Tandemly repeating peptide motifs and their secondary structure in *Ceratitis capitata* eggshell proteins Ccs36 and Ccs38

Amalia Aggeli, Stavros J. Hamodrakas* and Katia Komitopoulou

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

and Mary Konsolaki

Department of Biology, University of Crete, Heraclio 71110, Greece

(Received 18 October 1990; revised 4 April 1991)

Evidence from amino acid composition, Fourier transform analysis of primary structure and secondary structure prediction suggests a tripartite structure for Ceratitis capitata *eggshell proteins Ccs36 and Ccs38, which consists of a central domain and two flanking 'arms'. The proteins, apparently, contain tandemly repeating peptide motifs specific for each domain of the tripartite structure. The central domain of both proteins, which exhibits extensive sequence homology with the corresponding domains of* Drosophila melanogaster *proteins s36 and s38, is formed by tandem repeats of an octapeptide -X-X-X-Z-Z-Z-Z-Z- (where X = large hydrophobic residue and Z = β-turn former residue) and its variants. It is predicted to adopt a compact, most probably twisted, antiparallel β-pleated sheet structure of β-sheet strands regularly alternating with β-turns or loops. The central domains of Ccs36 and Ccs38 share structural similarities, but they are recognizably different. The 'arms' of the proteins presumably serving for protein and species-specific functions differ substantially from those of* Drosophila melanogaster. *In Ccs36, the C-terminal 'arm' is formed by, almost precise, tandem repeats of an octapeptide -Y-X-A-A-P-A-A-S- (X = G or S), whereas the N-terminal 'arm' contains repeats of the octapeptide -Z-Z-Z-A-X-A-A-Z- (X = Q, N or E and Z a β-turn former). In both 'arms' α-helices are predicted, alternating with β-turns. In the Ccs38 C-terminal 'arm' nonapeptide repeats of the form -Y-Z-Z-Z-Z-(G, A)-Z-Q-Z- (Z = A, G, P or S) are observed, whereas the N-terminal includes repeats of an octapeptide motif -x-y-G-z-G-u-G-v- (x = I, S, y = G, Q). The latter are predicted as β-sheet strands alternating with β-turns, whereas, for the former the evidence is conflicting. The presence of these motifs suggests periodical patterns of dityrosine crosslinks, which harden the eggshell rendering it insoluble. Fourier transform infrared spectroscopy data from intact* Ceratitis capitata *eggshells support the validity of prediction.*

Keywords: *Ceratitis capitata*; eggshell (chorion) protein structure; secondary structure prediction; Fourier analysis; infrared spectroscopy; tandemly repeating peptide motifs

Introduction

The Mediterranean fruitfly *Ceratitis capitata* (Diptera/Tephritidae) is an insect of economic importance, since more than 250 different fruit species are infested with its eggs. Its proteinaceous eggshell (chorion) has similar ultrastructural features to the eggshell of *Drosophila melanogaster*, which has been studied in detail[1-3]. The eggshell participates in important functions during oogenesis and embryogenesis including oocyte respiration, sperm entry-fertilization, thermal insulation, water-proofing, resistance to external mechanical pressures and hatching[4].

The *Ceratitis capitata* chorion is currently used as a model system in the study of programmed differential gene expression during development and of molecular evolution of dipteran chorion genes[5,6]. The proteins forming this multilayered protective and functional structure are produced and secreted solely by the follicle cells surrounding the oocyte, in a characteristic temporal and spatial pattern that reflects the transcriptional regulation of the respective genes. They are deposited onto the surface of the oocyte, where they self-assemble to form the eggshell, in the short time available for choriogenesis[4].

A set of six major *Ceratitis capitata* chorion proteins has been identified by polyacrylamide gel electrophoresis: Cs60, Cs45, Cs43, Cs30, Cs19 and Cs17; numbers indicate molecular weights in kDa[4]. Isolation of chorion cDNA clones, denoted by C1, C2, C4 and C5, reveals that the apparent molecular weights of the proteins encoded by the C1, C2, C4 and C5 clones are approximately 68, 43, 59 and 32 kDa, respectively[6].

The amino acid sequences of *Ceratitis capitata* chorion proteins encoded by the C2 (43 kDa) and C5 (32 kDa)

* To whom correspondence should be addressed.

0141-8130/91/050307-09

clones have been determined[5,6]. Exhibiting extensive sequence homology, they were found to correspond to proteins s36 and s38 of *Drosophila melanogaster*, respectively; therefore, they will be referred to, hereafter, as Ccs36 and Ccs38 correspondingly.

Our efforts are focused on attempts to determine the folding modes of chorion proteins and correlate the molecular models with the complex architectural plan displayed by the various eggshell layers. Ultimately, we want to elucidate the principles governing chorion architecture and their influence on its important functions; we have been using the eggshell of *Drosophila melanogaster* and its constituent proteins as a model system[7-10]. In this work, we present findings from the analysis of the information hidden in the amino acid sequences of *Ceratitis capitata* eggshell proteins Ccs36 and Ccs38, together with supporting experimental evidence from Fourier transform infrared spectroscopy data. Comparison with *Drosophila melanogaster* data highlights important structural and functional features of the molecules.

Experimental

Fourier transforms were obtained essentially as outlined by MacLachlan[11], 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[12].

The methods used for secondary structure prediction have been described in detail by Hamodrakas *et al.*[13] and Hamodrakas and Kafatos[14]. 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[15].

Ceratitis capitata flies, conditioned at 25°C, were lightly etherized after 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 300 individual chorions, thoroughly dried.

Infrared spectra were recorded on a Fourier Transform Bruker 113 v, 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 ground thoroughly in a vibrating mill, before mixing with KBr.

Results

The amino acid sequences of *Ceratitis capitata* chorion proteins Ccs36 and Ccs38[5,6] are shown in *Figure 1*.

A highly unusual amino acid composition and distribution of certain types of residues is clearly discernible in internal regions within each protein (*Table 1* and data not shown). Both proteins exhibit a tripartite structure: They appear to consist of a central 'domain' rich in Val, Pro and Lys, flanked by two 'arms' or 'tails' rich in Gly, Ala and Ser. In the C-terminal 'tail' of both Ccs36 and Ccs38, Tyr is abundant. Furthermore, the distribution of charged amino acid residues is uneven; positively charged amino acid residues (Lys, Arg) are clustered in the central domain, whereas negatively charged residues (Glu, Asp) are mostly confined in the arms of the molecules. Following the notation used for the corresponding proteins of *Drosophila melanogaster*[9], we define the borders of the central domain by the first and the last appearance of Lys in both cases. Thus, the central domain of Ccs36 contains residues 97 to 218, whereas that of Ccs38 residues 77 to 183 (*Figure 1*).

The results of secondary structure prediction (*Figure 2*) are in agreement with this definition. The central domain of both proteins is predicted to adopt a

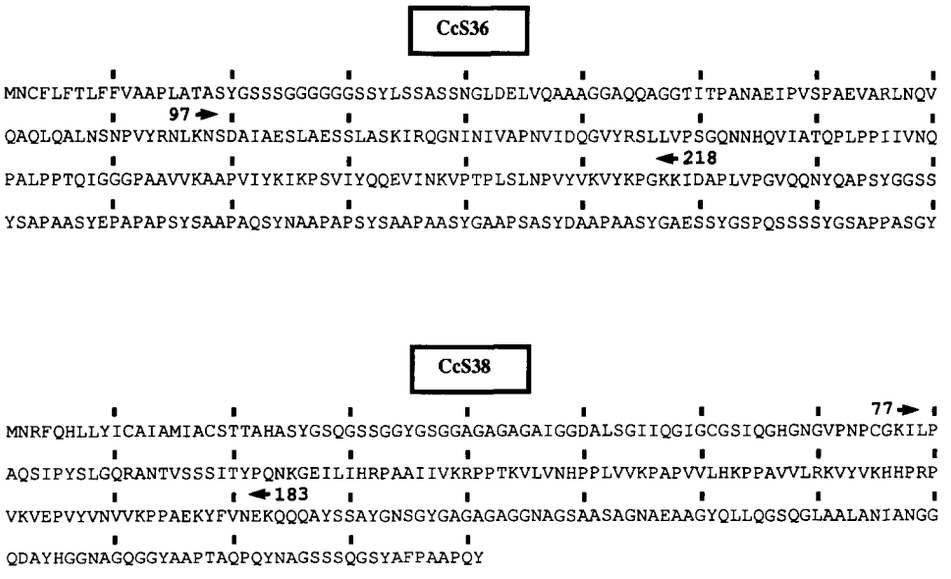
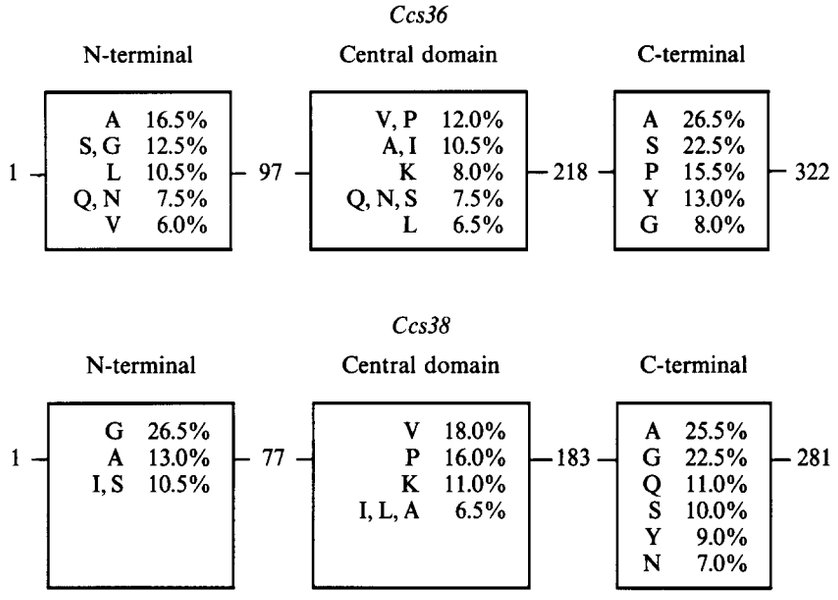


Figure 1 Amino acid sequences (one letter code) of *Ceratitis capitata* chorion proteins Ccs36 and Ccs38 obtained from refs. 5, 6. Arrows and numbers indicate the borders of the central domain in each protein (see text)

Table 1 Amino acid composition in the central domain and the flanking arms of *Ceratitis capitata* eggshell proteins Ccs36 and Ccs38. Residues scoring above 6% are shown



characteristic structure of short β-sheet strands, alternating with β-turns, reminiscent of the structure of the corresponding proteins in *Drosophila melanogaster* and, in some respects, of silkmoth chorion proteins and of *Schistosoma mansoni* eggshell protein[12,16,17]. The regular alternation of these secondary structure elements is more profound in Ccs38 and less evident in Ccs36 (*Figure 2*). In the 'arms' of the molecules α-helices are frequently predicted, connected by turns or loops. The non-random distribution of certain types of residues in the three distinct domains of the molecules and the existence of tandemly repeating secondary structure elements suggested a search for the existence of repeating motifs, a characteristic feature of fibrous protein structure. Thus, Fourier analysis of the sequences detects high intensity periodicities for certain types of residues or groups of residues in the three parts, and/or the entire length of the molecules. The most characteristic periodicities are listed in *Table 2*. In Ccs38, the analysis reveals strong 7–8 residue periodicities for residues abundant in the central domain (Val, Pro, Leu), both in the central domain and the entire protein. A 7–8 residue strong periodicity is also seen for β-turn former residues (G, P, N, S, C, K, W, Q, T, R, E; ref. 18) and for β-sheet former residues (V, L, I, F, W, Y, T,C; ref. 18) which are out of phase. In Ccs36 (data not shown), the evidence for the existence of similar periodicities is less convincing.

However, very strong indications suggest an 8-residue periodicity for Ala, Ser, Pro, Tyr and β-turn former residues in the C-terminal 'tail' of the molecule (*Table 2*). These results, combined with secondary structure prediction, can best be interpreted to imply the presence of the tandem imprecise repeats shown in *Figure 3*, for both molecules.

The central domain of the proteins consists of tandem repeats of an octapeptide (and its variants), having the general formula -X-X-X-Z-Z-Z-Z-Z-, where X is a large hydrophobic, β-sheet-former residue and Z a β-turn former residue. In Ccs38, this imprecise repeating motif is seen to have the specific pattern -X-X-X-p-p-P-P-β-, where p is usually a positively charged residue, P is proline and β a β-turn former residue.

In the Ccs36 C-terminal 'arm', a striking repeating pattern of the motif -Y-X-A-A-P-A-A-S- (X is usually Gly or Ser) is evident. The architectural plan of the C-terminal 'arm' of Ccs38 contains tandem repeats of a nonapeptide having the form: -Y-Z-Z-Z-Z-(G, A)-Z-Q-Z-, where Z is mostly Ala, Gly, Pro or Ser. An intriguing feature of both Ccs36 and Ccs38 C-terminal arms is the existence of an internal twofold axis of symmetry, which appears to relate two, most probably, structurally equivalent, halves of this domain (*Figure 4*).

In the N-terminal 'tails' of the molecules the presence of tandem imprecise repeats is also evident: Thus, the N-terminal 'tail' of Ccs36 contains repeats of an octapeptide of the form -Z-Z-Z-A-X-A-A-Z- (X is Gln, Asn or Glu and Z is usually a β-turn former residue) which, apparently, shares similarities with the repeating peptides of the C-terminal 'tail' having the motif -Y-X-A-A-P-A-A-S-. whereas in the N-terminal 'arm' of Ccs38 the presence of tandem repeats of the motif -x-y-G-z-G-u-G-v- (x = I, S, y = G, Q) can be seen (*Figure 3*).

In view of the evidence presented above, the imprecise repeats of the central domain of both Ccs36 and Ccs38 can be accommodated into the antiparallel β-pleated sheet model of alternating β-strands/β-turns or loops shown in *Figure 5*. This model, as might have been expected from the exceptionally high degree of homology of primary sequences in this region, which implies a remarkable evolutionary conservation[5,6], shows extensive similarities to the model proposed for the corresponding regions of *Drosophila melanogaster* proteins s36 and s38[9].

We shall not offer exact models for the protein 'arms', since the precise definition of the borders of secondary structure elements in these domains is difficult (see *Figures 2* and *3*). Nevertheless, some observations may have a heuristic value. Thus, the amino and carboxy

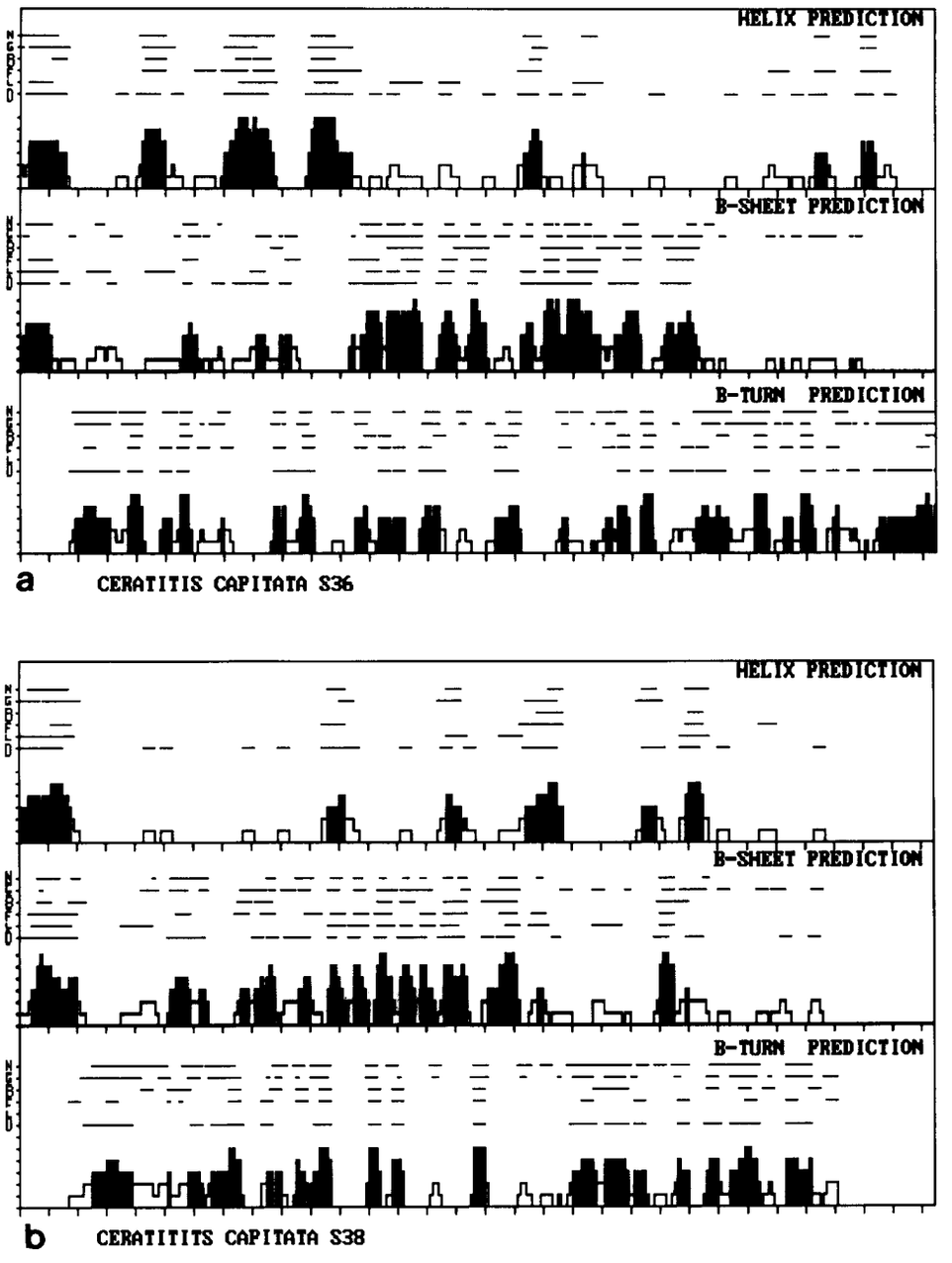


Figure 2 Secondary structure prediction plots for α-helix, β-sheet and β-turns, for *Ceratitis capitata* proteins Ccs36 (a) and Ccs38 (b). Individual predictions, as derived according to Nagano (N), Garnier *et al.* (G), Burgess *et al.* (B), Chou and Fasman (F), Lim (L) and Dufton and Hider (D), are shown by horizontal lines. Joint prediction histograms, constructed by tallying individual predictions, are also shown. The most probable structures, predicted by three or more methods, are shaded. The plots are 'hard' copies on a laser printer of a monitor screen

terminal 'arms' of protein Ccs36, containing the tandemly repeating peptides -Z-Z-Z-A-X-A-A-Z- and -Y-X-A-A-P-A-A-S- respectively, which apparently share homologies, appear to be formed of alternating α-helices and β-turns (the rich in Ala regions of the molecules constituting the α-helices). In contrast, the N-terminal arm of Ccs38, containing repeats of the peptide -x-y-G-z-G-u-G-v-, is predicted to consist of β-strands regularly alternating with β-turns. For the repeating nonapeptides of the Ccs38 C-terminal 'arm', the evidence for the existence of α-helix and/or β-sheet strands is conflicting.

The presence of all types of secondary structure in *Ceratitis capitata* chorion proteins is strongly supported by the Fourier transform infrared spectrum of *Ceratitis capitata* chorions (*Figure 6*). Absorption bands at 1634 cm^{-1} (amide I) and 1520 cm^{-1} (amide II) are indicative of antiparallel β-pleated sheet conformation, whereas the bands at 1650 cm^{-1} (amide I) and at 1539 and 1550 cm^{-1} (amide II) characterize α-helical secondary structure[19]. The band at 1643 cm^{-1} might indicate a different type of β-pleated sheet structure, or correspond to a collagen-like conformation in chorion proteins[19].

Discussion

These results exemplify the potential of analysing protein sequences for the recovery of tandem motifs. Tandemly repetitive peptide motifs have been found in the sequences

Table 2 The most characteristic residue periodicities in *Ceratitis capitata* 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

Ccs38 (1–281) Type of residue	Periodicity	Intensity	Phase angle
P	7.31	5.79	−137.0
V	7.53	4.57	168.1
R, K, H	7.42	4.68	−2.0
β-turn formers	7.76	6.14	146.2
β-sheet formers	7.53	8.34	134.4
	7.64	6.26	53.8
	7.76	5.03	−15.9
Ccs38 Central domain (77–183) Type of residue	Periodicity	Intensity	Phase angle
P	7.21	7.34	114.9
V	7.40	5.22	−178.1
L	7.64	4.90	119.0
β-turn formers	7.64	7.04	−88.6
β-sheet formers	7.64	14.12	78.5
Ccs36 C-terminal arm (219–320) Type of residue	Periodicity	Intensity	Phase angle
A	7.88	5.56	63.3
	4.03	6.95	−141.6
S	7.76	5.72	−108.1
	7.64	6.42	−69.5
P	7.88	4.07	79.7
Y	7.88	7.70	−102.5
β-turn formers	8.00	4.95	−159.2

of most fibrous proteins and play an important role in the formation of the fibrous structure[20,21]. 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[22]. The presence of specific peptide motifs of well defined conformation and function, though not necessarily of a repetitive nature, has also been documented in globular proteins[23]. An important question which always arises in such cases is what type of structure these peptides adopt.

A key aspect in the proposal of the compact antiparallel β-pleated sheet structure of the central domain of chorion proteins Ccs36 and Ccs38, presented in *Figure 5*, were the results of secondary structure prediction, combined with the out-of-phase periodical arrangement of β-sheet former residues with β-turn former residues (*Table 2*). The same approach has been applied successfully in the elucidation of the structure of feather keratin[24], the adenovirus fibre protein[25], silkmoth chorion proteins[12,16] and *Drosophila* chorion proteins[9].

The antiparallel β-pleated sheet model of the central domain of both Ccs36 and Ccs38, 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). It exhibits a remarkable evolutionary conservation in two species like *Drosophila* and *Ceratitis* estimated to have separated approximately 120 million years ago (ref. 5 and references therein), which, perhaps, signifies its crucial functional role; sequence comparison between *Ceratitis capitata* and *Drosophila melanogaster* s38 proteins reveals four variant residues in this domain only, whereas, there are no insertions or deletions[6]. Despite, limited sequence homology between Ccs36 and Ccs38 (estimated to be approximately 30%) in the central domain, the basic fold of the proteins appears to be the same (*Figure 5*). However, distinct differences in the distribution of certain types of residues (e.g. positively charged residues), in the turns or loops of the structure, perhaps, imply different roles for the two molecules.

The structural conservation of the central domain both in Ccs36 and Ccs38, in two Dipteran species (and also in a third species, *Drosophila virilis* – data not shown), leads to a search for conserved ultrastructural features in the chorions of these species. A chorion layer of unknown function, highly conserved in several Diptera, is the crystalline innermost chorionic layer (ICL); this layer consists of proteins secreted early in choriogenesis and contains two different types of similar globular subunits with diameters of approximately 3–4 nm[7,8,10]. The structurally similar central domains of the 'early' secreted proteins Ccs36 and Ccs38[4], which were detected to participate in the architecture of the ICL[25a], are ideal candidates for the construction of this structure. Preliminary model building[25b] suggests that the central domain of Ccs36 and Ccs38 can form a globular unit with a diameter of the order of 3–4 nm, if folded into a β-barrel type of structure, or if it is simply an antiparallel twisted β-pleated sheet; most β-sheets, even in fibrous proteins, are twisted (ref. 12 and references therein). The different role of Ccs38, which most probably acts as structural peroxidase crosslinking chorion proteins with

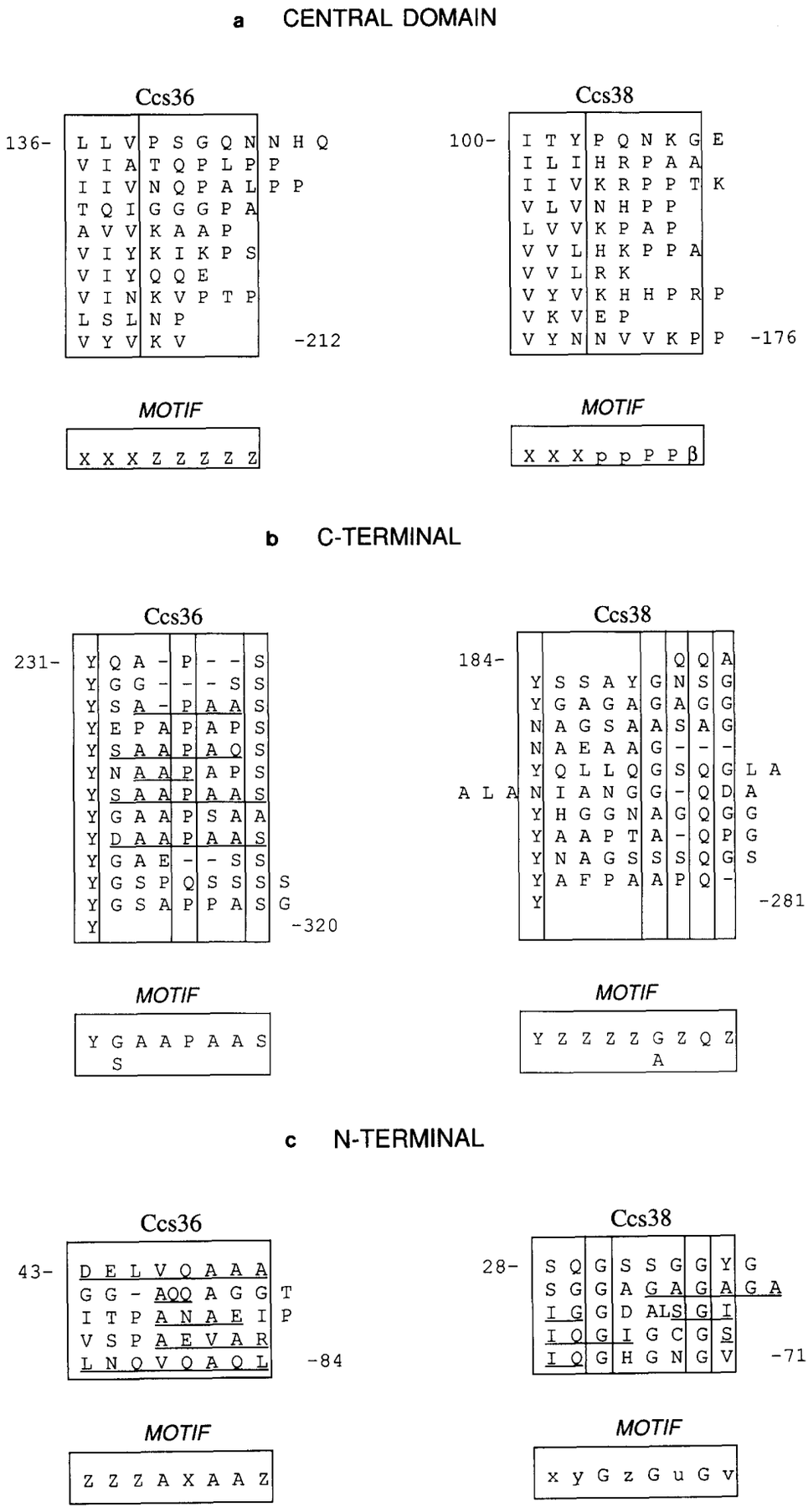


Figure 3 Tandem imprecise repeats and representative repeating peptide motifs in: (a) the central domain; (b) the C-terminal 'arm'; and (c) the N-terminal 'arm' of *Ceratitis capitata* eggshell proteins Ccs36 and Ccs38. Sequences should be read left to right, top to bottom. Numbers indicate actual amino acid positions. In (b), residues predicted as α-helix are underlined. In (c), for Ccs36, residues predicted as α-helix and for Ccs38 as β-sheet are also underlined. For further details see text

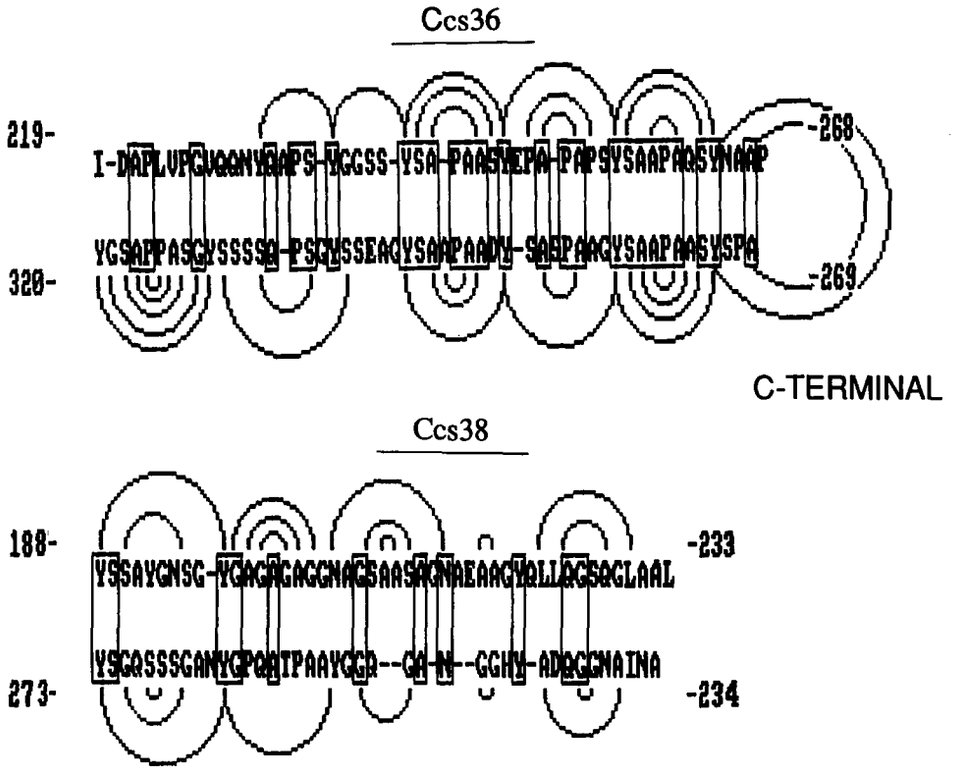


Figure 4 Symmetry in the C-terminal 'arm' of *Ceratitis capitata* eggshell proteins Ccs36 and Ccs38. A twofold axis of symmetry relates two, apparently homologous, halves. Identical residues in the two halves are boxed. The homology is clearly higher in Ccs36. In addition, twofold axes of symmetry are also present in the tandem repeats. Arcs connect symmetrically disposed residues. Numbers indicate actual amino acid positions

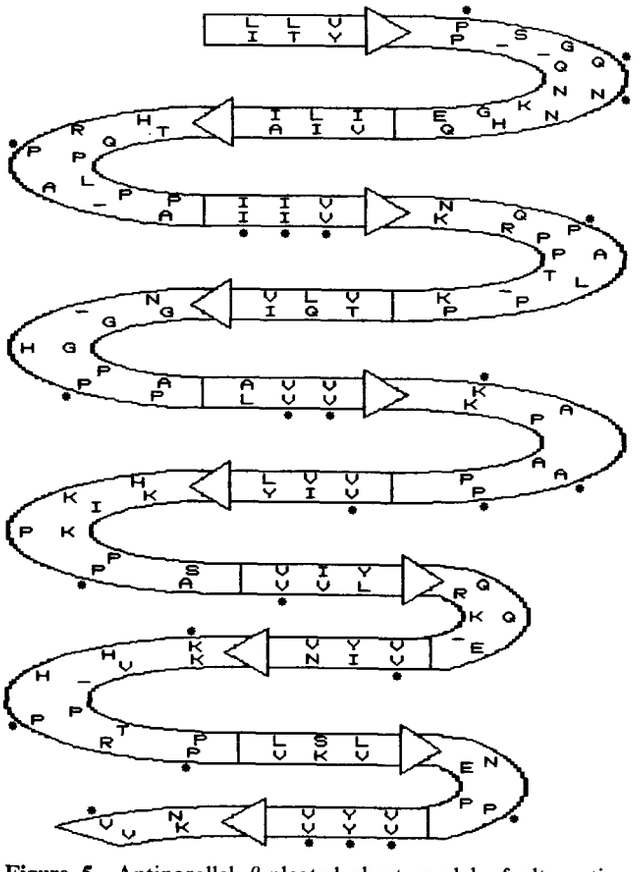


Figure 5 Antiparallel β-pleated sheet model of alternating β-strands (arrows)/β-turns or loops of the central domain of *Ceratitis capitata* eggshell proteins Ccs36 (right) and Ccs38 (left). Sequences should be read continuously beginning at the top. Asterisks indicate exact matches. For further details see text

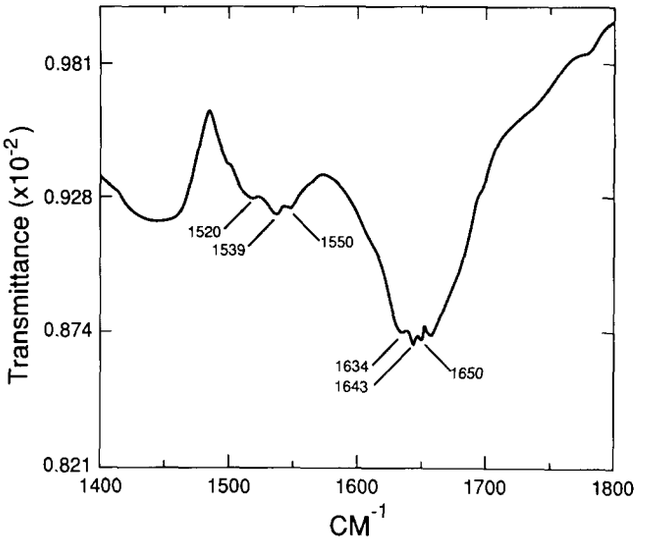


Figure 6 Fourier transform infrared spectrum of *Ceratitis capitata* chorion. The spectrum is the result of signal averaging of 100 scans, at 2 cm^{-1} resolution. Samples were in the form of KBr pellets, containing about 2% wt material (approximately 300 chorions), thoroughly ground in a vibrating mill, before mixing with KBr

di-tyrosine bonds in the late choriogenetic stages[25a], might be explained by the different, in detail, structure of the two molecules.

A search of the OWL database (Ver. 9.0: 25049 entries with 7.308.377 residues; Ref. 26) reveals that the tandem motifs of the central domain appear rarely in proteins. For example, the motif -X-X-X-p-p-P-P-β, where X is a large hydrophobic residue, p a positively charged residue

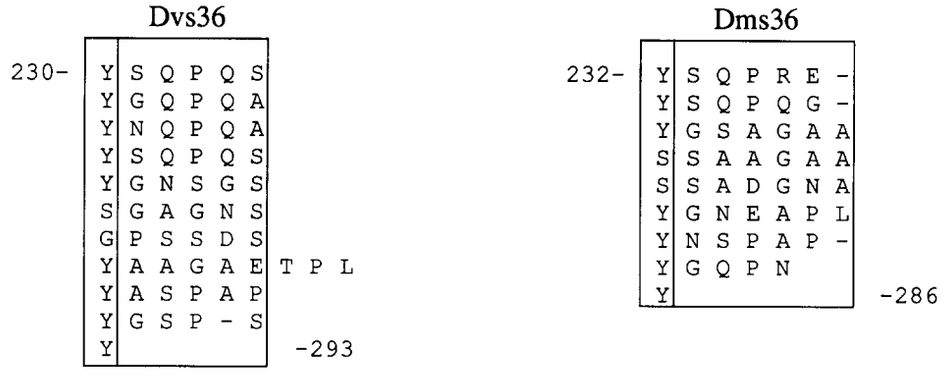


Figure 7 Tandemly repeating peptides in the C-terminal 'arm' of *Drosophila virilis* s36 (Dvs36) and *Drosophila melanogaster* s36 (Dms36)

and β a β-turn former appears three times only, in, apparently, unrelated proteins.

In strong contrast to the central domain, where the degree of evolutionary relatedness is high, the flanking 'arms' of the proteins share limited sequence homologies with the corresponding regions of *Drosophila melanogaster*[5,6]. Structural diversification is also apparent (compare *Figure 2* with *Figure 2* of Hamodrakas *et al.*[9]). The most prevalent feature of the protein 'arms' is the almost exact octapeptide periodicity appearing in the C-terminal domain of protein Ccs36, with the octapeptide having the form -Y-(G or S)-A-A-P-A-A-S-. A search of the OWL database verifies its unique appearance in *Ceratitis capitata* chorion protein Ccs36. Omitting Tyr, the motif shows an internal twofold symmetry axis passing through Pro. Interestingly, the periodicity is retained at the DNA level (a 24-nucleotide repeat is seen) and, also, the internal twofold symmetry axis in each nucleotide repeat (data not shown), which might have important evolutionary implications. The secondary structure of the octapeptide repeats in the C-terminal 'arm' of Ccs36 appears to consist of alternating short α-helices and β-turns. However, the α-helices might not be ideal α-helices: they are very short (one to one and a half turns), and they contain Pro, an α-helix breaker residue; Pro creates local kinks in α-helices[27]. Nevertheless, the overall structure of the octapeptide repeats is most probably a helical type of structure, currently under investigation utilizing computer modelling.

The eight-residue periodicity for Tyr is of special interest. Tyrosines are crosslinked through the action of a peroxidase, *in vivo*, by covalent di-tyrosine bonds, late in choriogenesis, to render chorion insoluble[3]. Therefore, the spatial disposition of Tyr every eight residues, may suggest a similar periodical pattern of di-tyrosine bond formation.

Apparently, the octapeptide tandem repeats of the Ccs36 C-terminal 'arm' are not a conserved feature in the sequences of *Drosophila melanogaster* and *Drosophila virilis*[5,9]; there is evidence for a hexa- or heptapeptide repeat in *D. melanogaster* and a hexapeptide repeat in *D. virilis* based on the periodicity of tyrosines in this domain, as shown in *Figure 7*. These peptides are rich in polar residues (Ser, Gln, Asn) and also Ala, Gly, Pro. Their existence suggests different di-tyrosine bonding patterns in these species.

The N-terminal arm of Ccs36, obviously, plays a different role than the C-terminal. Its octapeptide imprecise repeats -Z-Z-Z-A-X-A-A-Z-, predicted to form α-helices alternating with β-turns and containing the motif -A-X-A-A- found also in the C-terminal peptides, cannot participate in di-tyrosine crosslinks; they accommodate no tyrosines. Three tyrosines not included in the repeats of the N-terminal 'arm' are, however, potential sites for di-tyrosine bond formation.

In the Ccs38 N-terminal 'arm', the regular repeat of Gly every two residues in the tandem motifs of the peptide -x-y-G-z-G-u-G-v- is reminiscent of the Gly periodicities in silk fibroin, a β-sheet protein[28]. Interestingly, despite the fact that secondary structure prediction methods, with their inherent limitations, usually fail in the presence of short tandem motifs[14,18], these portions of the molecules are predicted to form β-sheet strands alternating with β-turns.

In recent works, Fourier transform infrared spectroscopy, frequently in conjunction with Raman spectroscopy, has empirically been demonstrated to be a powerful and reliable technique for the elucidation of protein secondary structure in similar systems[29]. Unfortunately, for *Ceratitis capitata*, 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. A similar effect has been observed for *Drosophila melanogaster* chorions and it might be attributed partly to the presence of the di-tyrosine crosslinks.

Further work is needed along more refined paths to relate the conformation of the tandem motifs of chorion proteins with the complex ultrastructural architecture of chorion and its important physiological functions.

References

1 Margaritis, L. H., Kafatos, F. C. and Petri, L. H. *J. Cell Sci.* 1980, **43**, 1

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, 1985a, Vol. 1, pp. 153–230

3 Margaritis, L. H. *Tissue and Cell* 1985b, **17(4)**, 553

4 Margaritis, L. H. *Can. J. Zool.* 1985c, **63**, 2194
5 Konsolaki, M., Komitopoulou, K., Tolias, P., King, D., Swimmer, C. and Kafatos, F. C. *Nucleic Acids Res.* 1990, **18(7)**, 1731
6 Tolias, P., Konsolaki, M., Komitopoulou, K. and Kafatos, F. C. *Dev. Biol.* 1990, **140**, 105
7 Hamodrakas, S. J., Margaritis, L. H. and Nixon, P. E. *Int. J. Biol. Macromol.* 1982, **4**, 25
8 Margaritis, L. H., Hamodrakas, S. J., Arad, T. and Leonard, K. R. *Biol. Cell* 1984, **52**, 279
9 Hamodrakas, S. J., Batrinou, A. and Christiphoratou, T. *Int. J. Biol. Macromol.* 1989, **11**, 307
10 Margaritis, L. H., Hamodrakas, S. J., Papassideri, I., Arad, T. and Leonard, K. R. *Int. J. Biol. Macromol.* 1991, **13**, 247
11 McLachlan, A. D. *Biopolymers* 1977, **16**, 1271
12 Hamodrakas, S. J., Ekmetzoglou, T. and Kafatos, F. C. *J. Mol. Biol.* 1985, **186**, 583
13 Hamodrakas, S. J., Jones, C. W. and Kafatos, F. C. *Biochim. Biophys. Acta* 1982, **700**, 42
14 Hamodrakas, S. J. and Kafatos, F. C. *J. Mol. Evol.* 1984, **20**, 296
15 Hamodrakas, S. J. *CABIOS* 1988, **4(4)**, 473
16 Hamodrakas, S. J., Bosshard, H. E. and Carlson, C. N. *Prot. Engineer.* 1988, **2(3)**, 201
17 Rodrigues, V., Chaundhri, M., Knight, M., Meadows, H., Chambers, A. E., Taylor, W. R., Kelly, C. and Simpson, A. J. G. *Mol. Biochem. Parasitol.* 1989, **32**, 7
18 Chou, P. Y. and Fasman, G. D. *Adv. Enzymol.* 1978, **47**, 45
19 Parker, F. S. 'Applications of Infrared Spectroscopy in Biochemistry, Biology and Medicine', Plenum Press, New York, 1971
20 Fraser, R. D. B. and McRae, T. P. in 'Conformation in Fibrous Proteins', Academic Press, New York, 1973
21 Parry, D. A. D. in 'Fibrous Protein Structure' (Eds J. M. Squire and P. J. Vibert), Academic Press, London, 1987
22 Parry, D. A. D., Fraser, R. D. B. and McRae, T. P. *Int. J. Biol. Macromol.* 1979, **1**, 17
23 Bairoch, A. 'Prosite: A Dictionary of Protein Sites and Patterns', University of Geneva, Fifth release, 1990
24 Fraser, R. D. B. and McRae, T. P. 'Proc. 16th Int. Ornithological Congress', Canberra, 1976, pp. 443–451, Australian Academy of Science, Canberra
25 Green, N. M., Wrigley, N. G., Russel, W. C., Martin, S. R. and McLachlan, A. D. *EMBO J.* 1983, **2(8)**, 1357
25a Margaritis, L. H. Unpublished data
25b Hamodrakas, S. J. Unpublished data
26 Akkrigg, D., Bleasby, A. J., Dix, N. I. M., Findlay, J. B. C., North, A. T. C., Parry-Smith, D., Wooton, J. C., Blundell, T. L., Gardner, S. P., Hayes, F., Islam, S., Sternberg, M. J. E., Thornton, J. M. and Tickle, I. J. *Nature* 1988, **335**, 745
27 Richardson, J. S. and Richardson, D. C. in 'Prediction of Protein Structure and the Principles of Protein Conformation' (Ed. G. D. Fasman), Plenum Press, New York, London, 1989, pp. 1–98
28 Marsh, R. E., Corey, R. B. and Pauling, L. *Biochim. Biophys. Acta* 1955, **16**, 1
29 Hamodrakas, S. J., Kamitsos, E. I. and Papadopoulou, P. G. *Biochim. Biophys. Acta* 1987, **913**, 163