Protein Engineering vol.2 no.3 pp.201–207, 1988

Structural models of the evolutionarily conservative central domain of silk-moth chorion proteins

Stavros J.Hamodrakas, Heinz E.Bosshard[1] and Christopher N.Carlson[1]

Department of Biochemistry, Cell and Molecular Biology and Genetics, University of Athens, Panepistimiopolis, Athens 157.01, Greece and [1]European Molecular Biology Laboratory, Meyerhofstrasse 1, Postfach 1022.09, 6900 Heidelberg, FRG

Silk-moth chorion proteins belong to a small number of families: A, B, C, Hc-A and Hc-B. The central domain is an evolutionarily conservative region in each family, of variable length and composition between families. This domain shows clear 6-fold periodicities for various amino acid residues, e.g. glycine. The periodicities, together with the well-documented prevalence of β-sheet and β-turn secondary structure of chorion proteins, strongly support a structural model in which four-residue β-strands alternate with β-turns, forming a compact antiparallel, probably twisted β-sheet. Conformational analysis, aided by interactive graphics refinement and recent experimental findings, further suggest that this structure consists of β-strands, alternating with I′ and II′ β-turns, and apparently forms the basis for the molecular and supramolecular assembly of chorion.

Key words: silk-moth chorion proteins/β-strand/β-sheet/β-turn/central domain

Introduction

The silk-moth eggshell or chorion is a complex extracellular proteinaceous formation. Its ultrastructure and morphogenesis have been studied extensively, in parallel with sequence analysis of its structural genes, and with studies aimed at elucidating the mechanisms that regulate sequential expression of these genes during development (reviewed by Kafatos *et al.*, 1977; Mazur *et al.*, 1982; Kafatos, 1983; Goldsmith and Kafatos, 1984; Regier and Kafatos, 1985).

Biochemically, the chorion is surprisingly complex: as many as 186 protein components have been resolved by two-dimensional gel electrophoresis, from the chorions produced by an individual *Antheraea polyphemus* moth (Regier *et al.*, 1980). It appears that the number of chorion genes is at least as high (Eickbush and Kafatos, 1982). However, most chorion genes are related: they belong to a small number of gene families (A, B, C, Hc-A, Hc-B . . .), each encompassing multiple genes that arose during evolution by reduplication followed by sequence divergence (see, for example, Regier *et al.*, 1978; Jones and Kafatos, 1980, 1982). The gene families are themselves related, and constitute a superfamily, with one branch encompassing the A and Hc-A families, and the other the B, Hc-B and C families (Rodakis and Kafatos, 1982; Regier *et al.*, 1983; Iatrou *et al.*, 1984; Rodakis *et al.*, 1984).

Primary sequences have been determined for all five of these gene families (references listed above). Sequence comparisons and predictions of secondary structure have revealed that chorion proteins have a tripartite structure (Hamodrakas *et al.*, 1982a; Regier *et al.*, 1983). A central domain is highly conserved within each family and can be recognized as homologous between families of the same branch. The flanking, N- and C-terminal domains or 'arms' are more variable, and are marked by the presence of tandemly repetitive peptides that are not apparent in the central domain. In agreement with the predictions, laser-Raman and X-ray diffraction studies show that β-sheet structure predominates in the chorion (Hamodrakas *et al.*, 1982b, 1983, 1984, 1986).

Ultrastructurally the chorion consists of fibrous layers parallel to the chorion surface (Smith *et al.*, 1971; Papanicolaou *et al.*, 1986, and references therein). Between adjacent layers the direction of the fibres differs by a constant angle, resulting in a helicoidal structure (cf. Bouligand, 1972) which is a biological analogue of a cholesteric liquid crystal (Mazur *et al.*, 1982). The structure changes dramatically during morphogenesis and also varies locally, consistent with the biochemical complexity and the multiple physiological functions of the eggshell (Kafatos, 1983).

Ultimately, we want to relate the primary sequences and tripartite composition of chorion proteins to the fibrous structures they assume, and the regular assemblies of these fibres. Although the chorion is very complex, this undertaking is facilitated by the evolutionary relatedness of the components. Limited evolutionary variations amount to information that can help specify further the fundamental molecular properties of these proteins.

In this work, we present a detailed analysis of central domain sequences from the B/Hc-B and C branches of the chorion superfamily. These sequences can be accommodated in a compact antiparallel β-sheet structure, similar to that appearing in the A/Hc-A branch (Hamodrakas *et al.*, 1985), which we suggest as the starting point for understanding the molecular and supramolecular assembly of the chorion. Detailed models of this structure are presented, obtained after a conformation analysis and refinement with an interactive graphics system, and attractive alternatives are also discussed. The results and models are in good agreement with recent experimental findings (Sibanda and Thornton, 1985; Hamodrakas *et al.*, 1986).

Methods

Chorion complementary DNA sequences were determined and converted to protein sequences as described (Tsitilou *et al.*, 1980; Jones and Kafatos, 1982; Rodakis and Kafatos, 1982; Rodakis *et al.*, 1982).

Pattern strengths for sequences arranged in rows of six residues were calculated by the method of McLachlan (1977). Fourier transforms were obtained essentially as outlined by McLachlan and Stewart (1987), using a Fortran 77 computer program, kindly provided by Dr A.D.McLachlan and suitably modified for use in the CYBER 170-730 computer of the University of Athens. Each sequence of N residues was represented as a linear array of N terms, with each term given a value of 1 or 0, according to whether the condition considered (e.g. presence of a Gly residue) was or was not satisfied. To increase resolution, this

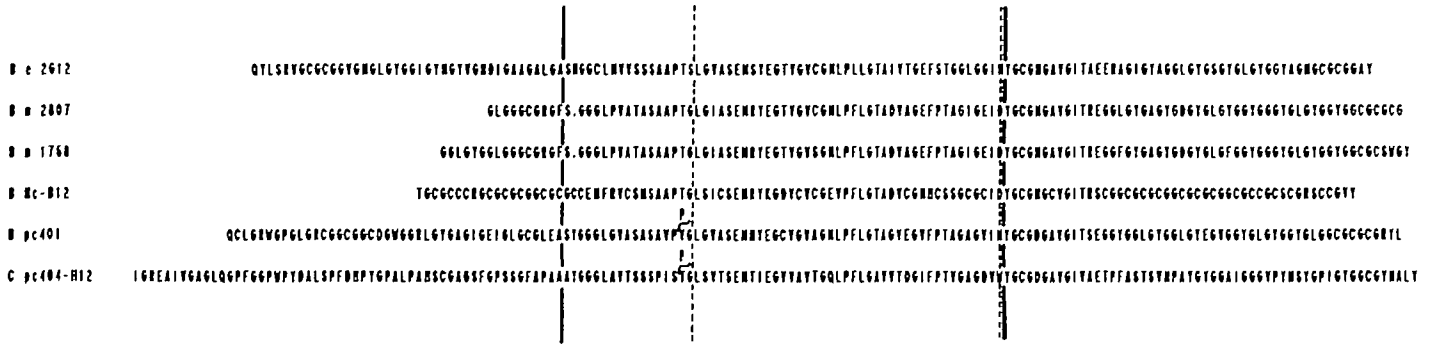

Fig. 1. Sequences of typical B, Hc-B and C proteins. Unbroken lines mark the borders of the central domain and broken lines flank the portions of the sequences considered in this work. Proteins B e 2G12, B m 2807, B m 1768 are representative proteins of the B family of *B.mori*, B Hc-B12 of the Hc-B family of *B.mori*, B pc401 of the B family of *A.polyphemus* and C pc404-H12 of the C family of *A.polyphemus* (references given in the Introduction).

a

	B e 2G12	B m 2807	B m 1768	B Hc-B12	B pc401	C pc404-H12
I	L G V A	L G I A	L G I A	L S I C	L G V A	L S V T
II	S E N S Y E	S E N R Y E	S E N R Y E	S E N R Y K	S E N M Y E	S E N T I E
III	G T V G V C	G T V G V C	G T V G V S	G D V C V C	G C V G V A	G V V A V T
IV	G N L P L L	G N L P F L	G N L P F L	G E V P F L	G N L P F L	G Q L P F L
V	G T A I V T	G T A D V A	G T A D V A	G T A D V C	G T A G V E	G A V V T D
VI	G E F S T G	G E F P T A	G E F P T A	G N M C S S	G V F P T A	G I F P T V
VII	G L G G I N	G I G E I D	G I G E I D	G C G C I D	G A G V I N	G A G D V W
	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6

b

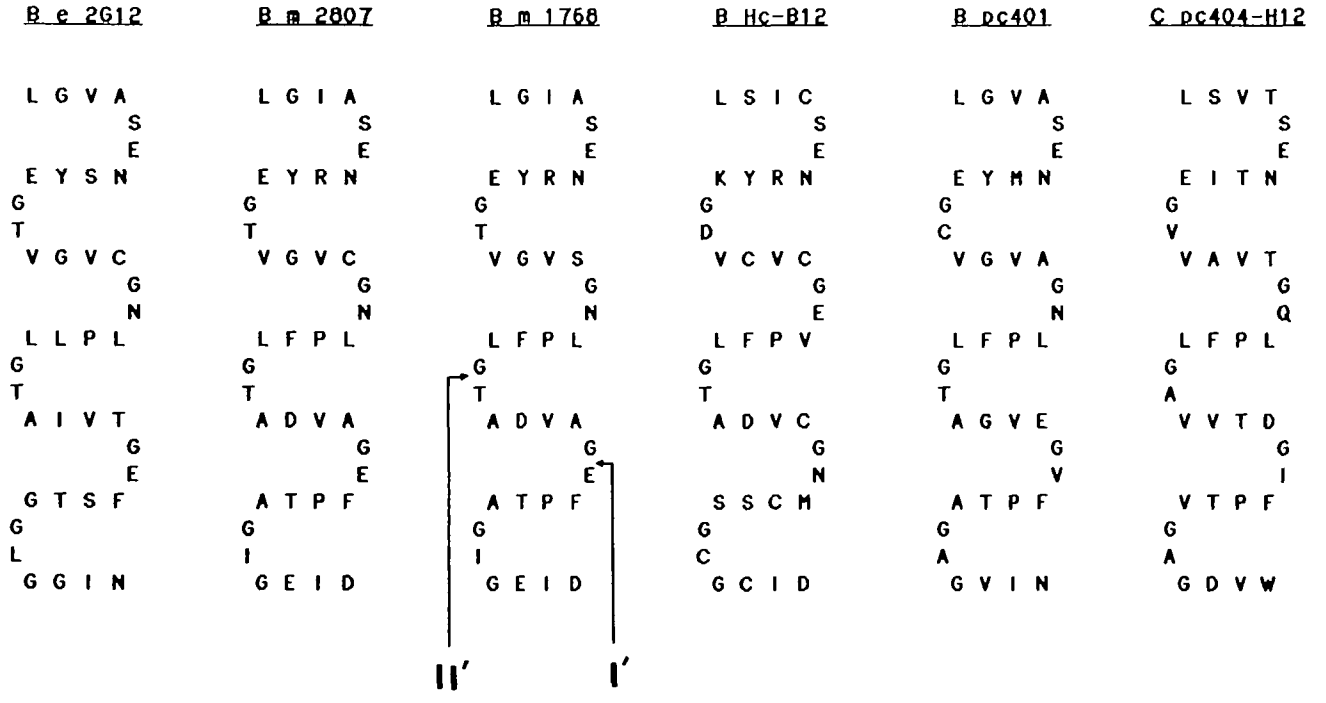

Fig. 2. (a) Regular amino acid distribution within the central domain. To reveal the 6-fold periodicities, sequences have been written in rows (numbered I–VII) of six residues each; they should be read left to right, top to bottom. Vertical columns (numbered 1–6) can thus be seen to have non-random prevalence of certain types of residues. A distinct pattern of Gly-*X*-large hydrophobic-*Y*-large hydrophobic-*Z* emerges, where *X* is usually a β-turn former and *Y* and *Z* other types of residues. (b) Antiparallel β-sheet model for the portion of the central domain considered. Sequences should be read continuously, beginning at the top. Tentative I′ and II′ β-turns alternate with four-residue β-strands. For further details, see text.

array was embedded in a larger array of zeros (McLachlan, 1977). In addition, partial Fourier inversion was employed, as suggested by Bear *et al.* (1978).

Modelling of the protein structures was performed by utilizing the interactive computer graphics facilities of the European Molecular Biology Laboratory, Heidelberg. An Evans and Sutherland Multipicture System was used, with colour and black and white displays, 256 kilowords of extended memory, and

various input and output devices. The system is hosted by a Digital Equipment Corp. VAX-11/785 computer. The interactive molecular modelling program EMBLFrodo, a descendant of the popular Frodo by T.A.Jones, was employed.

For the construction of our molecular models we have also used a Basic computer program, written for a Cannon AS100 microcomputer, with 256 Kbytes of memory and a screen resolution of 640 × 400 pixels, which constructs and manipulates space-filling or skeletal models of proteins. We further employed a version of PLUTO78 (W.D.S.Motherwell, Cambridge Crystallographic Data Centre, UK), installed for use in the CYBER 170-730 computer of the University of Athens.

Physical models were built with Pauling and Corey CPK and Kendrew skeletal kits.

Results

Definition of the sequences considered

Figure 1 presents typical chorion sequences of the B, Hc-B and C families. The region defined as the central domain is indicated, as is the region considered in the present analysis. The flanking arm sequences were not analysed since they are more variable.

We present the analysis of several sequences, representative of the central domain in the *Bombyx mori* Hc-B family and in the B and C family of both *A.polyphemus* and *B.mori*; our conclusions apply to all other available sequences, published and unpublished.

Hexad periodicities

The region considered is highly conserved among B, Hc-B and C proteins, and has not undergone deletions or insertions over >50 million years (Rodakis *et al.*, 1982). For preliminary analysis of periodicities a Fortran program was written (S.J. Hamodrakas, unpublished), which revealed a 6-fold periodicity for Gly. Accordingly, sequences were written out in rows of six residues (Figure 2a), and the significance of the non-random distribution of residues in the six columns thus generated was analysed with the method of McLachlan (1977), calculating the pattern strength, P. This measure is the difference between observed and randomly expected unevenness in the distribution of amino acid residues over the columns, divided by the standard deviation. Therefore, values of $P > 3.0$ are highly significant. The analysis established clear 6-fold periodicities for various types of residues which may be summarized by declaring that the hexapeptides have the general form:

Gly-*X*-large hydrophobic-*Y*-large hydrophobic-*Z*

with *X* usually a β-turn former residue and *Y*, *Z* of a more variable type (data not shown, but can be given by S.J.Hamodrakas to interested readers to assess the statistical significance of the repeats).

However, Fourier analysis of the sequences [as was performed for the A class of proteins by Hamodrakas *et al.* (1985)] does not show significant maxima for a period of approximately six residues, for most proteins (exceptions are the Hc proteins Hc-B12, Hc-B13 and the C protein pc404-H12, data not shown but available as above). Instead, it shows intense Fourier maxima usually for a period which is a fraction of 6, namely 6/3 = 2 or 6/2 = 3 residues (data not shown but available as above).

Interpretation of hexad periodicities

Following the arguments in detail by Hamodrakas *et al.* (1985) in the analysis of the A chorion proteins, we interpret the hexad periodicities appearing in the central domains of the B, Hc-B and C classes of proteins, by the alternating β-turn/β-strand model of an antiparallel β-pleated sheet shown in Figure 2b. In this figure, the horizontal rows, each containing four residues, represent antiparallel β-sheet strands (see also Discussion) and consist of the residues shown in columns 3–6 of Figure 2a. The end residues of these short strands also participate in β-turns, together with the vertically displayed dipeptides (columns 1 and 2 of Figure 2a). The latter represent the central residues $i + 1$ and $i + 2$ of the β-turns respectively. A similar model was proposed for the conservative central domain of the A proteins (Hamodrakas *et al.*, 1985).

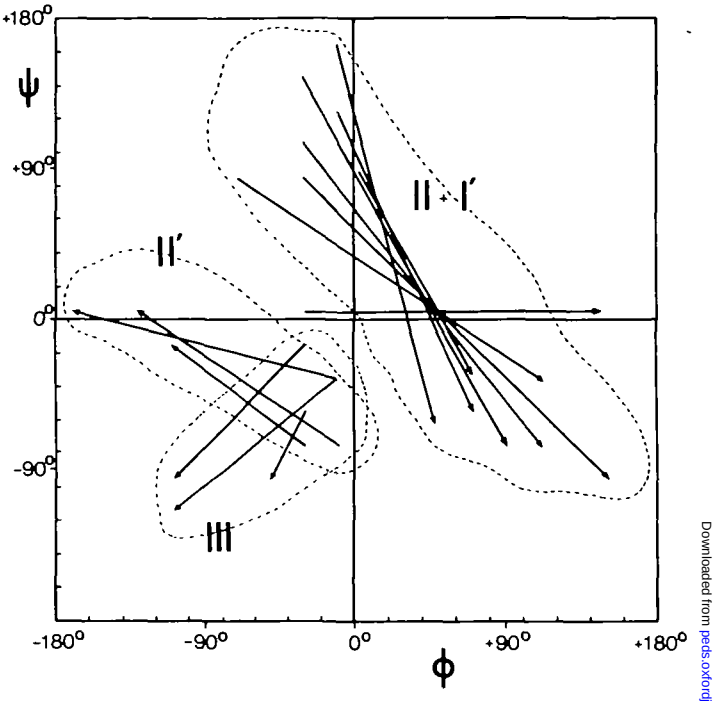

Fig. 3. Phi–psi plot showing the conformation of the turn residues $i + 1$, $i + 2$ by arrows, as were found by our conformational analysis of the decapeptide VAVAGELPVA, as described in the text. The beginning of each arrow marks residue $i + 1$ and the end residue $i + 2$. Three groups of possible β-turns were found: Group 1, which corresponds to distorted types of II and I′ turns (Chou and Fasman, 1977), Group 2, corresponding to distorted types of II′ turns, and Group 3, which corresponds to distorted types of III turns (see also Discussion). Presumably, these distorted turns are necessary in the model because the phi, psi angles of the sheet residues were kept constant.

β-Turn type determinations and modelling

To construct realistic models of the proposed antiparallel β-pleated structure for all chorion proteins, it was necessary to determine the types of the β-turns (Chou and Fasman, 1977) which give the silk-moth chorion protein central conservative domains their characteristic fold, shown in Figure 2b and in Figure 2(b) of Hamodrakas *et al.* (1985).

Since the conformation of residues i, $i + 3$ is constrained to have values corresponding to those of an antiparallel β-sheet (these residues are parts of the four-residue β-strands), this means that we had to determine the conformation of residues $i + 1$, $i + 2$, of each β-turn.

A similar analysis has been performed by Geddes *et al.* (1968) to determine the types of β-turns which generate the cross-β conformation in proteins. These authors arrived at the conclusion

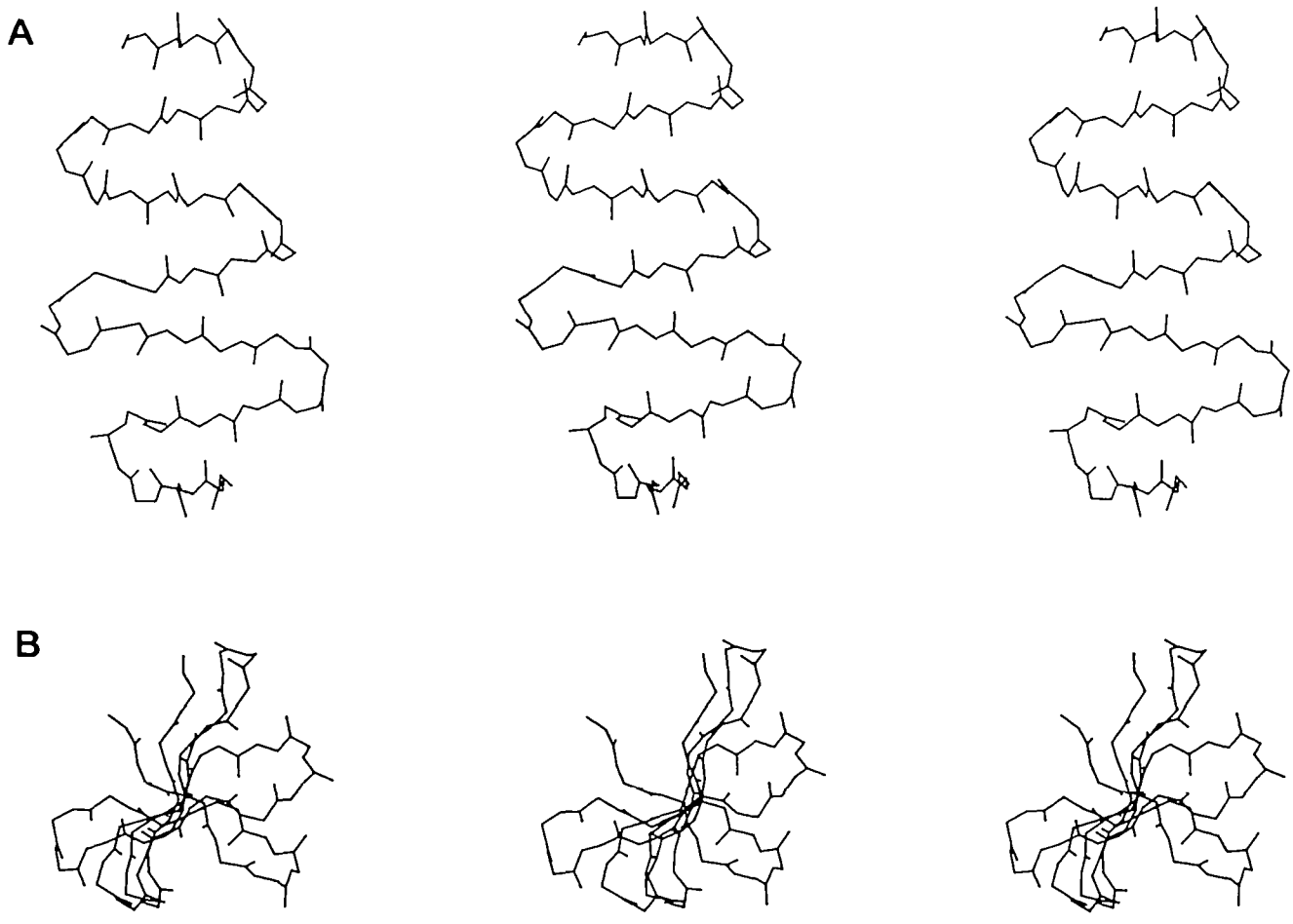

Fig. 4. A skeletal model, obtained from the interactive graphics system, showing the characteristic β-pleated sheet fold of the central conservative domain of the A protein pc609 (main chain and carbonyl oxygens only). (**A**) View perpendicular to the 'plane' of the β-sheet. (**B**) View perpendicular to the strands, parallel to the β-sheet 'plane'. The three-picture stereo system used in this figure enables readers with both normal and cross-over stereo vision to view the image. For normal vision select the left and centre images, for cross-over vision use the centre and right images.

that two types of turns were the most favourable for the formation of the cross-β conformation, which they denoted as fold A and fold B. They essentially correspond to the II′ and II types of Chou and Fasman (1977) respectively. In these folds, a hydrogen bond is formed between the -NH group of residue $i + 3$ with the -C =O group of residue i.

Our analysis, which in general terms resembled the analysis of Geddes *et al.* (1968), but was carried out with our own FORTRAN program, was carried out as follows.

(i) Representative decapeptides of the type forming the characteristic structure of Figure 2b were chosen, which contain two consecutive β-strands linked by a β-turn.

(ii) Backbone dihedral angles of the β-strand residues were set to phi = $-120°$ and psi = $+135°$. These values correspond to the centre of the allowed region for β-pleated sheets and they generate β-strands suitable for the formation of twisted β-pleated sheets (Schulz and Schirmer, 1978). We have chosen these values since most known β-sheets, both in globular and structural proteins, are twisted (Richardson, 1981; Lotz *et al.*, 1982; Fraser and McRae, 1976), and since experimental findings (Hamodrakas, 1984; Hamodrakas *et al.*, 1986) indicate that silk-moth chorion proteins contain twisted sheets.

(iii) The phi and psi angles of residues $i + 1$, $i + 2$ of the turns were varied in a systematic way, in order to generate allowed conformations without steric hindrance, and such that the -NH and =CO groups of the β-strands can create hydrogen bonds.

Several possible conformations were found from this analysis which are mostly distorted types of well-known β-turns (Figure 3). They can be classified into three major classes: type II or I′, type II′ and type III β-turns (see also Discussion). However, from model building, we selected β-turns II or I′ and II′, which, provided they alternate along the structure shown in Figure 2b, create satisfactory models for the conservative domains of silk-moth chorion proteins. In these models, favourable hydrogen bonds are formed between the NH group of residue $i + 3$ with the CO group of residue i.

The models obtained from this stereochemical analysis, assuming β-turns of the II or I′ and II′ type, alternating with four-residue β-strands along the sequence, were refined in detail on an interactive graphics system. They are shown for a representative protein in Figure 4.

The β-pleated sheet twist of these models is possibly exaggerated. This might be a consequence of the proline residues in two strands of each sheet, whose phi angle has been constrained to $\sim -60°$ in the models and of the phi and psi angles of the remaining strand residues being held at $-120°$ and $+135°$ respectively.

Discussion

Tandemly repetitive peptides have been found in the sequences of most fibrous proteins and play an important role in the formation of the fibrous structure (Fraser and McRae, 1973). Individual repeat units tend to be conformationally equivalent. If the equivalence is exact, a helical structure results, if not, the local conformations of the repeat units are likely to be similar (Parry *et al.*, 1979). An important question which always arises in such cases is what type of structure is formed by these repetitive peptides.

In the case of chorion proteins, which are products of the distinct but related multigene families A, B, Hc-A, Hc-B and C, and which are a characteristic example of protein engineering *in vivo*, we have convincingly shown for at least two related families, the A and the Hc-A, that their central conservative domains can be considered as constructed from repetitive hexapeptides which most probably form an antiparallel β-sheet. This structure is composed of four-residue β-strands, alternating with β-turns, with a ratio of 2:1 (Hamodrakas *et al.*, 1985). The model is somewhat reminiscent of feather keratin (Fraser and McRae, 1976) and of the cross-β-sheet structure in the shaft of the adenovirus fibre protein (Green *et al.*, 1983), as described in detail in a previous publication (Hamodrakas *et al.*, 1985).

A quantitative measurement of secondary structure in silk-moth chorion proteins by analysis of the amide I band of the laser-Raman spectrum of chorion, suggests that chorion protein components consist of 60−70% antiparallel β-sheet and 30−40% β-turns (Hamodrakas *et al.*, 1984). These estimates, although in good agreement with the proposed model structure, cannot be attributed solely to the central domain of the A and Hc-A families, since that domain accounts only for ~20% of total chorion mass. The models we present here for the central conservative domains of the B, Hc-B and C families (four-residue β-strands alternating with β-turns whose first and fourth residues are parts of the β-strands) are very similar to those of the As and the Hc-As and are in agreement with the percentages found by experiment: now, the conservative domains of all chorion protein families account for ~50% of total chorion mass.

To fully account for the observed experimental percentages of secondary structure, the question remains, of course, what structure the variable chorion protein 'arms' adopt. They also contain tandemly repetitive peptides, which is evident from the sequence, but different in nature from the peptides of the central region. This is a subject of an ongoing investigation.

A structure closely resembling the models presented in Figures 2b and 4 seems almost inevitable in view of the evidence. Perhaps, there are still ambiguities for a few minor points. For example, what is the exact type of β-turn which generates this structure? Unfortunately, the ambiguities cannot be resolved experimentally. Some authors suggest that this can be done by using laser-Raman spectroscopy (Krimm and Bandekar, 1986; and references therein). We have obtained laser-Raman spectra and studied them very carefully (Hamodrakas *et al.*, 1982b, 1984). Despite this, we have not been able to identify the exact type of turns unequivocally. Perhaps, it would be possible to obtain partial answers to the problem by synthesizing peptides representative of segments of the central domain of chorion proteins and solving their structure crystallographically. However, we have no guarantee that these peptides will fold to a conformation similar to the one *in vivo*.

It was very interesting to note that a recent systematic search of β-turns, connecting anti-parallel β-sheet strands in globular proteins (Sibanda and Thornton, 1985), has revealed that the type II′ and I′ turns clearly predominate. Out of the 29 β-turns observed experimentally, 15 were type I′, 10 type II′ and only four type I. The type I′ turn differs from the type II turn in the phi and psi angles of only one residue [see Figure 2 of Sibanda and Thornton (1985)]. Surprisingly, the type II turn is not observed to link adjacent antiparallel β-sheet strands in globular proteins.

A comparison of our theoretical search (Figure 3) with Figure 2 of Sibanda and Thornton (1985) clearly shows that our predicted structures do not differ substantially from those observed. Some of our type-II turns can be considered as distorted I′ turns (Figure 3 arrows). Sibanda and Thornton (1985) suggest that the abundance of the type-I′ turns is probably due to the fact that they have the correct twist to match the relative twist which is always observed between adjacent strands. Close study of our models verifies that this is indeed the case. Therefore, it appears that a further refinement of our proposed model is to adjust the phi and psi angles of only one residue, to belong to a type-I′ turn rather than a type-II turn. This modification leaves the remainder of the structure unaffected.

This model is further supported by the pattern of residues appearing in the β-turns. Sibands and Thornton (1985) have found that the observed β-turns are strongly selective for amino acid type. For a type-II′ turn they found that Gly predominates in the second position, whereas the third position is usually occupied by Ser, Thr or a polar residue; for a type-I′ turn, the second position is occupied mostly by Gly, Asp or Asn and the third usually by Gly. Close study of our models (Figure 2b) shows that for the type-II′ turns the pattern is almost ideal: Gly is usually found in the second position, whereas a Thr is usually found in the third. For the type-I′ (or II) turns the second position is occupied mostly by Gly, in agreement with the observed data, whereas, in the third position, charged (Glu) or polar (Asn) residues frequently appear.

The fact that the Fourier transform maxima for a hexapeptide periodicity in the central domain of the Bs are not particularly intense, but instead maxima for fractions (6/2 and 6/3) of this periodicity are very strong (see Methods), is not unusual for this method. A characteristic recent example is the analysis of the amino acid sequence of the nematode myosin rod, where Fourier transforms of the amino acid profiles show strong periodicities based on repeats of seven residues and 28 residues, as implied from the Fourier maxima which appear at 7/2 and 7/3 and also 28/2 and 28/3 residues (McLachlan and Karn, 1983). On the other hand, as mentioned above, certain Hc-B proteins (examples are Hc-B12 and Hc-B13) and the C protein pc404-H12 show clear 6-fold periodicities in their central domain, which further supports our proposed structure.

This structure, common to all chorion protein families and subfamilies, has the following characteristics.

(i) It is highly conservative in each family and subfamily. Greater variability in sequence and perhaps in secondary structure (although this is not certain), is seen in the remainder of the molecules ('arms').

(ii) In the short four-residue β-strands relatively 'small' residues (e.g. G) tend to alternate with 'bulky' residues in several cases, e.g. VAVA, VSIG, etc. This is reminiscent of the alternation of 'small' (G) and 'bulky' (A,S) residues to opposite sides of the β-sheet structure in silk fibroin (Marsh *et al.*, 1955) and may be important in chorion for the packing of β-sheets to form higher order structure.

(iii) Both faces of the proposed β-sheets have a pronounced

hydrophobic character, except for certain regions (e.g. in the strand VDFC), which might serve as sites for specific recognition. The existence of charged or polar residues in these regions shows that they should be counterbalanced by complementary residues in neighbouring β-sheets for the formation of hydrogen or salt bonds during morphogenesis of higher order structure. It is well known [for a recent analysis see Rashin and Honig (1984)] that this is always the case whenever a charged or polar residue is found in the interior of water-soluble globular proteins. A detailed analysis of protein−protein interactions is well under way (S.J. Hamodrakas, unpublished) for silk-moth chorion proteins.

(iv) Although both 'edges' of the proposed β-sheet structure, namely the two central residues of the β-turns, consist mostly of Gly and polar residues, it is clearly seen from Figure 2b that they show an uneven distribution of charges. Obviously, the right-hand side of the proposed sheets (Figure 2b) contains a number of charged residues (particularly Glu). The role of these charges has not been clarified yet but they appear to be very important for the formation of higher order structure.

(v) Certain residues occupy characteristic positions in this β-sheet structure. Cysteines are often found in β-turns, in positions favourable to create disulphide bonds, which crosslink adjacent protein molecules during the late choriogenetic stages (Hamodrakas *et al.*, 1984). Two prolines in the A, B and C proteins are always highly conserved and are always accompanied by large hydrophobic residues (I, V and L) in two β-strands which are distant by one strand. Their exact role is not yet fully understood. Prolines are not very common in β-strands and they do not favour β-sheet structure (Richardson, 1981). Sometimes their presence is marked by the formation of β-bulges (Richardson, 1981). Their appearance has seriously impaired our efforts to create satisfactory models of silk-moth chorion proteins. Perhaps, they simply serve to increase the sheet twist, if the β-sheets are actually twisted.

The presence of Gly with a periodicity of six residues was the first clue which led us to the proposed structure. However, this 6-fold periodicity has prompted N.M.Green (personal communication) to point out that a possible structure could be one formed by an alternation of γ and inverse-γ turns (Mathews, 1972; Nemethy and Printz, 1972), connecting five-residue β-strands. The Gly, repeated every sixth residue, could then occupy the central position of the three residue γ and inverse-γ turns.

Although the derived structures appear to be attractive, they have serious disadvantages as was checked with the graphics system and they are not described here.

Regarding the relation of the proposed structures (Figure 4) with the helicoidal, higher order structure of silk-moth chorion, a biological analogue of a cholesteric liquid crystal, we can say that the (twisted?, helical?) β-sheet ribbons of chorion proteins are the basis for the morphogenesis of the helicoidal architecture (Hamodrakas, 1984; Hamodrakas *et al.*, 1986). Rudall proposed some 30 years ago (1956) that a helicoidal structure can be formed from interactions of helical molecules.

The cross-sections of the β-sheet ribbons of the central domains of chorion proteins (Figure 4) must be ~30 Å (3 nm) in diameter, taking into account the side chains of the turn residues, as measured directly from molecular models and on the interactive graphics screen. They are in good agreement with the values indicated by X-ray diffraction and freeze-fracturing (Hamodrakas *et al.*, 1986).

The modes of packing of the proposed β-sheets are currently being studied, to determine the rules of formation of the helicoidal structure. These should be based on simple packing rules of β-sheets (Chothia and Janin, 1982; Cohen *et al.*, 1981).

Acknowledgements

S.J.H. thanks Dr A.D.McLachlan for providing him with his Fourier transform program and Drs A.Klug, A.Crowther and R Henderson for stimulating discussions during his visit to the MRC, Cambridge, UK. He also thanks Dr N.M.Green for pointing out the possibility of the γ-turn structure. Part of this work was made possible by an EMBO short-term fellowship to S.J.H, who thanks EMBL, Heidelberg, for hospitality. He also thanks the Athens Computing Laboratory for computing facilities and the Greek Ministry of Research and Technology and the University of Athens for financial support.

References

Bear,R.S., Adams,J.B. and Poultoy,J.W. (1978) *J. Mol. Biol.*, **118**, 123−126.
Blau,H.M. and Kafatos,F.C. (1979) *Dev. Biol.*, **72**, 211−225.
Bouligand,Y. (1972) *Tissue Cell*, **4**, 189−217.
Chothia,C. and Janin,J. (1982) *Biochemistry*, **21**, 3955−3965.
Chou,P.Y. and Fasman,G. (1977) *J. Mol. Biol.*, **115**, 135−175.
Cohen,F.E., Sternberg,M.J.E. and Taylor,W.R. (1981) *J. Mol. Biol.*, **148**, 253−272.
Eickbush,T.H. and Kafatos,F.C. (1982) *Cell*, **29**, 633−643.
Fraser,R.D.B. and McRae,T.P. (1973) In *Conformation in Fibrous Proteins.* Academic Press, New York.
Fraser,R.D.B. and McRae,T.P. (1976) *Proc. 16th Int. Ornithological Congress, Canberra.* Australian Academy of Science, Canberra, pp. 445−451.
Gedds,A.J., Parker,K.D., Atkins,E.D.T. and Beighton,E. (1968) *J. Mol. Biol.*, **32**, 343−358.
Goldsmith,M.R. and Kafatos,F.C. (1984) *Annu. Rev. Genet.*, **18**, 443−487.
Green,N.M., Wrigley,N.G., Russel,W.C., Martin,S.R. and McLachlan,A.D. (1983) *EMBO J.*, **2**, 1357−1365.
Hamodrakas,S.J (1984) *Int. J. Biol. Macromol.*, **6**, 51−53.
Hamodrakas,S.J., Jones,C.W. and Kafatos,F.C. (1982a) *Biochim. Biophys. Acta*, **700**, 42−51.
Hamodrakas,S.J., Asher,S.A., Mazur,G.D , Regier,J.C. and Kafatos,F.C. (1982b) *Biochim. Biophys. Acta*, **703**, 216−222.
Hamodrakas,S.J., Paulson,J.R., Rodakis,G.C. and Kafatos,F.C. (1983) *Int. J. Biol. Macromol.*, **5**, 149−153.
Hamodrakas,S.J., Kamitsos,E.I. and Papanicolaou,A. (1984) *Int. J. Biol. Macromol.*, **6**, 333−336.
Hamodrakas,S.J., Etmektzoglou,Th. and Kafatos,F.C. (1985) *J. Mol. Biol.*, **186**, 583−589.
Hamodrakas,S.J., Margaritis,L.H., Papassideri,I. and Fowler,A. (1986) *Int J Biol. Macromol.*, **8**, 237−242.
Iatrou,K., Tsitilou,S.G. and Kafatos,F.C. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4452−4456.
Jones,C.W. and Kafatos,F.C. (1980) *Cell*, **22**, 855−867.
Jones,C.W. and Kafatos,F.C. (1982) *J. Mol. Evol.*, **19**, 87−103.
Jones,C.W., Rosenthal,N., Rodakis,G.C. and Kafatos,F.C. (1979) *Cell*, **18**, 1317−1332.
Kafatos,F.C. (1983) In Subtelny,S. and Kafatos,F.C. (eds), *Gene Structure and Regulation in Development* Alan R.Liss, Inc., New York, pp. 33−61.
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. and Koehler,M. (1977) In Beerman,W. (ed.), *Results and Problems in Cell Differentiation.* Springer-Verlag, Berlin, Vol. 8, pp. 45−145.
Krimm,S. and Bandekar,J. (1986) *Adv. Protein Chem.*, **39**, 181−364.
Lotz,B., Gouthier-Vassal,A., Brack,A. and Magoshi,J. (1982) *J. Mol. Biol.*, **156**, 345−357.
Marsh,R.E., Corey,R.B. and Pauling,L. (1955) *Biochim. Biophys. Acta*, **16**, 1−34.
Mathews,B.W. (1972) *Macromolecules*, **5**, 818−819.
Mazur,G.D., Regier,J.C. and Kafatos,F.C. (1982) In Akai,H. and King,R.C. (eds), *Insect Ultrastructure.* Plenum Publishing Corp., New York, Vol. 1, pp. 150−183.
McLachlan,A.D. (1977) *Biopolymers*, **16**, 1271−1297.
McLachlan,A.D. and Stewart,M. (1976) *J. Mol. Biol.*, **103**, 271−298.
McLachlan,A.D. and Karn,J. (1983) *J. Mol. Biol.*, **164**, 605−626.
Nemethy,G. and Printz,M.P. (1972) *Macromolecules*, **5**, 755−758.
Papanicolaou,A.M., Margaritis,L.H. and Hamodrakas,S.J. (1986) *Can. J. Zool.*, **64**, 1158−1173.
Parry,D.A.D., Fraser,R.D.B. and McRae,T.P. (1979) *Int. J. Biol. Macromol.*, **1**, 17−22.
Rashin,A.A. and Honig,B. (1984) *J. Mol. Biol.*, **173**, 515−521.

Regier,J.C. and Kafatos,F.C. (1985) In Kerkut,G.A. and Gilbert,L.I. (eds), *Embryogenesis and Reproduction. Comprehensive Insect Physiology, Biochemistry and Pharmacology*. Pergamon Press, Oxford, Vol. 1, pp. 113–151.
Regier,J.C., Kafatos,F.C., Goodfliesh,R. and Hood,L. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 390–394.
Regier,J.C , Mazur,G.D. and Kafatos,F.C. (1980) *Dev. Biol.*, **76**, 286–304.
Regier,J.C., Mazur,G.D., Kafatos,F.C. and Paul,M. (1982) *Dev. Biol.*, **92**, 159–174.
Regier,J.C., Kafatos,F.C. and Hamodrakas,S.J. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1043–1047.
Richardson,J.D. (1977) *Nature*, **268**, 495–500.
Richardson,J.S. (1981) *Adv. Protein Chem.*, **34**, 167–339.
Rodakis,G.C. and Kafatos,F.C. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 3551–3555.
Rodakis,G.C , Moschonas,N.K. and Kafatos,F.C. (1982) *Mol. Cell. Biol.*, **2**, 554–563.
Rodakis,G.C., Lekanidou,R. and Eickbush,T.H. (1984) *J. Mol. Evol.*, **20**, 265–273.
Rudall,K.M. (1956) In *Lectures on the Scientific Basis of Medicine*. Athlone Press, London, Vol. 5, pp. 217–230.
Schulz,G.E. and Schirmer,R.H. (1978) In *Principles of Protein Structure*. Springer-Verlag, New York, Heidelberg and Berlin.
Sibanda,B.L. and Thornton,J.M. (1985) *Nature*, **316**, 170–174.
Smith,D.S., Telfer,W.H. and Neville,A.C. (1971) *Tissue Cell*, **3**, 477–498.
Tsitilou,S.G., Regier,J.C. and Kafatos,F.C. (1980) *Nucleic Acids Res.*, **8**, 1987–1997.
William,R.W. and Dunker,A.K. (1981) *J. Mol. Biol.*, **152**, 783–813.

Received on May 9, 1988; revised on July 15, 1988