



Pergamon

Insect Biochemistry and Molecular Biology 33 (2003) 331–343

*Insect  
Biochemistry  
and  
Molecular  
Biology*

www.elsevier.com/locate/ibmb

# Characterization and cDNA cloning of three major proteins from pharate pupal cuticle of *Manduca sexta*

R.J. Suderman\*<sup>a</sup>, S.O. Andersen\*<sup>b</sup>, T.L. Hopkins<sup>c,\*</sup>, M.R. Kanost<sup>a</sup>, K.J. Kramer<sup>d</sup>

<sup>a</sup> Department of Biochemistry, Kansas State University, Willard Hall, Manhattan, KS 66506-3706, USA

<sup>b</sup> August Krogh Institute, Copenhagen University, Universitetsparken 13, DK-2100 Copenhagen Ø, Denmark

<sup>c</sup> Department of Entomology, Kansas State University, 123 West Waters Hall, Manhattan, KS 66506-4004, USA

<sup>d</sup> Grain Marketing and Production Research Center, Agricultural Research Service, U.S. Department of Agriculture, Manhattan, KS 66502-2736, USA

Received 29 July 2002; accepted 22 November 2002

## Abstract

Three proteins, MsCP20, MsCP27 and MsCP36, that are secreted in greatest quantity into the pharate pupal cuticle of *Manduca sexta* (Hopkins et al., 2000) were purified and their amino acid sequences determined by mass spectrometry and Edman degradation. Although these proteins become sclerotized and insoluble in the pupal exoskeleton, their sequences contain features characteristic for proteins occurring in less sclerotized pliable cuticles, such as arthrodial membranes and soft larval cuticles. These proteins carry a secondary modification attached to a threonine residue, presumably an *O*-linked sugar moiety. cDNA clones of the genes for MsCP20, MsCP27 and MsCP36 were constructed from pharate pupal integument RNA. Close agreement was found between the amino acid sequences determined by Edman degradation and sequences deduced from the cDNA clones. The molecular masses determined by protein sequencing for MsCP20, MsCP27, and MsCP36 were 17,713, 17,448, and 29,582 Da, respectively, in close agreement with the masses deduced from the corresponding cDNA clones (17,711, 17,410, and 29,638 Da). Temporal expression analysis indicates that MsCP20 and MsCP36 transcripts are present at low levels early in the fifth larval stadium, followed by a large increase in abundance prior to pupal ecdysis. MsCP27 was not detected during development of the fifth larval instar, but its transcript, like those of MsCP20 and MsCP36, increased to a peak level just before pupal ecdysis. Only the MsCP36 transcript was detected in adults. These results support the hypothesis that these proteins are synthesized by the epidermis and are subsequently deposited into the cuticle during the larval–pupal transformation of *M. sexta* where they become sclerotized in the formation of pupal exocuticle.

© 2003 Elsevier Science Ltd. All rights reserved.

**Keywords:** Cuticle; Sclerotization; Tobacco hornworm; *Manduca sexta*; Insect; Lepidoptera; cDNA; Amino acid sequence; Nucleotide sequence; RNA expression; Epidermis; Development; Protein; Gene

## 1. Introduction

Proteins are a major structural component of the cuticular exoskeleton of insects, and they vary widely in numbers and properties in different types of cuticles (Andersen et al., 1995; Willis, 1996). Flexible and pliant cuticles such as found in soft bodied larvae and arthrodial membranes are largely unsclerotized and their proteins remain relatively soluble. Cuticles in sclerotized

regions of the exoskeleton are hardened and stiffened due to tanning reactions in their exocuticular layer by which the proteins become insoluble. Proteins that form exocuticle are mainly secreted before ecdysis into pharate or pre-ecdysial cuticle and they differ often in amino acid sequences and chemical properties from those proteins secreted after ecdysis, which make up the endocuticle (Roberts and Willis, 1980; Andersen and Højrup, 1987; Bærnholdt and Andersen, 1998; Hopkins et al., 2000). The functional and physical properties of each type of cuticle are therefore related in part to the types and quantities of proteins that make up their structure, as well as their subsequent interactions with other cuticular components such as chitin and the quinone metabolites

\* Corresponding author. Tel.: +1-785-532-4722; fax: +1-785-532-6232.

E-mail address: thopkins@ksu.edu (T.L. Hopkins).

of catechols (Hopkins and Kramer, 1992; Andersen et al., 1996; Kramer et al., 2001).

In the tobacco hornworm, *Manduca sexta*, nearly 100 proteins are secreted into pre-ecdysial or pharate pupal cuticle, which become largely insoluble as the cuticle becomes sclerotized (Hopkins et al., 2000). Of these, three major proteins, which according to their migration during sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) were called MsCP20, MsCP27 and MsCP36, were partially characterized as to N-terminal amino acid sequences and amino acid compositions. The sequences showed no close correspondence to proteins of known amino acid sequences, but their amino acid compositions did exhibit certain similarities to other cuticular proteins.

Because of the importance of these proteins in the formation of the pupal exocuticle and the sclerotization process, we have now determined their complete amino acid sequences by Edman degradation, and their deduced sequences from the cloned cDNAs. Temporal expression of their genes in abdominal integument in the last larval stadium, pharate pupal development, and the adult was followed by analysis of mRNA transcripts. All three genes were expressed at high levels in pharate pupal integument. MsCP20 and MsCP36 transcripts were present at low levels at the beginning of the fifth larval stadium, while MsCP36 was also detected in adults. MsCP27 transcript was detected only in the pupal stage.

## 2. Materials and methods

### 2.1. Purification of cuticular proteins

*M. sexta* larvae were reared as described by Bell and Joachim (1976) at 27 °C with a photoperiod of 16L:8D. Insects were collected as pharate pupae showing initial patterns of sclerotization within a few hours of ecdysis and frozen at –20 °C. Dissection and cleaning of the slightly brownish cuticle was performed as described previously (Hopkins et al., 2000).

The dried pharate cuticle (ca. 100 mg) was extracted in 15 ml 6 M urea, 0.1% trifluoroacetic acid (TFA) overnight at 4 °C. After filtration, the extract was subjected to gel filtration chromatography on a column of Sephacryl S200HR (2.6 × 90 cm, Pharmacia-Biotech, Uppsala, Sweden) using a mobile phase of 0.1% TFA. Fractions of ca. 8 ml were collected, and aliquots from the protein-containing fractions were subjected to one-dimensional SDS-PAGE on 15% gels (Andersen and Højrup, 1987). Fractions were pooled according to their content of major proteins.

### 2.2. Purification by RP-HPLC

Pure proteins were obtained by reverse phase high performance liquid chromatography (RP-HPLC) on a

Source 5RPC column (Pharmacia-Biotech, Uppsala, Sweden) of samples from the pooled protein fractions. The column was equilibrated with 10% acetonitrile, 0.1% TFA in ultrahigh quality (UHQ) water (A-buffer). The B-buffer was 70% acetonitrile, 0.1% TFA in UHQ water. During the first 2 min of elution, the concentration of B-buffer was increased from 0 to 30%, and during the following 40 min, it was increased linearly to 100%. The major fractions were collected manually and taken to dryness using a rotary evaporator.

Peptides obtained by digesting the proteins with proteases were separated by RP-HPLC on a Vydac C<sub>4</sub>-column as described previously (Andersen, 2000). The purity of the collected proteins was checked by one-dimensional SDS-PAGE and by matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS).

### 2.3. Enzymatic digestion

To obtain peptides of a reasonable size for sequencing studies, the proteins were digested with one or more of the following proteolytic enzymes: trypsin and endoproteinase Glu-C from Promega (Madison, WI), chymotrypsin and endoproteinase Asp-N from Boehringer (Mannheim, Germany). The digestions were performed as described previously (Andersen, 1998).

### 2.4. Mass spectrometry

MALDI-MS was performed at the Institute of Molecular Biology and Biochemistry, Odense University, as described previously (Jensen et al., 1998).

Plasma desorption mass spectrometry (PDMS) was performed on a BioIon 20 time-of-flight mass spectrometer (BioIon, Uppsala, Sweden). Peptides were dissolved in 0.1% TFA at a concentration of approximately 50 pmol/μl, and 5 μl applied to a nitrocellulose covered target (Roepstorff, 1993). Spectra were acquired for 0.1–1 × 10<sup>6</sup> fission events at a 15 kV acceleration voltage and calibrated based on the H<sup>+</sup> and Na<sup>+</sup> ions.

### 2.5. Sequence determination

The intact proteins and selected peptides were sequenced by Edman degradation using an Applied Biosystems 476A Protein Sequencer. Degradation, conversion and identification of the liberated phenylthiohydantoin amino acids were performed as described by the supplier.

### 2.6. RNA isolation

The dorsal integument (epidermis and cuticle) of bar stage pharate pupae (when localized sclerotization of thoracic cuticle was visible a few hours before ecdysis) was carefully dissected, and total RNA was isolated using a

Glassmax RNA Microisolation Spin Cartridge System (Life Technologies) or a Stratagene RNA Isolation Kit. Poly(A) RNA was purified using a Poly(A) Quik mRNA Isolation Kit (Stratagene). RNA used for northern blot analysis was prepared from the epidermis of fourth and fifth instar larvae, pupae, and adults using the method of Chomczynski and Sacchi (1987). Three insects were dissected and the RNA extracted was pooled for each data point. To facilitate the homogenization of the tissue, each sample was frozen in liquid nitrogen and then ground to a powder with a mortar and pestle prior to RNA isolation.

## 2.7. Isolation of cDNA clones

Degenerate primers were designed based on the amino acid sequences of cuticle proteins MsCP20, MsCP27 and MsCP36, and were used in reverse transcription-polymerase chain reaction (RT-PCR) experiments (Life Technologies Superscript OneStep RT-PCR System) to generate partial cDNA clones corresponding to each protein. Samples of 100 ng of RNA were reverse transcribed and then amplified by PCR using degenerate primers 552(+) and 553(–) to generate a 334 bp partial MsCP27 clone, and degenerate primers 550(+) and 551(–) to generate a 194 bp partial MsCP20 clone (see Table 1 for

primer sequences). RT-PCR was performed with primer concentrations of 0.2  $\mu$ M in a total volume of 25  $\mu$ l as follows: 50 °C, 30 min; 94 °C, 2 min; 35 cycles of 94 °C, 15 s; 50 °C, 30 s; 72 °C, 1 min; followed by a 10 min hold at 72 °C. After separation by agarose gel electrophoresis, no clear band of expected size could be detected. The reaction products (1  $\mu$ L) were then used as templates for a second PCR using the same primers as follows: for MsCP20, 35 cycles of 94 °C, 15 s; 45 °C, 30 s; 72 °C, 1 min; followed by a 10 min hold at 72 °C, for MsCP27, 45 cycles of 94 °C, 15 s; 50 °C, 30 s; 72 °C, 1 min; followed by a 10 min hold at 72 °C. In these reactions the primer concentrations were raised to 1.7  $\mu$ M and the Mg<sup>2+</sup> concentration raised from 1.2 to 1.9 mM. Separation by agarose gel electrophoresis of the MsCP27 reaction yielded a single band of expected size. To obtain a single band of expected size of MsCP20, the reaction product (1  $\mu$ L) from the MsCP20 PCR was used as a template for an additional PCR using the same primers as follows: 35 cycles of 94 °C, 15 s; 45 °C, 30 s; 72 °C, 1 min; followed by a 10 min hold at 72 °C. After separation by low melting agarose gel electrophoresis, the cDNA band was recovered and cloned into the pGEM-T vector (Promega). The two partial clones were sequenced using the dideoxynucleotide

Table 1  
Oligonucleotide primers used for polymerase chain reactions of *M. sexta* cuticle protein cDNAs

Primer name	Gene	Position	5'→3' Primer sequence
501(–)	RP-S3	603–619	GCCGTTCTTGCCCTGTT
504(+)	RP-S3	221–237	CGCGAGTTGACTTCGGT
550(+)	CP20	469–485	GAYGGNCARCARATHGC
551(–)	CP20	658–671	ATYACYATNGGYAA
552(+)	CP27	199–218	CARCARGAYGCNGARAARAA
553(–)	CP27	551–570	AAYAARAGNCCYACYATRAC
555(+)	CP36	436–452	TTYGCNTAYGAYTTYGA
556(+)	CP27	199–218	CAGCAGGATGCCGAGAAAAA
557(–)	CP27	479–498	GCCAGCAGCCTCATCTCTCT
560(–)	CP36	1019–1035	YTGNCRTRRAARTTNC
561(–)	CP20	558–579	AATGGAGCGCAGGATCTCCTCT
563(+)	CP36	520–542	TTYGCNTAYAARGGNGAYGAYGG
564(–)	CP36	682–699	RTANGTNCRCRTCTCNAC
565(–)	CP20	628–656	TAGCCGTTGGAGCCGCCGAGCCGCCATT
566(–)	CP27	294–321	TGGGCCTCCCGCCTCCTCAGCTCGGATA
567(+)	CP27	296–324	TCCGAGCTGAGGAGGCGGGAGGCCAGAA
569(+)	CP36	619–647	ACCCCGCCCCGATCCCTGAAGAGATCCT
570(–)	CP36	583–610	TGGGGAGGTGAGCGCCTTGGGACACGAA
571(–)	CP36	552–581	CCGTCTCGTCAGCAGTGTAGGTGATGCTA
572(+)	CP36	641–669	AGTCTTTGGAACAGAACGCTCGCGACGAA
575(+)	CP36	552–572	TAGCATCACCTACACTGCTGA
576(–)	CP36	661–681	AATACCGGACGCTTCGTGCGG
578(+)	CP20	255–276	GGCTGGCGCTTACGGAGGATCA
580(+)	CP36	1–21	GCACACAGCATCCGAACCACT
GR5P			CGACTGGAGCACGAGGACACTGA
GR5NP			GGACACTGACATGGACTGAAGGAGTA
GR3P			GCTGTCAACGATACGCTACGTAACG
GR3NP			CGCTACGTAACGGCATGACAGTG

See Sections 2.7 and 2.8 for more details and Fig. 7 for diagram

method (ABI Prism Big-Dye Terminator Cycle Sequencing Ready Reaction, PE Applied Biosystems).

To generate a 180 bp partial MsCP36 clone, RT-PCR (Life Technologies Superscript OneStep RT-PCR System) was carried out on 30 ng mRNA, using 2  $\mu$ M degenerate primer pair 555(+) and 560(–) as follows: 45 °C, 30 min; 94 °C, 2 min; 35 cycles of 94 °C, 20 s; 47 °C, 30 s; 72 °C, 1 min; followed by a 10 min hold at 72 °C. The following successive rounds of PCR were carried out to generate a single cDNA band of expected size using agarose gel electrophoresis. The reaction product (1  $\mu$ L) was used as a template for PCR using the same primers as follows: 95 °C, 1 min; 35 cycles of 95 °C, 30 s; 45 °C, 30 s; 70 °C, 1 min, followed by a 10 min hold at 70 °C. The reaction product (1  $\mu$ L) was used as a template for PCR using the same primers, increasing the annealing temperature to 48 °C and decreasing the number of cycles to 25. The reaction product (1  $\mu$ L) was used as a template for PCR using the same primers and increasing the annealing temperature to 52 °C. This reaction product was used as a template for PCR using nested degenerate primers 563(+) and 564(–) for 25 cycles of 95 °C, 30 s; 54 °C, 30 s; 70 °C, 1 min, followed by a 10 min hold at 70 °C. A cDNA band of expected size was recovered and cloned into the pGEM-T vector (Promega), and sequenced as described above.

The 194 bp partial MsCP20 clone was used to create an ( $\alpha$ -<sup>32</sup>P)-dCTP labeled probe to screen an *M. sexta* fifth instar bar-stage pharate pupal integument cDNA library (Clontech SMART cDNA Library Construction Kit). The library was screened using standard procedures according to Sambrook and Russell (2001). A 600-bp clone was recovered, sequenced as described above, and found to contain the 3' portion of the cDNA, including the poly(A) tail.

## 2.8. 5'- and 3'- RACE

The 5'-ends of MsCP20, MsCP27, and MsCP36 and the 3'-ends of MsCP27 and MsCP36 were obtained using rapid amplification of cDNA ends (RACE) (Invitrogen GeneRacer Kit). Briefly, 5  $\mu$ g of RNA was 5'-dephosphorylated using calf intestinal phosphatase. The RNA was then treated with tobacco acid phosphatase to remove the 5'-cap structure from intact, full-length mRNA, yielding 5'-phosphates on only full-length mRNA. The GeneRacer RNA Oligo was then ligated to the 5'-end of the mRNA using T4 ligase. A single stranded cDNA pool was created by reverse transcribing the RNA using the GeneRacer OligodT Primer and avian myeloblastosis virus reverse transcriptase. This cDNA pool served as template for subsequent PCR reactions using gene-specific primers near the 5'- (or 3'-) ends of the partial clones and the GeneRacer 5'- (or 3'-) oligo specific primer(s).

To obtain the 5'- and 3'-ends of the partial clones, 1

ng of single stranded cDNA per reaction was amplified by PCR using Platinum Taq High Fidelity DNA Polymerase (Invitrogen) in a total volume of 50  $\mu$ l under the following conditions: 94 °C, 2 min; 5 cycles of 94 °C, 30 s; 72 °C, 3 min; 5 cycles of 94 °C, 30 s; 70 °C, 3 min; and 25 cycles of 94 °C, 30 s; 68 °C, 30 s; 72 °C, 3 min; followed by 72 °C, 10 min. The primers (Table 1) 565(–) and GeneRacer 5'-primer (GR5P) were used for the 5'-end of MsCP20, 566(–) and GR5P for the 5'-end of MsCP27, 570(–) and GR5P for the 5'-end of MsCP36, and 567(+) and GeneRacer 3'-primer (GR3P) for the 3'-end of MsCP27. The reaction product for the 3'-end of MsCP27 yielded a band of expected size on low melting agarose gel, which was recovered, cloned into the pCR4-TOPO vector (Invitrogen), and sequenced as described above. The remaining 5'- and 3'-RACE experiments did not yield single cDNA bands of expected size on agarose gels. Subsequently, the reaction products (1  $\mu$ L) were used as templates for PCR (same temperature profile as before, except changing the final annealing temperature to 65 °C and lowering the extension temperature to 68 °C) using the following nested primer sets: primers 561(–) and GeneRacer 5'-nested primer (GR5NP) for the 5'-end of MsCP20, 566(–) and GR5NP for the 5'-end of MsCP27, and 571(–) and GR5NP for the 5'-end of MsCP36. These reaction products were subjected to electrophoresis on a low melting agarose gel and bands of expected size were recovered, cloned into the pCR4-TOPO vector (Invitrogen), and sequenced as described above. The 5' untranslated region of the MsCP36 5' RACE fragment was used to design primer 580(+), which was then used with primer GR3P for 3' RACE using 1 ng of single stranded cDNA as template. The full length sequence was amplified using Platinum Taq High Fidelity DNA Polymerase (Invitrogen) as follows: 94 °C, 2 min; 5 cycles of 94 °C, 30 s; 72 °C, 3 min; 5 cycles of 94 °C, 30 s; 68 °C, 3 min; and 35 cycles of 94 °C, 30 s; 61 °C, 30 s; 68 °C, 3 min; followed by 72 °C, 10 min. The reaction product was cloned and sequenced as described above.

## 2.9. Northern analysis of RNA

Samples of total RNA (20  $\mu$ g) were separated by electrophoresis in a 1.5% agarose gel containing formaldehyde according to Sambrook and Russell (2001). The RNA was transferred onto neutral nylon membrane (GeneScreen) and hybridized with an ( $\alpha$ -<sup>32</sup>P)-dCTP labeled cDNA probe constructed using the Multiprime DNA Labeling System (Amersham Pharmacia Biotech) with the appropriate cuticle protein cDNA as template. The membranes were washed under high stringency (1  $\times$  SSC, 0.1% SDS, 66 °C) and exposed to X-ray film for either 1 or 4 days.

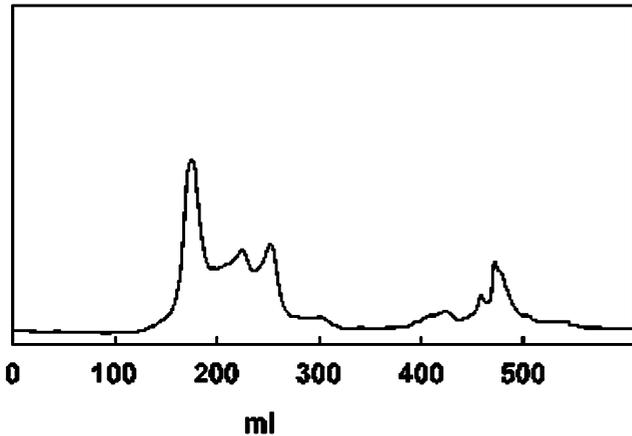


Fig. 1. Fractionation of extract from 100 mg dried pharate *Manduca sexta* pupal cuticle by gel-filtration on Sephacryl S200HR. Absorbance at 280 nm was recorded.

### 3. Results

#### 3.1. Purification

Carefully cleaned cuticles from pharate pupae of *M. sexta* were extracted by 6 M urea, and the extract was subjected to gel-filtration on a column of Sephacryl S200HR. The absorption of the eluate at 280 nm is shown in Fig. 1. Fig. 2 shows the protein pattern obtained by SDS gel-electrophoresis of aliquots from fractions 21 to 40 (corresponding to 154–308 ml in Fig. 1). The major cuticular protein, MsCP36, (Hopkins et

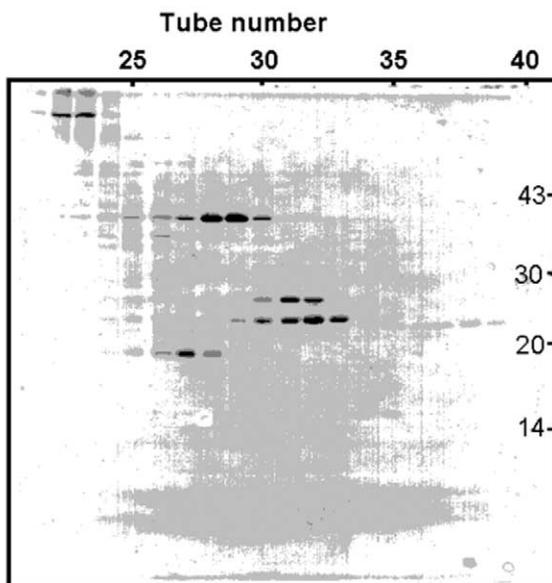


Fig. 2. One-dimensional SDS-PAGE of aliquots from fractions 21 to 40 obtained by gel-filtration of cuticular extract; the fractions correspond to 154–308 ml in Fig. 1. The positions and sizes (kDa) of protein markers are shown on the right-hand side of the gel.

al., 2000) is present in fractions 28–29, whereas MsCP20 and MsCP27 are present in fractions 31–33. Samples from these fractions were subjected to RP-HPLC (Fig. 3), and the major peaks were collected, each representing a pure protein according to analysis by SDS-electrophoresis and MALDI mass spectrometry. By MALDI mass spectrometry the mass of MsCP36, MsCP27, and MsCP20 was determined to be 29,771, 17,623, and 17,894 Da, respectively.

#### 3.2. Sequence determination

MsCP36: The first 41 residues from the N-terminus of MsCP36 were obtained by Edman degradation. After digestion of the protein with trypsin, 17 peptides were separated by RP-HPLC. Most of the peptides had molecular masses below 3000 and were sequenced completely without problems, and good agreement between measured and calculated masses was observed. How-

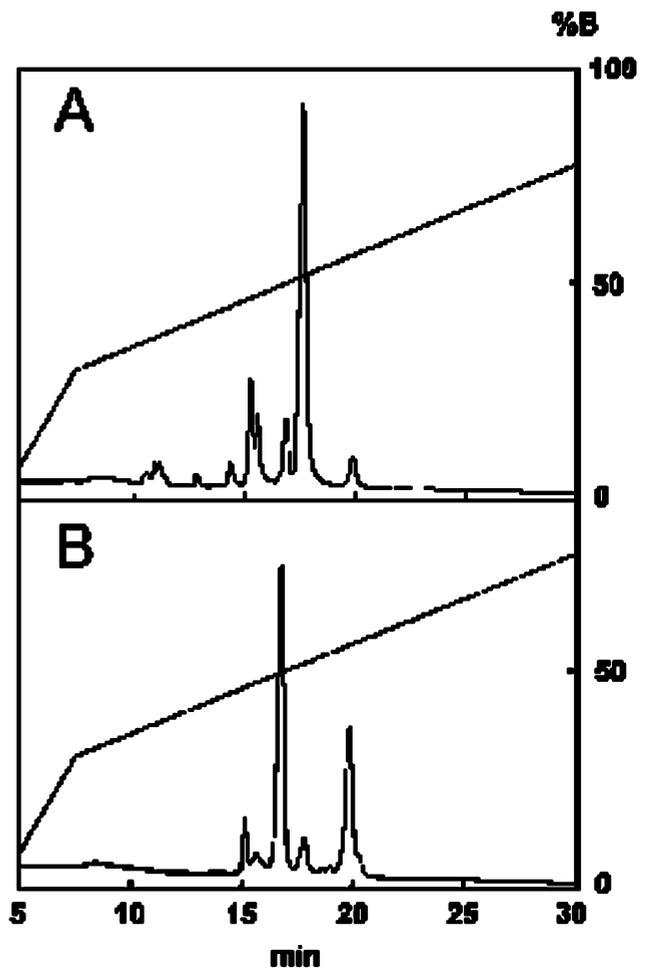


Fig. 3. A. RP-HPLC on Vydac C4 of aliquot from tube 29 (Fig. 2). Protein MsCP36 elutes at 17–18 min B. RP-HPLC on Vydac C4 of aliquot from tube 31 (Fig. 2). Protein MsCP27 elutes at 16–17 min, and MsCP20 elutes at 19–20 min.

ever, two of the tryptic peptides, T4 and T8, were problematic.

A mass of 5652 was obtained by PDMS for peptide T4 (residues 29–99 in the complete sequence), and only the first 30 residues could be determined unambiguously by Edman degradation, presumably due to an abundance of glycine in the peptide. Peptide T4 contains no cleavage sites for endoproteases Glu-C and Asp-N but does possess several potential cleavage sites for chymotrypsin. The chymotryptic peptides were purified, their masses were determined by PDMS and their sequences by Edman degradation. Calculated and measured masses for the individual chymotryptic peptides from T4 were in agreement, but only eight peptides were obtained, and all of the peptide masses did not add up to the measured mass of the parent peptide, indicating that some chymotryptic peptides were missing or occurring as repeated peptides in the parent peptide. This question was answered by partial chymotryptic digestion of T4 and mass determination of the resulting peptides. Comparison between the masses from the incomplete digest and the masses of the peptides from the complete chymotryptic digest combined with Edman degradation of the various peptides showed that T4 contains three consecutive copies of the peptide GGAGSGF, and the order of the chymotryptic peptides was then established. It was also observed that, besides the chymotryptic peptide GGGAGGAF, a small amount of a related peptide, GGGAGGGF, was present, indicating the possible occurrence of an isoform of the protein.

The other tryptic peptide to present problems was T8 (residues 155–191 in the complete sequence), which according to MALDI-MS has a relative mass of 4187. It was sequenced completely, and from the sequence its mass was calculated to be 3999, differing from the measured value by 188 mass units. Comparing the yields of phenylthiohydantoin obtained in the individual cycles during the Edman degradation of the peptide showed that the threonine residue in position 181 gave a very low yield, indicating that this residue may be conjugated, perhaps with a sugar moiety.

Protein MsCP36 was also digested with chymotrypsin and endoprotease Asp-N, and the resulting peptides were purified, mass determined and sequenced. The results confirmed the sequences obtained from the tryptic peptides and they provided sufficiently long overlaps to the tryptic peptides to allow them to be unambiguously arranged in order to give the complete sequence for the protein as shown in Fig. 4A. The sequence was confirmed by comparing the masses of the peptides obtained by endoprotease Glu-C digestion to the masses expected from the sequence. From the complete amino acid sequence, the mass of MsCP36 was calculated to be 29,582, which is 189 mass units lower than that determined by MALDI mass spectrometry.

MsCP20: The sequence of MsCP20 was determined

#### MsCP36

##### A.

DRLDNKYLFP RGNAGAGFGP GFGGAGPKGG GSGFPGAGSG FGGAGSGFGG  
AGSGFGGGAG GAFGHGGAGS GFGGQAGGYS GAGGAGGYSG GASQYSRGR  
GAASADANAQ ILRLNNEVTA EGFAYDFETS NGRIRADAQGV ATNGVQSQGS  
FAYKGGDDGD YSITYTADEN GFVPQGAHLP TFPPIIEEIL KSLEQNARDE  
AAGIVDDGTY RGEAGAGGA GYSYGAGGYS GGAGGAGGFG AGGAGRAGGA  
GGFAGGAGR AGGAGGFGAG GAGGAGGAGG FGAGAGGFGG RSGGFGGAA  
SRQYLAPNAG GRGSGSGNFN AQTGYQY

##### B

```

1      gcacacagcatccgaaccactaacatgaactgtctgctttagccgctgctcctcggtgtc
      M K L F V L A A V L G V
61      tgctggcagatgctcgacacaacgtacctcggccagagggcaagcagcgccggc
      C L A D R L D N K Y L P P R G N A G A G A
121     ttccggcctggctcggagagcgggactaaagggggcgccagcggctcggagggcgt
      F G P G F G G A G P K G G G S G F G G A
181     ggctcgggttccggaggtccggatcaggttttggcggcgcggtatcaggtcttggcgt
      G S G F G G A G S G F G G A G S G F G G
241     ggtccggagggccttttggccatggagggcgggatctggcttcggcggccaggtgga
      G A G G A F G H G G A G S G F G G Q A G
301     ggatcacagcggcctggcggagcggcgctcacagcgggtggtgctgacggacaatcacgc
      G Y S G A G G A G G Y S G G A S G Q Y S
361     ggccaggtggcggcctcctcgtgacgcaacggcgaatcctcagactgaacaatgag
      G R G G A A S A D A N A Q I L R L N N E
421     gtgacagcgaaggttccgctacgactccgacactccaacggcctccgctgacgct
      V T A E G F A Y D F E T S N G I R A D A
481     caggagtgccccaacggcgtgagctcggcggcagctcggctacaagggcgagdc
      Q G V A T N G V Q C G Q G S F A Y K G D D
541     ggccaagactacagcatcactcactgctgacgagacgggttctgctcccaagggcgt
      G Q D Y S I T Y T A D E N G F V P Q A
601     cacctcccaccccccccgatccctgaggagatcctcaagtcttggaaacaggaact
      H L P T P P P I P E E I L K S L E Q N A
661     cgcgacgaagctcgggtatttggcagtgacttaccggcggtaaggtcgggtgct
      R D E A A G I V D D G T Y R G E G A G
721     ggcgtcgtggcggctactcgggtgctggagatactctggagggcgtggtggtccgga
      G Y A G G Y S G A G G Y S G G A G G A G
781     ggctcgggtgctggaggtgctggccgctggagggcgtgagggctcggcaccggcgt
      G F G A G G A G R A G G A G G F G G
841     gctggtcggcggagggcgtggaggtcggcggcgggtgctgctggtgaggtggagc
      A G R A G G A G G F G A G G A G G A G G
901     gccggcgggttccgctgctggcggcgggtggttccgagggcgggtgattcgggt
      A G G P G A S A G G F G G R G S G F G
961     ggagcggcctcggcagctacctcggcccaacggcggcggcggcggcggcggcggc
      G A A S R Q Y L A P N A G G R G S G S G
1021    aactcaacggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggc
      N F N A Q T G Y Q Y
1081    gatgatgactgccaatgatgtgataaatttaacttccggacggatctaaactatctc
1141    ccggcgttggtaagttggccatgctgctgctcggcggcggcggcggcggcggcggcggc
1201    tttcaaatcattatttattgcatcttcttccccatcgtccccccggcggcggcggcggc
1261    cctctctcgtcagctcggcggcggcggcggcggcggcggcggcggcggcggcggcggc
1321    aaaaagttgtatagatcttcaacttcaatttggatctcaaaaatttcaaaaatttc
1381    atcttctcgtgttaataatagttgttaagttcttgtatatacgaacaataagggacag
1441    tagtagtgaagtaaggaagttaagtaaaaatgtataaaaatgtatagttatgtttaa
1501    ggattgtaatacagagatgactcaacgaatatacagacttaaaagcaaaaaaaaaaaaa
1561    aa

```

Fig. 4. A. Amino acid sequence of MsCP36 determined by mass spectrometry and Edman degradation. B. Nucleotide and deduced amino acid sequence of MsCP36 cDNA. The deduced amino acid sequence is shown below the cDNA sequence. The putative signal sequence is shaded. Bold residues denote conserved residues of the R&R consensus sequence (Rebers and Riddiford, 1988). Underlined residues denote differences with the amino acid sequence obtained from protein sequencing. This sequence has been deposited in the GenBank database, with accession number AF526296.

by the same strategy as that of MsCP36. The first 41 residues from the N-terminus were determined by Edman degradation of the intact protein. Trypsin digestion resulted in six peptides, one of which (T1) consisted of only two residues, and three could be sequenced completely without problems. Peptide T2 (residues 3–96) was too large, however, to be sequenced in a single run; its mass was 8631 according to MALDI. Samples of the peptide were digested either partially or completely with chymotrypsin, and by combining the results from mass determinations and sequence studies of peptides from the

two digests, it was possible to derive the complete sequence of peptide T2.

Peptide T4 could be sequenced completely, but the relative mass calculated from the sequence (4036) was 184 units lower than the mass determined by MALDI-MS (4220). It contains a threonine residue (position 142 in the intact protein), which gave a very low yield during sequencing, indicating that it may be conjugated.

Digestion of the intact protein with chymotrypsin and endoproteinase Asp-N provided the peptides necessary for arranging the tryptic peptides in an unambiguous order to give the complete sequence of the protein (Fig. 5A). The sequence was confirmed by comparing the masses of the peptides obtained by endoproteinase Glu-C digestion to the masses expected from the sequence. From the complete amino acid sequence the mass of MsCP20 was calculated to be 17,713, which is 181 mass units lower than that determined by MALDI mass spectrometry.

MsCP27: The sequence of MsCP27 was obtained by the same strategy used for MsCP20 and MsCP36—digestion with trypsin followed by mass determination and Edman degradation of the tryptic peptides, and

digestion of the protein with endoproteinase Asp-N and chymotrypsin provided the missing parts of the sequence and made it possible to arrange the tryptic peptides in the correct order. The tryptic peptide, T2, was obtained as two variants, differing in mass by 10 units, and Edman degradation showed that one of them contained a proline residue and the other a serine residue in position 9 in the peptides (corresponding to position 15 in the protein), which indicates that, in our hornworm colony, there are two allelic forms of this cuticle protein present. The mass of the tryptic peptide, T7, as determined by mass spectrometry (4898) was 188 mass units heavier than that calculated from the sequence (4710), and the yields of PTH-threonine obtained by Edman degradation was low, indicating that the threonine residue at position 116 in the protein was conjugated. The peptides obtained by digestion of the protein with the other proteases and containing threonine-116 all showed a similar discrepancy between calculated and measured masses. The complete sequence is shown in Fig. 6A. From the complete amino acid sequence the mass of MsCP27 was calculated to be 17,448, which is 175 mass units lower than that determined by MALDI mass spectrometry.

**MsCP20**

**A.**  
 GRLEPQYLFP RPNNGFVGGA GSGGNAGLGG GAGGSGAGAG GFGGSGAGAY  
 GSGGSGSHGG AQIPITSFEN VNNGDGSYHF SYETGNGISA QESGAPRAG  
 PEGLAVTAEQ AFSYRTPDQG QIALTYTADE NGFHPVGPLL PTPPIIEEI  
 LRSIEFNQRN PSSEGSYNGG NNGSGGNGY HY

**B.**

```

1      aaaaatcagctgacagctgaacaaggagcgctgtcctattagatcaatacacaactatc
61     aatatggcaaccactactgttgggttaactgcagctcctgatcgcggtgggctcgtgcggt
      M A T I L L V L T A V L I A V G S C G 1
121    cgtttggagccgcagtagctccccctcgccctgggaacgggttggaggtggcgagcgc
      R L E P Q Y L P P R P G N G F G G G A G 21
181    agcggggtaatgctggtttaggtggagagctggaggtcaggtgctggagctgggtgga
      S G G N A G L G G G A G G S G A G A G G 41
241    tttggcgatcagggctgcgcttacgggagatcaggcgctccaccggcggtggtgct
      F G G S G A G A Y G G S G G S H G G G A 61
301    caaatccattactagtttggagaactgaaacacggcgatgctccttaccattcaagt
      Q I P I T S F E N V N N G D G S Y H F S 81
361    tatgaaacgggtaacggcatctcagcccaggaggtggagctcctcggcgctgggct
      Y E T G N G I S A Q E S G A P R A P G P 101
421    gagggctcggcagtaaacgctgagggcgctctcgtaccgactccggagcggccagcag
      E G L A V T A E G A F S Y R T P D G Q Q 121
481    atcgcaactcacctacacggcagtagagacggctccaccctttgggacctcctcccc
      I A L T Y T A D E N G F H P L G P H L P 141
541    accctcccccaatcccagaggagatcctcgtcctcattgagtlcaacagacagaatcct
      T P P P I P E E I L R S I E F N R Q N P 161
601    tcttctgaaggatcacaacggaggaaatggcggtcctccggcgctccaacggctaccac
      S S E G S Y N G N G G S G S G S N N G Y H 181
661    tactgatctcattcaactaatttgcgttaattcaccacaaatgcaataaagtctc
      Y - 182
721    cataccttagatctaagatacggacttggcccactcagctccttattactatataata
781    tagcataaatcaaacgtaaacgagcagcttattcagaagaattcccaactcogcttcaa
841    tagcgaataaagatagggctcgggaatttagatctgtaacgctcgggttttatatgagactag
901    gcaacgattaaagtcgaattagagcctccttctctctgataaatgactacagcctcagc
961    aaatcgggactgtgataccacagcagcaacgagcactatttagcttaattataatfaatg
1021  taatttagcaatgtaaataggcatgagtataaaataaacgagcagattatataaaaaaaaa
1081  aaaaaaaaaaaaaa
    
```

Fig. 5. A. Amino acid sequence of MsCP20 determined by mass spectrometry and Edman degradation. B. Nucleotide and deduced amino acid sequence of MsCP20 cDNA. The deduced amino acid sequence is shown below the cDNA sequence. The putative signal sequence is shaded below the cDNA sequence. Bold residues denote conserved residues of the R&R consensus sequence (Rebers and Riddiford, 1988). Underlined residues denote differences with the amino acid sequence obtained from protein sequencing. This sequence has been deposited in the GenBank database, with accession number AF487520.

**MsCP27**

**A.**  
 AELPSRNYIP QDQGGPGPYPG PHGGIGGDTG NRRPQQDAEK NSNVVKQEQE  
 ISDSGNHYFG FETSNIRAE EAGGPEQAQG GYSYKGGDDGQ TYTLIYTSGE  
 GGFKPQGEHL PVAPPTPEAI LIALQONNER EAAGIFDDGQ YHPENNGQRN  
 PGASAGAGTG SGYHY

**B.**

```

1      gaaaaacagtttgggttcaattttggttttaagtcaccaccctaaaggaaccatgaattg
61     ataatctacgtaaccataatcgccctcactcgcgctgagcttccatcaaggaattat
      I I Y V T I I A L T S A A E L P S R N Y 8
121    atccctcaggatcaaggatcgggtccctaccaggccctcattggggcattggtggagac
      I P Q D Q G S G P Y P G P H G G I G G D 28
181    accggcaacagaaagcaccagcaagcagctgagaaaaattccaattggtgaaacagag
      T G N R R P Q Q D A E K N S N V V K Q E 48
241    caggaaatcagtgactctggaattatcactctggcttccaaacctcaaacggcctccga
      Q E I S D S G N Y H F G F E T S N G I R 68
301    gctgaggaggcggaggccagaaacggcgcaaggagggtactcgtacaaaggcagcagat
      A E E A G G P E Q A Q G G Y S Y K G D D 88
361    ggaacagacatacagcctgatctacacgtcgggagggcggttcaagccgcaaggagag
      G Q T Y T L I Y T S G E G G F K P Q G E 108
421    cacctgcagtcgcccaccacggcggcggcattctgatcgcgctcagcagaacagag
      H L P V A P P T P E A I L I A L Q Q N E 128
481    agagatgaggctcgtgcatattttagatggttcaatacccccgaacaaatggtcag
      R D E A A G I F D D G Q Y H P E N N G Q 148
541    agtagcccggtcctcggcagggcgtgtacaggcagcgggtaccactactgaagatta
      S R P G A S A G A G T G S G Y H Y - 165
601    ccaatggactgtacaataaactatagaagaacaactcagtgatatttgggttttacagt
661    ggtgatttggcaaacggtgatctgggaggttatttagattatatttatatttgaact
721    atgaagacattatttgaatgtaataaaatggtttgtatactagctgttataata
781    tcaaacgctatttttatagatagtgctgtaaaaataatttttctattcagcagcttatt
841    aaggaaaaatgagtaatttggtagtttggatttatatgaacacataatcgtatggt
901    tactgtaaacagaaatgaagatagattacatttactttatttattttttttttttta
961    taaatgaagatcacaggttaaaaaaaaaaaaaaaaaaaaaa
    
```

Fig. 6. A. Amino acid sequence of MsCP27 determined by mass spectrometry and Edman degradation. B. Nucleotide and deduced amino acid sequence of MsCP27 cDNA. The deduced amino acid sequence is shown below the cDNA sequence. The putative signal sequence is shaded below the cDNA sequence. Bold residues denote conserved residues of the R&R consensus sequence (Rebers and Riddiford, 1988). Underlined residues denote differences with the amino acid sequence obtained from protein sequencing. This sequence has been deposited in the GenBank database, with accession number AY083171.

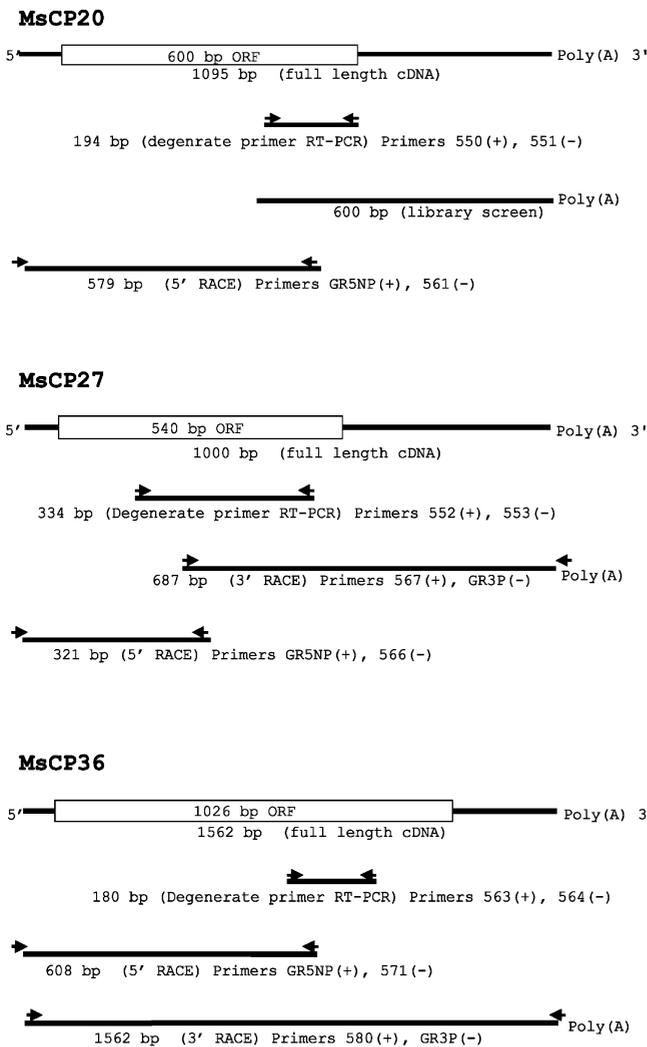


Fig. 7. Strategy used to clone cDNA for MsCP20, MsCP27, and MsCP36. Bars represent double stranded cDNA, arrows represent primers. Primer names are listed, with sense and antisense primers designated (+) and (-), respectively. See Table 1 for exact primer positions.

### 3.3. Cloning of MsCP36, MsCP20, and MsCP27 cDNAs

Degenerate primers were designed based on the amino acid sequences of MsCP36, MsCP20 and MsCP27, and used in RT-PCR experiments using fifth instar bar-stage pharate pupal integument RNA as template to isolate partial clones of each gene. The partial clones of MsCP36 and MsCP27 were each used as templates for 5'- and 3'-RACE and full-length cDNA sequences were obtained by assembly of overlapping sequences (Fig. 7). The partial clone for MsCP20 was used as a probe to screen an *M. sexta* bar-stage pharate pupal integument cDNA library, which yielded another partial clone that contained the poly(A) tail. RACE was employed to obtain the 5'-end of the clone.

The cDNA for MsCP36 contains 1562 nucleotides with a 1026-bp open reading frame encoding a 342-resi-

due polypeptide, including a putative secretion signal peptide of 15 residues (Fig. 4B). The putative positions P186, V219, and T255 correspond to the protein sequence positions I186, G219, and A255. The predicted mass of the deduced mature protein is 29,638 Da with a calculated pI of pH 5.00. MsCP36 cDNA hybridized with an approximately 1.5 kb RNA (Fig. 8).

The cDNA for MsCP20 contains 1095 nucleotides with a 600-bp open reading frame encoding a 200-residue polypeptide, including a putative secretion signal peptide of 18 residues (Fig. 5B). The putative amino acid sequence differed from the experimentally determined protein sequence (Fig. 5A) in two positions. The putative positions L136 and P147 correspond to the protein sequence positions V136 and I147. The predicted mass of the deduced mature protein is 17,711 Da with a calcu-

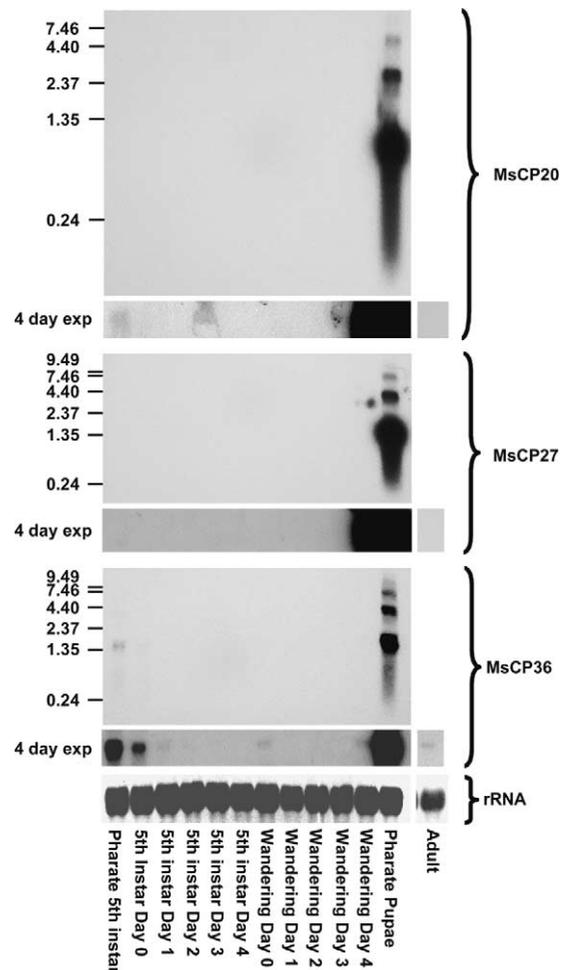


Fig. 8. Northern analysis of RNA from *M. sexta* pharate fifth instar, fifth instar, pharate pupal, and adult integuments. Total RNA (20  $\mu$ g) was separated by electrophoresis in agarose gel with formaldehyde, transferred to neutral nylon membrane, and hybridized with the appropriate  $^{32}$ P-labeled cDNA. Bars indicate the positions of the size markers (labeled in kilobases). Ribosomal RNA (rRNA) is shown on the nylon membrane stained with methylene blue. All X-ray film was exposed for 1 day, unless labeled as 4 day exposure.

lated pI of pH 4.89. The MsCP20 cDNA hybridized with an approximately 1.2 kb RNA (Fig. 8).

The cDNA for MsCP27 contains 1000 nucleotides with a 540-bp open reading frame encoding a 180-residue polypeptide, including a putative secretion signal peptide of 15 residues (Fig. 6B). This deduced amino acid sequence also differed from the experimentally determined protein sequence (Fig. 6A) in two positions. The putative positions S15 and S149 correspond to the protein sequence positions P15 and N149. The predicted mass of the deduced mature protein is 17,410 Da with a calculated pI of pH 4.57. MsCP27 cDNA hybridized with an approximately 1.0 kb RNA (Fig. 8).

The cDNA probes for all three transcripts also hybridized with RNAs of approximately 4 and 7 kb. This could be due to alternative splicing of larger transcripts, with the largest band representing the least spliced and the band of expected size representing the most spliced. Alternatively, because two larger and similar sized bands were detected for all of the transcripts (which are similar in primary structure), it is possible that the RNA assumes some secondary structure causing self association that results in anomalous migration patterns.

### 3.4. Temporal expression of MsCP36, MsCP20 and MsCP27 genes

To determine the transcript levels of MsCP36, MsCP20 and MsCP27, northern analysis was carried out on RNA samples collected each day from pharate fifth instar larvae to pharate pupae, as well as one adult day. MsCP36 transcripts were detected at low levels from pharate fifth instar larvae to fifth instar day 2 (Fig. 8). Transcripts appeared again in wandering day 0 larvae and wandering day 4 larvae before increasing to a maximum in pharate pupae. MsCP36 transcripts were also present at low levels in adults. Transcripts of MsCP20 were present at low levels in pharate fifth instar and fifth instar day 2 larvae, then rose dramatically in pharate pupae (Fig. 8). No transcript could be detected in adults. Transcripts of MsCP27 were only detected in pharate pupae (Fig. 8). All of these results are consistent with our earlier RT-PCR experiments using 35 cycles, which showed MsCP20 and MsCP36 to be present throughout the larval stadium rising to a maximum in pharate pupae and MsCP27 only present in pharate pupae (data not shown).

## 4. Discussion

Three major proteins, MsCP20, MsCP27 and MsCP36, from the cuticle of pharate pupae of *M. sexta* have been purified and characterized by complete sequence determination. They have relative masses of 17,894, 17,623 and 29,771 Da, respectively, according to

MALDI mass spectrometry, and their masses calculated from the protein sequences are 17,713, 17,448 and 29,582 Da, giving deficits in masses of 181, 175 and 189, respectively. Similar discrepancies between experimentally determined and calculated masses, ranging between 184 and 189 mass units, were also observed for some of the peptides obtained by proteolytic digestion. All of these peptides contained a threonine residue with a very low sequencing yield; that residue occurred in a cluster of proline residues at a position corresponding to a glycosylated threonine residue in an endocuticular protein (LmAbd-4) in *Locusta migratoria*. The threonine *O*-modification in LmAbd-4 has a relative mass of 203 and was identified as *N*-acetylgalactosamine (Talbo et al., 1991). Threonine residues, similarly located in a cluster of prolines and carrying a modification with a mass of 203, have been reported from cuticular proteins from *Schistocerca gregaria* (Andersen, 1998) and the crab, *Cancer pagurus* (Andersen, 1999). These modifications were tentatively identified as an *N*-acetylhexosamine. *O*-glycosylation site prediction analysis of all three of the *M. sexta* cuticular proteins using the NetOGlc 2.0 Prediction Server indicated that only the three threonine residues that exhibited low yields during Edman degradation, T181, T142 and T116 of MsCP36, 27, and 20, respectively, would be *O*-glycosylated (Hansen et al., 1997, 1998). Presumably, those threonine residues in the *Manduca* proteins are also conjugated to a sugar moiety. However, the differences between the measured and calculated masses of all three *Manduca* pupal proteins agree neither with that of an acetylated amino sugar nor an unmodified hexose. The identity of these moieties remains unknown.

The N-terminal sequences of MsCP20, MsCP27 and MsCP36 have been determined previously from preparations obtained using two different purification procedures (Hopkins et al., 2000). Comparison to the present complete sequences shows that the previous determinations conducted on proteins electroeluted from 1-D polyacrylamide gels are in better agreement with the actual sequences than are the determinations using PVDF blots of proteins resolved by 2-D IF-SDS-PAGE.

All three *Manduca* pupal proteins have a primary structure that is typical for many cuticular proteins from a variety of arthropod species. Their sequences are divided into three regions, where the central region contains the Rebers–Riddiford consensus sequence originally based upon six cuticular proteins representing three insect species, *Drosophila melanogaster*, *Sarcophaga bullata*, and *M. sexta* (Rebers and Riddiford, 1988). Fig. 9 shows an alignment of the central regions of the three pupal proteins and some selected cuticular proteins containing the elongated version of the Rebers–Riddiford sequence called RR-1 (Andersen, 1998). Both MsCP20 and MsCP36 contain a perfect Rebers–Riddiford sequence in the C-terminal part of their central region,



cantly lower than the masses calculated from their migration in SDS-PAGE. Similar mass discrepancies have regularly been observed for cuticular proteins (Andersen et al., 1995), and they are probably due to a tendency for cuticular proteins to bind less SDS than most globular proteins of the same size, resulting in slower migration in an electric field. The calculated isoelectric points for MsCP20, MsCP27 and MsCP36 are 4.8, 4.4 and 4.9, respectively, which are different from the experimentally determined pI-values of 5.4, 5.1 and 5.0. The reasons for these differences are unknown.

Most of the glycine residues in MsCP20 and MsCP36 are concentrated in the terminal regions, where they account for about 50% of the total number of residues. Proteins containing glycine-rich regions have been obtained from cuticles of various insect species, such as stiff cuticle from adult *Tenebrio molitor* (Charles et al., 1992; Bouhin et al., 1992; Mathelin et al., 1995), pupae of *D. melanogaster* (Apple and Fristrom, 1991), and pharate adult locusts (Andersen et al., 1993; Krogh et al., 1995). Despite similarities in composition between these proteins, little similarity is seen when the sequences of their glycine-rich regions are compared.

The central regions as well as the terminal regions in all three pupal cuticle proteins of *M. sexta* are predominantly hydrophilic, according to the analysis method of Kyte and Doolittle (1982). According to Garnier et al. (1978), the terminal regions will preferably fold into beta-turns and random coils, and MsCP36 contains two short regions (residues 108–126 and 188–202), which are predicted to fold into alpha-helices. The conserved central RR-sequences will probably fold into the beta-pleated sheet structures suggested by Iconomidou et al. (1999).

The regulation of cuticle protein gene expression by ecdysteroid and juvenile hormones has been determined for several larval cuticular proteins of *M. sexta* (Riddiford et al., 1986; Rebers et al., 1997; Horodyski and Riddiford, 1989). The molting hormones may also play a role in the synthesis and secretion of proteins for pupal cuticle formation. MsCP20 and MsCP36 transcripts are present at low levels during the molt from fourth to fifth instar indicating a role for these proteins in larval cuticle structure, but they then increase rapidly to peak levels in pharate pupae (Fig. 8). These proteins are the predominant ones secreted into pharate pupal cuticle, suggesting their importance in the structure of the pupal exoskeleton (Hopkins et al., 2000). The transcript for MsCP36 was the only one of the three present in the adult, suggesting some involvement in adult cuticle structure. The transcript for MsCP27, which was not detected in the fifth instar larvae or adult, also increased to a peak level in pharate pupae. Synthesis and secretion of MsCP27 appears to be specific for pharate pupal cuticle formation. Because transcripts of the three proteins increase to their highest levels during pharate pupal cuti-

cle formation, a period coinciding with the decrease in 20-hydroxyecdysone (20-HE) to very low titers, expression of the genes for these proteins appears to be regulated in a similar manner. The negative correlation between MsCP20, MsCP27 and MsCP36 transcript levels in pharate pupae and 20-HE titers suggests that these genes are expressed only after there is a decline of the hormone titer.

From the results presented here and from previous studies (Hopkins et al., 2000), we hypothesize that, in *M. sexta*, the three proteins MsCP20, MsCP27 and MsCP36 are synthesized in the epidermis during the larval–pupal transformation and are secreted in large quantities into the presumptive exocuticle of the pupa, where they take part in the sclerotization process (Schaefer et al., 1987; Okot-Kotber et al., 1996; Xu et al., 1997; Huang et al., 1997; Kerwin et al., 1999). MsCP20 and MsCP36 also are produced during development of fifth instar larvae suggesting their roles in larval cuticle formation (Wolfgang and Riddiford, 1981, 1986, 1987), whereas MsCP27 appears to be specifically associated with pupal cuticle.

## Acknowledgements

We thank Lene Skou, Odense University, for obtaining the MALDI mass spectra, and Dr Judith Willis, University of Georgia, and Dr Lynn Riddiford, University of Washington, for critical comments. Support from the Danish Natural Science Foundation, The Novo Nordisk Foundation, the Carlsberg Foundation and the U.S. National Science Foundation (MCB-9985959) is gratefully acknowledged. This is a cooperative investigation between Kansas State University (Contribution No. 02-443-J from the Kansas Agricultural Experiment Station), Copenhagen University and the USDA. All programs and services of the USDA are offered on a nondiscriminatory basis without regard to race, color, national origin, religion, sex, age, marital status or handicap. Mention of a proprietary product does not constitute a recommendation by the USDA.

## References

- Andersen, J.S., Andersen, S.O., Højrup, P., Roepstorff, P., 1993. Primary structure of a 14 kDa basic structural protein (Lm-76) from the cuticle of the migratory locust, *Locusta migratoria*. *Insect Biochemistry and Molecular Biology* 23, 391–402.
- Andersen, S.O., 1998. Amino acid sequence studies on endocuticular proteins from the desert locust, *Schistocerca gregaria*. *Insect Biochemistry and Molecular Biology* 28, 421–434.
- Andersen, S.O., 1999. Exoskeletal proteins from the crab, *Cancer pagurus*. *Comparative Biochemistry and Physiology* 123A, 203–211.
- Andersen, S.O., 2000. Studies on proteins in post-ecdysial nymphal cuticle of locust, *Locusta migratoria*, and cockroach, *Blaberus craniifer*. *Insect Biochemistry and Molecular Biology* 30, 569–577.

- Andersen, S.O., Højrup, P., 1987. Extractable proteins from abdominal cuticle of sexually mature locusts, *Locusta migratoria*. *Insect Biochemistry* 17, 45–51.
- Andersen, S.O., Højrup, P., Roepstorff, P., 1995. Insect cuticular proteins. *Insect Biochemistry and Molecular Biology* 2, 153–176.
- Andersen, S.O., Peter, M.G., Roepstorff, P., 1996. Cuticular sclerotization in insects. *Comparative Biochemistry and Physiology* 113, 689–705.
- Andersen, S.O., Rafn, K., Roepstorff, P., 1997. Sequence studies of proteins from larval and pupal cuticle of the yellow mealworm, *Tenebrio molitor*. *Insect Biochemistry and Molecular Biology* 27, 121–131.
- Apple, R.T., Fristrom, J.W., 1991. 20-Hydroxyecdysone is required for, and negatively regulates, transcription of *Drosophila* pupal cuticle protein genes. *Developmental Biology* 146, 569–582.
- Bærnholdt, D., Andersen, S.O., 1998. Sequence studies on post-ecdysial cuticular proteins from pupae of the yellow mealworm, *Tenebrio molitor*. *Insect Biochemistry and Molecular Biology* 28, 517–526.
- Bell, R.A., Joachim, F.G., 1976. Techniques for rearing laboratory colonies of tobacco hornworm and pink bollworm. *Annals Entomological Society of America* 69, 365–373.
- Binger, L.C., Willis, J.H., 1994. Identification of the cDNA, gene and promoter for the major protein form the flexible cuticles of the giant silkworm *Hyalophora cecropia*. *Insect Biochemistry and Molecular Biology* 24, 989–1000.
- Bouhin, H., Charles, J.-P., Quenedey, B., Courrent, A., 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 Molecular Biology* 1, 53–62.
- Charles, J.-P., Bouhin, H., Quenedey, 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. *European Journal of Biochemistry* 206, 813–819.
- Chomczynski, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Analytical Biochemistry* 162, 156–159.
- Christensen, A.M., Schaefer, J., Kramer, K.J., Morgan, T.D., Hopkins, T.L., 1991. Detection of cross-links in insect cuticle by REDOR NMR spectroscopy. *Journal of the American Chemical Society* 113, 6799–6802.
- Cox, D.L., Willis, J.H., 1985. The cuticular proteins of *Hyalophora cecropia* from different anatomical regions and metamorphic stages. *Insect Biochemistry* 15, 349–362.
- Garnier, J., Osguthorpe, D.J., Robson, B., 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *Journal of Molecular Biology* 120, 97–120.
- Hansen, J.E., Lund, O., Rapacki, K., Brunak, S., 1997. *O*-glycosylase version 2.0—a revised database of glycosylated proteins. *Nucleic Acids Research* 25, 278–282.
- Hansen, J.E., Lund, O., Tolstrup, N., Gooley, A.A., Williams, K.L., Brunak, S., 1998. NetOglyc: prediction of mucin type *O*-glycosylation sites based on sequence context and surface accessibility. *Glycoconjugate Journal* 15, 115–130.
- Hopkins, T.L., Kramer, K.J., 1992. Insect cuticle sclerotization. *Annual Review of Entomology* 37, 273–302.
- Hopkins, T.L., Krchma, L.J., Ahmad, S.A., Kramer, K.J., 2000. Pupal cuticle proteins of *Manduca sexta*: characterization and profiles during sclerotization. *Insect Biochemistry and Molecular Biology* 30, 19–27.
- Horodyski, F.M., Riddiford, L.M., 1989. Expression and hormonal control of a new larval cuticular multigene family at the onset of metamorphosis of the tobacco hornworm. *Developmental Biology* 132, 292–303.
- Huang, X., Xu, R., Hawley, M.D., Kramer, K.J., 1997. Model insect cuticle sclerotization: reactions of catecholamine quinones with the nitrogen-centered nucleophiles imidazole and *N*-acetylhistidine. *Bioorganic Chemistry* 25, 179–202.
- Iconomidou, V.A., Willis, J.H., Hamodrakas, S.J., 1999. Is  $\beta$ -pleated sheet the molecular conformation which dictates formation of helical cuticle? *Insect Biochemistry and Molecular Biology* 29, 285–292.
- Jensen, C., Andersen, S.O., Roepstorff, P., 1998. Primary structure of two major cuticular proteins from the migratory locust, *Locusta migratoria*, and their identification in polyacrylamide gels by mass spectrometry. *Biochimica et Biophysica Acta* 1429, 151–162.
- Kerwin, J.L., Turecek, F., Xu, R., Kramer, K.J., Hokins, T.L., Gatlin, C.L., et al. 1999. Mass spectrometric analysis of catechol-histidine adducts from insect cuticle. *Analytical Biochemistry* 268, 229–237.
- Kramer, K.J., Kanost, M.R., Hopkins, T.L., Jiang, H., Zhu, Y.C., Xu, R., et al. 2001. Oxidative conjugation of catechols with proteins in insect skeletal systems. *Tetrahedron* 57, 385–392.
- Krogh, T.N., Skou, L., Roepstorff, P., Andersen, S.O., Højrup, P., 1995. Primary structure of proteins from the wing cuticle of the migratory locust, *Locusta migratoria*. *Insect Biochemistry and Molecular Biology* 25, 319–329.
- Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydrophobic character of a protein. *Journal of Molecular Biology* 157, 105–132.
- Lampe, D.J., Willis, J.H., 1994. Characterization of a cDNA and gene encoding a cuticular protein from rigid cuticles of the giant silkworm, *Hyalophora cecropia*. *Insect Biochemistry and Molecular Biology* 24, 419–435.
- Mathelin, J., Bouhin, H., Quenedey, B., Courrent, A., Delachambre, J., 1995. Identification, sequence and mRNA expression pattern during metamorphosis of a cDNA encoding a glycine-rich cuticular protein in *Tenebrio molitor*. *Gene* 156, 259–264.
- Okot-Kotber, B.M., Morgan, T.D., Hopkins, T.L., Kramer, K.J., 1996. Catecholamine containing proteins from the pharate pupal cuticle of the tobacco hornworm, *Manduca sexta*. *Insect Biochemistry and Molecular Biology* 26, 475–484.
- Rebers, J.E., Riddiford, L.M., 1988. Structure and expression of a *Manduca sexta* larval cuticle gene homologous to *Drosophila* cuticle genes. *Journal of Molecular Biology* 203, 411–423.
- Rebers, J.E., Niu, J., Riddiford, L.M., 1997. Structure and spatial expression of the *Manduca sexta* MSCP14.6 cuticle gene. *Insect Biochemistry and Molecular Biology* 27, 229–240.
- Rebers, J.E., Willis, J.H., 2001. A conserved domain in arthropod cuticular proteins binds chitin. *Insect Biochemistry and Molecular Biology* 31, 1083–1093.
- Riddiford, L.M., Baeckmann, A., Hice, R.H., Rebers, J.E., 1986. Developmental expression of three genes for larval cuticular proteins of the tobacco hornworm, *Manduca sexta*. *Developmental Biology* 118, 82–94.
- Roberts, P.E., Willis, J.H., 1980. The cuticular proteins of *Tenebrio molitor*. II. Patterns of synthesis during postembryonic development. *Developmental Biology* 75 (59–69), 70–77.
- Roepstorff, P., 1993. 252-Californium plasma desorption time-of-flight mass spectrometry of peptides and proteins. *Methods in Molecular Biology* 17, 229–235.
- Sambrook, J., Russell, D., 2001. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.
- Schaefer, J., Kramer, K.J., Garbow, J.R., Jacob, G.S., Stejskal, E.O., Hopkins, T.L., et al. 1987. Aromatic cross-links in insect cuticle: detection by solid state  $^{13}\text{C}$  and  $^{15}\text{N}$  NMR. *Science* 235, 1200–1204.
- Talbo, G., Højrup, P., Rahbek-Nielsen, H., Andersen, S.O., 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 post-translationally modified proteins. *European Journal of Biochemistry* 195, 495–504.

- Willis, J.H., 1996. Metamorphosis of the cuticle, its proteins, and their genes. In: Gilbert, L.I., Tata, J.R., Atkinson, B.G. (Eds.), *Metamorphosis: Postembryonic Reprogramming of Gene Expression in Amphibian and Insect Cells*. Academic Press, San Diego, pp. 253–282.
- Willis, J.H., 1999. Cuticular proteins in insects and crustaceans. *American Zoologist* 39, 600–609.
- Wolfgang, W.J., Riddiford, L.M., 1981. Cuticular morphogenesis during continuous growth of the final instar larva of a moth. *Tissue Cell* 13, 757–772.
- Wolfgang, W.J., Riddiford, L.M., 1986. Larval cuticular morphogenesis in the tobacco hornworm, *Manduca sexta*. *Developmental Biology* 113, 305–316.
- Wolfgang, W.J., Riddiford, L.M., 1987. Cuticular mechanics during larval development of the tobacco hornworm, *Manduca sexta*. *Journal of Experimental Biology* 128, 19–33.
- Xu, R., Huang, X., Hopkins, T.L., Kramer, K.J., 1997. Catecholamine and histidyl protein cross-linked structures in sclerotized insect cuticle. *Insect Biochemistry and Molecular Biology* 27, 101–108.