



Pergamon

Insect Biochemistry and Molecular Biology 29 (1999) 285–292

*Insect  
Biochemistry  
and  
Molecular  
Biology*

# Is $\beta$ -pleated sheet the molecular conformation which dictates formation of helicoidal cuticle?

Vassiliki A. Iconomidou<sup>a</sup>, Judith H. Willis<sup>b</sup>, Stavros J. Hamodrakas<sup>a,\*</sup>

<sup>a</sup> Department of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Athens, 157 01, Greece

<sup>b</sup> Department of Cellular Biology, University of Georgia, Athens, GA 30602, USA

Received 21 September 1998; received in revised form 26 December 1998; accepted 6 January 1999

## Abstract

Over 100 sequences for cuticular proteins are now available, but there have been no formal analyses of how these sequences might contribute to the helicoidal architecture of cuticle or to the interaction of these proteins with chitin. A secondary structure prediction scheme (Hamodrakas, S.J., 1988. A protein secondary structure prediction scheme for the IBM PC and compatibles. CABIOS 4, 473–477) that combines six different algorithms predicting  $\alpha$ -helix,  $\beta$ -strands and  $\beta$ -turn/loops/coil has been used to predict the secondary structure of chorion proteins and experimental confirmation has established its utility (Hamodrakas, S.J., 1992. Molecular architecture of helicoidal proteinaceous eggshells. In: Case, S.T. (Ed.), Results and Problems in Cell Differentiation, Vol. 19, Berlin–Heidelberg, Springer Verlag, pp. 116–186 and references therein). We have used this same scheme with eight cuticular protein sequences associated with hard cuticles and nineteen from soft cuticles. Secondary structure predictions were restricted to a conserved 68 amino acid region that begins with a preponderance of hydrophilic residues and ends with a 33 amino acid consensus region first identified by Rebers and Riddiford (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). Both classes of sequences showed a preponderance of  $\beta$ -pleated sheet, with four distinct strands in the proteins from ‘hard’ cuticles and three from ‘soft’. In both cases, tyrosine and phenylalanine were found on one face within a sheet, an optimal location for interaction with chitin. We propose that this  $\beta$ -sheet dictates formation of helicoidal cuticle. © 1999 Elsevier Science Ltd. All rights reserved.

**Keywords:**  $\beta$ -sheet; Chitin; Cuticular proteins; Helicoids; Secondary structure prediction

## 1. Introduction

Several extracellular fibrous structures are known to have helicoidal architecture. Such structures include arthropod cuticle, vertebrate tendons, plant cell walls etc. The widespread occurrence of the helicoidal structure in spherical shells, such as eggshells, spore walls, cyst walls and others and its correlation with the mechanical strength it provides is intriguing.

Excellent reviews on the helicoidal architecture and its appearance in biological systems have been made by Bouligand (1972, 1978a,b, 1978) and Neville (1975, 1981, 1986). These works describe, in a beautiful and most comprehensive way, how helicoids are identified, how widespread they are, the basic molecular principles

of their formation and 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 several organelles and extracellular products are self-assembled 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. Therefore, it is important to determine in such cases the molecular mechanisms of self-assembly.

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 the cells which make them (Bouligand, 1978a,b; Neville, 1986).

Natural helicoidal composites occur in several combi-

\* Corresponding author. Tel.: +30 1 727 4545; fax: +30 1 723 1634.  
E-mail address: shamodr@cc.uoa.gr (S.J. Hamodrakas)

nations: polysaccharide fibers in a polysaccharide matrix (plant cell walls), polysaccharide fibers in a protein matrix (arthropod cuticle), protein fibers in a protein matrix (insect and fish eggshells) to mention just a few examples. In all cases, principles of molecular recognition should govern the self-assembly mechanisms (Neville, 1986).

The chorion (eggshell) of several Lepidoptera and fishes has been shown to be an example of a helicoidal composite of protein fibers in a protein matrix (Kafatos et al., 1977; Mazur et al., 1982). A secondary structure prediction scheme (Hamodrakas, 1988) was used to estimate the structure of proteins from different moth chorion protein families (Hamodrakas et al., 1982b). In all cases, there was a central domain consisting of a  $\beta$ -pleated sheet, predicted to form the basic structural unit in all chorion protein families. Experimental studies on intact chorions utilising laser-Raman and infrared spectroscopy as well as X-ray diffraction have confirmed the preponderance of  $\beta$ -pleated sheet in silkworm chorion proteins (Hamodrakas et al., 1982a; Hamodrakas et al., 1983). Thus, a common molecular denominator, the  $\beta$ -pleated sheet, apparently, dictates the self-assembly process in helicoidal proteinaceous eggshells, biological analogues of cholesteric liquid crystals (for a review see Hamodrakas, 1992).

Insect exo- and endo-cuticle also exhibit a helicoidal architecture (Neville, 1975). In contrast to the situation with chorion, that has only protein-protein interactions, chitin filaments are embedded in the proteinaceous matrix of cuticle. Also, cysteine, a major contributor to chorion stabilization via disulfide bonds, is absent from almost all insect cuticular proteins. Like chorion, dozens of different proteins can be extracted from insect cuticle, but, there is no sequence similarity between chorion and cuticular proteins. To date, there has been no systematic analysis of the secondary structure of these proteins, nor any substantial suggestion as to how they interact with chitin.

Sequences are now available for over 100 cuticular proteins from insects of five different orders. Several of the protein sequences in the databases are very similar. Some of these relatives appear to represent allelic forms, others have been shown to correspond to closely linked genes. Excluding duplicates with over 90% identity, there are about 80 unique sequences known.

Over half of the unique sequences bear a motif similar to that first identified by Rebers and Riddiford (1988) in seven cuticular proteins: G-x(8)-G-x(6)-Y-x-A-x-E-x-G-Y-x(7)-P-x(2)-P (where x represents any amino acid, and the value in parentheses indicates the number of residues). Additional sequences revealed minor differences and new conserved residues in this R&R Consensus. A modification of this consensus: G-x(7)-[DEN]-G-x(6)-[FY]-x-A-[DGN]-x(2,3)-G-[FY]-x-[AP]-x(6) is present in about 40 unique insect cuticular proteins

(Willis, 1999). None of these are identical in the consensus region. The R&R Consensus has also been identified in cuticular proteins from an arachnid and a crustacean (Norup et al., 1996; Kragh et al., 1997; Nousiainen et al., 1998), but these sequences are somewhat different (Willis, 1999) and are not included in our analysis.

Rebers and Riddiford (1988) suggested that the consensus would turn out to be a region of structural importance. Subsequently, Andersen et al. (1995) suggested that the motif might be involved in protein/chitin interaction. The modified R&R Consensus has been found in proteins associated with both hard (rigid) and soft (flexible) cuticles. Association is based on the isolation of the protein from that type of cuticle or the presence of its mRNA in the underlying epidermis. Nearest neighbor and parsimony analyses on the modified R & R Consensus region separated proteins from the two types of cuticle (Willis, 1999). We have not considered whether the protein is from endo- or exo-cuticle, another distinction that may be important in structural/functional relations (Baernholdt and Andersen, 1998).

The region N-terminal to the R&R Consensus is enriched in hydrophilic amino acids (Lampe and Willis, 1994; Andersen et al., 1995). In total, a stretch of approximately 68 amino acids appears to be conserved. We will refer to this as the 'extended R&R Consensus'.

Unlike the situation in chorion, where the  $\beta$ -pleated sheet is found in a central, evolutionarily conserved, domain (although there are indications that it might extend to more variable 'arms' of the proteins as well, Hamodrakas, 1992), the position of the extended R&R Consensus is variable. It is found at the carboxyl terminus of some cuticular proteins, in others it lies in the center of the molecule.

In this report, we present data that suggest that  $\beta$ -pleated sheet is most probably, the underlying molecular conformation of a large part of this extended R&R Consensus, especially the part which contains the R&R Consensus itself. We also propose that this conformation is essential in defining cuticle's helicoidal architecture, and most probably is involved in specific interactions of these proteins with the chitin filaments. Thus, perhaps, a universal architectural plan for helicoidal extracellular structures exists, based primarily on the packing interactions of  $\beta$ -pleated sheets.

## 2. Materials and methods

### 2.1. Secondary structure prediction

A secondary structure prediction scheme has been developed for the IBM-PC and compatibles (Hamodrakas, 1988), containing computer programs, making use of the protein secondary structure prediction algorithms of Nagano (1977a,b), Garnier et al. (1978),

Burgess et al. (1974), Chou and Fasman (1974a,b), Lim (1974a,b) and Dufton and Hider (1977). The results of the individual prediction methods are combined as described by Hamodrakas (1988) and Hamodrakas et al. (1982b), to produce joint prediction histograms for a protein, for three types of secondary structure:  $\alpha$ -helix,  $\beta$ -sheet and  $\beta$ -turns. The scheme requires uniform input for the prediction programs, produced by any word processor, spreadsheet, editor or database program and produces uniform output on a printer, a graphics screen or a file. The scheme is independent of any additional software and runs under DOS 2.0 or later releases. It has been tried successfully in a number of cases on fibrous proteins producing results in good agreement with experimental data, where individual prediction methods usually fail to produce meaningful results (Hamodrakas et al., 1982a; Hamodrakas and Kafatos, 1984; Hamodrakas, 1992).

## 2.2. Cuticular proteins

The 27 proteins chosen for inclusion in this analysis are a subset of 36 unique insect cuticular proteins with the extended R&R Consensus. They represent proteins from cuticles of different metamorphic stages and four different orders. A nearest neighbor cladogram based on the 32-33 amino acid R&R Consensus region separated the sequences into two groups corresponding to hard and soft cuticles. It also revealed that several proteins were very similar (Willis, 1999). For this analysis only one among two close ‘relatives’ was included. Three proteins that had an ambiguous placement on the cladogram and belonged to unknown or intermediate cuticle types were also excluded prior to analysis.

The sequences (extended R&R Consensus region) of the following proteins were used. ENTREZ accession numbers, and hence access to the original reference, are given in parentheses after each name. Some of the cuticular protein names include a designation of L(larva) or A(adult) after the initials for genus or species. Other proteins have no such designation but are known to be specific for a single stage. Others, such as MSCP14.6 and HCCP12 have been associated with cuticles of all three metamorphic stages.

### ‘Hard insect cuticle’:

*Anopheles gambiae*: AGCP1 (1245432), AGCP2b (2961110)

*Drosophila melanogaster*: EDG84 (477929)

*Hyalophora cecropia*: HCCP66 (1169133)

*Locusta migratoria*: LMCP7 (1345862), LMCP8 (117622)

*Tenebrio molitor*: TMCPA1A (1706191), TMCP20 (102879)

### ‘Soft insect cuticle’:

*Bombyx mori*: BMLCP17 (2204069), BMLCP22 (2204071)

*Drosophila melanogaster*: DMLCP1 (85223), DMLCP3 (85225), DMACP65A (1857602), DMLCP65a (1857600), DMLCP65c (1857593), DMLCP65d (1857495), DMLCP65e (1857604), DMLCP65f (1857606), DMCP78 (477634), DMCPGART (72265)

*Drosophila miranda*: DMiLCP1 (1707433)

*Hyalophora cecropia*: HCCP12 (1169129)

*Locusta migratoria*: LMCP4 (461860)

*Lucilia cuprinia*: LCCP1 (2565392), LCCP12 (2565394)

*Manduca sexta*: MSCP14.6 (1666245), MSLCP14 (84824)

## 3. Results

### 3.1. Sequence comparisons

The 67-68 amino acids of the extended R&R Consensus region were aligned as described in the legend to Fig. 1. One gap was inserted to preserve alignment of the R&R Consensus. It is apparent that this entire region from ‘hard’ cuticles is conserved. Ten residues N-terminal to the consensus are invariant, most others represent conserved substitutions. Within the 32-33 amino acids of the original consensus, there are 15 additional identities (Fig. 1, top). Seven additional, non-allelic, sequences can be assigned to the ‘hard’ group (Regier and Willis, unpublished observations). These sequences, from *Anopheles*, *Bombyx*, *Locusta* and *Tenebrio*, also have the same 25 identities, bringing the number of sequences in this ‘hard’ group to 15. The cuticular proteins from ‘soft’ cuticle are more variable, with only 6 identities all within the consensus itself. Conservative substitutions are abundant, however (Fig. 1, bottom). Clearly, proteins from the ‘hard’ class are more conserved, both within the strict consensus and in its N-terminal extension. The residues that are invariant in the ‘hard’ class are frequently quite diverse in the ‘soft’ class. The sequence conservation we identified is impressive, especially considering that proteins come from four different orders including both hemimetabolous and holometabolous insects. It is thus of great interest to learn if the secondary structure of these proteins is also conserved.

### 3.2. Secondary structure prediction

For each protein, individual predictions of  $\alpha$ -helix,  $\beta$ -pleated sheet and  $\beta$ -turns/coil/loops were made, according to the methods of Nagano, Garnier et al., Burgess et al., Chou and Fasman, Lim ( $\alpha$ -helix and  $\beta$ -pleated sheet only) and Dufton and Hider. Joint prediction histograms were then constructed, since they are more dependable



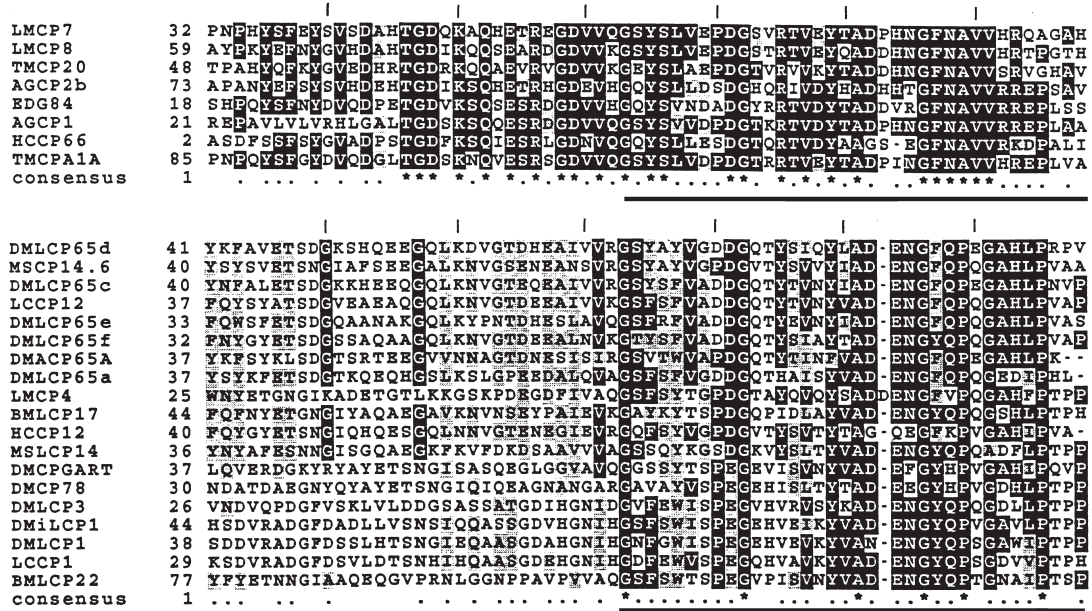


Fig. 1. Alignment of the extended R & R Consensus region (see Materials and methods) of several cuticular proteins. Top, Proteins from hard (rigid) cuticles, and, bottom, from soft (flexible) cuticles. See Materials and methods for abbreviations and ENTREZ accession numbers. Numbers to the left of each sequence are the residue number from the mature peptide of the first amino acid in the alignment; every tenth residue of marked at the top of each set. Signal cleavage sites were obtained from Andersen et al. (1995) or deduced according to the method of von Heijne (1987). Black-boxed residues are identical and gray-boxed residues represent conservative substitutions. Alignments were created with CLUSTAL W (Thompson et al., 1994) and shading was done with BOXSHADE 3.33 (Unix version). The extended R & R Consensus is shown at the bottom with (.) indicating conserved substitutions and (\*) indicating identities; the strict R & R Consensus is underlined.

than individual prediction schemes (Hamodrakas, 1988; Hamodrakas et al., 1982b and references therein; Hamodrakas and Kafatos, 1984). An average secondary structure prediction was made for the sequences of proteins from hard (Fig. 2A) and soft (Fig. 2B) cuticles. To accomplish this, all of the joint prediction histograms for each set were aligned and then the average value for each position was determined (rounded off to the nearest integer). The alignment was based on Fig. 1 and only involved inserting a single gap to keep the R&R consensus in register. The carboxy terminus of three of the proteins from ‘soft’ cuticles occurred one or two residues before the end of the extended consensus. A structure predicted by three or more methods (out of six, or out of five in the case of  $\beta$ -turns/loops) was considered probable and is shaded in Fig. 2. Below the joint prediction histograms, the sequences of *H. cecropia* cuticular protein 66 (HCCP66) (Fig. 2A) and *M. sexta* cuticular protein 14.6 (MSCP14.6) (Fig. 2B) are shown for direct comparison, as are the conserved residues belonging to the extended R&R Consensus.

Corresponding parts of the predicted secondary structure of each protein (by comparison with Fig. 1) and of the R&R Consensus can be identified by nominal residue position (68 nominal aligned positions in total). Thus, the R&R Consensus begins at nominal position 33 in families of cuticular proteins from both ‘hard’ and ‘soft’ cuticles.

The results of prediction of another, rather popular,

prediction algorithm, PHD, (Rost and Sander, 1993; Rost and Sander, 1994; Rost, 1996) on all (eight and nineteen respectively) cuticular proteins are also displayed below the joint prediction histograms, for comparison. The symbol E, represents predicted secondary structure of  $\beta$ -pleated sheet in this case, whereas gaps correspond to random coil or  $\beta$ -turns/loops. This algorithm claims to achieve prediction of protein secondary structure at better than 70% accuracy (Rost and Sander, 1993).

Our prediction package, clearly indicates that the extended R&R domain of cuticular proteins has a considerable proportion of  $\beta$ -pleated sheet structure and total absence of  $\alpha$ -helix (Figs. 2A and 2B). An  $\alpha$ -helix predicted near the end of the hydrophilic region, in the hard cuticle proteins (nominal positions 58–63 with sequence NAVVRK in protein HCCP66), is not considered as probable because there is a simultaneous and much stronger prediction for a  $\beta$ -strand in the region containing residues 58–62. We should note that the prediction algorithm PHD predicts also a  $\beta$ -strand in this region (Fig. 2A).

For the proteins derived from ‘hard’ insect cuticle, it is obvious from Fig. 2A that, in the region between nominal positions 20 and 68, containing the R&R consensus, four  $\beta$ -strands are predicted alternating with  $\beta$ -turns/loops. The four  $\beta$ -strands are predicted at nominal positions 30–32, 34–38, 44–49, 57–62 (relevant amino acid sequence of these segments for protein HCCP66: NVQ, QYSLL, QRTVDY and FNAVVR respectively).

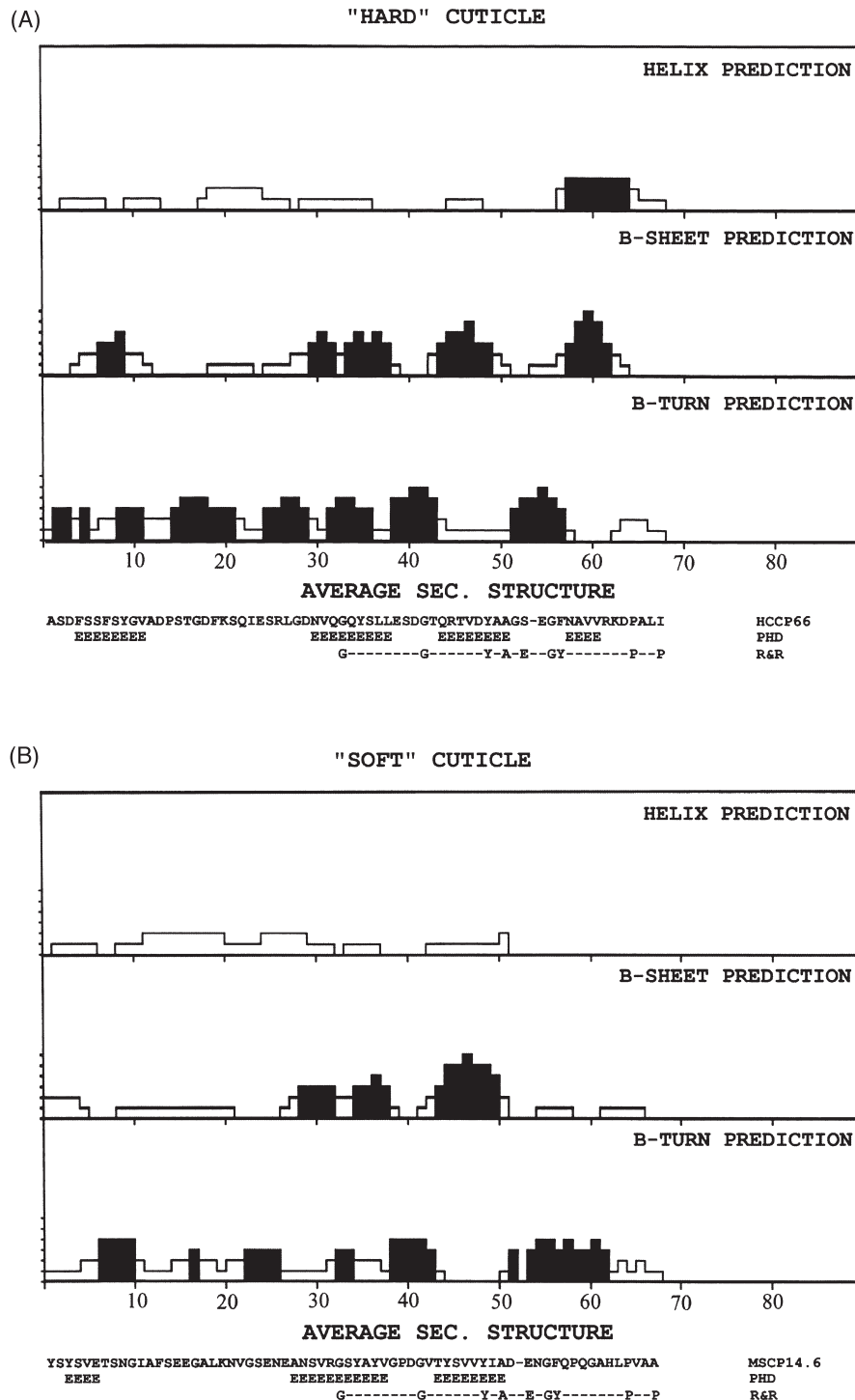


Fig. 2. Average secondary structure prediction plots for  $\alpha$ -helix,  $\beta$ -pleated sheet and  $\beta$ -turns/loops for the hydrophilic domains of the cuticular proteins from hard (top) and soft (bottom) cuticle of Fig. 1. Individual predictions were derived according to Nagano (1977a,b), Garnier et al. (1978), Burgess et al. (1974), Chou and Fasman (1974a,b), Lim (1974a,b) and Dufton and Hider (1977) (see Materials and methods) for each protein separately. Average joint prediction histograms for all cuticular proteins from hard and soft cuticle were constructed by averaging the scores for each residue from the individual predictions. The most probable structures predicted by three or more methods are shaded. The numbering corresponds to the 'consensus' of Figs. 1. The sequences of the extended consensus of *Hyalophora cecropia* cuticular protein 66 (HCCP66) and of *Manduca sexta* cuticular protein 14.6 (MSCP14.6) (see Fig. 1) are shown for direct comparison in parts (A) and (B), below the joint prediction histograms. R & R positions a slightly modified Rebers and Riddiford (1988) consensus. Predicted secondary structure according to a popular prediction algorithm, PHD, (Rost and Sander, 1993; Rost and Sander, 1994; Rost, 1996) on the eight cuticular proteins from hard cuticle (Figs. 1, top) and on the nineteen cuticular proteins from soft cuticle (Fig. 1, bottom) are also displayed below the joint prediction histograms, respectively, for comparison. The symbol E, represents predicted secondary structure of  $\beta$ -pleated sheet in this case, whereas gaps correspond to random coil or  $\beta$ -turns/loops. The plots are 'hard' copies, on a laser printer, of a monitor screen.

The four turn/loop regions are predicted at 25–29, 32–36, 39–43, 52–56 (sequence for protein HCCP66: SRLGD, QGQYS, ESDGT, GS-EG respectively).

In contrast, for the proteins derived from soft insect cuticle, which constitutes a far larger family, Fig. 2B shows that, in the region between nominal positions 20 and 68, containing the R&R consensus, three  $\beta$ -strands are predicted alternating with  $\beta$ -turns/loops. The three  $\beta$ -strands are predicted at nominal positions 29–32, 35–38 and 44–50 (relevant amino acid sequence of these segments for protein MSCP14.6: NSVR, YAYV, and TYSVVYI respectively). The four turn/loop regions are predicted at 23–26, 33–34, 39–43, 52–62 (sequence for protein MSCP14.6: GSEN, GS, GPDGV, D-ENGFQPQGA respectively).

Apparently, the cuticular proteins from ‘soft’ cuticles which lack the HxGFNAV V (residues 54–61) sequence that characterizes most of the proteins from ‘rigid’ cuticles (Fig. 1A), lack also the  $\beta$ -strand that is predicted to occur in this region.

Therefore, it appears that this region, which contains the R&R consensus, has a rather well defined secondary structure both in ‘hard’ and ‘soft’ cuticular proteins. Clearly, the primary structure invariance of this domain leads also to its structural invariance through evolution (that of a  $\beta$ -pleated sheet), indicating an important functional role. However, the proteins from ‘rigid’ cuticles appear to have a larger  $\beta$ -pleated sheet containing four  $\beta$ -strands alternating with turns/loops than those from ‘flexible’ cuticle, which apparently contain a  $\beta$ -pleated sheet of three strands alternating with turns/loops.

A rather interesting observation is that with both classes of cuticular proteins, the sheets show a ‘polar’ character, i.e. the ‘faces’ of the sheets seem to have a different character. It is well known that alternating residues along a strand point towards opposite faces of a  $\beta$ -sheet. Thus, especially along long  $\beta$ -strands of these proteins, we can see where ‘bulky’ hydrophobic and/or aromatic residues alternate with other (sometimes hydrophilic) residues. To facilitate visualization, we have underlined and bolded those aromatic or hydrophobic residues that would appear on one face of a strand. These would be for protein HCCP66, in the strands 34–38 QYSLL, 47–51 VDYAA, 57–61 FNAV V and in protein MSCP14.6, in the strands 35–38 SYA YV, and 44–50 TYSVVYI. The situation is reminiscent of the alternation of small (G) and ‘bulky’ (A,S) residues to opposite sides (faces) of the  $\beta$ -sheet structure (‘polar’  $\beta$ -sheet) in silk fibroin (Marsh et al., 1955) and might be important for the formation of higher order structure (see Discussion).

We also note that a  $\beta$ -strand is predicted at the beginning of the hydrophilic region (residues 7–9 for protein HCCP66 with sequence FSY and residues 1–5 for protein MSCP14.6 with sequence YSYSV). These also

show an alternation of aromatic or hydrophobic residues with hydrophilic residues.

#### 4. Discussion

Portions of eight cuticular protein sequences from hard insect cuticle and nineteen from soft insect cuticle, were analyzed for secondary structure predictions, as explained in Materials and Methods. The 67–68 amino acids used come from the hydrophilic domain of cuticular proteins (Lampe and Willis, 1994; Andersen et al., 1995) and the associated R&R Consensus (Rebers and Riddiford, 1988).

One rather striking observation is that all the three **invariant** glycines (G) of the R&R consensus correspond exactly at maxima of  $\beta$ -turn/loop prediction. Their evolutionary invariance therefore is well suited to their structural role. It is well known that glycines are good turn/loop formers (Chou and Fasman, 1978; Chou and Fasman, 1979). Characteristic examples of invariant Gly residues in protein sequences for structural reasons (usually to ensure the integrity of  $\beta$ -turns) are numerous in the literature. The most well known, although for different structural constraints, is that of collagen.

Another important observation is that the turn/loop regions of the region containing the R&R consensus, frequently, contain histidines (H). This is well exemplified by protein AGCP2b (Dotson et al., 1998) with turns/loops at 23–27 HETRH, 32–34 HGQ, 40–43 SDGH and 50–54 HADHH. The localization of the histidines at turns/loops (‘edges’) of a  $\beta$ -pleated sheet (histidines there are ‘exposed’), seems to be in excellent agreement with their possible roles: Histidines are either involved in cuticular sclerotization as they readily react with activated N-acetyldopamine residues, or are involved in the variations of the water-binding capacity of cuticle and the interactions of its constituent proteins thanks to the fact that small changes of pH can affect the ionization of their imidazole group (Andersen et al., 1995).

It is also interesting to observe that aromatic residues (Y and F) are frequently predicted to belong to  $\beta$ -strands and to belong to the same, hydrophobic, face of the sheets (see Results). Their aromatic rings (if the  $\beta$ -pleated sheets intervene and interact with chitin filaments) could well stack against faces of the saccharide rings of chitin (poly-N-acetyl-glucosamine). This type of interaction is fairly common in protein-saccharide complexes (Vyas, 1991; Hamodrakas et al., 1997; Tews et al., 1997).

It has already been speculated that a part of the hydrophilic domain and especially the one that contains the R&R consensus, may be involved in specific binding to some common components in cuticles, such as chitin filaments (Andersen et al., 1995). The fact that this invariant domain, as we demonstrate here, most probably



adopts a well defined structure, that of  $\beta$ -pleated sheet is in good agreement with this speculation. In this context, it is important to note that Fraenkel and Rudall, some fifty years ago (1947), provided evidence from X-ray diffraction that the protein associated with chitin in insect cuticle has a  $\beta$ -sheet type of structure.

The difference in size of the  $\beta$ -pleated sheets predicted for proteins of the hard cuticle (four strands) from those predicted for proteins of the soft cuticle (three strands), might account, in part, for the difference in mechanical properties between rigid and flexible cuticles. It is expected that rigid cuticles would have a larger number of regular interactions between chitin filaments and the  $\beta$ -pleated sheet, that could stabilize further the structure.

In the two-phase composite, helicoidal systems of silkmoth and fish chorions, we have proposed and experimentally have shown that  $\beta$ -pleated sheet is the conformation which dictates formation of the helicoidal architecture (Hamodrakas, 1992 and references therein). In analogy to these systems, it is tempting to speculate that, the  $\beta$ -sheet of the extended R&R conserved region of cuticular proteins predicted in this work, constitutes a part of the protein matrix intervening between the chitin filaments (Neville, 1975), in the two-phase (polysaccharide-protein) composite cuticle, and dictates its helicoidal architecture.

However, in contrast to silkmoth chorion protein central domains, where the  $\beta$ -sheets are very regular consisting of four-residue  $\beta$ -strands alternating with regular  $\beta$ -turns (Hamodrakas, 1992), the postulated  $\beta$ -strands of the cuticular proteins are not of uniform size or spacing. This might be related to the requirements for the geometry of the formed structures. Chorions have a strict, almost spherical shape; maybe they have to be built from uniform and regular units, whereas cuticles have various shapes and perhaps less stringent requirements for their building blocks.

A variety of experiments can be performed to test the validity of this hypothesis and also to throw some light to the secrets of cuticle architecture: To start with, the structure of individual cuticular proteins and cuticular protein domains should be examined carefully in detail and perhaps, also, of specific two-phase systems of cuticular protein domains together with polysaccharide analogs of chitin.

## References

- Andersen, S.O., Hojrup, P., Roepstorff, P., 1995. Insect cuticular proteins. *Insect Biochem. Molec. Biol.* 25, 153–176.
- Baernholdt, D., Andersen, S.O., 1998. Sequence studies on post ecdysial cuticular proteins from pupae of the yellow mealworm, *Tenebrio molitor*. *Insect Biochem. Molec. Biol.* 28, 517–526.
- 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. *A.C.S. 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, New York, pp. 261–297.
- Burgess, A.W., Ponnuswamy, P.K., Scheraga, H.A., 1974. Analysis of conformations of amino acid residues and prediction of backbone topography in proteins. *Isr. J. Chem.* 12, 239–286.
- Chou, P., Fasman, G.D., 1974a. Conformational parameters for amino acids in helical,  $\beta$ -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.
- Chou, P., Fasman, G.D., 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.* 47, 145–148.
- Chou, P., Fasman, G.D., 1979. Prediction of  $\beta$ -turns. *Biophys. Journal* 26, 367–384.
- 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. Molec. Biol.* 28, 459–472.
- Dufton, M.J., Hider, R.C., 1977. Snake toxin secondary structure predictions: structure activity relationships. *J. Mol. Biol.* 115, 117–193.
- Fraenkel, G., Rudall, K.M., 1947. The structure of insect cuticles. *Proc. Roy. Soc. B.* 34, 111–143.
- Garnier, J., Osguthorpe, D.J., Robson, B., 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* 120, 97–120.
- Hamodrakas, S.J., 1988. A protein secondary structure prediction scheme for the IBM PC and compatibles. *CABIOS* 4, 473–477.
- 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–Heidelberg, pp. 116–186.
- Hamodrakas, S.J., Asher, S.A., Mazur, G.D., Regier, J.C., Kafatos, F.C., 1982a. Laser-Raman studies of protein conformation in the silkmoth chorion. *Biochim. Biophys. Acta* 703, 216–222.
- Hamodrakas, S.J., Jones, C.W., Kafatos, F.C., 1982b. Secondary structure predictions for silkmoth chorion proteins. *Biochim. Biophys. Acta* 700, 42–51.
- Hamodrakas, S.J., Kafatos, F.C., 1984. Structural implications of primary sequences from a family of Balbiani ring encoded proteins in chironomus. *J. Mol. Evol.* 20, 206–303.
- 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. *J. Struct. Biol.* 118, 23–30.
- Hamodrakas, S.J., Paulson, J.R., Rodakis, G.C., Kafatos, F.C., 1983. X-ray diffraction studies of a silkmoth chorion. *Int. J. Biol. Macromol.* 5, 149–153.
- Kragh, M., Molbak, L., Andersen, S.O., 1997. Cuticular proteins from the lobster *Homarus americanus*. *Comp. Biochem. Physiol.* 118B, 147–154.
- Kafatos, F.C., Regier, J.C., Mazur, G.D., Nadel, M.R., Blau, H.M., Petri, W.H., Wyman, A.R., Gelinas, R.E., Moore, P.B., Paul, M., Efstratiadis, A., Vournakis, J.N., Goldsmith, M.R., Hunsley, J.R., Baker, B., Nardi, J., Koehler, M., 1977. The eggshell of insects: Differentiation-specific proteins and the control of their synthesis and accumulation during development. In: Beerman, W. (Ed.), *Results and Problems in Cell Differentiation*, vol. 8. Springer-Verlag, Berlin–Heidelberg, pp. 45–145.
- 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. Molec. Biol.* 24, 419–435.
- Lim, V.I., 1974a. Structural principles of the globular organization of protein chains. A stereochemical theory of globular protein secondary structure. *J. Mol. Biol.* 88, 857–872.

- Lim, V.I., 1974b. Algorithms for prediction of  $\alpha$ -helical and  $\beta$ -structural regions in globular proteins. *J. Mol. Biol.* 88, 873–894.
- Marsh, R.E., Corey, R.B., Pauling, L., 1955. The structure of silk fibroin. *Biochim. Biophys. Acta* 16, 1–34.
- Mazur, G.D., Regier, J.C., Kafatos, F.C., 1982. Order and defects in the silkmoth chorioin, a biological analogue of a cholesteric liquid crystal. In: King, R.C., Akai, H. (Eds.), *Insect Ultrastructure*, vol 1. Plenum Press, New York, pp. 150–185.
- Nagano, K., 1977a. Logical analysis of the mechanism of protein folding IV Supersecondary structures. *J. Mol. Biol.* 109, 235–250.
- Nagano, K., 1977b. Triplet information in helix prediction applied to the analysis of supersecondary structures. *J. Mol. Biol.* 109, 251–274.
- Neville, A.C., 1975. *Biology of the Arthropod Cuticle*. Springer-Verlag, New York.
- 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.
- Norup, T., Berg, T., Stenholm, H., Andersen, S.O., Hojrup, P., 1996. Purification and characterization of five cuticular proteins from the spider *Araneus diadematus*. *Insect Biochem. Molec. Biol.* 26, 907–915.
- Nousiainen, M., Rafn, K., Skou, L., Roepstorff, P., Andersen, S.O., 1998. Characterization of exoskeletal proteins from the American lobster *Homarus americanus*. *Comp. Biochem. Physiol.* 119, 189–199.
- 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.
- Rost, B., Sander, C., 1993. Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* 232, 584–599.
- Rost, B., Sander, C., 1994. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* 19, 55–55.
- Rost, B., 1996. PHD: Predicting one-dimensional protein structure by profile based neural networks. *Meth. In Enzym.* 266, 525–539.
- 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.
- 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 Res.* 22, 4673–4680.
- Von Heijne, G., 1987. *Sequence analysis in molecular biology: Treasure trove or trivial pursuit*, Academic Press Inc., New York.
- Vyas, N.K., 1991. Atomic features of protein-carbohydrate interactions. *Curr. Opin. Struct. Biol.* 1, 732–740.
- Willis, J.H., 1999. Cuticular proteins in insects and crustaceans. *Amer. Zoologist*. In press.