Sarcophagid Larval Proteins: Partial Sequence Homologies Among Three Cuticle Proteins and Related Structures of Drosophilids

William J. Henzel,^{1*} John E. Mole,² Kathleen Mulligan,¹ and Herbert Lipke¹

¹ Department of Biology, University of Massachusetts at Boston, Boston, Massachusetts 02125, USA

² Department of Biochemistry, University of Massachusetts Medical Center, Worcester, Massachusetts 01605, USA

Summary. Three structural proteins from the larval cuticle of Sarcophaga bullata have been sequenced at the amino terminus for 30-40 residues. We observed a high degree of homology with related proteins of Drosophila melanogaster, based on the previous findings of M. Snyder, J. Hirsh, and N. Davidson [(1981) Cell 25:165-177]. S. bullata protein SC1 had 65% homology with Drosophila isolate CP1, and SC6 showed 49% homology with CPX and 54% with CP2a. The three sarcophagid polypeptides also resembled each other with respect to mapped products of tryptic cleavage. The sites of posttranslational arylation required for puparium formation, namely histidyl and lysyl residues, were asymmetrically distributed in the sarcophagid samples. In SC1 the bulk of the loci of putative crosslinks lay beyond the 43-residue fragment. In SC6 half the histidines fell within the first 25% of the primary

Key words: Cyclorrhaphan cuticle – Secondary structure of arthropod integument – Dipteran diversity

Introduction

In cyclorrhaphid dipterans the pupal and preemergent adult stages are protected by the sclerotized puparial sheath (Fraenkel and Bhaskaran 1973). The

tanned matrix of the puparium is formed directly by the addition of unconjugated soluble polypeptides to the chitin-protein complex at the conclusion of the third larval instar. These structural components are then joined by aryl bridges and the synthesis of covalent chitin-protein crosslinks (Kimura et al. 1976; Sugumaran and Lipke 1982; Sugumaran et al. 1982). Although the morphological and endocrine events initiating pupariation are essentially uniform within the suborder and its taxonomic subdivisions (Borror et al. 1976), the varied environments to which this life stage is subjected in different taxa may require significant differences in the composition and bridging of the sheath components (Fraenkel and Hsiao 1967; Lipke et al. 1983). The differences reported to date are substantial. For example, the phytophagous calyptrate drosophilids elaborate five major and two minor proteins of M, 8-10 kilodaltons, with the β -carbon of the side chain of N-acetyldopamine as the site of union between crosslinked polypeptides (Driskell 1974; Fristrom et al. 1978; Chihara et al. 1982). In this genus asparagine and leucine make up the majority of the N-termini after removal of the signal peptides (Snyder et al. 1982). On the other hand, the sets of precursor proteins for the puparial cases of two carrion-eating Calyptratae, Calliphora vicina and Sarcophaga bullata, consist of a minimum of 11 and as many as 21 proteins with higher molecular weights, and a broader range of isoelectric points, than those of Drosophila (Hackman and Goldberg 1976; Lipke et al. 1981, Willis et al. 1981). These two blowflies, furthermore, exhibit greater diversity with respect to N-termini, since alanine, valine, tyrosine, glycine, and aspartic and glutamic acids are found at this locus in addition to asparagine and

Offprint requests to: Dr. M. Sugumaran, Department of Biology, University of Massachusetts at Boston, Harbor Campus, Boston, Massachusetts 02125, USA

^{*} Present address: Genentech, Inc., South San Francisco, California 94040, USA



Fig. 1. Anion-exchange separation of cuticle proteins of S. bullata

leucine (Sugumaran et al. 1982). A further distinction between these two groups concerns the interprotein bridge itself. Crosslinking in blowflies is accomplished by ring peptidylation of a quinonoid intermediate rather than side-chain substitution (Sugumaran and Lipke 1982, 1983a, b).

Diptera compared evolutionary trees constructed on the basis of immunological cross-reactivities with those based on morphological features (Collier and MacIntyre 1977; Beverley and Wilson 1982). Good agreement was reported between the two immunologically based dendrograms in spite of marked functional differences between the proteins chosen for microcomplement fixation, namely α -glycerophosphate dehydrogenase and larval hemolymph protein. The extent to which larval cuticle proteins conform to these trees remains to be determined, ideally by means of an independent criterion such as amino acid sequence homology (Wilson et al. 1977). In addition, the extent of primary structural similarity among cuticle polypeptides from a single calyptrate species has not been established in spite of the interest in patterns of metamorphosis and puparial assembly. In the course of studies on the mechanism of peptide arylation, three cuticle proteins from S. bullata have been purified and sequenced to such an extent that comparison with drosophilid material is possible. Although considerable sequence homology is seen, sufficient variation in primary and secondary structure is present to accommodate ecological specialization.

Materials and Methods

Larval Culture. Sarcophaga bullata puparia were purchased from Carolina Biological Supply Co. (Burlington, NC, USA). Following eclosion and mating, larviposition was initiated by introduction of swine liver into the cages. Maggots were reared at 28°C in batches of 10³ on 500 g commercial dog food containing 20% horsemeat. After 7 days the third instar entered the prepuparial stage, as expressed by emptying of the gut and abandonment of the moist medium. For synchronization, the maggots were transferred to covered jars containing water to a depth of 2 cm and held at 25°C for 18 h. The population was sampled for synchrony by placing a few individuals on dry sand and observing the interval to darkening of the spiracles and formation of the white puparial stage, usually 1-4 h. The maggots were washed with water, drained, and processed for separation of cuticles from viscera.

Purification of Larval Structural Proteins. Disruption of the larvae, isolation of the integument, and extraction of the soluble components in the presence of phenylthiourea and phenylmethanesulfonyl fluoride were done as previously described (Lipke et al. 1981). The material soluble at pH 5.0 was precipitated by addition of ethyl alcohol to 20% (vol/vol), redissolved in 50 mM Na₃ citrate-1 mM KCN, dialyzed for 72 h at 2°C against 500 volumes of the same buffer followed by three changes of water, and lyophilized.

Proteins were resolved by one of two procedures, depending on whether a high yield of selected components or a complete profile of structural proteins was required. When the full complement of skeletal polypeptides was desired, the proteins were isolated by replicated isoelectric focusing (IEF) disk gel electrophoresis followed by sodium dodecyl sulfate (SDS) gel electrophoresis to separate components of similar pI but disparate Mr (Lipke et al. 1981). The proteins prepared by this method are numbered 1-20 according to Lipke et al.'s (1981) notation. Since more material was required for sequencing, three of the most abundant components were isolated in high yield at the expense of lesser components by a simpler procedure. These three proteins, SC1, SC4, and SC6, were not resolved completely on the disk gels but were separable in homogeneous condition by highpressure liquid chromatography (HPLC). Batches of lyophilized cuticle protein (see above) were focused for 48 h at 8W constant power in lots of 240 mg on a Model 1415 flat-bed electrophoresis cell equipped with a 110 × 300 mm tray (Bio-Rad). Electrolytes were 0.1 M NaOH and 0.1 M H₃PO₄ for the cathode and anode, respectively. The gel was scanned by UV reflectance and sectioned, and proteins of pI 4.1-5.0 were pooled and eluted with 0.1 M NH₄HCO₃. The sample was resolved by anion-exchange HPLC to afford components SC1, SC4, and SC6, each of which exhibited a single N-terminus and single bands on SDS-polyacrylamide (9%) electrophoresis and on Agarose-Sephadex IEF (Laemmli 1970; Manrique and Lasky 1981). For anion-exchange chromatography a Brownlee AX-300 column (25 cm × 4.6 mm) was equilibrated with 0.1 M NH4HCO3 prior to loading with 10 mg of prefocused proteins. The initial flow rate was 1 ml/min, which was increased to 2 ml/min at 10 min. A linear 80-min gradient of 0.05 M NH4HCO3 to 80% 0.2 M NH4HCO3 was begun at 25 min. Protein SC4 eluted from the column after 13.4 min, SC1 at 61.4 min, and SC6 at 64.6 min (Fig. 1).

Peptide Mapping. Peptide maps were obtained by digestion of 0.5–0.7 mg of cuticle protein with 2% (wt/wt) TPCK trypsin in 100 μ l 0.1 M NH₄HCO₃ for 24 h at 37°C. The digest was separated on an Altex C₁₈ 5 μ column (25 cm × 4.6 mm) with a linear gradient of 0.1% H₃PO₄ to 50% CH₃CN in 50 min at a flow rate of 2.0 ml/min (Fullmer and Wasserman 1979).

Amino Acid Analysis. Proteins were hydrolyzed in 200 μ l constantly boiling 5.7 N HCl under vacuum for 24 h at 110°C. Tryptophan was assayed following hydrolysis in 3 M mercaptoethanesulfonic acid. Hydrolysates were analyzed on a Beckman 119C amino acid analyzer using an accelerated protein hydrolysate program.

Sequence Determination. In addition to electrophoretic homogeneity, the purity of the proteins isolated by flat-bed IEF and

 Table 1.
 Automated Edman degradation of cuticle proteins SC1,

 SC4, and SC6 from Sarcophaga bullata: successive yields^a

| Cycle | SC1 | Yield (nmol) | SC4 | Yield (nmol) | SC6 | Yield (nmol) |
|-------|--------|-----------------|--------------------|-----------------|--------|-----------------|
| 1 | Α | 6.75 | Y | 5 10 | N | 4.41 |
| 2 | L | 4.00 | G | 6.61 | F | 5.42 |
| 3 | E | 7.10 | ŭ | 10.7 | ñ | 3.60 |
| 4 | Т | 0.77 | F | 7 36 | ă | 4 49 |
| 5 | S | 0.36 | ĩ | 14.1 | N | 4 38 |
| 6 | N | 5.63 | D | 1 3 3 | v | 4.50 |
| 7 | G | 1.01 | N | 1.55 | r T | 4.08 |
| 8 | I | 2.90 | ŝ | 0.08 | v | |
| 9 | н | NO | ī | 24 | 6 | 0.23 |
| 10 | F | 4 90 | I V | 3.4 NO | s V | 0.23 |
| 11 | v | 2.90 | Å | | Y C | 3.14 |
| 12 | Α | 3.00 | n | 2.89 | 3 | NQ 2.20 |
| 13 | Ī | 1.02 | 0 | 1.04 | | 3.39 |
| 14 | G | 3 55 | Q. | 1.89 | Ŷ | 2.71 |
| 15 | Ğ | NO | E | 2.25 | G | 3.22 |
| 16 | Ď | 2.75 | G | 2.11 | v | 2.98 |
| 17 | Ē | 2.73 | H , | 0.9 | D | 2.12 |
| 18 | ้ห | NO | L | 1.94 | Q | 2.15 |
| 19 | Ĝ | 2.60 | E | 1.58 | F | 2,25 |
| 20 | Ň | 2.00 | G | 1.68 | K. | NQ |
| 21 | Î | 1.00 | D | 1.50 | Ŷ | 1.84 |
| 22 | н | 2.08 NO | (K/F) ^c | NQ | G | 2,46 |
| 23 | ä | 165 | Т | 0.2 | L | 1.6/ |
| 24 | N/F | 1.05 | W | 0.93 | E | 1.75 |
| 25 | F | 1.76 | V | 1.21 | L | 1.53 |
| 26 | â | 1.75 | V | 1.10 | D | 1.20 |
| 27 | w | 1.30 | K | NQ | (N) | 0.44 |
| 28 | v | 2 10 | G | 0.81 | (S) | 0.08 |
| 29 | Ť | 2.19 | D | 1.82 | (I) | 0.95 |
| 30 | T | 0.19 | Y | 0.51 | ĸ | NQ |
| 31 | 4 F | 0.27 | E | 0.48 | Α | 0.85 |
| 32 | G | 1.00 | (Y) | 0.51 | (D) | 0.67 |
| 33 | E C | 1.00 | (V) | 0.67 | Q | 0.56 |
| 34 | | 1.16 | S | NQ | (E) | 0.59 |
| 35 | N N | 0.28 | ĸ | NQ | G | 0.78 |
| 36 | Ď | 1.19 | (E) | 0.31 | (H) | NQ |
| 37 | T T | 0.73 | G | 0.33 | L | 0.44 |
| 38 | т Т | 0.58 | (K) | NQ | (E) | 0.50 |
| 39 | v | 0.21 | | | G | 0.70 |
| 40 | I V | 0.65 | | | (D) | 0.55 |
| 41 | v A | 1.32 | | | - | |
| 42 | A | 0.77 | | | | |
| 43 | D F | 0.65 | | | | |
| | E | 1.00 | | | | |

NQ, not quantitated

* Residues in parentheses were identified by only one HPLC system

^b Chlorobutane was lost for cycle 15; trailover in cycle 15 suggests that glycine is the probable residue

• Cycle 21 was identified by only one HPLC system, which was unable to resolve lysine and phenylalanine

HPLC was assessed by quantitation of N-terminal residues through two cycles of the manual Edman procedure (Zalut et al. 1980). Amino-terminal amino acid sequences were ascertained on 2–7 nmol material using automated Beckman 890C sequencers adapted for microsequencing as described by Bhown et al. (1981) except that the coupling buffer was 0.1 M Quadrol and simultaneous ethyl acetate and benzene extractions were performed following coupling. PTH amino acids were identified and quantitated by chromatography on two independent reverse-phase systems. The primary system used a 6 μ Zorbox ODS column



Fig. 2. Peptide mapping of larval proteins from S. bullata by HPLC

and a 12-min, 25%-52% acetonitrile gradient in sodium acetate, pH 4.8. Most identifications were confirmed using a 5 μ Ultrasphere ODS column and a linear methanol gradient (20%-45%). Data were quantitated with a Hewlett-Packard 3390A integrator using PTH norleucine as an internal standard. Values for PTH serine and PTH threonine are reported as absolute yields and were not corrected for conversion to the dehydro forms, which were the major products. Yields for repeated analyses were between 93% and 97%.

Results

Relatedness Among Structural Proteins

The borate-soluble components of sarcophagid cuticle constitute 60% of the total cuticle protein. This compartment has been fractionated into 21 distinct entities of generally similar properties and composition. All the isoelectric points fall between pH 4.0 and 6.2, with $M_r = 16-24$ kilodaltons. Sulfur-containing amino acids are lacking and tryptophan is rarely encountered. The remaining 40% of the protein is covalently joined to chitin, and constitutes a single insoluble polypeptide chain of 90 kilodaltons that includes active sites for propolyphenol oxidase and peroxidase (Lipke et al. 1981; Sugumaran et al. 1982). The similarities in the molecular weights and compositions of the soluble components permit an assessment of relatedness based on amino acid analysis; 220 was taken as the average residue number to establish the "strong test" criterion of $S\Delta Q \leq$ 38.6 (Marchalonis and Weltman 1971). It can be seen from Table 2 that proteins 6-12, which are grouped within the pH range 4.88-5.23, fall well below the critical $S\Delta Q$ level. Protein 10, for instance, is related to components 6, 7, 8, 9, and 11. For proteins 2-6 and 17-19, recovered at the extremes of the isoelectric point distribution, little or no similarity in composition is evident, suggesting functionally disparate roles in puparium formation. Protein 20 is atypical among the components of high

Table 2. Relatedness among cuticle proteins from S. bullata expressed as $S\Delta Q$ values

| Protein | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | SC1 | SC4 | SC6 |
|---------|-----|-----|-----|-----|-----|-----|-----|-----------|-----------|-----------|-----|-----|-----|-----|-----|-----|-----|-----------|-----|-----|-----------|
| 2 | 156 | 149 | 287 | 213 | 245 | 240 | 259 | 244 | 178 | 136 | 141 | 179 | 169 | 145 | 154 | 121 | 138 | 150 | 164 | 274 | 190 |
| 3 | | 53 | 261 | 190 | 179 | 180 | 208 | 192 | 126 | 136 | 381 | 229 | 225 | 142 | 124 | 28 | 215 | <u>36</u> | 223 | 360 | 194 |
| 4 | | | 102 | 81 | 73 | 78 | 100 | 77 | 45 | 77 | 355 | 125 | 131 | 71 | 65 | 21 | 113 | <u>26</u> | 132 | 213 | 95 |
| 5 | | | | 102 | 104 | 121 | 125 | 80 | 87 | 168 | 544 | 121 | 152 | 148 | 149 | 102 | 140 | 161 | 145 | 79 | 73 |
| 6 | | | | | 8 | 23 | 25 | <u>29</u> | 23 | 45 | 285 | 171 | 207 | 191 | 193 | 107 | 82 | 132 | 163 | 202 | 85 |
| 7 | | | | | | 8 | 14 | 21 | 21 | 46 | 325 | 200 | 240 | 204 | 189 | 102 | 96 | 128 | 150 | 210 | 73 |
| 8 | | | | | | | 5 | 18 | <u>20</u> | <u>33</u> | 303 | 209 | 254 | 211 | 179 | 98 | 79 | 134 | 133 | 207 | 70 |
| 9 | | | | | | | | 18 | <u>20</u> | <u>34</u> | 309 | 212 | 259 | 231 | 207 | 118 | 88 | 147 | 145 | 210 | 81 |
| 10 | | | | | | | | | 22 | 56 | 367 | 169 | 220 | 164 | 139 | 106 | 97 | 123 | 165 | 171 | 84 |
| 11 | | | | | | | | | | 22 | 297 | 129 | 161 | 149 | 140 | 57 | 71 | 77 | 108 | 172 | 56 |
| 12 | | | | | | | | | | | 138 | 181 | 208 | 201 | 184 | 68 | 60 | 111 | 110 | 232 | 86 |
| 13 | | | | | | | | | | | | 454 | 449 | 417 | 433 | 309 | 199 | 375 | 348 | 521 | 370 |
| 14 | | | | | | | | | | | | | 12 | 103 | 141 | 149 | 185 | 141 | 140 | 166 | 144 |
| 15 | | | | | | | | | | | | | | 100 | 161 | 155 | 216 | 138 | 165 | 203 | 184 |
| 16 | | | | | | | | | | | | | | | 28 | 97 | 194 | 62 | 235 | 239 | 106 |
| 17 | | | | | | | | | | | | | | | | 81 | 166 | 75 | 200 | 218 | 166 |
| 18 | | | | | | | | | | | | | | | | | 90 | 47 | 147 | 244 | 122 |
| 19 | | | | | | | | | | | | | | | | | | 167 | 115 | 165 | 81 |
| 20 | | | | | | | | | | | | | | | | | | | 191 | 279 | 155 |
| SC1 | | | | | | | | | | | | | | | | | | | | 142 | <u>34</u> |
| SC4 | | | | | | | | | | | | | | | | | | | | | 83 |

Proteins 2-20 were separated by focusing and dissociative gel electrophoresis according to Lipke et al. (1981). Proteins SC1, SC4, and SC6 were resolved by flat-bed IEF and HPLC (see Materials and Methods). Closeness of relationship is indicated by an S Δ Q value of \leq 38.6 ("strong test" criterion; Marchalonis and Weltman 1971) for an average residue number of 220. S Δ Q values for closely related proteins are underlined

Table 3. Amino acid composition (residues/mol) of cuticle proteins of *S. bullata* resolved by preparative flat-bed IEF and anionexchange HPLC

| | Protein | | | | | | | | | | |
|----------------------------|---------|-------|-------|--|--|--|--|--|--|--|--|
| Amino acid | SC1 | SC4 | SC6 | | | | | | | | |
| Aspartic acid [®] | 26.28 | 27.16 | 24.02 | | | | | | | | |
| Threonine | 7.85 | 5.16 | 5.91 | | | | | | | | |
| Serine | 13.30 | 7.97 | 9.88 | | | | | | | | |
| Glutamic acid ^a | 25.80 | 24.24 | 22.37 | | | | | | | | |
| Proline | 14.26 | 3.59 | 6.88 | | | | | | | | |
| Glycine | 25.00 | 19.97 | 17.14 | | | | | | | | |
| Alanine | 12.82 | 8.86 | 9.97 | | | | | | | | |
| Valine | 8.01 | 3.03 | 9.59 | | | | | | | | |
| Isoleucine | 5.93 | 1.01 | 3.20 | | | | | | | | |
| Leucine | 7.05 | 14.03 | 7.07 | | | | | | | | |
| Tyrosine | 10.58 | 6.96 | 9.78 | | | | | | | | |
| Phenylalanine | 11.70 | 3.70 | 6.78 | | | | | | | | |
| Histidine | 10.26 | 4.94 | 7.07 | | | | | | | | |
| Lysine | 9.94 | 12.34 | 10.07 | | | | | | | | |
| Cysteine | 0 | 0 | 0 | | | | | | | | |
| Methionine | 0 | 0 | 0 | | | | | | | | |
| Arginine | 0 | 0 | 0 | | | | | | | | |
| pI | 4.12 | 4.75 | 4.89 | | | | | | | | |
| M _r | 20.3 | 15.1 | 16.7 | | | | | | | | |

Includes amides

pI, since it shows a strong relationship with acidic components 3 and 4 and another basic protein, 18. That the two methods of protein resolution (disk gel and flat bed) favored different groups of polypeptides is evident from the absence of relatedness between the class 2-20 and SC1, SC4, and SC6. Relatedness among proteins was confirmed by independent approaches, namely protein separation by HPLC and comparison of the resolved polypeptides by peptide mapping (Fig. 2). It is apparent that SC4 and SC6 fragment in a similar manner under the influence of TPCK trypsin. SC1 resembles SC4 and SC6 with respect to components eluting from the column after 28 min, but clearly differs during the 15–20 min interval. This difference is expressed in amino acid titers as well, where the abundance of proline, isoleucine, phenylalanine, and histidine clearly distinguishes SC1 from the other two components (Table 3).

Sequence Homologies in Sarcophagid and Drosophilid Exoskeletons

The chemical and physiological bases of cuticle crosslinking in Diptera have been reviewed recently (Lipke et al. 1983). Crosslinking of cuticle proteins is the result of posttranslational modification of lysyl, histidyl, and tyrosyl residues of some or all of the cuticle proteins present in the larval stage. To identify putative bridge sites, we isolated and partially sequenced uncrosslinked larval cuticle proteins from *S. bullata* (Fig. 3).

The sequenced regions of SC1, SC4, and SC6 showed there to be strong homologies between the two cyclorrhaphan suborders (Fig. 3). From the partial sequence data SC4 appears to be identical to SC6, except that it lacks the first 19 residues of SC6.

| SC1 CP1 CP2 CP3 CP4 CP4 CP2A CP2A SC6 SC4 | ΝРРγ | PHLA | i S . P | L V | G S | R S | S E S D | D D N N N N | V V A E E E E | A C A C A C A C A C A N | | | S S E E K R K | R L L S A S | S V V D E Y | D D N S Q S | D D D Q O | | R A R A P R A P R A R A V R V | | G G G D G Q | F F F F F | D V V N K | S S S S S S S S S S S S S S S S S S S | | | | S S D D I D I D I D I D I D I D I D I D | | G I G I G I G I S S S S I I I I I | *E E E A R N K | V Q S A A V A | A A A S S V Q D D | I A A A Q I Q I Q I | | | E A A I V L L L |
|--|------|--------------------------------------|------------------|------------------|------------------------|--|-------------|----------------------------|---------------------------------|--|-----------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|-----------------------|---------------------------------|---|------------------|----------------------------|-----------------------|-----------------------|---------------------------------------|------------------|------------------|-------------|--|---------------------------------------|---|----------------------------------|---------------------------------|---|--|------------------|-------------|--------------------------------------|
| SC1 CP1 CP2 CP3 CP4 CP4 CP2 CP2A SC6 SC4 | | H H H H K N E E | G G G G G G G G | N N N N H E D D | I I I E[? | H G H G G G G G G G G G G V W W T W | | F F F V V V | GGEESK | / T S S S S S S V S V S V S V S V S V S V | I P P E I | E E E Y W Y | G G G G V V | EEVEA | A H H H K | | | I T V K V S V S K T | Y Y Y Y Y | V V K K | A A A A V | Dj N D V | | N G N G N G | Y Y Y Y | Q Q Q E | P P T | S Q S Q Q S G Y | A A A A A A A A A A A A A A A A A A A | W L P | I L K | P P P ? | T T T V | P F P F P F E A | P P P P | I I I | |
| CP1 CP2 CP3 CP4 | | P P P | E E A E | A A A A | | AR AR K | A A A | V V I I | | E E Q | S S A A | H H N H | P P P | P P S S | A A K | P I P I N I E I | E F E F ENC | H P H P H P | R R | н н | H H | ENC ENC |) | | | | | | | | | | | | | | |

Fig. 3. Comparison of partial and complete amino acid sequences of larval cuticle proteins from Sarcophaga bullata (SC series) and Drosophila melanogaster (CP series). Primary structures of Drosophila proteins from Snyder et al. (1981, 1982). The star in the SCI sequence indicates that residue 9 was deleted to maximize homology; brackets in CP2A sequence show where a space has been inserted to maximize homology, even though no residue occupies that position in CP2A

Sequence identity did not exceed 12% between proteins SC4 and SC1. Sarcophaga protein SC1 shows 65% homology with the major cuticle protein of Drosophila, CP1. SC6 shows 49% homology with CPX and 54% with CP2A. All of the Drosophila and Sarcophaga proteins share identical amino acid residues at position 6, 13, 18, and 34 of Sarcophaga protein SC6 in the alignment shown (Fig. 3). Sarcophaga protein SC6 and all the Drosophila proteins except CP2 have an N-terminal asparagine. Protein SC1 exhibited microheterogeneity at residue 24, with a nearly equimolar ratio of asparagine and glutamic

With the exception of the tryptophan at residue 23 of SC4, the phenylisothiocyanate procedure confirmed the amino acid analysis of acid hydrolysates by corroborating the absence of cyst(e)ine, methionine, and arginine from the isolates (Table 3). The trend toward acidic isoelectric points for the three blowfly proteins is expressed within the N-terminal peptides, since glutamic and aspartic acids were recovered in ninefold excess over lysine and histidine in the most acidic protein, SC1, and in 1.7-fold excess in SC4 and SC6. The data also show clearly that the recovery of amide nitrogen in the sequenced regions of the sarcophagid peptides is insufficient to counteract the influence of the acidic amino acid residues on the pI values of the larval proteins, since the acid/amide ratios are 9:3, 8:1, and 10:5 for the sequenced stretches of SC1, SC4, and SC6, respectively.

Discussion

Exoskeletal proteins of cyclorrhaphid dipterans perform two distinct functions in addition to fulfilling a common requirement as integumental components of the locomotor systems of the larvae. In the immature stage the integument is exposed to the characteristic pH, osmolarity, microbial complement, and physical consistency of the medium in which the maggots develop. The cuticle not only protects the organism from local conditions detrimental to epidermal function, but also adjusts excretory and transport activity to accommodate the external milieu. Later in metamorphosis, larval proteins serve as precursors of the puparium by exposing functional groups subject to posttranslational arylation and glycosylation. The puparium in turn is fabricated in accord with the rigor of the puparial environment and the duration of this stage. Although D. melanogaster and S. bullata share the taxonomic features of the Series Schizophora and the Section Myodaria, the bionomics of the two cyclorrhaphans differ considerably. Larvae of the fruit fly develop in carbohydrate-rich fruit pulp of acid reaction. Low pH is maintained during the period of ammonia production by the larvae. S. bullata forgoes the egg stage completely; larviposition and feeding occur directly on proteinaceous carrion, the temperature and pH of which increase dramatically as a result of microbial and larval catabolism and extraintestinal digestion. Thus, although they share many morphological and physiological systems, their larval cuticles must be specialized with respect to their structural proteins.

On the basis of sequence homology for the initial 30-40 residues, it is evident that the cuticle genes of the calypterate and acalypterate cyclorrhaphans share a common descent. Homology ranges from 40% to 65% (Fig. 3). Homology between SC4 and SC6 is suggestive of proteolytic generation of the smaller component. When they are prepared in the presence of sodium citrate and phenylmethanesulfonyl fluoride to inhibit chain degradation, the ratio of structural components is constant and in accord with a separate gene for SC4. This near identity of the structural members raises major questions concerning the function of the five related components of D. melanogaster and the 20 isolates from the blowfly. The maximum opportunities for specialization occur at the anatomical level. Although little differentiation is apparent among the individual strata of the endocuticle, inter- or intraspecific deposition of individual proteins may distinguish the distal from the more proximal regions of the body wall. Localization of polypeptides is also possible with respect to the dorsal and ventral aspects of the cylindrical larval form and between the principal annuli and the intersegmental membranes (Andersen et al. 1973; Cox and Willis 1982).

In the course of pupariation in *Sarcophaga*, lysine and histidine of the primary chain are modified by arylation, and thus afford basic adducts on acid hydrolysis (Sugumaran and Lipke 1982). This addition reaction, which constitutes the major crosslinking process, thereby decreases the solubility and susceptibility to proteolysis of all the puparial proteins in concert (Lipke et al. 1981; Willis et al. 1981). Within the stretch of SC1 sequenced, which constituted 23% of the protein, none of the lysine and 30% of the histidine were identified. In contrast, although the N-terminal peptide comprised 25% of SC6, 14% of the histidine was recovered in this region. Since radiometric studies indicate that 40% of the histidines of the primary chain are converted to aryl adducts chemically distinct from peptide histidyl (Sugumaran and Lipke 1982; Lipke et al., 1983), it follows that the residues altered during sclerotization are distributed nonuniformly within the primary structure.

Acknowledgments. We thank G. DeMatteo and M. Burke for devoted laboratory assistance. Michael Snyder (now at Stanford University), James Fristrom, and Don Silvert of the University of California, Berkeley, were most generous in providing sequence data prior to publication. The NIH provided funds via Research Grant AI-14753.

References

- Andersen SO, Chase AM, Willis JH (1973) The amino acid composition of cuticles from *Tenebrio molitor* with special reference to the action of juvenile hormone. Insect Biochem 3:171-180
- Beverley SM, Wilson AC (1982) Molecular evolution in Drosophila and higher Diptera. I. Micro-complement fixation studies of a larval hemolymph protein. J Mol Evol 18:251-264
- Bhown AS, Mole JE, Bennett JC (1981) An improved procedure for high-sensitivity microsequencing. Use of aminoethyl-aminopropyl glass beads in the Beckman sequencer and the Ultrasphere ODS column for phenylthiohydantoin amino acid identification. Anal Biochem 110:355-359
- Borror DJ, Delong DM, Triplehorn CA (1976) An introduction to the study of insects. Holt, Rinehart & Winston, New York.
- Chihara CJ, Silvert DJ, Fristrom JW (1982) The cuticle proteins of *Drosophila melanogaster*: Stage specificity. Dev Biol 89: 379–388
- Collier GE, MacIntyre RJ (1977) Microcomplement fixation studies on the evolution of alpha-glycerophosphate dehydrogenase within the genus *Drosophila*. Proc Natl Acad Sci USA 74:684–688
- Cox DL, Willis JH (1982) Cuticular proteins of Hyalophora cecropia: new insights into insect morphology and metamorphosis. In: Proceedings of the Entomological Society of America and Canada, 1982. Entomological Society of America and Canada, Toronto, p 63
- Driskell WJ (1974) The role of tyrosine in the tanning of the puparium of *Drosophila melanogaster*. Ph.D. thesis, California Institute of Technology, (University Microfilm no. 74-17, 942. Ann Arbor, Michigan, USA 48106)
- Fraenkel G, Baskaran A (1973) Pupariation and pupation in cylorrhaphous flies (Diptera): terminology and interpretation. Ann Ent Soc Am 66:418-422
- Fraenkel G, Hsiao C (1967) Calcification tanning and the role of ecdysone in the formation of the puparium of the facefly, *Musca autumnalis*. J Insect Physiol 13:1387-1394

- Fristrom JW, Hill RJ, Watt F (1978) The procuticle of Drosophila: heterogeneity of urea-soluble proteins. Biochemistry 19:3917-3924
- Fullmer SC, Wasserman RH (1979) Analytical peptide mapping by high performance liquid chromatography. J Biol Chem 254:7208-7212
- Hackman RH, Goldberg M (1976) Comparative chemistry of arthropod cuticular proteins. Comp Biochem Physiol [B] 55: 201-216
- Kimura S, Strout HV, Lipke H (1976) Peptidochitodextrins of Sacrophaga bullata: nonidentity of limit glycopeptides from larval and puparial cuticle. Insect Biochem 6:65-77
- Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-685
- Lipke H, Strout K, Henzel W, Sugumaran M (1981) Structural proteins of sarcophagid larval exoskeleton. Composition and distribution of radioactivity derived from [7-14C] dopamine. J Biol Chem 256:4241-4246
- Lipke H, Sugumaran M, Henzel W (1983) Mechanisms of sclerotization in dipterans. Adv Insect Physiol 17:1-84
- Manrique A, Lasky M (1981) Agarose-Sephadex: an improved matrix for preparative flatbed isoelectric focusing. Electrophoresis 2:315-320
- Marchalonis JJ, Weltman JE (1971) Relatedness among proteins: a new method of estimation and its application to immunoglobulins. Comp Biochem Physiol [B] 38:609-625
- Snyder M, Hirsh J, Davidson M (1981) The cuticle genes of

Drosophila: a developmentally regulated gene cluster. Cell 25: 165-177

- Snyder M, Hunkapiller M, Yuen D, Silvert D, Fristrom J, Davidson N (1982) Cuticle protein genes of Drosophila: structure, organization and evolution of four clustered genes. Cell 29:1027-1040
- Sugumaran M, Lipke H (1982) Crosslink precursors for the dipteran puparium. Proc Natl Acad Sci USA 79:2480-2484
- Sugumaran M, Lipke H (1983a) A new method for studying the ratio of quinone and B-sclerotization. Insect Biochem 13: 307-312
- Sugumaran M, Lipke H (1983b) Quinone methode formation from 4-alkylcatechols-a novel reaction catalyzed by cuticular polyphenol oxidase. FEBS Lett 155:65-68
- Sugumaran M, Henzel WJ, Mulligan K, Lipke H (1982) Chitinbound protein of sarcophagid larvae: metabolism of aromatic constituents. Biochemistry 21:6509-6515
- Willis JH, Regier JC, Debrunner BA (1981) The metamorphosis of arthropodin. In: Bhaskaran G, Friedman S, Rodrigue JA (eds) Current topics in endocrinology and nutrition. Plenum, New York, pp 27-46
- Wilson AC, Carlson SS, White TJ (1977) Biochemical evolution. Annu Rev Biochem 46:573-639
- Zalut C, Henzel WS, Harris WHJ (1980) Micro quantitative Edman manual sequencing. J Biochem Biophys Methods 3: 11 - 17
- Received April 25, 1984/Revised November 21, 1984