Journal of Insect Physiology 56 (2010) 1420–1426

Contents lists available at ScienceDirect

Journal of Insect Physiology

journal homepage: www.elsevier.com/locate/jinsphys

A possible structural model of members of the CPF family of cuticular proteins implicating binding to components other than chitin

Nikos C. Papandreou [a], Vassiliki A. Iconomidou [a], Judith H. Willis [b], Stavros J. Hamodrakas [a,*]

[a] Department of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Panepistimiopolis 157 01, Athens, Greece
[b] Department of Cellular Biology, University of Georgia, Athens, GA 30602, USA

ARTICLE INFO

Article history:
Received 25 January 2010
Received in revised form 5 April 2010
Accepted 5 April 2010

Keywords:
Cuticular proteins structural model
Lipocalins
Eight-stranded up and down antiparallel β-pleated sheet barrel
Insect sex pheromones
Cuticular lipids

ABSTRACT

The physical properties of cuticle are determined by the structure of its two major components, cuticular proteins (CPs) and chitin, and, also, by their interactions.

A common consensus region (extended R&R Consensus) found in the majority of cuticular proteins, the CPRs, binds to chitin. Previous work established that β-pleated sheet predominates in the Consensus region and we proposed that it is responsible for the formation of helicoidal cuticle. Remote sequence similarity between CPRs and a lipocalin, bovine plasma retinol binding protein (RBP), led us to suggest an antiparallel β-sheet half-barrel structure as the basic folding motif of the R&R Consensus. There are several other families of cuticular proteins. One of the best defined is CPF. Its four members in *Anopheles gambiae* are expressed during the early stages of either pharate pupal or pharate adult development, suggesting that the proteins contribute to the outer regions of the cuticle, the epi- and/or exo-cuticle. These proteins did not bind to chitin in the same assay used successfully for CPRs. Although CPFs are distinct in sequence from CPRs, the same lipocalin could also be used to derive homology models for one *A. gambiae* and one *Drosophila melanogaster* CPF. For the CPFs, the basic folding motif predicted is an eight-stranded, antiparallel β-sheet, full-barrel structure. Possible implications of this structure are discussed and docking experiments were carried out with one possible *Drosophila* ligand, 7(Z),11(Z)-heptacosadiene.

1. Introduction

Cuticle is a complex, bipartite, composite material that provides structural and mechanical support acting functionally as both skin and skeleton to arthropods (Neville, 1975; Vincent and Wegst, 2004). It is composed of chitin crystallites embedded in a proteinaceous matrix (Neville, 1975). However, the interaction of cuticular proteins with chitin crystallites and the detailed structure of insect cuticle have not yet been resolved.

Detailed analyses have revealed a wide variety of arthropod cuticular proteins (CPs) within and amongst species, which can be classified into almost a dozen families (Willis, 2010). Cuticular proteins exhibit certain sequence motifs that are present even in proteins from distantly related species and such conserved motifs have common and important roles for the proper function of cuticle (Andersen et al., 1995). These sequence motifs are utilized for the classification of cuticular proteins into families. The first identified motif was the "R&R consensus sequence" (Rebers and Riddiford, 1988). It has subsequently been extended and modifications recognized (Iconomidou et al., 1999; Willis et al., 2005; Willis, 2010). Experimental procedures have shown that cuticular proteins, which exhibit the "extended R&R consensus sequence" and that region alone have chitin-binding properties (Rebers and Willis, 2001; Togawa et al., 2004).

Secondary structure prediction and experimental data indicate that β-pleated sheet is most probably the molecular conformation of a large part of the extended R&R Consensus, especially the part that contains the R&R Consensus itself (Iconomidou et al., 1999, 2001). It was also proposed that this conformation is most probably involved in β-sheet-chitin crystallite interactions (Iconomidou et al., 1999, 2001). This proposal and experimental results are in agreement with earlier experimental findings and proposals that β-sheet should be involved in chitin–protein interactions (Fraenkel and Rudall, 1947; Atkins, 1985).

Unexpectedly, distant sequence similarities of the extended consensus from several CPR proteins were found with bovine plasma retinol binding protein, RBP (Hamodrakas et al., 2002). RBP (Protein Data Bank Accession Code 1FEN; Berman et al., 2000) belongs to the lipocalin class of proteins. Lipocalins are small, secreted proteins (160–200 residues). While sequence similarity among family members is low (frequently <20%), they have in common a tertiary structure, which consists of 8 strands forming

* Corresponding author. Tel.: +30 210 7274931; fax: +30 210 7274254.
E-mail address: shamodr@biol.uoa.gr (S.J. Hamodrakas).

0022-1910/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jinsphys.2010.04.002

an anti-parallel up and down β-barrel (Adam et al., 2008). A 3_{10} helix caps one extremity of the barrel (H1) and a second α-helix is parallel to its surface (H2). The interior of the cavity is large enough to permit the binding of a small, typically hydrophobic, molecule. Lipocalins adopt high binding specificity to their respective ligands. Lipocalins can also bind to receptors and participate in macromolecular complexes. They are involved in numerous functions such as the transport of molecules, implicated in development and physiology (e.g. retinoids, arachidonic acid), enzymatic synthesis, immunomodulation, olfaction, pheromone signaling and cell regulation (Flower, 1996).

Retinol binding protein has a β-sheet barrel as its basic structural motif (Zanotti et al., 1994). A large part of this β-sheet barrel is the part similar in sequence to the "extended R&R consensus" (Iconomidou et al., 1999) sequence of CPR proteins. Based on the sequence similarity of RBP with the "extended R&R consensus" region of several distinct CPR proteins, structural models were constructed by homology modeling, (Hamodrakas et al., 2002; Iconomidou et al., 2005; Willis et al., 2005). These models have several attractive features to serve as a chitin-binding structural motif in cuticle and to provide the basis for elucidating cuticle's overall architecture in detail.

Numerous cuticular proteins have been identified that lack the "R&R consensus sequence", and they appear to fall into several distinct protein families. These families are discussed in detail in Willis (2010). One of the best characterized is the CPF gene family first identified in *Tenebrio molitor* (Andersen et al., 1997). While this family was originally described with a 55-amino acid motif, as more sequences were analyzed that motif was shortened to about 44 aa (Togawa et al., 2007), so the name of the family could remain CPF. This analysis revealed another conserved motif at the C-terminus. Four proteins were identified in *Anopheles gambiae* with a domain that allowed them to be placed in the CPF family (Togawa et al., 2007). In *A. gambiae* the four CPFs genes are expressed just before pupal or adult ecdysis. If appearance of the mRNA is followed immediately by translation, then it is likely that these proteins contribute to the outer layers of pupal or adult cuticle, i.e. the epi- and/or exo-cuticle. Another clue to their location, within the cuticle, came from the finding that recombinant CPF proteins do not bind to chitin. Of course, absence of chitin binding in these assays does not rule out that these proteins might bind to chitin when they are in their native state and natural environment. Nonetheless, since chitin is not found in epi-cuticle (see Willis et al., 2005 for discussion of this point), the CPFs have a property appropriate for the epi-cuticle.

In this work, we propose an antiparallel β-sheet, up and down, full-barrel structure as the basic folding motif of these CPF proteins, having also noticed remote similarities to the lipocalins and having predicted their secondary structure, which shows abundant β-sheet structure (data not shown). Possible important implications of this structural model to the function of CPFs as possible repositories of sex-pheromones (Greenspan and Ferveur, 2000; Hall, 1994) or cuticular lipids are also discussed. Docking experiments were carried out with 7(Z),11(Z)-heptacosadiene, a known sex pheromone of *Drosophila melanogaster* (Antony et al., 1985), in order to learn if the predicted structures of the CPFs were compatible with such binding.

2. Materials and methods

The employment of the prediction algorithm PHD (Rost, 1996) on several CPR cuticular proteins revealed significant structural similarity of these proteins to the crystallographically solved structure of bovine plasma retinol binding protein (RBP, Protein Data Bank Accession Code 1FEN; Berman et al., 2000), which belongs to the class of lipocalins. The structure of bovine RBP was utilized as template for comparative modeling studies of two representatives of the CPF family of cuticular proteins. The sequences that were modeled were AgamCPF3 from *A. gambiae* (AGAP004690; ENTREZ accession number 118790289) (Togawa et al., 2007) and CG8541 from *D. melanogaster* (ENTREZ accession number 7295230). CG8541 had been identified as a CPF by Togawa et al. (2007) and is the top BLAST (Altschul et al., 1997) match to AgamCPF3. Pairwise alignments of these sequences with bovine RBP were produced with CLUSTALW (Thompson et al., 1994). The similarity matrix and the parameters used were the default values of CLUSTALW. The sequences had their signal peptide removed according to information provided from the SignalP3.0 server (Emanuelsson et al., 2007). Comparative modeling experiments were carried out with Modeller v9.2 (Sali and Blundell, 1993) based on the pairwise alignments and using as template the solved structure of retinol binding protein (Zanotti et al., 1994). Docking experiments were performed, separately, utilizing the program GRAMM (Vakser, 1996), with target protein structures the models of AgamCPF3 and DmelCG8541 utilizing as ligand 7(Z),11(Z)-heptacosadiene (*cis*, *cis*, 7,11-heptacosadiene; Fig. 1), the predominant female-specific sex pheromone of the fruit fly *D. melanogaster* (Antony et al., 1985). The coordinates of this pheromone were retrieved from Pherobase (El-Sayed, 2008). In addition, the software Autodock4.2 (Morris et al., 2009) was also employed for docking calculations of the models of AgamCPF3 and DmelCG8541 with heptacosadiene. The procedure was carried out considering the ligand both flexible (all rotational bonds were set free) and, also, rigid. In both cases, the protein side chains remain fixed. The AutodockTools program was utilized to generate the input files for docking. In both docking experiments, a grid map of dimensions $126 \times 126 \times 126$ points in *x*, *y* and *z* dimensions with a grid-spacing of 0.375 Å was built to cover the CPF3 molecule. Ten runs were generated by using Lamarckian genetic algorithm searches. The settings applied, were the default of the program, with an initial population of 150 randomly placed individuals, a maximum number of 2.5×10^6 energy evaluations, and a maximum number of 2.7×10^4 generations. A mutation rate of 0.02 and a crossover rate of 0.8 were chosen. Results differing by <2 Å in a positional root mean square deviation (rmsd) were clustered together.

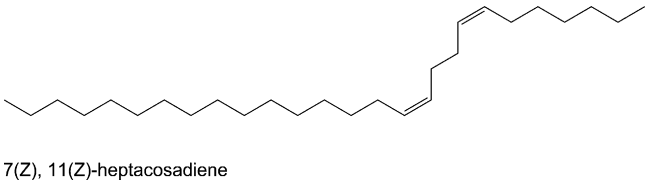

Fig. 1. The chemical structure of the *Drosophila melanogaster* female sex pheromone 7(Z),11(Z)-heptacosadiene (*cis*, *cis* 7,11-heptacosadiene).

3. Results

Pairwise alignments were performed utilizing CLUSTALW (Thompson et al., 1994) and involved CPF3 from *A. gambiae*, a representative of the CPF family (Fig. 2) and CG8541 (Fig. 3), a homologous sequence to CPF3 from *D. melanogaster* respectively (see Section 2), against bovine RBP. Sequence identity is 13% for the entire CPF3 of *A. gambiae* (121 residues of the entire secreted protein). Taking into account conservative substitutions sequence similarity approaches 50% (61 out of 121 residues). In the case of CG8541, sequence identity to RBP is 13%. This identity corresponds to 190 out of 257 residues of the entire secreted protein. The remaining 67 residues involve the N′- and C′-terminals that cannot be aligned due to the greater sequence length of CG8541 compared to that of RBP. Conservative substitutions represent 47% (89 out of 190 residues) of the CG8541 sequence in the region that could be aligned.

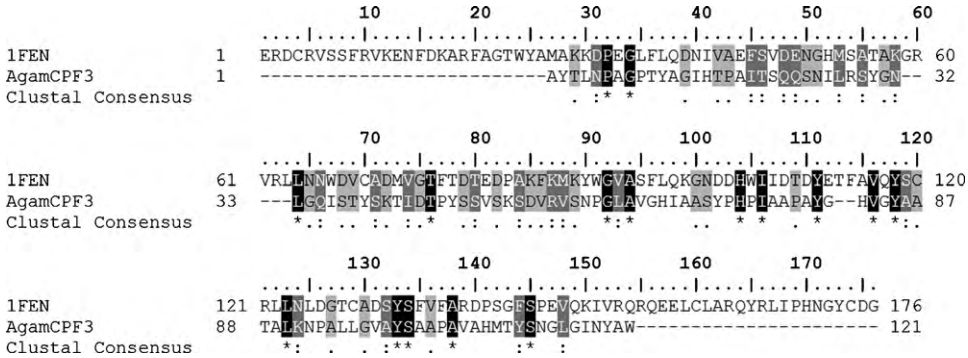

Fig. 2. CLUSTALW alignment of Retinol binding protein (PDB code 1FEN) with cuticular protein AgamCPF3 (*A. gambiae* CPF3) (Togawa et al., 2007). The sequences of both proteins are given in the one letter code. Numbering for CPF3 is that of the mature protein (16-residue signal peptide removed; http://biophysics.biol.uoa.gr/cuticleDB, Magkrioti et al., 2004). Black-boxed residues are identical and light and dark gray-boxed residues represent conservative substitutions at the 60% and 80% significance levels, respectively. These similarities are indicated on the bottom lines by (*) and (:) respectively.

Although sequence identity is rather low (below 20%) in all the above mentioned comparisons, levels of sequence similarity (ca. 50% conservative substitutions) are adequate for the construction of structural models of these proteins by comparative modeling. Utilizing as a template the experimentally determined structure of bovine RBP (Zanotti et al., 1994), we constructed models of AgamCPF3 and DmelCG8541, utilizing Modeller (Sali and Blundell, 1993). The model of AgamCPF3 is presented in Fig. 4. It comprises all the residues of the entire secreted protein, while the model of CG8541 (Fig. 5) comprises 190 out of 257 residues of the entire secreted protein. The derived models indicate that the basic folding motif of CPFs is most probably an eight-stranded, antiparallel, up and down, β-sheet full-barrel structure.

The models of AgamCPF3 and DmelCG8541 shown in Figs. 4 and 5, respectively, also show the location where 7(Z),11(Z)-heptacosadiene would complex with the respective proteins. They are the most favourable complexes, which were derived from a "high resolution" docking experiment utilizing the docking program GRAMM (Vakser, 1996). On both complexes, 7(Z),11(Z)-heptacosadiene fits rather comfortably in the interior of the β-barrels almost parallel to the barrel axis. It must be noted that both docking experiments provided a variety of different configurations of 7(Z),11(Z)-heptacosadiene relative to the interior of the barrels of the proposed models of AgamCPF3 and CG8541 that were thermodynamically favored, thus, indicating several favorable possible modes of interaction of 7(Z),11(Z)-heptacosadiene with both proteins. Additional docking experiments of 7(Z),11(Z)-heptacosadiene to AgamCPF3 and DmelCG8541, performed utilizing the highly popular (Sousa et al., 2006) software Autock4.2 (Morris et al., 2009), also indicate that the highly flexible heptacosadiene ligand, fits comfortably in the interior of the β-barrel "pocket" of AgamCPF3 and DmelCG8541, almost parallel to the barrel axis. This occurs considering both the ligand as rigid (in its minimum energy conformation) (Fig. 6a and b), or allowing the ligand to be flexible (all rotational bonds were set free) (Fig. 7a and b). Interestingly, this software indicates another possible binding site of the hydrophobic ligand heptacosadiene, to the outer surface of AgamCPF3 and this site coincides with the characteristic "outer hydrophobic cluster" (see Section 4) of the possible lipocalin-like AgamCPF3 (Fig. 8).

4. Discussion

Despite the evidence that the cuticular proteins bearing the R&R consensus, that is the CPR family, bind to chitin, the precise nature of the interactions of insect cuticular proteins and chitin is still,

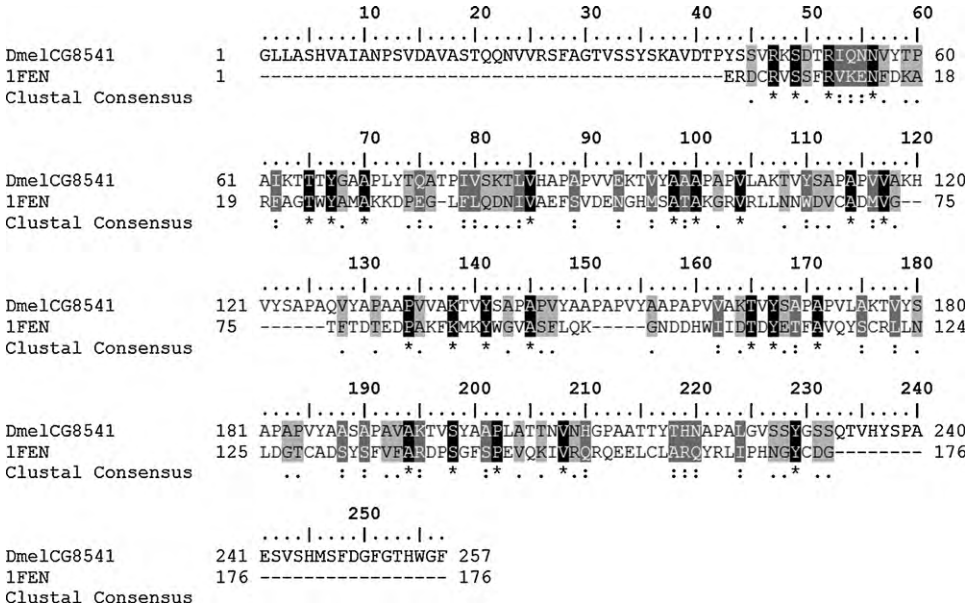

Fig. 3. CLUSTALW alignment of cuticular protein DmelCG8541 (*D. melanogaster* orthologue of CPF3) with retinol binding protein (PDB code 1FEN). The sequences of both proteins are given in the one letter code. Numbering for DmelCG8541 is that of the mature protein (18-residue signal peptide removed). Shading and symbols as in Fig. 2.

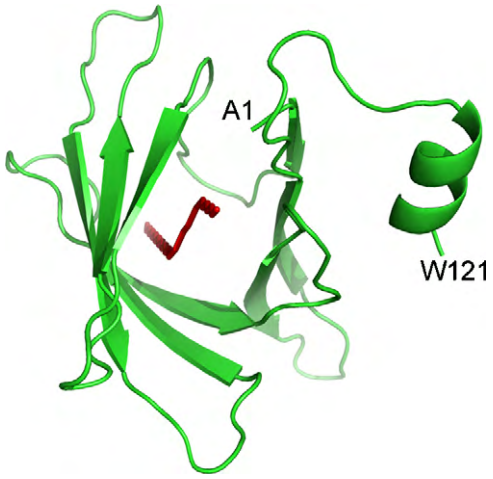

Fig. 4. A ribbon model of the cuticular protein AgamCPF3 structure (green), displayed using the software PyMOL (Delano, 2005). It was modelled on that of bovine retinol binding protein (RBP; PDB code 1FEN, Zanotti et al., 1994) utilizing the software Modeller v9.2 (Sali and Blundell, 1993) and the alignment details shown in Fig. 2. The entire secreted protein, from A1 to W121, is shown in the model. It is complexed with 7(Z),11(Z)-heptacosadiene (Fig. 1), a female sex pheromone of *D. melanogaster*, shown in red. The complex was derived from a "high resolution" docking experiment of 7(Z),11(Z)-heptacosadiene, considered rigid, in its minimum energy conformation (Fig. 1), with the model of AgamCPF3, utilizing the docking software GRAMM (Vakser, 1996) and the default parameters of the software. The 1st best solution is shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

experimentally, unknown. All analyses indicate that the CPR family is by far the most abundant protein family among cuticular proteins (Willis, 2010). An antiparallel β-sheet half-barrel structure has been proposed as the basic folding motif of the "extended R&R Consensus", and theoretical models of chitin–cuticular protein interactions have been proposed (Hamodrakas et al., 2002; Iconomidou et al., 2005). We arrived at this structural model for the CPRs observing the remote similarities of the CPRs to the lipocalins, which were suggested by the PhD software (Rost, 1996). Lipocalins share several common molecular recognition properties: the binding of small, mainly hydrophobic molecules (such as retinol), binding to specific cell-surface receptors and the formation of covalent and non-covalent complexes with other soluble macromolecules. Although they have been classified

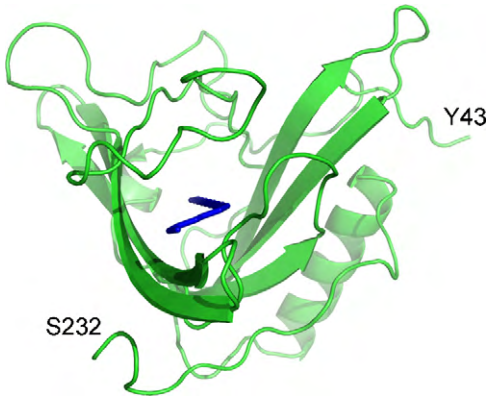

Fig. 5. A ribbon model of the cuticular protein DmelCG8541 structure (green), done as in Fig. 4, with the alignment details shown in Fig. 3. The model comprises 190 of 257 residues of the secreted protein, from Y43 to S232. It is complexed with 7(Z),11(Z)-heptacosadiene (Fig. 1), shown in blue. Details for the derivation of the complex are the same as in Fig. 4. The 1st best solution is shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

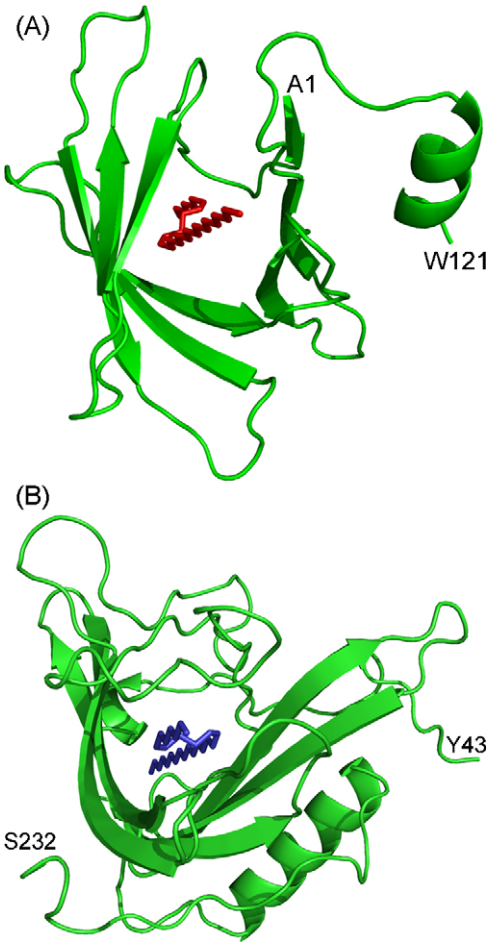

Fig. 6. (a) A ribbon model of the cuticular protein AgamCPF3 structure (green), constructed and displayed as in Fig. 4. The entire secreted protein, from A1 to W121, is shown in the model. It is complexed with 7(Z),11(Z)-heptacosadiene (Fig. 1), shown in red. The complex was derived from a docking experiment of 7(Z),11(Z)-heptacosadiene (Fig. 1), with the model of AgamCPF3, utilizing the docking software Autodock4.2 (Morris et al., 2009). The ligand, heptacosadiene, is inside the "pocket" of the β-barrel of AgamCPF3. The ligand was considered as rigid, in its minimum energy conformation, and other details of the docking procedure are described in detail in "Section 2". The ligand represents a cluster of 4 out of 10 best solutions (runs). (b) A ribbon model of the cuticular protein DmelCG8541 structure (green), constructed and displayed as in Fig. 5. The model comprises 190 of 257 residues of the secreted protein, from Y43 to S232. It is complexed with 7(Z),11(Z)-heptacosadiene (Fig. 1), shown in blue. Details of the docking experiment, which produced this complex, as in Fig. 6a above. The ligand represents a cluster of 7 out of 10 best solutions (runs). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

mainly as transport proteins it is now clear that they have various functions (Flower, 1996). Members of this family have frequently very remote similarities, having pairwise sequence identities very commonly below 20% (Flower et al., 2000). The lipocalin fold is a highly symmetrical all-β structure dominated by a single eight-stranded antiparallel up and down β-sheet barrel (Flower et al., 2000). Lipocalins are characterized by two hydrophobic "clusters" of residues the "inner" and the "outer cluster" (Adam et al., 2008). In Table 1 the two hydrophobic clusters ("inner" and "outer") for a characteristic member of the lipocalin family, retinol binding protein, are shown, together with the "clusters" of the two CPF cuticular proteins modeled, AgamCPF3 and DmelCG8541 and also a CPR cuticular protein HCCP12 (Hamodrakas et al., 2002). Shaded are the conserved hydrophobic residues. It is seen that the conservation is rather high, especially for the "outer" cluster, which strengthens the view that the CPFs may act in a similiar manner to lipocalins.

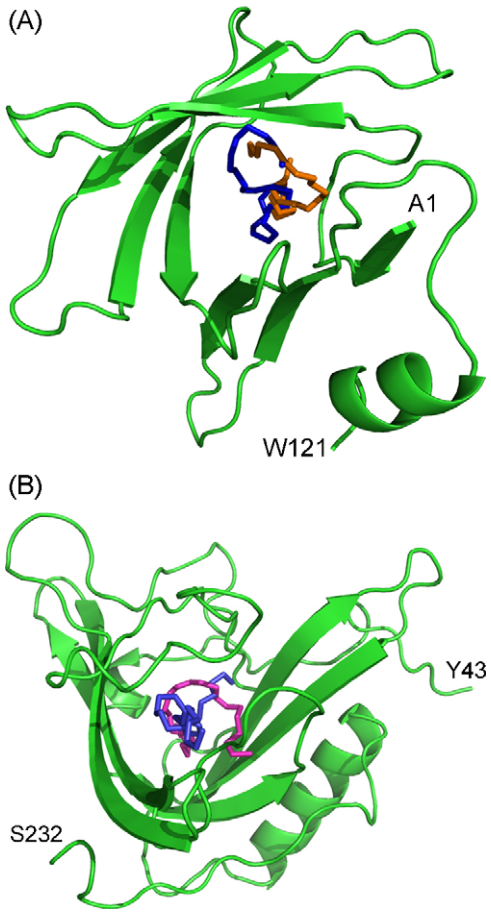

Fig. 7. (a) A ribbon model of the cuticular protein AgamCPF3 structure (green), constructed and displayed as in Fig. 4. The entire secreted protein, from A1 to W121, is shown in the model. It is complexed with 7(Z),11(Z)-heptacosadiene (Fig. 1). The complex was derived from a docking experiment of 7(Z),11(Z)-heptacosadiene (Fig. 1), with the model of AgamCPF3, utilizing the docking software Autodock4.2 (Morris et al., 2009). Two out of 10 best solutions (runs) for the ligand are shown in red and blue, respectively, inside the "pocket" of the β-barrel of AgamCPF3. The remaining 8 solutions also show the ligand to reside inside the "pocket". The heptacosadiene ligand was considered as flexible (all rotable bonds were set free) and other details of the docking procedure are described in detail in "Section 2". (b) A ribbon model of the cuticular protein DmelCG8541 structure (green), constructed and displayed as in Fig. 5. The model comprises 190 of 257 residues of the secreted protein, from Y43 to S232. It is complexed with 7(Z),11(Z)-heptacosadiene (Fig. 1). All other details of the docking experiment, which produced the complex, as in Fig. 7a. Two out of 10 best solutions (runs) for the ligand are shown in magenta and blue, respectively, inside the "pocket" of the β-barrel of DmelCG8541. The remaining 8 solutions also show the ligand to reside inside the "pocket". (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

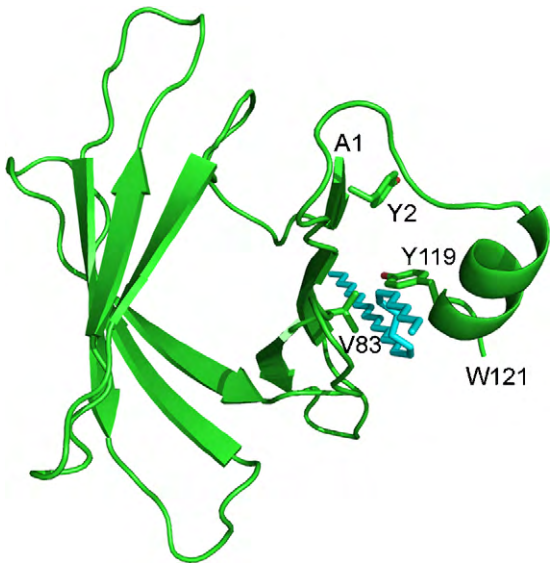

Fig. 8. A ribbon model of the cuticular protein AgamCPF3 structure (green), constructed and displayed as in Fig. 4. The entire secreted protein, from A1 to W121, is shown in the model. It is complexed with 7(Z),11(Z)-heptacosadiene (Fig. 1), shown in cyan. The complex was derived from a docking experiment of 7(Z),11(Z)-heptacosadiene (Fig. 1), with the model of AgamCPF3, utilizing the docking software Autodock4.2 (Morris et al., 2009). It shows the ligand, heptacosadiene, outside the β-barrel of AgamCPF3, in contact with the "hydrophobic outer cluster" (see Table 1). The side chains of three hydrophophic residues of the conserved "hydrophobic outer cluster" Y2, V83 and Y119 are shown. The heptacosadiene ligand was considered as rigid, in its minimum energy conformation, and other details of the docking procedure are described in detail, in "Section 2". The ligand represents the cluster of the remaining 6 out of 10 best solutions (see Fig. 6a). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

We are interested in elucidating the structural motifs and possible functions of cuticular proteins that belong to families where the 'extended R&R consensus' is absent. The first obvious choice was the CPF family of cuticular proteins (Togawa et al., 2007). This family of cuticular proteins is of particular interest because they are expressed just before pupal or adult ecdysis, suggesting that these families are most probably components of the outer layer of pupal and adult cuticles. That is, they are likely located in the epi- or exo-cuticle. Furthermore, although they can be modeled with a binding pocket, they did not to bind chitin in *in vitro* assays, which suggests that they serve other functions. Therefore, the question which arises is what is their functional role? One possible function is that they intercalate among the chitin crystallites and their associated proteins of the procuticle (exo- and endo-cuticle) or they are simply loosely bound to chitin chains. But this does not explain why they should form a binding pocket. Alternatively, if components of the epi-cuticle, they could perhaps bind, having similarities to the lipocalins, to the lipoidal molecules, which are known to act as female sex pheromones in certain insect species (Antony et al., 1985). In our attempts to see whether these cuticular proteins might bind to female sex pheromones it has been shown that this is possible, indeed in more than one possible modes (see Section 3), although not all of them energetically equally stable. These attempts clearly show that especially in *D. melanogaster* where there are known molecules acting as contact sex pheromones, as 7(Z),11(Z)-heptacosadiene, the formation of complexes between CPFs and the pheromone is possible and energetically favorable.

In *A. gambiae*, the molecular nature of sex pheromones remains unknown. However, complex formation between AgamCPF3 and 7(Z),11(Z)-heptacosadiene is also favored. Therefore, if a similar in structure to 7(Z),11(Z)-heptacosadiene sex pheromone exists in *A. gambiae* it could easily bind to AgamCPF3. In this respect is it provocative that Cassone et al. (2008) found that CPF3 may play a role in incipient species (M and S) differentiation in *A. gambiae*. CPF3 mRNA is present in far higher levels in the M than in the S form. Indeed it was the most differentially expressed gene in these microarray analyses of virgin females with differences ranging from 27-fold in laboratory populations, to but 2.5–3.5-fold in three natural populations. The amino acid sequence of the CPF3 protein is identical in the two forms (unpublished observations), so a quantitative difference in protein levels may be important. On the other hand, it is surprising that an epi-cuticular component would continue to be made and secreted into outer regions of the cuticle days after adult eclosion. Nonetheless, evidence for the presence of the CPF3 mRNA in 3-day-old males and non-blood fed females has been found in another microarray analysis and its level falls 5-fold 3 h after the blood meal (Marinotti et al., 2006). An alternative

Table 1
Distantly related lipocalins share two conserved clusters (“inner” and “outer”) of (usually hydrophobic residues Adam et al., 2008). The two “clusters” are shown for bovine retinol binding protein (PDB code: 1FEN), CPR cuticular protein HCCP12 (Hamodrakas et al., 2002 and refs. therein) and two CPF cuticular proteins, AgamCPF3 (Togawa et al., 2007) and its *D. melanogaster* orthologue DmelCG8541. The numbering of HCCP12, AgamCPF3 and DmelCG8541 is that of the mature proteins (16-, 16- and 18-residue signal peptides have been subtracted, respectively). Conserved hydrophobic residues are shaded.

Inner cluster				
Lipocalins (general positions)	1FEN (RBP bovine)	HCCP12	AgamCPF3	DmelCG8541
39	F20	–	–	I62
48	W24	–	–	T66
80	F45	–	I19	P88
91	M53	–	L27	V96
93	A55	T20	S29	A98
115	F77	V37	P46	V128
131	F86	F40	V55	A137
133	M88	Y42	V57	T139
156	H104	L58	H73	P160
158	I106	N60	I75	V162
168	A115	E69	H82	A171
170	Q117	R71	G84	V173
190	F137	Y88	P104	V193
192	R139	A90	V106	K195
Outer cluster				
Lipocalins (general positions)	1FEN (RBP bovine)	HCCP12	AgamCPF3	DmelCG8541
52	A28	–	Y2	A70
159	I107	V61	A76	A163
169	V116	V70	V83	P172
171	Y118	G72	Y85	L174
189	V136	–	A103	A192
205	V152	–	Y119	V208
220	Y165	–	–	A221

occupant of the CPF binding pocket might just be intracuticular lipids that are present throughout the cuticle. Several of these cuticular lipids have chemical structures very similar to the 7(Z),11(Z)-heptacosadiene sex pheromone (Hadley, 1981 and references therein). Therefore, they would fit easily into the pocket of the β-barrel of the CPFs, or bind to their “outer hydrophobic cluster” (see ‘Section 3’ and Figs. 4, 5, 6a, 6b, 7a, 7b, 8).

It should be emphasized that chitin chains do not fit into the pocket of the CPFs. Docking experiments indicate this fact clearly (Fig. 9), in agreement with experimental data (Togawa et al., 2007). They indicate, however, that the CPFs might interact loosely with chitin chains, with their β-strands lying parallel to the chitin chains, in agreement with experimental observations (Atkins, 1985). Further evidence against a role of the CPFs in direct binding to chitin comes from the information in Table 1. Protein–carbohydrate interactions involve aromatic residues (Quiocho, 1989; Vyas and Vyas, 1991; Elgavish and Shaanan, 1997; Hamodrakas et al., 1997; Svitil and Kirchman, 1998; Shen and Jacobs-Lorena, 1999) and in the cleft of the half-barrel model of HCCP12 there are three critical aromatic residues, F40, Y42 and Y88 (Hamodrakas et al., 2002). Comparable residues in the two CPFs, AgamCPF3 and DmelCG8541 are hydrophobic but not aromatic.

Another question that might be posed is: Why have we proposed a half-barrel model for CPRs and a full barrel for CPFs? The CPR model is based on the extended R&R Consensus for that is a region of the proteins that matches closely retinol binding protein. This region, generally is <70 aa, far too short to form a full barrel. On the other hand, the CPF match is far longer, and compatible with an eight-stranded, full barrel.

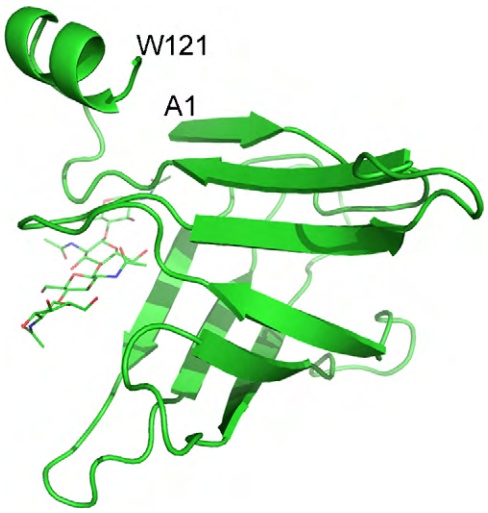

Fig. 9. A complex of AgamCPF3 (ribbon model shown in green) with a NAG tetramer (ball and stick model) in an extended conformation (taken as a chitin analog), derived from a “high resolution” docking experiment, utilizing the docking software GRAMM (Vakser, 1996) and the default parameters of the program, displayed using PyMol (Delano, 2005). The model presented is the “top of the list”, most favorable complex. Note that, the “chitin chain” runs parallel to the β-strands, of at least half of theβ-barrel, in agreement with experimentally derived data (Atkins, 1985). No solution was obtained with the “chitin chain” into the pocket of the β-barrel. The entire secreted protein, from A1 to W121, is shown in the model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

At this point it is essential to acknowledge that models are proposed to suggest further experiments. It is obvious that we cannot simply concentrate on proposing models for cuticular protein interactions and that there is an urgent need for experiments leading to the determination of the structure of cuticular proteins either by X-ray crystallography or NMR spectroscopy, alone and as complexes with oligosaccharide analogues of chitin, or female sex pheromones, or cuticular lipids. The determination of such structures will throw light on the complex structure of cuticle with its extraordinary physical and physiological properties and to protein interactions with chitin and other functional molecules, which play important roles to the structure and physiology of cuticle.

Acknowledgments

We thank the reviewers of this article for their useful criticism. NCP, VAI and SJH thank the University of Athens, Greece, for financial support. JHW was supported by a grant from the U.S. National Institutes of Health (AI55624).

References

Adam, B., Charloteaux, B., Beaufays, J., Vanhamme, L., Godfroid, E., Brasseur, R., Lins, L., 2008. Distantly related lipocalins share two conserved clusters of hydrophobic residues: use in homology modeling. BMC Structural Biology 8, 1.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25, 3389–3402.

Andersen, S.O., Hojrup, P., Roepstorff, P., 1995. Insect cuticular proteins. Insect Biochemistry and Molecular Biology 25, 153–176.

Andersen, S.O., Rafn, K., Roepstorff, P., 1997. Sequence studies of proteins from larval and pupal cuticle of the yellow meal worm, *Tenebrio molitor*. Insect Biochemistry and Molecular Biology 27, 121–131.

Antony, C., Davis, T.L., Carlson, D.A., Pechine, J.M., Jallon, J.M., 1985. Compared behavioral responses of male *Drosophila melanogaster* (Canton S) to natural and synthetic aphrodisiacs. Journal of Chemical Ecology 11, 1617–1629.

Atkins, E., 1985. Conformations in polysaccharides and complex carbohydrates. Proceedings of the International Symposium on Biomolecular Structure and Interactions, Journal of Biosciences 8 (Suppl.), 375–387.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein data bank. Nucleic Acids Research 28, 235–242.

Cassone, B.J., Mouline, K., Hahn, M.W., White, B.J., Pombi, M., Simard, F., Constantini, C., Bezansky, N.J., 2008. Differential gene expression in incipient species of *Anopheles gambiae*. Molecular Ecology 17, 2491–2504.

Delano, W.L., 2005. The PyMOL Molecular Graphics System. DeLano Scientific LLC 400, Oyster Point Blvd., Suite 213, South San Francisco, CA 94080-1918, USA.

Elgavish, S., Shaanan, B., 1997. Lectin–carbohydrate interactions: different folds, common recognition principles. Trends in Biochemical Sciences 22, 462–467.

El-Sayed, A.M., 2008. The Pherobase: Database of Insect Pheromones and Semiochemicals. .

Emanuelsson, O., Brunak, S., von Heijne, G., Nielsen, H., 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nature Protocols 2, 953–971.

Flower, D.R., 1996. The lipocalin protein family: structure and function. Biochemical Journal 318, 1–14.

Flower, D.R., North, A.C.T., Sansom, C.E., 2000. The lipocalin protein family: structural and sequence overview. Biochimica et Biophysica Acta 1482, 9–24.

Fraenkel, G., Rudall, K.M., 1947. The structure of insect cuticles. Proceedings of the Royal Society London B 34, 111–143.

Greenspan, R.J., Ferveur, J.F., 2000. Courtship in *Drosophila*. Annual Review of Genetics 34, 205–232.

Hadley, N.F., 1981. Cuticular lipids of terrestrial plants and arthropods: a comparison of their structure, composition and waterproofing function. Biological Reviews 56, 23–47.

Hall, J.C., 1994. The mating of a fly. Science 264, 1702–1714.

Hamodrakas, S.J., Kanellopoulos, P.N., Pavlou, K., Tucker, P.A., 1997. The crystal structure of the complex of concanavalin A with 4′-methylumbelliferyl-alpha-D-glucopyranoside. Journal of Structural Biology 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 Biochemistry and Molecular Biology 32, 1577–1583.

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 Biochemistry and Molecular Biology 31, 877–885.

Iconomidou, V.A., Willis, J.H., Hamodrakas, S.J., 1999. Is beta-pleated sheet the molecular conformation which dictates formation of helicoidal cuticle? Insect Biochemistry and Molecular Biology 29, 285–292.

Iconomidou, V.A., Willis, J.H., Hamodrakas, S.J., 2005. Unique features of the structural model of 'hard' cuticle proteins: implications for chitin-protein interactions and cross-linking in cuticle. Insect Biochemistry and Molecular Biology 35, 553–560.

Magkrioti, C.K., Spyropoulos, I.C., Iconomidou, V.A., Willis, J.H., Hamodrakas, S.J., 2004. cuticleDB: a relational database of Arthropod cuticular proteins. BMC Bioinformatics 5, 138.

Marinotti, O., Calvo, E., Nguyen, Q.K., Dissanayake, S., Ribeiro, J.M., James, A.A., 2006. Genome-wide analysis of gene expression in adult *Anopheles gambiae*. Insect Molecular Biology 15, 1–12.

Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J., 2009. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. Journal of Computational Chemistry 30, 2785–2791.

Neville, A.C., 1975. Biology of the Arthropod Cuticle. Springer-Verlag, New York.

Quiocho, F.A., 1989. Protein-carbohydrate interactions: basic molecular features. Pure and Applied Chemistry 61, 1293–1306.

Rebers, J.E., Riddiford, L.M., 1988. Structure and expression of a *Manduca sexta* larval cuticle gene homologous to *Drosophila* cuticle genes. Journal of Molecular Biology 203, 411–423.

Rebers, J.E., Willis, J.H., 2001. A conserved domain in arthropod cuticular proteins binds chitin. Insect Biochemistry and Molecular Biology 31, 1083–1093.

Rost, B., 1996. PHD: predicting one-dimensional protein structure by profile-based neural networks. Methods in Enzymology 266, 525–539.

Sali, A., Blundell, T.L., 1993. Comparative protein modeling by satisfaction of spatial restraints. Journal of Molecular Biology 234, 779–815.

Shen, Z., Jacobs-Lorena, M., 1999. Evolution of chitin-binding proteins in invertebrates. Journal of Molecular Evolution 48, 341–347.

Sousa, S.F., Fernandes, P.A., Ramos, M.J., 2006. Protein-ligand docking: current status and future challenges. Proteins: Structure, Function and Bioinformatics 65, 15–26.

Svitil, A.L., Kirchman, D.L., 1998. A chitin-binding domain in a marine bacterial chitinase and other microbial chitinases: implications for the ecology and evolution of 1,4-beta-glycanases. Microbiology 144, 1299–1308.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22, 4673–4680.

Togawa, T., Augustine Dunn, W., Emmons, A.C., Willis, J.H., 2007. CPF and CPFL, two related gene families encoding cuticular proteins of *Anopheles gambiae* and other insects. Insect Biochemistry and Molecular Biology 37, 675–688.

Togawa, T., Nakato, H., Izumi, S., 2004. Analysis of the chitin recognition mechanism of cuticle proteins from the soft cuticle of the silkworm, *Bombyx mori*. Insect Biochemistry and Molecular Biology 34, 1059–1067.

Vakser, I.A., 1996. Low-resolution docking: prediction of complexes for undetermined structures. Biopolymers 39, 455–464.

Vincent, J.F., Wegst, U.G., 2004. Design and mechanical properties of insect cuticle. Arthropod Structure and Development 33, 187–199.

Vyas, N., Vyas, K., 1991. Atomic features of protein–carbohydrate interactions. Current Opinion in Structural Biology 1 (5), 732–740.

Willis, J.H., 2010. Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomic era. Insect Biochemistry and Molecular Biology 40, 189–204.

Willis, J.H., Iconomidou, V.A., Smith, R.F., Hamodrakas, S.J., 2005. Cuticular Proteins. In: Gilbert, L.I., Iatrou, K., Gill, S. (Eds.), Comprehensive Insect Science, vol. 4Elsevier, Oxford, pp. 79–110, chapter 2.

Zanotti, G., Marcello, M., Malpeli, G., Folli, C., Sartori, G., Berni, R., 1994. Crystallographic studies on complexes between retinoids and plasma retinol-binding protein. Journal of Biological Chemistry 269, 29613–29620.