Structural and functional features of *Drosophila* chorion proteins s36 and s38 from analysis of primary structure and infrared spectroscopy

Stavros J. Hamodrakas*, Anthimia Batrinou and Tania Christophoratou

Department of Biochemistry, Cell and Molecular Biology and Genetics, University of Athens, Athens 157.01, Greece

(Received 10 January 1989; revised 8 May 1989)

Amino acid composition, Fourier transform analysis and secondary structure prediction methods strongly support a tripartite structure for Drosophila *chorion proteins s36 and s38. Each protein consists of a central domain and two flanking 'arms'. The central domain contains tandemly repetitive peptides, which apparently generate a secondary structure of β-sheet strands alternating with β-turns, most probably, forming a twisted β-pleated sheet or β-barrel. The central domains of s36 and s38 share similarities, but they are recognizably different. The flanking 'arms', with different primary and secondary structure features, presumably serve protein-specific functions. The possible roles of the protein domains for the establishment of higher order structure in* Drosophila *chorion and the possible function of the molecules are discussed. The predicted secondary structure of* Drosophila *chorion proteins s36 and s38 is supported by experimental information obtained from Fourier transform infrared spectroscopic studies of* Drosophila *chorions.*

Keywords: Drosophila; chorion protein structure; secondary structure prediction; Fourier analysis; infrared spectroscopy

Introduction

The *Drosophila* eggshell or chorion has been studied extensively both as a model system for the study of programmed, differential gene expression during development (Ref. 1 and references therein), and also in order to understand its morphogenesis and structure-function relationship[2–4].

The follicle cells of *Drosophila melanogaster* produce the structural proteins of the eggshell (chorion) according to a precise spatial and temporal programme. At the end of oogensis, the chorion proteins are synthesized by the follicular epithelial cells and secreted onto the surface of the oocyte, where they assemble to form the multilayered eggshell. A set of six major (s15, s16, s18, s19, s36 and s38; numbers indicate approximate molecular weights in kD) and more than 14 minor chorion proteins can be resolved by two-dimensional gel electrophoresis; subsets of these proteins are expressed in a temporally regulated mode during the 5 h of choriogenesis[5–7].

The sequences of all major chorion proteins have been determined: the lower molecular weight s15, s16, s18, s19, encoded by genes tightly clustered on the third chromosome, and s36, s38 by genes clustered on the X chromosome[8–10]. These genes are products of specific DNA replication (gene amplification). Proteins s15, s16, s18 and s19 show obvious homologies in their primary structure. This is also true for the s36 and s38 proteins. However, the amino acid sequences of s36 and s38 exhibit unique features not shown by s15, s18 and s19. Recent observations by Margaritis and collaborators[11] suggest that, in addition to its structural role, s38 might play a functional role, possibly cross-linking *Drosophila* chorion proteins by di-tyrosine and tri-tyrosine bonds during the late choriogenetic stages.

In this report we present characteristic structural features of the higher molecular weight major proteins s36 and s38, emerging from analysis of their sequences, which apparently dictate the way(s) the proteins are serving as structural and functional elements of *Drosophila* chorion. We also provide evidence indicating agreement between secondary structure prediction of proteins s36, s38 and experimental infrared spectroscopy data.

Experimental

Fourier transforms were obtained essentially as outlined by MacLachlan[12], using a Fortran 77 computer program. 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 array was embedded in a larger array of zeros[13].

The methods used for secondary structure prediction have been described in detail by Hamodrakas *et al.*[14] and Hamodrakas and Kafatos[15]. They have been developed into a fully computerized prediction scheme, which runs on the IBM PC/XT/AT and compatibles, under DOS 2.0 or later releases[16].

Drosophila melanogaster (Oregon-R) flies conditioned

* To whom correspondence should be addressed.

0141-8130/89/050307-07$03.00

at 25°C were lightly etherized after 2–4 days in culture to ensure that all developmental stages would be found in the ovaries. After dissection in distilled water, late stage 14 individual follicles were selected under a high power Zeiss stereomicroscope using fibre optics. The follicles were cut in half using fine forceps and washed several times in distilled water to remove the vitelline membrane and the remnants of the oocyte and the follicular epithelium. The samples used for infrared spectroscopy experiments were composed of approximately 500 individual chorions, thoroughly dried.

Infrared spectra were recorded on a Fourier Transform Bruker 113v, vacuum spectrometer. Each spectrum is the result of signal averaging of 100 scans at $2\,cm^{-1}$ resolution. Samples were in the form of KBr pellets, containing about 2% (w/w) material, which was thoroughly ground in a vibrating mill, before mixing with KBr.

Attempts to obtain laser-Raman spectra of chorions utilizing various spectral lines have failed: the samples showed a strong fluorescent background, masking completely the Raman signal even after prolonged laser irradiation.

Results

The amino acid sequences of *Drosophila* chorion proteins s36 and s38, shown in *Figure 1*, were obtained from the sequences of the genes encoding them[9].

The amino acid composition varies in internal regions within each protein (*Table 1*). A central domain rich in Lys, Pro and Val differs from two flanking 'arms' rich in Gly, Ala and Ser. The central domain in s36 extends from residue 99 to 222 and in s38 from residue 97 to 195. We arbitrarily define the central domain by the first and the last appearance of Lys in both cases.

In the central domains of s36 and s38 Pro is observed to have a non-random distribution, as well as Val and Lys. Fourier transform analysis detects a seven-residue periodicity of high intensity for Pro in both proteins and of lower intensity periodicities for Val and Lys only in s38 (*Table 2*). The same analysis also reveals a 7.6-fold periodicity in s38, for β-turn former residues (G, P, D, N, S, C, K, W, Y, Q, T, R, E; Ref. 17) and for β-sheet formers (V, L, I, F, W, Y, T, C; Ref. 17) which are out of phase. In s36 a nine-residue periodicity for β-turn formers and β-sheet formers is also out of phase.

Prediction of secondary structure shows the prevalence of short β-sheet strands alternating regularly with β-turns in the central domain (*Figure 2*). This is clear for protein s38, but less evident for s36. Therefore, Fourier analysis and prediction methods suggest that the central domains of s36 and s38 are composed of the following patterns of repetitive peptides (*Figure 3a*). These imprecise repeats appearing in the central domain of s36 and s38 can best be interpreted by the alternating β-turn or loop/β-strand model of an antiparallel β-pleated sheet shown in *Figure 3b*.

α-Helices are also predicted in the N- and C-terminal 'arms' of both s36 and s38. Apparently, they are abundant in the amino-terminal segment of s36. The remainder of the molecules are predicted to adopt random coil structure or β-turns (the methods do not discriminate between these two types of secondary structure; see also Refs. 14–16).

Supporting evidence for the presence of all types of secondary structure in *Drosophila* chorion proteins is given by the Fourier transform infrared spectrum of *Drosophila* chorion (*Figure 4*). Intense absorption bands at $1638\,cm^{-1}$ (amide I) and $1520\,cm^{-1}$ (amide II) are characteristic of antiparallel β-pleated sheet conformation, whereas the bands at $1653\,cm^{-1}$ (amide I) and at 1542 and $1559\,cm^{-1}$ (amide II) may indicate a considerable proportion of unordered (coil) or β-turns, or α-helical structure[18].

Discussion

The patterns emerging from the analysis of the amino acid sequences of both s36 and s38 *Drosophila* chorion

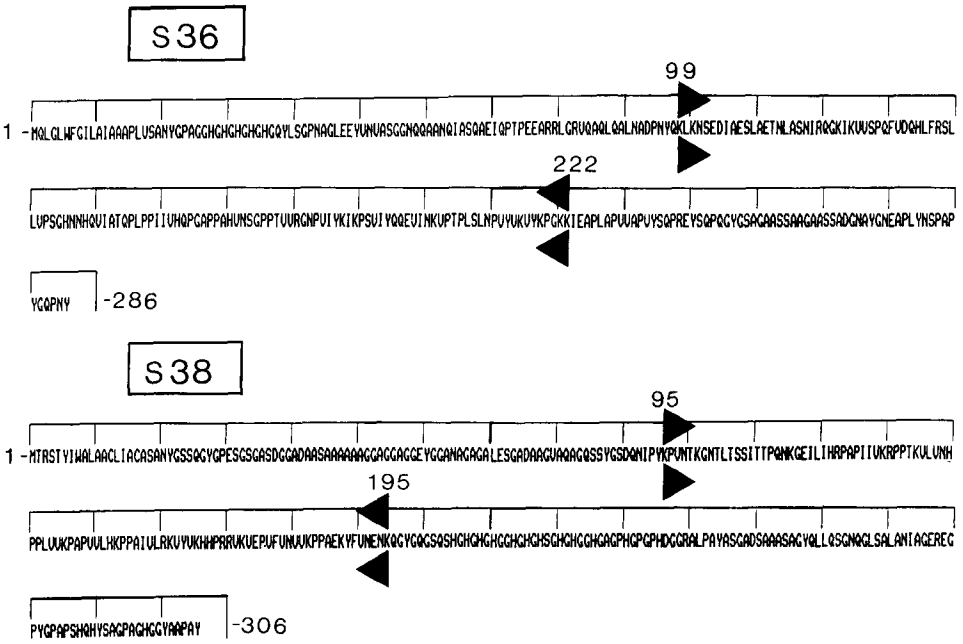

Figure 1 Amino acid sequences (one letter code) of *Drosophila* chorion proteins s36 and s38, obtained from Ref. 9. Arrows indicate the borders of the central domain in each protein (see text)

Table 1 Amino acid composition in the central domain and the flanking arms of proteins s36 and s38. Residues scoring above 6% are shown

s36

1-	Left arm	-99-	Central domain	-222-	Right arm	-286
	A: 16% G: 15% Q: 11%		V: 12% P: 13% K: 9%		A: 14% G: 10% S: 12.5% Y: 10%	
	L: 9% N: 7% P: 7%		I: 8% N: 8% L: 7% S: 7% Q: 6.5%		V: 9% P: 9%	

s38

1-	Left arm	-97-	Central domain	-195-	Right arm	-306
	A: 28% G: 22% S: 12%		V: 14.5% P: 16% K: 13%		A: 16% G: 28% S: 9% H: 13%	
	Y: 6%		N: 7% T: 6%		Y: 6% P: 8%	

Table 2 Residue periodicities in the central domain of *Drosophila* chorion proteins s36 and s38 detected by Fourier transform analysis. The probability of observing by chance an intensity, I, at any particular periodicity is exp(−I). Therefore, values of intensities greater than 3.0 are considered as significant

s38 Central domain (97-195)

Type of residue	Periodicity	Intensity	Phase angle
P	7.2	8.73	22
V	7.4	4.2	−153.8
K	3.76	4.79	39
β-turn formers	7.6	6.2	−142.3
β-sheet formers	7.6	10.3	23.5

s36 Central domain (99-222)

Type of residue	Periodicity	Intensity	Phase angle
P	6.8	5.63	163.5
β-turn formers[a]	8.98	6.23	154.2
β-sheet formers[a]	8.98	4.27	−51.9

[a] Weighted: the conformational parameters of Chou and Fasman were used as weights

proteins and from secondary structure prediction can best be interpreted as indicating that the central domain most probably forms a structure which consists of alternating β-sheet strands of three consecutive large hydrophobic residues (usually Val, Ile, Tyr or Leu) connected with turns or loops formed by two, usually consecutive, prolines and polar or positively charged residues (mostly Lys, Arg, His). The number of residues participating in each turn varies from two (normal β-turns?) to six, in s38 (*Figure 3b*). Protein s38 shows clearly a different type of structure than s36: the latter exhibits apparently less periodical primary and secondary structure features, where intervening stretches of residues disrupt the regular pattern.

This type of structure is reminiscent, in some respects, of the antiparallel β-pleated sheet structure of the central domains of silkmoth chorion proteins[13,19] and of the antiparallel β-pleated sheet structure shown by a major eggshell protein of *Schistosoma mansoni*[20]. It shares also similarities with the antiparallel β-pleated sheet ('cross-β') structure of the adenovirus fibre protein[21]. In the latter, the amphipathic nature of the sheet promotes dimerization: hydrophobic surfaces are packed together to form the shaft of the fibre, presumably composed of a dimer.

The model β-sheet structure of s36 and s38 might well be a twisted β-sheet structure or even a β-barrel generating a globular central 'core' for both proteins: most β-sheets in globular and also in fibrous proteins are twisted β-sheets (Ref. 19 and references therein). This hypothesis is further supported by the presence of two, usually consecutive, prolines in the turns connecting the β-strands, which may provide the necessary local twist to the polypeptide chain for the formation of a β-barrel structure.

It is perhaps interesting to note that in the chorion pillars, which presumably contain large amounts of both s36 and s38 proteins, secreted at the first stages of choriogenesis ('early' proteins), freeze fracturing with rotary shadowing reveals globular structures interconnected with fine fibres[7]. It is also important to observe that the crystalline in nature, innermost chorionic layer, whose structure was revealed by three-dimensional reconstruction techniques[3], is formed by globular protein domains, with a diameter of approximately 3–4 nm, connected with thinner 'arms'. These chorion proteins with apparent molecular weights of 30–40 kD[3], most probably, correspond to the 'early' proteins s36 and s38.

We estimate from physical models that the diameter of the globular β-barrels, which may correspond to the central domains of s36 and s38, is approximately of the same order of magnitude, namely 3–4 nm.

Recent work on the chorion proteins of the med fly *Ceratitis capitata*[22] indicates that the central domain is highly conserved in similar proteins found in the chorion

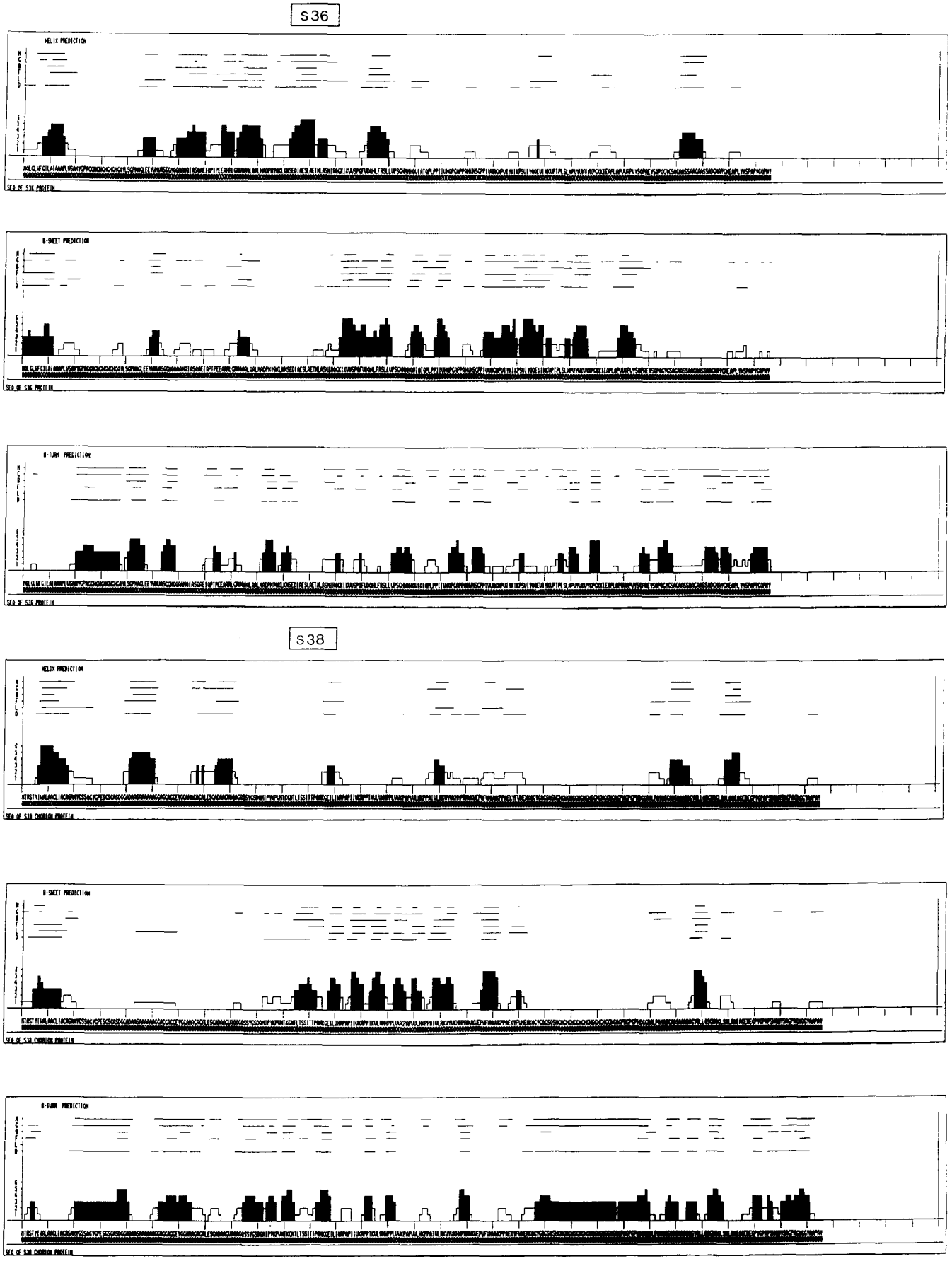

Figure 2 Secondary structure prediction plots for α-helix, β-sheet and β-turns, for *Drosophila* chorion proteins s38 and s36. Individual predictions, as derived according to Nagano (N), Garnier *et al.* (G), Burgess *et al* (B), Chou and Fasman (F), Lim (L) and Dufton and Hider (D), are shown by horizontal lines. Joint prediction histograms, constructed by tallying the individual predictions, are also shown. The most probable structures, predicted by three or more methods, are shaded. The plots are 'hard' copies on a dot matrix printer of a monitor screen

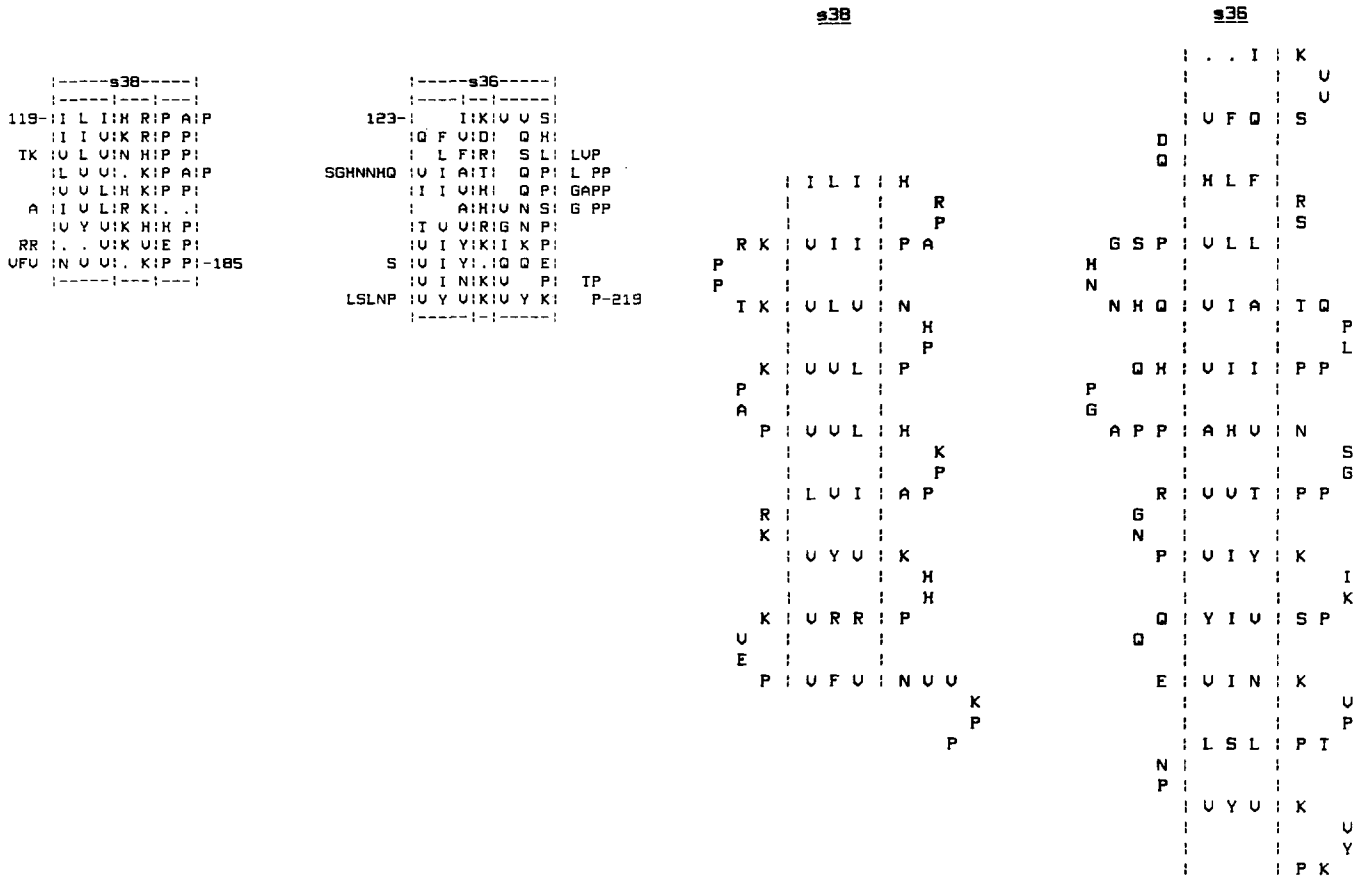

Figure 3 (a) Regular amino acid distribution within the central domain of *Drosophila* chorion proteins s38 and s36. Sequences should be read continuously, left to right, top to bottom. (b) Antiparallel β-pleated sheet model for the central domain of *Drosophila* chorion proteins s36 and s38. Sequences should be read continuously, beginning at the top. The β-strands contain three consecutive 'large' hydrophobic residues (outlined by broken vertical lines), whereas a non-random prevalence of two, frequently consecutive, prolines and of positively charged (Lys, Arg, His) or polar residues is seen in the turns or loops of the structure, which, apparently, is more regular for protein s38

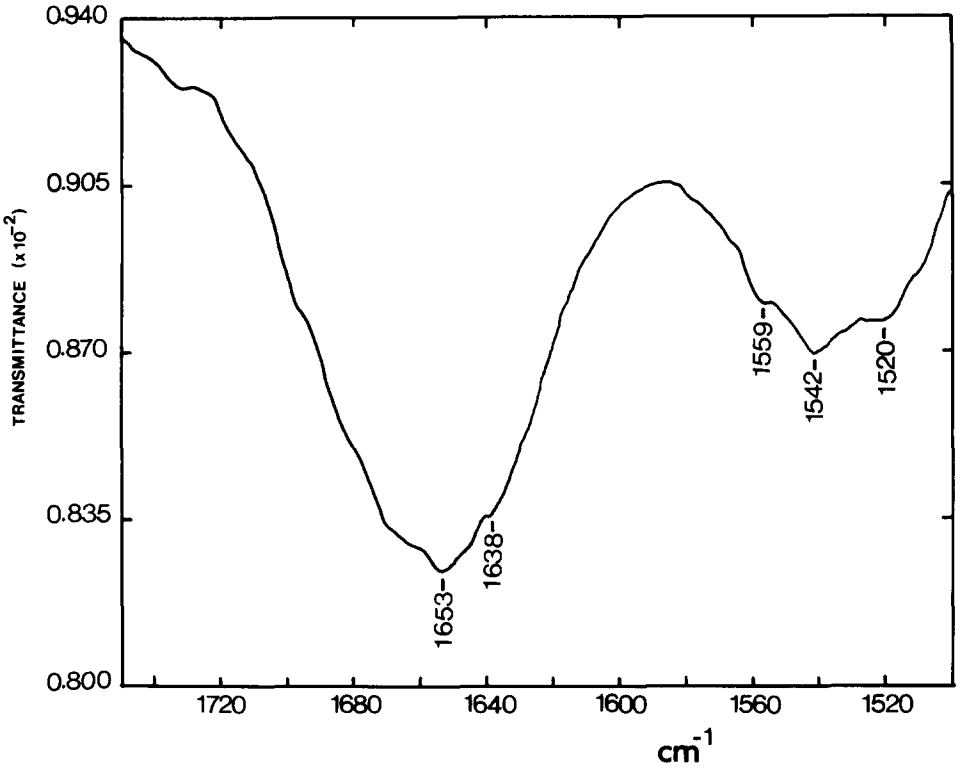

Figure 4 Fourier transform infrared spectrum of *Drosophila melanogaster* chorion. The spectrum is the result of signal averaging of 100 scans, at 2 cm^{-1} resolution. Samples were in the form of KBr pellets, containing about 2% wt material (approximately 500 chorions), thoroughly ground in a vibrating mill, before mixing with KBr

of this fly, as judged from extensive sequence homology, which perhaps suggests an important structural and/or functional role.

There is significant homology between the opposite arms of the two proteins. The left arm of s36 and the right arm of s38 both contain tandem repeats of the dipeptide Gly-His (five times in s36 and nine times in s38) which were, most probably, derived from gene duplication. Prediction indicates random coil structure or β-turns for these tandem motifs. However, since prediction methods fail for sequences that include short precise subrepeats[17], we attach no significance to the prediction for these sequences. The alternation of small (Gly) with bulky (His) residues in these stretches, reminiscent of the structure of silk fibroin[23], permits the hypothesis for an extended type of structure, with similar residues, in terms of volume, facing opposite sides of the extended structure. This possibly facilitates protein–protein interactions to form a higher order structure in chorion.

Similarly, the right arm of s36 and the left arm of s38 are characterized by long stretches of alanines, predicted as α-helices.

Both the α-helices of the long stretches of alanines and the structures of the tandem repeats of Gly-His of the protein arms, might correspond to the fibrous structures seen by freeze-fracturing[7], and the interconnecting arms of the globular domains revealed by 3D reconstruction in the crystalline innermost chorionic layer[3].

The *Drosophila* chorion undergoes a hardening process during the last developmental stages of oogenesis: this process which is accomplished through the action of a peroxidase *in vivo*, occurs by the formation of covalent crosslinks, di-tyrosine and tri-tyrosine bonds, between its constituent polypeptides[24]. Therefore, it is important to observe tyrosine distribution and localization in chorion proteins. Tyrosine is abundant in the arms of both s36 and s38, but it occurs rarely in the central domain. Apparently, the arms might participate in the formation of di-tyrosine and tri-tyrosine bonds, whereas it is further emphasized that the central domain serves for a functional and structural role. In this respect, an intriguing feature of the right arm of protein s36 (residues 235–286) is the appearance of a hexapeptide periodicity for Tyr in this region, with an intervening peptide enriched in Ala:

235- Y S Q P R E
Y S Q P Q G
Y G S A G A A S S A A G A A S S A D G N A
Y G N E A P L
Y N S P A P
Y G Q P N Y -286

Secondary structure prediction fails to indicate a regular structure for this portion of the sequence, apart from the intervening peptide, which is predicted to form an α-helix. Apparently, this portion of the protein s36 serves for a protein-specific function for the formation of di-tyrosine and tri-tyrosine bonds.

The fact that a similar region is absent for protein s38, perhaps, further signifies the different structural and functional roles of the two proteins.

Recent work by Margaritis and collaborators[11] suggests that s38 might play a functional role crosslinking chorion proteins in the late choriogenetic stages by di-tyrosine and tri-tyrosine bonds. It would, however, be premature to propose a detailed model for the functional parts of the molecule with the existing evidence.

Secondary structure predictions should always be undertaken with full awareness of their limitations, even in the case of globular proteins for which they were initially developed and applied[16,17,25]. However, comparisons of evolutionarily related sequences or of imprecise internal repeats are useful in overcoming some limitations. Limited variation reduces the 'noise' and frequently helps to identify structural features. This approach has been applied successfully by us in the case of the silkmoth chorion proteins[13,19] and also by Green *et al.* in the adenovirus fibre protein[21]. In both cases, repetitive secondary structure features have been elucidated, on the basis of internal periodicities corresponding to imprecise repeats, in conjunction with results obtained from secondary structure prediction.

The validity of our structural predictions for chorion proteins s36 and s38 was tested experimentally by the infrared spectroscopic studies of *Drosophila* chorions. The analysis of the infrared spectrum (*Figure 4*), clearly shows the presence of all types of secondary structure in *Drosophila* chorion proteins. This is in good agreement with the predictions for proteins s36 and s38 and also with the results of prediction for the lower molecular weight chorion proteins s15, s16, s18, s19 (our unpublished data): thus, although these proteins exhibit different characteristic structural and presumably functional features from s36 and s38, they appear to contain all types of secondary structure.

Nevertheless, more refined experimental and theoretical work is needed to correlate sequence, conformation and function of the various *Drosophila* chorion protein segments.

Acknowledgements

We express our gratitude to Dr E. I. Kamitsos for his invaluable help to obtain the infrared spectrum of *Drosophila* chorion.

References

1 Kafatos, F. C., Delidakis, C., Orr, W., Thireos, G., Komitopoulou, K. and Wong, Y.-C., Molecular and Developmental Biology, Alan R. Liss, Inc., New York, 1986, p. 85
2 Margaritis, L. H. in 'Comprehensive Insect Biochemistry, Physiology and Pharmacology', (Eds L. I. Gilbert and G. A. Kerkut), Pergamon Press, Oxford and New York, 1985, Vol. 1, pp. 153–230
3 Margaritis, L. H., Hamodrakas, S. J., Arad, T. and Leonard, K. R. *Biol. Cell* 1984, **52**, 279
4 Hamodrakas, S. J., Margaritis, L. H. and Nixon, P. E. *Int. J. Biol. Macromol.* 1982, **4**, 25
5 Petri, W. H., Wyman, A. R. and Kafatos, F. C. *Dev. Biol.* 1976, **49**, 185
6 Waring, G. L. and Mahowald, A. P. *Cell* 1979, **16**, 599
7 Margaritis, L. H., Kafatos, F. C. and Petri, L. H. *J. Cell Sci.* 1980, **43**, 1
8 Wong, Y. C., Pustell, J., Spoerel, N. and Kafatos, F. C *Chromosoma (Berlin)* 1985, **92**, 124
9. Spradling, A. C., De Cicco, D. V., Wakimoto, B. T., Levine, J. F., Kalfayan, L. J. and Cooley, L. *EMBO J.* 1987, **6**, 1045
10 Kafatos, F. C. Unpublished data
11 Margaritis, L. H. *et al.* In preparation
12 McLachlan, A. D. *Biopolymers* 1977, **16**, 1271
13 Hamodrakas, S. J., Ekmetzoglou, T. and Kafatos, F. C. *J. Mol. Biol.* 1985, **186**, 583
14 Hamodrakas, S. J., Jones, C. W. and Kafatos, F. C. *Biochim. Biophys. Acta* 1982, **700**, 42
15 Hamodrakas, S. J. and Kafatos, F. C. *J. Mol. Evol.* 1984, **20**, 296

16 Hamodrakas, S. J. *CABIOS* 1988, **4** (No. 4), 473
17 Chou, P. Y. and Fasman, G. D. *Adv. Enzymol.* 1978, **47**, 45
18 Parker, F. S. 'Applications of Infrared Spectroscopy in Biochemistry, Biology and Medicine', Plenum Press, New York, 1971
19 Hamodrakas, S. J., Bosshard, H. E. and Carlson, C. N. *Prot. Engineer.* 1988, **2** (No. 3), 201
20 Rodrigues, V., Chaudhri, M., Knight, M., Meadows, H., Chambers, A. E., Taylor, W. R., Kelly, C. and Simpson, A. J. G. *Mol. Biochem. Parasitol.* 1989, **32**, 7
21 Green, N. M., Wrigley, N. G., Russel, W. C., Martin, S. R. and McLachlan, A. D. *EMBO J.* 1983, **2** (No. 8), 1357
22 Komitopoulou, K. and Kafatos, F. C. Unpublished data
23 Marsh, R. E., Corey, R. B. and Pauling, L. *Biochim. Biophys. Acta* 1955, **16**, 1
24 Margaritis, L. H. *Tissue Cell* 1985, **17**(4) 553
25 Kabsch, W. and Sander, C. *FEBS Lett.* 1983, **155**, 179