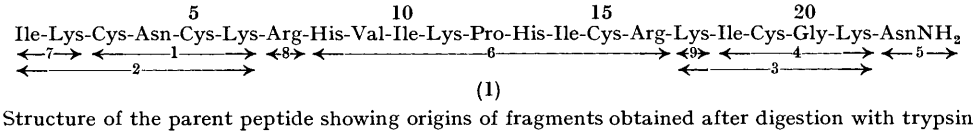
Identification by Mass Spectrometry of N^{ϵ}-Formyl-lysine Residues in a Peptide from Bee Venom

By **Shawn Doonan,* Andrew J. Garman, Jennifer M. Hanson, Alexander G. Loudon,** and **Charles A. Vernon,** The Christopher Ingold Laboratories, Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ

A mixture of variant forms of the peptide (1) has been isolated from bee venom. The variants have the same amino-acid sequence and arrangement of disulphide bridges as the parent peptide but differ from it in that the side chains of lysine residues at positions 2, 17, and possibly 21 are chemically modified. The modifying group has been identified, by mass spectrometry, as formyl. Each variant probably contains only one modified lysine residue.

THE peptide (1) comprises 2—3% by weight of the dried venom of the common European honey bee (*Apis mellifera*).[1,2] The peptide has the ability to degranulate mast cells[1] but has also been shown to have anti-inflammatory activity;[3] it has been named both MCD-peptide[1] and peptide-401.[2] The amino-acid sequence of the peptide was determined independently by two groups of workers.[4,5] The *C*-terminus is amidated; there are no free carboxy-groups and consequently the peptide is highly basic. The two disulphide bridges link residues 3 to 15 and 5 to 19 respectively.[3,6] In addition to the parent peptide, a variant form was detected in bee venom[1] and has subsequently been isolated.[2] This report describes investigations of the differences between the parent and variant forms.



(1)

Structure of the parent peptide showing origins of fragments obtained after digestion with trypsin

[1] H. Breithaupt and E. Habermann, *Arch. Pharm.*, 1968, **261**, 252.
[2] J. Gauldie, J. M. Hanson, F. D. Rumjanek, R. A. Shipolini, and C. A. Vernon, *European J. Biochem.*, 1976, **61**, 369.
[3] M. E. J. Billingham, J. Morley, J. M. Hanson, R. A. Shipolini, and C. A. Vernon, *Nature*, 1973, **245**, 163.
[4] P. von Haux, *Z. physiol. Chem.*, 1969, **350**, 536.
[5] C. A. Vernon, J. M. Hanson, and R. W. Brimblecome, B.P. 1,314,823/1969.
[6] J. Gauldie, J. M. Hanson, R. A. Shipolini, and C. A. Vernon, *European J. Biochem.*, 1978, **83**, 405.

RESULTS AND DISCUSSION

Peptide (1) and the variant appeared to have identical amino-acid compositions (Table 1). The variant migrated more rapidly (by a factor of approx. 1.7) than

TABLE 1

Amino-acids liberated by acid hydrolysis from peptide (1) and from the variant form. Values have been normalized to a recovery of 1 mol of glycine in each case.

	Time of hydrolysis			
	22h	96 h	24 h	96 h
Amino-acid	Peptide (1)		Variant	
Aspartic acid	2.03	2.19	1.89	2.03
Proline	1.20	1.27	1.06	1.17
Glycine	1.00	1.00	1.00	1.00
Valine	0.47	0.95	0.57	0.92
(Cystine)/2	3.83	3.61	3.53	3.66
Isoleucine	3.35	3.61	2.80	3.64
Lysine	4.74	4.89	4.05	4.76
Histidine	1.77	1.90	1.57	1.77
Arginine	1.94	2.00	1.79	1.91

peptide (1) on t.l.c. and was less basic as judged by electrophoresis at pH 3.5 and 6.5. These differences were maintained after reduction of disulphide bridges and reaction with iodoacetic acid (carboxymethylation) hence the peptides do not differ in the number or arrangement of disulphide brides. *N*-Terminal analysis of the variant gave the same *N*-terminal residue (Ile) in approximately the same yield as peptide (1); hence the *N*-terminus of the variant is not blocked. Sequence analysis of the reduced and carboxymethylated variant was carried out by the Edman method. The first 20 amino-acids were identified unambiguously and were the same as in peptide (1). Digestion of the variant with trypsin yielded asparagine amide [2] as one of the products showing that the *C*-terminal residue is the same as in peptide (1). Consistently, carboxypeptidase A was without effect on the peptide.

The results given up to this point are consistent only with the hypothesis that the amino-acid sequences of peptide (1) and of the variant are identical and that the peptides differ in that one or more of the amino-acid side chains in the variant carries a substituent. This substituent must either be labile to acid or such as not to be detectable by the techniques used for amino-acid analysis and sequence determination. Evidence that the modified residue was lysine came from the observation that after extensive digestion with pronase the yield of lysine from the variant was less than from peptide (1) whereas the yields of the other amino-acids released in major amounts (asparagine, isoleucine, histidine, and arginine) were identical in the two cases.

Both peptides were reduced and carboxymethylated, digested with trypsin, and the digests subjected to peptide mapping (chromatography in direction 1 and electrophoresis at pH 6.5 in direction 2) on cellulose thin-layer plates. Peptide (1) yielded nine major ninhydrin-positive spots. The same nine components were obtained after digestion of a larger amount of the peptide (40 mg) and separation of the digest by preparative paper chromatography and electrophoresis. The purified components were identified by *N*-terminal analysis, amino-acid analysis and, in some cases, by determination of their sequences. The fragments arose from cleavage at the points shown in (1); they were numbered in order of increasing mobility on electrophoresis at pH 6.5. In addition to the expected fragments, two more (2 and 3) were obtained arising from failure to cleave completely at lysine residues 2 and 17; no cleavage was observed at residue 11, no doubt because the adjacent residue is proline.

Digestion of the variant of peptide (1) yielded the same nine fragments as did the parent peptide plus two other major components not obtained in the previous digest. These extra components were found to have amino-acid compositions identical to those of peptides 3 and 7 obtained from the parent peptide, but in both cases the new fragments migrated more rapidly on chromatography and were less basic than those obtained from the parent peptide; these new fragments will be referred to as 3′ and 7′. Peptide 7′ gave isoleucine on *N*-terminal analysis and hence was not blocked on the α-amino-group. Given the other evidence described above, these results suggested that peptides 3′ and 7′ contained blocked lysine residues.

Low-resolution mass spectrometry seemed to be the method of choice for identification of the blocking group in these peptides. This method has been applied with success not only to sequence analysis of peptides containing the commonly occurring amino-acids,[7] but also to peptides with blocked *N*-terminal residues [8] and to peptides containing unusual amino-acids.[9] Peptides are rendered sufficiently volatile for mass spectrometry by acetylation [10] followed by permethylation with methyl iodide in the presence of methyl sulphinyl carbanion.[11] Since both acetyl and methyl were considered possible blocking groups for the lysine side chains in the present case (although methylation was much less likely since N^{ϵ}-methyl-lysine is stable to acid and should have been resolved from lysine on amino-acid analysis) it was necessary to consider preparing derivatives for mass spectrometry in which either the acetyl groups or the methyl groups introduced were deuteriated. In practice, analysis of peptides derivatized with CD_3I allowed identification of the blocking group.

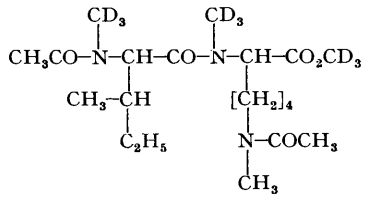
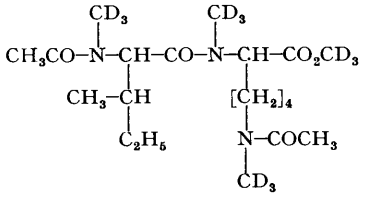
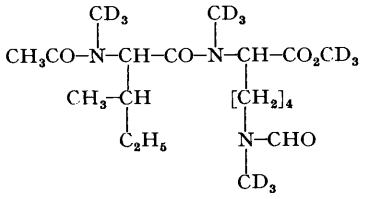
Peptides 3′ and 7′ were isolated individually from a trypsin digest of the reduced and carboxymethylated variant by using paper chromatography (R_F values 0.40 and 0.65 respectively). The peptides were acetylated, permethylated using CD_3I, and analysed by mass spectrometry over a range of source temperatures. Peptide 7′ (blocked Ile–Lys) yielded a clean mass spectrum at a

[7] H. R. Morris, D. H. Williams, and R. P. Ambler, *Biochem. J.*, 1971, **125**, 189.
[8] H. R. Morris and A. Dell, *Biochem. J.*, 1975, **149**, 754.
[9] H. R. Morris, A. Dell, T. E. Petersen, L. Sottrup-Jensen, and S. Magnusson, *Biochem. J.*, 1976, **153**, 663.
[10] D. W. Thomas, B. C. Das, S. D. Gero, and E. Lederer, *Biochem. Biophys. Res. Comm.*, 1968, **33**, 519.
[11] S. I. Hakamori, *J. Biochem.* (*Japan*), 1964, **55**, 205.

source temperature of *ca.* 170 °C. The most prominent features were peaks at *m/e* 173, 145, 131, and 103 corresponding to *N*-terminal isoleucine (Table 2), and

TABLE 2

Possible structure for derivatized peptide 7′ assuming that the natural blocking group is methyl, acetyl, or formyl

Natural blocking group	Derivatized peptide	*m/e* of parent ion
Methyl		408
Acetyl		411
Formyl		397

to fragmentation products arising from loss of CO, of CH_2CO, and of both these from the *N*-terminal ion. The parent ion had *m/e* 397 corresponding to a *C*-terminal ion mass of 224. This observation showed that the natural blocking group was neither acetyl nor methyl but was consistent with it being formyl (Table 2); the other evidence quoted above was fully in accord with this conclusion. The other major peak observed in the spectrum arose from loss of CO_2CD_3 from the parent ion (*m/e* 335). It should be pointed out that the results from mass spectrometry were also consistent with the *C*-terminal residue of peptide 7′ being ornithine (unblocked); after derivatization in this case the side-chain amino-group would have carried the substituents CD_3 and $COCH_3$. This possibility can be ruled out, however, since ornithine and lysine are well separated by the system used for amino-acid analysis and by the chromatographic procedure used to identify dansyl (5-dimethylaminonaphthalene-1-sulphonyl) amino-acids and amino-acid phenylthiohydantoins.

Fragment 3′ (blocked Lys–Ile–CMCys–Gly–Lys) yielded a mass spectrum at a source temperature of 230 °C that clearly showed it to be a mixture of two peptides. One yielded an *N*-terminal sequence ion at *m/e* 233 {corresponding to $[CH_3CO{-}N(CD_3){-}CHCO{-}(CH_2)_4{-}N(CD_3)CHO]^+$} and arising from a peptide in which the ε-amino-group of the *N*-terminal lysine was formylated; the other *N*-terminal ion was at *m/e* 247 {corresponding to $[CH_3CO{-}N(CD_3){-}CHCO{-}(CH_2)_4{-}N(CD_3)COCH_3]^+$} arising from a peptide in which the ε-amino-group of the *N*-terminal lysine was originally unmodified. Ions at *m/e* values of 219 and 205 arising from loss of CO from both the *N*-terminal ions were observed. In addition, ions at *m/e* values of 363 and 377 were observed corresponding to the *N*-terminal dipeptides from both components of the mixture. Although ions of higher *m/e* values were present in the spectrum, they could not be assigned with certainty. This was probably due to the presence of carboxymethylcysteine at position 3 in the peptides. The apparent presence of the two distinct peptides in fragment 3′ requires comment. It is unlikely to have arisen from partial deformylation and subsequent acetylation during derivatization since this did not occur when peptide 7′ was treated under the same conditions. Similarly, peptide 3′ could not have been contaminated with peptide 3 (the non-formylated version) since the chromatographic behaviour of the two was very different. It seems most probable that fragment 3′ was a mixture of two peptides, one formylated on the *N*-terminal lysine and the other formylated on the *C*-terminal residue. This hypothesis could not have been confirmed by mass spectrometry of fragment 3′ since the parent ions in the two cases would have had the same mass; it would be necessary to isolate peptides from the variants containing lysine residues 17 and 21 separately to clarify this point.

The conclusion to be drawn from these studies is that the variant of peptide (1) isolated from bee venom is a mixture of peptides each formylated on one of the lysine residues 2, 17, and probably 21. That the variant is not a single species doubly or triply formylated is shown by the isolation of non-formylated peptides containing these lysine residues from the trypsin digest. It is possible that doubly or triply blocked peptides are present in bee venom, but in small amounts. Formylation of lysine residues may not be unique to the variant of peptide (1) since a modified form of another major component of bee venom (apamin) also exists,[1,2] the properties of which again suggest that it contains a blocked lysine residue; the blocking group in this case has yet to be identified. It should be emphasized that the formyl groups in the variants occur naturally and were not introduced by the procedures used to isolate the components of the venom. Although isolation of peptide (1) is usually carried out by ion-exchange chromatography using formate buffers [2] the use of other buffer systems does not decrease the amount of the variants obtained. Similarly, incubation of purified peptide (1) in formic acid (0.1M) for several days at room temperature followed by lyophilization does not lead to variant forms; this treatment is similar to that used for initial separation of peptide material from the whole venom.[2] The venom itself does not contain formic acid,[12] and

[12] A. T. Tu, 'Venoms: Chemistry and Molecular Biology,' John Wiley, New York, 1977.

hence formylation does not occur during drying of the original material. The biological significance of formylation of peptide (1) (and perhaps of other components of the venom) is, however, obscure.

A point of general interest arises from this work in that the peptide bond *C*-terminal to the formyl-lysine residue at position 2 in the variant and also that at position 21 if this residue is indeed formylated, was susceptible to hydrolysis by trypsin. This is contrary to the accepted views about the specificity of trypsin and clearly merits further investigation.

EXPERIMENTAL

Analytical Methods.—Thin-layer chromatography was carried out using Merck cellulose plates. The chromatograms were developed for about 18 h using butan-1-ol–pyridine–acetic acid–water (90 : 60 : 18 : 72; v/v) as solvent. Analytical electrophoresis (70 V cm^{-1}, 15—20 min) was similarly carried out using cellulose thin-layer plates with buffers at pH 3.5 or 6.5 (pyridine–acetic acid). Peptide mapping combined chromatography in the first direction with electrophoresis at pH 6.5 in the second. Samples for amino-acid analysis were hydrolysed with HCl (6M) at 105 °C under vacuum for 22—96 h, and then analysed using a Technicon Auto Analyzer. *N*-Terminal analysis was carried out by the dansyl method,[13] the dansyl amino-acids being identified by t.l.c. on polyamide sheets (7.5 cm × 7.5 cm) using the solvent systems described by Hartley.[14] Amino-acid sequence analysis of reduced and carboxymethylated peptide (1) and its variant was carried out by a modification of the Edman method; [15] the modification has been described in detail elsewhere.[6] Sequence analysis of peptides obtained after proteolytic digestion of parent peptides was carried out by the dansyl Edman method.[14]

Isolation of Peptides from Bee Venom.—Peptide (1) and the variant were isolated as previously described.[2] Final traces of peptide (1) were removed from the sample of the variant by preparative paper chromatography (Whatman 3MM paper, solvent system as above). The product was homogeneous to analytical chromatography.

Trypsin Digestion of Peptides.—Samples of peptides were reduced and carboxymethylated before digestion with trypsin. The following example is typical. The variant peptide (14.5 mg, 5.6×10^{-6} mol) was incubated in Tris–HCl buffer (0.2M, pH 8.6, 2.9 ml) containing urea (8M) and dithioerythritol (19.1 mg, 0.124 mmol) for 4 h under N_2. Iodoacetic acid (49.7 mg, 0.267 mmol) in NaOH (2M; 0.148 cm^{-3}) was added and reaction allowed to proceed for 15 min. Low-molecular-weight material was removed from the peptide by gel filtration through a column of Sephadex G-25 equilibrated with water and the peptide recovered by lyophilization. The modified peptide (10 mg) was dissolved in *N*-ethylmorpholine–HCl buffer (0.1M, pH 8.0, 10 ml). Trypsin (Sigma, DCC treated, 0.2 mg) was added and the reaction mixture incubated at 37 °C for 3 h. The proteolysis products were recovered by lyophilization.

Purification of Proteolysis Products.—This was generally done by an initial fractionation of the digest by preparative paper chromatography (Whatman 3MM paper, solvent system as above) followed by further purification by high-voltage paper electrophoresis (pH 6.5). Peptides for mass spectrometry were obtained by using paper chromatography alone since it was not necessary for the samples to be completely pure.

Derivatization and Mass Spectrometry.—Dimethyl sulphoxide was dried over CaH_2 for 4 days and then distilled from CaH_2 under reduced pressure at 60 °C. A solution of methyl sulphinyl carbanion [11] was prepared by incubation of NaH (0.1 g) with dry dimethyl sulphoxide (2 ml) in an oven at 80 °C for 30 min. Peptides (*ca.* 1 μmol) were acetylated by incubation in MeOH–acetic anhydride (4 : 1 v/v; 0.5 ml) for 3 h at room temperature. Samples were dried *in vacuo* and then permethylated by incubation in dimethyl sulphoxide (*ca.* 0.2 ml) containing methyl sulphinyl carbanion solution (*ca.* 0.6 ml) and CD_3I (99 atom %, B.O.C.; *ca.* 0.2 ml). Reaction was allowed to proceed for 2 min [16] after which time water (1 ml) was added. Derivatized peptides were extracted from the reaction mixture with chloroform (1 ml) and the extract washed with water (2 × 1 ml). The extract was dried *in vacuo*.

Mass spectra of derivatized peptides were recorded using an AEI MS12 spectrometer at 70 eV over a range of source temperatures from 150—250 °C.

We thank the N.R.D.C. (J. M. H.) and the S.R.C. (A. J. G.) for financial support.

[7/1871 *Received*, 25*th October*, 1977]

[13] W. R. Gray, *Methods Enzymol.*, 1967, **11**, 469.
[14] B. S. Hartley, *Biochem. J.*, 1970, **119**, 805.
[15] P. Edman, *Ann. New York. Acad. Sci.*, 1960, **88**, 602.
[16] H. R. Morris, R. J. Dickinson, and D. H. Williams, *Biochem. Biophys. Res. Comm.*, 1973, **51**, 247.