J. Mol. Biol. (1985) **186**, 583–589

Amino Acid Periodicities and Their Structural Implications for the Evolutionarily Conservative Central Domain of some Silkmoth Chorion Proteins

Stavros J. Hamodrakas, Thanos Etmektzoglou

Department of Biochemistry, Cell and Molecular Biology and Genetics
University of Athens, Panepistimiopolis Kouponia
Athens 15701, Greece

and Fotis C. Kafatos

Department of Cellular and Developmental Biology
Harvard University, 16 Divinity Avenue
Cambridge, MA 02138, U.S.A.
and
Institute of Molecular Biology and Biotechnology and Department of Biology
University of Crete, Heraclio, Crete, Greece

(Received 5 March 1985)

The central domain is an evolutionarily conservative region that is invariant in length in the A and Hc-A families of silkmoth chorion proteins. This domain shows strong sixfold periodicities for various amino acid residues, such as glycine and large non-polar residues. The periodicities and their phase relationships, together with the documented prevalence of β-sheets and β-turns in the chorion, strongly support a secondary structure model in which short (4-residue) β-sheet strands alternate with β-turns, forming a compact antiparallel, probably twisted β-sheet. This structure should be important for the establishment of higher order structure in the chorion.

1. Introduction

The silkmoth 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; Reiger & Kafatos, 1985; Goldsmith & Kafatos, 1984).

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 & 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 (e.g. see Regier *et al.*, 1978; Jones & 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 (Regier *et al.*, 1983; Rodakis & Kafatos, 1982; Rodakis *et al.*, 1984; Iatrou *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.*, 1982*a*; 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, amino and carboxy-terminal domains or "arms" are more variable, and are marked by the presence of obvious, tandemly repetitive peptides that are not apparent in the central

0022-2836/85/230583-07 $03.00/0

© 1985 Academic Press Inc. (London) Ltd.

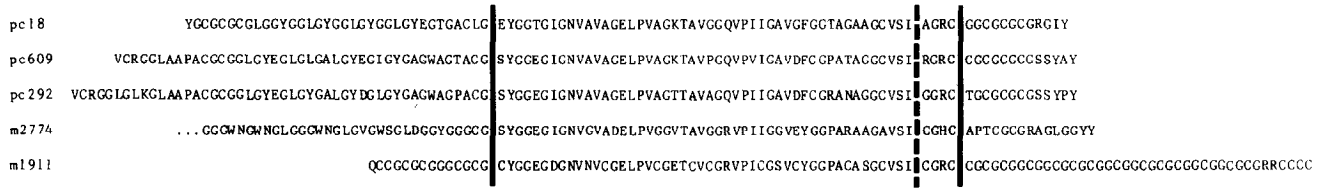

Figure 1. Sequence of typical A and Hc-A proteins determined from complementary DNA clones: pc18, pc609 and pc292 are members of the A family from *A. polyphemus* (Jones & Kafatos, 1982), m2774 is an A component from *B. mori* (Rodakis *et al.*, 1982) and m1911 is an Hc-A component from *B. mori* (Rodakis & Kafatos, 1982). Unbroken vertical lines flank the portions of the sequence considered in this paper, and a broken line indicates a slightly different border for the previously defined central domain (Jones & Kafatos, 1982). Dots indicate that the m2774 sequence is incomplete.

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.*, 1982*b*, 1983, 1984).

Ultrastructurally, the chorion consists of fibrous layers parallel to the chorion surface (Smith *et al.*, 1971; Kafatos *et al.*, 1977; Regier *et al.*, 1982; Mazur *et al.*, 1982). 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 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 the proteins.

In this paper, we present a detailed analysis of central domain sequences from the A/Hc-A branch of the chorion superfamily. These sequences can be accommodated in a compact, anti-parallel β-sheet structure, which we suggest as the starting point for understanding the molecular and supramolecular assembly of the chorion.

2. Materials and Methods

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

Pattern strengths for sequences arranged in rows of 6 residues were calculated by the method of McLachlan (1977). Fourier transforms were obtained essentially as outlined by McLachlan & Stewart (1976), using a computer program kindly provided by Dr A. D. McLachlan and suitably modified for use in the CDC 6600 Computer of the Greek Ministry of Agriculture. 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 G residue) was or was not satisfied. To increase resolution, this array was embedded in a larger array of zeros (McLachlan, 1977). Partial Fourier inversion was also performed as suggested by Bear *et al.* (1978).

3. Results

(a) *Definition of the sequences considered*

Figure 1 presents five typical chorion sequences of the A and Hc-A families. The region defined as the central domain is indicated, as is the nearly identical region considered in the present analysis. The flanking arm sequences were not analysed, since they are much more variable and, by inspection, are seen not to conform to the patterns observed in the central region. We present the analysis of five sequences representative of the central domain in the *Bombyx mori* Hc-A family, and in the A family of both *A. polyphemus* and *B. mori*; our conclusions apply to all other available sequences of these families, published and unpublished.

(b) *Hexad periodicities*

The region considered is 52 residues long; it is highly conserved among A and Hc-A proteins, and has not undergone deletions or insertions, over more than 50 million years (Rodakis *et al.*, 1982). Preliminary analysis of periodicities within it were undertaken by a computer program, revealing a sixfold periodicity for G†, and two overlapping sixfold periodicities for V. Accordingly, sequences were written out in rows of six residues (Fig. 2(a)), and the significance of the non-random distribution of residues in the six columns thus generated was analysed according to McLachlan (1977), by 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

† Amino acid residues are listed throughout according to the IUPAC-IUB 1-letter code as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; Y, Tyr; V, Val; W, Trp.

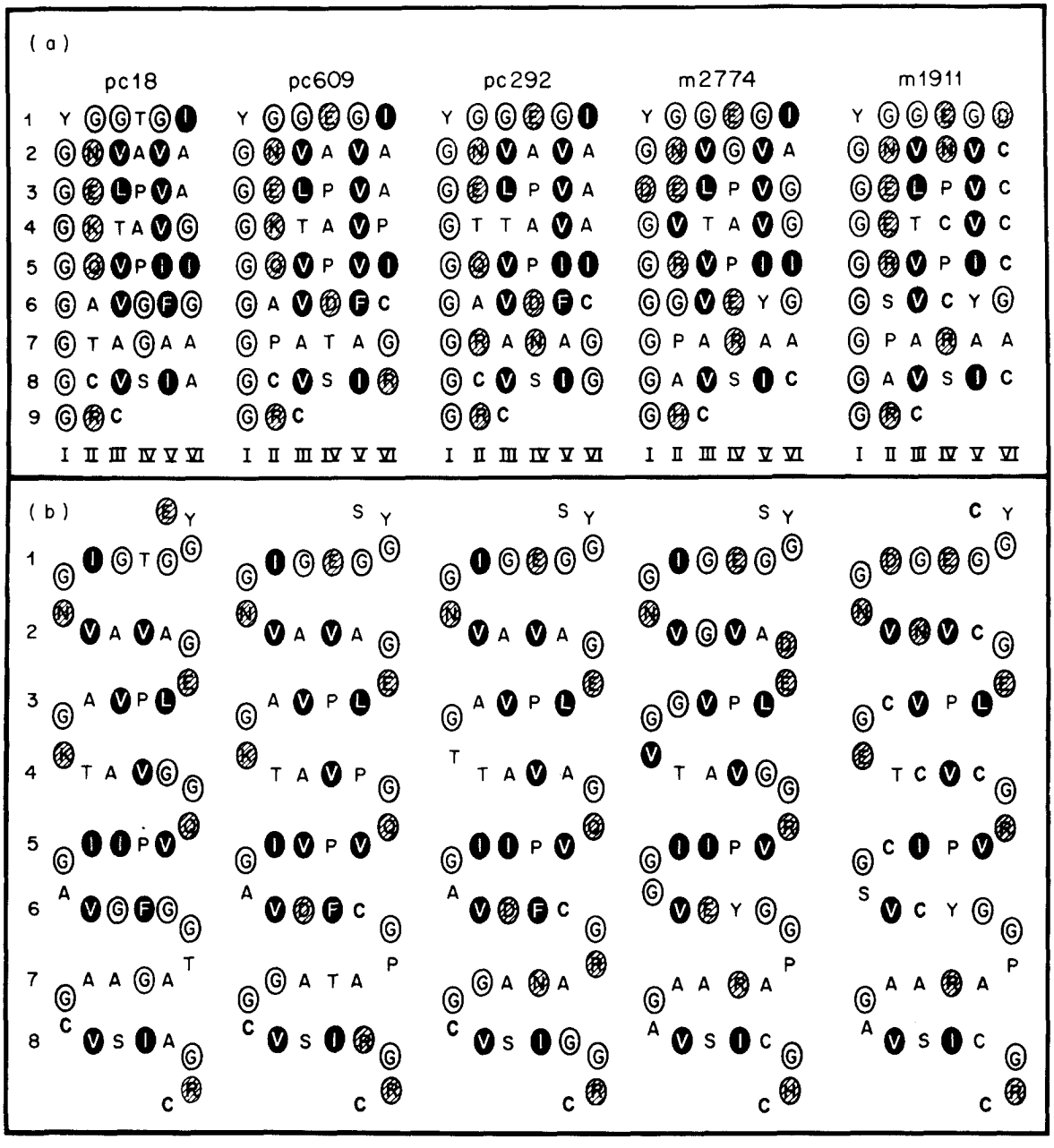

Figure 2. (a) Regular amino acid distribution within the central domain. To reveal the sixfold periodicities, sequences have been written in rows (numbered 1 to 9) of 6 residues each; they should be read left to right, top to bottom. Vertical columns (numbered I to VI) can thus be seen to have non-random prevalence of certain types of residues, which are distinctively marked (see also Fig. 3). (b) Anti-parallel β-sheet model for the central domain. Sequences should be read continuously, beginning at the top. Each row of (a) corresponds to a vertical β-turn and a similarly numbered horizontal β-sheet strand. For further details, see the text.

standard deviation. Therefore, values of P greater than 3·0 are highly significant. The analysis established clear sixfold periodicities for various types of residues, which may be summarized in terms of four classes.

(1) G.
(2) Charged or large polar residues, R, K, H, D, E, N, Q.
(3) Large non-polar residues, V, L, I, F.
(4) Miscellaneous, S, T, Y, A, P.

The pattern strengths for the first three classes are highly significant (ranges of P values for the various proteins 3·38 to 7·75, 2·87 to 7·98 and 3·75 to 6·55, respectively). In particular, the first column is occupied almost exclusively by class 1 (G), the second and fourth columns largely by classes 2 and 4, and the third and fifth columns largely by class 3. The sixth column is somewhat more variable. Inspection of the sequences (Fig. 2(a)) reveals that the first row is the most different from the rest. If that row is omitted, the pattern is even more striking.

Similar periodicities were detected by Fourier transform analysis. Both the Fourier transforms and the pattern strengths scored the periodicities of G and V residues as significant with especially high

Table 1

Residue periodicities in the central domain

Pattern strength analysis

Type of residue	pc18	pc609	pc292	m2774	m1911
G	9·02	12·60	11·41	6·34	12·60
β-Turn†	6·62	6·70	6·17	4·86	9·46
V	4·07	4·82	4·07	2·97	4·07
β-Sheet‡	7·19	6·11	6·48	5·97	4·94

Fourier transform analysis

Type of residue	pc18	pc609	pc292	m2774	m1911
G	3·48	6·41	7·61	5·07	6·60
	(98°)	(113°)	(110°)	(114°)	(111°)
β-Turn†	3·11	4·55	4·49	2·88	3·31
	(144°)	(131°)	(139°)	(115°)	(109°)
V	3·40	3·43	3·40	2·36	3·20
	(−86°)	(−75°)	(−87°)	(−96°)	(−87°)
β-Sheet‡	4·04	4·49	5·11	3·23	7·76
	(−47°)	(−51°)	(−44°)	(−50°)	(−45°)

Values for rows 2 to 9 in Fig. 2(a).

Top: Pattern strength (P) values are shown for a periodicity of 6 residues; values are expressed in standard deviation units.

Bottom: Fourier transforms for a periodicity of 5·82 residues (a value giving consistent maxima at approximately 6 residues). Each entry includes an intensity (I) value and a phase angle in parentheses. I values are scaled as recommended by McLachlan & Stewart (1976); the probability of observing by chance an intensity, I, at any particular periodicity is $\exp(-I)$.

† β-Turn formers, G, P, D, N, S, C, K, W, Y, Q, T, R, E; Chou & Fasman (1978).

‡ β-Sheet formers, V, L, I, F, W, Y, T, C; Chou & Fasman (1978).

consistency (Table 1). Furthermore, the Fourier method showed that periodicities for G and V (or V, L, I, F) are strongly out of phase (Table 1; see also Fig. 3).

(c) *Interpretation of hexad periodicities*

Predictions of secondary structure in the central region, based on a wide variety of methods, consistently indicated the prevalence of β-sheet and β-turn structure (Hamodrakas *et al.*, 1982*a*), in agreement with laser Raman and X-ray diffraction data (Hamodrakas *et al.*, 1982*b*, 1983, 1984). Because of uncertainties about the laser Raman signature of β-turns, and their relative resistance to correct prediction (a consequence of their sensitivity to non-local interactions), the prevalence of β-turns was initially reported as conjectural. However, there is a method (Williams & Dunker, 1981) that can yield quantitative estimates of the proportion of six types of secondary structure from the profile of the amide I band in laser Raman spectra. When this method was used to reanalyse our previously published spectra between 1630 cm^{-1} and 1700 cm^{-1}, it clearly indicated that in the chorion as a whole, β-sheets are anti-parallel and quantitatively dominant (encompassing 57 to 72% of the residues), and that β-turns are also prevalent (18 to 31%); the extent of random coil or α-helical structure appeared limited (see also Hamodrakas *et al.*, 1984).

These features cannot be attributed firmly to the central domain of the A and Hc-A families, since that domain accounts for only about 20% of the total chorion mass. However, the estimates in combination with the original predictions and the significant similarities between A-like and B-like chorion proteins (Jones *et al.*, 1979; Tsitilou *et al.*, 1980; Hamodrakas *et al.*, 1982*a*; Regier *et al.*, 1983; Regier & Kafatos, 1984) suggested that the observed periodicities in the central domain might correspond to periodic β-turn and β-sheet structures. In this connection, it was noted that the out-of-phase G and V (or V, L, I, F) residues have high propensities for β-turn and β-sheet structures, respectively.

Periodicities were again calculated, by both the pattern strength and the Fourier transform methods, for two groups of residues, known to be β-turn and β-sheet formers, respectively. As shown in Table 1, strong periodicities were observed for both groups, with a period of approximately six residues, and a phase difference of approximately 180°. Thus, we were led to the alternating β-turn/β-strand model shown in Figure 2(b). In Figure 2(b), the horizontal rows, of four residues each, represent anti-parallel β-sheet strands (but see Discussion), and consist of the residues shown in columns III to VI in Figures 2(a) and 3. The end-residues of these short strands also participate in β-turns, together with the vertically displayed dipeptides (columns I and II of Figs 2(a) and 3); the latter represent the central residues ($i+1$ and $i+2$, respectively) in the β-turns. This model is considered further in the Discussion.

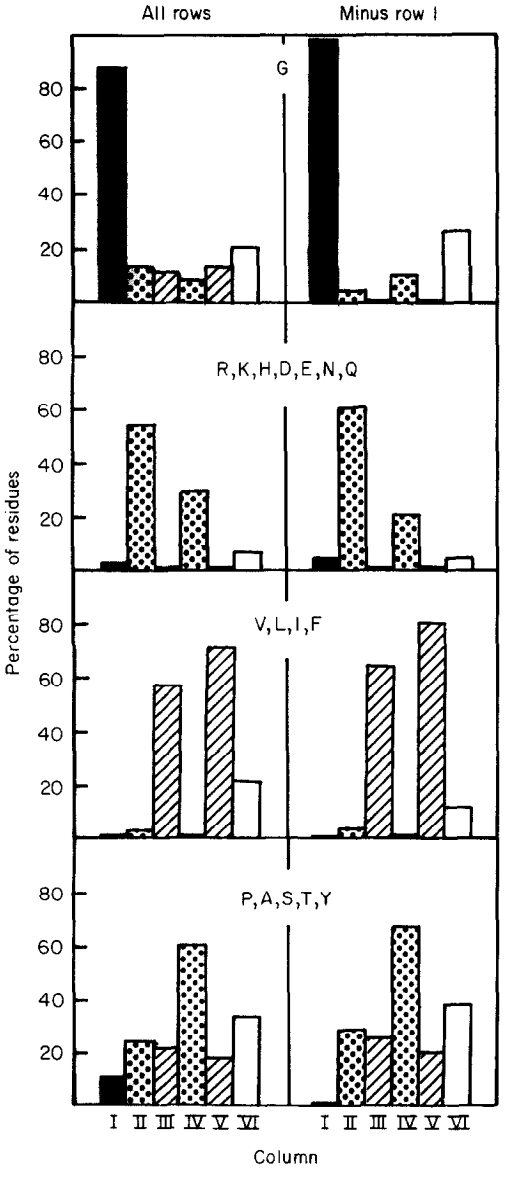

Figure 3. Histograms of the regular amino acid distribution within the central domains of all 5 sequences shown in Fig. 1. Percentages of residues were calculated for the 6 columns of Fig. 2(a), with or without the entries of row 1. Different shading is used for each column. Column I is largely G, the 3rd and 5th columns are occupied mostly by large non-polar residues (V, L, I, F), and the 2nd and 4th columns are mostly occupied by charged or large polar residues (R, K, H, D, E, N, Q), or by the P, A, S, T, Y class of residues. Column VI is the most variable.

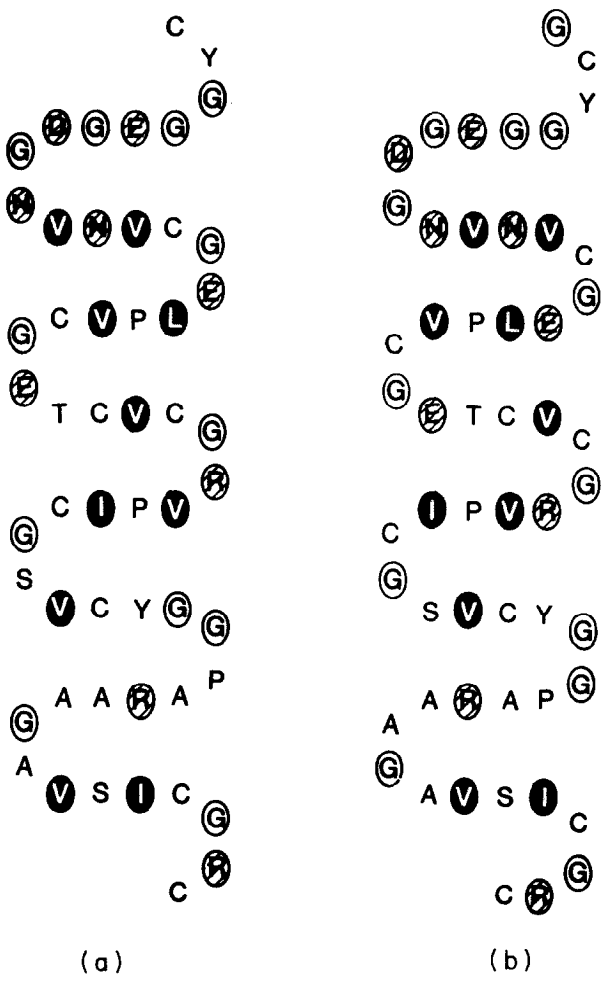

Figure 4. Anti-parallel β-sheet models for the central domain, differing by the exact locations of the β-turns. The model on the left is identical to those of Fig. 2(b), and shows the G residues (column I of Figs 2(a) and 3) in position $i+1$ of the β-turns. The model on the right shows the G residues in position $i+2$ of the β-turns. In both models, large hydrophobic side-chains occur in both faces of the sheet (see Discussion). This feature may be important for hydrophobic packing of the β-sheets.

4. Discussion

The evolutionary conservation and length invariance of the central domain of A and Hc-A chorion proteins suggest that this domain assumes a precise and functionally important three-dimensional structure. Furthermore, its internal periodicities suggest the existence of a corresponding repetitive substructure. The nature of the periodic residues and their phase relationships, together with the documented prevalence of anti-parallel β-sheet and β-turn structures in the chorion as a whole, lead to the structural model presented in Figure 2(b). In that model, short (4-residue) anti-parallel β-sheet strands are connected by β-turns; the end-residues of the turn (i and $i+3$) also belong to the β-sheet strands, thus accounting for the sixfold periodicity. The model is somewhat reminiscent of Silver Gull feather keratin (Fraser & McRae, 1976), where an eightfold periodicity in β-sheet propensities and a similar but out-of-phase periodicity of random coil propensities reflects the structure of a β-sheet consisting of short anti-parallel strands. Similarly, the model is reminiscent of the cross-β-sheet structure in the shaft of the adenovirus fibre protein (Green *et al.*, 1983); in that case, the anti-parallel β-sheet strands are usually five or six residues long and, together with two β-turns that punctuate them, result in an observed 15-fold periodicity.

A structure resembling the model in Figure 2(b) seems almost inevitable, in view of all the evidence. The issues that require further discussion are: (1) the conformation of individual strands; (2) the type and precise location of the turns; and (3) the overall properties of the domain, including possible supersecondary structures and intermolecular packing.

The strands numbered 2, 4, 5, 6, and 8 in Figure 2(b) most clearly have β-sheet conformation. They are predicted as such by a variety of methods, without reference to the periodicities discussed here (Hamodrakas *et al.*, 1982*a*). Furthermore, they show an alternating pattern characteristic of many β-sheets, namely large hydrophobic residues (V, I, F) alternating mostly with smaller non-polar or uncharged polar residues (A, P, C, G, S, T, N; occasionally with D, E, R).

Strand 3 most probably also has a β-sheet conformation. In the original predictions (Hamodrakas *et al.*, 1982*a*), an α-helical structure was predicted for this segment, interrupted or destabilized because of the P contained therein. However, we note that this segment also consists of large hydrophobic residues (V, L), alternating with smaller non-polar or uncharged polar residues (A, P, C, G). We suggest that this case represents a failure of the usual predictive methods, caused largely by the proximity of two residues with low β-sheet propensities (P and A), and of a residue (L) that is usually found in α-helical rather than β-sheet conformation. We note that both P and A are found, individually, in other strands of the same molecule, which are unambiguously in β-sheet conformation (Fig. 2(b)).

The conformations of strands 1 and 7 are more problematical. For strand 1, predictive methods suggest the existence of a β-turn, or overlapping β-turn and β-sheet structures (Hamodrakas *et al.*, 1982*a*). Strand 7 is not predicted as β-sheet (Hamodrakas *et al.*, 1982*a*), and does not show the same alternating pattern of residues that was noted above. However, in this case A residues are invariably found in the positions where V, I or F residues would have been expected. Thus, we can describe the pattern of this strand as small hydrophobic (A) rather than large hydrophobic (V, I, F) residues alternating with A, G, T, N, R residues, thus relating it to the pattern of the unambiguously β-sheet strands. It should be recalled that, although A has a low β-sheet propensity in globular proteins, the β-sheet strands of fibroin exhibit frequent alternation of A and G residues. In summary, we suggest that strand 7 also has β-sheet structure, although it is possible to postulate a coiled or even α-helical conformation, resulting in a βξβ supersecondary structure (two parallel β-strands, 6 and 8, connected by a non-β crossover connection).

Although it seems clear that in the central domain β-turns alternate with the β-sheet strands, approximately as shown in Figure 2(b), neither the exact location nor the type of the β-turns can be specified. Two alternative locations are possible for each turn, related by a one-residue translation (Fig. 4). In the first alternative exemplified by Figures 2(b) and 4 (left), the nearly invariant G residues (column I in Fig. 2(a)) occupy the second position in each turn, and the charged or large polar residues (column II in Fig. 2(a)) occupy the third position. In the second alternative exemplified by Figure 4 (right), these residues are shifted by one, to the third and fourth positions, respectively. Overlapping alternatives for turns are not uncommon in protein structure, and a correct choice between them is not always easy.

Using a data base of 408 β-turns in 29 proteins resolved at atomic dimensions, Chou & Fasman (1979) proposed two criteria for correctly predicting four contiguous residues as β-turns: (1) the β-turn potential, P_t (a measure of how frequently these residues occur in β-turns, relative to all residues found in β-turns) should have a value greater than 1·00; and (2) the β-turn probability at that location, p_t (a measure of how frequently these four residues occur in β-turns, relative to their total occurrence in the proteins of the data base) should be equal or greater than $0{\cdot}75 \times 10^{-4}$. Although in overall terms the probability criterion favours the second alternative ($0{\cdot}74(\pm 0{\cdot}10) \times 10^{-4}$ for all turns in all 5 chorion proteins *versus* $0{\cdot}36(\pm 0{\cdot}04) \times 10^{-4}$ for the first alternative), we are hesitant about a definitive choice. The two choices are essentially equivalent with respect to β-turn potential, with most individual turns meeting the criterion of $P_t > 1{\cdot}0$. Furthermore, the probability criterion favours the first alternative for the Hc-A protein, m1911, and the second for the proteins of the A family. The first alternative is supported also by preliminary model building (unpublished results): it corresponds to a β-sheet with surfaces of more pronounced hydrophobicity than the sheet corresponding to the second alternative. Hydrophobicity should facilitate packing of the sheets in higher-order structures.

It is well-known that various types of β-turns are possible, and some are substantially different than others in conformational terms, and in preference for specific residues in specific positions (Chou & Fasman, 1978). For example, in type I′, II and III′ turns, G tends to be located in the third position, whereas in type I, II′ or III turns, G is usually found in the second position. The Chou & Fasman (1979) criteria were developed with a data base of a variety of turns. Thus, model building and physical measurements on synthetic peptides representative of the central domain will be necessary for choosing between the alternative structures shown in Figure 4, as well as for specifying the prevalent type of β-turn.

The chorion structure is finalized at the end of morphogenesis by the formation of disulphide bonds (Blau & Kafatos, 1979). Although most cysteine residues are found near the protein termini, in Hc-A proteins such as m1911 they are prevalent also in the central domain. In our models, it is

notable that the cysteine residues of the central domain tend to be found at the ends of the β-strands or within the β-turns, i.e. in positions that are accessible from both sides of the β-sheet and favourable for formation of disulphide bridges.

With respect to the overall structure of the central domain and its possible interactions with other chorion domains and proteins, little can be said with confidence. One notable aspect is that, in most reasonable models, both faces of the proposed β-sheets would include large hydrophobic residues; that can be seen from Figures 2(b) and 4, bearing in mind that in the common types of turns, the side-chains of the end residues (1 and 4) point towards the same face, and that in β-sheet strands successive side-chains point towards opposite faces. A second significant possibility is that the proposed β-sheets may be twisted, and that the twist may be related to the well-known helicoidal architecture of the silkmoth chorion (Mazur *et al.*, 1982; Regier *et al.*, 1980, 1982; Hamodrakas, 1984). Twisted β-sheets are common in both globular proteins (Chothia, 1973) and fibrous proteins (Lotz *et al.*, 1982). It should be recalled that in twisted β-sheets successive strands form an angle of approximately 20° (Chothia & Janin, 1981), leading to a left-handed helical structure when the sheet is viewed along a direction perpendicular to the β-strands (Chothia, 1973). The possibility that the proposed sheets are folded into a β-barrel (Richardson, 1977) is not very likely, because of the small number of residues per strand and the short length of the interstrand connections (two residues). The significance of the possible twist, and the interactions of residues on the two faces of the sheet with other chorion structures, can be evaluated only after a more specific structural model is developed.

We are grateful to Dr A. D. McLachlan for help and for providing us with the Fourier transform program. We thank the Greek Ministry of Agriculture Computer Laboratory for computing facilities and the Greek Ministry of Research and Technology and the NIH for financial support. Part of this work was made possible by an EMBO short-term fellowship to S.J.H.

References

Bear, R. S., Adams, J. B. & Poultoy, J. W. (1978). *J. Mol. Biol.* **118**, 123–126.

Blau, H. M. & Kafatos, F. C. (1979). *Develop. Biol.* **72**, 211–225.

Bouligand, Y. (1972). *Tissue Cell*, **4 (2)**, 189–217.

Chothia, C. (1973). *J. Mol. Biol.* **75**, 295–302.

Chothia, C. & Janin, J. (1981). *Proc. Nat. Acad. Sci., U.S.A.* **78**, 4146–4150.

Chou, P. Y. & Fasman, G. D. (1978). *Advan. Enzymol.* **47**, 45–148.

Chou, P. Y. & Fasman, G. D. (1979). *Biophysics*, **26**, 385.

Eickbush, T. H. & Kafatos, F. C. (1982). *Cell*, **29**, 633–643.

Fraser, R. D. B. & McRae, T. (1976). *Proc. 16th Int. Ornithological Congr., Canberra*, pp. 443–451, Australian Academy of Science, Canberra.

Goldsmith, M. R. & Kafatos, F. C. (1984). *Annu. Rev. Genet.* **18**, 443–487.

Green, N. M., Wrigley, N. G., Russell, W. C., Martin, S. R. & 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. & Kafatos, F. C. (1982*a*). *Biochim. Biophys. Acta*, **700**, 42–51.

Hamodrakas, S. J., Asher, S. A., Mazur, G. D., Regier, J. C. & Kafatos, F. C. (1982*b*). *Biochim. Biophys. Acta*, **703**, 216–222.

Hamodrakas, S. J., Paulson, J. R., Rodakis, G. C. & Kafatos, F. C. (1983). *Int. J. Biol. Macromol.* **5**, 149–153.

Hamodrakas, S. J., Kamitsos, E. I. & Papanikolaou, A. (1984). *Int. J. Biol. Macromol.* **6**, 333–336.

Iatrou, K., Tsitilou, S. G. & Kafatos, F. C. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 4452–4456.

Jones, C. W. & Kafatos, F. C. (1980). *Cell*, **22**, 855–867.

Jones, C. W. & Kafatos, F. C. (1982). *J. Mol. Evol.* **19**, 87–103.

Jones, C. W., Rosenthal, N., Rodakis, G. C. & Kafatos, F. C. (1979). *Cell*, **18**, 1317–1332.

Kafatos, F. C. (1983). In *Gene Structure and Regulation in Development*. 41st Symp. Soc. Devel. Biol. (Subtelny, S. & Kafatos, F. C., eds), pp. 33–61, Alan R. Liss, Inc., New York.

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). In *Results and Problems in Cell Differentiation* (Beermann, W., ed.), vol. 8, pp. 45–145, Springer-Verlag, Berlin.

Lotz, B., Gouthier-Vassal, A., Brack, A. & Magoshi, J. (1982). *J. Mol. Biol.* **156**, 345–357.

Mazur, G. D., Regier, J. C. & Kafatos, F. C. (1982). In *Insect Ultrastructure* (Akai, H. & King, R. C., eds), vol. 1, pp. 150–183, Plenum Publishing Corp., New York.

McLachlan, A. D. (1977). *Biopolymers*, **16**, 1271–1297.

McLachlan, A. D. & Stewart, M. (1976). *J. Mol. Biol.* **103**, 271–298.

Regier, J. C. & Kafatos, F. C. (1985). In *Embryogenesis and Reproduction. Comprehensive Insect Physiology, Biochemistry and Pharmacology* (Kerkut, G. A. & Gilbert, L. I., eds), vol. 1, pp. 113–151, Pergamon Press, Oxford.

Regier, J. C., Kafatos, F. C., Goodfliesh, R. & Hood, L. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 390–394.

Regier, J. C., Mazur, G. D. & Kafatos, F. C. (1980). *Develop. Biol.* **76**, 286–304.

Regier, J. C., Mazur, G. D., Kafatos, F. C. & Paul, M. (1982). *Develop. Biol.* **92**, 159–174.

Regier, J. C., Kafatos, F. C. & Hamodrakas, S. J. (1983). *Proc. Nat. Acad. Sci., U.S.A.* **80**, 1043–1047.

Richardson, J. D. (1977). *Nature (London)*, **268**, 495–500.

Rodakis, G. C. & Kafatos, F. C. (1982). *Proc. Nat. Acad. Sci., U.S.A.* **79**, 3551–3555.

Rodakis, G. C., Moschonas, N. K. & Kafatos, F. C. (1982). *Mol. Cell Biol.* **2**, 554–563.

Rodakis, G. C., Lecanidou, R. & Eickbush, T. H. (1984). *J. Mol. Evol.* **20**, 265–273.

Smith, D. S., Telfer, W. H. & Neville, A. C. (1971). *Tissue Cell*, **3**, 477–498.

Tsitilou, S. G., Regier, J. C. & Kafatos, F. C. (1980). *Nucl. Acids Res.* **8**, 1987–1997.

William, R. W. & Dunker, A. K. (1981). *J. Mol. Biol.* **152**, 783–813.

Edited by R. Huber