4.2 Cuticular Proteins

J H Willis, University of Georgia, Athens, GA, USA
V A Iconomidou, University of Athens, Athens, Greece
R F Smith, Northern Arizona University, Flagstaff, AZ, USA
S J Hamodrakas, University of Athens, Athens, Greece

4.2.1. Introduction

In the previous 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 article 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 multigene families, and the even more limited genomic information revealed that similar genes were adjacent on a chromosome.

Progress in less than two decades has been spectacular, but not surprising given the advances in relevant techniques. Elegant immunolocalization analyses have solved the problem of the sources of cuticular proteins. Over 300 cuticular protein sequences are now available from six orders and over 20 species of insects. Listed in **Tables 1** and **2** are all but those that come exclusively from the annotation of the genomes of *D. melanogaster* and *Anopheles gambiae* or SilkBase, an extensive expressed sequence tag (EST) project in *Bombyx mori* (Mita *et al.*, 1999; SilkBase, 2003). Throughout this chapter, proteins are referred to by the names used in **Tables 1** and **2**. These tables also provide a gi number or an identifier from SwissProt that provides access to the complete sequence and the relevant references. Recombinant proteins have

Table 1 Characteristics of "structural" cuticular proteins that have been completely sequenced

Order/species	*Protein name*	*Number of amino acids*[a]	*Type*	*AAP(A/V) Repeats*[b]	*Other features*[c]	*Sequence method*[d]	*Identifier*[e]
Coleoptera							
Apriona germari	LCP10.7	87	RR-1			CT	16226511
	LCP12.3	114	RR-1			CT	21617523
	LCP12.6	120	RR-1			CT	21617525
Tenebrio molitor	TM-LCP-A1A	174	RR-2	2N 1C	2[AAP(I/L)]	DS	1706191
	TM-LCP-A2B	117	RR-2	1N 1C		DS	1706192
	TM-LCP-A3A	134	RR-2	1N 1C	1[AAP(I/L)]	DS	1706194
	TM-PCP-C1B	161		9	51 aa motif 1[AAP(I/L)]	DS	1706197
	TM-E1A	243		22		DS	913040
	TM-F1A	154		3	5 [AAP(I/L)]	DS	998953
	TM1-F1B	158		4	4 [AAPL]	DS	998954
	TM-F1C	162		4	5 [AAP(I/L)]	DS	998955
	TM-PCP-G1A	211		5	51 aa motif	DS	1706198
	TM-H1C = TMLPCP22	195		4	51 aa motif 1[AAP(I/L)]	CT DS	3123202*
	PCP5.8	53				DS	7511760
	PCP9.2	92				DS	7511761
	PCP15.6	161		5	1[AAP(I/L)]	DS	7441995
	PCP16.7	166				DS	7511759
	TMLPCP-23	214		2	51 aa motif 1[AAP(I/L)]	CT	3121955*
	TMLPCP29	276	RR-3	11N	1 (18 residue motif) 1[AAP(I/L)]	CT	2275132*
	ACP17	167		2	High G	CT	1078986*
	ACP20	191	RR-2		High G	CT	102879
	ACP-22	180	RR-2		2(6G), 1(8G), 1(15G)	CT	113012*
Dictyoptera							
Blaberus craniifer	BC-NCP1	87			Has 6 Cys, forming 3 potential S-S	DS	3023587
	BC-NCP2	99			8 As; 2 (18 residue motif)	DS	3023589
	BC-NCP4	127			2 (18 residue motif)	DS	3023588
	BC-NCP5	145			2 (18 residue motif)	DS	P82118
	BC-NCP6	139	RR-3		1 (18 residue motif)	DS	P82119
	BC-NCP7	145	RR-3		1 (18 residue motif)	DS	P82120
	BC-NCP8	195	RR-2			DS	P82121
	BC-NCP9	34				DS	P82122
Diptera							
Anopheles gambiae	AGCP2A	214	RR-2	1N	4 [(S/A)APIAH]	CT	2961109*
	AGCP2B	222	RR-2	1N	5 [(S/A)APIAH]	CT	2961110*
	AGCP2C	214	RR-2	2N	5 [(S/A)APIAH]	CT	2961111*
	AGCP2D	215	RR-2	2N	4 [(S/A)APIAH]	CT	2961113
Drosophila melanogaster	Dacp-1	116		3		CT	7580461*
	LCP-1	114	RR-1			DS CT	17380379*
	LCP-2	110	RR-1			DS CT	117634*
	LCP-3	96	RR-1			DS CT	117635*
	LCP-4	96	RR-1			DS CT	117636*
	PCP = GART INTRON	166	RR-1			CT	157483*
	ACP65A	85	RR-1			CT	1857602*
	LCP65Aa	83	RR-1			CT	1857600*
	LCP65Ab1	86	RR-1			CT pDS	1857597*
	LCP65Ab2	86	RR-1			CT pDS	1857595*
	LCP65Ac	91	RR-1			CT	1857593*
	LCP65Ad	90	RR-1			CT	1857495*
	LCP65Ae	83	RR-1			CT	1857604*
	LCP65Af	84	RR-1			CT	1857606*
	LCP65Ag1	87	RR-1			CT pDS	1857608*
	LCP65Ag2	87	RR-1			CT pDS	1857610*
	EDG-78	106	RR-1			CT	117639*
	EDG-84	171	RR-2	1C		CT	117640*

Table 1 Continued

Order/species	Protein name	Number of amino acids[a]	Type	AAP(A/V) Repeats[b]	Other features[c]	Sequence method[d]	Identifier[e]
	EDG91	138			8 (2G) 6(3G) 3 (4G)	CT	17380419*
	Ccp84Aa	188	RR-2	1N 6C		CT	4389433*
	Ccp84Ab	204	RR-2	1N 7C		CT	4389434*
	Ccp84Ac	199	RR-2	2C		CT	4389435*
	Ccp84Ad	182	RR-2	1N 6C		CT	4389436*
	Ccp84Ae	191	RR-2	1N 3C		CT	4389437*
	Ccp84Af	134	RR-2	1N		CT	4389438*
	Ccp84Ag	173	RR-2	6C		CT	4389439*
	Cry	457	RR-2		7Q, internal M	CT	22946279*
Drosophila miranda	LCP-1	122	RR-1			CT	3023591*
	LCP-2	110	RR-1			CT pDS	3023592
	LCP-3	96	RR-1			CT	231917*
	LCP-4	96	RR-1			CT	231917*
	LCP-3Y	96	RR-1			CT pDS	386246
Drosophila pseudoobscura	CP = GART INTRON	177	RR-1			CT	435017*
Drosophila simulans	DS-06238.4-like	188	RR-2			CT	9966434*
Drosophila yakuba	DS-06238.4-like	194	RR-2			CT	9966436*
Lucilia cuprina	CUT1	102	RR-1			CT	2565392
	CUT12	89	RR-1			CT	2565394
Hemiptera							
Myzus persicae	CP	208	RR-2	1N 4C	[(S/K)APAY] 1N 10C	CT	16798648
Aphis fabae	CP	208	RR-2	3C	[(S/K)APAY] 1N 12C	CT	29124938
Aphis gossypii	CP	203	RR-2	1N 3C	[(S/K)APAY] 10C	CT	29124934
Brevicoryne brassicae	CP	208	RR-2	1N 2C	[(S/K)APAY] 1N 12C	CT	29124932
Lipaphis erysimi	CP	208	RR-2	1N 3C	[(S/K)APAY] 1N 11C	CT	29124930
Rhopalosiphum maidis	CP	210	RR-2	2N 2C	[(S/K)APAY] 13C	CT	29124936
Lepidoptera							
Bombyx mori	PCP	235		3	3 (18 residue motif)	CT	1169137*
	BMWCP1A	205	RR-2	5N		CT	12862579
	BMWCP1B	205	RR-2	5N		CT	12862581
	BMWCP2	231	RR-2	9N		CT	12862583
	BMWCP3	215	RR-2	7N		CT	12862585
	BMWCP4	226	RR-2	3N		CT	12862587
	BMWCP5	274	RR-2	2N 1C		CT	12862589
	BMWCP6	186	RR-2	2N 2C		CT	
	BMWCP7A	157	RR-2	1N		CT	12862593
	BMWCP7B	157	RR-2	1N		CT	12862595
	BMWCP8	221	RR-2			CT	12862597
	BMWCP9	158	RR-1			CT	12862599
	BMWCP10	295	RR-1		Internal M	CT	23096118
	EDG84A homologue	180	RR-2	1N 1C		CT	3608259
	BMCP17	127	RR-1	1N		CT pDS	2204069
	BMCP18	89	RR-1	1N		CT pDS	5360249*
	BMCP22	158	RR-1			CT pDS	2204071
	BMCP30	223	RR-1			CT pDS	6634056*
	GCP1	149			17(GGY)	CT	15146344
Galleria melonella	PCP-52	338		1	Has 2 Cys; 21% Ala	CT	1086154
Helicoverpa armigera	LCP-1	95	RR-1		VVVV	CT	3913261
Hyalophora cecropia	HCCP12	89	RR-1			CT pDS	1169129*

Continued

Table 1 Continued

Order/species	*Protein name*	*Number of amino acids*[a]	*Type*	*AAP(A/V) Repeats*[b]	*Other features*[c]	*Sequence method*[d]	*Identifier*[e]
	HCCP66	112	RR-2			CT pDS	1169133*
Manduca sexta	LCP-14	109	RR-1			CT	117623
	CP14.6	90	RR-1			CT	3121956*
	LCP16/17	123	RR-1			CT	3121953*
	CP20	182	RR-1		7(GG) 3(GGG); glycosylated	CT DS	19548965
	CP27	165	RR-1		Glycosylated	CT DS	19743772
	CP36	327	RR-1		32(GG) 1(GGG); glycosylated	CT DS	22000820
Orthoptera							
Locusta migratoria	LM-ACP-abd4	116	RR-1		3 glycosylated forms	DS	461860
	LM-ACP-abd5	82	RR-1		Glycosylated	DS	3913394
	LM-ACP7	131	RR-2	2N 3C		DS	998751
	LM-ACP8	148	RR-2	5N		DS	84730
	LM-ACP19	157	RR-2	6N 1C	GYL motif	DS	1345864
	LM-ACP21	169	RR-2	6N 3C		DS	3287770
	LM-ACP38	163		14	GYL motif	DS	72263
	LM-NCP55	33		2		DS	446069
	LM-NCP62	88				DS	446070
	LM-ACP63	157		8	GYL motif	DS	1169130
	LM-ACP64	152		7	GYL motif	DS	1169131
	LM-ACP65	145		7	GYL motif	DS	1169132
	LM-ACP67a	98		6	51 aa motif	DS	416850
	LM-ACP67b	104		6	51 aa motif	DS	542520
	LM-ACP70	88		4	GYL motif	DS	416852
	LM-ACP76	139		6	GYL motif; 51 aa motif	DS	1169134
	LM-ACP79a	131		3	5 [GGG(L/Y)]	DS	1168721
	LM-ACP79b	131		3	5 [GGG(L/Y)]	DS	1168722
	LM-NCP4.9	46		1		DS	P82168
	LM-NCP5.1	48		1		DS	P82169
	LM-NCP6.4	62				DS	P82170
	LM-NCP9.5	90				DS	P82171
	LM-NCP18.7	193			4 (18 residue motif)	DS	P82165
	LM-NCP21.3	200		10		DS	P82167
	LM-NCP19.8	200	RR-2	3N 5C		DS	P82166
Schistocerca gregaria	SGAbd-1	184	RR-1		Glycosylated; (PPPPPPP)	DS	7511754
	SGAbd-2	135	RR-1		Glycosylated	DS	7511755
	SGAbd-3	119	RR-1		Glycosylated	DS	7441999
	SGAbd-4	116	RR-1		Glycosylated	DS	7441997
	SGAbd-5	82	RR-1		Glycosylated	DS	3913395
	SGAbd-6	82	RR-1		Glycosylated	DS	7511756
	SGAbd-8	139	RR-1		Glycosylated; internal M	DS	7511757
	SGAb-9	129	RR-1		Glycosylated	DS	7441998

[a]Sequence length of mature peptide; signal peptides were deleted using data from authors or SignalP V2.0 (http://www.cbs.dtu.dk/services/SignalP-2.0/).

[b]If a protein has an R&R Consensus, location of the AAP(A/V) repeats is given relative to the consensus.

[c]See Section 4.2.3.2.2 for description of these features.

[d]DS, direct sequencing of protein, a p indicates only a partial (generally N-terminal) sequence was obtained; CT, conceptual translation of a cDNA, genomic region, or EST product.

[e]Protein sequences and additional annotation can be found at http://www3.ncbi.nlm.nih.gov/Entrez/index.html. Sequences that have an identifier that begins with a letter can be found at http://us.expasy.org. An asterisk indicates that genomic sequence information is available.

Table 2 Characteristics of some nonstructural proteins that have been found in cuticle

Species	*Protein name*	*Number of amino acids*[a]	*Function*	*Sequence method*[b]	*Identifier*[c]
Schistocerca gregaria	Putative carotene binding protein	250	Transfers carotene into cuticle	DS	13959527
Caliphora vicinia	ARYLPHORIN A4	743	Found in cuticle	CT	114232
	ARYLPHORIN C223	743		CT	114236
Drosophila melanogaster	YELLOW	520	Positions melanin pigment in cuticle	CT	140623
Bombyx mori	CECROPIN A	41	Defense protein	CT	2493573
	CECROPIN B	41	Defense protein	CT	1705754
	PROPHENOLOXIDASE	675	Melanization enzyme	CT	13591614
Calpodes ethlius	CECP22	169	Cuticle digestion	CT	4104409
Manduca sexta	ARYLPHORINα	684		CT	114240
	ARYLPHORINβ	687		CT	1168527
	INSECTICYANIN A	189	Blue pigment	CT	102968
	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

[a]Sequence length of mature peptide; signal peptides were deleted using data from authors or SignalP V2.0 (http://www.cbs.dtu.dk/services/SignalP-2.0/).
[b]DS, direct sequencing of protein; CT, conceptual translation of a cDNA, genomic region, or EST product.
[c]Protein sequences and additional annotation can be found at: http://www3.ncbi.nlm.nih.gov/Entrez/index.html.

revealed a function for a highly conserved domain that was present in those first protein sequences discussed by Silvert. Structural predictions have elucidated the basis for this function. It is these developments that will be the focus of this review.

4.2.2. Cuticle Structure and Synthesis

4.2.2.1. Cuticle Morphology

4.2.2.1.1. 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 not only serves to protect the underlying epidermis from molting fluid enzymes that begin to digest the old cuticle, but, 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 4.2.2.1.3) 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 of the inner regions of the cuticle under the term "procuticle," encompassing both preecdysial and postecdysial secretions. The procuticle, then, is the region that combines chitin and cuticular proteins in various combinations and becomes sclerotized (see **Chapter 4.4**) and pigmented to varying degrees. This is the region depicted in electron micrographs showing stacks of precisely oriented lamellae. According to Locke (2001), apical microvilli bend in concert across the epithelial sheet and this movement serves to orient the laminae that will form lamellae. 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.

4.2.2.1.2. 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 intrainstar 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).

4.2.2.1.3. 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 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 *D. 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 ethlius* 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 more 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 (review: 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.*, 1996). Only two proteins with known sequence are among this group, TMACP22 and TMLPCP22.

Csikos *et al.* (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.

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, a lectin that recognizes *N*-acetylglucosamine, while other antigens failed to show this distribution. Notably, only C36 isolated from cuticle reacted with wheat germ agglutinin on lectin blots. Based on this evidence Locke *et al.* (1994) suggest that the isolated protein may have obtained its *N*-acetylglucosamine from chitin.

4.2.2.1.4. 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 (see **Chapter 3.7**). 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, 1980b; 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.*, 1996). 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 pupalike in morphology and the second with adult features. The adultlike cuticle was shown with immunolocalization to have ACP22 (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 ACP22.

Zhou and Riddiford (2002) used Northern analysis to characterize the somewhat nondescript cuticles made by *D. melanogaster* that had been manipulated by misexpressing the gene, *broad*, that codes for a transcription factor that appears before the larval–pupal molt in flies and moths. By following mRNAs for the adult cuticular protein ACP65A or the pupal cuticular protein Edg78E, they were able to demonstrate the essential role of *broad* in directing pupal development and thereby helped clarify the perplexing action of juvenoids in the higher Diptera.

4.2.2.2. The Site of Synthesis of Cuticular Proteins

One of the unresolved issues addressed in Silvert'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. Silvert 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 is to use Northern analysis to learn in what tissues and at which stages mRNA is present for a particular cuticular protein. 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 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, 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 solublized 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 almost all mature cuticular proteins lack methionine residues (see Section 4.2.3.2.1). The initiator methionine will be lost along with the entire signal peptide. Clear differences in labeling patterns with ^{35}S-methionine and ^{3}H-leucine have been found, with none of the major proteins from pharate adult cuticle of *D. melanogaster* or from larval cuticles of *Hyalophora 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. Results from several studies are summarized in **Table 3**. *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 disc from young larvae (Gu and Willis, 2003). A limitation of the technique is that cRNA probes sometimes bind to the cuticle itself, possibly obscuring detection of mRNA in the underlying epidermis (Fechtel *et al.*, 1989; Gu and 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 3**.

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 Northern analyses, 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 nonepidermal cells and by their inability to address heterogeneity of cell types within the epidermis.

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, carried out in *Calpodes*, revealed four classes of exported proteins that are handled by the epidermis.

Table 3 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*
Bombyx mori	BMLCP18	RR-1	Imaginal discs	EST		Gu and Willis (2003)
Drosophila melanogaster	EDG-78	RR-1	Larval and imaginal cells of prepupa	ISH		Fechtel *et al.* (1989)
Drosophila melanogaster	EDG-84	RR-2	Imaginal disc cells	ISH		Fechtel *et al.* (1989)
Drosophila melanogaster	PCP	RR-1	Prepupal thorax and abdomen	ISH		Henikoff *et al.* (1986)
Hyalophora cecropia	HCCP12	RR-1	Soft cuticle; imaginal discs	CD and ISH		Cox and Willis (1985), Gu and Willis (2003)
Hyalophora cecropia	HCCP66	RR-2	Hard cuticle	CD and ISH		Cox and Willis (1985), Gu and Willis (2003)
Locusta migratoria	LM-ACP7	RR-2	Hard cuticle	CD		Andersen *et al.* (1995a)
Locusta migratoria	LM-ACP8	RR-2	Hard cuticle	CD		Andersen *et al.* (1995a)
Locusta migratoria	LM-ACP19	RR-2	Hard cuticle	CD		Andersen *et al.* (1995a)
Manduca sexta	CP14.6	RR-1	Soft cuticle	ISH		Rebers *et al.* (1997)
Manduca sexta	LCP16/17	RR-1	Soft cuticle	ISH		Horodyski and Riddiford (1989)
Tenebrio molitor	ACP17	Glycine-rich	Hard cuticle	ISH	Strongest post-ecdysis	Mathelin *et al.* (1995, 1998)
Tenebrio molitor	ACP20	RR-2	Hard cuticle	ISH	Primarily pre-ecdysis	Charles *et al.* (1992)
Tenebrio molitor	ACP-22	RR-2	Hard cuticle	ISH, mAB	Pre-ecdysis	Bouhin *et al.* (1992a, 1992b)
Tenebrio molitor	TMLPCP22	51 aa motif	Hard and soft cuticle pre-ecdysis, then only soft cuticle	ISH, mAB	Primarily pre-ecdysis	Rondot *et al.* (1998)
Tenebrio molitor	TMLPCP23	51 aa motif	Hard and soft cuticle	ISH	Only pre-ecdysis	Rondot *et al.* (1998)
Tenebrio molitor	TMLPCP29	RR-3 and 18-residue motif	Hard and soft cuticle, except not posterior borders of sclerites	ISH	Post-ecdysis	Mathelin *et al.* (1998)

[a]For *in situ* hybridization, cuticle type was determined by nature of cuticle overlying the epidermis.
[b]CD, careful dissection prior to extraction of proteins; ISH, *in situ* hybridization used to localize mRNA; mAB, monoclonal antibody immunolocalization; EST, from *Bombyx* EST project (Mita *et al.* 1999).

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 h 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, hence has 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, T class, for proteins transported into cuticle, but not synthesized by the epidermis, was identified. 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 the 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.

4.2.2.3. 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 nonintegumental 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).

4.2.3. Classes of Proteins Found in Cuticles

4.2.3.1. Nonstructural Proteins

Nonstructural proteins that have been identified in cuticle are listed in **Table 2**.

4.2.3.1.1. 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 (see **Chapter 4.8**). 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* has been localized with immunocytochemistry in cuticles destined to become melanized (Kornezos and Chia, 1992). Thus it was found in association with larval mouth hooks, denticle belts, and Keilin's organs. Mutants of the gene *yellow* 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 β-alanyl-dopamine synthase, to allow melanin to be deposited. Flybase (2003) reports that 740 different alleles of *yellow* have been described, in 542 references beginning in 1916. The complete sequence of *yellow* has been determined for 13 species of *Drosophila* in addition to *D. melanogaster*. An examination of *yellow* expression revealed that both *cis*- and *trans*-regulation are responsible for differences in pigmentation patterns among different species (Wittkopp *et al.*, 2002). 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) (see **Chapter 4.4**).

Another cuticular protein implicated in pigmentation, putatively β-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 JH-binding proteins (see **Chapter 3.7**), as well as odorant-binding proteins (see **Chapter 3.15**). Wybrandt and Andersen (2001) suggest that it is involved in the transport of carotenes into epidermis and then the cuticle.

4.2.3.1.2. Enzymes Some of the enzymes involved in sclerotization have been identified in cuticle. Since they are discussed by Andersen (see **Chapter 4.4**) they will not be considered here.

Some enzymes that belong to the molting fluid become evident as the electrophoretic banding pattern of cuticular proteins 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 protein (CECP22). 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).

4.2.3.1.3. Defense proteins Also found in the cuticle are components of the insect defense system. In one study, cuticle was removed from *Bombyx* larvae 24 h 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 prophenoloxidase 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 prophenoloxidase was localized throughout the epicuticle and procuticle, and in a conspicuous orderly array on the basal side of the helicoidal chitin lamellae. An extraepidermal source is likely for this enzyme since no labeling was found in the epidermis, nor 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.

4.2.3.1.4. Arylphorins The final class of nonstructural proteins is the arylphorins, proteins with high content of aromatic amino acids and some lipid. These proteins, assumed to be hemolymph proteins, have been of special interest since the discovery by Scheller *et al.* (1980) that although calliphorin (the arylophorin from *Calliphora*) was found in cuticle, it seemed to come from the hemolymph, because radioactively 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 *Calpodes* integumental sheets *in vitro* that appeared in both cuticle and media. Thus arylphorin appeared to be a bidirectionally 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) and in the Golgi complexes of the fat body, and to show by quantifying 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).

4.2.3.2. Structural Proteins

4.2.3.2.1. Overview Slightly less 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.*, 1995a). As of June 2003, in addition to the nonstructural cuticular proteins discussed above, there are now 139 sequences available for what are postulated to be structural proteins. These numbers do not include almost 200 more that have been identified by protein prediction programs used to annotate the *D. melanogaster* and *Anopheles gambiae* genomes. These have been omitted because their annotation is still in a state of flux. All 139 sequences and some of their key features are listed in **Table 1**. Marcu and Locke (1998) have also published tabular summaries of a smaller number of cuticular proteins.

Unfortunately, cuticular protein terminology is not uniform. Most workers have included the initials of the genus and species. Some have named their proteins based on their molecular mass, others on the order in which they obtained them. Many *D. melanogaster* proteins have been designated by the chromosomal band to which a gene-specific probe hybridized. Some have been named after their sequence similarity to a particular cuticular protein from another species. Capitalization and the use of hyphens are erratic. Two different groups have worked on proteins from *Tenebrio*, and given two different names to one protein. Some of the names have included a designator for genus, species, and metamorphic stage (e.g., TMLCP-A1A). Although this designation is an accurate indication of the stage from which the protein was purified, it can inadvertently support the misconception that a particular cuticular protein is stage-specific. In *Tenebrio*, larval and pupal cuticular proteins are indistinguishable electrophoretically (Andersen, 1975; Roberts and Willis, 1980a; Lemoine and Delachambre, 1986; Andersen *et al.*, 1995b), and molecular analyses of several *Tenebrio* cDNAs found that all expressed in pupae are also expressed in larvae (Mathelin *et al.*, 1998; Rondot *et al.*, 1998). Furthermore, proteins that are a major component of the cuticle of one stage can be a minor component of another (Cox and Willis, 1985; Willis, 1986). A final complication is whether two almost identical proteins are allelic variants or distinct proteins. In some cases an "isoform" has been described. Genomic sequences, however, have revealed that stretches coding for proteins with very similar or indeed identical sequence may be linked on a chromosome (Charles *et al.*, 1997; Dotson *et al.*, 1998) (see Section 4.2.4.2). Thus the finding of "isoforms" may reflect distinct genes and hence distinct proteins.

Table 1 includes proteins from discrete genes even when two or more may have the same amino acid sequence. In **Table 1**, proteins isolated from cleaned cuticles were counted as cuticular proteins, as were proteins whose nucleic acid sequences were obtained using partial protein sequences or antibodies raised against cuticular protein to select corresponding cDNAs. In addition, **Table 1** contains numerous proteins that had been classified as cuticular proteins by their "discoverers" because their sequence, or a part thereof, was similar to a cuticular protein already in the databases. For many of those in the latter category, the source of the cDNA that led to the sequence came from integumental epidermis or imaginal discs or a cDNA hybridized to epidermal RNA in a Northern analysis. For some, especially those from *Tenebrio* studied by Delachambre's group, confirmation came from *in situ* hybridization of specific probes (see Section 4.2.2.1.3 and **Table 3**). But for many, sequence similarity served as the sole criterion.

Most of the structural cuticular proteins whose sequences were known in 1995 came from the efforts of Svend Andersen and his group, and a significant fraction (42%) still does. All of their sequences come from direct sequencing of purified cuticular proteins. Most of the other protein sequences come from sequencing cDNA or genomic DNA. For these, the length of the mature proteins can only be deduced by subtracting the amino acids of the signal peptide. In a few cases, N-terminal sequence data is available to assure that the signal peptide has been correctly identified. In cases where this information was not available, or when the original submissions did not provide this information, it was determined using the program SignalP (Nielsen *et al.*, 1997; Nielsen and Krogh, 1998). All lengths in **Table 1** represent the mature, processed protein.

One notable feature of the structural cuticular proteins is that almost all lack cysteine and methionine residues in the mature protein; the five exceptions to this are indicated in **Table 1** and **Figure 4**. Andersen (see **Chapter 4.4**) suggests that the reactivity of cystine and cysteine with *ortho*-quinones could interfere with sclerotization.

Most of the cuticular proteins identified to date are quite short. Those less than 100 amino acids account for 27% of the 139 sequences in **Table 1**, while those between 100 and 199 account for an additional 52%. Only three proteins have more than 300 amino acids. The largest is the gene for

a *D. melanogaster* corneal lens protein (Cry, drosocrystallin) with 457 amino acids, and a perfect RR-2 consensus (Janssens and Gehring, 1999). (See Section 4.2.3.2.3 for discussion of this consensus.) A cDNA of the appropriate size has been described (gi:2143072). The next largest is for the only cuticular protein characterized from *Galleria* (Kollberg *et al.*, 1995). This protein is unusual in that its only resemblance to known cuticular proteins is an abundance of alanine residues, and it is unique in having two cysteine residues. Yet its cDNA was selected with a polyclonal antibody raised against pupal cuticular proteins. The third largest protein is MSCP36; this is a high glycine protein that has a RR-1 consensus.

4.2.3.2.2. Motifs found in cuticular proteins The review by Andersen *et al.* (1995a) was the first to assemble a variety of motifs found in cuticular proteins. The occurrence of such motifs is given in **Table 1** and summarized in **Table 4**. Most common of these is a 28-residue region, first recognized by Rebers and Riddiford (1988) in seven cuticular proteins that is commonly referred to as the R&R Consensus. The original R&R Consensus is part of a longer conserved sequence – pfam00379 – and it is now apparent that there are three distinct forms of the extended R&R Consensus. These matters are discussed in detail below (see Section 4.2.3.2.3).

After the R&R Consensus, the next most common motifs were repeats of A-A-P-(A/V). These repeats were found in cuticular proteins both with and lacking the R&R Consensus; in sequences with the R&R Consensus they may occur N- or C-terminal to the extended Consensus. They are found in 46% of the sequences in **Table 1**. Thus, while abundant in cuticular proteins, they certainly are not diagnostic for this class of protein.

Andersen *et al.* (1995a) recognized several sequences with stretches of glycine, leucine, and tyrosine, beginning G-Y-G-L- or G-L-L-G. In **Table 1**, they are combined under the designation, G(Y/L) motifs. 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. There is an 18-residue motif found in seven cuticular proteins from four orders of insects (and two crustaceans), and its consensus has been described (Andersen, 2000). It occurs in proteins with and without the R&R Consensus. Also reported was a 51-residue motif so far identified only in cuticular proteins from *Locusta* and *Tenebrio*. It has not been found in proteins with the R&R Consensus (Andersen *et al.*, 1997). Other short repeats have been found in a limited number of proteins, from a single species. Proteins with the various motifs are identified in **Table 1**.

The basic sequences of the three long repeats are:

Original R&R Consensus: [G-x(8)-G-x(6)-Y-x(2)-A-x-E-x-G-F-x(7)-P-x-P.]
18 amino acid repeat: [(PV)-x-D-T-P-E-V-A-A-A-(KR)-A-A-(HF)-x-A-A-(HY).]
51 amino acid repeat: [x(6)-A-x(9)-R-S-x-G-x(4)-V-S-x-Y-x-K-x(2)-D-x(3)-S-S-V-x-K-x-D-x-R-x(2)-N-x(3).]

With this nomenclature, x is any amino acid, the number in parentheses represents the number of amino acids, and multiple letters in parentheses indicate that either of two amino acids may be present. This is the format used by MOTIF (2003), a resource that lets you search a given motif against various databases.

Andersen (2000) presented a model where proteins with the R&R Consensus bind to chitin and the other structural proteins remain free in the interfilament space.

4.2.3.2.3. Proteins with the R&R consensus The R&R Consensus is a common feature of cuticular proteins from all six orders of insects examined to date and it has also been recognized in cuticular proteins from arachnids and crustaceans (review: Willis, 1999).

Three distinct forms of the consensus have been recognized and named by Andersen (1998, 2000) RR-1, RR-2, and RR-3 (**Figures 1–3**). RR-1 is present in 51 (37%) of the proteins in **Table 1**.

Table 4 Summary characteristics of the 139 "structural" cuticular proteins that have been completely sequenced

	Number with AAP(AV)	*Number lacking AAP(AV)*	*Number 18-residue motif*	*Number 51-residue motif*	*Mean number H+K in extended R&R region (range)*	*Number with Met/Cys residues*	*Total proteins in class*
RR-1	2	49	0	0	3.7 (0–9)	3/0	51
RR-2	36	8	0	0	7.4 (2–19)	0/0	44
RR-3	1	2	3	0	6.3 (5–8)	0/0	3
Not RR	26	15	4	7		0/2	41

RR-1-bearing proteins have been isolated from flexible cuticles, while RR-2 proteins have been associated with hard cuticle. This generalization, based on relatively few cases (**Table 3**), has been used to link proteins to cuticle types in the absence of any other evidence.

The RR-1 proteins have the essential features of the original consensus (**Figure 1**). The vast majority have the short sequence -Y-x-A-x-E-x-G-(FY)-x(7)-P. N-terminal to this region the sequences diverge, but most have a series of three aromatic residues, such as Y-x-F-x-Y, that begins about 32 amino acids N-terminal to the start of the R&R Consensus (**Figure 1**). Another distinguishing feature defined by Andersen (1998) is a glutamic acid residue found in a conserved position [Y-x-A-x-**E**-G-(FY)] in 90% of the sequences in **Figure 1**.

The RR-2 proteins have a considerably extended consensus, first recognized by Bouhin *et al.* (1992a) and Charles *et al.* (1992). What is extraordinary about the RR-2 consensus is its conservation across six orders of insects. Only two single amino acid gaps are required to accommodate all 44 RR-2 sequences with this variant listed in **Table 1**. Twenty-two of the 70 residues in the extended consensus (31%) are virtually invariant and an additional 21 are represented by a single amino acid in over half of the proteins (**Figure 2**). The first few RR-2 sequences identified suggested that the residues G-F-N-A-V-V would be diagnostic (Andersen, 1998). The identification of more RR-2 sequences revealed that that region of the consensus is not perfectly conserved. Rather, all of the sequences have G-F-x-A-x-V, a configuration found in none of the RR-1 sequences.

There are other differences between RR-1- and RR-2-bearing proteins. Most of the RR-2 (82 %) have at least one A-A-P-(AV) motif, while only 2 (4%) of the RR-1 type have this motif (**Table 1**). Histidine and lysine residues can be more abundant in the extended consensus region of RR-2 proteins (**Table 4**, **Figures 1** and **2**). Only one RR-1 protein has been found with more than six histidine plus lysines in this region, while seven or more were present in 20 of the 44 RR-2 sequences. There is an invariant lysine in all RR-2 sequences and several other positions appear to be favorable for either of these amino acids. Over half of the RR-2 proteins, but only a quarter of the RR-1 proteins, have histidine as their final or penultimate C-terminal amino acid. Histidines and lysines are known to be reactive sites for sclerotizing agents (see **Chapter 4.4**), so it is possibly significant that proteins from "hard" sclerotized cuticles would have these amino acids in abundance. The number of potential sclerotization sites may also be related to whether a cuticle can grow by intussusception, something that would be impossible if the proteins were extensively cross-linked. The six aphid cuticular proteins, while all of the RR-2 type, have relatively few histidines plus lysines (only four or five); this paucity may reflect the need for cuticular expansion with the type of feeding and brooding of progeny that occurs in these animals.

An RR-3 form of the consensus has been based on three sequences from insects and two from other arthropods (Andersen, 2000). A tentative consensus (**Figure 3**) was constructed from the sequence alignment in Andersen (2000).

Whole genome sequencing has led to the need to classify annotated sequences. A valuable website, Pfam (2003a) has used hidden Markov modeling to define motifs characteristic of particular classes of proteins (Bateman *et al.*, 2002). When a protein sequence is searched against all known (and predicted) proteins using the BLAST server (Blast, 2003), 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 is Pfam00379, a 68 amino acid sequence that includes the extended R&R Consensus. It also goes under the name "chitin_bind_4," for reasons that will become apparent later (see Section 4.2.5.4). The pfam00379 was obviously based on proteins of both RR-1 and RR-2 classes, for it matches neither particularly well (**Figure 3**). This makes it particularly useful for a preliminary classification of a putative cuticular protein sequence.

The pfam00379 is found in 70% of the 139 cuticular proteins in **Table 1**, i.e., all the RR-1, RR-2, and RR-3 sequences. A complete listing of all sequences with pfam00379 can be found at Pfam (2003b) or at ENTREZ (2003), where you search Domains for pfam00379. There are no nonarthropod sequences with this consensus. Now that two species of insects (*D. melanogaster* and *A. gambiae*) have had their genomes completely sequenced and pfam00379 is being used to recognize cuticular proteins, the representation of proteins bearing this motif will be disproportionate. As of August 2003, 90 *D. melanogaster* sequences beyond those listed in **Table 1** have been found to have pfam00379, and using the ENTREZ site, *A. gambiae* had over 100 beyond those in **Table 1**.

Pfam00379 so far has been found to occur only once in a given protein, with the notable exception of 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). In the insect proteins, this motif has been found near the N- or C-terminus, or within the protein.

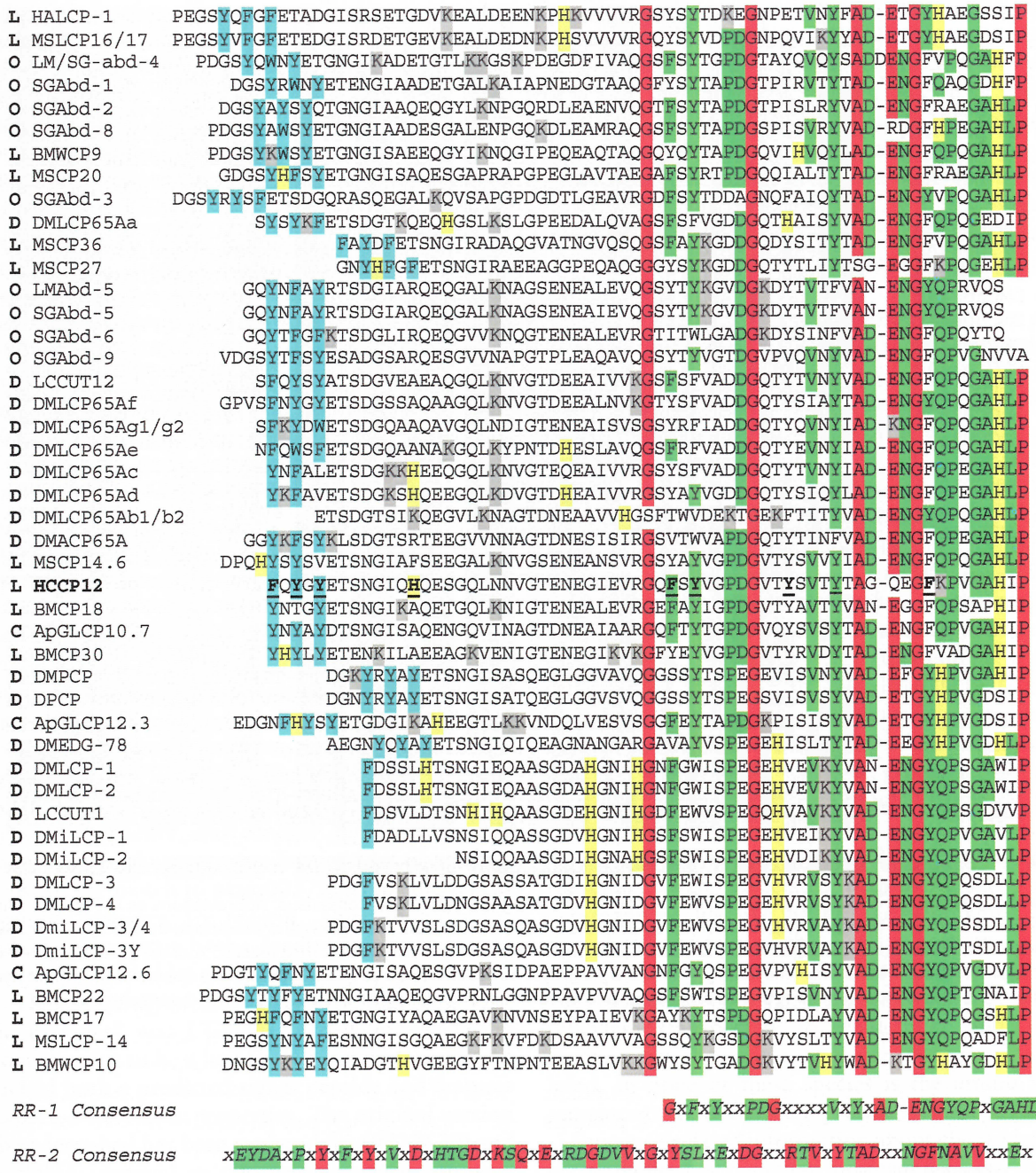

Figure 1 Alignment of the pfam00379 regions of 51 proteins with the RR-1 consensus. The pfam 00379 regions for RR-1 proteins were aligned with ClustalW (http://clustalw.genome.ad.jp/); only one internal gap was used to allow direct comparison with the RR-2 consensus. Orders of insects are: (**C**) Coleoptera, (**D**) Diptera, (**L**) Lepidoptera, (**O**) Orthoptera. Abbreviations for proteins as in **Table 1**. Four pairs of identical sequences are each presented on a single line. Red represents amino acids present in at least 95% of the proteins, green in the majority. Histidines are shown in yellow, lysines in light gray. A common triad of aromatic residues is shown in light blue. Bolded and underlined are several residues from HCCP12 that are shown in **Figure 5a–c**. An RR-1 consensus based on these sequences is given. The bottom line gives the consensus for RR-2 proteins from **Figure 2**, except that two single amino acid gaps needed to accommodate three atypical sequences were eliminated.

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 4.2.5) is only a beginning. Essential properties of cuticle remain to be explained, and important questions raised in the older literature about various means of achieving cuticle plasticity and the importance of hydration in cuticle stabilization must not be forgotten (Vincent, 2002 and references therein).

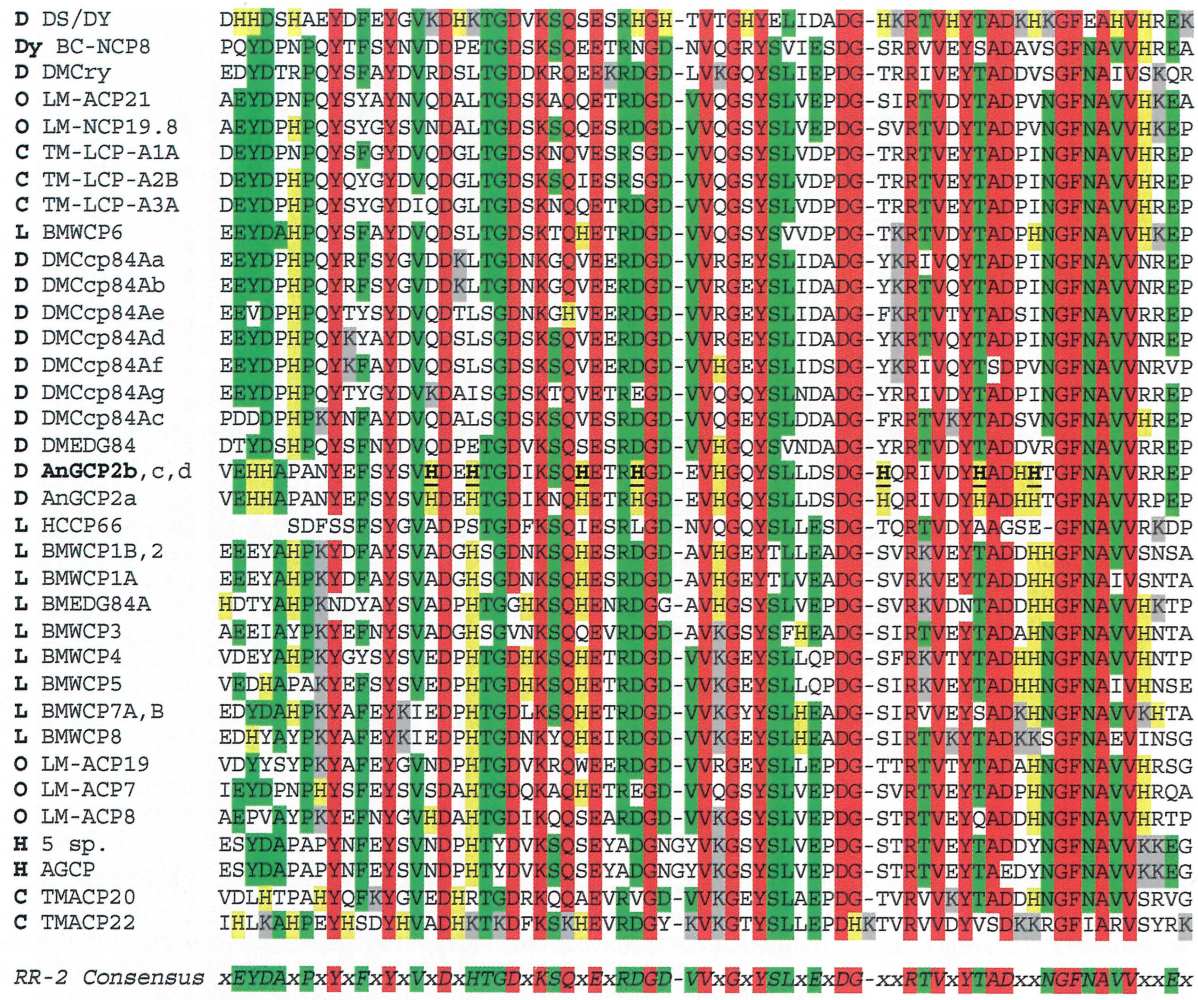

Figure 2 Alignment of the pfam00379 region of 44 cuticular proteins with the RR-2 consensus. Abbreviations as in **Figure 1** plus Dictyoptera (**Dy**) and Hemiptera (**H**). All hemipteran proteins except the one from *Aphis gossypii* are indicated by [***H*** 5 sp]. The order of the proteins was based on alignment by ClustalW. Red represents amino acids present in at least 95% of the proteins, green in the majority. Histidines are shown in yellow, lysines in light gray. The seven histidines in AGCP2b that are discussed in the text (see Section 4.2.5.4) and modeled in **Figure 5d** are bolded and underlined, beginning with residue 99. An RR-2 consensus is given.

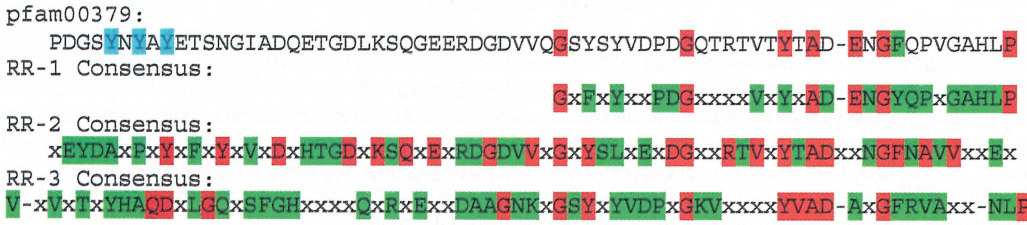

Figure 3 Consensus regions from the three types of RR cuticular proteins plus pfam00379. For pfam00379, the three aromatic residues found in RR-1 and RR-2 sequences (**Figures 1** and **2**) are shown in light blue. Indicated in red is the original R&R Consensus; the dash (-) was inserted to facilitate alignment with the other sequences. The green F was a Y in the original. RR-1 and RR-2 consensuses are from **Figures 1** and **2**, respectively. Red residues were found in at least 95% of the sequences, green in at least 50%. RR-3 was taken from Andersen (2000) using the sequences from the three insects and two arthropods arachnid and crustacean, that was identified as RR-3. Red residues were found in all five RR-3 sequences, green in three or four. The first and third dashes were inserted to allow alignment among all RR-3 sequences; the second was necessary to allow alignment with the RR-2 sequence.

4.2.3.2.4. Resilin The name resilin has been given to the rubberlike proteins responsible for the elasticity of jumping fleas and vibrating wings. Resilins are characterized by a high percentage of glycine (35–40%) and proline (7–10%). They are cross-linked with di- and tri-tyrosine residues (Andersen and Weis-Fogh, 1964). An intact protein corresponding to resilin has never been isolated from cuticle, presumably reflecting its insolublility after cross-linking. Ardell and Andersen (2001) used short peptide sequences that they had obtained from resilin-bearing regions of *Schistocerca* cuticle to probe the annotated *D. melanogaster* genome. Two candidate proteins had good matches to the locust peptides and to some that Lombardi and Kaplan (1993) had obtained from resilin in *Periplaneta americana*.

Ardell and Andersen concluded that predicted protein CG15920 (gi:24654243) of 620 amino acids was most likely to be a true proresilin (the non-cross-linked version), for in addition to the peptide matches, it had 35% glycine and 11% proline. 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 β-spiral, a common form for proteins with elastic properties (Ardell and Andersen, 2001). A cDNA (gi:27820115) is available that corresponds to most of the genomic sequence, but it lacks 45 internal amino acids. These correspond to the predicted second exon in the genomic sequence and contain almost the entire match to the locust proresilin peptides as well as most of the match to pfam00379. All of the repeat regions, however, are present. Thus, in order for CG15920 to match the locust resilin, it would have to be coded by an alternatively spliced form.

The second *D. melanogaster* gene that had "resilin peptides" was CG9036. Ardell and Andersen considered it to be a less likely candidate because it lacked both a predicted signal peptide and features expected for elastic properties. The original version they described has been replaced (gi:19922620) and a cDNA identical to this new sequence has been obtained. The new version has a predicted signal peptide of 19 amino acids. The mature protein of 198 amino acids has a pfam00379 region (occupying one-third of the mature protein) and is 20% glycine and 10% proline; it has neither of the repeats found in the other candidate protein.

Proof that either of these sequences is proresilin will require localization of the mRNA or protein to the tendons shown to have resilin in Diptera (Andersen and Weis-Fogh, 1964). The special properties of resilin justify further work to establish its sequence.

4.2.3.2.5. Glycosylation of cuticular proteins Glycosylation of cuticular proteins was first reported by Trim (1941) and in limited subsequent reports (review: Cox and Willis, 1987b). In recent years, posttranslational 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 matrix-assisted laser desorption ionization – mass spectrometry (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 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; 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 *Calpodes*, all the BD peptides (see Section 4.2.2.2) extracted from the cuticle were associated with α-D-glucose and α-D-mannose, just like most of the hemolymph proteins but very few of the C class 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* (see **Table 1**) 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 *Schistocerca 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.

4.2.4. Genomic Information

4.2.4.1. 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 were for *D. melanogaster* cuticular proteins LCP-1, -2, -3, and -4, and 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 acid. *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, for LCP-3 and -4, 85%, with similarity between the two groups about 60%. For the noncoding 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 unusually placed 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.

4.2.4.2. Chromosomal Linkage of Cuticular Proteins 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 are not known for all the members. A summary of such linked genes is presented in **Table 5**. Many more instances will become known as annotation of the *A. gambiae* and *D. melanogaster* genomes is completed.

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 (**Table 5**) were identified in a stretch of 22 kb with the direction of transcription, or more accurately the strand used, was

$$><<<<<<\quad >>>>>> .$$

The third gene in the cluster appeared to be a pseudogene. Several important features were found. First, the number of *Lcp-b* genes within the cluster was variable among different strains of *D. melanogaster*, two in the original line (with identical coding sequences), one in another, and three in a third. On the other hand, three copies of the *Lcp-g*

Table 5 Linked cuticular protein genes

Species	*Length of DNA examined*	*Number of genes found*	*Protein names*	*Reference*
Anopheles gambiae	17.4 kb	3	AGCP2a, 2b, 2c	Dotson *et al.* (1998)
Drosophila melanogaster	9 kb	6	DM-LCP1, 1ψ, 2, 3, 4 and one other	Snyder *et al.* (1982)
Drosophila melanogaster	20.5 kb	8	EDG-84, 84Aa, Ab, Ac, Ad, Ae, Af, Ag	Apple and Fristrom (1991), Kaufman *et al.* (1990), genome annotation
Drosophila melanogaster	22 kb	12	ACP65A, LCP65Aa, aψ, b1, b2, c, d, e, f, g1, g2, g3	Charles *et al.* (1997, 1998)
Manduca sexta	20 kb	3	LCP16/17 + 2 not named	Horodyski and Riddiford (1989)
Tenebrio molitor	3.9 kb	2	TMLPCP 22, 23	Rondot *et al.* (1998)

genes (also with 100% sequence identity) were found in all three strains. Comparison of cDNA sequences and genes revealed that *Lcp-b1* and *-b2* lacked introns. Both *Lcp-b1* and *-b2* had tracts of As at the 3′ end of the genes, as well as short flanking direct repeats. These features are consistent with the *Lcp-b* genes arising in this cluster by retrotransposition. The sequence data also indicated that *Acp* and *Lcp-a* lacked introns. The rest of the genes had introns but not the common one interrupting the signal peptide (see Section 4.2.4.3).

Evidence for gene conversion between the *Lcp-c* gene and those on the right side of the cluster was also found after a careful analysis of the sequences (Charles *et al.*, 1997).

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 difference 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).

4.2.4.3. Intron Structure of Cuticular Protein Genes

Genomic sequence data is available for 45 cuticular proteins in **Table 1**. None of these has more than two introns; most have only one, and these introns are in a very conserved positions. Thirty-five of these proteins have an intron that interrupts the signal peptide. In 25 of these sequences, interruption occurs after four amino acids (12 bp); in the remainder, two to eight amino acids are coded for before the intron begins. The PSORT tool (Psort, 2003) calculates the location of discrete regions of a potential signal peptide using modifications of McGeoch's method (McGeoch, 1985) and reports this information as "PSG" data. These short stretches, confined to the first exon, were shown to be identical to the N-terminal positively charged region in 20 cases; all but four of the rest were but one amino acid longer. Whether this correlation of the coding region of the first exon with the N-region of the signal peptide is because it is so short, or because it reflects something more fundamental awaits further exploration. Most of the putative cuticular protein genes in the annotated *A. gambiae* database are missing their initiator methionine, probably because it resides in a short exon, coding for these few amino acids of the signal plus 5′ untranslated nucleotides. The programs, unfortunately, are not yet trained to recognize a configuration with such a short open reading frame.

Genes for four of the proteins listed in **Table 1** (HCCP12 and MSLCP 16/17, MSCP14.6, and TMACP22) have two introns, one interrupting the signal peptide, the other occurring shortly after the beginning of the pfam00379 region. The intron-bearing *D. melanogaster* genes in the cluster at 65A (see above) have their sole intron at the internal position. This 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 *Drosophila* cuticular protein whose gene is located within the region corresponding to the first intron of *Gart*, a gene that encodes three proteins involved in the purine pathway. The gene for this RR-1 protein (*Gart Intron*) 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).

4.2.4.4. 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 (Riddiford, 1994). It would be expected, therefore, that some might have hormone response elements (see **Chapter 3.5**). Imperfect matches to ecdysteroid response elements (EcREs) from *D. melanogaster* have been 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).

Both *Bombyx* PCP 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).

As more genomic sequence information becomes available, identification of regulatory elements and verification of their action is certain to be productive.

4.2.5. Interaction of Cuticular Proteins with Chitin

Ever since the R&R Consensus was recognized in 1988, it has been predicted that it must be playing an important role in cuticle. As more and more sequences were discovered with the Consensus, and as it was learned that it also is present in cuticles formed by arachnids and crustaceans, this prediction became more likely. Several workers suggested that the role of the R&R Consensus might be to bind to chitin (Bouhin *et al.*, 1992a; Charles *et al.*, 1992; Andersen *et al.*, 1995a).

Four complementary routes have been followed to learn more about the function of this consensus region. The first was to analyze it with appropriate programs to generate predictions of 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*. Third, the tertiary structure of the extended Consensus has been modeled, and the fourth route was a direct experimental approach to test whether the extended Consensus could bind to chitin.

4.2.5.1. Secondary Structure Predictions

Prediction of secondary structure was carried out on the extended R&R Consensus region (67–68 amino acids, the pfam00379 region) of cuticular proteins representing different metamorphic stages and four different orders (Iconomidou *et al.*, 1999). For each protein, individual predictions of α-helix, β-sheet, and β-turn/coil/loops were carried out using several different predictions programs. These predictions on individual proteins were combined to produce joint prediction histograms for the two classes of proteins. (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. There appeared to be four β-strands in the RR-2 proteins and only three in the RR-1. Three other features were immediately apparent:

1. The three invariant glycines of the original R&R Consensus correspond exactly at the maxima of β-turn/loop predictions, and it is well known that glycines are good turn/loop formers (Chou and Fasman, 1974a, 1974b).
2. With both classes of cuticular proteins, the sheets showed an amphipathic character, i.e., one face is polar, the other nonpolar. 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 amino acids 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).
3. The turn/loop regions frequently contained histidines. This would place them "exposed" at the "edges" of a β-pleated sheet. Histidines are involved in cuticular sclerotization (see **Chapter 4.4**) and are involved in the variations of the water-binding capacity of cuticle and the interactions of its constituent proteins. This occurs because small changes of pH can affect the ionization of their imidazole group (Andersen *et al.*, 1995a).

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

4.2.5.2. 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 8 M 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). 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 spectra 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 bands could be attributed to side-chain vibrations of amino acids with aromatic rings, tyrosine, phenylalanine, and tryptophan. Bands in the amide I region (1600–1700 cm^{-1}) of the Raman spectra of the extracted cuticle proteins and of the difference spectrum exhibited a well-defined maximum at 1669 cm^{-1}, typical of β-sheet structure. The absence of bands at ~1650 cm^{-1} indicates that α-helical structures are not favored. The amide III range (1230–1320 cm^{-1}) is relatively free from side group vibrations and, thus, highly diagnostic of secondary structure. The extracted proteins had a doublet at 1241 and 1268 cm^{-1}; the former can be assigned to β-sheet and the latter to β-turns or coil.

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 protein 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 (see Section 4.2.5.1).

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 two-dimensional 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).

4.2.5.3. Modeling of Chitin-Binding Domains of Cuticular Proteins

Secondary structure prediction and experimental data summarized above (see Sections 4.2.5.1 and 4.2.5.2) indicated that β-pleated sheet is most probably the underlying molecular conformation of a large part of the extended R&R Consensus, especially the part which contains the R&R Consensus itself, 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, a distant (20%) sequence similarity was found between RR-1-bearing cuticular proteins and the crystallographically determined C-terminal, β-sheet barrel portion, of bovine plasma retinol-binding protein (RBP). When, following alignment, both conservative substitutions and identities were combined, the similarity rises to 60% of the total HCCP12 sequence (Hamodrakas *et al.*, 2002). This similarity allowed the construction, by "homology" modeling, of a structural model of the "extended R&R consensus" (Hamodrakas *et al.*, 2002). This modeling was successful even though it is clear that RBP and the R&R Consensus-bearing cuticular

proteins are not strictly homologous, for HCCP12 lacks the N-terminal region that is conserved in members of the lipocalin superfamily to which RBP belongs. The original model (**Figure 5a**) comprises the C-terminal 66 residues (out of 89 in total) of HCCP12 and corresponds to the "extended R&R consensus" (see Section 4.2.3.2.3).

Does this model fit both major classes of RR proteins? Stereo plots of this model of HCCP12 (**Figure 4a**) can be compared to comparable models of two RR-2 proteins (HCCP66 and AGCP2b) (**Figure 4b** and **c**). These models demonstrated that the extended R&R Consensus of both "soft" and "hard" cuticle proteins may easily adopt the proposed conformation.

How would this proposed structure interact with chitin? A low-resolution docking experiment 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, rather comfortably, at least one extended chitin chain (**Figure 4d**) (Hamodrakas *et al.*, 2002). The features revealed by secondary structure predictions (see Section 4.2.5.1) and by experimental spectroscopic analysis (see Section 4.2.5.2) work exceedingly well with this model. It 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 (seven-residue) two-turn α-helix at the C-terminus of the extended R&R Consensus of HCCP12, starting and ending with two proline residues, present in 60% of the "soft" cuticle proteins in **Figure 1**. 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). The structures of these two different chitin-binding proteins cannot be superimposed, however, and show no sequence similarity.

For this review, the proposed half-barrel model (see **Chapter 4.8**). has been used as a basis for more detailed docking experiments. A "high-resolution" docking experiment with the same tetramer was performed and the results are displayed and discussed in **Figure 5**. A new possibility emerges from this "high-resolution" experiment: the chitin chains can run either parallel to the β-strands of the half β-barrel model (**Figure 5b** and **c**), in good agreement with the observations of Atkins (1985), or perpendicular to the β-strands (**Figure 5a**). **Figure 5b** and **c** also provide an instructive view as to how a twisted helicoidal structure might arise from a close packing interaction of half-β-barrel models of cuticle

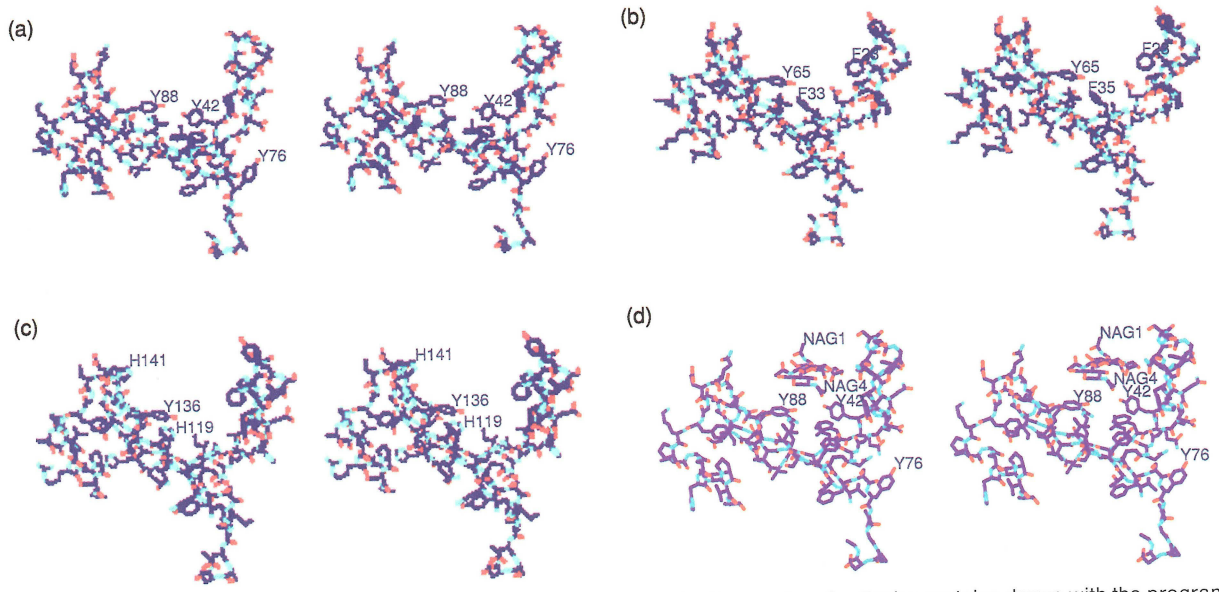

Figure 4 Stero pairs of cuticular proteins and their interaction with chitin. Stero pairs of cuticular proteins drawn with the program O (Jones *et al.*, 1991). The numbering scheme used is that of the unprocessed proteins. (a) View of the "soft" cuticle protein HCCP12. (b) View of the "hard" cuticle protein HCCP66. The terminal Ile83 residue could not be modeled and is not shown. (c) View of "hard" cuticle protein AGCP2b. His140 and the terminal Val155 residues could not be modeled and are not shown. (d) HCCP12 shown 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. (Reprinted with permission from 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, © Elsevier.)

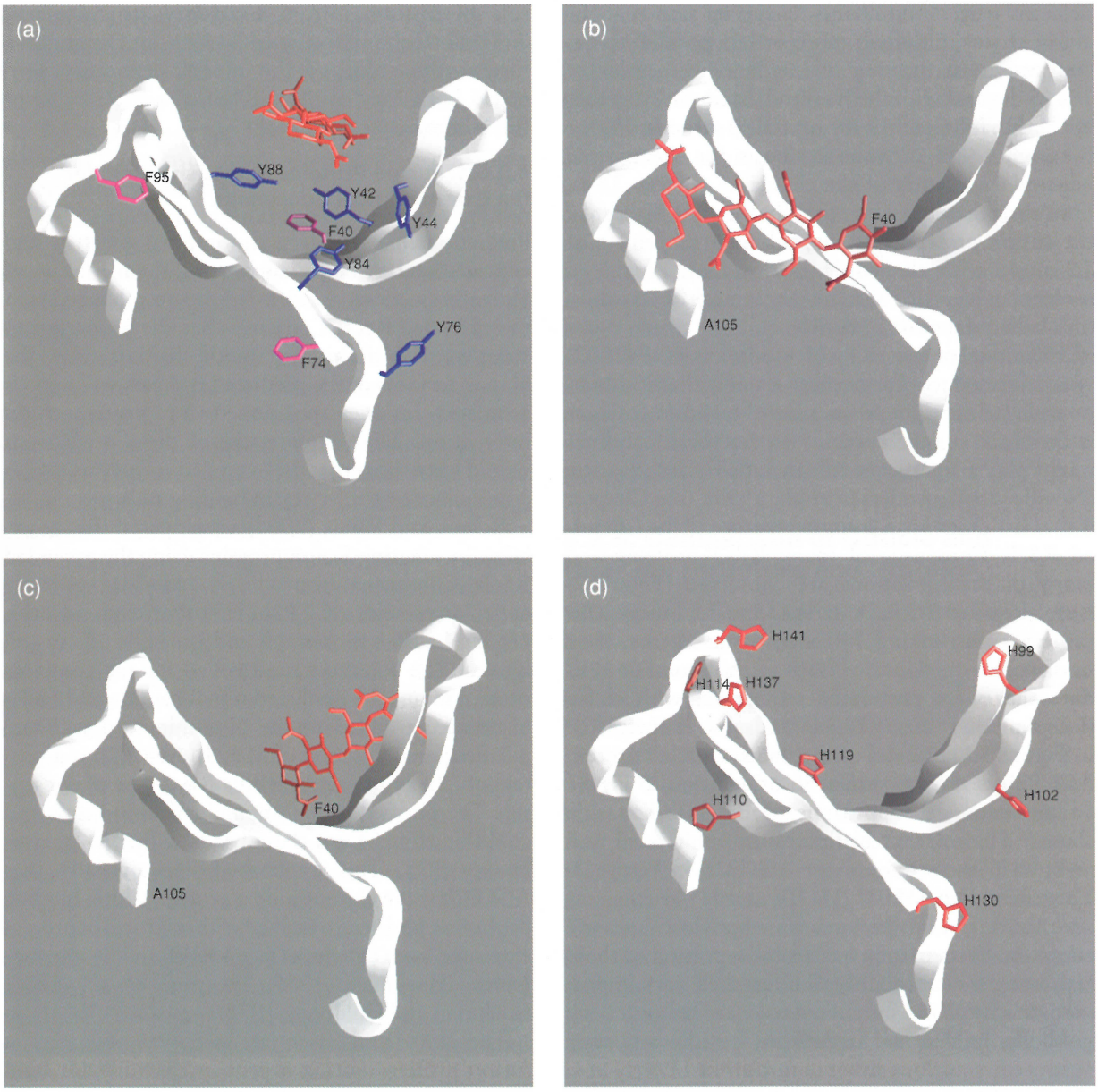

Figure 5 Ribbon models of cuticular proteins derived from homology modeling. (a) A ribbon model of cuticle protein structure, displayed using GRASP (Nicholls *et al.*, 1991). The structure of the representative "soft" cuticle protein HCCP12 was modeled on that of bovine retinol binding protein (RBP; 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. These are: F40, Y42, Y44, H52, F74, Y76, Y84, Y88, and F95, underlined and bolded in **Figure 1**. 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 (a view similar to **Figure 4d**). (b) and (c) Two more possible complexes of HCCP12 with an 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, both in (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 "hard cuticle" 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. The corresponding residues are underlined in the sequence for AGCP2b in **Figure 2**.

proteins with chitin chains. It can be seen that the chitin chains, although more or less parallel to the β-strands, are forming an angle of the order of 10–15 degrees with the β-strands, and this, together with the inherent twist of the β-strands in the β-barrel, could provide the basis for the twisted helicoidal structure of the cuticle in general. Thus, both the inherent twist of the half-barrel β-sheet of the cuticle proteins and its packing arrangement at an angle with the chitin chains may provide a molecular basis for the morphological observation of a helicoidal twist in cuticle.

The model proposed by Hamodrakas *et al.* (2002) was subjected to a further test, namely that it should provide for the right positioning of histidine residues in the "hard cuticle" proteins, so that these histidines might play a significant role in cuticle sclerotization (Neville, 1975; Andersen *et al.*, 1995a (see **Chapter 4.4**). Histidines are a common feature of the extended R&R Consensus of many RR-2 proteins and many of their positions are conserved (**Figure 2**) (see Section 4.2.3.2.3). If they are to function in protein cross-linking by sclerotizing agents, they must reside on opposite faces to the aromatic residues that were postulated to interact with chitin (Iconomidou *et al.*, 1999; Hamodrakas *et al.*, 2002). In **Figure 5d**, a model of the "hard cuticle" protein AGCP2b is shown, similar in orientation to **Figure 5a**, indicating the positions of the histidine side chains. The relevant histidines are underlined and in bold in the sequence for AGCP2b in **Figure 2**. Three histidines, H102, H110, and H141 (the second, third, and last) are at sites where histidines are common. Such an interspecies conservation of these histidines, most probably signifies their very important structural and functional role (see below).

All the bolded and underlined histidines occupy "exposed" positions either in turns (like H99, H114, H137, H141), or at the "edges" of the half-β-barrel or its periphery (like H102, H110, H130), in excellent positions to be involved in cuticular sclerotization, readily reacting with activated *N*-acetyldopamine residues. 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.*, 1995a).

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 model previously proposed both for "soft" and "hard" cuticle proteins (Hamodrakas *et al.*, 2002).

4.2.5.4. 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 4.2.5.1–4.2.5.3) 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 *A. gambiae* putative cuticular protein, AGCP2b. First they expressed this protein in *Escherichia 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 (Dotson *et al.*, 1998). AGCP2b is a protein of 222 amino acids, with a RR-2 type of consensus. The purified protein bound to chitin beads and could be eluted from these beads with 8 M 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, as shown in **Figure 6**.

The 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,

APANYEFSYSVHDEH**T**G**D**IKSQHETRHGDEVH*GQYSLLDSDGHQRIVD***Y***HADHHTG***F***NAVVRREP*

Figure 6 The pfam00379 region of AGcP26 used to construct fusion proteins. For details, see text.

40 amino acids (underlined in **Figure 6**) with the strict R&R Consensus (shown in italics) did not bind chitin. Nor did the full construct when either the Y and F (bolded and highlighted) of the strict R&R Consensus or the T and D (highlighted) of the extended consensus were "mutated" to alanine (Rebers and Willis, 2001).

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.

Chitinase, some lectins, and proteins from peritrophic membranes all bind chitin (review: Shen and Jacobs-Lorena, 1999). What is unique about the extended R&R Consensus is that it lacks 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. Rebers and Willis (2001) suggested that this conservation (see **Figures 1** and **2**) could well be due to the need to preserve a precise conformation of the chitin-binding domain in the absence of stabilizing disulfide bonds.

In addition to establishing a function of the extended R&R Consensus, these experiments also provided confirmation of key elements in the models discussed above (see Section 4.2.5.4). Substitution of the two conserved aromatic residues 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 5a**), 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 5d**). 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.

4.2.5.5. Summary

Four different types of data have been presented above (see Section 4.2.5) analyzing the extended R&R Consensus: secondary structure predictions of antiparallel β-sheets (see Section 4.2.5.1), experimental spectroscopic evidence from cuticles and cuticle extracts for the predominance of such β-sheets in cuticular protein conformation (see Section 4.2.5.2), models showing organization of the consensus into a half β-barrel with a groove that can accommodate chitin (see Section 4.2.5.3), and direct demonstration that the extended consensus is sufficient to confer chitin binding on a protein (see Section 4.2.5.4). 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.

4.2.6. Comparison of Cuticle and Chorion: Structure and Proteins

Silkmoth and fish chorions (eggshells) and cuticle are known to have a helicoidal architecture (Neville, 1975; Hamodrakas, 1992). Excellent reviews on helicoidal architecture and its appearance in biological systems have been made by Bouligand (1972, 1978a, 1978b) and Neville (1975, 1981, 1986). These works describe, in a beautiful and most comprehensive way, how helicoids are identified, how widespread they are, and the basic molecular principles of their formation as well as their geometrical, physical, and biological properties. The close analogy between the helicoidal structures of (usually extracellular) biological materials and the structure of cholesteric liquid crystals suggests that these structures self-assemble according to a mechanism that is very similar to the process allowing materials to form liquid crystals. Apparently, helicoids should pass through a liquid crystalline phase before solidifying. It is assumed that this occurs in the assembly zone during cuticle formation. Self-assembling systems are important in biology, as they are economical in energy terms, requiring neither enzymatic control nor the expenditure of energy-rich bonds. They are particularly appropriate for building extracellular skeletal structures outside of the cells that secrete the components (Bouligand, 1978a, 1978b; Neville, 1986).

Silkmoth chorion is produced by the follicular cells that surround the oocyte (Regier and Kafatos, 1985 and references therein). Fish eggshell is mainly produced by the oocyte, with minor contributions from the follicular cells (Hamodrakas, 1992 and references therein) and cuticle is produced by the epidermis.

Natural helicoidal composites occur in several combinations such as polysaccharide fibers in a polysaccharide matrix (plant cell walls), polysaccharide fibers in a protein matrix (arthropod cuticle), and protein fibers in a protein matrix (insect and fish eggshells). In all cases, principles of molecular recognition and weak intermolecular interactions should govern the self-assembly mechanisms (Neville, 1986).

In silkmoth chorion, disulfide bonds, and in fish eggshell, isopeptide bonds between the side chains of R-K and D-E, are major contributors to stabilization. These covalent bonds, however, are totally absent in cuticle, where stabilization occurs via protein–chitin interaction and by cross-linking by sclerotization compounds (see **Chapter 4.4**).

It is clear that the main characteristic of chorion proteins is the presence of exact, tandemly repeating hexapeptide motifs that adopt a characteristic antiparallel β-pleated sheet structure. This is the main structural unit of silkmoth chorion fibrils and, apparently, the molecular denominator, which dictates formation of the helicoidal architecture (Hamodrakas, 1992). The ellipsoidal shape of silkmoth chorions is, most probably, due to the fact that the basic building-blocks, chorion protein fibrils, are so uniform in shape. By contrast, in cuticle, despite the fact that there are regions of the molecules rich in tandem repeats of certain motifs (see Section 4.2.3.2.2), the sequences are mainly characteristic of globular proteins, and cuticle may adopt all sorts of shapes depending on the local needs of the arthropods producing it. The majority of cuticular proteins contain a conserved domain, rich in a characteristic antiparallel β-pleated sheet structure, a half β-barrel (see Section 4.2.5.3) which again should serve as the molecular denominator determining the helicoidal structure of cuticle, interacting with chitin crystalline chains and giving rise to a plethora of architectural plans as needed locally.

Apparently, an antiparallel β-pleated sheet type of structure is the common molecular denominator, that dictates the helicoidal architecture adopted by the chorion of Lepidoptera and fish and also by the arthropod cuticle.

4.2.7. Summary and Future Challenges

This review has summarized the wealth of information about cuticular proteins amassed since Silvert's review in 1985. Most striking is that the 35-fold increase in sequences for structural cuticular proteins has revealed that the majority has a conserved domain (pfam00379) that is an extended version of the R&R Consensus. A group of proteins that appears to contribute to hard cuticles have a highly conserved extended consensus (RR-2). It is now known that RR-2 proteins interact with chitin and we can predict in some detail the features of their sequence that confer this property. It is not known whether the RR-1 proteins are as effective in binding chitin. We have not yet begun to analyze how the regions outside the consensus contribute to cuticular properties, nor have we learned how the proteins lacking the consensus but with other conserved features contribute to cuticle structure.

Cuticular proteins with pfam00379 are one of the largest multigene families found in *Drosophila* (Lespinet *et al.*, 2002). We need to learn 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. 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, 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.

A wealth of information is available already 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.

Acknowledgments

We appreciate the helpful comments of John Rebers, Augustine Dunn, and John S. Willis.

References

Agrawal, O.P., Scheller, K., **1986**. The formation of the chitin-arylphorin complex *in vitro*. In: Muzzarelli, R., Jeuniaux, C., Gooday, G.W. (Eds.), Chitin in Nature and Technology. Plenum Press, New York, pp. 316–320.

Andersen, S.O., **1975**. Cuticular sclerotization in the beetles *Pachynoda epphipiata* and *Tenebrio molitor*. *J. Insect Physiol. 21*, 1225–1232.

Andersen, S.O., **1998**. Amino acid sequence studies on endocuticular proteins form 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 cranifer*. *Insect Biochem. Mol. Biol. 30*, 569–577.

Andersen, S.O., Hojrup, P., Roepstorff, P., **1995a**. Insect cuticular proteins. *Insect Biochem. Mol. Biol. 25*, 153–176.

Andersen, S.O., Rafn, K., Krogh, T.N., Hojrup, P., Roepstorff, P., **1995b**. Comparison of larval and pupal cuticular proteins in *Tenebrio molitor*. *Insect Biochem. Mol. Biol. 25*, 177–187.

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.

Andersen, S.O., Weis-Fogh, T., **1964**. Resilin: a rubber-like protein in arthropod cuticle. *Adv. Insect Physiol. 2*, 1–65.

Apple, R.T., Fristrom, J.W., **1991**. 20-Hydroxyecdysone is required for, and negatively regulates, transcription of *Drosophila* pupal cuticle protein genes. *Devel. Biol. 146*, 569–582.

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. Int. Symp. Biomol. Struct. Interactions, Suppl. J. Biosci. 8*, 375–387.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., *et al.*, **2002**. The Pfam protein families database. *Nucleic Acids Res. 30*, 276–280.

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 silkmoth *Hyalophora cecropia*. *Insect Biochem. Mol. Biol. 24*, 989–1000.

BLAST, **2003**. http://www.ncbi.nlm.nih.gov/BLAST/

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 Mol. 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. *Devel. Biol. 149*, 112–122.

Bouligand, Y., **1972**. Twisted fibrous arrangements in biological materials and cholesteric mesophases. *Tissue and Cell 4*, 189–217.

Bouligand, Y., **1978a**. Cholesteric order in biopolymers. *ACS Symposium Series 74*, 237–247.

Bouligand, Y., **1978b**. Liquid crystalline order in biological materials. In: Blumstein, A. (Ed.), Liquid Crystalline Order in Polymers. Academic Press, London, pp. 261–297.

Browder, L.W., Wilkes, J., Rodenhiser, D.I., **1992**. Preparative labeling of proteins with [^{35}S]methionine. *Anal. Biochem. 204*, 85–89.

Carter, D., Locke, M., **1993**. Why caterpillars do not grow short and fat. *Int. J. Insect Morphol. Embryol. 22*, 81–102.

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, *Calpodes ethlius* Stoll (Lepidoptera, Hesperiidae). *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.

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.

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.

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.

ENTREZ, **2003**. http://www3.ncbi.nlm.nih.gov/Entrez/index.html

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.

Flybase, **2003**. Genetic data on *Drosophila*. http://flybase.bio.indiana.edu/

Fraenkel, G., Rudall, K.M., **1947**. The structure of insect cuticles. *Proc. Roy. Soc. B 134*, 111–143.

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, D., Fristrom, J.W., **1993**. The metamorphic development of the adult epidermis. In: Bate, M., Martinez Arias, A. (Eds.), The Development of *Drosophila melanogaster*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 843–897.

Fristrom, J.W., Hill, R.J., Watt, F., **1978**. The procuticle of *Drosophila*: heterogeneity of urea-soluble proteins. *Biochemistry 17*, 3917–3924.

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 Devel. 1*, 996–1004.

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. *Int. 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.

Hackman, R.H., **1955**. Studies on chitin. 3. 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.

Hamodrakas, S.J., **1992**. Molecular architecture of helicoidal proteinaceous eggshells. In: Case, S.T. (Ed.), Results and Problems in Cell Differentiation, vol. 19. Springer-Verlag, Berlin, pp. 116–186.

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., *et al.*, **2002**. Identification of *Drosophila melanogaster* yellow-f and yellow-f2 proteins as dopachrome-conversion enzymes. *Biochem. J. 368*, 333–340.

Henikoff, S., Eghtedarzadeh, M.K., **1987**. Conserved arrangement of nested genes at the *Drosophila Gart* locus. *Genetics 117*, 711–725.

Henikoff, S., Keene, M.A., Fechtel, 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 A 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. *Devel. Biol. 132*, 292–303.

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., **1999**. Is β-pleated sheet the molecular conformation which dictates formation of helicoidal cuticle? Insect Biochem. *Mol. Biol. 29*, 285–292.

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.

Janssens, H., Gehring, W.J., **1999**. Isolation and characterization of *drosocrystallin*, a lens crystallin gene of *Drosophila melanogaster*. *Devel. Biol. 207*, 204–214.

Jones, T.A., Zou, J.Y., Cowan, S.W., Kjeldgaard, M., **1991**. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A47*, 110–119.

Kalinich, J.F., McClain, D.E., **1992**. An *in vitro* method for radiolabeling proteins with ^{35}S. *Anal. Biochem. 205*, 208–212.

Kaufman, T.C., Seeger, M.A., Olsen, G., **1990**. Molecular and genetic organization of the Antennapedia gene complex of *Drosophila melanogaster*. *Adv. Genet. 27*, 309–362.

Kollberg, U., Obermaier, B., Hirsch, H., Kelber, G., Wobert, P., **1995**. Expression cloning and characterization of a pupal cuticle protein cDNA of *Galleria mellonella* L. Insect Biochem. *Mol. Biol. 25*, 355–363.

Konig, M., Agrawal, O.P., Schenkel, H., Scheller, K., **1986**. Incorporation of calliphorin into the cuticle of the developing blowfly, *Calliphora vicinia*. *Roux's Arch. Devel. 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.

Lampe, D.J., Willis, J.H., **1994**. Characterization of a cDNA and gene encoding a cuticular protein from rigid cuticles of the giant silkmoth, *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., Delachambre, J., **1986**. A water-soluble protein specific to the adult cuticle in *Tenebrio*: its use as a marker of a new programme expressed by epidermal cells. *Insect Biochem. 16*, 483–489.

Lemoine, A., Millot, C., Curie, G., Delachambre, J., **1989**. A monoclonal antibody against an adult-specific cuticular protein of *Tenebrio molitor* (Insecta, Coleoptera). *Devel. 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. Devel. 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 insecticyanin in the tobacco hornworm, *Manduca sexta*. *Eur. J. Biochem. 205*, 491–499.

Locke, M., **1998**. Epidermis. In: Harrison, F.W., Locke, M. (Eds.), Microscopic Anatomy of Invertebrates, vol. 11A. Wiley–Liss, New York, pp. 75–138.

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.

Lombardi, E.C., Kaplan, D.L., **1993**. Preliminary characterization of resilin isolated from the cockroach, *Periplanata americana*. *Mat. Res. Soc. Proc. 292*, 3–7.

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, Hesperridae). *J. Insect Physiol. 45*, 861–870.

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.

McGeoch, D.J., **1985**. On the predictive recognition of signal peptide sequences. *Virus Res. 3*, 271–286.

Mita, K., Morimyo, M., Okano, K., Shimada, T., Maeda, S., **1999**. The construction of EST database for genome analysis of *Bombyx mori*. *RIKEN Rev. 22*, 63–67.

Molnar, K., Borhegyi, N.H., Csikos, G., Sass, M., **2001**. The immunoprotein scolexin and its synthesizing sites: the midgut epithelium and the epidermis. *Acta Biol. Hungarica 52*, 473–484.

MOTIF, **2003**. http://motif.genome.ad.jp/MOTIF2.html

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.

Nash, W.G., **1976**. Patterns of pigmentation color states regulated by the *y* locus in *Drosophila melanogaster*. *Devel. Biol. 48*, 336–343.

Neville, A.C., **1975**. Biology of the Arthropod Cuticle. Springer-Verlag, Berlin.

Neville, A.C., **1981**. Cholesteric proteins. *Mol. Cryst. Liq. Cryst. 76*, 279–286.

Neville, A.C., **1986**. The physics of helicoids: multidirectional "plywood" structures in biological systems. *Phys. Bull. 37*, 74–76.

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.

Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne, G., **1997**. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng. 10*, 1–6.

Nielsen, H., Krogh, A., **1998**. Prediction of signal peptides and signal anchors by a hidden Markov model. *Proc. Int. Conf. Intell. Syst. Mol. Biol. 6*, 122–130.

Palli, S.R., Locke, M., **1987**. The synthesis of hemolymph proteins by the larval epidermis of an insect *Calpodes ethlius* (Lepidoptera: Hesperiidae). *Insect Biochem. 17*, 711–722.

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.

Pfam, **2003a**. http://www.sanger.ac.uk/Software/Pfam/

Pfam, **2003b**. http://pfam.wustl.edu/cgi-bin/getdesc?acc=PF00379

Psort, **2003**. http://psort.ims.u-tokyo.ac.jp/

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.

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.

Regier, J.C., Kafatos, F.C., **1985**. Molecular aspects of chorion formation. In: Kerkut, G.A., Gilbert, L.I. (Eds.), Comprehensive Insect Physiology, Biochemistry, and Pharmacology, vol. 1. Pergamon Press, Oxford, pp. 113–151.

Riddiford, L.M., **1994**. Cellular and molecular actions of juvenile hormone I: general considerations and premetamorphic 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., **1980a**. The cuticular proteins of *Tenebrio molitor*. 1. Electrophoretic banding patterns during postembryonic development. *Devel. Biol. 75*, 59–69.

Roberts, P.E., Willis, J.H., **1980b**. Effects of juvenile hormone, ecdysterone, actinomycin D, and mitomycin C on the cuticular proteins of *Tenebrio molitor*. *J. Embryol. Exp. Morph. 56*, 107–123.

Rondot, I., Quennedey, B., Courrent, A., Lemoine, A., Delachambre, J., **1996**. Cloning and sequencing of a cDNA encoding a larval–pupal-specific cuticular protein in *Tenebrio molitor* (Insecta, Coleoptera): developmental expression and effect of a juvenile hormone analogue. *Eur. J. Biochem. 235*, 138–143.

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. *Devel. Biol. 107*, 420–431.

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 vicinia*. *Z. Naturforsch. 35c*, 387–389.

Scholer, H.R., **1991**. Octamania: the POU factors in murine development. *Trends Genet. 7*, 323–328.

Shen, Z., Jacobs-Lorena, M., **1999**. Evolution of chitin-binding proteins in invertebrates. *J. Mol. Evol. 48*, 341–347.

SilkBase, **2003**. Expressed sequence tags in *Bombyx mori*. http://www.ab.a.u-tokyo.ac.jp/silkbase

Silvert, D.J., **1985**. Cuticular proteins during postembryonic development. In: Kerkut, G.A., Gilbert, L.I. (Eds.), Comprehensive Insect Physiology, Biochemistry, and Pharmacology, vol. 2. Pergamon Press, Oxford, pp. 239–254.

Snyder, M., Hunkapiller, M., Yuen, D., Silvert, D., Fristrom, J., *et al.*, **1982**. Cuticle protein genes of *Drosophila*: structure, organization and evolution of four clustered genes. *Cell 29*, 1027–1040.

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.

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.

Trim, A.R.H., **1941**. Studies in the chemistry of insect cuticle. 1. 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., *et al.*, **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.

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: Locke, M., Smith, D.S. (Eds.), Insect Biology in the Future. Academic Press, London, pp. 369–383.

Willis, J.H., **1986**. The paradigm of stage-specific gene sets in insect metamorphosis: time for revision! *Arch. Insect Biochem. Physiol.* (Suppl. 1), 47–57.

Willis, J.H., **1996**. Metamorphosis of the cuticle, its proteins, and their genes. In: Gilbert, L.I., Atkinson, B.G., Tata, J. (Eds.), Metamorphosis/Post-Embryonic Reprogramming of Gene Expression in Amphibian and Insect Cells. Academic Press, London, pp. 253–282.

Willis, J.H., **1999**. Cuticular proteins in insects and crustaceans. *Amr. Zool. 39*, 600–609.

Willis, J.H., Rezaur, R., Sehnal, F., **1982**. Juvenoids cause some insects to form composite cuticles. *J. Embryol. Exp. Morph. 71*, 25–40.

Wittkopp, P.J., Vaccaro, K., Carroll, S.B., **2002**. Evolution of *yellow* gene regulation and pigmentation in *Drosophila*. *Curr. Biol. 12*, 1547–1556.

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.

Wolfgang, W.J., Riddiford, L.M., **1986**. Larval cuticular morphogenesis in the tobacco hornworm, *Manduca sexta*, and its hormonal regulation. *Devel. Biol. 113*, 305–316.

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., *et al.*, **1994**. Crystallographic studies on complexes between retinoids and plasma retinol-binding protein. *J. Biol. Chem. 269*, 29613–29620.

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.

Relevant Website

http://bioinformatics2.biol.uoa.gr – A relational database of arthropod cuticular proteins established by C.K. Magkrioti, I.C. Spyropoulos, V.A. Iconomidou, J.H. Willis, and S.J. Hamodrakas.