The Structure of a Toxic Octapeptide from the Larvae of Sawfly

Dudley H. Williams, Sitthivet Santikarn, Francesco De Angelis, Richard J. Smith, and David G. Reid University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW Peter B. Oelrichs Department of Primary Industries, Animal Research Institute, Yeerongpilly, Queensland, Australia John K. MacLeod

Research School of Chemistry, Australian National University, Canberra, Australia

The amino acid sequence of an octapeptide from sawfly (*Lophyrotoma interrupta*) larvae, which acts as a liver toxin, has been determined as C_6H_5COAla -Phe-Val-Ile-Asp-Asp-Glu-Gln. The amino acids, Ala, Phe, Asp⁵, and Glu⁷ residues, have the D absolute configuration.

The poisoning of cattle by ingestion of larvae of an Australian species of sawfly (*Lophyrotoma interrupta*) is a serious problem in several of the grazing areas of Queensland.^{1,2} Generally, the disorder occurs in later winter or spring when there is a heavy infestation of the silver-leaf ironbark tree (*Eucalyptus melanophloia*), the principal host of the sawfly larvae. Large accumulations of dead larvae may collect on the ground. Animals eat such material, and develop a considerable liking for it. It has been suggested that it may act as a protein source. However, the end result is poisoning, and critically affected animals usually die within two days.

The first report of the isolation of a toxic factor is due to Leonard.³ An improved method for isolation of the toxin was reported by Oelrichs et al.⁴ They reported the isolation of an essentially pure toxin with an LD_{100} of 2 mg/kg. Amino acid analysis showed the presence of the following amino acids (molar ratios given in parentheses): Asp (1.89), Glu (1.93), Ala (1.00), Val(1.16), Ile(0.82), Phe(0.92) (taking Ala = 1.00). The *N*-terminus of the putative peptide was found to be blocked, but mass spectrometry established benzoyl as the N-terminal-blocking group and determined the partial sequence C₆H₅-COAlaPheValIle This partial sequence was established through the occurrence of the expected acyl sequence ions at m/z 105, 176, 323, 422, and 535 in the mass spectrum of the toxin, and at m/z 105, 190, 351, 464, and 591 in the mass spectrum of the permethylated toxin. On the assumption that the amino acid analysis accounts for all the components of the toxin (now named lophyrotomin⁴), its structure is therefore represented by $C_6H_5COAlaPheValIle(Asx_2Glx_2)$, where Asx and Glx represent aspartyl or asparagyl, and glutamyl or glutaminyl, residues respectively. The remaining problems in the structure determination are therefore (i) to establish the sequence and nature of the last four residues, (ii) to establish the absence of other entities in the toxin, and (iii) to determine the absolute configurations of the amino acids. The experiments which determine these points are now described.

Results and Discussion

(i) N.M.R. Studies.—In order to obtain information as to the overall composition of the toxin, its 270-MHz ¹H n.m.r. spectrum was obtained, using a *ca*. 0.03M-solution in $(CD_3)_2$ -SO. The spectra were obtained at various temperatures in the range 25—75 °C, and both before and after the addition of D₂O in order to determine the number of NH resonances in the sample. Although electrophoretic experiments indicate

 Table 1. Relative numbers of protons in various environments in lophyrotomin

δ (p.p.m.)	No. of protons ^a	Assignment
0.5 1.0	10	CH ₃
0.51.9	19	$CH(CH_3)_2$ CH(CH_1)CH_2CH_1
1.7-3.2	12	$-CH_2CO^-$ (×2)
39-17	74	$-CH_2CH_2CO^-$ (×2)
7.1-8.0	11.7	Aromatic H
6.99.0	9 ± 1 b	NH

^a Values normalised relative to 19 protons in the 0.5–1.6 region. ^b Value obtained by D_2O exchange.

appearance of a very broad, multi-proton resonance in the region 4.8-6.5 p.p.m. at temperatures in the range 45-75 °C. The results of integration of the proton spectrum are given in Table 1.

On the assumption that lophyrotomin is an octapeptide containing two aspartic acid and two glutamic acid residues, then the number of protons appropriate to rows 1-5 of Table 1 are 19:12:8:10:8. Thus, within the limits of accuracy of the spectral integrations, the ¹H n.m.r. spectrum indicates that lophyrotomin is indeed an octapeptide but with the proviso that *either* the four (Asx + Glx) residues have carboxylic acid side chains or three of these residues have carboxylic acid side chains and one is present as Asn or Gln (or possibly iso-Asn or iso-Gln at the C-terminus). These possibilities correspond to eight and ten exchangeable NH protons, respectively. The latter possibility is supported by the observation in a 400-MHz spectrum [see section (v)], of two relatively sharp NH proton singlets at 6.20 and 6.85 p.p.m. Upon irradiation of one of these resonances, transfer of saturation to the other is observed, consistent with their assignment to a CONH₂ group.

The overall composition of lophyrotomin was also investigated by ¹³C n.m.r. at 67.8 MHz in D₂O solution. The spectrum was recorded with proton-noise decoupling. The S : N ratio in the δ 170—186 p.p.m. region was insufficient to give a count of the carbonyl carbons, but signals in the 128—138 p.p.m. region were consistent with the presence of two aromatic rings. Signals at δ 51.0, 52.3 (×2), 54.9, 55.3, 55.9, 59.4, and 60.2 p.p.m. indicated the presence of eight α -CH carbons. The δ 10—40 p.p.m. region contained fifteen clearly resolved



The octapeptide structure requires fifteen resonances in this region.

(ii) Sequence Studies and Molecular Weight Determination by Mass Spectrometry.-Since sequence information had not been obtainable beyond the first four residues of the toxin, selective cleavage of the peptide backbone was attempted, to be followed by mass spectral sequencing of the fragments. Attempts to cleave the peptide enzymatically (chymotrypsin, papain, carboxypeptidase A, carboxypeptidase Y, thermolysin) were unsuccessful. This failure was due to one or more of the amino acids not possessing the L-absolute configuration (as confirmed subsequently). Attempted hydrazinolysis (hydrazine hydrate at 130 °C for 30 min) also failed to cause significant cleavage. However, reaction with 10% methanolic HCl for 10 days at room temperature gave a mixture containing both unchanged toxin and methyl esters of amine hydrochlorides. This mixture was subjected to direct analysis by mass spectrometry, components of the mixture being partially separated by fractional volatilisation from the direct insertion probe. Peaks which appeared to contain useful structural information, or intense peaks whose origin was obscure, were selected for further study. Among these cases were: at a source temperature of 220 °C, m/z 116 and 259; at 240 °C, m/z 388, and 429; and at 260 °C, m/z 487. Metastable peaks established at m/z 388 and 487 both lose methanol to give m/z 356 and 455, respectively.

All these peaks, except m/z 429, were initially interpreted as corresponding to the ions (1)—(5). For clarity, the structural inferences are also indicated. It will be evident that the sequence of Asp and Glu residues in (3)—(5) is arbitrary. Since these interpretations cannot give the order of the Glu and Asp amino acids by reference to any one ion, only comparison between ions indicates how they might be joined together.



The ambiguity in potential information from the m/z 487 moiety (5) arises since the charged (right-hand terminal) fragment could be $-N=CHCO_2Me$ [rather than as written in (5)], and in this case the ' in chain ' aming saids would both have

and in this case the 'in-chain' amino acids would both have to be glutamic acids. The interpretations are supported by the mass shifts of the ions (1)—(5) in the mass spectra of products from reaction in CD₃OD-HCl (followed by brief treatment with H₂O to back-exchange ND to NH). The ions then appear at m/z 119, 265, 397, 364, and 497, consistent with their containing 1, 2, 3, 2, and 3 methoxy groups, respectively.

Persuasive as these interpretations appear, being based on an established favourable fragmentation of simple peptides,⁵ they are incorrect except in the case of (1). This is established by high resolution measurements which are given in Table 2.

Except for (1), the compositions of the ions differ from the structural formulae accompanying (2)—(5) in so far as a nitrogen atom must be replaced by CH_2 . In order to deduce the origins of these related ions, is it clearly best to consider first that of lowest mass (m/z 259). The composition of this ion establishes that it contains three double-bond equivalents, two of which must be associated with the carbonyls of ester groups (see above); it is highly probable that the third is due to a carbonyl of the peptide backbone. Since the ion is otherwise saturated, it must correspond to a radical-cation. The

Table 2. High resolution data on ions (1)—(5), and on the decomposition product, m/z 455, of (5)

Ion (m/z)	Composition	Error (p.p.m.) ª	Ion (m/z)	Composition	Error (p.p.m.) ª
(1) (116)	C ₄ H ₁₀ NO ₂	1	(4) (358)	C17H30N2O6	1
(2) (259)	C ₁₂ H ₂₁ NO ₅	0	(5) (487)	C ₂₂ H ₃₇ N ₃ O ₄	4
(3) (388)	$C_{17}H_{28}N_2O_8$	6	(455)	C ₂₁ H ₃₃ N ₃ O ₈	0
				•.•	

" Differences between measured mass and the mass calculated from the given composition.

number of carbon atoms demands that all, or major parts, of at least two amino acids be incorporated in the ion. Therefore, formation of the ion must occur via a process which results in an amino acid losing its amino group. All these requirements can be satisfied if m/z 259 corresponds to (2'). The ions m/z388, 258 and 487 then correspond to (3'), (4'), and (5'), respectively. All can be generated in a process in which a hydrogen rearranges to the α -carbon of an amino acid with associated loss of its amino group; for example, (6) \longrightarrow (7), which assumes (without evidence) that the fragmentation is electronimpact induced, rather than thermally induced with subsequent ionisation of a fragment.

These interpretations lead to a self-consistent sequence for the octapeptide of $C_6H_5COAlaPheValIleAsxAsxGlxGlx$.

Since the above sequence for an octapeptide is based on (i) a novel and unexpected fragmentation and (ii) a lack of direct observation of the two Glx residues, independent evidence for the sequence was sought and found. N-Benzoylalanine methyl ester (8) is produced in relatively large amounts upon treating lophyrotomin with 10% methanolic HCl for 10 days at room temperature. This relatively volatile component is identified from peaks in its mass spectrum at m/z 105, 148, and 207; these peaks appear at m/z 105, 148, and 210 when the methanolysis is carried out in CD₃OD-HCl (followed by back exchange of ND for NH), and this component shows an abundant quasi-molecular ion (MH^+) in its NH₃ chemical ionisation spectrum at m/z 208. A corollary of this finding is that the product mixture should contain the amine hydrochlorides of one or more peptides containing N-terminal phenylalanine. Two aliquots of the mixture from methanolysis and CD₃OD-HCl were therefore taken; one was acetylated with Ac₂O-MeOH (1:3) and the other with $[^{2}H_{6}]$ -Ac₂O-MeOH (1:3). The former product showed abundant sequence ions at m/z190, 289, 402, 534, and 666. The ions correspond to the acylium ions formed by cleavage, as indicated in (9). The ions. Table 3. High resolution data on m/z 196 and 339

Ion	Composition	Error (p.p.m.)
<i>m</i> / <i>z</i> 196	C ₉ H ₁₀ N ₁ O ₄	0
<i>m</i> / <i>z</i> 339	$C_{15}H_{19}N_2O_7$	0



latter undergoes two successive losses of MeOH to m/z 307 and 275, with appropriate metastable peaks at m/z 278.0 and 246.3. The mass difference (143 a.m.u.) between m/z 196 and 339 corresponds to glutamic acid residue (as its methyl ester); this inference is confirmed by high resolution measurements (Table 3).

Since m/z 196 shifts to m/z 199 when methanolysis is carried out with methanol containing a CD₃ group, this ion must contain one methyl ester function. This requirement, and its composition, are satisfied if the ion arises by in-chain cleavage at an Asp residue (10) \longrightarrow (11) which in turn is attached at its C-terminal side to a Glx residue which appears as pyro-Glu in the resulting ion. Thus, this well precedented fragmentation ⁶ at Asp confirms the formerly presumed sequence \cdots AsxGlxGlx in lophyrotomin.

The molecular weight of lophyrotomin, and an independent sequence determination of its consistituent amino acids was







Figure 2. Negative ion FAB mass spectrum of lophyrotomin in the region m/z 410–1050

Table 4. Amino acid composition after treatment with TIB

Amino acid ratio ^a	Starting material				Control
	Ac-Gln	Ac-iso-Gln	2,4-DAB °	Liver toxin	Liver toxin
Ala				1.00	1.00
Phe				0.96	0.91
Val				0.94	1.16
Ile				1.01	0.82
Asp				1.93	1.89
Glu	0.3 *	0.3 ^b		1.19	1.93
2,4-DAB	0.6 ^b		1.0 "	0.89	

" As 4-oxobutyric acid lacks a primary amino group, it does not form a coloured derivative with ninhydrin and thus is not detected in the amino acid analyser. ^b Relative to the amount of the starting material. ^c 2,4-DAB, Lys, and Orn are not resolved on this analyser.

occur in the digestive system, or quickly absorbed. The alternative approach was to remove the benzoyl blocking group, or any *N*-terminal fragments, enzymically (thermolysin, subtilisin, papain) or chemically (partial acid hydrolysis with

27 8

of the Hofmann rearrangement products with naphthalene-1sulphonyl chloride prior to acid hydrolysis. The resulting 1-naphthylsulphonylated amino acids were analysed by reverse phase h.p.l.c. The advantage of this method is that both





Figure 3. ¹H N.m.r. spectrum of lophyrotomin at 400 MHz in (CD₃)₂SO with sodium 4,4-dimethyl-4-silapentanesulphonate (DSS as internal reference

gc-ms on a 'Chirasil-Val' column. The amino acid derivatives were identified by their retention times as well as their positive and negative ion CI-MS, while their absolute configurations were determined by co-chromatography with authentic derivatives. The hydrolysate contained D- and L-Glu in the ratio *ca*. 93 : 7, and D- and L-DAB in the ratio *ca*. 5 : 95. Thus the penultimate residue, Glu⁷, has the D-configuration.

The presence of D-amino acids in liver toxin strongly indicates that the peptide was generated by a non-ribosomal biosynthetic pathway. Therefore it cannot be assumed that all the aminodicarboxylic acids, D- and L-Asp and D-Glu, form peptide bonds through the α -carboxy groups. The possibility that has and/or Clu mouths arrest in the possibility



	Table 5. Isotopic abundance and relative amounts of some D- and L-amino acid derivatives in LT-C and LT-D	Table 6. N.m.r. assignments of the spectrum in Figure 3 Signal Assignment
<u>.</u>		
й —		
		,
1		
• • •		
	<u> [a</u>	
· ·		
\$		
<u> </u>		
	0	
1		
i aliy		
- {		
ŧŧ.		
<u>.</u>		
•1		
,= 	<u>→</u> • <u>x</u>	
a		
1 		
i		
<u>, 1</u>		
₹		
)		
-		
\$		

methylsilane was employed as internal standard, and for ${}^{13}C$ spectra in D₂O, sodium 3-trimethylsilyl[2,2,3,3- ${}^{2}H_{4}$]propionate (TSP) was used as internal standard.

Electron impact mass spectra were obtained on an AEI MS9 instrument operating at 70 eV with a resolving power (10% valley definition) of *ca*. 1 500. Simple mixtures of peptides, produced by partial hydrolysis of lophyrotomin in HCl-MeOH, were partially resolved by repeated introduction of the sample into the mass spectrometer source over the temperature range 200—300 °C. High resolution measurements on selected peaks (to provide atomic compositions) were carried out at a resolving power of *ca*. 20 000, using peaks in the spectrum of heptacosafluorotributylamine to provide reference masses.

Chemical ionisation (CI) mass spectra were recorded on a VG 70-70 instrument, using ammonia as the reagent gas. FAB mass spectra were obtained on a Kratos MS 50 mass spectrometer. Samples of the toxin in glycerol were hombarded with followed by u.v. absorption at 260 nm. The eluate was freezedried to give the pure acidic form of the toxin $(MH^+ m/z 1040; Figure 1)$.

Partial Acid-hydrolysis of Lophyrotomin with HCl-MeOH.— Lophyrotomin (2.5 mg) was left in a solution of 10% HCl-MeOH (2 ml) at room temperature for 10 days. The reagents were then removed by evaporation at room temperature. The product was analysed by mass spectrometry, as indicated in the text, employing a temperature gradient at the probe as described above.

This hydrolysis was repeated on a separate sample of lophyrotomin, using 10% HCl-CD₃OD as the reagent.

Acylation of the Product of Partial Methanolysis.—The product (2.0 mg) from treatment with 10% HCl-CD₃OD, as described above, was dissolved in methanol (1.2 ml), and the resulting solution separated into two equal portions. To one

		,

solved in ethyl acetate (100 µl) in preparation for g.l.c. analysis on a ' Chirasil Val ' column (25 m \times 0.24 mm I.D.) in an Erba Science Fractovap 4160 capillary gas chromatograph, programmed for a temperature gradient of 4 °C/min from 80 to 180 °C, with 4-min holds at 80 and 180 °C. The above solution (1.5 μ l) was used in a splitless injection. The carrier gas was hydrogen at a pressure of ca. 9 lb in⁻² and the injector temperature was 300 °C. The retention times of the derivatives established the conclusions given in the text.

Hofmann Degradation, Amino Acid Analysis, and Configuration at Glu⁷ and Ala¹.—Material and instruments. 2,4-Diaminobuytric acid hydrochloride, naphthalene-1-sulphonyl chloride (Sigma), I,I-bis(trifluoroacetoxy)iodobenzene and glutamine (Fluka) were used as supplied. H.p.l.c. grade water was obtained from BDH. H.p.l.c. grade propan-2-ol from Fisons was distilled in all-glass apparatus before use. Unless otherwise stated, all solvents and chemicals used were AnalaR grade.

Amino acid analyses were carried out on a Beckman amino acid analyser model 119CL, and h.p.l.c. analyses on a Du Pont 850 Liquid Chromatographic System with a μ -Bondapak C₁₈ $(0.7 \times 30 \text{ cm}; \text{Waters Associates})$ reverse phase column. G.l.c. was carried out on a Perkin Elmer F-17, adapted to take capillary columns. The carrier gas was hydrogen with a head pressure of ca. 160 kN/m². Make-up gas was nitrogen at a pressure of ca. 12 kN/m², and a flame ionisation detector was used. The column temperature was programmed as for the determination of configuration of Asp⁵, and the injector temperature was 250 °C.

Hofmann rearrangement with I,I-bis(trifluoroacetoxy)iodobenzene. Lophyrotomin (100 nmol) was dissolved in 50% aqueous acetonitrile (40 μ l) and an equimolar amount (>100 nmol, 0.044 mg) of TIB dissolved in ca. 10 µl of the same solvent was added. The mixture was stirred at room temperature for 3 h and 0.01M-HCl (ca. 150 µl) was then added. The solution was extracted once with diethyl ether and the aqueous phase then freeze-dried overnight.

Total acid hydrolysis. Samples (ca. 100 nmol) were sealed in Pyrex test tubes with 6м-HCl (ca. 0.5 cm³, amino acid analysis grade), and heated at 110 °C for 14 h. The solvent was removed by rotary evaporation.

Treatment with naphthalene-1-sulphonyl chloride. Starting material (100 nmol) was dissolved in 0.2M-NaHCO₃ solution (100 μ l) and a freshly prepared aqueous acetone solution (0.25% w/v) of naphthalene-1-sulphonyl chloride (200 µl) was added. The mixture was incubated at 45 °C for 30 min and the solvent was then removed under reduced pressure.

H.p.l.c. of 1-naphthylsulphonylated amino acids. The reverse phase column was kept at 45 °C (flow rate 2.00 cm³/min) and the u.v. absorption at 340 nm was monitored. The aqueous solvent used was 5% v/v aqueous acetic acid, and propan-2-ol was used as organic modifier. The programmed composition of organic modifier was as follows: isocratic at 5% v/v for 10 min, followed by a linear gradient from 5 to 25% v/v in 20 min, then constant solvent composition at 25% v/v for 5 min, and then another linear gradient from 25 to 100% v/v in 5 min.

G.l.c. analysis. Using the conditions outlined in this section, and the derivatisation conditions for the preparation of Ntrifluoroacetyl isopropyl esters given previously, the amino acids from the Hofmann degradation product were shown to include D- and L-Glu (93:7) and D- and L-DAB (ca. 5:95).

These assignments were based upon co-chromatography with authentic samples and, in addition, upon gc-ms analysis. This latter analysis was performed on a Finnigan 4000 GC-MS with 6110 data system. Detection was by electron impact and, in a separate run, by alternating positive ion and negative ion chemical ionisation. Integration of the appropriate peaks gave the ratios for D:L Glu and D:L DAB shown in parentheses above. Experiments, carried out on a total hydrolysate of the toxin, established the absolute configurations D-Ala, D-Phe, L-Val, L-Ile, equimolar amounts of L- and D-Asp, and equimolar amounts of L- and D-Glu. This last experiment confirms the earlier assignments of absolute configuration of seven amino acids using the prolyl ester method, and establishes the D-configuration at Ala¹.

Acknowledgements

We thank Dr. Bengt Sandberg for help, Dr. Alan Jones and his staff for use of the facilities of the National NMR Centre, A.N.U., Canberra, Dr. E. Houghton and Mr. P. Teale for running gc-ms on the equipment at the Racecourse Security Services Labs., Newmarket, and Dr. John Shaw, DSIR, Palmeston North, New Zealand, for complementary analyses of amino acid derivatives on a ' Chirasil-Val ' column.

One of us (D. H. W.) thanks the Australian National University for financial support, and the Royal Society for the award of a Commonwealth Bursary to undertake the initial parts of the work. We acknowledge support from the S.E.R.C. (U.K.) and the Royal Society for the research carried out in Cambridge.

References

- 1 F. H. S. Roberts, Queensl. Agric. J., 1932, 37, 41.
- 2 L. L. Callow, Queensl. Agric. J., 1955, 81, 155.
- 3 G. J. Leonard, Toxicon, 1972, 10, 597.
- 4 P. B. Oelrichs, P. J. Vallely, J. K. MacLeod, J. Cable, D. E. Kiely, and R. E. Summons, *Lloydia*, 1977, **40**, 209. 5 K. Biemann, 'Mass Spectrometry,' McGraw-Hill, 1962, pp.
- 291-296.
- 6 H. R. Morris, D. H. Williams, G. Midwinter, and B. S. Hartley, Biochem. J., 1974, 141, 701.
- 7 M. Barber, R. S. Bordoli, R. D. Sedgwick, and A. N. Tyler, J. Chem. Soc., Chem. Commun., 1981, 325.
- 8 D. H. Williams, C. V. Bradley, G. Bojesen, S. Santikarn, and L. C. E. Taylor, J. Am. Chem. Soc., 1981, 103, 5700.
- 9 A. S. Radhakrishna, M. E. Parham, R. M. Riggs, and G. M. Loudon, J. Org. Chem., 1979, 44, 1746.
- 10 See, for example, E. Gil-Av and B. Feibush, Tetrahedron Lett., 1967, 3345.
- 11 J. C. Dabrowiak and D. W. Cooke, Anal. Chem., 1971, 43, 791.
- 12 R. Liardon, S. Ledermann, and U. Ott, J. Chromatogr., 1981, 203. 385.
- 13 S. Kusumoto, M. Matsukura, and T. Shiba, Chem. Lett., 1981, 1117.
- 14 H. Iwase and A. Murai, Chem. Pharm. Bull., 1974, 22, 8; ibid., 1977. 25, 285.
- 15 G. N. Holcomb, S. A. James, and D. N. Ward, Biochemistry, 1968, 7, 1291.
- 16 O. Jardetzky and G. C. K. Roberts, 'NMR in Molecular Biology,' Academic Press, New York, 1981, p. 151.
- 17 A. Bundi and K. Wüthrich, Biopolymers, 1979, 18, 285.

Received 28th October 1982; Paper 2/1820