difficulties associated with the sequence of four consecutive Asx residues in positions 60–64. The first approach to assignment of amide side-chains in this region was by isolation and identification of phenylthiohydantoin derivatives during degradation of the appropriate tryptic peptide (see Methods). The procedure used to convert 2-anilino-5-thiazolinones to phenylthiohydantoin derivatives invariably produces partial deamination of asparagine to aspartic acid and, since the recovery of phenylthiohydantoin asparagine at the position corresponding to residue 61 was low, this was interpreted as being due to carry-over of asparagine from the previous residue. A re-examination of this region of the structure by measurements of mobilities at pH 6.5 of peptide L-6 after each step of Edman degradation showed clearly that residue 61 is asparagine. The earlier inclusion of an extra leucine residue between residues 63 and 64 also arose from difficulties due to the repeated Asx sequence. Accumulation of Asx-terminal fragments consequent on incomplete degradation after seven cycles of the Edman reaction on peptide T-7 made interpretation of results at the C-terminal end of the peptide difficult. Peptide L-6 proved easier to sequence since the Asx residues occurred near the N-terminus; moreover, quantitative amino acid analysis (Table 4) showed that the peptide did not contain leucine. Hence we are confident that the revised sequence in this region is correct.

The specificity of action of *A. mellea* protease on protein substrates is not yet completely defined [10] and the results presented here provide further information on this point. Cleavage was observed N-terminal to all lysine residues except Lys¹⁴ and Lys¹¹⁶. Lys¹¹⁶ is followed in the sequence by a proline residue which

presumably blocked cleavage at this point; cleavage did occur, however, at Lys¹¹⁸ which is preceded by the proline residue. Failure to cleave at Lys¹⁴ may be attributed to the fact that carbohydrate moiety of phospholipase A₂ is attached to residue 13 (see later). Hydrolysis was not observed at any of the four arginine residues. Cleavage by the protease at aminoethylcysteine residues occurred readily in most cases. Again a C-terminal proline residue prevented hydrolysis (at Cys³⁷). At positions 88 and 89 where lysine and aminoethyl cysteine were adjacent, the main cleavage point was N-terminal to the lysine residue, although a small amount of the dipeptide AetCys-Tyr was also found in the digest.

Phospholipase A₂ from bee venom contains bound carbohydrate residues. The results in Table 5 showed that the carbohydrate moiety contains fucose, galactose, mannose and glucosamine (presumed to be *N*-acetylated) in the molar ratio 1:1:8:4. It appeared, however, that the actual content of fucose and galactose is only 0.5 residue each per polypeptide chain. This was consistent with the recovery of 2 moles of glucosamine on amino acid analysis of both peptides T-1 and C-2; this must, however, be considered the lower limit owing to possible destruction of glucosamine during acid hydrolysis. The low content of fucose and galactose indicates that the carbohydrate moiety is inhomogeneous. This is consistent with the observation that, although apparently homogeneous with respect to the polypeptide chain, phospholipase A₂ gives two approximately equal precipitin bands on immunodiffusion analysis (M. A. Yomtov, unpublished observations).

The detection of carbohydrate in both peptides T-1 and C-2 shows that the point of attachment is an amino acid residue common to both, *i.e.* residues 9 to 14. It is commonly observed [25] that *N*-acetylglucosamine-containing carbohydrate moieties are linked to asparagine residues in the sequence Asn-X-Ser/Thr. Such a sequence is found in the present case at positions 13–15; hence we assume that residue 13 is the point of attachment.

It remains to comment on the molecular weight of the enzyme. The peptide molecular weight is 14555, and this is increased to 15800 if the carbohydrate content is taken into account. A value of 20000 was obtained from ultracentrifugation measurements [2]. This value was calculated using $\bar{v} = 0.723$ as obtained by the method of Cohn and Edsall [26] from the amino acid composition. Recalculation of $\bar{v}$ taking the carbohydrate content into account [27] leads to a value of 0.713 and to a molecular weight of 19200. A discrepancy of about 20% still remains between these values of the molecular weight. It may be that the revised value of $\bar{v}$ is still too high and

indeed there seems to be no theoretical justification for the practice of obtaining values of $\bar{v}$ from the sum of contribution of individual residues even though the values obtained appear, in some cases, to be close to those arrived at by direct measurements.

We thank Dr A. R. Berry for carrying out the carbohydrate analyses. Financial support from the Medical Research Council (G.L.C.) and from the Science Research Council (R.C.C.) is gratefully acknowledged as is the support of NRDC and the award of a Senior EMBO Fellowship to R.A.S.

REFERENCES

- De Haas, G. H., Postema, N. M., Nieuwenhuizen, W. & Van Deenen, L. L. M. (1968) *Biochim. Biophys. Acta*, **159**, 103–117.
- Shipolini, R. A., Callewaert, G. L., Cottrell, R. C., Doonan, S., Vernon, C. A. & Banks, B. E. C. (1971) *Eur. J. Biochem.* **20**, 459–468.
- Wells, M. A. & Hanahan, D. J. (1969) *Biochemistry*, **8**, 414–424.
- De Haas, G. H., Bonsen, P. P. M., Pieterse, W. A. & Van Deenen, L. L. M. (1971) *Biochim. Biophys. Acta*, **239**, 252–266.
- De Haas, G. H., Postema, N. M., Nieuwenhuizen, W. & Van Deenen, L. L. M. (1968) *Biochim. Biophys. Acta*, **159**, 118–129.
- De Haas, G. H., Slotboom, A. J., Bonsen, P. P. M., Van Deenen, L. L. M., Maroux, S., Puigserver, A. & Desnuelle, P. (1970) *Biochim. Biophys. Acta*, **221**, 31–53.
- De Haas, G. H., Slotboom, A. J., Bonsen, P. P. M., Nieuwenhuizen, W., Van Deenen, L. L. M., Maroux, S., Dlouha, V. & Desnuelle, P. (1970) *Biochim. Biophys. Acta*, **221**, 54–61.
- Shipolini, R. A., Callewaert, G. L., Cottrell, R. C. & Vernon, C. A. (1971) *FEBS Lett.* **17**, 39–40.
- Walton, P. L., Turner, R. W. & Broadbent, D. (1972) *U.K. Pat. No. 1263956*.
- Doonan, S., Doonan, H. J., Hanford, R., Vernon, C. A., Walker, J. M., Bossa, F., Barra, D., Carloni, M., Fassella, P., Riva, F. & Walton, P. L. (1974) *FEBS Lett.* **38**, 229–233.
- Roverly, M. (1967) *Methods Enzymol.* **11**, 231–236.
- Cole, R. D. (1967) *Methods Enzymol.* **11**, 315–317.
- Butler, P. J. G., Harris, J. I., Hartley, B. S. & Leberman, R. (1969) *Biochem. J.* **112**, 679–689.
- Waley, S. G. & Watson, J. (1953) *Biochem. J.* **55**, 328–337.
- Leggett Bailey, J. (1967) *Techniques in Protein Chemistry*, p. 23, Elsevier, London.
- Easley, C. W. (1965) *Biochim. Biophys. Acta*, **107**, 386–388.
- Acher, R. & Crocker, C. (1952) *Biochim. Biophys. Acta*, **9**, 704–705.
- Gray, W. R. (1967) *Methods Enzymol.* **11**, 469–475.
- Woods, K. R. & Wang, K. T. (1967) *Biochim. Biophys. Acta*, **133**, 369–370.
- Crowshaw, K., Jessup, S. J. & Ramwell, P. W. (1967) *Biochem. J.* **103**, 79–85.
- Hartley, B. S. (1970) *Biochem. J.* **119**, 805–822.
- Edman, P. & Begg, G. (1967) *Eur. J. Biochem.* **1**, 80–91.
- Brenner, M., Niederwieser, A. & Pataki, G. (1961) *Experientia*, **17**, 145–153.
- Bolton, C. H., Clamp, J. R. & Hough, L. (1965) *Biochem. J.* **96**, 5C–6C.
- Spiro, R. G. (1970) *Annu. Rev. Biochem.* **39**, 599–638.
- Cohn, E. J. & Edsall, J. T. (1943) *Proteins, Amino Acids and Peptides*, p. 370, Reinhold, New York.
- Gibbons, R. A. (1966) *Glycoproteins* (Gottschalk, A., ed.) pp. 29–95, Elsevier, Amsterdam.

R. A. Shipolini, G. L. Callewaert, and C. A. Vernon, Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, Great Britain

R. C. Cottrell's present address: Department of Biochemistry, University College London, Gower Street, London WC1E 6BT, Great Britain