Protein Science (1995), *4*:394–404. Cambridge University Press. Printed in the USA.

Isoforms of a cuticular protein from larvae of the meal beetle, *Tenebrio molitor*, studied by mass spectrometry in combination with Edman degradation and two-dimensional polyacrylamide gel electrophoresis

SOPHIE HAEBEL,[1] CHARLOTTE JENSEN,[1] SVEND OLAV ANDERSEN,[2] AND PETER ROEPSTORFF[1]

[1] Department of Molecular Biology, Odense University, DK-5230 Odense M, Denmark
[2] August Krogh Institute, Copenhagen University, DK-2100 Copenhagen Ø, Denmark

(RECEIVED November 14, 1994; ACCEPTED December 30, 1994)

Abstract

Simultaneous sequencing, using a combination of mass spectrometry and Edman degradation, of three approximately 15-kDa variants of a cuticular protein extracted from the meal beetle *Tenebrio molitor* larva is demonstrated. The information obtained by matrix-assisted laser desorption ionization mass spectrometry (MALDI MS) time-course monitoring of enzymatic digests was found essential to identify the differences among the three variants and for alignment of the peptides in the sequence. To determine whether each individual insect larva contains all three protein variants, proteins extracted from single animals were separated by two-dimensional gel electrophoresis, electroeluted from the gel spots, and analyzed by MALDI MS. Molecular weights of the proteins present in each sample could be obtained, and mass spectrometric mapping of the peptides after digestion with trypsin gave additional information. The protein isoforms were found to be allelic variants.

Keywords: allelic variants; electroelution; gel electrophoresis; insect cuticle proteins; mass spectrometry; protein sequencing

Structural proteins purified from insect sources are often microheterogeneous due to single amino acid replacement (Andreasen et al., 1993), to differences in glycosylation (Talbo et al., 1991), or to a varying number of short, repeated motifs. Pronounced heterogeneity due to a varying number of repeats has been described among moth chorion proteins (oothecins) (Regier & Kafatos, 1985; Pau et al., 1986), where the variants can be observed as separate spots on two-dimensional electrophoresis gels. A more limited heterogeneity has been observed for locust cuticular proteins (Andreasen et al., 1993; unpublished observations from current work of our laboratory), and the differences between the isoforms are so small that they are not resolved by electrophoresis or isoelectric focusing.

Protein isoforms obtained by batch purification of cuticular proteins from insects either may be allelic variants of the same gene, and thus represent population polymorphism, or may be products from closely related genes, which are all present in the genome of each individual. This question can only be answered by the use of techniques allowing accurate characterization of proteins obtained from single animals.

The progress in recent years in mass spectrometric methods for characterization of peptides and proteins should allow one to obtain sufficient information to identify the various isoforms of a protein from the amount that can be eluted from a single electrophoretic gel. The amount of protein contained in a single gel spot is typically around a microgram. Due to high sensitivity and good tolerance toward impurities, matrix-assisted laser desorption ionization mass spectrometry should be the ideal tool to obtain molecular mass information on the intact proteins as well as mass spectrometric peptide maps from such a small amount of protein isolated from a gel.

Reprint requests to: Peter Roepstorff, Department of Molecular Biology, Odense University, Campusvej 55, DK-5230 Odense M, Denmark; e-mail: roe@pr-group.ou.dk.

Abbreviations: ESI, electrospray ionization; MS, mass spectrometry; MALDI, matrix-assisted laser desorption ionization; RP-HPLC, reverse-phase high pressure liquid chromatography; 2D, two-dimensional; TmlF1, cuticular protein from *Tenebrio molitor* larvae present in fraction F1; TFA, trifluoroacetic acid; PTH, phenylthiohydantoin; PVDF, polyvinyldifluoride; TPCK, *N*-tosyl-L-phenylalanine chloromethyl ketone.

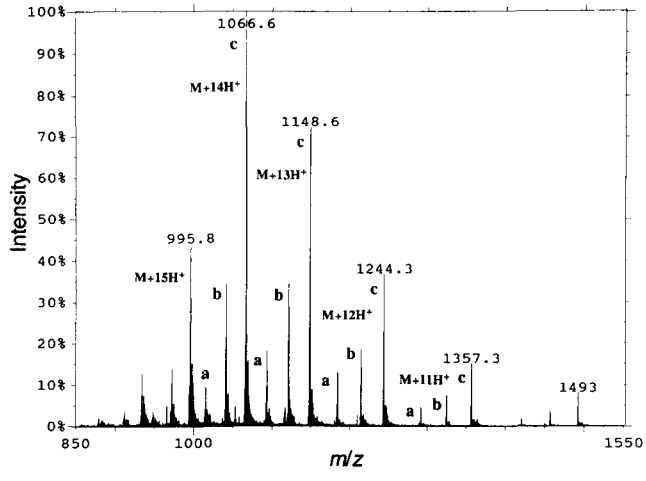

Fig. 1. Electrospray mass spectrum of the Tml-F1 protein mixture obtained by batch purification, indicating the presence of three variants Tml-F1a, Tml-F1b, and Tml-F1c, named according to increasing mass.

In the course of our characterization of proteins from insect cuticles, we have obtained three isoforms of a larval cuticular protein from the meal beetle *Tenebrio molitor*. Their amino acid sequences were determined by a general strategy combining the use of MS and Edman degradation (Roepstorff & Højrup, 1993). Mass determination of the protein after electroelution from a 2D gel of cuticular extracts demonstrated that it is possible unambiguously to identify the variants present in single individuals. Identification of variants was confirmed by tryptic mapping of the proteins in the electroeluted spots. Twelve animals have been analyzed, and they contain either only one of the three isoforms or two of them in approximately equal amounts, indicating that they are allelic variants.

Results

Primary structure of the Tml-F1 protein

The Tml-F1 protein is one of the major components of *T. molitor* larvae cuticular extract. F1 refers to the F and the first fractions of two successive chromatographic separations.

Molecular weight determination of protein Tml-F1

The ESI spectrum (see Fig. 1) shows the presence of a mixture of three protein variants that could not be separated by RP-HPLC (C4 column). The three variants in the F1 fraction were previously termed a, b, and c, according to increasing mass (Andersen et al., 1994b). The measured molecular weights of the variants, their relative abundance, and their mass differences with respect to the most abundant species are given in Table 1.

N-terminal sequence

Automatic Edman degradation of the protein allowed identification of the first 40 residues from the N-terminus with one unidentified amino acid residue in position 34. No clear minor sequences could be detected, suggesting either that the sequences of the three compounds do not differ up to that point of the sequence or that minor signals are lost in the background.

Table 1. *Average molecular weights, mass differences with respect to the most abundant component (Tml-F1c), and relative amounts of the three different protein variants detected*[a]

Protein variant	Measured ave. M_r	Mass difference	Relative amount (%)	Calculated ave. M_r[b]
Tml-F1c	14,919.9	0	60	14,919.1
Tml-F1b	14,567.5	352.4	30	14,566.7
Tml-F1a	14,201.7	718.2	10	14,202.3

[a] Relative amounts of protein variants detected in the mixture were determined according to the abundance of the corresponding peaks in the electrospray mass spectrum shown in Figure 1.
[b] Calculated based on the later determined sequence.

Amino acid analysis

Amino acid analysis of several samples gave the results summarized in Figure 2. The occurrence of about five arginines and no lysines led to the choice of trypsin for digestion of the protein.

Tryptic digestion

Tryptic digestion of the protein mixture yielded seven major peptides (Table 2), two co-eluting (T2 and T2a) (Fig. 3). The sum of the masses of the most abundant peptides, T1-T2-T3-T4-T5, give an average mass of 14,924 Da, which is within 4 Da of the molecular weight determined by ESI MS for the most abundant variant, Tml-F1c. Moreover, peptide T1b, present in minor quantities as suggested by a smaller HPLC peak, has a mass 353 Da below that of T1 and might therefore replace T1 in the sequence of the minor compound Tml-F1b. This possibility is confirmed by sequencing of T1 and T1b because the two amino acid sequences are identical except for one missing repetitive unit, IAAP, in T1b (see Table 2).

The sequence obtained from the intact protein identifies peptide T1 (T1b for Tml-F1b) to be the N-terminus as well as T2 to be the following peptide. The absence of an arginine residue at the C-terminus of peptide T5 suggests that it is C-terminus in the protein. Therefore, only T3 and T4 must be placed in the correct order.

Their alignment is possible with the help of two larger peptides observed in the MALDI spectrum of a peptide mixture obtained by incomplete tryptic digestion (Fig. 4). A peptide corresponding to T3 plus T4 and a peptide with a mass corresponding to T4 plus T5 are found in the spectrum, showing that the peptide order is T1-T2-T3-T4-T5 for Tml-F1c and T1b-T2-T3-T4-T5 for Tml-F1b. The 716.3-Da mass difference between T2 and T2a and their relative intensities in the corresponding MALDI spectrum also suggest that T2a might replace T2 in the sequence of the third low abundance component Tml-F1a.

All tryptic peptides except for T2 and T2a could be entirely sequenced by N-terminal Edman degradation (see Table 2). The expected presence of approximately three tyrosine residues and several leucines led to the choice of chymotrypsin for further subdigestion of the mixture of T2 and T2a, which were not separated by HPLC.

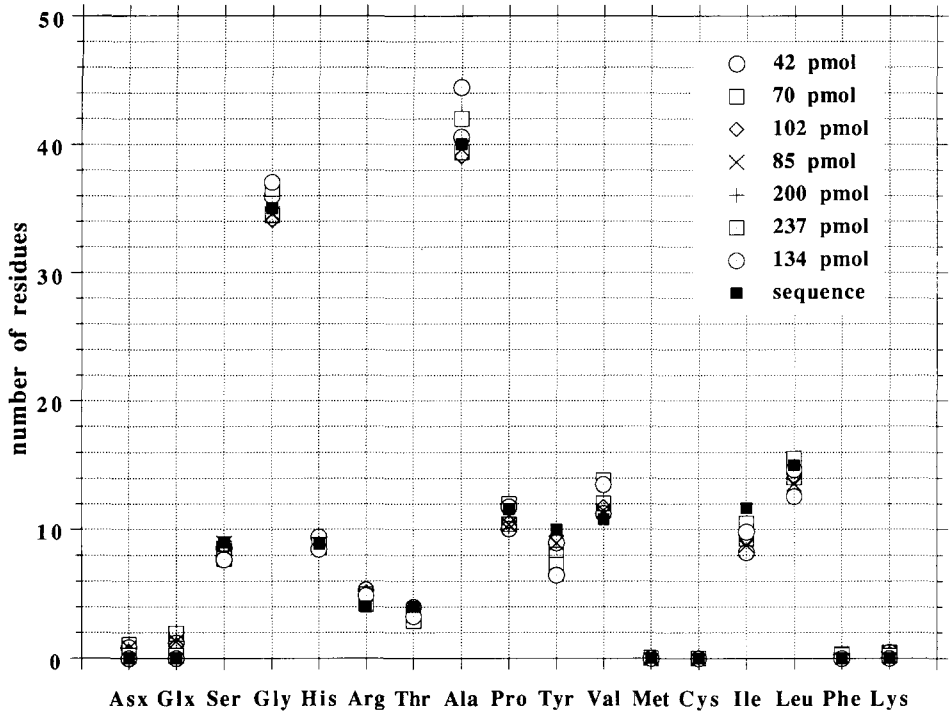

Fig. 2. Results of multiple amino acid analysis performed on various amounts of the Tml-F1 protein mixture. Filled squares represent the amino acid composition calculated for the mixture of the three variants in the relative abundance indicated by Figure 1 and based on their primary sequences.

Chymotryptic digestion

HPLC separation of the chymotryptic digest of the T2/T2a peptide mixture resulted in eight or nine major peaks (Fig. 5). The amino acid sequences of the corresponding peptides were determined by Edman degradation, and five peptides (TC1 to TC5) were found to result from cleavage at tyrosine or leucine residues (Table 3), whereas the remaining peptides resulted from additional cleavage at histidine residues. The summed molecular weights of these five peptides gives a mass of 5,673.5 Da, which is 1,451 Da below that of T2. This mass difference would correspond to a missing peptide of 1,469 Da, which, according to Table 3, corresponds precisely to the mass of peptide TC2. The very large HPLC peak corresponding to this peptide, and its occurrence in an area of the protein consisting of several repeating units, suggests that it is repeated twice in the sequence. No peptide corresponding to the minor component Tml-F1a was detected.

Time-course chymotryptic digestion of T2 and T2a

To confirm that peptide TC2 is repeated twice in T2, the chymotryptic digestion was monitored in time course. The digestion was followed by successive MALDI spectra and stopped at a time suitable to isolate the peptides of interest (50 min at 37 °C) by addition of an equal amount of a 0.1% TFA solution. To study the difference between T2 and T2a, fractions enriched in each peptide were prepared (see Materials and methods). Results from the "T2-enriched" fraction (Fig. 6A) show the presence of the peptides listed in Table 4, which can all be attributed to combinations of two or more of the already observed peptides TC1 to TC5. Thus, peptides TC13 to TC16 confirm the presence of

Table 2. *Major peptides obtained from tryptic digestion of Tml-F1 and their sequences as determined by automatic Edman degradation, mass calculated from the sequence, and mass measured by MALDI MS*[a]

Peptide	Sequence	Measured ave. M_r	Calculated ave. M_r	Error (%)
T1	GLIGAPIAAPIAAPLATSVVSTR	2,147.1[b]	2,146.5	0.028
T1b	GLIGAPIAAPLATSVVSTR	1,794.0[b]	1,794.1	0.006
T2	TIHAAPVAVAHAAPLAVAHAAPVAVXHAA...	7,124.2[b]	7,121.2[c]	0.042
T2a	?	6,407.9[b]	6,404.4[c]	0.055
T3	SLYGGYGSGLGIAR	1,370.7[b]	1,370.5	0.015
T4	STPGGYGSGLIGGAYGSGLIGGGLYGAR	2,516.4[b]	2,515.7	0.028
T5	YGLGAPALGHGLIGGAHLY	1,837.6[b]	1,837.1	0.027

[a] T2 and T2a elute in the same HPLC peak.
[b] Internal calibration on human insulin.
[c] Calculated based on the later determined sequence.

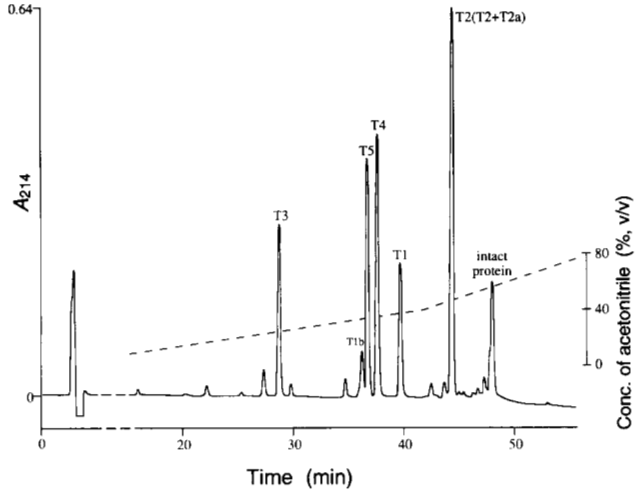

Fig. 3. HPLC separation of peptides derived from tryptic digestion of protein Tml-F1. Dotted line shows the gradient.

two TC2 units one after the other and allow alignment of the peptides in the order: TC1-TC2-TC2-TC3-TC4-TC5.

The repetition of two TC2 units is further confirmed by N-terminal sequencing of fragment TC13 (TC2-TC2-TC3), giving the following sequence:

TC2 TC2 TC3

|AVAHAAPVAVAHAAPL||AVAHAAPVAVAHAAPL||AVAHAA . . .

In the sequence of T2, the sequence motif VAVAHAAP (716.8 Da) is repeated three times along the protein portion, covering chymotryptic peptides TC1-TC2-TC2. The absence of one of these repeats would explain the measured mass difference of approximately 717 Da between Tml-F1c and Tml-F1a.

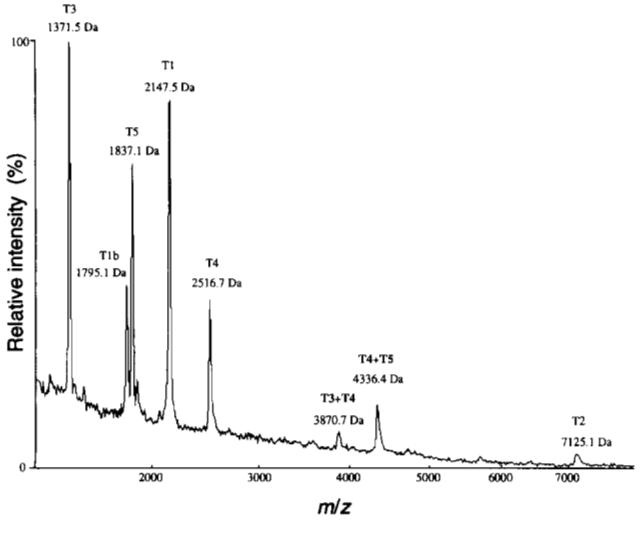

Fig. 4. MALDI spectrum of the peptide mixture resulting from incomplete tryptic digestion of protein Tml-F1. Internal calibration was performed using the masses of peptides T3 and T4 as calculated from their previously determined sequence (Table 2).

Table 3. *Major peptides obtained from chymotryptic digestion of the T2/T2a mixture in the naturally occurring 6:1 proportion*

Peptide	Sequence	Measured ave. M_r	Calculated ave. M_r	Error (%)
TC1	TIHAAPVAVAHAAPL	1,440.8[a]	1,438.7	0.11
TC2	AVAHAAPVAVAHAAPL	1,467.8[a]	1,465.7	0.14
TC3	AVAHAAPAIAY	1,056.3[a]	1,054.2	0.2
TC4	GGYGSGIIGGAY	1,072.4[a]	1,071.1	0.12
TC5	GGGLIGSR	716.6[b]	715.8	0.11

[a] Measured by MALDI with internal calibration on human insulin.
[b] Measured by plasma desorption mass spectrometry as described by Andersen et al. (1986).

The MALDI spectrum of the incomplete chymotryptic digest of the "T2a-enriched" T2/T2a mixture, presented in Figure 6B (peptides listed in Table 5), allows localization of the missing amino acid sequence. For TC14 (TC1-TC2-TC2), TC13 (TC2-TC2-TC3), and TC12 (TC1-TC2), an approximately 717-Da lower mass peak, corresponding to the Tml-F1a isoform can be clearly seen. This result locates the missing group of eight amino acids to the first TC2 peptide. This conclusion is confirmed because no peak 717 Da below that of TC11 (TC2-TC3) is observed.

The lack of a VAVAHAAP repeat in the first TC2 peptide was further confirmed by N-terminal sequencing of the "T2a-enriched" peptide mixture. In position 23, both valine and leucine were observed, Val corresponding to TC2 (Tml-F1c) and Leu to TC2a (Tml-F1a):

TC1 TC2/TC2a

|TIHAAPVAVAHAAPL||AVAHAAP$^{V}_{L}$AVA . . .

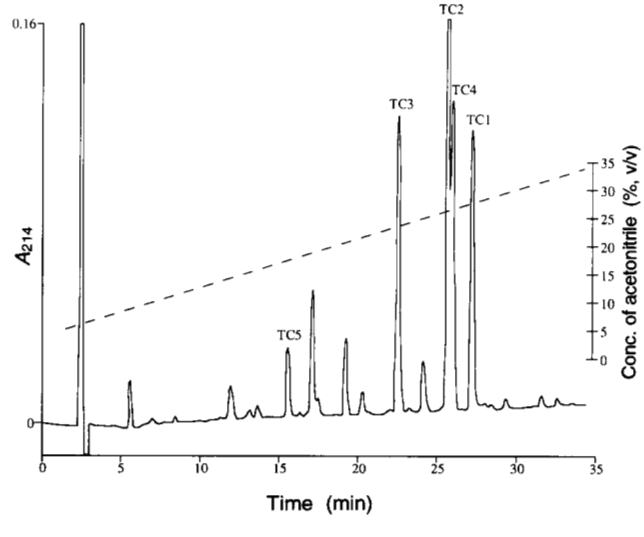

Fig. 5. HPLC separation of peptides obtained from chymotryptic subdigestion of the mixture of tryptic peptides T2/T2a.

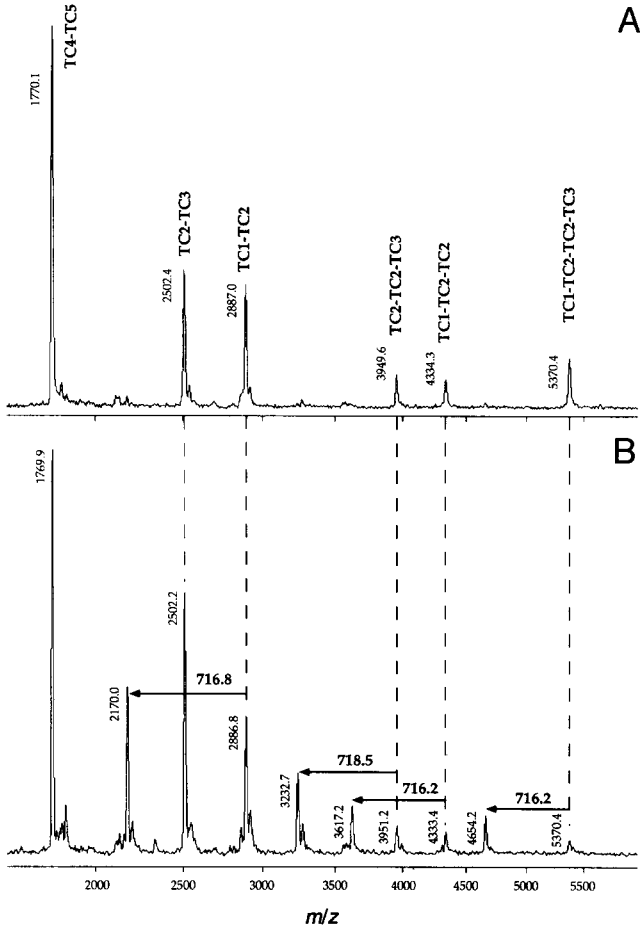

Fig. 6. MALDI spectrum of the peptide mixture obtained after a 50-min chymotryptic digestion of almost pure peptide T2 (**A**) and after digestion (same experimental conditions as A) of a mixture of the T2 and T2a peptides in molar proportions of approximately 1:2 (**B**).

As a result of the investigations described above, the sequence of all three protein variants is established as given in Figure 7.

Mass spectrometric characterization of proteins electroeluted from 2D gels of single animal extracts

The protein Tml-F1 used for sequencing was purified from cuticles collected from a large number of animals. Therefore, it cannot be decided whether the different forms are isoforms present in all animals or represent allelic differences. Preparative isolation of the protein from single animals was not possible. Therefore, the protein extracts from single animals were separated by 2D-PAGE (Fig. 8) and the spots corresponding to the Tml-F1 protein electroeluted for further investigation. A first series of experiments was performed on six different animals.

For each electroeluate, a small sample of 0.8 μL of a total volume of approximately 400 μL was analyzed by MALDI MS. The results are shown in the upper part of Table 6, and one of the spectra is shown in Figure 9A.

Table 4. *Larger mass peptides obtained from uncompleted chymotryptic digestion of almost pure T2 (see Fig. 6A) allowing assignment of the peptide sequence TC1-TC2-TC2-TC3-TC4-TC5 for T2*

Peptide number	Peptide groups	Measured ave. M_r	Calculated ave. M_r	Error (%)
TC10	TC4-TC5	1,769.1[a]	1,768.9	−0.011
TC11	TC2-TC3	2,501.4[a]	2,501.9	0.019
TC12	TC1-TC2	2,886.0[a]	2,886.4	0.014
TC13	TC2-TC2-TC3	3,948.6[a]	3,949.6	0.025
TC14	TC1-TC2-TC2	4,333.3[a]	4,334.1	0.018
TC15	TC1-TC2-TC2-TC3	5,369.4[a]	5,370.3	0.017
TC16	TC1-TC2-TC2-TC3-TC4	6,423.3[b]	6,423.4	0.002

[a] Internal calibration on the mass calculated from the sequence of peptide TC5 and on the measured mass of the HPLC-purified peptide TC15 (TC1-TC2-TC2-TC3) (internal calibrant: human insulin).

[b] This peptide was not observed on the spectrum shown in Figure 6, but was measured independently after HPLC purification (internal calibrant: human insulin).

The spectra obtained have sufficient resolution ($\Delta m/m \approx 85$) to distinguish the two isoforms, a and b, separated by 352 Da. In some spectra, small adduct ions at approximately M + 260 and M + 800 were observed. They can be attributed to adducts from SDS ($C_{12}O_4H_{25}S^-$) and Coomassie blue, respectively. Fortunately, none of these adducts interfered with the interpretation of the spectra. However, the peaks corresponding to Tml-F1c and Tml-F1b are unusually broad and, in some spectra, split into a doublet, which indicates another poorly resolved compound approximately 60 Da above the mass of the protein. To investigate the origin and nature of this compound, tryptic digestion was performed by adding 0.5 μg of trypsin to 30 μL of eluted protein solution. A spectrum of the resulting peptide mixture was recorded from a 0.8-μL aliquot of the digest (Fig. 9B). All tryptic peptides corresponding to the present protein variants, as listed in Table 2, could be detected (although the peak for T4 is very weak). A series of products of incomplete digestion corresponding to different combinations of tryptic peptides could also be observed. A 71-Da adduct ion is observed for T1 as well as for T1b. This adduct ion has been observed by several authors and interpreted as a product of a reaction with nonpolymerized acrylamide during electrophoresis. Several authors (Chiari et al., 1992; Hall et al., 1993; Le Maire et al., 1993) reported the observation of acrylamidated cysteine residues. Klarskov et al. (1994) found a peptide from an in-gel digestion with a mass increase of 71 Da, although the peptide did not contain cysteine. They suggested that either the ϵ-amino group of lysine or the N-terminal α-amino group was acrylamidated. Because the peptide was not the N-terminus of the protein, a possible reaction with the N-terminal group could only have occurred after enzymatic digestion and not during gel separation. Because, in our case, peptides T1 and T1b do not contain cysteine or lysine, the higher mass component is most likely caused by acrylamidation of the N-terminal α-amino group of the protein. Because acrylamidation of the N-terminus has not been unambiguously demonstrated previously, it is likely not to be a frequent phenomenon. However, its occurrence might explain occasional failure to sequence electroeluted or electroblotted

Table 5. *Larger mass peptides obtained from uncompleted chymotryptic digestion of a ca. 1:2 mixture of T2/T2a (see Fig. 6B), allowing assignment of the peptide sequence TC1-TC2a-TC2-TC3-TC4-TC5 for T2a*

Peptide number	Peptide groups	Measured ave. M_r	Calculated ave. M_r	Measured mass difference	Error (%)
TC10	TC4-TC5	1,768.9[a]	1,768.9		0
TC11	TC2-TC3	2,501.2[a]	2,501.9		0.028
TC12[b]	TC1-TC2a	2,169.0[a]		716.8	
TC12	TC1-TC2	2,885.8[a]	2,886.4		0.021
TC13[b]	TC2a-TC2-TC3	3,231.7[a]		718.5	
TC13	TC2-TC2-TC3	3,950.2[a]	3,949.6		−0.015
TC14[b]	TC1-TC2a-TC2	3,616.2[a]		716.2	
TC14	TC1-TC2-TC2	4,332.4[a]	4,334.1		0.039
TC15[b]	TC1-TC2a-TC2-TC3	4,653.2[a]		716.2	
TC15	TC1-TC2-TC2-TC3	5,369.4[a]	5,370.3		0.017

[a] Internal calibration as for Table 4.
[b] Peptides originate from T2a (Tml-F1a), because they differ by approximately 717 Da compared to the corresponding peptides originating from T2 (Tml-F1c).

proteins. The observation of this reaction might be explained by the fact that it is favored by low steric hindrance due to the presence of a glycine as the N-terminal residue or the high pK of the α-amino group.

The experiments demonstrate that the mass spectrometric peptide maps obtained from tryptic digests can be used to resolve ambiguities in identification of a protein from its molecular peak alone. Such ambiguities might (as here) be caused by protein

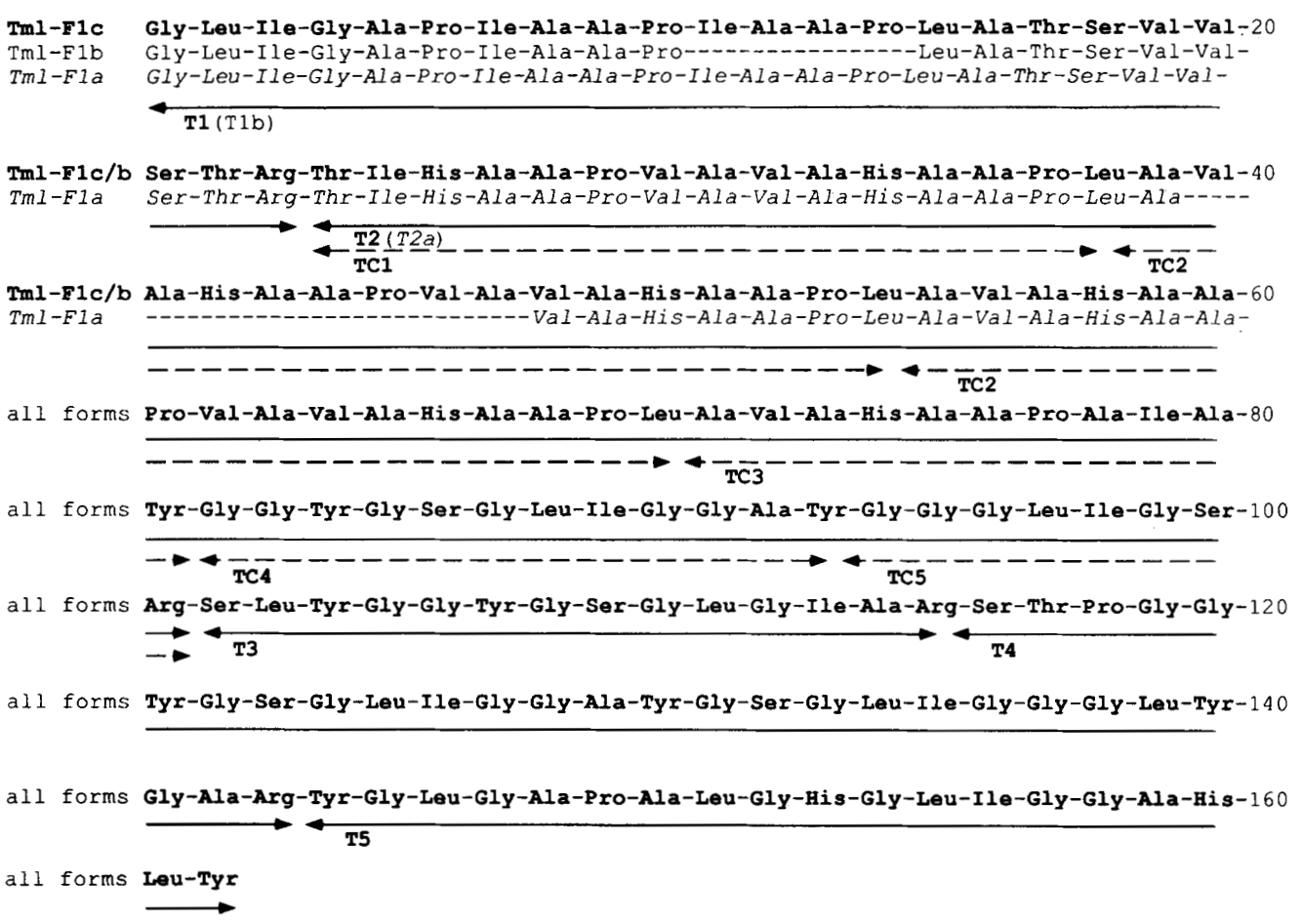

Fig. 7. Amino acid sequences of the three variant proteins Tml-F1c, Tml-F1b, and Tml-F1a.

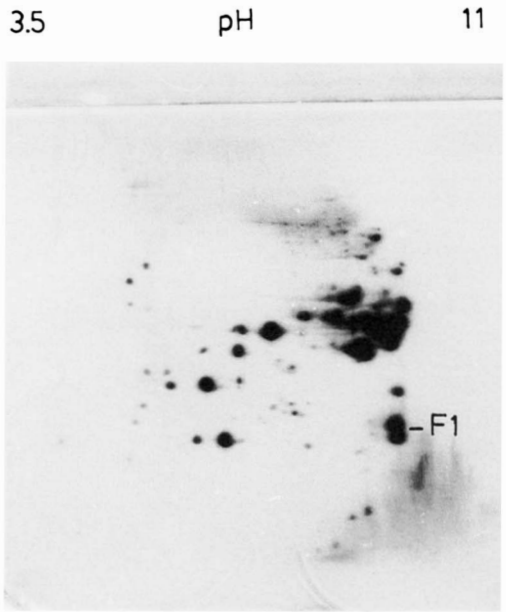

Fig. 8. 2D-PAGE of the protein mixture extracted from the cuticle of a single animal. The gel shown is gel number 10 containing variants a and c and two poorly separated spots are visible. When variants a and b are present in the same gel, only one spot is observed.

modification during gel electrophoresis or may be due to insufficient mass accuracy. Moreover, in this specific study, the presence or absence of the peptides specific for the protein variants a, b, and c allowed a complementary and unambiguous identification of the proteins extracted from the gel.

The results in Table 6 show that, among the six animals studied, four contained variants c and b, one only c, and one only b. Unfortunately, none of the studied animals contained variant a. It was, therefore, decided to pursue the investigation on another series of six animals. This time we attempted to avoid acrylamidation by adding mercaptoacetic acid as a scavenger to the buffer used for the second electrophoresis dimension. The results are presented in Table 6 (gels 7–12), and examples of spectra of the intact protein as well as of its tryptic peptide map are shown in the lower part of Figure 9. The positive effect of the scavenger can be seen in both spectra because the molecular peak is narrower (resolution ≈ 150) and the adduct to T1 in the peptide map spectrum is of much lower intensity. From Table 6, it can be seen that the extent of acrylamidation dropped from an average level of 55% to an average of 7%. Among the second group of investigated animals, three contained variants c and b, two only c, and one a and c. The results strongly indicate that any single animal contains either two or only one of the three protein variants.

Discussion

The amino acid sequence has been obtained for three isoforms of the *T. molitor* larval cuticular protein, Tml-F1. The sequences were determined according to a previously described strategy (Roepstorff & Højrup, 1993) by combining information obtained by MS with information from Edman degradation. The major component, F1c, has a relative molecular mass of 14,919 Da and is 162 residues long, whereas the two minor compounds, F1b and F1a, are 158 and 154 residues long, respectively. The sequence of F1b corresponds to that of F1c after deletion of residues 11–14 (Ile-Ala-Ala-Pro), whereas the sequence of F1a corresponds to F1c after deletion of residues 40–47 (Val-Ala-Val-Ala-His-Ala-Ala-Pro).

The sequences are characterized by the lack of aspartic and glutamic acid and their amides, the sulfur-containing amino acids, as well as lysine, phenylalanine, and tryptophan. The N-terminal half of the proteins is dominated by alanine, valine,

Table 6. *Summary of results obtained from the 12 electroeluted Tml-F1 gel spots*[a]

Gel	Tml-F1c (14,919 Da)	Tml-F1b (14,567 Da)	Tml-F1a (14,202 Da)	T1 (2,146.5 Da)	T1b (1,794.1 Da)	T2 (7,121.2 Da)	T2c (6,404.4 Da)	"Acrylamidation" (%)
1	14,925	14,567	–	2,148.4	1,796.3	7,119.7	–	55
2	14,924	14,570	–	2,149.9	1,797.2	7,126.7	–	50
3	14,941	–	–	2,149.0	–	7,124.1	–	20
4	14,933	14,645[b]	–	2,151.3	1,797.7	7,127.7	–	65
5	14,910	14,570	–	2,148.8	1,796.6	7,129.0	–	70
6	–	14,642[b]	–	–	1,797.9	7,126.9	–	65
7	14,918	14,564	–	2,148.0	1,795.1	7,123.9	–	0
8	14,922	14,567	–	2,148.5	1,796.4	7,121.1	–	0
9	14,903	–	–	2,151.9	–	7,130.2	–	15
10	14,912	–	14,204	2,150.9	–	7,133.6	6,407.8	15
11	14,921	14,576	–	2,148.0	1,795.3	7,114.0	–	0
12	14,928	–	–	2,149.6	–	7,126.2	–	15

[a] Columns 2–4 give the molecular weights of the intact protein variants as determined by MALDI MS. Columns 5–8 list the tryptic peptides confirming the identification of the variants and showing the extent of "acrylamidation," which is listed in column 9. Gels 1–6 were run without scavenger and acrylamidation of the N-terminal was always observed, whereas gels 7–12 were run with a scavenger in the second dimension, which reduced the extent of acrylamidation.

[b] No clearly separated peak doublet was observed. Because these proteins are extensively acrylamidated, the peak is assumed to correspond to the acrylamidated form ($MH^+ + 71$).

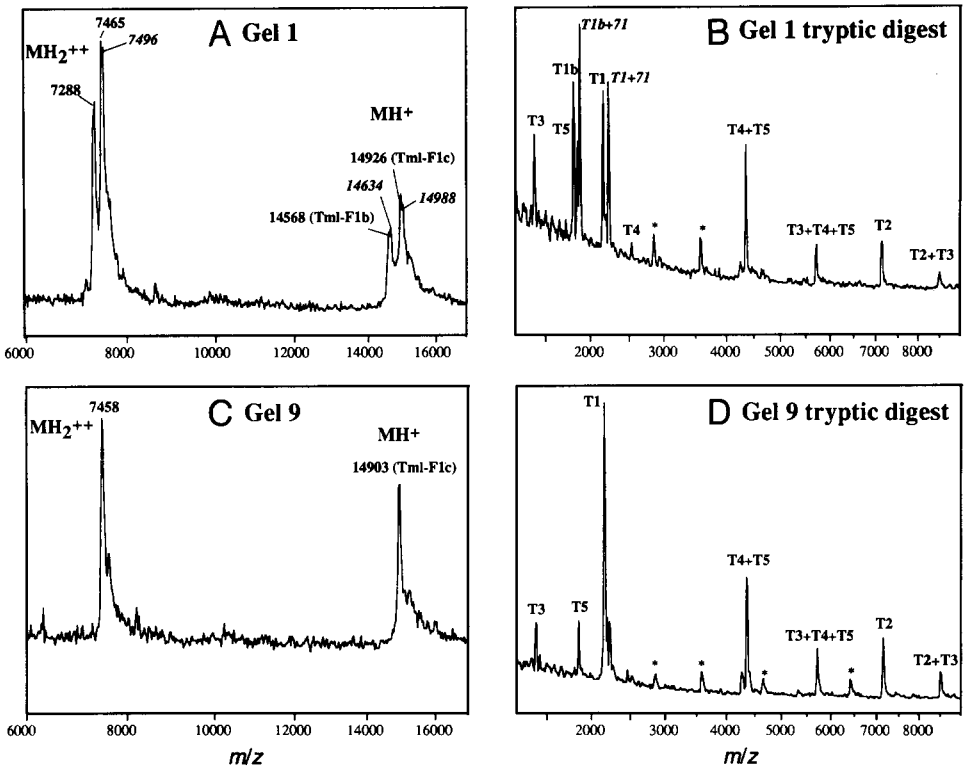

Fig. 9. Examples of the MALDI spectra obtained from 2 among the 12 electroeluted Tml-F1 gel spots. In the upper left spectrum (**A**), obtained from gel 1, the two protein variants c and b can be observed. Moreover, the spectrum of the tryptic digest (**B**) allows one to observe the peaks resulting from partial "acrylamidation" of the N-terminal peptides T1 and T1b. Lower spectra (**C** and **D**) are obtained from gel 9. They show that only the Tml-F1c variant is present and that the extent of acrylamidation is much lower. Ions corresponding to acrylamidated proteins or peptides are labeled in italics.

proline, and histidine, whereas the C-terminal is dominated by glycine and tyrosine.

Figure 10 shows a dot-plot where the sequence of F1c is plotted against itself. This presentation shows the different compositions of the two halves of the molecule. It also shows that shorter sequence motifs occur repeatedly in both halves. Variants of an octapeptide sequence are repeated seven times in the N-terminal region (see Fig. 11A). Variants of another octapeptide occur five times in the C-terminal region (see Fig. 11B).

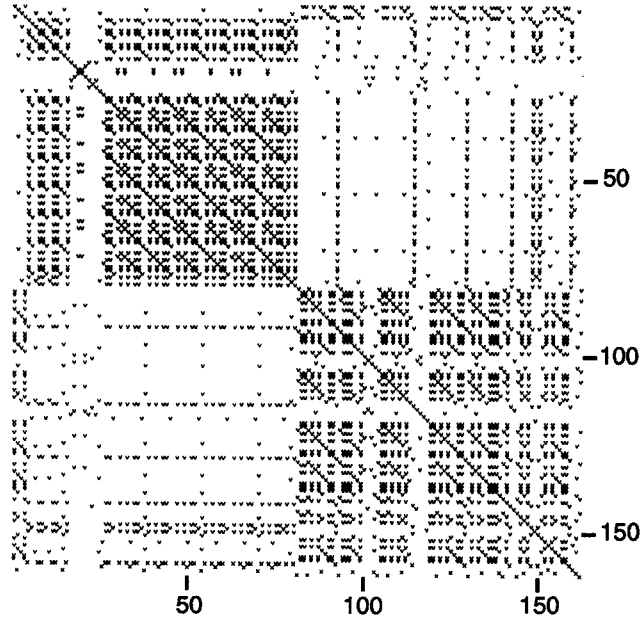

Fig. 10. Identity dot-plot of protein Tml-F1c versus itself. (Identical residues are represented by a dot.) This shows that the protein contains two regions of distinct amino acid composition and that short sequence motifs are repeated several times.

The lower M_r isoforms lack half of the first octapeptide motif in the N-terminal region (F1b) and one copy of the N-terminal repeat sequence completely (F1a). No differences are present in the glycine-rich C-terminal region.

The complete amino acid sequence has recently been obtained for another *T. molitor* cuticular protein, E1a, occurring in both larvae and pupae. Protein E1a contains 40% alanine, and it bears some resemblance to the N-terminal region of the F1 proteins, although no convincing homologies are present. The amino acid sequences for two cuticular proteins from adult *T. molitor* have been deduced from the corresponding nucleic acid sequences (Bouhin et al., 1992; Charles et al., 1992). One of the proteins has a tripartite structure with a glycine-rich N- and C-terminal region, whereas the other has a bipartite structure with a glycine-rich N-terminal region. No obvious sequence homologies can be observed between the various glycine-rich regions.

The function of the two types of sequence regions found in the F1 proteins is not known. The occurrence of similar regions in some cuticular proteins from other species might suggest that these types of sequences contribute to give the cuticle the optimal mechanical properties. Proline tends to occur with nearly regular spacing in the alanine-rich region, presumably resulting in regular turns in the chain conformation; also, the glycine-rich regions will presumably have a conformation dominated by turns. It has been suggested that such a structure may contribute to giving the cuticles an optimal combination of stiffness and flexibility (Andersen et al., 1994a).

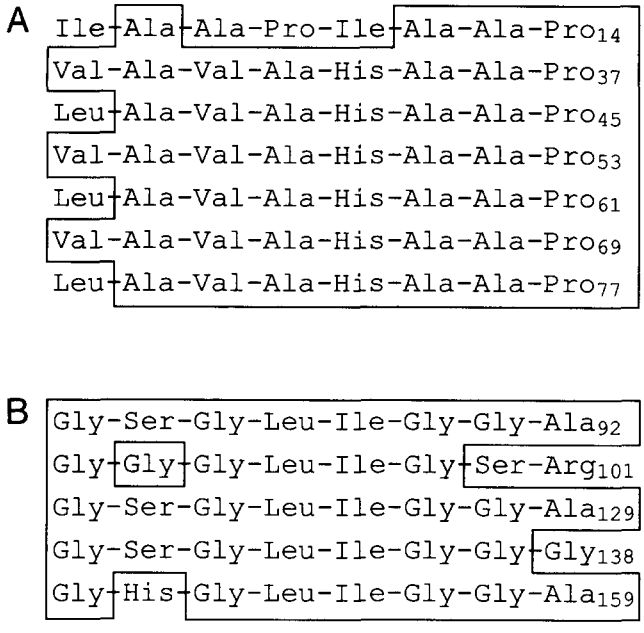

Fig. 11. Internal homology in protein Tml-F1c. Identical residues are boxed.

The F1 proteins were obtained from a large batch of mealworms (ca. 3 g), and the occurrence of the three isoforms could indicate genetic polymorphism in the population. Mass spectrometric analysis of the proteins obtained from single individuals by 2D gel electrophoresis revealed that an animal contains either only one of the variants or two of them in nearly equal amounts. All three isoforms were not observed in a single animal, indicating that the three forms are coded for by allelic variants of the same gene and that some of the animals were homozygotes, whereas others were heterozygotes. The intensity of the a, b, and c peaks in the ESI spectrum of the protein isolated from a large number of animals (Fig. 1) reflects the statistical occurrence of the genes responsible for a, b, and c, provided that the ion yields for the three variants are the same. In our experience, this is the case for closely related proteins. Although the 12 investigated animals may not be enough to be statistically significant, the number of times the individual variants have been observed – 1 (5%) for a, 8 (40%) for b, and 11 (55%) for c – corroborates the distribution (10%/30%/60%) deduced from the ESI MS spectrum in Figure 1 and reported in Table 1.

The precision of the mass spectrometric molecular mass determination of the intact protein by ESI MS is very good (0.01%). For MALDI MS, two different mass calibration methods were used. All spectra of peptides were calibrated using human insulin or other known peptides as internal calibrant. The errors on the measurements are below 0.1% for the tryptic peptides and slightly higher for the chymotryptic peptides. The masses of gel-eluted proteins are all determined using a theoretical calibration based on the instrumental parameters because high precision is not required to identify the already known peptides or proteins. Surprisingly, the precision is generally within 0.1%, although a few deviations up to 0.35% were observed. The sensitivity of the combination of protein electroelution from 2D-PAGE and mass spectrometric measurements is adequate for practical protein studies, although the absolute sensitivity is difficult to estimate because several unknown factors are involved. First, the protein amount in the different gel spots is not known and varies between spots. The intensity of Coomassie blue staining gives a rough estimate of 1-4 μg per spot. For a 15-kDa protein, this amount represents between 70 and 270 pmol. The second unknown is the yield of electroelution, which at the highest can be estimated to 90% (according to the manufacturer, elution yields between 70% and 90% can be expected for stained and fixed proteins). The protein is eluted into a final volume of 400 μL buffer solution, corresponding to 150–600 fmol/μL. The spectra shown in Figure 8 were obtained from 0.8 μL of this solution, corresponding to 120–500 fmol of protein. The good signal-to-noise ratio in the spectra indicates that the sensitivity limit has not been reached.

The same estimate of the protein concentration shows that a very high enzyme/substrate ratio is needed for digestion of the eluted protein with trypsin. The molar excess of trypsin was between 1 and 4, which is two orders of magnitude higher than the ratio used for standard digestions in the sequencing procedure. Complete cleavage of all susceptible bonds, as observed under standard digestion conditions, is never obtained in the digests of electroeluted proteins. Furthermore, prolongation of the digestion time in excess of 1-2 h did not result in further cleavage. We do not know why digestion of electroeluted proteins is less effective. The presence of SDS, either protecting the protein from digestion or inhibiting the action of trypsin, seems to be the most likely explanation.

Materials and methods

Sequencing of the Tml-F1 protein

Materials

Non-TPCK-treated porcine trypsin (EC 3.4.21.4) was a gift from Novo Nordisk A/S, Bagsværd, Denmark; chymotrypsin (EC 3.4.21.1) came from Boehringer Mannheim, Germany. Water used was "ultra-high" quality (UHQ) water purified on an Elgastat UHQ (Elga Ltd., High Wycombe Bucks, UK). Chemicals and solvents for sequencing were from Knauer (Berlin, Germany). Solvents for amino acid analysis were from Applied Biosystems (Foster City, California). All other chemicals were analytical grade from various manufacturers.

Preparative isolation of the cuticle protein

Proteins were isolated from a laboratory culture of mealworms by a combination of ion-exchange and gel chromatography as described previously (Andersen et al., 1994b).

Amino acid analysis

Norleucine (450 pmol) was added as internal standard to 40–250 pmol of protein and lyophilized in small Eppendorf tubes. Tubes were placed in sealed screw-cap vials containing 100–200 μL 6 N HCl, 0.1% phenol, and the vials were flushed with nitrogen before the pressure was reduced to less than 1 Torr. Hydrolysis was carried out at 110 °C for 18–20 h.

Hydrolysates were lyophilized and redissolved in 20 μL 0.025% EDTA prior to analysis on an ABI amino acid analyzer consisting of a model 420A derivatizer, a model 130A separa-

tion unit, and a model 920A data analysis module (Applied Biosystems).

Digestion with trypsin

Preparative tryptic digestion was carried out on 150 μg (approximately 10 nmol) of protein. The protein was first dissolved in 15 μL 8 M urea and diluted to 150 μL with pH 7.8, 100 mM NH_4HCO_3 buffer. Trypsin (2 μg) dissolved in 20 μL of the same buffer was added to give a final enzyme/substrate ratio of 1:75 (w/w), corresponding to a molar ratio of 1:120. The digestion was performed at 37 °C for 90 min before being stopped by addition of an equivalent volume of a 0.1% TFA solution. Digestions were performed in siliconized Eppendorf tubes to prevent protein from sticking to the walls of the tube, where it would be inaccessible to the enzyme. The Eppendorfs were siliconized by rinsing with a solution of polymethylhydrogensiloxane in trichloroethylene (Struers Kebo Lab, Albertslund, Denmark) followed by rinsing twice with water and drying overnight at 37 °C. Lower enzyme/substrate ratio (1:500, w/w) and shorter digestion times (15 min at room temperature) were used to observe products of incomplete enzymatic cleavage corresponding to the sum of different tryptic peptides.

Digestion with chymotrypsin

Subdigestion of a mixture of the tryptic peptides T2 and T2a (molar ratio approximately 6:1) was performed on ca. 3 nmol (21 μg) of the mixture isolated by HPLC after tryptic digestion of the intact protein. The peptide mixture was dissolved in 5 μL 8 M urea and diluted to 50 μL with pH 7.8, 100 mM NH_4HCO_3 buffer before addition of 1 μg of chymotrypsin. This corresponds to an enzyme/substrate ratio of approximately 1:20 (w/w) or a molar ratio of 1:70. The digestion was performed in a siliconized Eppendorf tube for 2.5 h at 37 °C.

To separate the two peptides, T2 and T2a, the unresolved HPLC peak was collected in five fractions. Fractions 1 and 2, which contain a majority of T2a (as shown by the corresponding MALDI mass spectra), and fractions 3–5, which contain almost pure T2, were pooled.

The subdigestions (room temperature, enzyme/substrate ratio 1:60, w/w, 1:210 mol/mol) of those two fractions (250 pmol of "T2a-enriched" peptide mixture and 1 nmol of almost pure T2) were monitored by MALDI MS. Spectra of the enzyme substrate mixtures were taken at different digestion times (10, 20, 35, 50 min). For the duration of the MALDI MS measurement, the digestion reaction was stopped by placing the sample in a freezer at −18 °C. After 50 min, the spectrum indicated the desired level of cleavage. The reaction was stopped at this time by addition of an equivalent volume of 0.1% TFA, and the peptides were separated on RP-HPLC.

Purification of peptides

The peptides generated by enzymatic digestion were separated by RP-HPLC on a Nucleosil 300-10-C18 column thermostated at 60 °C using an acetonitrile gradient in 0.1% TFA (see Figs. 3, 5). The HPLC system consisted of two LKB 2150 pumps (Pharmacia Biotechnology, Uppsala, Sweden), a PC controller, and an LKB 2151 variable wavelength detector operating at 214 nm.

ESI MS

ESI MS was performed using a Vestec (Houston, Texas) single quadrupole electrospray mass spectrometer. The instrument was calibrated using standard proteins of appropriate molecular masses. Samples were dissolved in 1% acetic acid, 50% methanol to a concentration of 0.1–0.6 μg/μL and 3 μL was introduced into the mass spectrometer at a flow rate of 0.5 μL/min using a syringe pump. Spectra were recorded with a scan rate of 10 s/scan with a mass window of m/z 800–2,000. An appropriate number of scans were averaged, and the molecular mass was calculated from the "multiply charged" molecular ions using the instrument software (Technivent, Maryland Heights, Missouri).

MALDI MS

MALDI mass spectra were acquired on a prototype linear time-of-flight (24 kV acceleration potential) MALDI instrument from Applied Biosystems Sweden A.B. (Uppsala). The instrument is equipped with a nitrogen laser operating at a wavelength of 337 nm, and data acquisition is performed by a LeCroy model 9450A digital oscilloscope with a sampling rate of 300 MHz. The system is controlled by a Macintosh computer.

α-Cyano-4-hydroxy cinnamic acid was used as matrix. Equal volumes of matrix (15 μg/μL in 70% acetonitrile) and sample solutions (approximately 1 μL) were mixed on the target and dried. The sample solution contained between a few hundred femtomoles to several picomoles of protein or peptide per microliter. Spectra from 20–100 laser shots were acquired and averaged.

When internal calibration using human insulin or "previously determined" peptide masses was performed, it is specified in the tables or captions. Otherwise, the theoretical calibration based on instrument parameters was used.

Edman degradation

The intact protein and selected peptides were sequenced by Edman degradation using a Knauer (Berlin) model 810 or 816 pulsed-liquid sequencer. From each sample, 50–500 pmol was immobilized on a "polybrene-coated" PVDF membrane. Degradation and conversion were performed as described by the supplier. Identification of the liberated PTH-amino acid derivatives was performed by RP-HPLC using, respectively, an ABI 140A HPLC system or a Knauer model 120A PTH analyzer, both equipped with Knauer narrow-bore columns (250 × 2 mm) for separation of PTH-amino acid derivatives.

Nomenclature

Peptides isolated from the digestion of the protein or peptide mixtures have a prefix indicating the enzyme used (T for trypsin, TC for subdigestion of a tryptic peptide by chymotrypsin), followed by a number and, in some cases, by a lowercase letter (b or a). The numbers of the peptides follow their position in the protein (or peptide in the case of subdigestion) counted from the N-terminus of the intact protein (or peptide). The peptide name is followed by a letter when the peptide originates from one of the lower abundance isoforms, b or a, of the protein. Amino acids are abbreviated using the commonly accepted one- or three-letter codes.

Electroelution of protein spots from 2D-PAGE gels obtained from single animals

Preparation and electrophoresis of cuticle extracts

The cleaned larval cuticle from each single animal was extracted in 0.5 mL 4 M urea in 0.1% TFA at 4 °C and separated

by 2D gel electrophoresis as described previously (Andersen et al., 1986) with the following minor modifications. Approximately one-fifth of the extract from one animal was loaded on the gel. Electrofocusing in the first dimension was performed in the pH region 3.5–11 for 8 h, and the voltage was increased stepwise from 150 to 500 V. Eighteen percent polyacrylamide gels were used for SDS-PAGE in the second dimension, and 0.2% mercaptoacetic acid was in some experiments added to the sample buffer to act as a scavenger for nonpolymerized acrylamide. Gels were stained in 0.12% Coomassie blue R overnight.

Electroelution

Spots corresponding to protein F1 were cut out from the gels and electroeluted one by one in a Bio-Rad model 422 electroeluter, allowing simultaneous elution of up to six samples. The elution was performed for 6 h at 10 mA/sample in a 50 mM NH_4HCO_3 volatile buffer containing 0.06% SDS. To reduce the final SDS concentration of the samples, the buffer in the reservoir was changed to 50 mM NH_4HCO_3 without SDS and the electroelution continued overnight at reduced current (7 mA/sample). The eluted protein was collected in a volume of approximately 400 μL of buffer between the frit and the 3.5-kDa cutoff membrane in the electroeluter. The final SDS concentration is difficult to evaluate but is supposed to be lower than the concentration of the first elution buffer.

Molecular weight determination and peptide mapping for protein eluted from 2D-PAGE gel spots

Tryptic digestion

Tryptic digestion of the eluted protein was performed by adding 0.5 μg of trypsin to 30 μL of protein solution from the gel eluter and allowing the reaction to proceed for 2 h at 37 °C.

MALDI MS

MALDI MS of samples eluted from gel spots was performed by depositing 0.8–1 μL of protein solution from the gel eluter onto the target followed by the same volume of first 2% TFA and then matrix solution (α-cyano-4-hydroxy cinnamic acid; 15 μg/μL in 70% acetonitrile). The sample was dried at atmospheric pressure.

Samples prepared from eluted proteins require higher laser irradiances than conventionally purified proteins. The resulting high abundance of lower mass matrix ions tends to saturate the channel plate detector with concomitant reduction of the detection efficiency of the higher mass signals. This problem is eliminated in the instrument used because the detector voltage is turned on with a time delay. This capability has been found to be essential to obtain spectra from electroeluted proteins.

Acknowledgments

This project was supported through grants from the Danish Natural Science and Technical Science Research Councils. The DANVIS program of the Danish Research Academy and the Swiss National Science Foundation are acknowledged for financial support to S.H.

References

Andersen SO, Højrup P, Roepstorff P. 1986. Characterization of cuticular proteins from the migratory locust, *Locusta migratoria*. *Insect Biochem 16*:441–447.

Andersen SO, Højrup P, Roepstorff P. 1994a. Insect cuticular proteins. *Insect Biochem Mol Biol*. Forthcoming.

Andersen SO, Rafn K, Krogh TN, Højrup P, Roepstorff P. 1994b. Comparison of larval and pupal cuticular proteins in *Tenebrio molitor*. *Insect Biochem Mol Biol*. Forthcoming.

Andreasen L, Højrup P, Andersen SO, Roepstorff P. 1993. Combined plasma desorption mass spectrometry and Edman degradation applied to simultaneous sequence determination of isoforms of structural proteins from cuticle of *Locusta migratoria*. *Eur J Biochem 217*:267–273.

Bouhin H, Charles JP, Quennedey B, Delachambre J. 1992. Characterization of a cDNA clone encoding a glycine-rich cuticular protein of *Tenebrio molitor*: Developmental expression and effect of a juvenile hormone analogue. *Insect Mol Biol 1*:53–62.

Charles JP, Bouhin H, Quennedey B, Courrent A, Delachambre J. 1992. cDNA cloning and deduced amino acid sequence of a major, glycine-rich cuticular protein from the coleopteran *Tenebrio molitor*. Temporal and spatial distribution of the transcript during metamorphosis. *Eur J Biochem 206*:813–819.

Chiari M, Righetti PG, Negri A, Ceciliani F, Ronchi S. 1992. Preincubation with cysteine prevents modification of sulfhydryl groups in proteins by unreacted acrylamide in a gel. *Electrophoresis 13*:882–884.

Hall SC, Smith DM, Masiarz FR, Soo VW, Tran HM, Epstein LB, Burlingame AL. 1993. Mass spectrometric and Edman sequencing of lipocortin I isolated by two-dimensional SDS/PAGE of human melanoma lysates. *Proc Natl Acad Sci USA 90*:1927–1931.

Klarskov K, Roecklin D, Bouchon B, Sabatié J, Van Dorsselaer A, Bischoff R. 1994. Analysis of recombinant *Schistosoma mansoni* antigen tSmp28 by on-line liquid chromatography–mass spectrometry combined with sodium dodecyl sulfate polyacrylamide gel electrophoresis. *Anal Biochem 216*:127–134.

Le Maire M, Deschamps S, Møller JV, Le Caer JP, Rossier J. 1993. Electrospray ionization mass spectrometry on hydrophobic peptides electroeluted from sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Application to the topology of the sarcoplasmic reticulum Ca^{2+} ATPase. *Anal Biochem 214*:50–57.

Pau RN, Weaver RJ, Edwards-Jones K. 1986. Regulation of cockroach oothecin synthesis by juvenile hormone. *Arch Insect Biochem Physiol Suppl 1*:59–73.

Regier JC, Kafatos FC. 1985. Molecular aspects of chorion formation. In: Kerkut GA, Gilbert LI, eds. *Comprehensive insect physiology biochemistry and pharmacology*. Oxford, UK: Pergamon Press. pp 113–152.

Roepstorff P, Højrup P. 1993. A general strategy for the use of mass spectrometric molecular weight information in protein purification and sequence determination. In: Imahori K, Saiyama F, eds. *Methods in protein sequence analysis*. New York: Plenum Press. pp 149–156.

Talbo G, Højrup P, Rahbek-Nielsen H, Andersen SO, Roepstorff P. 1991. Determination of the covalent structure of an N- and C-terminally blocked glycoprotein from endocuticle of *Locusta migratoria*. Combined use of plasma desorption mass spectrometry and Edman degradation to study posttranslationally modified proteins. *Eur J Biochem 195*:495–504.