

5 Cuticular Proteins

Judith H Willis

Department of Cellular Biology,
University of Georgia, Athens, GA, USA

Nikos C Papandreou

Department of Cell Biology and Biophysics,
Faculty of Biology, University of Athens,
Athens, Greece

Vassiliki A Iconomidou

Department of Cell Biology and Biophysics,
Faculty of Biology, University of Athens,
Athens, Greece

Stavros J Hamodrakas

Department of Cell Biology and Biophysics,
Faculty of Biology, University of Athens,
Athens, Greece

© 2012 Elsevier B.V. All Rights Reserved

5.1. Introduction	134
5.2. Cuticle Structure and Synthesis	135
5.2.1. Cuticle Morphology	135
5.2.2. The Site of Synthesis of Cuticular Proteins	137
5.2.3. Tracheal Cuticular Proteins	139
5.3. Classes of Proteins Found in Cuticles	139
5.3.1. Non-Structural Proteins	139
5.3.2. Structural Proteins	141
5.3.3. Motifs Found in Cuticular Proteins that do not Define Families	150
5.3.4. Glycosylation of Cuticular Proteins	150
5.4. Genomic Information	151
5.4.1. Introduction	151
5.4.2. Chromosomal Linkage of Cuticular Protein Genes	151
5.4.3. Intron Structure of Cuticular Protein Genes	152
5.4.4. Regulatory Elements	153
5.5. Interactions of Cuticular Proteins with Components of Cuticle	153
5.5.1. Secondary Structure Predictions	153
5.5.2. Experimental Studies of Cuticular Protein Secondary Structure	154
5.5.3. Modeling of Cuticular Proteins	154
5.5.4. Fusion Proteins Establish a Role for the Extended R&R Consensus	159
5.5.5. Members of Other Cuticular Protein Families Analyzed for Chitin Binding	159
5.5.6. Summary of Interaction Studies	160
5.6. Summary and Future Challenges	160

5.1. Introduction

In the first edition of this series, Silvert (1985) outlined several major areas of uncertainty regarding cuticular proteins. The questions raised were: Were proteins extracted from cuticle authentic cuticular proteins, or might some be contaminants of adhering cells and hemolymph? Was the epidermis the sole site of synthesis of cuticular proteins, or were some synthesized in other tissues and transported to the cuticle? What was the relation among cuticular proteins of various developmental stages? Did cuticular proteins share common structural features?

That review presented all the cuticular protein sequence data then available – four complete and three partial sequences from *Drosophila melanogaster*, and one partial sequence from *Sarcophaga bullata*. The considerable sequence similarity seen with those limited data indicated that cuticular protein genes belonged to multi-gene families, and the even more limited genomic information revealed that similar genes were adjacent on a chromosome.

Progress over the next two decades was impressive, but not surprising, given the advances in relevant techniques. Elegant immunolocalization analyses solved the

problem of the sources of cuticular proteins. The 2005 version of this article reported that over 300 cuticular protein sequences had been recognized, from 6 orders and over 20 species of insects. Progress since 2005 has been predominantly in identifying new cuticular protein sequences. Since then, several whole genomes have been annotated and hundreds of new sequences have been posted, based on EST (expressed sequence tag) analyses. It is no longer possible to list the available sequences. Rather, this chapter will summarize what was learned from these data. Some areas of analysis have not progressed far since the 2005 version, and will be repeated here without revision.

5.2. Cuticle Structure and Synthesis

5.2.1. Cuticle Morphology

5.2.1.1. Terminology The descriptive terms used here to describe the regions of cuticle have been simplified according to Locke's (2001) cogent suggestions for new nomenclature. He proposes the use of the term "envelope" to describe the outermost layer of cuticle, rather than the previous term "cuticulin." At the start of each molt cycle, the smooth apical plasma membrane forms microvilli with plaques at their tips where the new envelope assembles. This discrete layer of 10–30 nm serves not only to protect the underlying epidermis from molting fluid enzymes that begin to digest the old cuticle, but also, as Locke points out, affects "resistance to abrasion and infection, penetration of insecticides, permeability, surface reflectivity, and physical colors." The sequences and properties of its constituent proteins remain unknown.

Next formed is the epicuticle, about 1 μm in thickness. This chitin-free layer (but see Section 5.2.1.3) is stabilized by quinones. It was formerly referred to as the "inner epicuticle," with cuticulin being the outer.

Former arguments about the precise distinction between exo- and endo-cuticle are eliminated by Locke's lumping together of the inner regions of the cuticle under the term "procuticle," encompassing both pre-ecdysial and post-ecdysial secretions. The procuticle, then, is the region that combines chitin and cuticular proteins in various combinations, and becomes sclerotized (Andersen, 2010a; see also Chapter 6 in this volume) and pigmented to varying degrees. This is the region depicted in electron micrographs showing stacks of precisely oriented lamellae. According to Locke (2001), it is the secretion of chitin fibers by apical microvilli that apparently bend in concert across the epithelial sheet to orient the laminae that gather into lamellae (Neville, 1975; Locke, 1998). Alternative mechanisms have been proposed by Moussian (2010), involving movement of the "chitin synthesis complex" across the cell surface or merely self-assembly. While knowledge of the process of secreting and assembling such

a highly ordered structure is limited, details about the proteins associated with the lamellae are now voluminous.

5.2.1.2. Growth of the cuticle within an instar Central to the issue of cuticle structure is the important fact that considerable cuticle growth can occur during an intermolt period (Williams, 1980), some of it by a smoothing out of macro- and microscopic folds and pleats (Carter and Locke, 1993). During intra-instar growth, new cuticular proteins are interspersed among the old, necessitating a model of chitin–protein and protein–protein interactions that will permit such intussusception (Condoulis and Locke, 1966; Wolfgang and Riddiford, 1986).

5.2.1.3. Localization of cuticular proteins within the cuticle Precise localization of cuticular proteins within the cuticle, and even within cellular organelles, has been made possible with immunogold labeling of electron-microscopic sections. Here, a specific primary antibody is bound to the sections and visualized with a secondary antibody conjugated to colloidal gold particles.

Antibodies have been raised against extracts of whole cuticle or isolated electrophoretic bands, and the specificity of each antibody ascertained with Western blots. While each polyclonal antibody raised against a single band was specific for the immunizing protein, monoclonals raised against cuticular extracts frequently reacted with more than one electrophoretic band.

One concern with immunolocalization is that as cuticular proteins become modified in the cuticle by binding to chitin or by becoming sclerotized, the immunizing epitopes might become masked – a problem that should be more serious with monoclonal than with polyclonal antibodies. All groups recognized that while the presence of an antigen is significant, its absence may reflect no more than such masking.

This concern is significant when one considers results of immunolocalization in the assembly zone, the region of cuticle directly above the microvilli. It is here that chitin secreted from the tips of the microvilli interacts with cuticular proteins secreted into the perimicrovillar space. Immunolocalization studies revealed only a few of the cuticular proteins within the perimicrovillar space, but the same ones and others were abundant in the assembly zone directly above it (Locke *et al.*, 1994; Locke, 1998). The authors' conclusion was that the assembly zone "is where we should expect proteins to unravel and expose most epitopes in preparation for assuming a new configuration as they stabilize in the maturing cuticle." Wolfgang *et al.* (1986, 1987) found two *Drosophila melanogaster* cuticular proteins exclusively in this zone, and suggested they might function in cuticle assembly. Locke *et al.* (1994) point out that it was common for antibodies raised against *Calpodes ethilus* proteins to react more strongly with the assembly zone than with more mature regions of cuticle,

where sclerotization and chitin binding might mask epitopes. Thus, further substantial evidence than the failure to detect a protein in more mature regions is needed to confirm that it belonged exclusively to the assembly zone.

It was known from earlier work on protein and mRNA distribution that cuticles from different metamorphic stages and different anatomical regions had different cuticular proteins, and that there may be a change in cuticular proteins synthesized by a single cell within a molt cycle (for review, see Willis, 1996). Such a transition in proteins synthesized is especially apparent at the time of ecdysis, and, in some insects, late in the instar. Consistent with this, immunolocalization revealed different proteins in morphologically distinct early and late lamellae in *D. melanogaster* pupae, and *Tenebrio molitor* and *Manduca sexta* larvae (Doctor *et al.*, 1985; Fristrom *et al.*, 1986; Wolfgang and Riddiford, 1986; Wolfgang *et al.*, 1986; Lemoine *et al.*, 1989, 1993; Bouhin *et al.*, 1992a, 1992b; Rondot *et al.*, 1998). Only two proteins with known sequence are among this group: TMACP22 (P26968.1) and TMLPCP22 (P80686.2).

Csikos and colleagues (1999) have used immunohistochemistry to follow some of *Manduca's* cuticular proteins throughout the molt cycle. These proteins are obviously in a dynamic state as they move from epidermis to cuticle to molting fluid to fat body, and then apparently back to cuticle via the hemolymph. More detailed studies are needed to learn if the same molecules make the return trip, and whether their initial passage from molting fluid into the hemolymph is solely via uptake and then basal secretion by the epidermis, or whether the midgut plays a role, since lepidopteran larvae drink their molting fluid (Cornell and Pan, 1983).

The findings with epicuticle, the first region to be secreted beneath the envelope, were complex. None of the monoclonal antibodies that recognized *Tenebrio* cuticular proteins reacted with epicuticle (Lemoine *et al.*, 1990). On the other hand, arylphorin from *Calpodes* has been localized to epicuticle and no other cuticular region (Leung *et al.*, 1989), and several proteins, of unknown sequence, were found both in the epicuticle and in the lamellar regions of the procuticle in *D. melanogaster* (Fristrom *et al.*, 1986) and *Calpodes* (Locke *et al.*, 1994). This finding of cuticular proteins in both epicuticle and lamellar regions was surprising, since the epicuticle had always been described as lacking chitin (*cf.* Fristrom *et al.*, 1986, Fristrom and Fristrom, 1993) and thus was expected to have unique proteins. A study of moth olfactory sensilla detected chitin in the procuticle with gold-conjugated wheat germ agglutinin, but it was not found in the epicuticle (Steinbrecht and Stankiewicz, 1999).

In addition to temporal differences in the secretion of cuticular proteins by single cells, there may be regional differences in the cuticle secreted by single cells. Individual epidermal cells of the articulating membranes

(intersegmental membranes) in *Tenebrio* secrete a cuticle with sclerotized cones embedded in softer cuticle. Two of the classes of monoclonal antibodies raised against *Tenebrio's* larval and pupal cuticular proteins recognized proteins in these cones. The same antibodies recognized proteins in cuticles in other regions that were destined to be sclerotized. Different antibodies recognized the proteins in the softer cuticle (Lemoine *et al.*, 1990, 1993).

Locke *et al.* (1994) were able, using carefully reconstructed sections of *Calpodes* larval cuticle, to distinguish one protein (C36) that was found with the same distribution as the chitin microfibrils that had been visualized with wheat germ agglutinin (WGA), a lectin that recognizes N-acetylglucosamine, while other antigens failed to show this distribution. Notably, only C36 isolated from cuticle reacted with WGA on lectin blots. Based on this evidence, Locke *et al.* (1994) suggest that the isolated protein may have obtained its N-acetylglucosamine from chitin.

5.2.1.4. Cuticles formed following disruption of normal metamorphosis

Treatment of many insects with juvenile hormone (JH) causes them to resynthesize a cuticle with a morphology characteristic of the current metamorphic stage, rather than the next. Thus, in *Tenebrio*, treatment of pupae with JH prior to pupal-adult apolysis causes the formation of a second pupa rather than an adult. Earlier work revealed that these second pupae had proteins with the same electrophoretic mobility as those extracted from normal pupae (Roberts and Willis, 1980; Lemoine *et al.*, 1989). A combination of Northern analysis and *in situ* hybridization demonstrated that second pupae have the same cuticular protein mRNAs and protein localization as normal pupae (Lemoine *et al.*, 1993; Rondot *et al.*, 1998). Adult cuticular proteins are not deposited in these cuticles, and the adult mRNAs do not appear (Lemoine *et al.*, 1989, 1993; Bouhin *et al.*, 1992a, 1992b; Charles *et al.*, 1992). Some JH-treated *Tenebrio* pupae form two cuticles, the first pupal-like in morphology and the second with adult features. The adult-like cuticle was shown with immunolocalization to have TMACP22 (P26968.1) (Bouhin *et al.*, 1992a). If JH is applied too late to form a perfect second pupa, the next cuticle formed will be a composite with morphological features of two metamorphic stages (Willis *et al.*, 1982). Bouhin *et al.* (1992b) found that all the epidermal cells laying down such a composite cuticle had mRNAs for TMACP22.

Zhou and Riddiford (2002) used Northern analysis to characterize the somewhat nondescript cuticles made by *D. melanogaster* that had been manipulated by mis-expressing the gene *br* (*broad*), which codes for a transcription factor that first appears before the larval/pupal molt in flies and moths. By following mRNAs for the adult cuticular protein ACP65A (CG10297) or the pupal cuticular protein Edg78E (CG7673), they were able to

demonstrate the essential role of *br* in directing pupal development, and thereby clarified the perplexing action of juvenoids in the higher Diptera.

5.2.2. The Site of Synthesis of Cuticular Proteins

One of the unresolved issues addressed in Silver's (1985) review was the site of synthesis of cuticular proteins. This might appear to be a trivial issue, for one would expect that the epidermis that underlies the cuticle would synthesize the cuticular proteins. There are, however, reports in the literature that proteins found in the hemolymph were present in cuticle, and even that labeled proteins injected into the hemolymph would appear in cuticle. Silver discussed the possibility that the injected protein had been broken down and resynthesized so that the cuticular protein was labeled solely because its constituent amino acids had come from a labeled pool.

Five methods have now provided data that address the site of synthesis of cuticular proteins. The most common method is to learn in what tissues and at which stages mRNA is present for a particular cuticular protein by detecting its presence via Northern analysis, RT-PCR, or qRT-PCR. This method is so common that specific examples will not be given. The second method is to incubate epidermis or integument *in vitro* with radioactive amino acids, separate the proteins, and compare the electrophoretic mobility of the labeled proteins to proteins isolated from cleaned cuticles. A third method is to isolate mRNAs from tissues and translate these *in vitro* with commercially available wheat germ extracts or rabbit reticulocytes, and compare the translation products to known cuticular proteins. The fourth method is *in situ* hybridization, and the fifth is immunolocalization to visualize proteins within the endoplasmic reticulum and Golgi apparatus.

The first three methods suffer from the possibility that tracheae and adhering tissues, fat body, muscles, and hemocytes contribute to the mRNA pool. Both labeling methods suffer from the problem that cuticular proteins are notoriously sensitive to solubilizing buffer and gel conditions (pH, urea concentration) (Cox and Willis, 1987a), and, unless cuticular protein standards and labeled translation products are mixed prior to electrophoresis, they may not show identical electrophoretic mobility even in adjacent lanes. Some workers have precipitated labeled translation products with antibodies raised against extracts of cuticle or individual cuticular proteins, then solubilized the precipitate, run it on a gel, and detected the labeled product with fluorography. Csikos *et al.* (1999) used Western blots of translation products to identify cuticular proteins. Since cuticular proteins are destined for secretion from cells, they have a signal peptide that is cleaved before the protein is secreted into the cuticle. Hence, translation products made *in vitro* will be larger than the protein extracted from cuticle. There are two methods to

circumvent this problem. The translation products can have their signal peptides cleaved by adding a preparation of canine microsomes, or antibodies against cuticular proteins (specific or against an extract) can be used to precipitate the translation products before they are solubilized and run on a gel. Either method allows some certainty in the comparison of these *in vitro* translation products with authentic cuticular proteins. It was also found that some commercial preparations of wheat germ extract have endogenous signal peptide processing activity (Binger and Willis, 1990).

Frequently, ^{35}S -methionine was used for metabolic labeling of integument and for *in vitro* translation. This is an unfortunate choice, as most mature cuticular proteins lack methionine residues (see section 5.3.2.1). The initiator methionine will be lost, along with the entire signal peptide. Clear differences in labeling patterns with ^{35}S -methionine and ^3H -leucine have been found, with none of the major proteins from pharate adult cuticle of *D. melanogaster* or from larval cuticles of *H. cecropia* showing methionine labeling (Roter *et al.*, 1985; Willis, 1999). Why, then, did several studies find all of the known cuticular proteins labeled with methionine? Perhaps the finding that ^{35}S -methionine can donate its label to a variety of amino acids in preformed proteins (Browder *et al.*, 1992; Kalinich and McClain, 1992) explains its appearance, and suggests that it needs to be used with caution for such studies with cuticular proteins.

The fourth method is *in situ* hybridization, where specific mRNAs can be identified in the epidermis. *In situ* hybridization allows one to be somewhat more discerning about the site of synthesis of a cuticular protein, because it is possible to monitor the presence or absence of a particular mRNA at the level of an individual cell. With this technique, integument is fixed and sectioned, and then probed with a labeled cDNA or cRNA, allowing the identification of particular regions of the epidermis by examining the morphology of the overlying cuticle. With most detection methods, contaminating tissues and precise regions of the epidermis can be identified, and the presence of the particular mRNA in them can be assessed. Thus, this technique identifies the location of the mRNAs recognized by the specific probe used. It was this technique that revealed the precision with which mRNAs are produced, for abrupt boundaries of expression occur between sclerites and intersegmental membranes (Rebers *et al.*, 1997), or at muscle insertion zones (Horodyski and Riddiford, 1989), or next to specialized epidermal cells (Horodyski and Riddiford, 1989; Rebers *et al.*, 1997). This technique even revealed the presence of mRNA for cuticular proteins in epithelia of imaginal discs from young larvae (Gu and Willis, 2003). A limitation of the technique is that some cRNA probes bind to the cuticle itself, possibly obscuring detection of mRNA in the underlying epidermis (Fechtel *et al.*, 1989, Gu and

Table 1 Evidence for the Association of Location or Type of Cuticle and Sequence Class of Some Cuticular Proteins

Species	Protein	Sequence Class	Localization ^a	Nature of Evidence ^b	When Deposited	Reference
<i>Bombyx mori</i>	BMLCP18	RR-1	Imaginal discs	EST		Gu and Willis (2003)
<i>Drosophila melanogaster</i>	EDG-78	RR-1	Larval and imaginal cells of prepupa	ISH		Fechtel <i>et al.</i> (1989)
<i>Drosophila melanogaster</i>	EDG-84	RR-2	Imaginal disc cells	ISH		Fechtel <i>et al.</i> (1989)
<i>Drosophila melanogaster</i>	PCP	RR-1	Prepupal thorax and abdomen	ISH		Henikoff <i>et al.</i> (1986)
<i>Hyalophora cecropia</i>	HCCP12	RR-1	Soft cuticle; imaginal discs	CD and ISH		Cox and Willis (1985), Gu and Willis (2003)
<i>Hyalophora cecropia</i>	HCCP66	RR-2	Hard cuticle	CD and ISH		Cox and Willis (1985), Gu and Willis (2003)
<i>Locusta migratoria</i>	LM-ACP7	RR-2	Hard cuticle	CD		Andersen <i>et al.</i> (1995)
<i>Locusta migratoria</i>	LM-ACP8	RR-2	Hard cuticle	CD		Andersen <i>et al.</i> (1995)
<i>Locusta migratoria</i>	LM-ACP19	RR-2	Hard cuticle	CD		Andersen <i>et al.</i> (1995)
<i>Manduca sexta</i>	CP14.6	RR-1	Soft cuticle	ISH		Rebers <i>et al.</i> (1997)
<i>Manduca sexta</i>	LCP 16/17	RR-1	Soft cuticle	ISH		Horodyski and Riddiford (1989)
<i>Tenebrio molitor</i>	ACP17	Glycine-rich	Hard cuticle	ISH	Strongest post-ecdysis	Mathelin <i>et al.</i> (1995, 1998)
<i>Tenebrio molitor</i>	ACP20	RR-2	Hard cuticle	ISH	Primarily pre-ecdysis	Charles <i>et al.</i> (1992)
<i>Tenebrio molitor</i>	ACP-22	RR-2	Hard cuticle	ISH, mAB	Pre-ecdysis	Bouhin <i>et al.</i> (1992a, 1992b)
<i>Tenebrio molitor</i>	TMLPCP22	51 aa motif	Hard and soft cuticle pre-ecdysis, then only soft cuticle	ISH, mAB	Primarily pre-ecdysis	Rondot <i>et al.</i> (1998)
<i>Tenebrio molitor</i>	TMLPCP23	51 aa motif	Hard and soft cuticle	ISH	Only pre-ecdysis	Rondot <i>et al.</i> (1998)
<i>Tenebrio molitor</i>	TMLPCP29	RR-3 and 18-residue motif	Hard and soft cuticle, except not posterior borders of sclerites	ISH	Post-ecdysis	Mathelin <i>et al.</i> (1998)

^aFor *in situ* hybridization, cuticle type was determined by nature of cuticle overlying the epidermis.

^bCD, careful dissection prior to extraction of proteins; ISH, *in situ* hybridization used to to localize mRNA; mAB, monoclonal antibody immunolocalization; EST, from *Bombyx* EST project (Mita *et al.* 2002).

Willis, 2003). Fechtel *et al.* (1989) found this artifact to be cuticle-type- as well as strand- and probe-specific. Results from several species are summarized in **Table 1**.

The fifth method, immunolocalization, was described earlier in conjunction with localization of specific proteins within the cuticle, but it can also be used to identify the site of synthesis by looking for a particular protein within the endoplasmic reticulum or Golgi apparatus (Sass *et al.*, 1994a, 1994b).

The results from mRNA detection, metabolic tissue labeling, and *in vitro* translations reveal that all cuticular proteins with known sequences or for which specific probes are available are synthesized by the integumental preparations. Different proteins are synthesized at different times in a molt cycle, and in different anatomical regions, and there are some cuticular proteins whose synthesis is stage-specific. Differences in the presence of mRNA parallel the appearance of labeled proteins, indicating that

much of the temporal and spatial control of cuticular protein synthesis is at the level of transcription. As mentioned above, however, all three of these methods are limited by the possible contamination of tissues by non-epidermal cells, and by their inability to address heterogeneity of cell types within the epidermis.

A microarray analysis of isolated hemocytes from *An. gambiae* revealed the presence of mRNA for nine cuticular proteins (Baton *et al.*, 2009). Transcripts for *AgamCPR26* and *AgamCPR90* were significantly higher in adults challenged with heat-killed *Micrococcus luteus* than in naïve individuals. A massive study on hemocytes in *D. melanogaster* found significant levels of transcript for *DmLLCP1-4* in hemocytes from both naïve and bacteria-challenged larvae (Irving *et al.*, 2005).

Studies that have combined tissue labeling or *in vitro* translations with immunolocalization have at last clarified

the relationship between hemolymph and cuticular proteins with identical electrophoretic and immunological properties. The most comprehensive studies of protein trafficking were carried out in *Calpodes*, and revealed four classes of exported proteins that are handled by the epidermis.

These findings are so important that the experimental methodology is worth discussing. The first approach used was to seal sheets of final instar integument into a bathing chamber so there could be no leakage from the cut edges of the tissue, and then find what proteins were made in a 2-hour exposure to ^{35}S -methionine. Three classes of proteins were identified with this procedure; one was secreted exclusively into the cuticle (C class), a second appeared in the bathing fluid and hence had been secreted basally (B class), while the third was secreted in both directions (BD class) (Palli and Locke, 1987). Immunolocalization of numerous other *Calpodes* proteins (of unknown sequence) confirmed the existence of these three routing classes of epidermal proteins. A fourth class, the T class, was identified for proteins transported into cuticle but not synthesized by the epidermis. Its presence eliminated any concerns that the classes might be artifacts from labeling with ^{35}S -methionine (Sass *et al.*, 1993).

One member of the T class (T66) was studied in more detail. It was localized by immunogold throughout the cuticle, and, although found in epidermal cells, was not found in association with the Golgi apparatus, confirming its transcellular transport, rather than synthesis by the epidermis. A subsequent study identified the exclusive site of its synthesis as spherulocytes (Sass *et al.*, 1994a).

Whether the BD proteins are secreted from both apical and basal borders of epidermal cells is still not clear. Locke (1998, 2003) now favors the possibility that all secretion is apical, where the Golgi are concentrated, and that the secreted proteins are subsequently taken back into the cell from perimicrovillar space and transported in vesicles to the basal surface, where the contents are released into the hemolymph.

In conclusion, it is now clear that the epidermis can synthesize both cuticular and hemolymph proteins. It can also transport proteins made in tissues other than epidermis from hemolymph to cuticle.

5.2.3. Tracheal Cuticular Proteins

An often-neglected source of cuticle in insects is the tracheal system. Since tracheae are associated with all insect tissues, caution is needed in interpreting the significance of the presence of mRNAs or cuticular proteins from non-integumental tissues. Cox and Willis (1985) recognized that some of the proteins from tracheae had the same isoelectric points as proteins isolated from integumentary cuticle. A further study was carried out a decade later by Sass *et al.* (1994b), combining electrophoretic

analysis with immunogold labeling. Chitin was localized with wheat germ agglutinin, and found in all regions of tracheae and tracheoles except the taenidial cushion. Antibodies that had been raised against individual electrophoretic bands from integumentary extracts represented proteins from all four classes of integumentary peptides. Some C proteins, those from the surface cuticle, were found associated with chitin, but only in taenidia; other C proteins were in the general matrix, with and without chitin. The B and BD peptides were only found in the taenidial cushion, the region lacking chitin. It appears that hemolymph peptides that are synthesized by the epidermis may be tracheal cuticle precursors. The one T protein studied (T66, made in spherulocytes) was also found in the general matrix. An important insight from this study was the conclusion that: "The extremely thin tracheal epithelium suggests that transepithelial transport might supply proteins to the tracheal cuticle more evenly than Golgi complex secretions" (Sass *et al.*, 1994b). Analysis of tracheal morphogenesis is an active field that has recently been reviewed (Centanin *et al.*, 2010; Moussian, 2010; see also Ghabrial *et al.*, 2003). Little information is available about the proteins that contribute to tracheae. Gasp (a member of the CPAP-3 family) was found to be restricted to tracheae in *D. melanogaster* (Barry *et al.*, 1999), but transcripts from its clear ortholog in the lepidopteran *Choristoneura fumiferana* were associated primarily with the body surface epidermis (Nisole *et al.*, 2010).

5.3. Classes of Proteins Found in Cuticles

5.3.1. Non-Structural Proteins

Some representative non-structural proteins that have been identified in cuticle are listed in Table 2.

5.3.1.1. Pigments Proteins from three classes of pigments used in cuticle – insecticyanins and two different yellow proteins – have been sequenced. The insecticyanins are blue pigments made by the epidermis and secreted into both hemolymph and cuticle. They are easily extracted from cuticle with aqueous buffers. Members of the lipocalin family, they are present as tetramers with the gamma isomer of biliverdin IX situated in a hydrophobic pocket. In the cuticle, in cooperation with carotenes, they confer green coloration. Their structure has been determined to 2.6 Å by X-ray diffraction (Holden *et al.*, 1987), making them structurally the best characterized cuticular proteins. Two genes code for insecticyanins in *Manduca* (Li and Riddiford, 1992).

The yellow protein in *D. melanogaster*, coded by the *y* gene (CG3757), has been localized with immunocytochemistry in cuticles destined to become melanized (Kornezov and Chia, 1992). Thus, it was found in

Table 2 Characteristics of Some Non-Structural Proteins that have been found in Cuticle

Species	Protein Name	Number of Amino Acids ^a	Function	Sequence Method ^b	Identifier ^c
<i>Schistocerca gregaria</i>	Putative carotene binding protein	250	Transfers carotene into cuticle	DS	13959427
<i>Caliphora vicina</i>	ARYLPHORIN A4	743	Found in cuticle	CT	114232
	ARYLPHORIN C223	743		CT	114236
<i>Drosophila melanogaster</i>	YELLOW	520	Positions melanin pigment in cuticle	CT	140623
<i>Bombyx mori</i>	CECROPIN A	41	Defense protein	CT	2493573
	CECROPIN B	41	Defense protein	CT	1705754
	PROPHENOLOXIDASE	675	Metanization enzyme	CT	13591614
<i>Calpodes ethilus</i>	CECP 22	169	Cuticle digestion	CT	4104409
<i>Manduca sexta</i>	ARYLPHORIN α	684		CT	114240
	ARYLPHORIN β	687		CT	1168527
	INSECTICYANIN A	189	Blue pigment	CT	124151
	INSECTICYANIN B	189	Blue pigment	CT	124527
	SCOLEXIN A	279	Serine protease immune protein	CT	4262357
	SCOLEXIN B	279	Serine protease immune protein	CT	4262359

^aSequence length of mature peptide; signal peptides were deleted using data from authors or SignalIP V2.0(<http://www.cbs.dtu.dk/services/SignalIP> 2.0/).

^bDS, direct sequencing of protein; CT, conceptual translation of a cDNA, genomic region, or EST product.

^cProtein sequences and additional annotation can be found at: <http://www.ncbi.nlm.nih.gov/protein>

association with larval mouth hooks, denticle belts, and Keilin's organs. Mutants of y lack black pigment in the affected cuticular region. Mutant analysis revealed two classes of mutants; those that affect all types of cuticle at all stages, and those affecting only particular areas of specific stages. At least 40 different adult cuticular structures could express their color independently (Nash, 1976), and the regulatory regions responsible for some of the stage and regional specificity have been identified (Geyer and Corces, 1987). The yellow protein has been described as a structural component of the cuticle that interacts with products from the gene, *ebony*, a beta-alanyl-dopamine synthase, to allow melanin to be deposited. Flybase (<http://flybase.bio.indiana.edu/>) reports that 1005 different alleles of y have been described, in 775 references, beginning in 1916. The complete sequence of y has been determined for 13 species of *Drosophila* in addition to *D. melanogaster*. An examination of y expression revealed that both cis- and trans-regulation are responsible for differences in pigmentation patterns among different species (Wittkopp *et al.*, 2002, 2009). There is no evidence for a known chitin-binding domain in the yellow protein; the only domain recognized is pfam03022 (major royal jelly protein). Although the sequence for yellow is 37% identical and 56% similar to a dopachrome conversion enzyme from *Aedes aegypti* that is involved in the melanotic encapsulation immune response, yellow itself evidently is devoid of enzyme activity (Han *et al.*, 2002). As a further complication, there are 13 other genes in *D. melanogaster* related to y ; most of their products do not seem to affect

pigmentation. *T. castaneum* also has 14 y homologs, and a comprehensive study using RNAi and mass spectrophotometric analyses revealed diverse activities, many involving the cuticle, with only the ortholog of *Dmely* having a role in cuticle melanization (Arakane *et al.*, 2010). Orthologs of *Dmely* also play a role in cuticle pigmentation in *B. mori* and *Papilio xuthus* (Futahashi and Fujiwara, 2005; Futahashi *et al.*, 2008).

Another distinct cuticular protein (P82886.1) implicated in pigmentation, putatively beta-carotene binding, has been isolated from extracts of cuticle from mature adult *Schistocerca gregaria* using column chromatography to isolate a protein that was yellow in color. It bears significant sequence similarity to various insect juvenile-hormone-binding proteins, as well as odorant-binding proteins. Wybrandt and Andersen (2001) suggest that it is involved in transport of carotenes into epidermis and then the cuticle.

5.3.1.2. Enzymes Some of the enzymes involved in sclerotization have been identified in cuticle. Since they are discussed by Andersen in Chapter 6, and in a recent review (Andersen, 2010a), they will not be considered here.

Some enzymes that belong to the molting fluid become evident as the electrophoretic banding pattern of proteins isolated from cuticle changes as *Calpodes* initiates molting at the end of the fifth instar, with the most conspicuous change being the appearance of a band of 19 kDa. Antibodies raised against this protein were used to isolate a cDNA from a library cloned in an expression vector.

The conceptual translation revealed a “cuticular molt protein” (AAD02029.1, also called CEPP22). Its sequence suggested it might have amidase activity. Further analysis revealed that the protein was present in the cuticle before each molt, and was also found in molting fluid. [Marcu and Locke \(1998, 1999\)](#) present evidence that this protein may be activated by proteolysis, and speculate that it may function to cleave an amidic bond between N-acetylglucosamine from chitin and amino acids in cuticular proteins.

Enzymes involved in digesting the old cuticle are temporary residents in cuticle. These include proteases and chitinases. Their interaction is discussed by [Marcu and Locke \(1998\)](#).

5.3.1.3. Defense proteins Also found in the cuticle are components of the insect defense system. In one study, cuticle was removed from *Bombyx* larvae 24 hours after they had been abraded with emery paper and exposed to bacteria. The antibacterial peptide, cecropin, was purified from the cuticles ([Lee and Brey, 1994](#)). Both pro-phenoloxidase and a zymogen form of a serine protease capable of activating it have been extracted from *Bombyx* larval cuticle. Colloidal gold secondary antibodies revealed that the pro-phenoloxidase was localized throughout the epi- and procuticle, and in a conspicuous orderly array on the basal side of the helicoidal chitin lamellae. An extra-epidermal source is likely for this enzyme since no labeling was found in the epidermis, and neither was mRNA detected in the epidermal cells. It is assumed to function in the melanization that occurs in response to injury ([Ashida and Brey, 1995](#)).

[Molnar et al. \(2001\)](#) presented immunological evidence for a protein related to the defense protein scolexin in the cuticle of *Manduca*. This protein exists in two forms in *Manduca*, but the antibody used did not distinguish between them.

The cuticle also appears to be the repository for a peptide (HCP; GI:240104242; 2RPS_A) that stimulates aggregation and movement of hemocytes in the moth *Pseudaletia separate* (*Mythimna separate*) ([Nakatogawa et al., 2009](#)).

5.3.1.4. Arylphorins The final class of non-structural proteins is the arylphorins, proteins with high content of aromatic amino acids and some lipid. These proteins, first identified from hemolymph, have been of special interest since the discovery by [Scheller et al. \(1980\)](#) that although calliphorin (the arylphorin from *Calliphora*) was found in cuticle it seemed to come from the hemolymph, because labeled calliphorin injected into the hemolymph appeared in cuticle. But there is also evidence that the epidermis is capable of synthesizing arylphorins, for [Riddiford and Hice \(1985\)](#) had detected arylphorin mRNA in the epidermis of *Manduca*.

[Palli and Locke \(1987\)](#) used an anti-arylphorin antibody to identify an 82-kDa protein made in *Calpodex* integumental sheets *in vitro* that appeared in both cuticle and media; thus, arylphorin appeared to be a bi-directionally secreted integumentary protein. Next, colloidal gold secondary antibodies were used to visualize the location of anti-arylphorin in ultrathin sections of various tissues ([Leung et al., 1989](#)). The resolution afforded by this method made it possible to recognize arylphorin in epicuticle (but not lamellar cuticle) in the Golgi complexes of the fat body, and to show by quantitating gold particles that it was also found in Golgi complexes of epidermis, midgut, pericardial cells, and hemocytes, as well as the meshwork of fibrous cuticle in tracheae. Thus, while the possibility remains that some arylphorin is transported from hemolymph to cuticle, it need not be, for the epidermis itself is capable of synthesizing and secreting this protein. These studies further demonstrated that a given protein can be synthesized by multiple tissues. Whether it is the same gene that functions in all tissues remains to be determined.

The role of arylphorin remains unknown. It is generally assumed to be participating in sclerotization because of its high tyrosine content. Is it degraded in the cuticle so that its constituent amino acids are released, or does it remain an integral part of the cuticle? The latter is favored by the available evidence because calliphorin has been shown to bind strongly to chitin *in vitro* ([Agrawal and Scheller, 1986](#)), and no breakdown products were detected after injection of labeled calliphorin ([Konig et al., 1986](#)).

5.3.2. Structural Proteins

5.3.2.1. Overview and families of cuticular proteins

More than a decade ago, a comprehensive and insightful review of cuticular proteins presented the complete sequence and full citation for all 40 cuticular proteins known at that time, and identified features that remain their hallmarks ([Andersen et al., 1995](#)). Most of the structural cuticular proteins whose sequences were known in 1995 came from the efforts of Svend Andersen and his group, and were based on direct sequencing of purified cuticular proteins. These data provided the starting point for subsequent analyses, for features identified in those early studies led to the assignment of predicted protein sequences as corresponding to putative structural cuticular proteins. The 2005 version of this review provided information about 139 cuticular proteins, and many were based on sequences from cDNAs or short stretches of genomic DNA. Some of the sequences had been confirmed, indeed had their isolation guided, by N-terminal sequences from proteins isolated from cuticle. Now, annotation of several insect genomes is complete. There are EST projects for multiple insect species. Fortunately, proteomic analyses on cuticle preparations have

confirmed that many of the sequences designated as coding for cuticular proteins are indeed coding for authentic rather than putative cuticular proteins. Proteomic studies also identified new families of cuticular proteins. A few analyses of mutant forms or animals with RNAi depleted transcripts have added to the confirmation of specific roles for specific cuticular proteins.

Thus, the number of structural cuticular proteins sequences has increased from fewer than 200 to several thousand, which recently have been organized into 13 fairly well-defined families, with several more as yet not classified (Willis, 2010). Some general comments on cuticular protein nomenclature will be followed by a definition of, and comments about, each family.

While nomenclature of cuticular proteins is not standardized it is improving, and the recognition and definition of distinct families (Willis, 2010) should aid in establishing relationships of the cuticular proteins within and among species. Now that we know that multiple genes may code for proteins with very similar or even identical sequences, it is recommended that the early practice of naming proteins with numbers that correspond to presumed orthologs in other species be abandoned until annotation of whole genomes is complete. Also unwise is calling them LCP or ACP because they were first identified in a larva or adult, because in many cases stage-specificity vanished as further studies were carried out. If one has a whole-genome sequence, the proteins in each family can be named in the order that the genes are located on chromosomes, but so far that practice has only been followed for *Bombyx*. At the very least, the prefix for the family followed by a number provides a useful and informative name. A four-letter abbreviation for the genus and species should precede that name when a paper deals with more than a single species.

A final complication is whether two almost identical proteins are allelic variants, or products of two distinct genes. In some cases an “isoform” has been described. Genomic sequences, however, revealed that stretches coding for proteins of almost identical sequence may be linked on a chromosome (Charles *et al.*, 1997; Dotson *et al.*, 1998; Cornman *et al.*, 2008; Cornman and Willis, 2008, 2009; Futahashi *et al.*, 2008). Only when one has a well-annotated genome is it possible to learn if two similar sequences represent distinct genes or alleles of a single gene. As one goes from ESTs to genomes, expansion and contraction of names will probably occur. A more extensive discussion of cuticle protein nomenclature can be found in a recent review (Willis, 2010).

The 2005 version of this review included a table that listed all known cuticular protein sequences except for those from whole-genome analyses that were just becoming available. Such a table would now exceed the length of this version, for each of the sequenced insect genomes has well over 100 structural cuticular proteins, and the EST

data for dozens more insects also have numerous proteins that are their homologs. **Table 3** gives a numerical summary of the cuticular proteins in some of the annotated genomes. Details on the characteristics of the families are discussed below. One interesting feature on numbers was unearthed by Cornman (2009), who compared numbers of CP genes in seven *Drosophila* species and compared them to numbers in the other two Diptera whose genomes are well annotated. Numbers of the CPR family in the *Drosophila* species ranged from 100 to 104 genes, while *Ae. aegypti* had about 50% more than the 156 in *An. gambiae*. The divergence time between the two lower Diptera is estimated to be 95 my, while members of the genus *Drosophila* are believed to have shared a common ancestor about 40 my ago.

Most of the proteins now described as cuticular were classified by their “discoverers” or computer-driven annotation because their sequences (or a part thereof) were similar to a cuticular protein already in the databases; obviously, such proteins should only be described as putative cuticular proteins until additional evidence is available. Over 90% of the *An. gambiae* cuticular proteins have been confirmed as authentic because peptides corresponding to them were found in extracts of cuticles by tandem mass spectrometry (He *et al.*, 2007). A smaller number of the *Bombyx* proteins have also been confirmed using chitin-binding proteins as starting material (Tang *et al.*, 2010), and many more are known to be authentic based on pre-genomic analyses (Futahashi *et al.*, 2008). Proteomics analyses are being carried out for *Tribolium* (Dittmer, personal communication). The presence of a signal peptide is essential for a cuticular protein, and coupled with compelling sequence similarity is strong evidence that the proteins have been correctly classified as putative cuticle proteins.

One feature of structural cuticular proteins frequently mentioned is that they lack cysteine and methionine residues in the mature protein; Andersen (2005) suggested that the reactivity of cystine and cysteine with orthoquinones could interfere with sclerotization. Thus, the recent finding of the CPAP1 and CPAP3 families with one or three easily recognizable domains each with six cysteines revealed an unappreciated type of cuticular protein. Moreover, the CPCFC family first recognized with BcNCP1 has two or three similar motifs each with two conservatively spaced cysteine residues. While many cuticular proteins are quite short (<200 amino acids), that early generalization too needs revision. There is an enormous CP, dumpy (CG33196), with 22,971 amino acids, that anchors muscle to cuticle in *D. melanogaster*. Among the more conventional cuticular proteins, even the CPR family in *An. gambiae* has 16% of its mature proteins with between 200 and 300 amino acids, while 10% have over 300, with the largest (AgamCPR140) having 837 (Cornman *et al.*, 2008). This large cuticular protein

Table 3 Approximate Number of Genes in Different Cuticular Protein Families in Species with Manual Annotation of Cuticular Proteins in Whole-Genome Data

	CPR	CPF + CPFL	TWDL	CPLCG	CPLCW	CPLCA	CPLCP	CPG	APIDERMIN	CPAP1	CPAP3 (OBSTRUCTOR)	CPCFC	OTHER	TOTAL
Section of chapter	3.2.2	3.2.3	3.2.4	3.2.5	3.2.6	3.2.7	3.2.8	3.2.9	3.2.10	3.2.11	3.2.11	3.2.12	3.2.13	
<i>An. gambiae</i>	156	11	12	27	9	3	4 +23?	0	0	0	7	1	10	240+
<i>B. mori</i>	148	5	4	0	0	0	7	18*	0	0	1	1	33	217
<i>D. melanogaster</i>	101	3	27	3	0	11	5	0	0	2	6	1	?	159
<i>A. mellifera</i>	32	3	2	0	0	0	2	0	3	0	5	0	?	47
<i>N. vitripennis</i>	62	4	2	0	0	0	3	0	3	0	6	0	?	80
<i>T. castaneum</i>	101	8	3	2	0	0	4	0	0	10	7	2	?	137

*Gly-Rich family from *Bombyx* is really a composite of possibly three families (see text). The 6 that have been identified as CPLCPs were deleted from this number, and only the 18 restricted to lepidoptera that have several GGY repeats were included. Absence of additional defining features prevented searches in other groups.

Sources: Togawa *et al.*, 2007; Futahashi *et al.*, 2008; Cornman *et al.*, 2008, Cornman and Willis, 2009, Jasrapuria *et al.*, 2010; *T. castaneum* based on personal communication from N. Dittmer and M. Kanost; Willis (unpublished).

Table 4 Presence of Cuticular Protein Families and Features in Different Groups of Insects

	<i>Diptera</i>		<i>Lepido-</i>	<i>Coleo-</i>	<i>Hymeno-</i>	<i>Hemi-</i>	<i>Ortho-</i>	<i>Dictyo-</i>	<i>Phthira-</i>	<i>Collembola</i>
	<i>Brachycera</i>	<i>Nematocera</i>								
CPR	+	+	+	+	+	+	+	+	+	+
CPF/CPFL	+	+	+	+	+	+	+	+	+	id
TWDL	+	+	+	+	+	+	+	+	+	id
CPLCA	+	+	no	no	no	no	no	no	no	id
CPLCG	+	+	no	+	no	no	no	+	no	id
CPLCW	no	+	no	no	no	no	no	no	no	id
CPLCP	+	+	+	+	+				+	
GPG			+							
apidermin	no	no	no	no	+	no	id	id	no	id
CPAP 1	+	+	+	+	+	+	+	+	+	+
CPAP3	+	+	+	+	+	+	+	+	+	+
CPCFC	+	+	+	+	no	+	+	+	+	+
18 aa motif	+	+	+	+	+	+	+	+	+	id
CP with >3 AAP[AVL]	+	+	+	+	+	+	+	id	+	id

This table is revised from Willis, 2010.

Final syllable ptera was removed from names of most orders.

Data were obtained from Blast searches in addition to analyses in: Togawa et al., 2007; Futahashi et al., 2008; Cornman et al. 2008, Cornman and Willis, 2009; Carmon et al., 2007, Jasrapuria et al., 2010.

id = insufficient data available to record absence; empty boxes indicate that motifs were insufficiently well defined to allow a search.

has an ortholog in *Pediculus humanus* (XP_002432942.1) of the same length. The situation in *Bombyx* is somewhat similar to *Anopheles*; of 148 CPR family members, 22% have 200–300 amino acids and 13% have over 300, with the largest (BmorCPR146) having 1618 (Futahashi et al., 2008).

Thirteen families of cuticular proteins have now been recognized, and the characteristics and history of each will be described below. Two – CPR and CPF – were recognized early. The proteomics study of He et al. (2007) revealed peptides from several dozen more possible cuticular proteins. These were annotated and their temporal expression patterns determined, and they have been separated into five distinct families, described in detail in Cornman and Willis (2009). Most of these proteins have extensive regions of Low sequence Complexity, and have been named CPLC followed by a final initial to designate one of four distinct families. The fifth low complexity family retained the original name TWDL. Most of the families described in that paper have turned up in other insect orders. As mentioned above, three families of cuticular proteins with conserved cysteine residues have been identified: CPAP1 and CPAP3 (Jasrapuria et al., 2010); and CPCFC. There are glycine-rich cuticular proteins that do not belong to any of these families, a small family (apidermin) so far restricted to Hymenoptera (Kuchar-ski et al., 2007), and then a few other cuticular protein sequences that have not yet been assigned to families. It is intriguing that members of most of these families are restricted to arthropods, some in only one or two insect orders, while others are fairly widely distributed (Table 4).

Many cuticular protein sequences are available at the website CuticleDB (<http://bioinformatics2.biol.uoa.gr/cuticleDB/index.jsp>), which allows a variety of different search strategies (Magkrioti et al., 2004).

A convenient way to illustrate diagnostic features is with WebLogos (Schneider and Stephens, 1990; Crooks et al., 2004), and these will be presented in Figures 1–3. A summary of these features that can be used for an initial BLAST search to learn if a database has cuticular protein sequences is available in Supplementary Information File 1 in Willis (2010).

5.3.2.2. CPR family: Proteins with the R&R consensus

By far the most common family of cuticular proteins is that containing the R&R Consensus. The name comes from a 28-aa motif, first recognized by Rebers and Riddiford (1988) in six cuticular proteins. The original R&R Consensus is part of a longer conserved sequence, pfam00379. A valuable website, Pfam (<http://pfam.janelia.org/family/>), has used hidden Markov modeling to define motifs characteristic of particular classes of proteins (Bateman et al., 2002). In accordance with recent nomenclature, this extended consensus region of about 63 amino acids will be referred to hereafter as the R&R Consensus. When a protein sequence is searched against non-redundant protein sequences using blastp at the BLAST server (<http://www.ncbi.nlm.nih.gov/BLAST/>), the first information that is presented is an indication of matches to pfam entries. The pfam sequence that allows annotators to classify a protein as a cuticular protein in the CPR family is pfam00379, a 68-aa sequence that includes the

extended R&R Consensus. It also goes under the name “chitin_bind_4,” for reasons that will become apparent in section 5.5.4. Pfam00379 was obviously based on proteins of both RR-1 and RR-2 classes, for it matches neither particularly well. This makes it particularly useful for a preliminary classification of a putative cuticular protein sequence.

An indication of the importance of the R&R Consensus comes from the Pfam website. It reports 2456 distinct proteins with the Consensus from 67 different species of arthropods (<http://pfam.janelia.org/family?acc=PF00379#tabview=tab6>). This is an underestimate, because close to 100 sequences from Hymenoptera are absent. The CPR family is restricted to arthropods. The one exception, *Xenopus* NP_001090156.1, is due to a contaminating sequence from *Drosophila erecta* (Willis, 2010).

While 98% of the entries have only a single occurrence of the R&R Consensus, the exceptions are interesting. The most notable exception is a protein from the tailfin of the prawn *Penaeus japonicus*. The entire sequence of this protein is made up of 14 consecutive pfam00379 motifs (Ikeya *et al.*, 2001). A protein from the horseshoe crab *Tachyplesus tridentatus* (BAE44187.1) has five Consensus regions (Iijima *et al.*, 2005), and the current annotation of the *Ixodes scapularis* genome reports several instances in a single predicted protein. Manual annotation revealed that most insect genes predicted to code for a protein with more than a single Consensus region actually coded for multiple proteins, easily recognized by standard markers of gene and transcript boundaries. There remains a small number of insect proteins that genuinely appear to have two Consensus regions, and the only one with three has orthologs in several species (Willis, 2010). When only a single Consensus region is present, it can be found near the N- or C-terminus, or within the protein. Three distinct forms of the Consensus have been recognized and named by Andersen (1998, 2000): RR-1, RR-2, and RR-3. RR-1-bearing proteins have been isolated from flexible cuticles, while RR-2 proteins have been associated with hard cuticle. This generalization was based on relatively few cases, and it has also been suggested that RR-2 proteins will contribute to exocuticle while RR-1 will be found predominantly in endocuticle (Andersen, 2000). This issue has not been resolved, even with the extensive expression data that are now available (Togawa *et al.*, 2007). Hopefully, immunolocalization data (see section 5.2.1.3) will prove helpful. The RR-2 Consensus region is far more conserved in length and sequence than the one from RR-1 proteins, as can be seen in the WebLogos in **Figure 1**. The website CuticleDB provides a tool using Hidden Markov Modeling to learn if a protein is RR-1 or RR-2 (Karouzou *et al.*, 2007).

Within the CPR family, numerous proteins can be identified that have orthologs in several species, some with distinct Consensus regions and other features (Cornman and Willis, 2008; Zhang and Pelletier, 2010).

The wealth of information on cuticular protein sequences and the unraveling of how the structure of some contributes to the interaction of chitin and protein (see section 5.5) is only a beginning. Essential properties of cuticle remain to be explained, and important questions raised in the older literatures about various means of achieving cuticle plasticity and the importance of hydration in cuticle stabilization must not be forgotten (Vincent, 2002, and references therein).

An especially interesting member of the CPR family is the resilin gene. The name “resilin” has been given to the rubber-like proteins responsible for the elasticity of jumping fleas and vibrating wings. Analysis of resilin-bearing cuticles in froghoppers (*Aphrophora alni* and *Philaenus spumarius*) concludes that resilin can function in two quite different ways. It is used:

as an energy buffer in rhythmically active, fast mechanical movements, such as those of the wings during flight or the tymbals in cicadas . . . The almost perfect elastic recovery of resilin and its extreme resistance to mechanical fatigue mean that it can return nearly all of the power put into it for the next cycle of movement. The second role . . . is in providing a flexible material that is combined with the stiffer chitinous cuticle in a composite structure.

(Burrows *et al.*, 2008)

The first identification of a complete sequence for resilin was carried out by Ardell and Andersen (2001), who used peptides from locust (*Schistocerca gregaria*) and cockroach (*Periplaneta americana*) resilin to identify a likely homolog in *D. melanogaster*. The peptides came from the R&R Consensus region. The protein they identified was CG15920. Its 18 N-terminal copies of a 15-residue repeat and 13 C-terminal copies of a 13-residue repeat were predicted to contribute to a beta-spiral, a common form for proteins with elastic properties (Ardell and Andersen, 2001). The corresponding gene produces two transcripts; one lacks over two-thirds of the start of the Consensus region. Two groups have studied the physical properties of CG15920 and its repeat regions, and showed that they have the properties one would expect of highly elastic proteins (Elvin *et al.*, 2005; Qin *et al.*, 2009).

The identification of resilin in other species is complicated. Two recent analyses, one brief (Willis, 2010) and the other detailed (Andersen, 2010b), emphasize the difficulties and reach different conclusions about some possible homologs. Andersen emphasizes the need for repeat regions that would underlie the elastic properties, while Willis focused on the R&R Consensus region that showed such conservation between *Schistocerca*, *Periplaneta*, and *D. melanogaster*. Both conclude that an authentic resilin gene should code for the Consensus, although alternative splicing may eliminate it in some of its transcripts. A major complication is that Lyons *et al.* (2007)

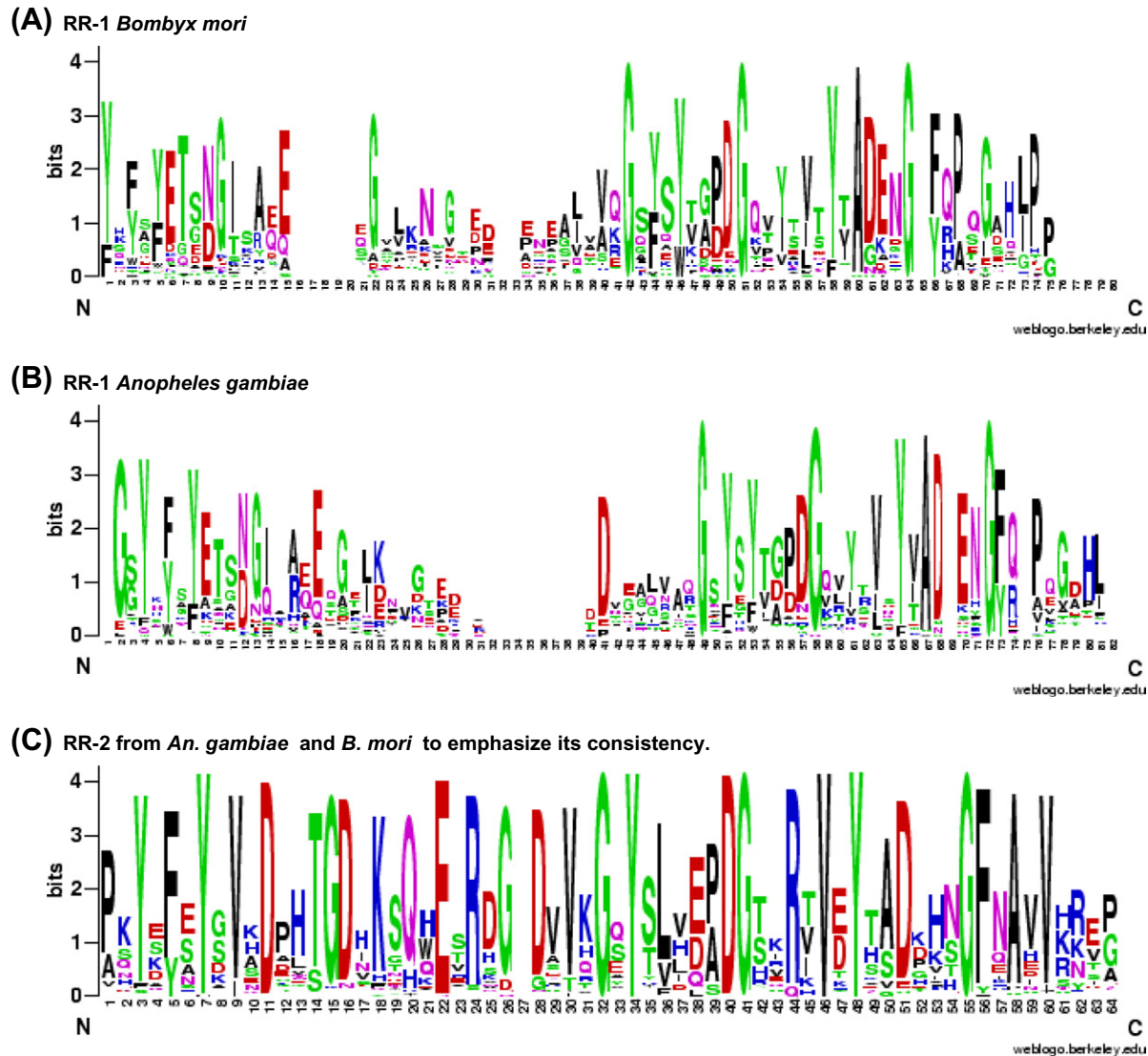


Figure 1 Comparison of the highly conserved RR-2 Consensus with the more variable one from RR-1 proteins. WebLogos were constructed at <<http://www.weblogo.berkeley.edu/logo.cgi>> (Schneider and Stephens, 1990; Crooks *et al.*, 2004). (A) RR-1 Consensus regions from 52 sequences from *B. mori*. (B) RR-1 Consensus regions from 51 sequences from *An. gambiae*. (C) WebLogo constructed from 87 *B. mori* and 101 *An. gambiae* sequences. Panels reprinted from Willis (2010), with permission.

found highly elastic physical properties of an *An. gambiae* protein coded for by an EST BX61916.1, but the corresponding gene, AGAP002367, lacks the Consensus region and is most closely related to a *D. melanogaster* protein (CG7709) that has been characterized as a mucin. The *Anopheles* protein with the closest similarity (74%) to the Consensus region of Dmelresilin is AgamCPR152, but this protein lacks any repeats. Andersen identified AgamCPR140, which has many repeats, as a possible resilin, but the Consensus region is only 34% similar to Dmelresilin.

One consistent property of resilin is its ability to fluoresce due to di- and tri-tyrosine cross-links (Andersen and Weis-Fogh, 1964), so a combination of studies that

establish anatomical location of candidate proteins along with the physical properties should make it possible to sort out what sequences are truly resilin, and identify possible differences when it is serving its different roles.

5.3.2.3. CPF and CPFL families A motif corresponding to a 51-aa repeat first recognized by Andersen *et al.* (1997) has been identified in a modified form in at least 9 orders of insects. However, the common repeat is somewhat shorter, at 42–44 amino acids, so the original name for the CPF family has been retained, with the F now referring to forty rather than to fifty. A detailed discussion of this family can be found in Togawa *et al.* (2007), where it is pointed out that in addition to the conserved motif of

about 42 amino acids, the proteins are also similar in the amino acids near their carboxyl-termini.

The C-terminal region characteristic of the CPF family has also been found in other cuticular proteins that lack the defining consensus. [Togawa et al. \(2007\)](#) named these CPFL, for CPF-like. All four CPF proteins and six of the seven CPFL proteins in *An. gambiae* have been verified as authentic cuticular proteins, based on shared peptides identified in a tandem mass spectrometry analysis of cast cuticles ([He et al., 2007](#)).

CPF and/or CPFL proteins have been identified throughout the hexapoda, including collembola and diplura ([Table 4](#)), but not yet in Crustacea or Chelicerata.

5.3.2.4. TWDL family One of the families of low-complexity proteins previously had been identified in *D. melanogaster* and named TWDL after the tubby phenotype in one of its mutants that reminded the authors of Tweedle Dee ([Guan et al., 2006](#)). There are 27 members of this family in *D. melanogaster*, 12 in *An. gambiae*, and fewer in other insects ([Table 3](#)). Their relationships are discussed in detail in [Cornman and Willis \(2009\)](#). Four conserved regions were defined by Guan and colleagues, and they remain diagnostic of the proteins across the Insecta ([Figure 2](#)). One member of the TWDL family (BmorCPT1) has been identified in a proteomics analysis of larval chitin-binding proteins, and a recombinant version binds chitin in an *in vitro* assay ([Tang et al., 2010](#)).

5.3.2.5. CPLCG family The largest of the new CP families is CPLCG, recognized by a conserved G-x(2)-H-(x2)-P ([Cornman and Willis, 2009](#)). The x residues are restricted to just a few amino acids, and a sequence logo, encompassing a longer stretch of conserved amino acids, is shown in [Figure 2](#). Two members of this family had been reported in *D. melanogaster* ([Qiu and Hardin, 1995](#)), 3 are now recognized, along with 27 in *An. gambiae*. The *D. melanogaster* sequences had been named Dacp-1 and -2, but members of the family are not restricted to adults, and the CPLCG name is more accurate. Furthermore, the family is not restricted to the Diptera, but was identified in other orders of insects and the crustacean *Daphnia* ([Tables 3, 4](#)).

5.3.2.6. CPLCW family Another small family, CPLCW, appears to be restricted to mosquitoes ([Table 4](#)). The WebLogo ([Figure 2](#)) shows the invariant W after which it was named, but several other amino acids in a 29-aa region are also almost invariant. Its nine genes are clustered in *An. gambiae* interspersed among some members of the CPLCG family, but the protein sequences of CPLCG and CPLCW families are distinct, having an average similarity of only 20% ([Cornman and Willis, 2009](#)).

5.3.2.7. CPLCA family The CPLCA family has from 13% to 26% alanine residues, but this number is not higher than in some members of other families; rather, the family is best identified by the presence of the retinin domain (pfam04527), although the *D. melanogaster* protein retinin is an outlier in the phylogeny of the group ([Cornman and Willis, 2009](#)). A WebLogo more typical of the group has been created ([Figure 3](#)). While the first published account of this family ([Cornman and Willis, 2009](#)) stated that it is restricted to Diptera, there is clearly an EST in *Daphnia* that has a sequence corresponding to the WebLogo (FE341353.1).

5.3.2.8. CPLCP family This is the most problematic of the cuticular protein families. Peptides corresponding to four genes turned up in the proteomics analysis of proteins from larval head capsules and cast pupal cuticles of *An. gambiae* ([He et al. 2007](#)). An additional 23 genes coding for related proteins are also present in *An. gambiae*, but none have yet been confirmed by proteomics, although their expression profiles resemble those of authentic low complexity cuticular proteins ([Cornman and Willis, 2009](#)). Members of the family have a high density of PV and PY pairs, but additional features described by Cornman and Willis appear to be restricted to mosquitoes where both *Aedes* and *Culex* have been found to have larger families ([Cornman and Willis, 2009; Willis, 2010](#)).

5.3.2.9. CPG, the glycine-rich protein family A group of 28 genes enriched in GGGG or GGxGG repeats was described in *B. mori* ([Futahashi et al., 2008](#)), but the group appears heterogeneous because six proteins with only zero to three repeats appear to belong to the CPLCP family; these were identified after that paper was published ([Willis, 2010](#)). Another subset of 18 appears to be lepidopteran-specific, and these can appropriately be designated as CPGs (see [Willis, 2010](#), Supplementary Material 2, for details).

5.3.2.10. Apidermin family Three apidermins, small (6.1–9.2 kDa), highly hydrophobic, and with at least 30% alanine content were described in *Apis mellifera* ([Kucharski et al., 2007](#)), and now three have been found in *Nasonia*, but as presently annotated, they are much larger (23–39 kDa). Members of the family do not have an obvious structure; rather, they were recognized by chromosomal linkage, and their role in the cuticle was confirmed with RT-PCR on cuticle-forming tissue. At present they have only been identified in Hymenoptera ([Table 4](#)). Their designation as a family thus is based on the initial publication, not the normal criterion of shared sequence similarity, and so it is not possible to evaluate the significance of numerous EST sequences from the beetle, *Diaprepes abbreviates* that are somewhat similar to *A. mellifera* apidermin 1 (e.g., CN474619.1).

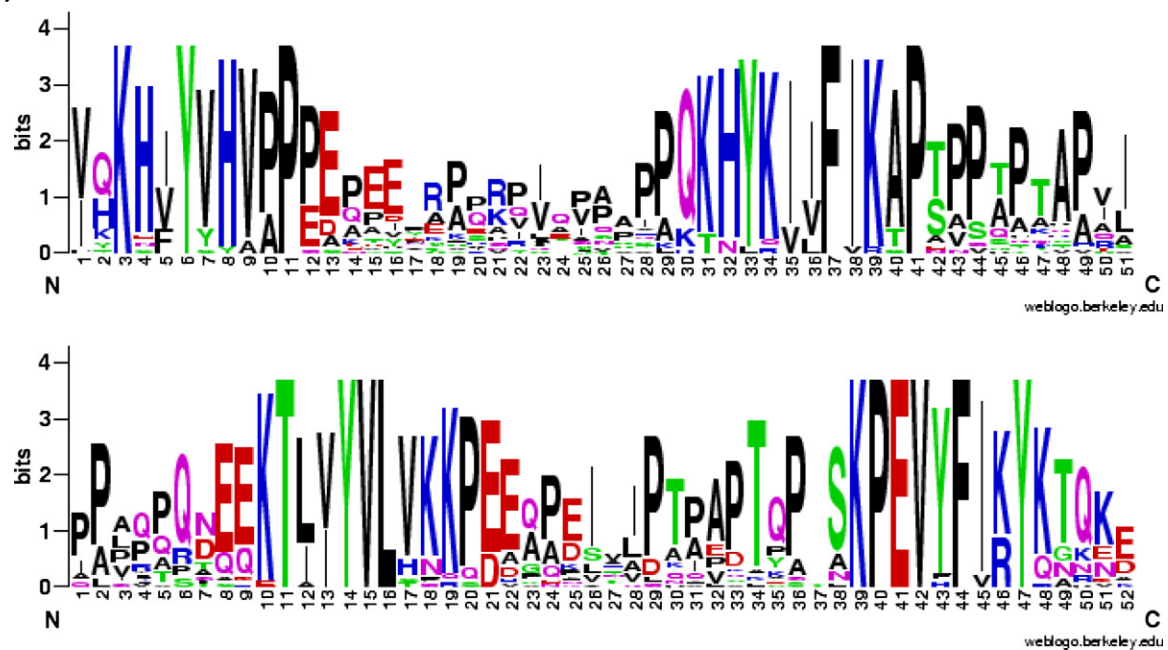
(A) TWDL FAMILY (sequence is continuous for 103 residues)**(B) CPLCG FAMILY****(C) CPLCW FAMILY**

Figure 2 WebLogos (see **Figure 1**) for three cuticular protein families. (A) TWDL family. Twenty-four sequences from eight species in six orders of insects were used. The continuous sequence was split to facilitate recognition of the four conserved regions. (B) CPLCG family. Note the highly conserved GHPG at residues 5, 8, 11, 14. Eighty-six sequences from dipterans were used. (C) CPLCW family. The 26 CPLCW sequences of this mosquito-restricted family were used. Unlike other WebLogos, the alignment for this one required gaps of five or eight amino acids between positions 16 and 25 to accommodate the longer *Ae. aegypti* sequences. Panels (A) and (B) reprinted from Willis (2010), with permission; some modified from Cornman and Willis (2009). More details are in those references.

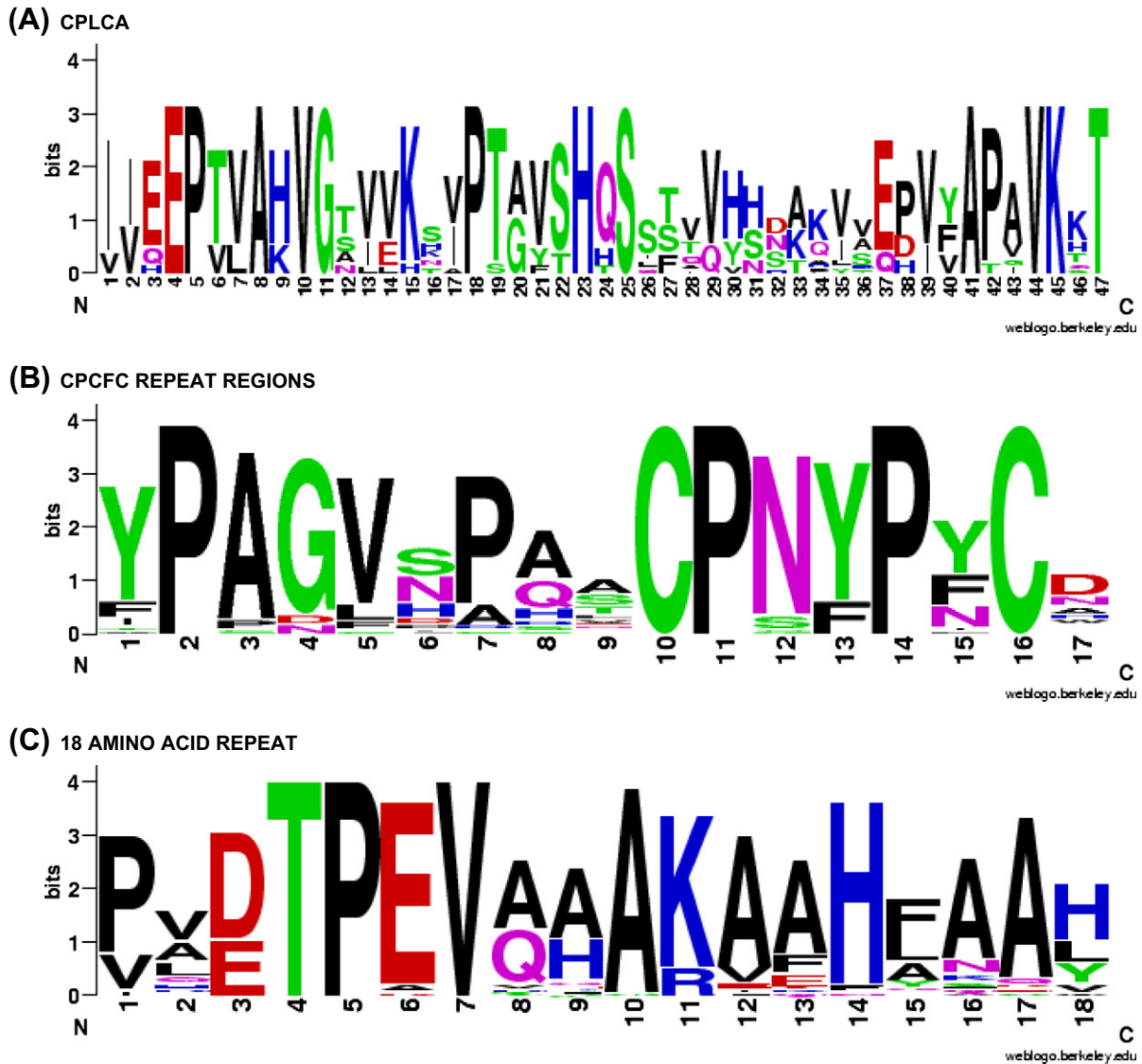


Figure 3 WebLogos (see Figure 1) for two cuticular protein families and one motif. (A) CPLCA family. The WebLogo is based on three sequences from each of four species, *An. gambiae*, *Ae. aegypti*, *C. pipiens*, and *D. melanogaster*, that had the closest match to AgamCPLCA1. This region corresponds to the retinin domain. (B) WebLogo for CPCFC family. Data from the single occurrence of this protein in individual genera of eight insect orders, plus the two occurrences in *T. castaneum* and *Heliconius melpomene*. All three (two in Coleoptera and Lepidoptera) repeat regions from each protein were used. (C) The 18 amino acid repeat from 40 sequences from 26 proteins from 5 insect orders and 2 crustaceans. Panels (A) and (C) reprinted from Willis (2010), with permission.

5.3.2.11. CPAP1 and CPAP3 families A recent publication has identified two more families of cuticular proteins: CPAP1 and CPAP3 (Jasrapuria *et al.*, 2010). They are unusual in that they have multiple cysteine residues, an amino acid rarely found in cuticular proteins from the other families. The families were named because they resemble some peritrophins, hence are peritrophin-like, but the spacing of the cysteines is distinct. The names come from Cuticular Proteins Analogous to Peritrophins. Comparable groups of six cysteines have been demonstrated to form a chitin-binding domain called the “peritrophin A domain,” or “type 2 chitin-binding domain”

(ChtBD2), with the six cysteines assumed to form three disulfide bridges. An exhaustive search for proteins with this domain was carried out in *Tribolium* accompanied by RT-PCR analysis of their temporal and spatial distributions. It yielded, in addition to members of the two new families of cuticular proteins, several genuine peritrophins, as well as chitinases and chitin deacetylases. It is assumed that the ChtBD2 domains in all these proteins bind chitin, but this has only been demonstrated experimentally for a chitinase (Arakane *et al.*, 2003) and a CPAP3 protein from another species (Nisole *et al.*, 2010; see also section 5.5.4). So far, the CPAP1 family,

with only one ChtBD2 domain, has only been identified in beetles, but the CPAP3 family, with three ChtBD2 domains, is more widespread (Table 4). Indeed, its motifs are found outside the arthropods (Jasrapuria *et al.*, 2010). The founding members of the CPAP3 family were a group of proteins, named obstructers, in *D. melanogaster* (Barry *et al.*, 1999; Behr and Hoch, 2005), among them a protein, Gasp, found in tracheae.

5.3.2.12. CPCFC family There is a third cuticular protein family with well-conserved cysteine residues. The founding member is BcNCP1, first identified in *Blaberus craniifer* (Jensen *et al.*, 1997). It has three repeat regions, each with a pair of cysteines separated by five other amino acids; the first and fourth amino acids in each repeat are proline. In a recent publication (Willis, 2010) family status was not recognized, because at that time there were only single occurrences of BcNCP1 orthologs in any species, and a family must have paralogs within a species. Now that criterion has been met in *Heliconius melpomene* and *Tribolium*, and likely in another beetle, *Diaprepes abbreviatus*, each with two related genes. This chapter recognizes family status for these paralogs. Several other species of beetles and moths have good orthologs, and in every case the middle cys-bearing region is missing. We are naming this family CPCFC in recognition of the two or three pairs of cysteines that are separated by five amino acids. A WebLogo is shown in Figure 3.

5.3.2.13. Cuticular proteins not assigned to families There remain some cuticular proteins that have not reached the criterion for belonging to families. Among them are three proteins identified with proteomics in *An. gambiae* (described in Cornman and Willis, 2009), and a group called CPH (cuticular protein hypothetical) in *Bombyx mori* (Futahashi *et al.*, 2008). Some of the CPH can now be assigned to families; others remain unclassified.

5.3.3. Motifs Found in Cuticular Proteins that do not Define Families

The review by Andersen *et al.* (1995) was the first to assemble a variety of motifs found in cuticular proteins. It is now possible to distinguish among two classes of motifs. The first defines a family such as the CPR and CPF families, while the second includes motifs that occur commonly in cuticular proteins but are found in more than one family; many of these are very short. It is this second class that will be discussed in this section.

The most common short motif described by Andersen *et al.* (1995) was A-A-P-(A/V). Once cuticular proteins of *An. gambiae* had been annotated, it was necessary to expand that motif to A-A-P-(A/V/L). While one or two instances of that motif are found in many proteins,

especially chorion proteins, the occurrence of three or more in a single protein appears to be restricted to cuticular proteins (Willis, 2010). The function of this motif was discussed by Andersen *et al.* (1995) who concluded:

A relevant feature of the Ala-Ala-Pro-Ala motif appears to be a strong tendency to form turns; several conformations can be present in equilibrium, indicating low energy barriers between the conformations. When the sequence occurs regularly in a protein, as it does in many of the CPs as well as in other structural proteins, it can be suggested that the result will be proteins folded in a more or less regular helix, which is easily and reversibly deformed by external forces, thereby resembling elastin.

Andersen *et al.* (1995) recognized several sequences with stretches of glycine, leucine, and tyrosine, beginning G-Y-G-L- or G-L-L-G. Other cuticular proteins are also high in glycine, but with less regular motifs; these are designated by the number of consecutive Gs. Proteins enriched in glycine residues are found in a variety of structures, such as plant cell walls, cockroach ootheca, and silk (see Bouhin *et al.*, 1992a, for discussion). Subsequent to their 1995 review, Andersen and his colleagues recognized two additional motifs.

Three copies of an 18-residue motif were found in a *B. mori* protein (PCP, now named BmorCPH31) by Nakato *et al.* (1992). Subsequently, Andersen (2000) recognized the repeat in a small number of cuticular proteins from four orders of insects and two crustaceans. A sequence logo based on its occurrence in 27 proteins from 5 orders of insects and 2 crustaceans is shown in Figure 3. These proteins include some with the R&R Consensus, especially those assigned as RR-3, as well as others, like BmorCPH31 and four other *B. mori* cuticular proteins, that do not have this Consensus (Futahashi *et al.*, 2008).

A recent analysis (Cornman, 2010) analyzed the short motifs GYR and YLP in several *Drosophila* species in relation to cuticular proteins and other classes of proteins.

5.3.4. Glycosylation of Cuticular Proteins

Glycosylation of cuticular proteins was first reported by Trim in 1941, and then in limited subsequent reports (see Cox and Willis, 1987b, for review). In recent years, post-translational modifications of cuticular proteins have been determined by staining gels with periodic acid Schiff (PAS), by using labeled lectins to probe blots of electrophoretically separated proteins, or by discovering discrepancies in masses of peptide fragments experimentally determined by MALDI-MS analysis and calculated from Edman sequencing.

Most of the major cuticular proteins seen on gels stained with Coomassie Blue are not recognized by PAS or lectins, while some minor ones are glycosylated. This was true

for *H. cecropia*, where PAS staining revealed glycosylated proteins in extracts of flexible cuticles of *H. cecropia* and a screen with eight lectins revealed the presence of mannose and N-acetylgalactosamine, with more limited binding to N-acetylglucosamine, galactose, and fucose, in a few of the proteins from all stages (Cox and Willis, 1987b). A comparable study in *Tenebrio* revealed one major band of water-soluble larval and pupal cuticular proteins that had N-acetylglucosamine, and a few other bands were weakly visualized with lectins; none of the proteins from adult cuticle reacted with the lectins (Lemoine *et al.*, 1990). In another Coleopteran, *Anthonomus grandis*, glycosylation was found in cuticular proteins extracted from all three metamorphic stages (Stiles, 1991). In yet another coleopteran, *T. castaneum*, the BioRad Immun-Blot kit for glycoprotein detection revealed multiple bands on a blotted 1D SDS gel; none of the abundant bands below 30 kDa were stained (Missios *et al.*, 2000). In *Calpodes*, all the BD peptides but very few of the C class proteins (see section 5.2.2) extracted from the cuticle were associated with α -D-glucose and α -D-mannose, just like most of the hemolymph proteins. Some of each class appeared to be modified with N-acetylglucosamine. T66, a protein synthesized in spherulocytes, transported to epidermis, and then secreted into the cuticle, however, was not glycosylated. In none of these species is the amino acid sequence of a glycosylated protein known.

Sequence-related information about glycosylation is available for cuticular proteins isolated from locusts and *Manduca* where the direct analysis of residues had been used. In *Locusta migratoria*, one to three threonine residues were modified in the protein LM-ACP-abd4. In each case, the modification was with a moiety with a mass of 203, identified as N-acetylglucosamine (Talbo *et al.*, 1991). Each of the three threonine residues occurred in association with proline (FPTPPP, LATLPPTPE). All eight of the cuticular proteins that have been sequenced from *S. gregaria* nymphs had evidence for glycosylation with a moiety with a mass of 203, all at a threonine residue found in a cluster of prolines (Andersen, 1998). Three proteins recently isolated from *Manduca* were similarly shown to be glycosylated on threonines also in proline-rich regions. Surprisingly, in these cases masses of the adducts were varied (184, 188, and 189) and their nature was not determined (Suderman *et al.*, 2003). In all of these cases, the available evidence indicates that the threonine residues had been O-glycosylated. The significance of such glycosylation awaits further elucidation.

5.4. Genomic Information

5.4.1. Introduction

The first four cuticular proteins whose complete sequences were determined were also the first to have their genes described (Snyder *et al.*, 1982). The wealth of experimental

detail and thoughtful discussion in that paper make it a classic in the cuticular protein literature. These four genes for *D. melanogaster* cuticular proteins LCP-1, -2, -3, and -4 were found to occupy 7.9-kb of DNA, along with what appeared to be a pseudogene. Each gene had a single intron, and that intron interrupted the protein-coding region between the third and fourth amino acids. *LCP-1* and -2 were in the opposite orientation of *LCP-3* and -4. The nucleic acid sequences in the protein coding regions for LCP-1 and -2 were 91% identical, and for LCP-3 and -4 were 85% identical, with similarity between the two groups about 60%. For the non-coding regions of the mRNAs, the 5' upstream regions had more sequence similarity than the 3' downstream. A consensus poly(A) addition site, AATAAA, was found for two of the genes, 110 bp from the stop codon, while similar but not identical sequences (AATACA, AGTAAA) were found for the other two. The four genes were all expressed in the third instar, and several short, shared elements were found in their 5' regions upstream from the transcription start site. Snyder *et al.* (1982) also speculated on the origin of the cluster through gene duplication and inversion. These features of those four genes (coding for RR-1 proteins) have turned out to be the common elements of most of the cuticular protein genes that are known – hence linkage, shared and divergent orientation, an intron that interrupts the signal peptide, presence of a pseudogene in the cluster, atypical poly(A) addition sites, and divergence of 3'-untranslated regions have been found for cuticular protein genes in Diptera, Lepidoptera, and Coleoptera.

5.4.2. Chromosomal Linkage of Cuticular Protein Genes

In addition to the four *D. melanogaster* genes discussed in the previous section, several more instances of linked cuticular proteins genes were described prior to sequencing entire genomes. In some cases the evidence for these genes was restricted to cross-hybridization of the genomic fragment, and complete sequences were not known for all the members.

A detailed analysis of the cluster of genes at 65A allowed Charles *et al.* (1997, 1998) to describe important features that most likely contributed to the multiplication and diversification of cuticular protein genes. Twelve genes were identified in a stretch of 22 kb, with the direction of transcription, or more accurately the strand used, being: > < < < < < > > > >. The third gene in the cluster appeared to be a pseudogene. Several important features were found: the number of *Lcp-b* genes within the cluster was variable among different strains of *D. melanogaster*; and some genes lacked introns, had tracks of As at the 3' end and short flanking direct repeats. These features are consistent with their having arisen by retrotransposition.

Now that there are complete sequence data for the entire 65A region the situation has been shown to be even more complex, and comparison with six other *Drosophila* species has provided new insights (Cornman, 2009). Eighteen CPR genes are present in the 65A region of *D. melanogaster*, seven of these are present in most or all seven species as one-to-one orthologs, with their chromosomal order conserved. Others have orthologs only within one of the two species groups analyzed. Others are found scattered among the array, with paralogs only within one or two species, and this analysis, of course, could not deal with the variation in copy number within a species. Cornman confirmed the findings of Charles *et al.* (1997) that some of the genes lacked introns, and assessed the possibility the latter raised that retroposition played a role in the formation of this array, but concluded that “retrogenes do not appear to contribute substantially to the distinctive pattern of evolution within these arrays.”

The consequences of gene duplication in terms of gene expression are an important issue. It could be that duplicated genes were preserved to boost the amount of product made in the short period that the single-layer epidermis is secreting cuticle. Alternatively, duplication may allow for precise regulation of expression of genes both spatially and temporally. Subtle differences in protein sequence may be advantageous for particular structures. A detailed analysis of mRNA levels with Northern blot analysis demonstrated that some members of the 65A cuticular protein cluster have quite different patterns of expression. *Acp* was expressed only in adults. Expression was not detected for *Lcp-a*; all other *Lcp* genes were expressed in all larval stages, and all but *Lcp-b* and *-f* also contributed to pupal cuticle (Charles *et al.*, 1998).

One of the major findings to come out of whole-genome sequencing was two different forms of chromosomal linkage of genes for cuticular proteins. Data from *An. gambiae*, *D. melanogaster*, and *B. mori* revealed that many CP genes are found adjacent to one another. Such genes have been described as being in tandem arrays, and both RR-1 and RR-2 genes are clustered in this manner, always in separate arrays (Cornman *et al.*, 2008).

In mosquitoes, there are numerous instances of sequence clusters – groups of genes that are very similar in sequence. Members are generally, but not always, found adjacent within a tandem array. Eight clusters (with 4–16 members) of RR-2 genes were identified in *An. gambiae* (Cornman *et al.*, 2008), and comparable clusters were also present in *Ae. aegypti* and *Culex pipiens*. There are no clusters in *D. melanogaster* coding for more than three proteins with almost identical sequences, but three small sequence clusters with a total of 15 RR-2 genes are present in the *B. mori* genome (Futahashi *et al.*, 2008). The suggestion was made that the *Anopheles* sequence clusters serve to facilitate accumulation of mRNA in a brief period of time, while the *Bombyx* workers speculated that different

members of the clusters might be used to build specific structures (Futahashi and Fujiwara, 2008; Futahashi *et al.*, 2008). A detailed analysis of sequence clusters in *An. gambiae* can be found in Cornman and Willis (2008).

It is not only CPR genes that are found in tandem arrays. There is a large tandem array on chromosome 3R in *An. gambiae* that has all 27 CPLCG genes and all 9 CPLCW genes. Members of the two families are interspersed, and in the array are an additional 10 unrelated genes. Twelve of the CPLCG genes belong to a sequence cluster, and, despite their considerable similarity (86% identity at the nucleotide level), they are dispersed throughout the tandem array and interspersed with the CPLCW genes that form another sequence cluster with at least 92% sequence identity at the protein level (Cornman and Willis, 2009).

5.4.3. Intron Structure of Cuticular Protein Genes

Genomic sequence data are now available for hundreds of cuticular proteins. Intron position has only been analyzed in detail for *An. gambiae* (Cornman and Willis, 2008) and *B. mori* (Futahashi *et al.*, 2008). An early prediction that genes for cuticular proteins would have no more than 2 introns was incorrect, for several have been identified with 5 or more, and one, *BmorCPR146*, has 13. Nonetheless, the number is usually low, averaging 2.3 for *An. gambiae* CPRs and 2.4 for that family in *B. mori*. Cuticular proteins in other *An. gambiae* cuticular protein families generally have only two exons. The most common position for the first intron is interrupting the signal peptide. Whether this conserved position represents something important awaits further exploration, but there are several ways the intron might be important (Charles, 2010). One possibility is that it contains information needed for transcription. Direct evidence that this is the case for one gene comes from an analysis of the *DmelACP65A* gene. Expression is suppressed in the absence of the intron that occurs after coding for the first four amino acids of the signal peptide, and is restored if the intron is added upstream of the transcription start site (Bruey-Sedano *et al.*, 2005).

Another common position is at or near the start of the aromatic triad. Some genes that lack the intron interrupting the signal peptide have one in this region. An early analysis of intron position led Charles *et al.* (1997) to postulate that the primitive condition for introns in insect cuticular proteins would be two; over time, some genes lost one, some the other, and some lost both or arrived in the genome by retrotransposition.

There is also a *D. melanogaster* cuticular protein whose gene is located within the region corresponding to the first intron of *Gart* (now named *ade3*, CG31628), a gene that encodes proteins involved in the purine pathway. The gene for this RR-1 protein (*Pcp*, CG3440) is read off the opposite strand and has its own intron, conventionally

placed interrupting the signal peptide (Henikoff *et al.*, 1986). A comparably placed gene with 70% amino acid sequence identity is found in *D. pseudoobscura* (Henikoff and Eghtedarzadeh, 1987).

5.4.4. Regulatory Elements

One of the attractions of studying cuticular proteins is that they are secreted at precise times in the molt cycle, and are thus candidates for genes under hormonal control (Ridiford, 1994; Togawa *et al.*, 2008). It would be expected, therefore, that some might have hormone response elements. Imperfect matches to ecdysteroid response elements (EcRE) from *D. melanogaster* were found on two of its cuticular protein genes – *EDG78* and *EDG84* (Apple and Fristrom, 1991). These genes are activated in imaginal discs exposed to a pulse of ecdysteroids, but if exposed to continuous hormone, no message appears. The two cuticular protein genes that have been studied in *H. cecropia* have regions close to their transcription start sites that resemble EcREs (Binger and Willis, 1994; Lampe and Willis, 1994), and upstream from *MSCP14.6* are also two regions that match (Rebers *et al.*, 1997).

It is now apparent that the regulatory regions controlling response to ecdysteroids encompass more than just an EcRE. Indeed, the EcRE itself can also recognize β -FTZ-F1, a protein induced in response to ecdysteroid stimulation that has been shown to be a major regulator of cuticular protein synthesis (reviewed in Charles, 2010). Charles (2010) discusses the evidence for this and other transcription factors (BR, DHR38, OCT, SVP) that bind upstream of cuticular protein genes.

Both *Bombyx PCP* (now *BmorCPH31*) and *H. cecropia HCCP66* have response elements for members of the POU family of receptors (Nakato *et al.*, 1992; Lampe and Willis, 1994). POU proteins are transcription factors used for tissue-specific regulation in mammals (Scholer, 1991). Gel mobility shift experiments established that there was a protein in epidermal cells that could bind to this element (Lampe and Willis, 1994).

Numerous additional genes, some of them coding for transcription factors, have been implicated in cuticle formation in *D. melanogaster* (Moussian, 2010).

Now that genomic sequence information is available, identification of regulatory elements and verification of their action is underway.

5.5. Interactions of Cuticular Proteins with Components of Cuticle

One of the most challenging aspects in the study of the cuticle is the elucidation of interactions among cuticular proteins and cuticle's non-proteinaceous components. The most abundant is the CPR family, which is characterized by the presence of the R&R Consensus (see section

5.3.2.2). The abundance of sequences bearing the R&R Consensus in cuticles formed by every species of arthropod examined led several workers to suggest that the role of the R&R Consensus might be to bind to chitin, and this has now been confirmed with recombinant proteins (see section 5.5.4).

Of particular interest among the other families of cuticular proteins that lack the R&R Consensus is the CPF gene family, now recognized by a 44-aa sequence motif (Togawa *et al.*, 2007) (see section 5.3.2.3). As discussed below, CPFs may interact with other components of cuticle, such as sex pheromones (Hall, 1994; Greenspan and Ferveur, 2000) or cuticular lipids, acting as possible repositories (Papandreou *et al.*, 2010).

Various approaches have been followed to gather information about the interactions of cuticular proteins with other components of cuticle. The first was to analyze cuticular protein sequences with appropriate software to predict their secondary structure. The second approach was to use spectroscopic techniques on cuticular components to gain information about the conformation of their protein constituents *in situ*, and compare experimental information with predictions. Third, the tertiary structures of cuticular proteins have been modeled, and the fourth route was a direct experimental approach to test whether proteins exhibiting the extended Consensus could bind to chitin. Such analyses are restricted primarily to the CPR and CPF families, the only ones to be discussed here.

5.5.1. Secondary Structure Predictions

Prediction of secondary structure was carried out on the extended R&R Consensus region of the cuticular proteins now classified in the CPR family (see Iconomidou *et al.*, 1999, for details of proteins analyzed, programs used, and pictorial representation of results.)

The results indicated that the extended R&R domain of cuticular proteins has a considerable proportion of β -pleated sheet structure and a total absence of α -helix. Other features revealed include the presence of glycines and histidines at the predicted β -turn/loop regions. Glycines are considered good turn/loop formers (Chou and Fasman, 1974a, 1974b), while histidines, which in this case are “exposed,” are certainly involved in cuticular sclerotization and in the variations of the water-binding capacity of cuticle and the interactions of its constituent proteins (Andersen, 2005). Also, the β -sheets exhibit an amphipathic character – i.e., one face is polar, the other non-polar. Alternating residues along a strand point in the opposite direction on the two faces of a β -sheet. With these proteins, it is the aromatic or hydrophobic residues that alternate with other, sometimes hydrophilic, residues. The aromatic rings are thus positioned to stack against faces of the saccharide rings of chitin. This type of interaction

is fairly common in protein–saccharide complexes (Vyas, 1991; Hamodrakas *et al.*, 1997; Tews *et al.*, 1997).

The suggestion that cuticular proteins adopt a β -sheet conformation is not new. Fraenkel and Rudall (1947) provided evidence from X-ray diffraction that the protein associated with chitin on intact cuticle has a β -type of structure.

5.5.2. Experimental Studies of Cuticular Protein Secondary Structure

The next step in probing the structure of cuticular proteins involved direct measurements on intact cuticles, on proteins extracted from them with a strong denaturing buffer with 8M guanidine hydrochloride, and on the extracted cuticle. The cuticles came from the flexible abdominal cuticle of larvae of *H. cecropia*, and extracts have HCCP12, a RR-1 protein, as a major constituent (Cox and Willis, 1985; Binger and Willis, 1994). The same prediction programs described above were used on the sequence for HCCP12, and it indicated that the entire protein had a considerable proportion of β -pleated sheet and total absence of α -helix. Fourier-transform Raman spectroscopy (FT-Raman), attenuated total reflectance infrared spectroscopy (ATR FT-IR), and circular dichroism spectroscopy (CD) were carried out on these preparations (Iconomidou *et al.*, 2001). These techniques eliminated problems that had been found previously with more conventional laser-Raman spectra due to the high fluorescent background associated with cuticle.

The FT-Raman spectra of both the intact and extracted cuticle were dominated by the contribution of bands due to chitin. Certain features of the Raman spectrum of the intact cuticle signified the presence of proteins. The protein contribution to the spectrum of intact cuticle was revealed by subtracting the spectrum of the extracted cuticle, after scaling the discrete chitin bands of both preparations. The comparison of this difference spectrum to that from the isolated proteins revealed striking similarities, suggesting that the former gave a reliable physical picture of the cuticle protein vibrations in the native state. While Iconomidou *et al.* (2001) presented a detailed analysis of the spectra and the basis for each assignment, only a few features will be reviewed here. Several of the spectral bands could be attributed to side-chain vibrations of amino acids with aromatic rings, tyrosine, phenylalanine, and tryptophan; others were typical of β -sheet structure and others could be assigned to β -turns or coil. The absence of bands at characteristic positions indicates that α -helical structures are not favored.

Results from ATR-FT-IR spectra from the extracted proteins were in good agreement with their FT-Raman spectra. These spectra had been obtained on lyophilized samples. The CD spectrum, on the other hand, was obtained with proteins solubilized in water. Detailed

analysis of the CD spectrum indicated a high percentage (54%) of β -sheet conformation with a small contribution of α -helix (~13%). The contributions of β -turns/loops and random coil were estimated as 24% and 9%, respectively (Iconomidou *et al.*, 2001). These results demonstrated that the main structural element of cuticle proteins is the antiparallel β -pleated sheet. Comparable results were obtained from lyophilized proteins and intact cuticles, and from proteins in solutions, thus negating the concern that lyophilization might increase the β -sheet content of proteins as discussed by Griebenow *et al.* (1999). These direct measurements confirm the results from secondary structure prediction discussed above in section 5.5.1.

These findings are in accord with the prediction of Atkins (1985) that the antiparallel β -pleated sheet part of cuticular proteins would bind to α -chitin. His proposal was based mainly on a 2D lattice matching between the surface of α -chitin and the antiparallel β -pleated sheet structure of cuticular proteins.

There seem to have been several independent solutions in nature whereby chitin binds to protein; in all, surface aromatic residues appear to be significant (Shen and Jacobs-Lorena, 1999). In several cases, β -sheets have been implicated. The chitin-binding motifs of two lectins studied at atomic resolution contain a two-stranded β -sheet (Suetake *et al.*, 2000). In bacterial chitinases, an antiparallel β -sheet barrel has also been postulated to play an important role in “holding” the chitin chain in place to facilitate catalysis. Four conserved tryptophans on the surface of the β -sheet are assumed to interact firmly with chitin, “guiding” the long chitin chains towards the catalytic “groove” (Perrakis *et al.*, 1997; Uchiyama *et al.*, 2001).

5.5.3. Modeling of Cuticular Proteins

5.5.3.1. CPR protein models Secondary structure prediction and experimental data summarized above (see sections 5.5.1 and 5.5.2) indicated that β -pleated sheet is most probably the underlying molecular conformation of the members of the CPR family, and that this conformation is most probably involved in β -sheet/chitin-chain interactions of the cuticular proteins with the chitin filaments (Iconomidou *et al.*, 1999, 2001). Can this information be translated into a three-dimensional model?

Unexpectedly, distant sequence similarities of the extended R&R Consensus from several CPR proteins were found with a lipocalin, bovine plasma retinol-binding protein (RBP) (Hamodrakas *et al.*, 2002). Lipocalins are members of a family of extracellular proteins, typically small (160–200 residues), with low sequence similarity among family members (frequently <20%). They exhibit several common molecular recognition properties, and, while they were classified mainly as transport proteins, it is now clear that they have various functions (Flower,

1996). The lipocalin fold is a highly symmetrical all- β structure dominated by a single eight-stranded antiparallel up-and-down β -sheet barrel (Flower *et al.*, 2000). Fairly recently, it was found that lipocalins are characterized by two hydrophobic “clusters” of residues, the “inner” and the “outer” clusters (Adam *et al.*, 2008).

The first attempt utilized HCCP12, an RR-1 protein leading to a construction of a structural model that corresponds to the “extended R&R Consensus” (Hamodrakas *et al.*, 2002). The original model (Figure 4A) comprises the C-terminal 66 residues (out of 89 in total) of HCCP12, and has many advantages since it corresponds to the full sequence of the “extended R&R Consensus” (see section 5.3.2.2). This work was extended to RR-2 proteins, leading to comparable results, as shown for AGCP2b (AgamCPR97) in Figure 4D (Iconomidou *et al.*, 2005).

Low-resolution docking experiments of an extended N-acetylglucosamine tetramer to the model of HCCP12, utilizing the docking program GRAMM (Vakser, 1996), revealed that the proposed model for cuticle proteins accommodates, perpendicularly to the half-barrel β -strands, at least one extended chitin chain (Figure 4A) (Hamodrakas *et al.*, 2002).

Homology modeling results indicate that the basic structural motif of the CPR family is an antiparallel β -sheet structure with a “cleft” full of conserved aromatic residues that form “flat” hydrophobic surfaces on one “face,” perfectly positioned to stack against faces of the saccharide rings of chitin. One unpredicted feature in the model is a short two-turn α -helix at the C-terminus of the extended R&R Consensus. This C-terminal part of the model is reminiscent in some respects of the chitin-binding domain of an invertebrate chitin-binding lectin, a two stranded β -sheet followed by a helical turn (Suetake *et al.*, 2000). More detailed docking experiments (Iconomidou *et al.*, 2005), utilizing GRAMM (Vakser, 1996), showed that chitin protein chains may run parallel to the β -strands of the half- β -barrel (Figures 4B, 4C). Thus, β -barrels of cuticle proteins may intervene between the long chitin chains in cuticle without disrupting continuity. This parallel arrangement of cuticle protein β -strands with the chitin chains agrees with observations made by Atkins, over 20 years ago (Atkins, 1985), from X-ray diffraction patterns.

The inherent twist of the half-barrel β -sheet of the cuticle proteins and its observed packing arrangement at an angle with the chitin chains may provide a molecular basis for the morphological observation of a helicoidal twist in cuticle. These models were also subjected to analysis (Iconomidou *et al.*, 2005) of the positions of histidine residues, since they might play a role in cuticle sclerotization (Neville, 1975; Andersen *et al.*, 1995; Andersen, 2005) and appear to be very conserved in RR-2 sequences.

The general remarks that arise from the analysis are that histidines are positioned “exposed” either in turns

or at the edges of the half-barrel or its periphery, permitting interactions with chitin involved this way in cuticle sclerotization (Figure 4D). Alternatively, they could be involved in the variations of the water-binding capacity of cuticle and the interactions of its constituent proteins, because small changes of pH can affect the ionization of their imidazole groups (Andersen *et al.*, 1995).

These observations are in excellent agreement with the predictions made several years ago for the role of histidines from secondary structure predictions (Iconomidou *et al.*, 1999), and strengthen further the value of the models previously proposed for CPR proteins (Hamodrakas *et al.*, 2002).

5.5.3.2. CPF protein models The next obvious step involved attempts at elucidating the structural motifs and possible functions of cuticular proteins that belong to families where the “extended R&R Consensus” is absent. An appropriate choice was the CPF family of cuticular proteins (see section 5.3.2.3) (Togawa *et al.*, 2007). This family of cuticular proteins is of particular interest because they are expressed just before pupal or adult ecdysis, suggesting that these families are most probably components of the outer layer of pupal and adult cuticles – that is, they are likely located in the epi- or exo-cuticle. Actually, the epicuticle is one cuticular region that lacks chitin, suggesting that the CPF family of proteins may interact with components of the cuticle other than chitin.

Similarly to CPR proteins, members of the CPF family share significant sequence similarity to the crystallographically solved structure of bovine retinol-binding protein (RBP), which belongs to the class of lipocalins. The models of two proteins, AgamCPF3 from *Anopheles gambiae* and a CPF homolog, CG8541, from *Drosophila melanogaster*, were constructed based on this similarity (Papandreou *et al.*, 2010). The derived models (Figures 5A and 5B) indicate that the basic folding motif of CPFs is most probably an antiparallel, up-and-down, β -sheet full-barrel structure, unlike the proposed half-barrel for the CPR family.

The next step involved a high-resolution experiment, utilizing GRAMM (Vakser, 1996), of the proposed model with a NAG tetramer (Papandreou *et al.*, 2010). The results (Figure 6B) indicated that the tetramer does not fit into the binding pocket of the CPFs; rather, the CPFs might interact loosely with chitin chains, with their β -strands lying parallel to the chitin chains, in agreement with experimental observations (Atkins, 1985). Further evidence against a role of the CPFs in direct binding to chitin comes from failure of recombinant CPF proteins to bind to chitin (Togawa *et al.*, 2007). Comparative structural information in the paper by Papandreou *et al.* (2010) also indicates that carbohydrates should not bind in the pocket. Protein-carbohydrate interactions involve aromatic residues, and in the cleft of the

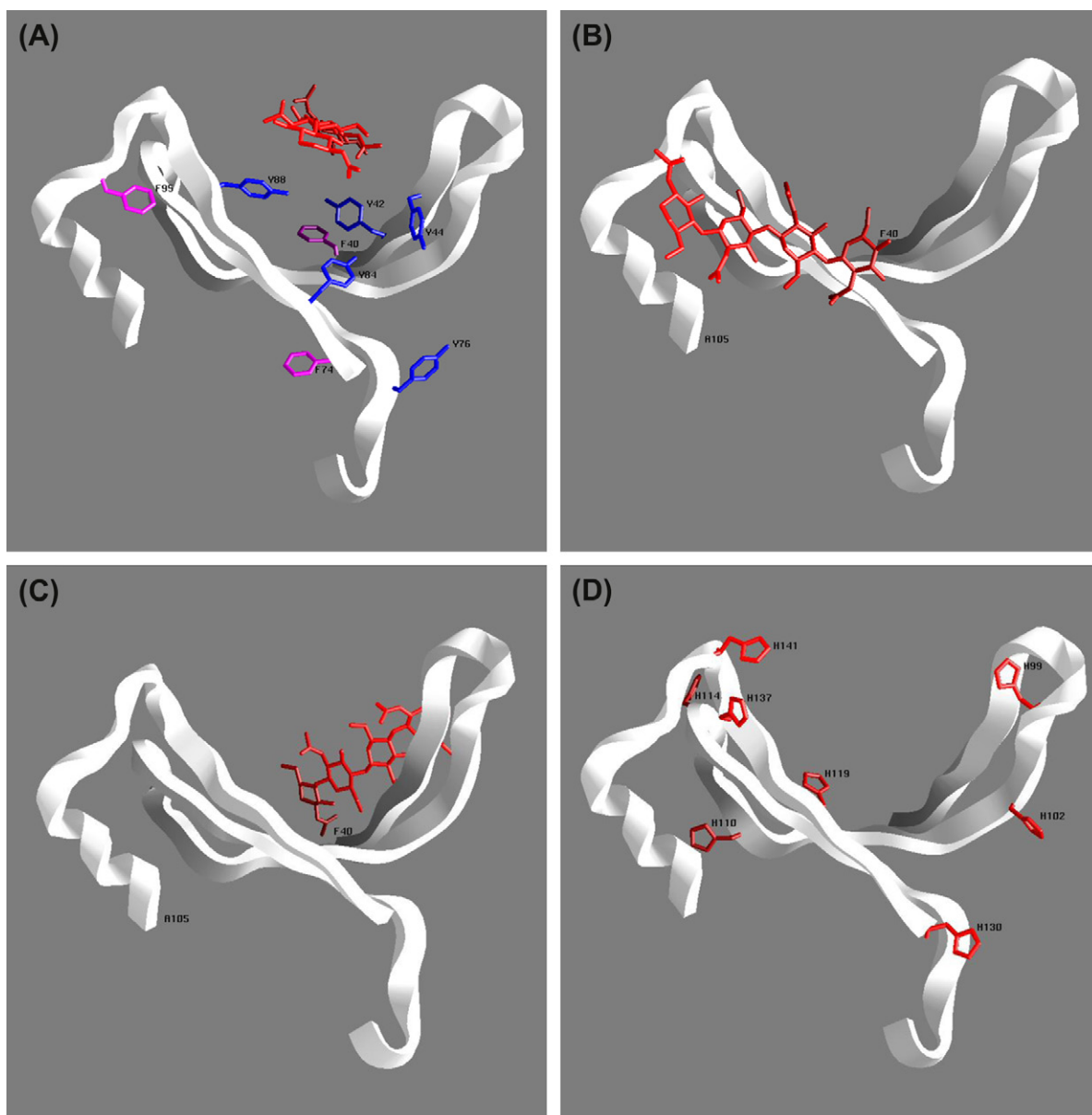


Figure 4 Ribbon models of cuticular proteins derived from homology modeling. (A) A ribbon model of cuticle protein structure, displayed using GRASP (Nicholls *et al.*, 1991). The structure of the representative RR-1 cuticle protein HCCP12 was modeled on that of bovine retinol-binding protein (RBP; PDB code 1FEN) (Zanotti *et al.*, 1994), utilizing the program WHAT IF (Vriend, 1990). Further details are in Hamodrakas *et al.* (2002). The side chains of several aromatic residues are shown and numbered, following the numbering scheme of the unprocessed HCCP12 sequence. The model structure has a “cleft” full of aromatic residues, which form “flat” surfaces of aromatic rings (upper side), ideally suited for cuticle protein–chitin chain interactions, and an outer surface (lower side), which should be important for protein–protein interactions in cuticle. The model is a complex of HCCP12 with an N-acetyl glucosamine (NAG) tetramer in an extended conformation. The complex was derived from a “low-resolution” docking experiment of a NAG tetramer, in an extended conformation, with the model of HCCP12, utilizing the docking program GRAMM (Vakser, 1996) and the default parameters of the program. (B) and (C) Two more possible complexes of HCCP12 with a NAG tetramer in an extended conformation derived from a “high-resolution” docking experiment, utilizing the program GRAMM (Vakser, 1996) and the default parameters of the program for “high resolution.” The two models presented in (B) and (C) are the two “top on the list,” most favorable complexes, whereas third on the list is a structure similar to that of (A). The one in (B) has the NAG tetramer more or less parallel to the last β -strand of the HCCP12 half β -barrel model, whereas that in (C) has the NAG tetramer more or less parallel to the first β -strand of the HCCP12 half β -barrel model. Note that in both (B) and (C) the chitin chain runs parallel to the β -strands, whereas in (A) the chain is arranged perpendicular to the β -strands. (D) A display of a model of the RR-2 protein AGCP2b. The numbering is that of the unprocessed protein. Histidine (H) side chains are shown as “ball and sticks,” in red, with their corresponding numbering (from Willis *et al.*, 2005).

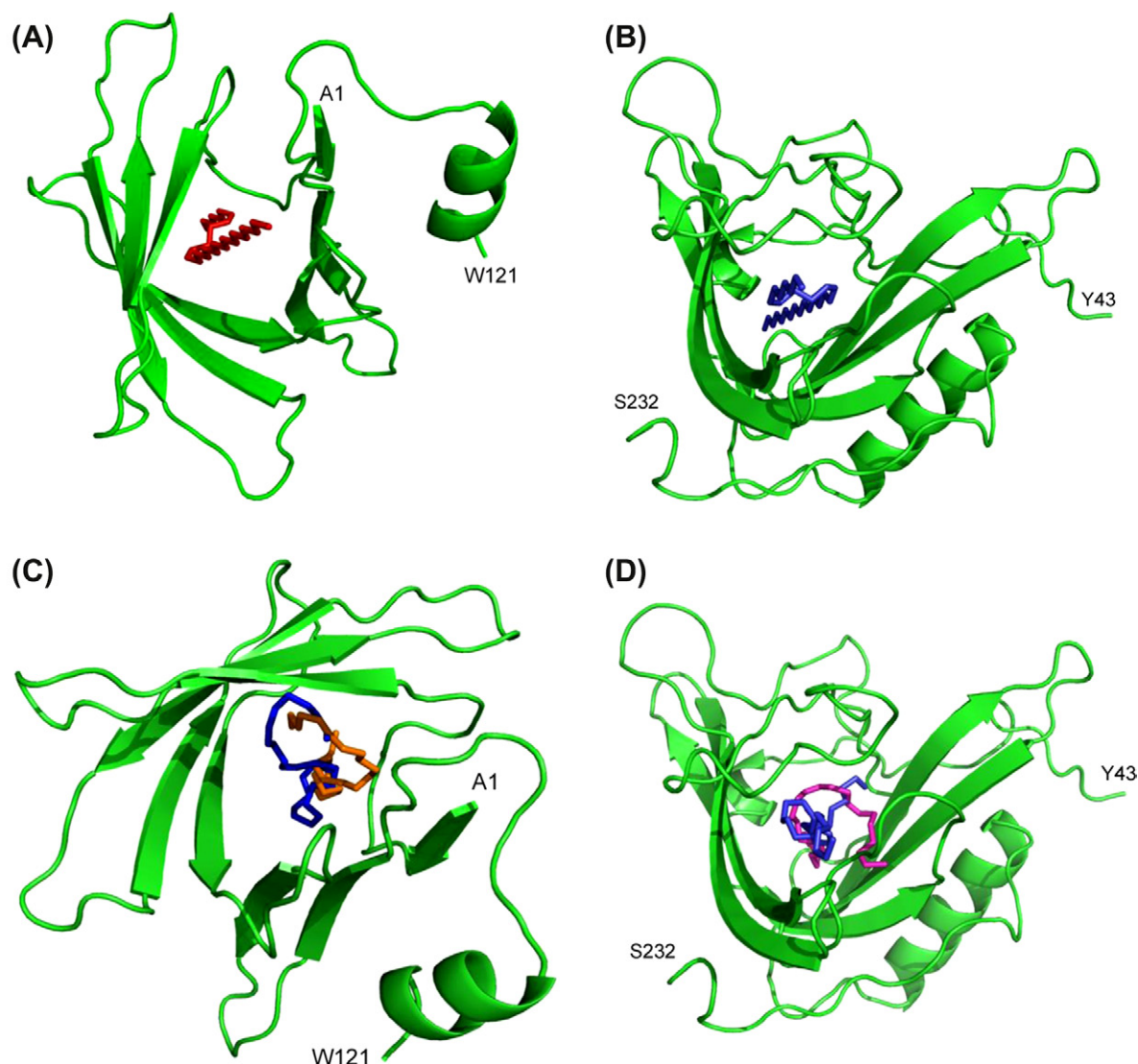


Figure 5 (A) A ribbon model of the cuticular protein AgamCPF3 structure (green), displayed using the software PyMOL (Delano, 2005). The model was modeled on that of bovine retinol-binding protein (RBP; PDB code 1FEN (Zanotti *et al.*, 1994) utilizing the software Modeller v9.2 (Sali and Blundell, 1993). The entire secreted protein, from A1 to W121, is shown in the model. It is complexed with 7(Z),11(Z)-heptacosadiene (7,11-HD), shown in red. The complex was derived from a docking experiment of 7,11-HD, with the model of AgamCPF3, utilizing the docking software Autodock4.2 (Morris *et al.*, 2009). The ligand is inside the “pocket” of the β -barrel of AgamCPF3. The ligand was considered as rigid, in its minimum energy conformation. The ligand represents a cluster of 4 out of 10 best solutions (runs). (B) A ribbon model of the CPF protein DmelCG8541 structure (green), constructed and displayed as in **Figure 5(A)**. The model comprises 190 of 257 residues of the secreted protein, from Y43 to S232. It is complexed with 7,11-HD, shown in blue. Details of the docking experiment that produced this complex are as in **Figure 5(A)**. The ligand represents a cluster of 7 out of 10 best solutions (runs). (C) A ribbon model of the cuticular protein AgamCPF3 structure (green), constructed and displayed as in **Figure 5(A)**. The entire secreted protein, from A1 to W121, is shown in the model. The complex was derived from a docking experiment of 7,11-HD, with the model of AgamCPF3, utilizing the docking software Autodock4.2 (Morris *et al.*, 2009). Two out of 10 best solutions (runs) for the ligand are shown in red and blue, respectively, inside the “pocket” of the β -barrel of AgamCPF3. The remaining eight solutions also show the ligand to reside inside the “pocket.” The 7,11-HD ligand was considered as flexible (all rotatable bonds were set free). (D) A ribbon model of the cuticular protein DmelCG8541 structure (green), constructed and displayed as in **Figure 5(B)**. The model comprises 190 of 257 residues of the secreted protein, from Y43 to S232. It is complexed with 7,11-HD. All other details of the docking experiment that produced the complex are as in **Figure 5(A)**. Two out of 10 best solutions (runs) for the ligand are shown in magenta and blue, respectively, inside the “pocket” of the β -barrel of DmelCG8541. The remaining eight solutions also show the ligand to reside inside the “pocket.” Reproduced from Papandreou *et al.* (2010), with permission.

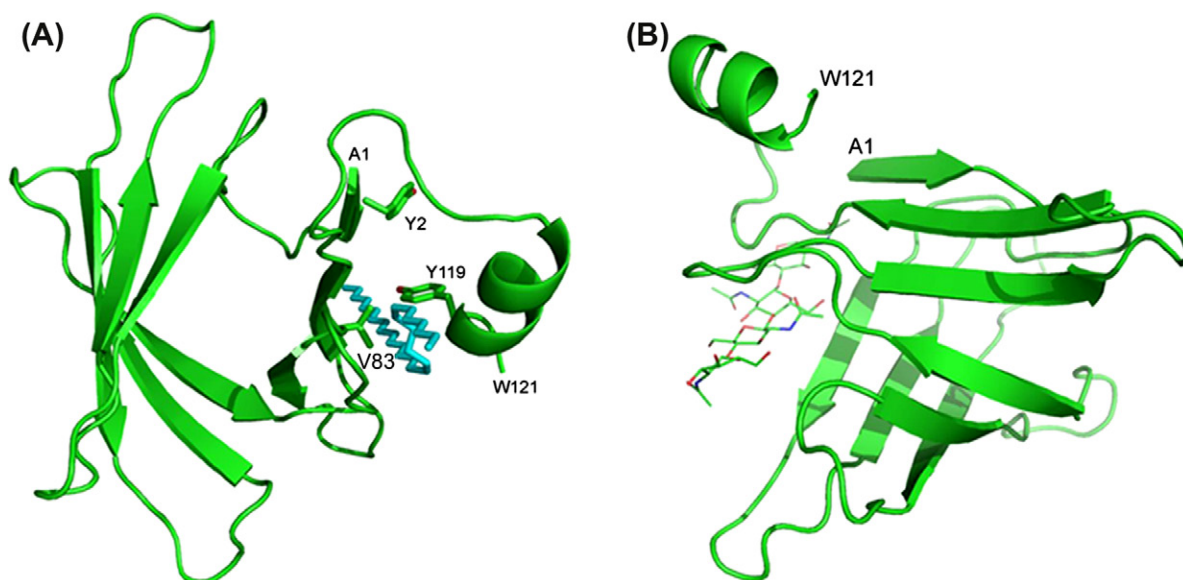


Figure 6 (A) A ribbon model of the cuticular protein AgamCPF3 structure (green), constructed and displayed as in **Figure 5(A)**. The entire secreted protein, from A1 to W121, is shown in the model. The complex was derived from a docking experiment of 7,11-HD (shown in cyan) with the model of AgamCPF3, utilizing the docking software Autodock4.2 (Morris *et al.*, 2009). It shows the ligand, outside the β -barrel of AgamCPF3, in contact with the “hydrophobic outer cluster” (see **Table 1** in Papandreou *et al.*, 2010). The side chains of three hydrophobic residues of the conserved “hydrophobic outer cluster,” Y2, V83, and Y119, are shown. The ligand was considered as rigid, in its minimum energy conformation. The ligand represents the cluster of the remaining 6 out of 10 best solutions (see **Figure 5(A)**). (B) A complex of AgamCPF3 (ribbon model shown in green) with a NAG tetramer (ball and stick model) in an extended conformation (taken as a chitin analog). The complex was derived from a “high resolution” docking experiment, utilizing the docking software GRAMM (Vakser, 1996) and the default parameters of the program, displayed using PyMol (Delano, 2005). The model presented is the “top of the list,” most favorable complex. Note that the “chitin chain” runs parallel to the β -strands, of at least half of the β -barrel, in agreement with experimentally derived data (Atkins, 1985). No solution was obtained with the “chitin chain” into the pocket of the β -barrel. The entire secreted protein, from A1 to W121, is shown in the model. Reproduced from Papandreou *et al.* (2010), with permission.

half-barrel model of HCCP12 are three critical aromatic residues, (Hamodrakas *et al.*, 2002); the model of HCCP66 has four aromatic residues in its cleft (Iconomidou *et al.*, 2005), whereas comparable residues in the two CPFs, AgamCPF3 and DmelCG8541, are hydrophobic but not aromatic.

Therefore, the questions that arise are, what is the functional role of the CPF proteins, and what fits within the cavity of the barrel? One possible function is that they intercalate among the chitin crystallites and chitin-binding proteins of the procuticle. However, this does not explain why they should form a binding pocket. Alternatively, if CPFs are components of the epicuticle, they could perhaps bind, as lipocalins do, to the lipoidal molecules, which are known to act as female contact sex pheromones in certain insect species (Antony and Jallon, 1982; Antony *et al.*, 1985) and are primarily located in the epicuticle (Andersen, 1979). We attempted to dock 7(Z), 11(Z)-heptacosadiene (7,11-HD), the predominant female-specific sex pheromone of *D. melanogaster* (Antony *et al.*, 1985), to the derived models of the *D. melanogaster* CPF protein, CG8541, utilizing GRAMM (Vakser, 1996) and Autodock4.2 (Morris *et al.*, 2009). The pheromone was considered both as rigid and flexible. Docking results

showed that this interaction is possible, indeed energetically favorable, and that 7,11-HD could fit into the binding pocket of the β -barrel or in the outer hydrophobic cluster (**Figures 5B, D**).

Complex formation between AgamCPF3 and 7,11-HD is also favored, although the molecular nature of sex pheromones in *An. gambiae*, if they exist, remains unknown, suggesting that a similar structure could easily bind to AgamCPF3, either inside the pocket (**Figures 5A, C**) or outside (**Figure 6A**). Microarray analyses have found significantly different levels of CPF3 transcript in adults of the incipient species M and S, and within the same form following a blood meal or in response to mating (Cassone *et al.*, 2008; Marinotti *et al.*, 2006; Rogers *et al.*, 2008). On the other hand, it is surprising that an epicuticular component would continue to be made and secreted into outer regions of the cuticle days after adult eclosion. An alternative occupant of the CPF binding pocket might just be intracuticular lipids that are present throughout the cuticle. Several of these cuticular lipids have chemical structures very similar to 7,11-HD (Hadley, 1981). Therefore, they would fit easily into the pocket of the β -barrel of the CPFs, or bind to their “outer hydrophobic cluster” (**Figures 5, 6A**).

Why do the proposed models correspond to a half-barrel model for CPRs and a full barrel for CPFs? The CPR Consensus region alone (<70 aa) was used, for that is the region of the protein that matches closely to retinol-binding protein; it is far too short to form a full barrel. By contrast, the CPF match is far longer, and compatible with a full barrel.

5.5.4. Fusion Proteins Establish a Role for the Extended R&R Consensus

Predictions of secondary and tertiary structure and experimental evidence supporting them (discussed above in sections 5.1–5.3) established that the extended R&R Consensus has the properties to serve as a chitin-binding motif. In particular, the planar surfaces of the predicted β -sheets will expose aromatic residues positioned for protein–chitin interaction. The ultimate test of these predictions would be to show that the extended consensus region is sufficient to confer chitin binding on a protein.

Rebers and Willis (2001) investigated this possibility by creating fusion proteins using the extended R&R Consensus from the *An. gambiae* putative cuticular protein, AGCP2b (Dotson *et al.*, 1998; now annotated as AgamCPR97). First, they expressed this protein in *E. coli* and isolated it from cell lysates. The construct used coded for the complete protein minus the predicted signal peptide, and had a histidine tag added to the N-terminus to facilitate purification. AGCP2b is a protein of 222 amino acids, with an RR-2 type of consensus. The purified protein bound to chitin beads, and could be eluted from these beads with 8M urea or boiling SDS. This established unequivocally that AGCP2b was a chitin-binding protein. Chitin binding previously had been obtained with mixtures of protein extracted from cuticles of two beetles and *D. melanogaster* (Hackman, 1955; Fristrom *et al.*, 1978; Hackman and Goldberg, 1978).

The next, and essential, step was to create a fusion protein uniting a protein that did not bind to chitin with the extended R&R Consensus region. Such a fusion was created between glutathione-S-transferase (GST) and 65 amino acids for AGCP2b – covering the region of pfam00379, the extended R&R Consensus:

APANYEFSYSVHDEHTGDIKSQHETR
HGDEVH GQYSLLDSDGHQRIVD
YHADHHTGFNAVVRREP

GST and the fusion protein were each affinity purified using a glutathione-sepharose column. GST alone did not bind to chitin but the fusion protein did, requiring denaturing agents for release.

Other experiments defined in more detail the requirements for converting GST into a chitin-binding protein. A shorter fragment of AGCP2b, 40 amino acids (underlined above), with the strict R&R Consensus (shown in

italics) did not bind chitin. Nor did the full construct when either the Y and F (highlighted) of the strict R&R Consensus or the T and D (bolded) of the extended consensus were “mutated” to alanine (Rebers and Willis, 2001).

In addition to establishing a function of the extended R&R Consensus, the experiments with “mutant” forms also provided confirmation of key elements in the models discussed in section 5.5.3. Substitution of the two conserved aromatic residues, postulated to be contact points with chitin, abolished chitin binding. With the TD “mutations,” alanines were substituted for two other conserved residues. These flank a glycine that is conserved in position in the “extended consensus” of all hard and many soft cuticles (Iconomidou *et al.*, 1999). According to the proposed model (Figure 4A), these two polar residues would point away from the hydrophobic “cleft” and thus should not participate in chitin binding. It should be noted, however, that this glycine is located at a sharp turn, at the end of the second β -strand (in the vicinity of H102 of Figure 4D). The substitution of two polar residues by two alanines may result in destruction of this turn and to improper folding, thus leading to a structure not capable of binding chitin.

These experiments established, at last, that the extended R&R Consensus is sufficient to confer chitin-binding properties on a protein, and thereby resolved years of speculation on the importance of this region. Since then, comparable experiments have been done with other proteins in the CPR family, both RR-1 and RR-2 forms; all confirm that the extended R&R Consensus can bind chitin (Togawa *et al.*, 2004, 2007; Qin *et al.*, 2009).

5.5.5. Members of Other Cuticular Protein Families Analyzed for Chitin Binding

Data are now available that identify members of other CP families as capable of binding chitin. Most notable was the finding that a recombinant BmorCPT1 bound to chitin (Tang *et al.*, 2010). Given that both CPAP1 and CPAP3 families have ChtBD2 domains (see section 5.3.2.11), it is expected that their members will also bind chitin, but this has only been demonstrated experimentally for a recombinant form of the gasp homolog, a member of the CPAP3 family, from *Choristoneura fumiferana* (Nisole *et al.*, 2010).

In contrast, using the same methodology, Togawa *et al.* (2007) failed to demonstrate that either AgamCPF1 or AgamCPF3 could bind to chitin. While the CPR proteins are easily purified after expression in *E. coli*, the CPF proteins required use of the Pierce Refolding Kit® for proper solubilization. Therefore, their failure to bind could be due to improper refolding, although the information from homology modeling is consistent with a lack of chitin binding (see section 5.5.3.2).

Chitinase, some lectins and proteins from peritrophic membranes all bind chitin (for review, see [Shen and Jacobs-Lorena, 1999](#)). What is unique about the extended R&R Consensus and members of the TWDL family is that they lack cysteine residues. These residues serve essential roles in the other types of chitin-binding proteins, forming disulfide bonds that hold the protein in the proper configuration for binding. While these other chitin-binding proteins have weak sequence similarities to one another, they do not approach the sequence conservation seen in the R&R Consensus throughout the arthropods, or the TWDL consensus in the groups in which it is found. [Rebers and Willis \(2001\)](#) suggested that the conservation of the R&R Consensus (shown in **Figure 1**) could well be due to the need to preserve a precise conformation of the chitin-binding domain in the absence of stabilizing disulfide bonds, and the same reasoning could now be applied to the TWDL sequences where consensus regions are evident (**Figure 2A**).

5.5.6. Summary of Interaction Studies

Four different types of data have been presented in section 5.5 analyzing the extended R&R Consensus: secondary structure predictions of anti-parallel β -sheets (section 5.5.1), experimental spectroscopic evidence from cuticles and cuticle extracts for the predominance of such β -sheets in cuticular protein conformation (section 5.5.2), models showing organization of the consensus into a half β -barrel with a groove that can accommodate chitin (section 5.5.3.1), and direct demonstration that the extended consensus is sufficient to confer chitin binding on a protein (section 5.5.4). These four types of data are all in agreement that the highly conserved amino acid sequence of the extended R&R Consensus forms a novel chitin-binding domain, albeit one that displays an essential feature of other proteins that interact with chitin – namely, the presentation of aromatic residues in a planar surface. Crystal structures of the cuticular protein–chitin complex are needed to assure that these inferences are correct.

5.6. Summary and Future Challenges

This chapter has summarized the wealth of information about cuticular proteins amassed since Silver's review in 1985. Most striking is that the several hundred-fold increase in sequences for structural cuticular proteins has revealed that the majority have a conserved domain (pfam00379) that is an extended version of the R&R Consensus. We now know that proteins with the R&R Consensus interact with chitin, and we can predict in some detail the features of their sequence that confer this property. We have not yet begun to analyze how the regions outside the Consensus contribute to cuticular properties. There is also direct experimental evidence that

a member of the TWDL family also binds chitin. However, we have yet to learn how proteins from other families contribute to cuticle structure, or how members of the different families interact with one another and other constituents of the cuticle.

Cuticular protein transcripts are turning up as major indicators of differential gene expression in analyses of insecticide resistance ([Vontas et al., 2007](#); [Zhang et al., 2008](#); [Awolola et al., 2009](#)), desiccation resistance ([Zhang et al. 2008](#)), resistance to heavy metals ([Shaw et al., 2007](#); [Roelofs et al., 2009](#)), response to changing photoperiod ([Gallot et al. 2010](#)), and even strain differences and mating ([Cassone et al., 2008](#); [Rogers et al., 2008](#)). Obviously, we need to understand how the individual cuticular proteins are contributing in such major ways to such important events.

Cuticular proteins with pfam00379 are one of the largest multigene families found in *D. melanogaster* ([Lepinet et al., 2002](#)), and their numbers are far larger in mosquitoes and *Bombyx*. We need more information about whether this multiplicity serves to allow rapid synthesis of cuticle, or whether different genes are used to construct cuticles in different regions. If the latter, the question becomes whether subtle differences in sequence are important for different cuticular properties, or if gene multiplication has been exploited to allow precise temporal and spatial control. We also need to learn how two hymenopterans, *Apis mellifera* and *Nasonia vitripennis*, manage with far fewer genes for cuticular proteins; is it their protected larval and pupal stages, or something else? The elegant immunolocalization studies that have been carried out were done with antibodies against proteins whose sequences for the most part are unknown. Now that we recognize that several genes may have almost identical sequences, we have to be very careful in designing specific probes for use in Northern analyses, for *in situ* hybridization, for qRT-PCR, and for immunolocalization, if our goal is to learn the use to which each individual gene is put.

Cuticular protein sequences are certain to be described in ever-increasing numbers as more insect genomes are analyzed. Describers need to be careful to submit to databases an indication of whether assignment as a cuticular protein is based on sequence alone, or on some type of corroborating evidence. It would be helpful if there were a more consistent system for naming cuticular proteins. At the very least, each protein should have a designation of genus and species, and a unique number ideally preceded by the gene family name – e.g., AgamCPR52, DmelTWDL12.

A wealth of sequence information is available already for cuticular proteins, but many challenges lie ahead for those who wish to continue to further our understanding of how the diverse forms and properties of cuticle are constructed extracellularly as these proteins self-assemble in proximity to chitin, and make specific contributions to the properties of the exoskeleton.

References

- Adam, B., Charlotiaux, B., Beaufays, J., Vanhamme, L., Godfroid, E., et al. (2008). Distantly related lipocalins share two conserved clusters of hydrophobic residues: Use in homology modeling. *BMC Struct. Biol.*, 8, 1.
- Agrawal, O. P., & Scheller, K. (1986). The formation of the chitin-arylphorin complex *in vitro*. In R. Muzzarelli, C. Jeuniaux, & C. W. Gooday (Eds.), *Chitin in Nature and Technology* (pp. 316–320). New York, NY: Plenum Press.
- Andersen, S. O. (1979). Biochemistry of insect cuticle. *Annu. Rev. Entomol.*, 24, 29–61.
- Andersen, S. O. (1998). Amino acid sequence studies on endocuticular proteins from the desert locust *Schistocerca gregaria*. *Insect Biochem. Mol. Biol.*, 28, 421–434.
- Andersen, S. O. (2000). Studies on proteins in post-ecdysial nymphal cuticle of locust, *Locusta migratoria*, and cockroach *Blaberus craniifer*. *Insect Biochem. Mol. Biol.*, 30, 569–577.
- Andersen, S. O. (2005). Cuticular sclerotization and tanning. In L. I. Gilbert, K. Iatrou, & S. S. Gill (Eds.), *Comprehensive Molecular Insect Science* (Vol. 4, pp. 145–170). Amsterdam, The Netherlands: Elsevier.
- Andersen, S. O. (2010a). Insect cuticular sclerotization: A review. *Insect Biochem. Mol. Biol.*, 40, 166–178.
- Andersen, S. O. (2010b). Studies on resilin-like gene products in insects. *Insect Biochem. Mol. Biol.*, 40, 541–551.
- Andersen, S. O., & Weis-Fogh, T. (1964). Resilin. A rubberlike protein in arthropod cuticle. *Adv. Insect Physiol.*, 2, 1–65.
- Andersen, S. O., Hojrup, P., & Roepstorff, P. (1995). Insect cuticular proteins. *Insect Biochem. Mol. Biol.*, 25, 153–176.
- Andersen, S. O., Rafn, K., & Roepstorff, P. (1997). Sequence studies of proteins from larval and pupal cuticle of the yellow meal worm, *Tenebrio molitor*. *Insect Biochem. Mol. Biol.*, 27, 121–131.
- Antony, C., & Jallon, J. M. (1982). The chemical basis for sex recognition in *Drosophila melanogaster*. *J. Insect Physiol.*, 28, 873–880.
- Antony, C., Davis, T. L., Carlson, D. A., Pechine, J. M., & Jallon, J. M. (1985). Compared behavioral responses of male *Drosophila melanogaster* (Canton S) to natural and synthetic aphrodisiacs. *J. Chem. Ecol.*, 11, 1617–1629.
- Apple, R. T., & Fristrom, J. W. (1991). 20-hydroxyecdysone is required for, and negatively regulates, transcription of *Drosophila* pupal cuticle protein genes. *Develop. Biol.*, 146, 569–582.
- Arakane, Y., Zhu, Q., Matsumiya, M., Muthukrishnan, S., & Kramer, K. J. (2003). Properties of catalytic, linker and chitin-binding domains of insect chitinase. *Insect Biochem. Mol. Biol.*, 33, 631–648.
- Arakane, Y., Dittmer, N. T., Tomoyasu, Y., Kramer, K. J., Muthukrishnan, S., et al. (2010). Identification, mRNA expression and functional analysis of several yellow family genes in *Tribolium castaneum*. *Insect Biochem. Mol. Biol.*, 40, 259–266.
- Ardell, D. H., & Andersen, S. O. (2001). Tentative identification of a resilin gene in *Drosophila melanogaster*. *Insect Biochem. Mol. Biol.*, 31, 965–970.
- Ashida, M., & Brey, P. T. (1995). Role of the integument in insect defense: Pro-phenol oxidase cascade in the cuticular matrix. *Proc. Natl. Acad. Sci. USA*, 92, 10698–10702.
- Atkins, E. D. T. (1985). Conformations in polysaccharides and complex carbohydrates. *Proc. Intl. Symp. Biomol. Struct. Interactions, Suppl. J. Biosci.*, 8, 375–387.
- Awolola, T. S., Oduola, O. A., Strode, C., Koekemoer, L. L., Brooke, B., & Ranson, H. (2009). Evidence of multiple pyrethroid resistance mechanisms in the malaria vector *Anopheles gambiae* sensu stricto from Nigeria. *Trans. R. Soc. Trop. Med. Hyg.*, 103, 1139–1145.
- Barry, M. K., Triplett, A. A., & Christensen, A. C. (1999). A peritrophin-like protein expressed in the embryonic tracheae of *Drosophila melanogaster*. *Insect Biochem. Mol. Biol.*, 29, 319–327.
- Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., et al. (2002). The Pfam protein families database. *Nucleic Acids Res.*, 30, 276–280.
- Baton, L. A., Robertson, A., Warr, E., Strand, M. R., & Dimopoulos, G. (2009). Genome-wide transcriptomic profiling of *Anopheles gambiae* hemocytes reveals pathogen-specific signatures upon bacterial challenge and *Plasmodium berghei* infection. *BMC Genomics*, 10, 257.
- Behr, M., & Hoch, M. (2005). Identification of the novel evolutionary conserved *obstructor* multigene family in invertebrates. *FEBS Lett.*, 579, 6827–6833.
- Binger, L. C., & Willis, J. H. (1990). *In vitro* translation of epidermal RNAs from different anatomical regions and metamorphic stages of *Hyalophora cecropia*. *Insect Biochem.*, 20, 573–583.
- Binger, L. C., & Willis, J. H. (1994). Identification of the cDNA, gene and promoter for a major protein from flexible cuticles of the giant silkworm *Hyalophora cecropia*. *Insect Biochem. Mol. Biol.*, 24, 989–1000.
- Bouhin, H., Charles, J. -P., Quennedey, B., Courrent, A., & Delachambre, J. (1992a). Characterization of a cDNA clone encoding a glycine-rich cuticular protein of *Tenebrio molitor*: Developmental expression and effect of a juvenile hormone analogue. *Insect Molec. Biol.*, 1, 53–62.
- Bouhin, H., Charles, J. -P., Quennedey, B., & Delachambre, J. (1992b). Developmental profiles of epidermal mRNAs during the pupal–adult molt of *Tenebrio molitor* and isolation of a cDNA encoding an adult cuticular protein: Effects of a juvenile hormone analogue. *Develop. Biol.*, 149, 112–122.
- Browder, L. W., Wilkes, J., & Rodenhiser, D. I. (1992). Preparative labeling of proteins with [³⁵S]methionine. *Anal. Biochem.*, 204, 85–89.
- Bruey-Sedano, N., Alabouvette, J., Lestrade, M., Hong, L., Girard, A., et al. (2005). The *Drosophila* ACP65 A cuticle gene: Deletion scanning analysis of cis-regulatory sequences and regulation by DHR38. *Genesis*, 43, 17–27.
- Burrows, M., Shaw, S. R., & Sutton, G. P. (2008). Resilin and chitinous cuticle form a composite structure for energy storage in jumping by frog hopper insects. *BMC Biol.*, 6, 41.
- Carter, D., & Locke, M. (1993). Why caterpillars do not grow short and fat. *Intl. J. Insect Morphol. Embryol.*, 22, 81–102.
- Cassone, B. J., Mouline, K., Hahn, M. W., White, B. J., Pombi, M., et al. (2008). Differential gene expression in incipient species of *Anopheles gambiae*. *Mol. Ecol.*, 17, 2491–2504.
- Centanin, L., Gorr, T. A., & Wappner, P. (2010). Tracheal remodelling in response to hypoxia. *J. Insect Physiol.*, 56, 447–454.

- Charles, J. -P. (2010). The regulation of expression of insect cuticle protein genes. *Insect Biochem. Mol. Biol.*, 40, 205–213.
- Charles, J. -P., 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.
- Charles, J. -P., Chihara, C., Nejad, S., & Riddiford, L. M. (1997). A cluster of cuticle protein genes of *Drosophila melanogaster* at 65A: Sequence, structure and evolution. *Genetics*, 147, 1213–1226.
- Charles, J. -P., Chihara, C., Nejad, S., & Riddiford, L. M. (1998). Identification of proteins and developmental expression of RNAs encoded by the 65A cuticle protein gene cluster in *Drosophila melanogaster*. *Insect Biochem. Mol. Biol.*, 28, 131–138.
- Chou, P., & Fasman, G. D. (1974a). Conformational parameters for amino acids in helical, β -sheet and random coil regions calculated from proteins. *Biochemistry*, 13, 211–221.
- Chou, P., & Fasman, G. D. (1974b). Prediction of protein conformation. *Biochemistry*, 13, 222–245.
- Condoulis, W. V., & Locke, M. (1966). The deposition of endocuticle in an insect, *Calpodex ethlius* Stoll (Lepidoptera: Hesperidae). *J. Insect Physiol.*, 12, 311–323.
- Cornell, J. C., & Pan, M. L. (1983). The disappearance of moulting fluid in the tobacco hornworm, *Manduca sexta*. *J. Exp. Biol.*, 107, 501–504.
- Cornman, R. S. (2009). Molecular evolution of *Drosophila* cuticular protein genes. *Plos ONE*, 4, e8345.
- Cornman, R. S. (2010). The distribution of GYR- and YLP-like motifs in *Drosophila* suggests a general role in cuticle assembly and other protein–protein interactions. *PLoS ONE*, 5, e12536.
- Cornman, R. S., & Willis, J. H. (2008). Extensive gene amplification and concerted evolution within the CPR family of cuticular proteins in mosquitoes. *Insect Biochem. Mol. Biol.*, 38, 661–676.
- Cornman, R. S., & Willis, J. H. (2009). Annotation and analysis of low-complexity protein families of *Anopheles gambiae* that are associated with cuticle. *Insect Mol. Biol.*, 18, 607–622.
- Cornman, R. S., Togawa, T., Dunn, W. A., He, N., Emmons, A. C., & Willis, J. H. (2008). Annotation and analysis of a large cuticular protein family with the R&R Consensus in *Anopheles gambiae*. *BMC Genomics*, 9, 22.
- Cox, D. L., & Willis, J. H. (1985). The cuticular proteins of *Hyalophora cecropia* from different anatomical regions and metamorphic stages. *Insect Biochem.*, 15, 349–362.
- Cox, D. L., & Willis, J. H. (1987a). Analysis of the cuticular proteins of *Hyalophora cecropia* with two dimensional electrophoresis. *Insect Biochem.*, 17, 457–468.
- Cox, D. L., & Willis, J. H. (1987b). Post-translational modifications of the cuticular proteins of *Hyalophora cecropia* from different anatomical regions and metamorphic stages. *Insect Biochem.*, 17, 469–484.
- Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: A sequence logo generator. *Genome Res.*, 14, 1188–1190.
- Csikos, G., Molnar, K., Borhegyi, N. H., Talian, G. C., & Sass, M. (1999). Insect cuticle, an *in vivo* model of protein trafficking. *J. Cell Sci.*, 112, 2113–2124.
- Delano, W. L., 2005. *The PyMOL Molecular Graphics System*, DeLano Scientific LLC, South San Francisco, CA 94080–1918, USA.
- Doctor, J., Fristrom, D., & Fristrom, J. W. (1985). The pupal cuticle of *Drosophila*: Biphasic synthesis of pupal cuticle proteins *in vivo* and *in vitro* in response to 20-hydroxyecdysone. *J. Cell Biol.*, 101, 189–200.
- Dotson, E. M., Cornel, A. J., Willis, J. H., & Collins, F. H. (1998). A family of pupal-specific cuticular protein genes in the mosquito *Anopheles gambiae*. *Insect Biochem. Mol. Biol.*, 28, 459–472.
- Elvin, C. M., Carr, A. G., Huson, M. G., Maxwell, J. M., Pearson, R. D., et al. (2005). Synthesis and properties of cross-linked recombinant pro-resilin. *Nature*, 437, 999–1002.
- Fechtel, K., Fristrom, D. K., & Fristrom, J. W. (1989). Prepupal differentiation in *Drosophila*: Distinct cell types elaborate a shared structure, the pupal cuticle, but accumulate transcripts in unique patterns. *Development*, 106, 649–656.
- Flower, D. R. (1996). The lipocalin protein family: Structure and function. *Biochem. J.*, 318, 1–14.
- Flower, D. R., North, A. C., & Sansom, C. E. (2000). The lipocalin protein family: Structural and sequence overview. *Biochim. Biophys. Acta.*, 1482, 9–24.
- Fraenkel, G., & Rudall, K. M. (1947). The structure of insect cuticles. *Proc. R. Soc. B*, 134, 111–143.
- Fristrom, D., & Fristrom, J. W. (1993). The metamorphic development of the adult epidermis. In M. Bate, & A. Martinez Arias (Eds.). *The Development of Drosophila melanogaster* (pp. 843–897). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Fristrom, D., Doctor, J., & Fristrom, J. W. (1986). Procuticle proteins and chitin-like material in the inner epicuticle of the *Drosophila* pupal cuticle. *Tissue and Cell*, 18, 531–543.
- Fristrom, J. W., Hill, R. J., & Watt, F. (1978). The procuticle of *Drosophila*: Heterogeneity of urea-soluble proteins. *Biochemistry*, 17, 3917–3924.
- Futahashi, R., & Fujiwara, H. (2005). Melanin-synthesis enzymes coregulate stage-specific larval cuticular markings in the swallowtail butterfly, *Papilio xuthus*. *Dev. Genes Evol.*, 215, 519–529.
- Futahashi, R., & Fujiwara, H. (2008). Identification of stage-specific larval camouflage associated genes in the swallowtail butterfly *Papilio xuthus*. *Dev. Genes Evol.*, 218, 491–504.
- Futahashi, R., Okamoto, S., Kawasaki, H., Zhong, Y. S., Iwanaga, M., et al. (2008). Genome-wide identification of cuticular protein genes in the silkworm, *Bombyx mori*. *Insect Biochem. Mol. Biol.*, 38, 1138–1146.
- Gallot, A., Rispe, C., Leterme, N., Gauthier, J. P., Jaubert-Possamai, S., & Tagu, D. (2010). Cuticular proteins and seasonal photoperiodism in aphids. *Insect Biochem. Mol. Biol.*, 40, 235–240.
- Geyer, P. G., & Corces, V. G. (1987). Separate regulatory elements are responsible for the complex pattern of tissue-specific and developmental transcription of the *yellow* locus in *Drosophila melanogaster*. *Genes and Develop.*, 1, 996–1004.

- Ghabrial, A., Luschnig, S., Metzstein, M. M., & Krasnow, M. A. (2003). Branching morphogenesis of the *Drosophila* tracheal system. *Annu. Rev. Cell Dev. Biol.*, 19, 623–647.
- Greenspan, R. J., & Ferveur, J. F. (2000). Courtship in *Drosophila*. *Annu. Rev. Genetics*, 34, 205–232.
- Griebenow, K., Santos, A. M., & Carrasquillo, K. G. (1999). Secondary structure of proteins in the amorphous dehydrated state probed by FTIR spectroscopy. Dehydration-induced structural changes and their prevention. *Intl. J. Vibr. Spectrosc.*, 3, 1–34.
- Gu, S., & Willis, J. H. (2003). Distribution of cuticular protein mRNAs in silk moth integument and imaginal discs. *Insect Biochem. Mol. Biol.*, 33, 1177–1188.
- Guan, X., Middlebrooks, B. W., Alexander, S., & Wasserman, S. A. (2006). Mutation of TweedleD, a member of an unconventional cuticle protein family, alters body shape in *Drosophila*. *Proc. Natl. Acad. Sci. USA*, 103, 16794–16799.
- Hackman, R. H. (1955). Studies on chitin. III. Absorption of proteins to chitin. *Austral. J. Biol. Sci.*, 8, 530–536.
- Hackman, R. H., & Goldberg, M. (1978). The non-covalent binding of two insect cuticular proteins by a chitin. *Insect Biochem.*, 8, 353–357.
- Hadley, N. F. (1981). Cuticular lipids of terrestrial plants and arthropods: A comparison of their structure, composition and waterproofing function. *Biol. Rev.*, 56, 23–47.
- Hall, J. C. (1994). The mating of a fly. *Science*, 264, 1702–1714.
- Hamodrakas, S. J., Kanellopoulos, P. N., Pavlou, K., & Tucker, P. A. (1997). The crystal structure of the complex of Concanavalin A with 4-methylumbelliferyl- α -D-glucopyranoside. *J. Struct. Biol.*, 118, 23–30.
- Hamodrakas, S. J., Willis, J. H., & Iconomidou, V. A. (2002). A structural model of the chitin-binding domain of cuticle proteins. *Insect Biochem. Mol. Biol.*, 32, 1577–1583.
- Han, Q., Fang, J., Ding, H., Johnson, J. K., Christensen, B. M., & Li, J. (2002). Identification of *Drosophila melanogaster* yellow-f and yellow-f2 proteins as dopachrome-conversion enzymes. *Biochem. J.*, 368, 333–340.
- He, N., Botelho, J. M., McNall, R. J., Belozarov, V., Dunn, W. A., et al. (2007). Proteomic analysis of cast cuticles from *Anopheles gambiae* by tandem mass spectrometry. *Insect Biochem. Mol. Biol.*, 37, 135–146.
- Henikoff, S., & Eghedarzadeh, M. K. (1987). Conserved arrangement of nested genes at the *Drosophila* Gart locus. *Genetics*, 117, 711–725.
- Henikoff, S., Keene, M. A., Fichtel, K., & Fristrom, J. W. (1986). Gene within a gene: Nested *Drosophila* genes encode unrelated proteins on opposite DNA strands. *Cell*, 44, 33–42.
- Holden, H. M., Rypniewski, W. R., Law, J. H., & Rayment, I. (1987). The molecular structure of insecticyanin from the tobacco hornworm *Manduca sexta* L. at 2.6 Å resolution. *EMBO J.*, 6, 1565–1570.
- 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. *Develop. Biol.*, 132, 292–303.
- Iconomidou, V. A., Willis, J. H., & Hamodrakas, S. J. (1999). Is β -pleated sheet the molecular conformation which dictates formation of helicoidal cuticle? *Insect Biochem. Mol. Biol.*, 29, 285–292.
- Iconomidou, V. A., Chryssikos, G. D., Gionis, V., Willis, J. H., & Hamodrakas, S. J. (2001). “Soft”-cuticle protein secondary structure as revealed by FT-Raman, ATR-FT-IR and CD spectroscopy. *Insect Biochem. Mol. Biol.*, 31, 877–885.
- Iconomidou, V. A., Willis, J. H., & Hamodrakas, S. J. (2005). Unique features of the structural model of “hard” cuticle proteins: Implications for chitin–protein interactions and cross-linking in cuticle. *Insect Biochem. Mol. Biol.*, 35, 553–560.
- Iijima, M., Hashimoto, T., Matsuda, Y., Nagai, T., Yamano, Y., et al. (2005). Comprehensive sequence analysis of horseshoe crab cuticular proteins and their involvement in transglutaminase-dependent cross-linking. *FEBS J.*, 272, 4774–4786.
- Ikeya, T., Persson, P., Kono, M., & Watanabe, T. (2001). The DD5 gene of the decapod crustacean *Penaeus japonicus* encodes a putative exoskeletal protein with a novel tandem repeat structure. *Comp. Biochem. Physiol. B*, 128, 379–388.
- Irving, P., Ubeda, J. M., Doucet, D., Troxler, L., Lagueux, M., et al. (2005). New insights into *Drosophila* larval haemocyte functions through genome-wide analysis. *Cell Microbiol.*, 7, 335–350.
- Jasrapuria, S., Arakane, Y., Osman, G., Kramer, K. J., Beeman, R. W., & Muthukrishnan, S. (2010). Genes encoding proteins with peritrophin A-type chitin-binding domains in *Tribolium castaneum* are grouped into three distinct families based on phylogeny, expression and function. *Insect Biochem. Mol. Biol.*, 40, 214–227.
- Jensen, U. G., Rothmann, A., Skou, L., Andersen, S. O., Roepstorff, P., & Hojrup, P. (1997). Cuticular proteins from the giant cockroach, *Blaberus craniifer*. *Insect Biochem. Mol. Biol.*, 27, 109–120.
- Kalinich, J. F., & McClain, D. E. (1992). An *in vitro* method for radiolabeling proteins with ^{35}S . *Anal. Biochem.*, 205, 208–212.
- Karouzou, M. V., Spyropoulos, Y., Iconomidou, V. A., Cornman, R. S., Hamodrakas, S. J., & Willis, J. H. (2007). *Drosophila* cuticular proteins with the R&R Consensus: Annotation and classification with a new tool for discriminating RR-1 and RR-2 sequences. *Insect Biochem. Mol. Biol.*, 37, 754–760.
- König, M., Agrawal, O. P., Schenkel, H., & Scheller, K. (1986). Incorporation of calliphorin into the cuticle of the developing blowfly, *Calliphora vicina*. *Roux's Arch. Develop. Biol.*, 195, 296–301.
- Kornezos, A., & Chia, W. (1992). Apical secretion and association of the *Drosophila* yellow gene product with developing larval cuticle structures during embryogenesis. *Mol. Gen. Genet.*, 235, 397–405.
- Kucharski, R., Maleszka, J., & Maleszka, R. (2007). Novel cuticular proteins revealed by the honey bee genome. *Insect Biochem. Mol. Biol.*, 37, 128–134.
- 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 Biochem. Mol. Biol.*, 24, 419–435.
- Lee, W. -J., & Brey, P. T. (1994). Isolation and identification of cecropin antibacterial peptides from the extracellular matrix of insect integument. *Anal. Biochem.*, 217, 231–235.
- Lemoine, A., Millot, C., Curie, G., & Delachambre, J. (1989). A monoclonal antibody against an adult-specific cuticular protein of *Tenebrio molitor* (Insecta: Coleoptera). *Develop. Biol.*, 136, 546–554.

- Lemoine, A., Millot, C., Curie, G., & Delachambre, J. (1990). Spatial and temporal variations in cuticle proteins as revealed by monoclonal antibodies, immunoblotting analysis and ultrastructural immunolocalization in a beetle, *Tenebrio molitor*. *Tissue and Cell*, 22, 177–189.
- Lemoine, A., Millot, C., Curie, G., Massonneau, V., & Delachambre, J. (1993). Monoclonal antibodies recognizing larval- and pupal-specific cuticular proteins of *Tenebrio molitor* (Insecta: Coleoptera). *Roux's Arch. Dev. Biol.*, 203, 92–99.
- Lespinet, O., Wolf, Y. I., Koonin, E. V., & Aravind, L. (2002). The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Res.*, 12, 1048–1059.
- Leung, H., Palli, S. R., & Locke, M. (1989). The localization of arylphorin in an insect, *Calpodes ethlius*. *J. Insect Physiol.*, 35, 223–231.
- Li, W., & Riddiford, L. M. (1992). Two distinct genes encode two major isoelectric forms of insecticynin in the tobacco hornworm, *Manduca sexta*. *Eur. J. Biochem.*, 205, 491–499.
- Locke, M. (1998). Epidermis. In F. W. Harrison, & M. Locke (Eds.). *Microscopic Anatomy of Invertebrates* (Vol. 11A, pp. 75–138). New York, NY: Wiley-Liss.
- Locke, M. (2001). The Wigglesworth Lecture: Insects for studying fundamental problems in biology. *J. Insect Physiol.*, 47, 495–507.
- Locke, M. (2003). Surface membranes, Golgi complexes and vacuolar systems. *Annu. Rev. Entomol.*, 48, 1–27.
- Locke, M., Kiss, A., & Sass, M. (1994). The cuticular localization of integument peptides from particular routing categories. *Tissue and Cell*, 26, 707–734.
- Lyons, R. E., Lesieur, E., Kim, M., Wong, D. C., Huson, M. G., et al. (2007). Design and facile production of recombinant resilin-like polypeptides: Gene construction and a rapid protein purification method. *Protein Eng. Des. Sel.*, 20, 25–32.
- Magkrioti, C. K., Spyropoulos, I. C., Iconomidou, V. A., Willis, J. H., & Hamodrakas, S. J. (2004). cuticleDB: A relational database of Arthropod cuticular proteins. *BMC Bioinformatics*, 5, 138.
- Marcu, O., & Locke, M. (1998). A cuticular protein from the moulting stages of an insect. *Insect Biochem. Mol. Biol.*, 28, 659–669.
- Marcu, O., & Locke, M. (1999). The origin, transport and cleavage of the molt-associated cuticular protein CECP22 from *Calpodes ethlius* (Lepidoptera: Hesperidae). *J. Insect Physiol.*, 45, 861–870.
- Marinotti, O., Calvo, E., Nguyen, Q. K., Dissanayake, S., Ribeiro, J. M., & James, A. A. (2006). Genome-wide analysis of gene expression in adult *Anopheles gambiae*. *Insect Mol. Biol.*, 15, 1–12.
- Mathelin, J., Bouhin, H., Quennedey, 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.
- Mathelin, J., Quennedey, B., Bouhin, H., & Delachambre, J. (1998). Characterization of two new cuticular genes specifically expressed during the post-ecdysial molting period in *Tenebrio molitor*. *Gene*, 211, 351–359.
- Mita, K., Morimyo, M., Okano, K., Koika, Y., Nohata, J., Suzuki, M. G., & Shimada, T. (2002). Construction of an EST database for *Bombyx Mori* and its applications. *Current Sci.* 83, 426–431.
- Missios, S., Davidson, H. C., Linder, D., Mortimer, L., Okobi, A. O., & Doctor, J. S. (2000). Characterization of cuticular proteins in the red flour beetle, *Tribolium castaneum*. *Insect Biochem. Mol. Biol.*, 30, 47–56.
- Molnar, K., Borhegyi, N. H., Csikos, G., & Sass, M. (2001). The immunoprotein scolexin and its synthesizing sites – the midgut epithelium and the epidermis. *Acta Biologica Hungarica*, 52, 473–484.
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., et al. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.*, 30, 2785–2791.
- Moussian, B. (2010). Recent advances in understanding mechanisms of insect cuticle differentiation. *Insect Biochem. Mol. Biol.*, 40, 363–375.
- Nakato, H., Izumi, S., & Tomino, S. (1992). Structure and expression of gene coding for a pupal cuticle protein of *Bombyx mori*. *Biochim. Biophys. Acta.*, 1132, 161–167.
- Nakatogawa, S., Oda, Y., Kamiya, M., Kamijima, T., Aizawa, T., et al. (2009). A novel peptide mediates aggregation and migration of hemocytes from an insect. *Curr. Biol.*, 19, 779–785.
- Nash, W. G. (1976). Patterns of pigmentation color states regulated by the γ locus in *Drosophila melanogaster*. *Develop. Biol.*, 48, 336–343.
- Neville, A. C. (1975). *Biology of the Arthropod Cuticle*. New York, NY: Springer-Verlag.
- Nicholls, A., Sharp, K. A., & Honig, B. (1991). Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.*, 11, 281–296.
- Nisole, A., Stewart, D., Bowman, S., Zhang, D., Krell, P. J., et al. (2010). Cloning and characterization of a Gasp homolog from the spruce budworm, *Choristoneura fumiferana*, and its putative role in cuticle formation. *J. Insect Physiol.*, 56, 1427–1435.
- Palli, S. R., & Locke, M. (1987). The synthesis of hemolymph proteins by the larval epidermis of an insect *Calpodes ethlius* (Lepidoptera: Hesperidae). *Insect Biochem.*, 17, 711–722.
- Papandreou, N. C., Iconomidou, V. A., Willis, J. H., & Hamodrakas, S. J. (2010). A possible structural model of members of the CPF family of cuticular proteins implicating binding to components other than chitin. *J. Insect Physiol.*, 56, 1420–1426.
- Perrakis, A., Ouzounis, C., & Wilson, K. S. (1997). Evolution of immunoglobulin-like modules in chitinases: Their structural flexibility and functional implications. *Folding and Design*, 2, 291–294.
- Qin, G., Lapidot, S., Numata, K., Hu, X., Meirovitch, S., et al. (2009). Expression, cross-linking, and characterization of recombinant chitin binding resilin. *Biomacromolecules*, 10, 3227–3234.
- Qiu, J., & Hardin, P. E. (1995). Temporal and spatial expression of an adult cuticle protein gene from *Drosophila* suggests that its protein product may impart some specialized cuticle function. *Develop. Biol.*, 167, 416–425.
- Rebers, J. F., & Riddiford, L. M. (1988). Structure and expression of a *Manduca sexta* larval cuticle gene homologous to *Drosophila* cuticle genes. *J. Mol. Biol.*, 203, 411–423.

- Rebers, J. E., & Willis, J. H. (2001). A conserved domain in arthropod cuticular proteins binds chitin. *Insect Biochem. Mol. Biol.*, 31, 1083–1093.
- Rebers, J. E., Niu, J., & Riddiford, L. M. (1997). Structure and spatial expression of the *Manduca sexta* MSCP14.6 cuticle gene. *Insect Biochem. Mol. Biol.*, 27, 229–240.
- Riddiford, L. M. (1994). Cellular and molecular actions of juvenile hormone I. General considerations and premorphic actions. *Adv. Insect Physiol.*, 24, 213–274.
- Riddiford, L. M., & Hice, R. H. (1985). Developmental profiles of the mRNAs for *Manduca* arylphorin and two other storage proteins during the final larval instar of *Manduca sexta*. *Insect Biochem.*, 15, 489–502.
- Roberts, P. E., & Willis, J. H. (1980). Effects of juvenile hormone, ecdysterone, actinomycin D, and mitomycin C on the cuticular proteins of *Tenebrio molitor*. *J. Embryol. Exp. Morph.*, 56, 107–123.
- Roelofs, D., Janssens, T. K., Timmermans, M. J., Nota, B., Marien, J., et al. (2009). Adaptive differences in gene expression associated with heavy metal tolerance in the soil arthropod *Orchesella cincta*. *Mol. Ecol.*, 18, 3227–3239.
- Rogers, D. W., Whitten, M. M., Thailayil, J., Soichot, J., Levashina, E. A., & Catteruccia, F. (2008). Molecular and cellular components of the mating machinery in *Anopheles gambiae* females. *Proc. Natl. Acad. Sci. USA*, 105, 19390–19395.
- Rondot, I., Quennedey, B., & Delachambre, J. (1998). Structure, organization and expression of two clustered cuticle protein genes during the metamorphosis of an insect *Tenebrio molitor*. *Eur. J. Biochem.*, 254, 304–312.
- Roter, A. H., Spofford, J. B., & Swift, H. (1985). Synthesis of the major adult cuticle proteins of *Drosophila melanogaster* during hypoderm differentiation. *Develop. Biol.*, 107, 420–431.
- Salí, A., & Blundell, T. L. (1993). Comparative protein modeling by satisfaction of spatial restraints. *J. Molec. Biol.*, 234, 779–815.
- Sass, M., Kiss, A., & Locke, M. (1993). Classes of integument peptides. *Insect Biochem. Mol. Biol.*, 23, 845–857.
- Sass, M., Kiss, A., & Locke, M. (1994a). Integument and hemocyte peptides. *J. Insect Physiol.*, 40, 407–421.
- Sass, M., Kiss, A., & Locke, M. (1994b). The localization of surface integument peptides in tracheae and tracheoles. *J. Insect Physiol.*, 40, 561–575.
- Scheller, K., Zimmermann, H. -P., & Sekeris, C. E. (1980). Calliphorin, a protein involved in the cuticle formation of the blowfly, *Calliphora vicina*. *Z. Naturforsch.*, 35 c, 387–389.
- Schneider, T. D., & Stephens, R. M. (1990). Sequence logos: A new way to display consensus sequences. *Nucl. Acids Res.*, 18, 6097–6110.
- Scholer, H. R. (1991). Octamania: The POU factors in murine development. *Trends Genet.*, 7, 323–328.
- Shaw, J. R., Colbourne, J. K., Davey, J. C., Glaholt, S. P., Hampton, T. H., et al. (2007). Gene response profiles for *Daphnia pulex* exposed to the environmental stressor cadmium reveals novel crustacean metallothioneins. *BMC Genomics*, 8, 477.
- Shen, Z., & Jacobs-Lorena, M. (1999). Evolution of chitin-binding proteins in invertebrates. *J. Mol. Evol.*, 48, 341–347.
- Silvert, D. J. (1985). Cuticular proteins during postembryonic development. In G. A. Kerkut, & L. I. Gilbert (Eds.). *Comprehensive Insect Physiology Biochemistry and Pharmacology* (Vol. 2, pp. 239–254). Oxford, UK: Pergamon Press.
- 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.
- Steinbrecht, R. A., & Stankiewicz, B. A. (1999). Molecular composition of the wall of insect olfactory sensilla – the chitin question. *J. Insect Physiol.*, 45, 785–790.
- Stiles, B. (1991). Cuticle proteins of the boll weevil, *Anthonomus grandis*, abdomen: Structural similarities and glycosylation. *Insect Biochem.*, 21, 249–258.
- Suderman, R. J., Andersen, S. O., Hopkins, T. L., Kanost, M. R., & Kramer, K. J. (2003). Characterization and cDNA cloning of three major proteins from pharate pupal cuticle of *Manduca sexta*. *Insect Biochem. Mol. Biol.*, 33, 331–343.
- Suetake, T., Tsuda, S., Kawabata, S., Miura, K., Iwanaga, S., et al. (2000). Chitin-binding proteins in invertebrates and plants comprise a common chitin-binding structural motif. *J. Biol. Chem.*, 275, 17929–17932.
- Talbo, G., Hojrup, 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. *Eur. J. Biochem.*, 195, 495–504.
- Tang, L., Liang, J., Zhan, Z., Xiang, Z., & He, N. (2010). Identification of the chitin-binding proteins from the larval proteins of silkworm, *Bombyx mori*. *Insect Biochem. Mol. Biol.*, 40, 228–234.
- Tews, I., Scheltiga, T., Perrakis, A., Wilson, K. S., & Dijkstra, B. W. (1997). Substrate-assisted catalysis unifies 2 families of chitinolytic enzymes. *J. Am. Chem. Soc.*, 119, 7954–7959.
- Togawa, T., Nakato, H., & Izumi, S. (2004). Analysis of the chitin recognition mechanism of cuticle proteins from the soft cuticle of the silkworm, *Bombyx mori*. *Insect Biochem. Mol. Biol.*, 34, 1059–1067.
- Togawa, T., Dunn, W. A., Emmons, A. C., & Willis, J. H. (2007). CPF and CPFL, two related gene families encoding cuticular proteins of *Anopheles gambiae* and other insects. *Insect Biochem. Mol. Biol.*, 37, 675–688.
- Togawa, T., Dunn, W. A., Emmons, A. C., Nagao, J., & Willis, J. H. (2008). Developmental expression patterns of cuticular protein genes with the R&R Consensus from *Anopheles gambiae*. *Insect Biochem. Mol. Biol.*, 38, 508–519.
- Trim, A. R. H. (1941). Studies in the chemistry of insect cuticle. I. Some general observations on certain arthropod cuticles with special reference to the characterization of the proteins. *Biochem. J.*, 35, 1088–1098.
- Uchiyama, T., Katouno, F., Nikaidou, N., Nonaka, T., Sugiyama, J., & Watanabe, T. (2001). Roles of the exposed aromatic residues in crystalline chitin hydrolysis by chitinase A from *Serratia marcescens* 2170. *J. Biol. Chem.*, 276, 41343–41349.
- Vakser, I. A. (1996). Low-resolution docking: Prediction of complexes for undetermined structures. *Biopolymers*, 39, 455–464.

- Vincent, J. F. V. (2002). Arthropod cuticle: A natural composite shell system. *Composites: Part A*, 33, 1311–1315.
- Vontas, J., David, J. P., Nikou, D., Hemingway, J., Christophides, G. K., et al. (2007). Transcriptional analysis of insecticide resistance in *Anopheles stephensi* using cross-species microarray hybridization. *Insect Mol. Biol.*, 16, 315–324.
- Vriend, G. (1990). WHAT IF: A molecular modeling and drug design package. *J. Mol. Graph.*, 8, 52–56.
- Vyas, N. K. (1991). Atomic features of protein–carbohydrate interactions. *Curr. Opin. Struct. Biol.*, 1, 723–740.
- Williams, C. M. (1980). Growth in insects. In M. Locke, & D. S. Smith (Eds.). *Insect Biology in the Future* (pp. 369–383). London, UK: Academic Press.
- Willis, J. H. (1996). Metamorphosis of the cuticle, its proteins, and their genes. In L. I. Gilbert, B. G. Atkinson, & J. Tata (Eds.). *Metamorphosis/Post-Embryonic Reprogramming of Gene Expression in Amphibian and Insect Cell* (pp. 253–282). London, UK: Academic Press.
- Willis, J. H. (1999). Cuticular proteins in insects and crustaceans. *Amer. Zool.*, 39, 600–609.
- Willis, J. H. (2010). Structural cuticular proteins from arthropods: Annotation, nomenclature, and sequence characteristics in the genomics era. *Insect Biochem. Mol. Biol.*, 40, 189–204.
- Willis, J. H., Rezaur, R., & Sehna, F. (1982). Juvenoids cause some insects to form composite cuticles. *J. Embryol. Exp. Morph.*, 71, 25–40.
- Willis, J. H., Ionomidou, V. A., Smith, R. F., & Hamodrakas, S. J. (2005). Cuticular proteins. In L. I. Gilbert, K. Iatrou, & S. S. Gill (Eds.). *Comprehensive Molecular Insect Science* (Vol. 4, pp. 79–109). Amsterdam, The Netherlands: Elsevier.
- Wittkopp, P. J., Vaccaro, K., & Carroll, S. B. (2002). Evolution of *yellow* gene regulation and pigmentation in *Drosophila*. *Current Biol.*, 12, 1547–1556.
- Wittkopp, P. J., Stewart, E. E., Arnold, L. L., Neidert, A., Haerum, B. K., et al. (2009). Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in *Drosophila*. *Science*, 326, 540–544.
- Wolfgang, W. J., & Riddiford, L. M. (1986). Larval cuticular morphogenesis in the tobacco hornworm, *Manduca sexta*, and its hormonal regulation. *Develop. Biol.*, 113, 305–316.
- Wolfgang, W. J., Fristrom, D., & Fristrom, J. W. (1986). The pupal cuticle of *Drosophila*: Differential ultrastructural immunolocalization of cuticle proteins. *J. Cell Biol.*, 102, 306–311.
- Wolfgang, W. J., Fristrom, D., & Fristrom, J. W. (1987). An assembly zone antigen of the insect cuticle. *Tissue and Cell*, 19, 827–838.
- Wybrandt, G. B., & Andersen, S. O. (2001). Purification and sequence determination of a yellow protein from sexually mature males of the desert locust, *Schistocerca gregaria*. *Insect Biochem. Mol. Biol.*, 31, 1183–1189.
- Zanotti, G., Marcello, M., Malpeli, G., Folli, C., Sartori, G., & Berni, R. (1994). Crystallographic studies on complexes between retinoids and plasma retinol-binding protein. *J. Biol. Chem.*, 269, 29613–29620.
- Zhang, J., Goyer, C., & Pelletier, Y. (2008). Environmental stresses induce the expression of putative glycine-rich insect cuticular protein genes in adult *Leptinotarsa decemlineata* (Say). *Insect Mol. Biol.*, 17, 209–216.
- Zhang, J., & Pelletier, Y. (2010). Characterization of cuticular chitin-binding proteins of *Leptinotarsa decemlineata* (Say) and post-ecdysial transcript levels at different developmental stages. *Insect Mol. Biol.*, 19, 517–525.
- Zhou, X., & Riddiford, L. M. (2002). Broad specifies pupal development and mediates the “status quo” action of juvenile hormone on the pupal–adult transformation in *Drosophila* and *Manduca*. *Development*, 129, 2259–2269.