Plasma-desorption mass spectrometry as an aid in protein sequence determination

Application of the method on a cuticular protein from the migratory locust (*Locusta migratoria*)

Klaus KLARSKOV,* Peter HØJRUP,* Svend Olav ANDERSEN† and Peter ROEPSTORFF*‡

*Department of Molecular Biology, University of Odense, DK-5230 Odense M, Denmark, and †Department of Biological Chemistry A, University of Copenhagen, DK-2100 Copenhagen Ø, Denmark

The complete amino acid sequence of a structural protein, protein 8, isolated from the pharate cuticle of the locust *Locusta migratoria* was determined. Protein 8 contains 148 amino acid residues and has an M_r of 15224. By the extensive use of information obtained by plasma-desorption mass spectrometry (p.d.m.s.) it was possible to reduce the need for conventional sequence determination and to improve the reliability of the results. On the basis of the determined M_r of the intact protein all the peptides that constitute the complete sequence could be isolated from a time-course enzymic digestion. The isolated peptides were sequenced by using a combination of Edman degradation and carboxypeptidase digestion monitored by p.d.m.s. The alignment of the peptides was established from the time-course digestion and further verified by a second enzymic digestion. The primary structure of the protein consists of two hydrophilic and two hydrophobic regions. The hydrophobic regions are enriched in alanine, valine and proline and dominated by a repetitive sequence Ala-Ala-Pro-(Ala/Val). The sequence strengthens the view that the cuticle proteins belong to a unique family of structural proteins.

INTRODUCTION

The cuticle of the migratory locust (*Locusta migratoria*) contains more than 100 different structural proteins, of which the majority are basic [1]. *N*-Terminal sequence determination and amino acid analysis of a number of the proteins covering a pI range of 6.4–10.6 and an M_r range of 15200–36800 indicate that these proteins belong to a new protein family [2]. Cysteine, methionine and tryptophan are absent from the proteins and the content of phenylalanine is very low [1,2]. The proteins investigated contain a short *N*-terminal part that is rich either in hydrophilic residues or in glycine, tyrosine and leucine. This region is followed by a region dominated by alanine. The complete amino acid sequence of one of the proteins, protein 38 (pI 10.2, M_r 15322), has previously been obtained by using a combination of Edman degradation and ^{252}Cf plasma-desorption mass spectrometry (p.d.m.s.) [3]. The protein investigated in the present study, protein 8, belongs to the group of slightly acidic proteins (pI 6.4) and contains an *N*-terminal region with some resemblance to that of protein 38 [2].

In conventional protein chemistry, sequence determination of a protein is carried out by sequencing all peptides obtained from a chemical or enzymic degradation of the protein. The order of the peptides in the protein is established by sequence determination of overlapping peptides obtained from a second degradation of the intact protein. Consequently, in order to establish the complete primary structure, the protein has to be examined at least twice.

By extensive use of the p.d.m.s. technique [4] it is

possible to reduce the need for conventional sequence determination. We have employed the M_r information obtained by this technique in the analysis of a cuticular protein, protein 8, isolated from locust cuticle.

The primary structure of protein 8 was established from a single enzymic digest by conventional Edman degradation in combination with M_r information and amino acid analysis. The *C*-terminal sequence of the peptides obtained from the digestion was determined by carboxypeptidase digestion in combination with p.d.m.s. The order of the peptides was established by time-course enzymic digestion monitored by p.d.m.s. and further confirmed by analysis of a second digest by p.d.m.s. The M_r values of the peptides combined with the knowledge of enzyme specificity were used in a computer program that, on the basis of the M_r , can search a linear protein sequence for all possible matches.

MATERIALS AND METHODS

Enzymes and special reagents

Carboxypeptidase Y and carboxypeptidase MII were obtained from Carlsberg Biotechnology, Copenhagen, Denmark. Carboxypeptidase Y was used as received (19 mg/ml in water), but carboxypeptidase MII, received as a 6 mg/ml solution in 200 mM-NaCl/50 mM-sodium acetate buffer, pH 4.0, was dialysed before use against 50 mM-ammonium acetate buffer, pH 3.8. Trypsin and carboxypeptidase B were gifts from Novo Industries, Bagsvaerd, Denmark. *Staphylococcus aureus* proteinase was from ICN Immunobiologics, Lisle, IL, U.S.A.

Abbreviation used: p.d.m.s., ^{252}Cf plasma-desorption time-of-flight mass spectrometry.

‡ To whom correspondence should be addressed.

These sequence data have been submitted to the EMBL/GenBank Data Libraries.

Nucleosil 300 (5 μm particle size) resin was from Macherey-Nagel, Duren, Germany. Spherosorb ODS2 (5 μm particle size) resin was from Phase Separations, Queensferry, Clwyd, U.K. Shandon Hypersil ODS2 (3 μm particle size) resin was from Shandon Southern Products, Runcorn, Cheshire, U.K. Vydac C₁₈ (10 μm particle size) was from The Separation Group, Hesperia, CA, U.S.A.

The water used was ultra-high-quality water from an Elgastat UHQ (Elga, High Wycombe, Bucks., U.K.).

Solvents and reagents used for h.p.l.c. and the Sequenator were from Rathburn Chemicals, Walkerburn, Borders, Scotland, U.K. The solvents were used as obtained.

Amino acid analysis

Samples (1–10 nmol) were freeze-dried in glass tubes (5 mm $\times$ 50 mm) and each was placed in a 22 ml screw-cap vial equipped with a Mininert valve. Approx. 0.5 ml of a 6 M-HCl/0.1 % phenol solution was placed in the bottom of each vial and the samples were hydrolysed under reduced pressure for 23 h at 110 °C. The amino acid analysis were performed on an LKB Alpha amino acid analyser with post-column formation of *o*-phthalaldehyde derivatives and fluorescence detection by Hans Frede Hansen, Rigshospitalet, Copenhagen, Denmark.

Purification of protein 8

Protein 8 was extracted from either wings or total cuticle. The extraction and primary purification procedure have been described elsewhere [2]. The protein was further purified by using h.p.l.c. equipment consisting of two Waters 6000 pumps, a Waters 680 gradient controller, a Waters 440 u.v. detector with an extended wavelength module and filters for 214 and 254 nm.

The column (8 mm $\times$ 120 mm) was packed with Vydac C₁₈ (10 μm particle size) resin and the protein was eluted with a linear gradient of 10–50 % (v/v) propanol-2-ol in 0.1 % (v/v) trifluoroacetic acid at 50 °C.

Time-course tryptic digestion

Analytical. A 200 μg portion of the purified protein was dissolved in 80 μl of 0.2 M-triethylamine and the pH was adjusted to 10.5 with HCl. Trypsin (2 μg ; enzyme/substrate ratio 1:100, w/w) was added and 16 μl portions were withdrawn at 5, 15, 30, 60, 120 and 180 min. The digestion was terminated by addition of 3 μg of soya-bean trypsin inhibitor.

Preparative tryptic digestion. A 5–10 mg portion of the protein was digested as described above and the reaction was terminated after 15 min.

Digestion with *S. aureus* proteinase

A 5–10 mg portion of the protein was digested at 37 °C with *S. aureus* proteinase (enzyme/substrate ratio 1:10, w/w) in 100 mM-ammonium acetate adjusted to pH 4.0 with acetic acid. The digestion was terminated by freezing the solution. Subdigestion of peptide T4 was carried out in 100 mM-ammonium bicarbonate buffer adjusted to pH 7.8 with acetic acid. Then 50–100 pmol of sample was withdrawn and analysed by p.d.m.s.

Purification of peptides from the enzymic digest

Peptides derived from the trypsin and *S. aureus*-proteinase digestions were separated by using the same

h.p.l.c. equipment as used for the purification of protein 8 (see above).

The tryptic-digest peptides were separated initially on a column (4 mm $\times$ 250 mm) packed with Nucleosil 300 (10 μm particle size) with a gradient of 2–50 % (v/v) acetonitrile in 0.1 % (v/v) trifluoroacetic acid. The fractions that, as judged from the p.d.m.s. spectra, were not pure, were further purified on a column (4 mm $\times$ 250 mm) packed with either Nucleosil 100 (10 μm particle size) or Spherosorb ODS2 (5 μm particle size) with appropriate gradients of propan-2-ol in 0.1 % (v/v) trifluoroacetic acid.

The peptides obtained from the *S. aureus*-proteinase digestion were separated similarly to the tryptic-digest peptides except for the use of a linear gradient of 5–18 % (v/v) propan-2-ol in 0.1 % (v/v) trifluoroacetic acid.

Determination of amino acid sequences

Automatic Edman degradation. Tryptic-digest peptides were sequenced on a Beckman 890 B/C sequencer. A 25–75 nmol portion of each sample was applied to the cup coated with 2.5 mg of Polybrene. The degradations were performed with a 0.25 M-Quadrol buffer. The extraction and conversion have been described elsewhere [5].

Manual sequencing. Manual sequencing of peptides was performed by Edman degradation in the gas phase as described by Brandt & Frank [6]. A 1–10 nmol portion of each peptide dissolved in 0.1 % (v/v) trifluoroacetic acid was applied to a glass filter coated with 75 μg of Polybrene.

Identification of phenylthiohydantoin derivatives. The amino acid phenylthiohydantoin derivatives were identified on Kontron h.p.l.c. equipment with a column (4 mm $\times$ 120 mm) packed with Shandon Hypersil ODS (3 μm particle size) resin. The derivatives were eluted with a gradient of 14–42 % (v/v) acetonitrile in 28 mM-sodium acetate buffer, pH 3.9–4.0, containing 5 % (v/v) tetrahydrofuran.

M.s.

The mass spectra were recorded on a Bio Ion 10k plasma-desorption mass spectrometer (Bio Ion AB, Uppsala, Sweden). The spectra were accumulated for 0.5×10^6 – 1×10^6 fission events corresponding to approx. 7–15 min. Further details have been described elsewhere [7].

Sample preparation for m.s.

The samples were analysed by using nitrocellulose-prepared targets [8] produced by electrospraying 100–200 μl of a nitrocellulose solution in acetone (2 $\mu\text{g}/\mu\text{l}$) on to a Mylar foil (1 cm diameter).

The samples (50–300 pmol) were dissolved in 12.5 % (v/v) acetonitrile in 0.1 % (v/v) trifluoroacetic acid. A 2 μl portion of the solution was applied to the nitrocellulose-prepared target. The excess of solvent was either left to dry at room temperature or dried by using the spin-drying technique [9].

Enzyme digestion *in situ*

Enzyme digestion *in situ* was carried out by applying 2 μl of a 1 $\mu\text{g}/\mu\text{l}$ solution of enzyme in either 50–100 mM-ammonium acetate or ammonium bicarbonate adjusted to an appropriate pH with acetic acid. The digestion was

terminated by letting the enzyme solution dry followed by washing the target with 10–50 μ l of 0.1% (v/v) trifluoroacetic acid.

Nomenclature of peptides

The peptides isolated from the enzymic digests of protein 8 have a prefix indicating the enzyme used: T for trypsin and A for *S. aureus* proteinase.

RESULTS

Determination of the M_r of protein 8

The mass spectrum obtained from the protein adsorbed on nitrocellulose is shown in Fig. 1. The single-charged molecular ion is absent but the double-charged and triple-charged ions are readily identified, a feature often observed for proteins desorbed from nitrocellulose [8]. By taking the average of the M_r values as calculated on the basis of the double-charged (M_r 15222) and triple-charged (M_r 15240) ions the M_r of the protein was determined to be 15231 ± 15 .

N-Terminal sequence

The N-terminal 56 amino acid residues of protein 8 (except for residue 54) have previously been determined by automatic Edman degradation of the intact protein [2]. The identity of 55 of the 56 N-terminal residues was verified, and in addition residue 54 was shown to be glutamic acid.

Time-course tryptic digestion

The purpose of performing a time-course tryptic digestion of the protein was to obtain large peptides suitable for Edman degradation and to determine the linear order of these peptides.

The peptides were separated by h.p.l.c. (Fig. 2) and the M_r values of each fraction were determined by p.d.m.s. (Table 1). The sum of the M_r values of the peptides T1, T2, T3, T4, T5 and T6 minus $5H_2O$ is 15235.4. This value corresponds within the experimental error to the M_r determined by p.d.m.s. of the intact protein. As the experimental error is much smaller than the mass of the smallest amino acid residues the peptides of T1–T6 must constitute the complete sequence of protein 8.

Fig. 1. Positive-ion p.d.m.s. mass spectrum of protein 8

The spectrum is background-subtracted. Counts/bin is an absolute measure of ion abundance for a given mass increment called a bin.

Fig. 2. H.p.l.c. separation of peptides derived from a time-course tryptic digestion of protein 8 (a) after 5 min, (b) after 15 min and (c) after 180 min

From the time-course analysis shown in Fig. 2 it can be seen that the three tryptic-digest peptides T2, T4 and T4.1 + T4.2 appeared and disappeared during the digestion. The peptide T2 (M_r 1980.1) can be shown to be degraded to two smaller tryptic-digest peptides T2.1 and T2.2 with M_r values 633 and 1369 respectively. The amount of peptides T2.1 and T2.2 increased during the digestion whereas the amount of T2 decreased after an initial increase. The tryptic degradation pattern of the peptide T4 was determined by performing a tryptic digestion of the isolated peptide directly on a nitrocellulose-prepared target as described previously [10]. The p.d.m.s. spectrum of the peptide is shown in Fig. 3(a) (before the tryptic digestion) and Fig. 3(b) (after the

Table 1. M_r values of tryptic-digest peptides as determined by p.d.m.s.

Peptide	Calculated M_r	Measured M_r
T4 + T2	5654.1	5652.8
T1 + T5	4876.4	4874.8
T4.1	1143.3	1144.7
T4.2	1866.0	1866.4
T4.3	717.7	N.D.*
T2.1	631.7	633.0
T2.2	1367.4	1369.0
T1	2072.2	2073.4
T2	1981.1	1983.4
T3	1468.8	1469.8
T4	3691.0	3693.0
T5	2822.3	2822.8
T6	3278.6	3280.0
Sum of T1, T2, T3, T4, T5 and T6 - 5 × H ₂ O	15325.4 90	15325.4 90
Total M_r determined by p.d.m.s.	15223.9	15235.4 15230.8

* Not determined.

Fig. 3. P.d.m.s. mass spectra of (a) peptide T4 before addition of enzyme and (b) of peptide T4 after tryptic digestion *in situ* in 100 mM-ammonium bicarbonate buffer, pH 7.8, for 20 min at room temperature.

tryptic digestion *in situ*). As can be seen from the spectrum, the digestion resulted in mainly two peptides T4.1 and T4.2 with M_r values 1144.7 and 1866.4 respectively. It was concluded that the tryptic-digest peptide T4 is gradually degraded during the tryptic digestion of the intact protein peptide to T4.1 + T4.2 and a small peptide T4.3 that could be found neither in the h.p.l.c. separation nor in the p.d.m.s. spectra.

The best yield of peptides T1–T6, and in particular peptide T4, was obtained after 15 min digestion. This digestion time was used for a preparative digest.

Determination of the sequences of the tryptic-digest peptides

Peptide T6 contains the *N*-terminal 33 residues, as confirmed by M_r (Table 1), amino acid composition (Table 2) and the products from 12 automatic Edman cycles. Peptide T3 covers the following 16 residues, as confirmed by M_r , amino acid composition and the products from 13 automatic Edman degradations.

In the present system used for automatic Edman degradation, it was only possible to identify histidine and arginine residues if they occurred close to the *N*-terminus of the peptide. This is probably due to a low yield in the extraction with chlorobutane in the Sequenator and a broad peak shape in the h.p.l.c. analysis. The largest tryptic peptide T4 contains 33 residues and covers the sequence residues 50–82. From the automatic Edman degradations it was not possible to determine the histidine residues in positions 68 and 71 and the *C*-terminal arginine residues. The glutamine, serine and glutamic acid residues in positions 78, 79 and 80 were only tentatively determined.

The result from a carboxypeptidase MII digestion *in situ* is shown in Fig. 4. The alanine residue in position 81 is confirmed and the presence of a serine residue in position 79 is established. The *C*-terminal arginine residue and the glutamic acid residue in position 80 were only tentatively determined by this method because of an insufficient precision in the mass determination. The *C*-terminal arginine residue was determined by amino acid analysis of peptide T4 and manual gas-phase Edman degradation of the peptide, which covers the residues 81–103, obtained from an *S. aureus*-proteinase digestion of protein 8.

A time-course *S. aureus*-proteinase digestion of peptide T4 resulted in the degradation pattern shown in Fig. 5. The peptide was hydrolysed to two peptides with M_r values 1437.8 and 2275.9 corresponding to the peptides Ala⁵⁰–Glu⁶² and Phe⁶³–Arg⁸². The peptide Phe⁶³–Arg⁸² was further degraded to Phe⁶³–Glu⁸⁰ and the *C*-terminal dipeptide Ala–Arg, thereby confirming the presence of a glutamic acid residue in position 80. After approx. 5 h of *S. aureus*-proteinase digestion two peptides with M_r values 852.6 and 1216.2 covering the sequences Phe⁶³–Asp⁶⁹ and Ala⁷⁰–Glu⁸⁰ appeared. The peptides were isolated by h.p.l.c. and the peptide Ala⁷⁰–Glu⁸⁰ was submitted to a tryptic digestion followed by manual gas-phase Edman degradation of the mixture. The histidine residue in position 71 and the glutamine residue in position 78 were determined from the manual Edman cycles. The histidine residue in position 68 was obtained from the difference between the M_r of the known sequence and the M_r determined by p.d.m.s. of the peptide Phe⁶³–Asp⁶⁹.

The remaining three peptides T2, T1 and T5 cover the

Table 2. Amino acid compositions of tryptic-digest peptides used for elucidation of the structure of protein 8, the amino acid composition of which is included

Analytical values are molar proportions without correction for destruction, slow release or impurities. Values below 0.2 are omitted. Proline residues were not determined (N.D.). The numbers in parentheses are values obtained from sequence analysis.

Amino acid	Composition (mol of residues/mol)						Protein 8
	Peptide T1	Peptide T2	Peptide T3	Peptide T4	Peptide T5	Peptide T6	
Asx	2.7 (4)	2.7 (3)	—	3.2 (3)	—	2.3 (2)	11.9 (12)
Glx	2.1 (2)	1.1 (1)	—	5.2 (5)	—	2.4 (2)	11.0 (10)
Ser	—	1.8 (3)	—	0.8 (1)	—	2.1 (1)	5.6 (5)
Gly	0.8 (1)	2.6 (3)	—	1.8 (2)	0.8 (1)	2.6 (3)	8.9 (10)
His	1.9 (2)	—	—	2.0 (2)	3.8 (4)	0.9 (1)	8.9 (9)
Arg	1.0 (1)	1.1 (1)	—	1.5 (1)	—	—	1.8 (3)
Thr	0.8 (1)	0.8 (1)	—	1.0 (1)	1.9 (2)	—	4.6 (5)
Ala	1.8 (2)	—	7.8 (8)	5.1 (5)	7.8 (8)	9.9 (11)	31.8 (35)
Pro	N.D. (0)	N.D. (1)	N.D. (2)	N.D. (2)	N.D. (5)	N.D. (3)	N.D. (13)
Tyr	1.0 (1)	1.0 (1)	1.1 (1)	3.0 (3)	1.0 (1)	4.0 (4)	11.0 (11)
Val	2.3 (3)	2.7 (3)	4.1 (4)	4.2 (4)	5.7 (6)	3.4 (3)	21.7 (23)
Met	—	—	—	—	—	—	—
Cys	—	—	—	—	—	—	—
Ile	—	—	—	1.0 (1)	0.9 (1)	1.0 (1)	2.9 (3)
Leu	—	1.0 (1)	—	—	—	1.0 (1)	2.1 (2)
Phe	1.0 (1)	—	—	1.0 (1)	—	—	2.1 (2)
Lys	1.0 (1)	1.1 (1)	1.0 (1)	2.0 (2)	—	0.9 (1)	5.4 (5)

Fig. 4. P.d.m.s. mass spectrum of peptide T4 after carboxypeptidase MI digestion in 50 mM-ammonium acetate buffer, pH 4.1, for 20 min at room temperature

sequence residues 83–148. Peptide T5 was concluded to be the C-terminal peptide because it was the only tryptic peptide recovered in major amounts that did not contain a C-terminal lysine or arginine residue.

Except for the histidine residues in positions 110 and 142, the sequence of all three peptides could be obtained by automatic Edman degradation and carboxypeptidase digestion monitored by p.d.m.s. (Fig. 6). The histidine residues in positions 110 and 142 were established from the difference between the M_r determined by p.d.m.s. of the corresponding peptide and the calculated M_r of the known amino acid sequences. They were further verified by amino acid analysis (Table 2).

Alignment of tryptic-digest peptides

The order of two of the tryptic-digest peptides was

Fig. 5. Cleavage pattern obtained by time-course digestion of peptide T4 with *S. aureus* proteinase

known from the previously determined N-terminal sequence of the intact protein [2]. The sequence of the N-terminal 56 residues was furthermore sufficient to link peptide T4 to peptide T3 because peptide T4 is the only peptide containing an N-terminal identical with the last five residues determined by N-terminal sequencing of the intact protein.

From the time-course tryptic digestion of protein 8

Fig. 6. Primary structure of protein 8 from the cuticle of *Locusta migratoria* and the peptides used to construct the amino acid sequence

►, Sequence determined by Edman degradation; ◄, sequence determined by carboxypeptidase digestion monitored by p.d.m.s. Dotted half-arrowheads are used when the amino acid residues were only tentatively identified.

two peptides with M_r values 5652.8 and 4874.8 determined by p.d.m.s. were found (Table 1). The peptides could, on the basis of M_r information, the specificity of trypsin and a computer program that can search an amino acid sequence for possible matches, be assigned to cover the sequences 50–101 and 102–148 with calculated M_r values of 5654.1 and 4876.4 respectively. The assignment of the peptide (M_r 5652.8) was further confirmed by obtaining the sequence of the four *N*-terminal residues by manual Edman degradation. Thus the peptide consists of peptide T4 followed by peptide T2. The peptide (M_r 4874.8) corresponds to the tryptic-digest peptide T1 covering the sequence 102–119 and the *C*-terminal peptide T5. No further verification of the peptide was performed because it was only found in a minor amount from the time-course analysis.

In order to verify that the primary alignment of the sequenced tryptic-digest peptides was correct, a second enzymic digestion of protein 8 was carried out. On the basis of the sequence, relatively large fragments could be expected if the digestion was performed with *S. aureus* proteinase at pH 4.0. Large peptides are preferred because of the lesser complexity and decreased probability of peptides being lost either on h.p.l.c. or when the sample is analysed by p.d.m.s.

H.p.l.c. separation of the peptides obtained from the *S. aureus*-proteinase digestion is shown in Fig. 7. The M_r values of the peptides determined by p.d.m.s. are shown in Table 3. The two *N*-terminal dipeptides Ile-Glu (A1) and Asp-Glu (A2) were not found, most probably

Fig. 7. H.p.l.c. separation of peptides derived from an *S. aureus*-proteinase digestion of protein 8

... The nomenclature refers to the text and Table 3.

Table 3. M_r values of *S. aureus*-proteinase-digest peptides as determined by p.d.m.s.

Peptide	Calculated M_r	Measured M_r
A3.1	1208.3	1208.1
A3.2	1135.3	1136.2
A3.2.2	590.7	591.5
A3.3	1468.8	1468.0
A3.4	1902.2	1901.4
A4+A5	4565.8	4565.9
A6	4547.1	4545.1

Fig. 8. P.d.m.s. spectra of (a) peptide A6 before digestion with trypsin and (b) of peptide A6 after tryptic digestion *in situ* for approx. 20 min

The experimental conditions are as described for Fig. 3(b).

because they are eluted in the solvent front of the h.p.l.c. However, the *N*-terminal sequence of protein 8 is well established from the Edman degradation of the intact protein and of the tryptic peptide T6.

A large *S. aureus*-proteinase-digest fragment (A3) covering the sequence residues 5–62 (M_r 5660.5) was expected from the sequence of the protein and the primary alignment of the tryptic peptides. The peptide was not found, but four minor fragments that by addition of their determined M_r values resulted in an M_r corresponding to that of the predicted peptide A3 were obtained. Two of the smaller peptides A3.3 and A3.4 covering the sequence residues 29–44 (M_r 1468.8) and 45–62 (M_r 1902.2) confirm the order of the tryptic-digest

peptides T6, T3 and T4. The sequence Tyr-Ala-Ala-Pro seems furthermore to be a specific cleavage site for *S. aureus* proteinase under the experimental conditions used in this study, with cleavage occurring between the two alanine residues.

The alignment of the tryptic-digest peptides T2 and T1 was confirmed by two larger peptides (M_r values 4565.8 and 4547) from the *S. aureus*-proteinase digestion that covers the sequence residues 63–104 and 105–148 respectively.

All the peptides obtained from the *S. aureus*-proteinase digestion of protein 8 were submitted to one to four manual Edman degradations. These were carried out to confirm their proposed position based on the M_r information and the computer program mentioned above. The position of peptide A4+A5, which confirms the alignment of the tryptic peptides T2 and T1, was furthermore established from an *S. aureus*-proteinase digestion at pH 7.8. The digestion resulted in the two expected fragments A4 and A5, as revealed from the M_r information obtained by p.d.m.s. Only one manual Edman degradation cycle was carried out on peptide A6, but since a tryptic digestion *in situ* resulted in a peptide (M_r 1742.6) corresponding to the sequence residues 105–119 and the *C*-terminal tryptic-digest peptide T5, the position of peptide A6 was unambiguously confirmed (Fig. 8).

DISCUSSION

In the sequence determination of a protein, protein 8, from *Locusta migratoria* a novel sequence strategy has been applied. The central idea is to employ M_r determination by m.s. as much as possible in order to reduce the need for conventional sequence determination.

The starting point is determination of the M_r of the intact protein. Although mass determination by p.d.m.s. cannot be carried out with unit mass resolution in this mass range, the experimental error is considerably smaller than the M_r of the smallest amino acid residue (glycine, M_r 57.0518). This precision thus eliminates the possibility of missing part of the sequence in the final structure elucidation.

Protein 8 belongs to the group of slightly acidic cuticular proteins, as revealed by gel electrophoresis [1]. The protein was found to contain 148 amino acid residues and has an isotopically averaged mass of 15224, which is in good agreement with the M_r determined by p.d.m.s. (M_r 15230 $\pm$ 15). The M_r obtained here is considerably lower than that previously estimated from SDS/polyacrylamide-gel electrophoresis (M_r 19700) [2]. The amino acid composition as determined by amino acid analysis is shown in Table 2 and is in agreement with the composition derived from the sequence as shown in Fig. 6. The content of proline was not determined here, but the content as derived from Fig. 6 is in accordance with what has previously been reported [2].

On the basis of only the M_r information, it was possible to verify that all the peptides constituting the protein were isolated from a single time-course digestion. It was furthermore possible to determine the order of the peptides, thus identifying the complete sequence of protein 8 from a single enzymic digest. The final alignment of the sequence was verified by peptides obtained from an *S. aureus*-proteinase digestion of the protein. These peptides were rapidly identified by p.d.m.s. in