Primary structure of a structural protein from the cuticle of the migratory locust, *Locusta migratoria*

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 isolated from pharate cuticle of the locust *Locusta migratoria* was determined. The protein has an unusual amino acid composition: 42% of the residues are alanine and only 14 of the 20 common amino acid residues are present. The primary structure consists of regions enriched in particular amino acid residues. The *N*-terminal region and a region close to the *C*-terminus are enriched in glycine. The rest of the protein is dominated by alanine, except for two short regions enriched in hydrophilic residues. Almost all the proline residues are situated in the alanine-rich regions in a conserved sequence 'A-A-P-A/V'. An internal duplication has taken place covering most of the protein except for the glycine-rich regions. Owing to the unusual features of the protein a combination of automated Edman degradations and plasma-desorption m.s. was used to determine the complete sequence. The protein does not show sequence homology to other proteins, but proteins divided into regions enriched in the same kind of amino acid residues have been isolated from other insect structures.

INTRODUCTION

Insect cuticle is composed of two main components, chitin and protein. The chitin is present in the form of long chains of α -chitin, bound firmly together by hydrogen bonds to form long fibres with a diameter of approx. 2.8 nm (28 Å). The fibres are always laid down in a matrix of proteins in a helicoidal or unidirectional manner (Neville, 1975).

Most adult insect cuticle is stabilized by sclerotization, a process in which a phenolic compound [*N*-acetyl-dopamine (*N*-acetyl-2,4-hydroxyphenylalanine) or β -alanyldopamine] is secreted into the integument, where polymerization and reaction with proteins take place after enzymic oxidation. This makes the proteins difficult to extract, presumably owing to extensive cross-linking with the phenols (Andersen, 1985).

As it is difficult to solubilize the proteins present in adult cuticle, most characterization of cuticle proteins has been carried out on unsclerotized larval and pupal cuticle (Fristrom *et al.*, 1978; Lipke & Henzel, 1981; Willis *et al.*, 1981; Silvert *et al.*, 1984). Mainly M_r values, isoelectric points and amino acid compositions have been determined, but from *Drosophila* a number of cuticle proteins from the late third-instar larva have been sequenced (Snyder *et al.*, 1982). All the proteins isolated from soft larval cuticle have low isoelectric points (pI below 6), have low M_r values (8000–25000) and are rich in polar amino acid residues.

In the migratory locust sclerotization takes place soon after the adult insect has emerged from the old nymphal cuticle. In locusts caught while emerging from the old nymphal cuticle (pharate locusts), 90% of the proteins in the cuticle (which accounts for 70% of the dry weight) can be extracted (Andersen & Barrett, 1971). The non-extractable proteins are intimately linked to the chitinous fibres and little is known about this part. The extractable

proteins from the pharate locust cuticle have turned out to be different from the proteins isolated from soft larval and pupal cuticle (Andersen, 1979). The proteins destined to become sclerotized are rather basic (pI from 6 to 11), are dominated by small non-polar amino acid residues and have slightly higher M_r values than the proteins isolated from soft larval cuticle.

In preparations of larval or pupal cuticles the number of proteins can be low; in the *Drosophila* third-instar larval cuticle only five major proteins have been reported (Snyder *et al.*, 1982), whereas about 100 proteins can be seen in two-dimensional gels of *Locusta* cuticle proteins (Andersen *et al.*, 1986). The diversity of proteins in locust cuticle is presumably due neither to polymorphism nor to secondary modifications, as most of the proteins appear to be separately coded entities (Højrup *et al.*, 1986).

Sequence studies on proteins isolated from hard cuticle have so far only been concerned with *N*-terminal sequences (Højrup *et al.*, 1986). Below we report the first complete sequence of a protein from hard cuticle.

MATERIALS AND METHODS

Enzymes and special reagents

Chymotrypsin (EC 3.4.21.1) was from Merck, Darmstadt, Germany. Trypsin (EC 3.4.21.4) was a gift from Novo, Bagsværd, Denmark. Vydac C18 resin was from The Separation Group, Hesperia, CA, U.S.A. Nucleosil ODS (10 μ m particle size) resin was from Macherey-Nagel, Düren, Germany. Spherosorb ODS2 (5 μ m particle size) was from Phase Separations, Deeside, U.K.

Amino acid analysis

Peptide and protein samples (0.5–2 nmol) were hydrolysed in 6 M-HCl containing 0.1% phenol (110 °C

† To whom correspondence should be addressed.

for 18–24 h in sealed tubes under reduced pressure). After being freeze-dried the hydrolysates were dissolved in 0.06 M-citrate buffer, pH 2.20, and portions (one-half to one-twentieth) were loaded on to a Beckman 121MB amino acid analyser equipped with a 0.28 cm × 16 cm column packed with LA/28/10 (8 µm particle size) ion-exchange resin (Locarte Co., London, U.K.). The results for the whole protein represent the average of three separate hydrolysates, and the result for the peptides represent single-sample analysis. No corrections were made for hydrolytic loss or incomplete hydrolysis.

Purification of protein no. 38

Cuticle was obtained from a laboratory stock of migratory locusts sampled while moulting to the adult stage. The extraction and purification of the protein were carried out as described by Andersen (1980) and Højrup *et al.* (1986).

Digestion with trypsin

A 10 mg portion of purified protein no. 38 was dissolved in 2 ml of 0.2 M-triethylamine and the pH adjusted to 10.7 with HCl. Trypsin (0.5 mg) dissolved in 0.01 M-HCl was added and the digestion was allowed to proceed at 37 °C for 2 h until terminated by the addition of 0.5 mg of soya-bean trypsin inhibitor.

Digestion with chymotrypsin

A 10 mg portion of protein was dissolved in 1 ml of 4 M-urea/0.1 M-NH₄HCO₃, pH 8.3. Chymotrypsin (0.5 mg) was dissolved in an equal volume of 0.1 M-NH₄HCO₃, pH 8.3, and added under stirring to give a final enzyme/substrate ratio of 1:20. The reaction was carried out at 37 °C for 2 h and stopped by acidification and separation on h.p.l.c.

Purification of peptides

All peptide separations were carried out with a gradient h.p.l.c. equipment consisting of two Waters 6000 pumps, a Waters 680 gradient controller, a Rheodyne injector, a Waters 440 u.v. detector with Extended Wavelength Module and filters for 214 nm and 254 nm. The column was placed in a heated air oven thermostatically controlled at 50 °C.

After digestion, the peptides were separated initially on a column (8 mm × 120 mm) packed with Vydac C18 (10 µm particle size) resin, with a gradient of 5–50% (v/v) propan-2-ol in 0.1% trifluoroacetic acid, pH 1.9. Further separation of the isolated pools was carried out either on the same column or by using a 4 mm × 250 mm column packed with Nucleosil ODS (10 µm particle size) material with appropriate gradients of propan-2-ol in either 0.1% trifluoroacetic acid or 0.1 M-triethylammonium acetate, pH 5.5.

N-Terminal analysis

N-Terminal amino acids were identified manually by two-dimensional t.l.c. analysis of dansyl (5-dimethylaminonaphthalene-1-sulphonyl) derivatives on polyamide sheets (Gray, 1967).

Determination of M_r by m.s.

The M_r values of large peptides obtained by enzymic digestion were determined by m.s. on a Bio-Ion Bin 10K ²⁵²Cf-plasma-desorption time-of-flight mass spectrometer. The peptides (1–2 nmol) were dissolved in 5 µl of

Table 1. Amino acid composition of protein no. 38 from the cuticle of *Locusta migratoria*

Experimental values are means for three HCl hydrolyses.
—, Below detection limit.

Amino acid	Composition (mol of residue/mol)	
	Acid hydrolysis	Sequence analysis
Asx	2.1	2*
Thr	4.0	4
Ser	3.5	4
Glx	2.4	2*
Pro	16.9	17
Gly	11.1	12
Ala	64.0	69
Cys	—	0
Val	16.6	16
Met	—	0
Ile	7.0	7
Leu	6.4	6
Tyr	15.0	15
Phe	—	0
Lys	1.0	1
His	3.0	3
Trp	—	0
Arg	5.0	5
M_r		15320.6

* All found in the amidated form.

acetic acid and electrosprayed on to a 1 µm-thick aluminium foil target, and the spectra were recorded on the mass spectrometer with an accelerating voltage of 16 kV and a resolution of 2 µs. For a detailed description of the instrumentation and the principles see Sundqvist *et al.* (1984).

Determination of amino acid sequences

Peptides were sequenced either on a Beckman 890C sequencer (as described by Sottrup-Jensen *et al.*, 1984) or on an Applied Biosystems gas-phase Sequenator with chemicals and program (MHTFA1) as recommended by the manufacturer. In the first case 10–20 nmol of peptide was applied to the sequencer, and the amino acid phenylthiohydantoin derivatives were identified on a Hewlett-Packard HP 1084B h.p.l.c. instrument with a 4 mm × 250 mm column packed with Spherosorb ODS2 (5 µm particle size) and eluted with a gradient of ethanol in 33 mM-sodium acetate buffer, pH 5.20 (Sottrup-Jensen *et al.*, 1980). In the second case 1–2 nmol of peptide was used for each sequence run, and the amino acid phenylthiohydantoin derivatives were identified on an HP 1084B instrument with an IBM cyano column (4 cm × 250 mm) and a gradient of acetonitrile in 35 mM-sodium acetate buffer, pH 5.2, containing 1% (v/v) tetrahydrofuran (Hunkapiller & Hood, 1983; Esch, 1984).

Nomenclature

The peptides isolated from digests of protein no. 38 have a prefix indicating the enzyme used (T for trypsin, C for chymotrypsin) followed by a number. The numbers of the peptides follow their positions in the protein counted from the N-terminus of the intact protein.

Fig. 1. Initial fractionations of a tryptic digest of protein no. 38 (a) and a chymotryptic digest of protein no. 38 (b) by reverse-phase h.p.l.c.

The 8 mm $\times$ 120 mm column was packed with Vydac RP18 (10 μ m particle size) material. The peptides were eluted with a gradient of propan-2-ol in 0.1% trifluoroacetic acid (-----). The peaks containing the indicated peptides were collected and further purified or sequenced directly as mentioned in the Results section.

RESULTS

N-Terminal sequence

The *N*-terminal sequence of the protein was derived from automatic Edman degradations of the intact protein. There were 79 cycles of degradation carried out and the product of most cycles was unambiguously identified. The arginine residues in steps 47 and 53 were not seen owing to low recovery of the basic amino acid residues, but evidence for these residues was found in purified peptides. From step 55 we were unable to determine the positions containing alanine residues owing to a rising background of this residue. Their presence was, however, established in the sequence of the tryptic peptide T3. All other residues were identified up to and including Ile-79.

Tryptic peptides

According to the amino acid composition (Table 1), protein no. 38 contains five arginine residues and one

lysine residue, and should thus give rise to seven tryptic peptides. Because of low solubility of the protein in the pH range 5–10 the tryptic digest was carried out at pH 10.7, which is close to the limit for enzymic activity of trypsin (Greene & Bartelt, 1977).

Separation of the digest on h.p.l.c. (Fig. 1a) revealed only five peaks, four containing a single peptide sufficiently pure for direct sequencing, and one peak upon rechromatography being shown to contain two peptides, one the C-terminal fragment and the other a hexapeptide that occurs twice in the sequence. The amino acid compositions of all tryptic peptides are shown in Table 2.

Peptide T1 contains the first 47 residues of the protein, and as the sequence was known from the *N*-terminal sequence run it was only checked for amino acid composition, *N*-terminal residue and M_r . Peptide T2 occurs twice in the sequence, as residues 48–53 and residues 115–120. The verification for the double occurrence was obtained from the chymotryptic digest.

Table 2. Amino acid compositions of peptides used for elucidation of the structure of protein no. 38

Analytical values are molar proportions without correction for destruction, slow release or impurities. Values less than 0.2 are omitted. Numbers in parentheses are values obtained from sequence analysis. The peptides are obtained from tryptic (T) and chymotryptic (C) digests.

Amino acid	Composition (mol of residue/mol)									
	T1	T2	T3	T4	T5	T6	C1	C2	C3	C4
Asx	1.08	—	—	0.91 (1)	0.39	—	—	—	—	—
Thr	0.89 (1)	0.95 (1)	—	1.00 (1)	—	—	0.92 (1)	—	0.84 (1)	—
Ser	0.97 (1)	0.88 (1)	—	1.00 (1)	0.56	—	0.86 (1)	—	0.82 (1)	—
Glx	—	1.00 (1)	—	—	0.39	—	1.00 (1)	—	0.89 (1)	—
Pro	4.81 (5)	—	4.81 (5)	1.92 (2)	5.00 (5)	—	—	—	—	0.86 (1)
Gly	4.46 (5)	—	2.28 (2)	—	4.86 (5)	—	0.38	—	—	1.07 (1)
Ala	15.69 (19)	0.98 (1)	16.23 (19)	11.04 (12)	15.00 (17)	—	1.08 (1)	1.21 (1)	1.32 (1)	3.00 (3)
Cys	—	—	—	—	—	—	—	—	—	—
Val	5.91 (6)	—	2.15 (2)	4.01 (4)	3.40 (3)	1.16 (1)	0.90 (1)	—	—	2.23 (2)
Met	—	—	—	—	—	—	—	—	—	—
Ile	1.93 (2)	0.97 (1)	1.66 (2)	—	1.13 (1)	—	0.83 (1)	—	0.94 (1)	—
Leu	1.19 (1)	—	2.00 (2)	—	2.92 (3)	—	1.05 (1)	—	2.00 (2)	—
Tyr	3.70 (4)	—	3.08 (3)	1.71 (2)	4.67 (5)	1.00 (1)	—	0.93 (1)	—	0.86 (1)
Phe	—	—	—	—	—	—	—	—	—	—
Lys	—	—	—	—	0.65 (1)	—	—	—	—	0.94 (1)
His	0.91 (1)	—	—	0.91 (1)	1.05 (1)	—	—	—	—	—
Trp	—	—	—	—	—	—	—	—	—	—
Arg	0.88 (1)	1.08 (1)	0.84 (1)	1.02 (1)	—	—	1.71 (2)	1.00 (1)	1.77 (2)	—

Table 3. M_r values of tryptic peptides determined by plasma-desorption time-of-flight m.s.

Peptide	Calculated mass	Measured mass	Ion identified
T1	4324.0	4347.6	$(M + Na)^+$
T2	674.8	676.5	$(M + H)^+$
		698.0	$(M + Na)^+$
T3	3264.8	3311.4	$(M - H + 2Na)^+$
T4	2383.7	2428.7	$(M - H + 2Na)^+$
T5	3828.4	3850.7	$(M + Na)^+$
T6	280.0	302.8	$(M + Na)^+$

Peptide T3 contains 36 residues and covers the sequence residues 54–89, which were all identified. Peptide T4 contains 24 residues; during sequencing the C-terminal arginine residue was only tentatively identified, but ^{252}Cf -plasma-desorption m.s. gave an M_r consistent with the 23 identified residues plus a C-terminal arginine residue (Table 3). Peptide T5 is located after the second occurrence of peptide T2, and containing 41 residues it covers residues 121–161. Peptide T6 is the dipeptide Val-Tyr and was the only peptide recovered in major amounts that did not contain arginine or lysine, and it is thus concluded to be the C-terminal peptide.

Chymotryptic peptides

Digestion of protein no. 38 with chymotrypsin was performed in 2 M-urea in order to overcome the low solubility of the protein. As the number of potential cleavage sites for chymotrypsin is much higher than for

trypsin, the h.p.l.c. separation was much more complex (Fig. 1b). The digest was scanned for peptides containing arginine and lysine by making amino acid analysis on each peak isolated.

Only three peaks containing arginine (C1–C3) and one containing lysine (C4) were found (Fig. 1b). Peptides C2 and C4 did not need further purification, whereas peptides C1 and C3 were purified to homogeneity by reverse-phase h.p.l.c. in a column packed with Nucleosil ODS and with a gradient of propan-2-ol in 0.1% trifluoroacetic acid. The amino acid compositions of the chymotryptic peptides shown in Fig. 3 are included in Table 2.

Sequencing of peptide C1 showed that it covers residues 47–55, thus connecting the tryptic peptides T1, T2 and T3. As all the residues except for the arginine residues were determined during the 79 cycles carried out from the N-terminal sequence run, the position of the peptide is unambiguous. Arg-89 is situated with two tyrosine residues close by, with the result that only a one-residue overlap is possible. The tripeptide C2 was the only peptide found overlapping this arginine residue. Peptide C3 covers residues 114–122 and, although the ultimate leucine was only obtained from the amino acid composition, the position of the peptide in the sequence is unambiguous. The last chymotryptic peptide covers residues 155–163, thus linking peptide T5 to C-terminal peptide T6.

M.s.

When the large tryptic fragments were being sequenced it was difficult to determine the ultimate basic residues. In order to verify the proposed sequence the M_r values were determined for all the tryptic fragments by using plasma-desorption time-of-flight m.s. (Fig. 2 and Table

Fig. 2. Positive-ion plasma-desorption time-of-flight mass spectrum of the molecular-ion region of peptide T3

Fig. 3. Primary structure of protein no. 38 from the cuticle of *Locusta migratoria* and of the peptides used to construct the amino acid sequence

Peptides are numbered from the *N*-terminus, with tryptic peptides having the prefix T and the chymotryptic peptides the prefix C. Unbroken arrows indicate sequenced residues, dashed arrows indicate tentative identification.

1). In most of the spectra the peptides were associated with Na^+ to various degrees. This is a normal occurrence, and when taken into account the measured values are in good agreement with the calculated masses for all the peptides.

DISCUSSION

The cuticle of the pharate locust contains close to 100 different structural proteins (Andersen *et al.*, 1986). The eight proteins that have so far been partially sequenced show extensive homology (Højrup *et al.*, 1986). As they also have similar solubility properties and amino acid compositions they are considered to constitute a family of proteins that are closely related evolutionary and may have similar function.

Analysis of extracts from small pieces of cuticle from different body parts has revealed that, whereas some proteins are localized to specific areas, others seem to be generally distributed in the cuticle. Protein no. 38, on which we here report the complete amino acid sequence, belongs to the latter group, as it can be found in most of the cuticle destined to become sclerotized.

The primary sequence was deduced from Edman degradations of the intact protein and sequence determination of peptides from enzymic digests with trypsin and chymotrypsin. The complete sequence is shown in Fig. 3, together with the peptides finally used for elucidation of the structure. In a parallel experiment chymotryptic and peptic peptides were sequenced by fast-atom-bombardment and electron-impact m.s. (Roepstorff *et al.*, 1985, and unpublished work). These data

verified 62% of the sequence. Protein no. 38 consists of 163 residues corresponding to an M_r of 15320.6 (isotopically averaged mass). This M_r is considerably lower than the value previously reported (21 600), which was estimated from electrophoresis in gels containing SDS. Verification of the M_r has, however, been obtained from plasma-desorption time-of-flight m.s. (P. Højrup, S. O. Andersen & P. Roepstorff, unpublished work).

The composition of the protein as derived from Fig. 3 is in good agreement with the amino acid analysis data (Table 1); it is highly unusual, as alanine amounts to 42% of the total, and only 14 of the 20 common amino acid residues are present. The basic pI (10.2) of the protein is due to the complete absence of acidic side chains. Although the composition is unusual, it is similar to that of other structural proteins isolated from the pharate locust cuticle (Højrup *et al.*, 1986).

The sequence shown in Fig. 3 shows a pronounced asymmetric distribution of amino acid residues. Three types of regions are found. The *N*-terminal (residues 1–6) is dominated by glycine, as is a region close to the C-terminal (residues 146–151). Most of the central part of the protein is dominated by alanine (content close to 50%), except for two short hydrophilic regions (positions 41–53 and 108–120). Regions homologous to the *N*-terminal glycine-rich region and to the following alanine-rich region were also found in the previously partly sequenced proteins.

In the eight *N*-terminal sequences reported previously several repeats of a small four-residue motif (Ala-Ala-Pro-Ala) were found, as were duplications of two slightly longer areas (eight and 13 residues). The motif Ala-

Fig. 4. Internal homology in protein no. 38

Identical residues are boxed.

Ala-Pro-Ala is also present in protein no. 38, where it occurs ten times in the sequence and further six times with an alanine exchange for a valine residue (usually in the ultimate position). In the alanine-rich regions, the proline residues are most frequently spaced with a distance of six residues. The significance of this repeating unit is unknown, but it must have a profound effect on the secondary structure and might result in a repeating three-dimensional structure. Protein no. 38 is constructed of a duplication that covers all of the alanine-rich region. This duplication is shown in Fig. 4. In order to obtain maximal homology two insertions of one and six residues respectively had to be made.

Secondary-structure prediction by the combined methods of Garnier *et al.* (1978), Levitt (1978) and Chou & Fasman (1974) indicates that the *N*-terminal glycine-rich and the hydrophilic regions are in β -sheet conformation. A β -turn and/or β -sheet may be present close to the *C*-terminus. The secondary structure of the alanine-rich regions is predicted to be in α -helical conformation. As the many proline residues cannot be accommodated in α -helices, the α -helices would then start with and terminate at proline residues and thus result in short helices of approx. $1\frac{1}{2}$ turns.

As the other proteins from locust cuticle appear to share the same structural elements, it seems plausible that their total sequences, like that of protein no. 38, are constructed of regions dominated by a few residues. Their secondary structures may also be very similar.

A search for homologous sequences in other proteins was carried out at the Carlsberg Laboratory by using the National Institute of Health protein sequence data base, but no related proteins were found. Although the sequence of protein no. 38 appears to be unique for this family of cuticular proteins, there may be some structural relationship to other structural proteins of insect origin, such as the chorion proteins from silkworm eggs (Hamodrakas *et al.*, 1982) and the cockroach oothecal proteins (Pau, 1984). In both types of proteins the *N*- and *C*-terminal regions are enriched in glycine, and the central regions are enriched in alanine and valine. None of these proteins has the proline residues placed in the constant environment characteristic for the locust cuticular proteins, and presumably there is no evolutionary relationship between these families of structural proteins.

Since the function of the structural proteins is to be integral components of composite protective materials, the precise interactions between a protein molecule and its immediate neighbour molecules, be they other proteins or chitin, will be of importance to the mechanical and other properties of the material. Since samples of cuticle from various body regions of a locust show distinct differences in protein composition and have

different mechanical properties, it can be supposed that the properties of the cuticle are modulated by regulation of the composition of the mixture of proteins deposited into the cuticle. A more detailed study of how these cuticular proteins are folded, and what factors can influence their mutual interactions, is needed.

This work was supported by the Danish Natural Science Research Council, the Carlsberg Foundation and the Novo Foundation. We are grateful to Hans Frede Hansen, Rigshospitalet, for help in part of the sequence work, to Flemming Poulsen, Carlsberg Laboratory, for searching the National Institutes of Health data base for homologous sequences, to Miss Susanne Klitgaard and Miss Lisbeth Poulsen for technical assistance and to Mrs. Kæthe Jørgensen and Mrs. Ulla Pedersen for typewriting the manuscript.

REFERENCES

- Andersen, S. O. (1979) *Annu. Rev. Entomol.* **24**, 29–61
- Andersen, S. O. (1980) in *Cuticular Techniques in Arthropods* (Miller, T. A., ed), pp. 185–215, Springer-Verlag, New York, Heidelberg and Berlin
- Andersen, S. O. (1985) in *Comprehensive Insect Physiology, Biochemistry and Pharmacology* (Kerkut, G. A. & Gilbert, L. I., eds.), vol. 3, pp. 59–74, Pergamon Press, Oxford, New York, Toronto, Sydney, Paris and Frankfurt
- Andersen, S. O. & Barrett, F. M. (1971) *J. Insect Physiol.* **17**, 69–83
- Andersen, S. O., Højrup, P., Roepstorff, P. (1986) *Insect Biochem.*, in the press
- Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 222–245
- Esch, F. S. (1984) *Anal. Biochem.* **136**, 39–47
- Fristrom, J. W., Hill, R. J. & Watt, F. (1978) *Biochemistry* **17**, 3917–3924
- Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120
- Gray, W. (1967) *Methods Enzymol.* **11**, 469–475
- Greene, L. J. & Bartelt, D. C. (1977) *Methods Enzymol.* **47**, 170–174
- Hamodrakas, S. J., Jones, C. W. & Kafatos, F. C. (1982) *Biochim. Biophys. Acta* **700**, 42–51
- Højrup, P., Andersen, S. O. & Roepstorff, P. (1986) *Eur. J. Biochem.* **154**, 153–159
- Hunkapiller, M. W. & Hood, L. E. (1983) *Methods Enzymol.* **91**, 486–493
- Levitt, M. (1978) *Biochemistry* **17**, 4277–4285
- Lipke, J. & Henzel, W. (1981) *Insect Biochem.* **11**, 445–451
- Neville, A. C. (1975) *Biology of the Arthropod Cuticle*, pp. 159–234, Berlin, Springer-Verlag
- Pau, R. N. (1984) *Biochim. Biophys. Acta* **782**, 422–428
- Roepstorff, P., Højrup, P. & Møller, J. (1985) *Biomed. Mass Spectrom.* **12**, 181–189
- Silvert, D. J., Doctor, J., Quesada, L. & Fristrom, J. W. (1984) *Biochemistry* **23**, 5767–5774

- Snyder, M., Hunkapiller, M., Yuen, D., Silvert, D., Fristrom, J. & Davidson, N. (1982) *Cell* (Cambridge, Mass.) **29**, 1027–1040
- Sottrup-Jensen, L., Petersen, T. E. & Magnusson, S. (1980) *Anal. Biochem.* **107**, 456–460
- Sottrup-Jensen, L., Stepanik, T. M., Jones, C. M., Lønblad, P. B., Kristensen, T. & Wierzbicki, D. M. (1984) *J. Biol. Chem.* **259**, 8293–8303
- Sundqvist, B., Kamensky, I., Håkansson, P., Kjellberg, J., Salehpour, M., Widdiyasekera, S., Fohlman, J., Peterson, P. A. & Roepstorff, P. (1984) *Biomed. Mass Spectrom.* **11**, 242–257
- Willis, J. H., Regier, J. C., Debrunner, B. A. (1981) in *Current Topics in Insect Endocrinology and Nutrition* (Baskaran, G., Friedman, S. & Rodriguez, J. G., eds.), pp. 27–46, Plenum Press, New York

Received 17 October 1985/18 December 1985; accepted 10 February 1986