Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallographic identification of an unexpected protein complex in silkworm haemolymph

Agnieszka J. Pietrzyk,[a] Anna Bujacz,[b] Jochen Mueller-Dieckmann,[c] Malgorzata Łochynska,[d] Mariusz Jaskolski[a,e] and Grzegorz Bujacz[a,b]*

[a]Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland, [b]Institute of Technical Biochemistry, Faculty of Biotechnology and Food Sciences, Technical University of Lodz, Stefanowskiego 4/10, 90-924 Lodz, Poland, [c]Biocenter Klein Flottbeck, University of Hamburg, Ohnhorststrasse 18, 22609 Hamburg, Germany, [d]Institute of Natural Fibers and Medicinal Plants, Wojska Polskiego 71B, 60-630 Poznan, Poland, and [e]Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Grunwaldzka 6, 60-780 Poznan, Poland

Correspondence e-mail: grzegorz.bujacz@p.lodz.pl

Received 7 June 2013
Accepted 5 August 2013

PDB Reference: SP2–SP3 complex, 4l37

The first crystal structure of a complex formed by two storage proteins, SP2 and SP3, isolated from their natural source, mulberry silkworm (*Bombyx mori* L.) haemolymph, has been determined. The structure was solved by molecular replacement using arylphorin, a protein rich in aromatic amino-acid residues, from oak silkworm as the initial model. The quality of the electron-density maps obtained from the X-ray diffraction experiment allowed the authors to detect that the investigated crystal structure was composed of two different arylphorins: SP2 and SP3. This discovery was confirmed by N-terminal sequencing. SP2 has been extensively studied previously, whereas only a few reports on SP3 are available. However, to date no structural studies have been reported for these proteins. These studies revealed that SP2 and SP3 exist in the silkworm body as a heterohexamer formed by one SP2 trimer and one SP3 trimer. The overall fold, consisting of three haemocyanin-like subdomains, of SP2 and SP3 is similar. Both proteins contain a conserved N-glycosylation motif in their structures.

1. Introduction

The mulberry silkworm (*Bombyx mori*) is the main producer of natural silk fibres. Like all holometabolous insects, *B. mori* undergos complete metamorphosis, *i.e.* a transformation from larvae to moth *via* a pupal stage. The pupa, enclosed in a cocoon of raw silk, digests itself and the larval tissues are disintegrated. The main source of materials for the histogenesis of adult tissues is the pupal (perivisceral) fat body containing protein granules (Riddiford & Law, 1983). The storage proteins present in the granules are the main supply of nitrogen (Riddiford & Law, 1983; Levenbook, 1985; Kanost *et al.*, 1990). The accumulation of storage proteins occurs in the haemolymph of fifth-instar larvae, when the proteins are synthesized in the larval (peripheral) fat body. After synthesis they are secreted to the haemolymph and stored in the perivisceral fat body as soon as spinning starts (Vanishree *et al.*, 2005).

The majority of lepidopteran storage proteins are hexamerins, *i.e.* protein oligomers formed by six subunits. They can be divided into three subgroups: arylphorins, methionine-rich hexamerins and moderately methionine-rich hexamerins (Tojo *et al.*, 2012). All of them are evolutionarily related to the arthropod haemocyanin family (Telfer & Kunkel, 1991). Each of the subunits forming the hexamer has a molecular weight of about 80 kDa. In the 1980s two mulberry silkworm haemolymph hexamerins were isolated and biochemically characterized (Sakurai *et al.*, 1988; Fujii *et al.*, 1989). *B. mori* storage

protein 1 (SP1) belongs to the methionine-rich hexamerins (Sakurai *et al.*, 1988), whereas *B. mori* storage protein 2 (SP2) is an arylphorin (Fujii *et al.*, 1989). This classification is based on the content of methionine, which is 11.1% in the complete sequence of SP1 (Sakurai *et al.*, 1988), and aromatic residues (mainly phenylalanine and tyrosine), which constitute 19.0% of the total amino-acid composition of SP2 (Fujii *et al.*, 1989). The term 'arylphorin' was introduced by Telfer *et al.* (1983) and refers to proteins with a high content of amino acids with an aryl moiety.

The level of storage-protein synthesis is sex-dependent and these proteins are often called sex-specific (Ogawa & Tojo, 1981). The content of storage proteins in the female pupal fat body is around 60% of the total proteins, while the male fat body contains only 20% of these proteins (Hou *et al.*, 2010). SP1 is only expressed in females, whereas SP2 is synthesized in silkworms of both sexes. However, the level of SP2 synthesis in females is much higher than in males (Fujii *et al.*, 1989). A quantitative analysis of the main pupal proteins from the beginning to the end of this phase (Ogawa & Tojo, 1981) revealed that SP1 and SP2 disappear completely towards the end of this stage in females. At the same time, vitellogenin, the main yolk protein, is synthesized in amounts nearly equal to the digested SP1 and SP2; thus, it was suggested by Ogawa & Tojo (1981) that the storage proteins provide components for vitellogenin synthesis. The level of SP2 decreases in males during pupation, but SP2 is still detectable at the end of the pupal period (Fujii *et al.*, 1989).

Both the silkworm storage proteins SP1 and SP2 have been extensively studied over the last three decades. It is noteworthy that SP2 is also an inhibitor of apoptosis and that the addition of SP2 to human cell lines improves viability (Rhee *et al.*, 2007; Yu *et al.*, 2013). A detailed review of lepidopteran storage proteins, including SP1 and SP2, has been presented by Tojo *et al.* (2012). Nevertheless, there is at least one more silkworm storage protein, SP3, that has not been characterized to date. The gene of SP3 was found in the silkworm genome (The International Silkworm Genome Consortium, 2008) and the presence of this uncharacterized protein, which shares a similar expression pattern with SP2 and can be classified as an arylphorin (20.7% aromatic residues in the amino-acid sequence), was also noticed during silkworm proteome studies (Hou *et al.*, 2010). However, these are the only available reports on SP3.

In this work, we present the crystal structure of a complex between the SP2 and SP3 proteins at 2.9 Å resolution. The hexameric complex is formed by a trimer of SP2 and a trimer of SP3. It crystallized from a mixture containing all three storage proteins (SP1, SP2 and SP3) isolated from their natural source: silkworm haemolymph. At the outset of our structural studies, we were convinced that the hexamerin consisted exclusively of SP2 subunits. However, detailed analysis of the electron-density maps revealed that the structure was a complex of two different silkworm arylphorins. Thus, this paper presents another example of successful protein identification by X-ray crystallography. Two other similar (albeit simpler) case studies of this type have recently been reported by our group (Pietrzyk *et al.*, 2012; Pietrzyk, Bujacz, Muller-Dieckmann *et al.*, 2013). A review of 'crystallographic sequencing from electron-density maps', illustrated by examples from other studies, has also been published (Pietrzyk, Bujacz, Jaskolski *et al.*, 2013). The SP2–SP3 complex represents the first crystal structure of mulberry silkworm storage proteins. The finding that the hexamerin is formed as a heterohexamer of two different arylphorins is entirely unexpected. Only one other crystal structure of an arylphorin (APA), that isolated from the oak silkworm *Antheraea pernyi* (PDB entry 3gwj; Ryu *et al.*, 2009), is available in the PDB. The present study makes a significant contribution to our knowledge about silkworm storage proteins in general and SP3 in particular.

2. Materials and methods

2.1. Haemolymph collection, protein purification and crystallization

B. mori haemolymph was collected from fifth-instar larvae of the mulberry silkworm *B. mori* L. as described previously (Pietrzyk *et al.*, 2011). The larvae were not segregated according to sex and the haemolymph was stored as 2.0 ml aliquots collected from five to seven insects.

A sample containing the storage proteins was obtained using a two-step purification protocol. Firstly, the haemolymph proteins were separated according to their molecular weight using gel filtration, as described previously (Pietrzyk *et al.*, 2011). Briefly, the haemolymph was passed through a Superdex 200 prep-grade column (XK 16/100, Amersham Biosciences) equilibrated with 100 m*M* NaCl, 10 m*M* Tris pH 7.3, 0.025 m*M* 1-phenyl-2-thiourea. The peak fraction containing the hexamerin (molecular weight of ∼500 kDa) was concentrated and further purified using a Q Sepharose ion-exchange column (XK 16/10, Amersham Biosciences) equilibrated with 30 m*M* NaCl, 10 m*M* Tris pH 7.3. A stepwise elution gradient was used for protein separation and the fraction containing the storage proteins was eluted with 450 m*M* NaCl. The collected peak fraction was concentrated to 100 mg ml^{-1} and the buffer was exchanged to 10 m*M* Tris pH 7.3. This sample was used for crystallization trials.

Initial screening for crystallization conditions was performed at the High Throughput Crystallization (HTX) Facility at EMBL Hamburg, Germany. The crystallization trials were carried out in sitting drops by vapour diffusion using commercially available screens. The initial crystals grew in 2.0 *M* ammonium sulfate at 293 K. The conditions were further optimized using the hanging-drop technique and Additive Screen (Hampton Research). The crystals used in the diffraction experiment were obtained using 2.0 *M* ammonium sulfate as a precipitant and sodium thiocyanate as an additive at a final concentration of 0.02 *M*.

2.2. X-ray data collection and processing

X-ray diffraction data were collected on beamline X12 of the DESY synchrotron, Hamburg, Germany and on beamline

Table 1
Diffraction data-collection and refinement statistics for SP2–SP3.

Values in parentheses are for the highest resolution shell.

Space group	$P6_322$
Unit-cell parameters (Å)	a = 192.8, c = 180.8
Protein molecules per asymmetric unit	2
V_M (Å^3 Da^{-1})	2.97
Solvent content (%)	58.6
X-ray data collection	
Temperature (K)	100
Radiation source	BL14.2, BESSY
Wavelength (Å)	0.918
Detector	MAR CCD 225
Crystal-to-detector distance (mm)	290
Rotation range (°)	0.5
Total rotation (°)	75
Exposure per image (s)	35.0
Resolution (Å)	48–2.9 (3.0–2.9)
Mosaicity (°)	0.44
Intensities measured	404092
Unique reflections	44232
R_{merge}† (%)	10.8 (63.9)
Multiplicity	9.1 (9.4)
$\langle I/\sigma(I)\rangle$	23.2 (5.0)
Completeness (%)	99.7 (100.0)
Refinement	
R_{work}/R_{free}‡ (%)	16.58/22.26
R_{free} test-set count	1096
Protein/solvent atoms	11373/136
R.m.s. deviations	
Bond lengths (Å)	0.013
Bond angles (°)	1.66
Average B factor (Å^2)	43.1
Average B factor for solvent molecules (Å^2)	39.6
Ramachandran φ/ψ	
Favoured (%)	91.0
Disallowed (%)	0.1
PDB code	4l37

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of observation i of reflection hkl. ‡ $R_{work} = \sum_{hkl} ||F_{obs}| - |F_{calc}||/\sum_{hkl}|F_{obs}|$ for all reflections, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively. R_{free} is calculated analogously for the test reflections, which were randomly selected and excluded from the refinement.

14.2 of the BESSY synchrotron, Berlin, Germany (Mueller *et al.*, 2012). The detector was a MAR 225 CCD and the data were recorded from a single crystal using the rotation method with an oscillation of 0.5° at a temperature of 100 K. Prior to data collection, the crystals were briefly transferred into a cryoprotectant solution containing 70% Tacsimate pH 5.5, mounted in nylon loops and vitrified in a stream of cold nitrogen gas. Salts of carboxylic acids, which are the components of Tacsimate solution, are good cryoprotectants for crystals grown from ammonium sulfate (Bujacz *et al.*, 2010). The collected images were indexed, integrated and scaled using *XDS* and *XSCALE* (Kabsch, 2010*a*,*b*). Crystal parameters and data-collection statistics are given in Table 1.

2.3. Structure determination and refinement

The hexamerin structure was solved by molecular replacement with the *Phaser* software (McCoy *et al.*, 2007) using the structure of arylphorin (APA) from oak silkworm (PDB entry 3gwj; Ryu *et al.*, 2009) as a model. The final model of the SP2–SP3 complex was generated by iterative rounds of manual model building using *Coot* (Emsley *et al.*, 2010) and refinement using the *phenix.refine* package in *PHENIX* (Adams *et al.*, 2010) or *REFMAC* (Murshudov *et al.*, 2011) with the inclusion of TLS parameters (Painter & Merritt, 2006). The progress of the refinement was monitored and validated using 1094 reflections set aside for R_{free} testing (Brünger, 1992). The geometry of the model was assessed using *PROCHECK* (Laskowski *et al.*, 1993).

2.4. LC/MS/MS

The sample containing the storage proteins was first separated by gel electrophoresis according to Laemmli (1970). A single band corresponding to a molecular weight of 80 kDa was cut out and subjected to sequence analysis using liquid chromatography and electrospray ionization tandem mass spectrometry (LC/MS/MS; McCormack *et al.*, 1997) at the Proteomics Core Facility, EMBL Heidelberg, Germany.

2.5. N-terminal sequencing analysis

Two samples were subjected to Edman degradation. The first was a concentrated sample of the storage proteins after chromatographic purification. The second sample contained proteins collected from six-month-old crystallization drops from which the investigated crystal had grown. The collected material contained proteins from dissolved crystals and from the solution surrounding the crystals. Prior to analysis, the sample was centrifuged in order to remove any precipitate present in the drops. Both samples were separated by gel electrophoresis according to Schägger & von Jagow (1987). The proteins were transferred onto a PVDF Immobilon membrane PSQ 0.22 µm (Millipore). The single protein bands (two from the first sample and one from the second sample) corresponding to a molecular weight of about 80 kDa were cut out and subjected to Edman degradation cycles on a fully automated Procise 491 sequencer (Applied Biosystems) in the BioCentrum, Krakow, Poland.

3. Results and discussion

3.1. Purification, crystallization and initial sample characterization

The sample containing the storage proteins isolated from their natural source, mulberry silkworm haemolymph, was obtained using a two-step purification protocol (Figs. 1*a* and 1*b*). At the beginning of our studies, the exact contents of the sample were unknown. During the second purification step (ion-exchange chromatography) the fractions were collected from a narrow elution peak (Fig. 1*b*). However, SDS–PAGE electrophoresis revealed that at least two proteins were present in the sample (Fig. 1*c*). The amount of protein present in the lower (smaller mass) band (marked 2 in Fig. 1*c*) was significantly higher. The content of the contaminating protein in the sample was variable and the upper band (marked 1 in Fig. 1*c*) visualized on the SDS–PAGE gel was broader or narrower depending on the composition of the initial haemolymph sample used for purification. MALDI-TOF MS analysis was carried out to determine the molecular weight of

the sample (Fig. 1*f*). The molecular mass of the main protein was established to be 84.2 kDa, whereas the molecular mass of the other protein was slightly higher at 86.9 kDa.

LC/MS/MS analysis was performed in order to identify the main protein present in the sample. The lower band corresponding to the main protein was cut out from the SDS–PAGE gel and subjected to MS analysis. The analysis indicated that the protein was SP2 at the 75% level of confidence (protein score). The protein score was derived using the *Mascot* search engine (Matrix Science; http://www.matrixscience.com) from the scores of all cleaved peptides. A score above 45% indicates a high level of homology.

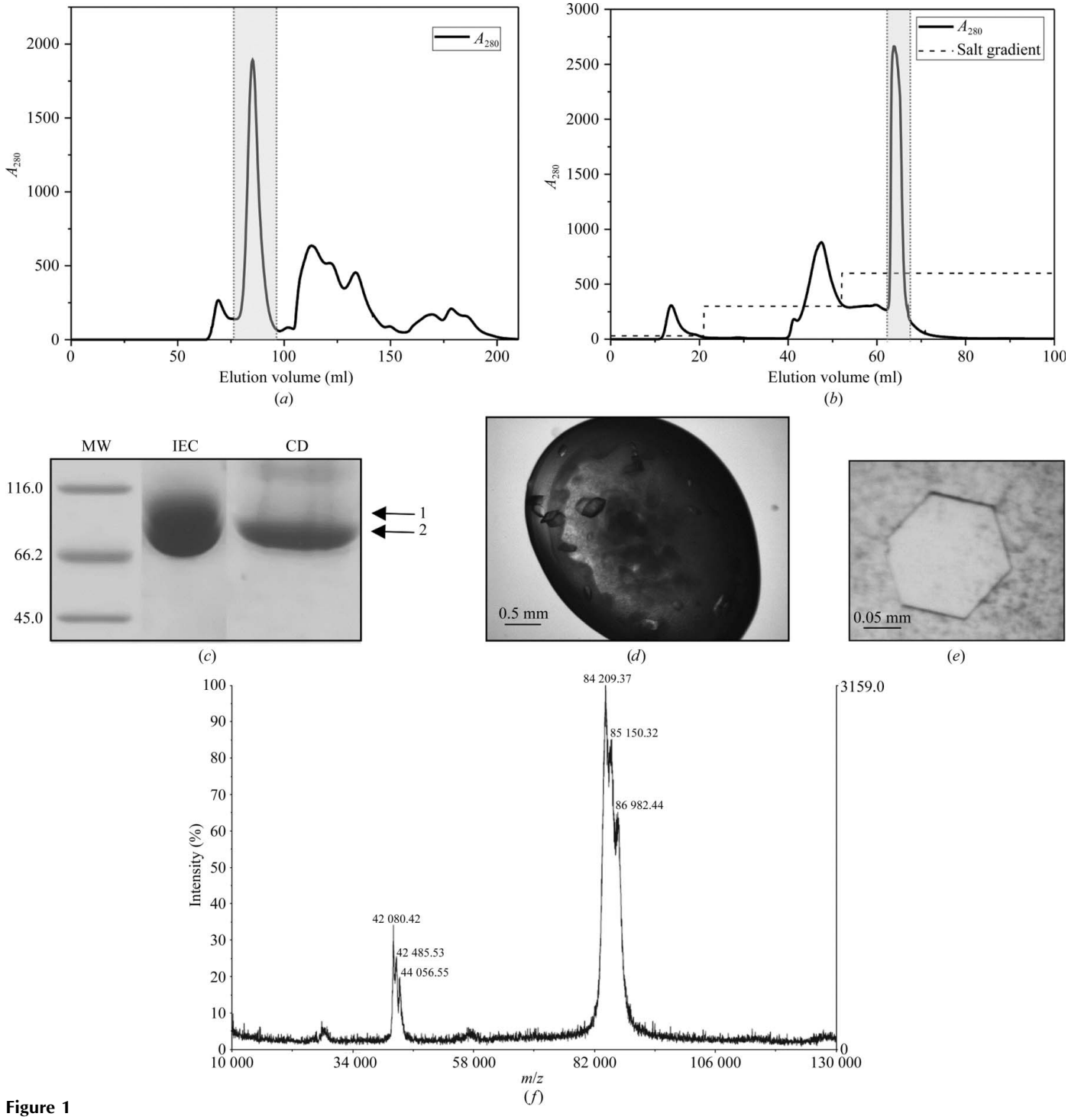

Figure 1
Purification, crystallization and initial sample characterization. Elution peaks from two purification steps, gel-filtration (*a*) and ion-exchange chromatography (*b*), are shaded. (*c*) SDS–PAGE electropherogram of the peak fraction collected during ion-exchange chromatography (IEC) and of a sample collected from crystallization drops (CD) (for details, see text); lane MW contains Protein Molecular Weight Marker (Fermentas); the upper and lower bands are marked 1 and 2, respectively. (*d*) Crystals of the SP2–SP3 complex precipitated in 2.0 *M* ammonium sulfate. (*e*) Crystals of the SP2–SP3 complex precipitated in 2.0 *M* ammonium sulfate and 0.02 *M* NaSCN. (*f*) MALDI-TOF MS spectrum of the IEC peak fraction.

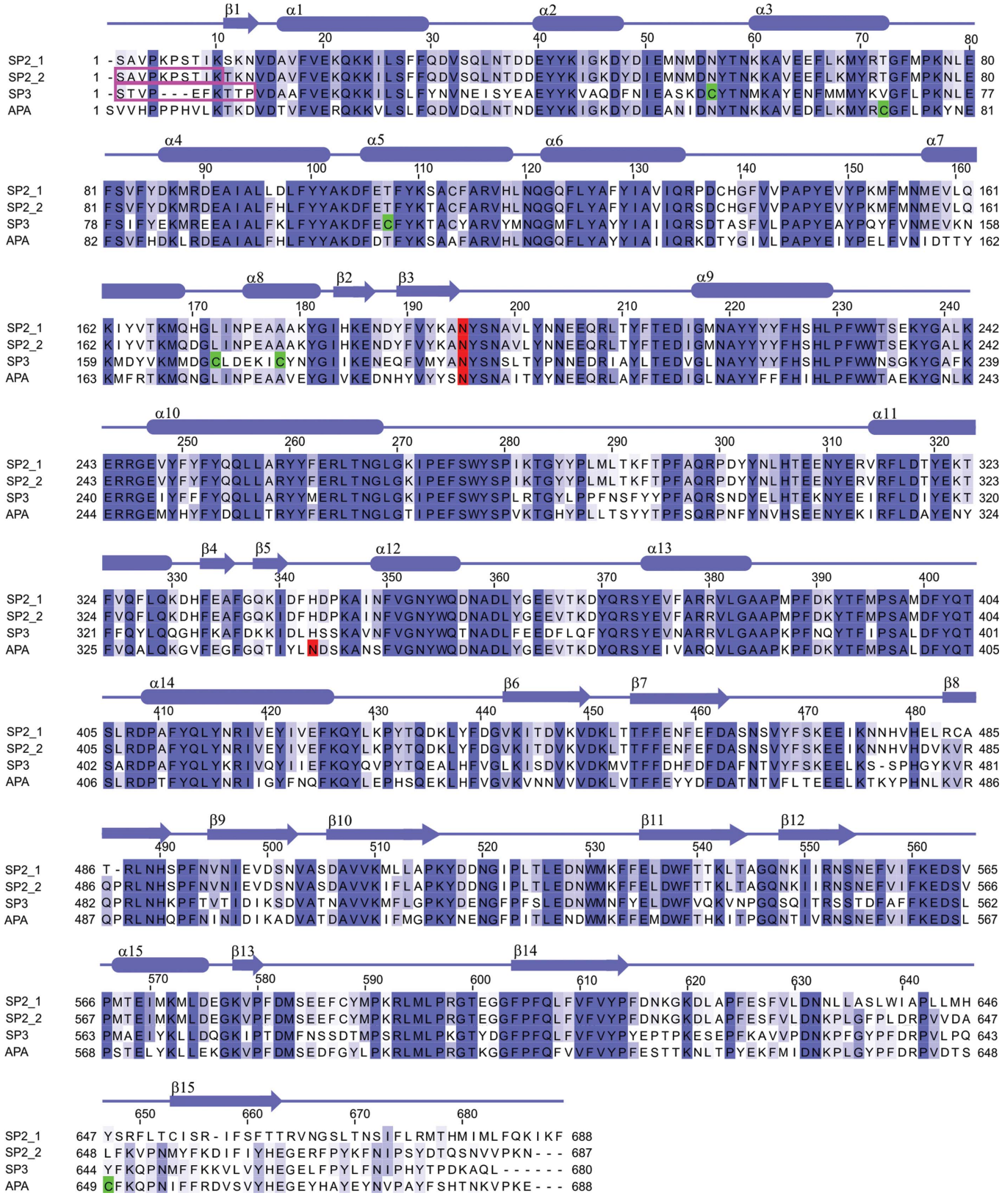

Figure 2
Amino-acid sequence alignment of arylphorins SP2_1, SP2_2, SP3 and APA (UniProt accession codes P20613, Q1HPP4, H9JHM9 and Q7Z1F8, respectively) calculated in *ClustalW* (http://www.ebi.ac.uk/Tools/msa/clustalw2/). The alignment is coloured according to identity (dark blue) and similarity (light blue) using *Jalview* (http://www.jalview.org/; Waterhouse *et al.*, 2009). The N-terminal sequences of SP2 and SP3 established by Edman degradation are boxed. The displayed sequences correspond to mature proteins without signal peptides. The glycosylation sites are highlighted in red and cysteine residues which form disulfide bridges are highlighted in green. The secondary-structure elements are assigned as α-helices (cylinders) and β-strands (arrows).

Table 2
Sequence homology among mulberry (SP*x*) and oak silkworm (APA) storage proteins.

The table presents percentage sequence similarities (and identities) among different sequences of mulberry and oak silkworm storage proteins calculated using *BLAST* (http://blast.ncbi.nlm.nih.gov/Blast.cgi; Altschul *et al.*, 1990). UniProt_1 and UniProt_2 refer to the two different sequences of SP2 with accession codes P20613 and Q1HPP4, respectively.

	SP1	SP2 (UniProt_1)	SP2 (UniProt_2)	SP3
APA	52 (33)	85 (69)	87 (70)	78 (61)
SP1		50 (32)	51 (33)	48 (32)
SP2 (UniProt_1)			91 (90)	78 (63)
SP2 (UniProt_2)				79 (65)

Table 3
Progress of model rebuilding and refinement.

The listed *R* factors are the values obtained in a round of refinement after each significant change of the model. Water molecules were added gradually at different stages and therefore there is no single step of solvent modelling.

	R_{work} (%)	R_{free} (%)
Molecular replacement, sequence SP2_1 (chains *A*, *B*)	29.2	36.2
Manual model rebuilding, sequence SP2_1 (chains *A*, *B*)	25.5	32.9
Sequence changed to SP2_2 (chains *A*, *B*)	22.8	29.3
Manual rebuilding of loop regions	19.4	27.4
Incorporation of all visible sugar moieties (ten sugar units)	18.7	25.7
Sequence of chain *A* changed to SP3	17.2	23.5
Final refinement	16.6	22.3

In order to separate the two proteins present in the sample, other chromatographic techniques were tested. Initial crystallization trials were concurrently performed and it turned out that crystallization itself could be a further purification step. The protein crystals grown in 2.0 *M* ammonium sulfate were buried in a brown precipitate in the crystallization drops (Fig. 1*d*). The protein material was collected from the drops and was centrifuged to remove the precipitate. SDS–PAGE analysis of this sample showed a single band (Fig. 1*c*), indicating that the sample purity was improved. However, the crystals had irregular prismatic shapes and only diffracted X-rays to about 3.0 Å resolution. The collected data could be indexed in space groups *C*2 or *C*222, with the Matthews volume indicating the presence of 3–4 protein molecules in the asymmetric unit. The intensity triage procedure indicated that the data were twinned. Further optimization of crystallization conditions with Additive Screen (Hampton Research) yielded crystals with a hexagonal plate morphology (Fig. 1*e*) belonging to space group $P6_322$. The crystals were obtained in the presence of 0.02 *M* sodium thiocyanate and diffracted X-rays to 2.9 Å resolution with no indication of twinning.

3.2. Structure determination and sequencing from electron-density maps

The X-ray diffraction data collected from a single hexagonal crystal were used for molecular-replacement calculations, which yielded a hexamerin crystal structure. In this approach, the amino-acid sequence of SP2 available in UniProt (http://www.uniprot.org; accession code P20613) was used and the crystal structure of arylphorin (APA) from oak silkworm (PDB entry 3gwj; Ryu *et al.*, 2009) was utilized as a model. The amino-acid sequences of SP2 and APA have 85% similarity and 69% identity (Table 2; Fig. 2). The asymmetric unit of the solved crystal structure contained two protein chains. A biological hexamer could be generated by the application of the crystallographic triad (3) symmetry to the dimer.

The first cycles of refinement, carried out with *PHENIX* (Adams *et al.*, 2010), and manual model rebuilding in *Coot* (Emsley *et al.*, 2010) according to the electron-density maps, indicated that the amino-acid sequence of the C-terminal fragment (residues 634–688) of the UniProt SP2 sequence P20613 (SP2_1) was completely wrong. At this point, another SP2 sequence (SP2_2) was found under accession code Q1HPP4. The two deposited sequences of SP2 differ by 10% (Table 2; Fig. 2), which is a very significant level. The discussed C-terminal part is completely different in the two variants. The C-terminal part of the model was significantly improved when the SP2 sequence was switched to Q1HPP4. The *R* factors decreased by about 3% (Table 3).

Further iterative rounds of refinement and manual rebuilding produced a model whose C^{α} atoms could easily be traced in the maps, indicating that the overall fold was correct. However, more than 20 positive and several negative peaks in $F_o - F_c$ difference electron-density maps were conspicuously outstanding at the 5σ level near the amino-acid side chains of subunit *A* (Fig. 3), suggesting sequence-related problems. The most intriguing observation was that no similar peaks were present at the corresponding positions of subunit *B* (Fig. 3). The sequence databases were screened again with emphasis on the ambiguities detected in subunit *A*. As a result, the sequence of another mulberry silkworm arylphorin, SP3, was found in the UniProt database with the accession code H9JHM9. The SP3 sequence matched the electron-density maps of chain *A* perfectly. Most of the positions where high positive difference peaks had been observed earlier corresponded to aromatic residues of SP3. Moreover, the sequence of SP2 contains only three cysteine residues, which do not form disulfide bonds, whereas two disulfide bridges are formed among the five cysteine residues present in the SP3 sequence. The assignment of the SP3 sequence to subunit *A* improved the refinement statistics significantly. The *R* factors decreased by about 2% (Table 3). All of these observations brought us to the final conclusion that the solved crystal structure represents a heterohexameric complex composed of two different arylphorins. Although the resolution of the X-ray diffraction data was not very high (2.9 Å resolution), sequencing from the electron density was possible thanks to the different positions of aromatic residues in the SP2 and SP3 sequences and to the high content of these residues in both proteins.

In order to confirm the results obtained from crystallography, N-terminal sequencing by chemical methods was conducted. The contents of the sample collected after ion-exchange chromatography and the material collected from the crystallization drops that had produced the crystals were analyzed. Two bands were present in the SDS–PAGE gel for the first sample. The analysis unequivocally indicated that the

upper (heavier) band (the contaminant) corresponded to SP1, with the N-terminal sequence $_1$SAISGGYGTM$_{10}$. The result for the lower band was more intricate (Fig. 4). However, its analysis revealed that the band contained all three storage proteins, *i.e.* SP1, SP2 and SP3. The N-terminal sequencing performed for the single band of the sample collected from the crystallization drops returned exactly the same result, indicating that SP1 had not completely precipitated during crystallization. However, the presence of SP1 in the soluble protein fraction did not prevent the crystallization of the SP2–SP3 complex. The N-terminal sequencing also provided new information about SP3. The protein contains an N-terminal signal peptide and our analysis revealed that the first amino-acid residue of the mature protein is a serine residue from the sequence $_1$STVP.... The amino-acid sequence of the signal peptide (MKTVLILAGLIALALS) is consistent with the signal-peptide sequences of other storage proteins (Fujii *et al.*, 1989) and other silkworm haemolymph proteins (Sakai *et al.*, 1988). The positively charged N-terminus (Lys residue) is followed by a number of hydrophobic residues in a sequence signature that is typical for secretion into the haemolymph (Shimada *et al.*, 1985). The cleavage site of the signal peptide of SP3 is located between two serine residues and in SP2 the cleavage occurs between the same residues (Fujii *et al.*, 1989). Both proteins were isolated from haemolymph; therefore, we worked with the mature forms of both of them.

The presence of variable amounts of SP1 as a contaminant in the samples is explained by the fact that this protein is female-specific (Sakurai *et al.*, 1988); the larvae were not segregated according to sex in our haemolymph isolation protocol.

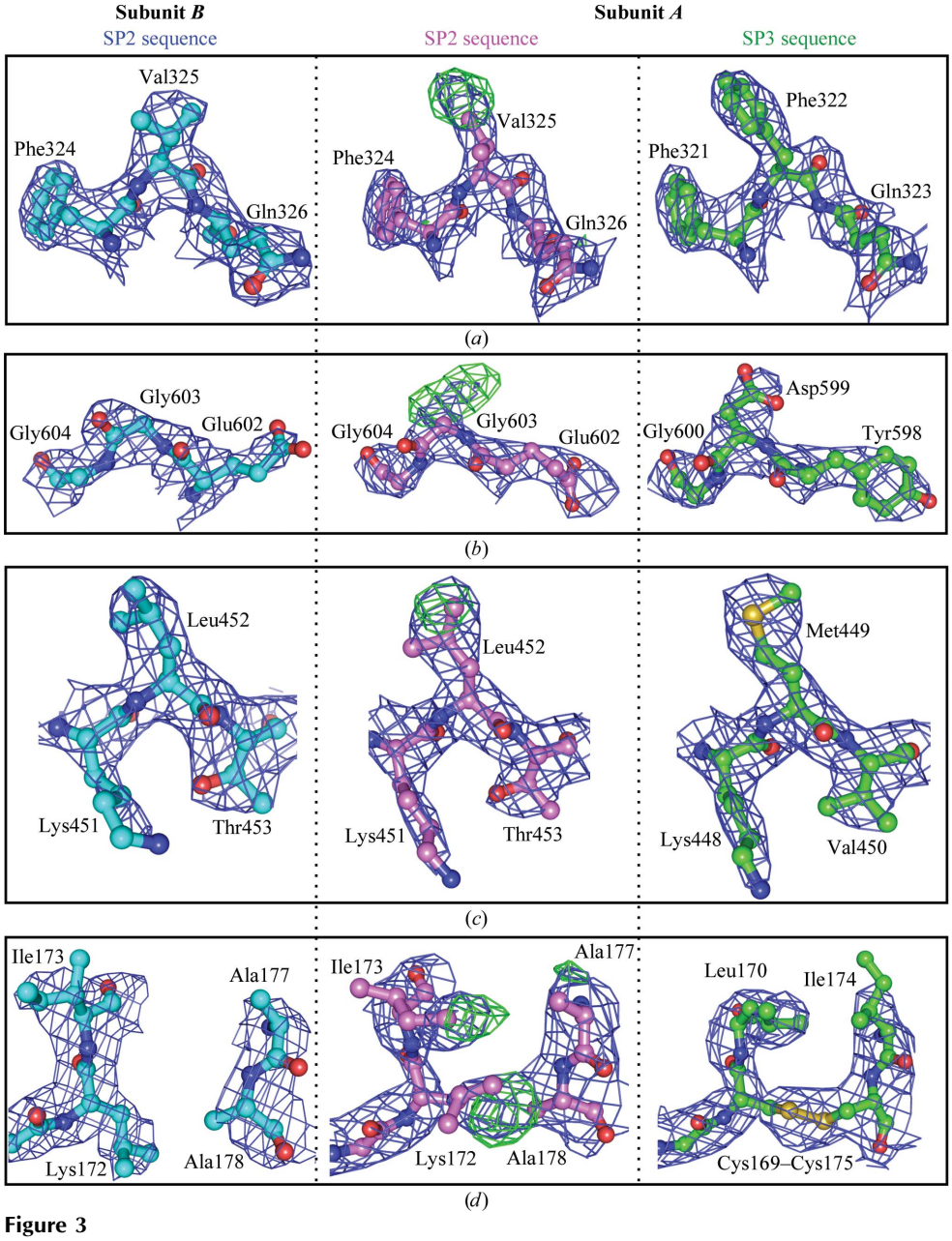

Figure 3
The electron-density maps indicated that the crystallized hexamerin consists of two different proteins. A number of peaks in the difference electron-density maps of chain *A* suggested that the SP2 sequence is not correct for this protein. Four representative sites are illustrated in (*a*–*d*) showing the situation in chain *B* (cyan), where the SP2 sequence matches the electron-density maps perfectly, the corresponding site in chain *A* with the SP2 sequence (magenta) of the model and after its correction to match the SP3 sequence (green). The $2F_o - F_c$ maps (blue) are contoured at the 1.0σ level and the $F_o - F_c$ maps (green) are contoured at the 3.0σ level.

3.3. Overall structure of the SP2–SP3 complex

The crystal structure of the SP2–SP3 complex was refined to final R and R_{free} factors of 16.6 and 22.3%, respectively. The asymmetric unit is comprised of two protein molecules (Fig. 5*a*): one molecule of SP2 (protomer *B*) and one of SP3 (protomer *A*). The biological assembly can be described as a heterohexamer composed of three SP2 molecules and three SP3 molecules. Each trimer is located on the threefold component of the crystallographic 6_3 axis (Fig. 5*b*) and the whole heterohexameric assembly has C_3 point symmetry.

The SP2–SP3 crystal structure (Figs. 6*a*–6*d*) contains 667 of the 687 residues of mature SP2 (9–675) and 674 of the 680 residues of mature SP3 (6–679). The overall fold of both SP2 and SP3 is consistent with the architecture

of APA (Ryu *et al.*, 2009) and of arthropod haemocyanins. The models of both SP2 and SP3 can be divided into three subdomains (Figs. 6*a*–6*d*): the haemocyanin_N domain (http://pfam.sanger.ac.uk; accession code PF03722), the haemocyanin_M domain (PF00372) and the haemocyanin_C domain (PF03723). The haemocyanin_N subdomain, located at the N-termini of SP2 (residues 17–156) and SP3 (14–153), consists only of α-helices (α1–α6 in SP topology). The second subdomain, haemocyanin_M (residues 157–441 in SP2 and 154–438 in SP3), is composed of several α-helices (α7–α14) and β-strands (β2–β5). The haemocyanin_C subdomain, located at the C-termini (residues 442–675 in SP2 and 439–679 in SP3), is classified as an Ig-like domain and contains mostly β-strands (β6–β15), forming an elongated β-barrel plus one α-helix (α15).

Although the overall folds of SP2 and SP3 are similar, the conformation of a number of loops is different in the two protein molecules. The most significant difference is observed for the loop containing residues 617–631 (SP2) or 614–628 (SP3). The distance between the C^{α} atoms of Asp622 (SP2) and the corresponding Glu618 (SP3) from the top of this loop in superposed chains *A* and *B* is 17.01 Å. The r.m.s.d. value for C^{α} superposition (Figs. 6*a*–6*c*) of chains *A* (SP3) and *B* (SP2) is 0.83 Å. Moreover, disulfide bridges are only present in SP3. The disulfide-paired residues are Cys53–Cys104 and Cys169–Cys175.

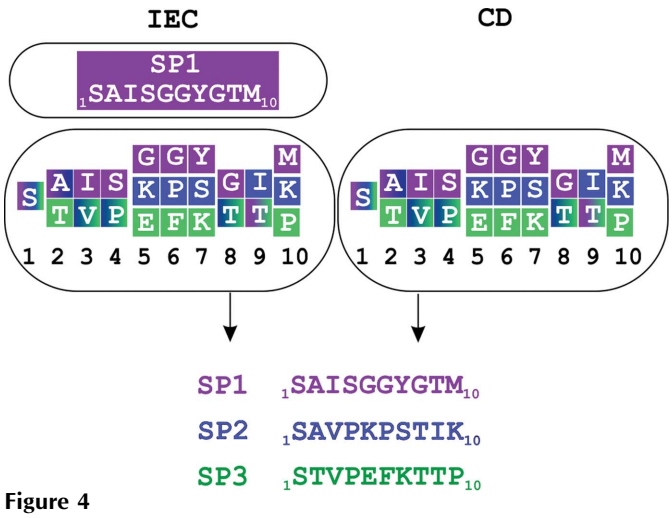

Figure 4
The results of N-terminal sequencing of the SDS–PAGE bands in Fig. 1(*c*). The sequences from Edman degradation for the first decapeptide for a particular band are displayed within the 'boundary' of this band. The colours purple, blue and green represent the sequences of SP1, SP2 and SP3, respectively. The fact that the SP1 protein is present as two bands could be speculatively explained by different glycosylation forms. It can be connected to the results of MALDI-TOF MS (see Fig. 1*f*), in which besides the 84.2 kDa peak corresponding to the components of the hexamerin, two additional peaks (85.1 and 86.9 kDa) could be assigned to SP1 glycosylation forms.

3.4. Arylphorin characteristic glycosylation motif

The glycosylation site found in SP2 (Asn195) and SP3 (Asn192) is conserved in the arylphorin family and is not present in other types of storage proteins (Ryu *et al.*, 2009). The N-glycan linked to chain *A* is located in a deep cleft at the protomer interface between molecules *A* and *B* and an adjacent molecule *B* from the same hexamer. Conversely, the N-glycan of chain *B* occupies the cleft between molecules *B*

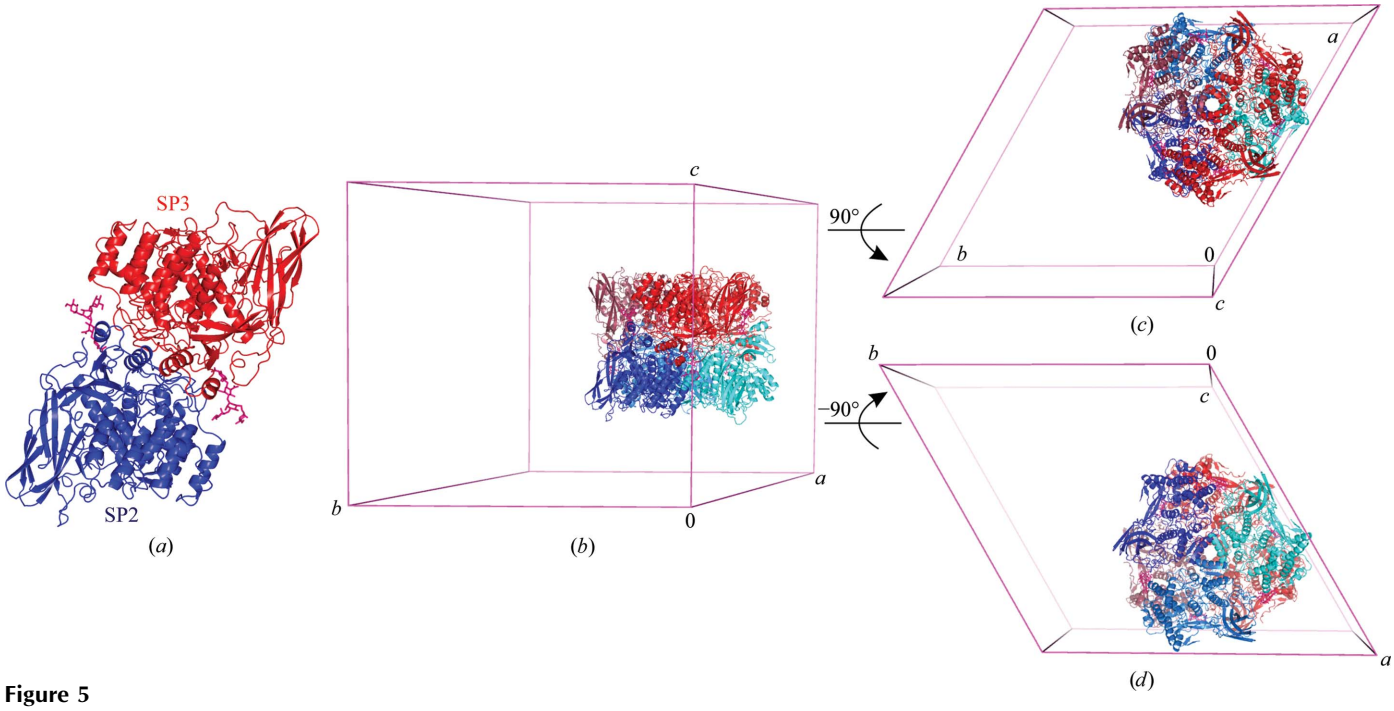

Figure 5
Asymmetric unit and unit-cell contents. (*a*) The asymmetric unit is comprised of two protein chains: SP2 (subunit *B*, blue) and SP3 (subunit *A*, red). (*b*) The biological assembly, a heterohexamer, is shown in the unit cell; symmetry-related copies of SP2 and SP3 are shown in shades of blue and red, respectively. (*c*) The unit cell after a 90° rotation about *x* (horizontal line), with the SP3 trimer facing the viewer. (*d*) The unit cell after a −90° rotation about *x*, with the SP2 trimer facing the viewer.

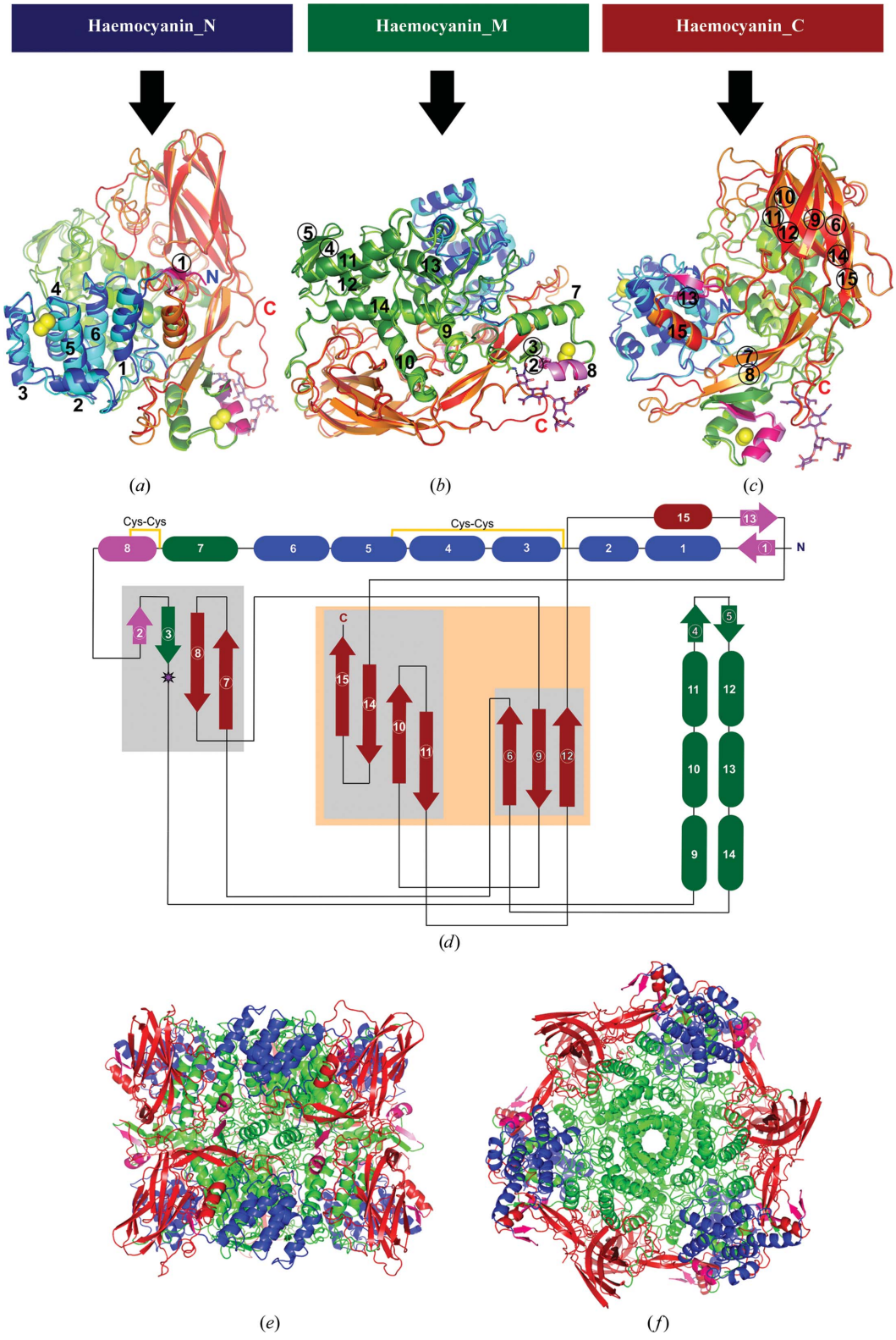

Figure 6
The overall structure of SP2 and SP3. The secondary-structure elements corresponding to haemocyanin_N, haemocyanin_M and haemocyanin_C are shown in blue, green and red, respectively. The secondary-structure elements found only in arylphorins are shown in magenta. (*a*, *b*, *c*) Cartoon representations of the SP2 and SP3 fold. The SP2 (blue–green–red) and SP3 (cyan–light green–orange) chains are superposed in three views to emphasize each of the subdomains: haemocyanin_N (*a*), haemocyanin_M (*b*) and haemocyanin_C (*c*). Four S atoms of cysteine residues forming disulfide bridges in the SP3 structure are shown as yellow spheres. (*d*) A topology diagram of SP2 and SP3. Cylinders and arrows represent α-helices and β-strands, respectively. The lengths of the α-helices and β-strands and of the loops are not commensurate with the numbers of amino-acid residues in these elements. The secondary-structure elements forming β-sheets and an elongated β-barrel are highlighted in grey and orange, respectively. The glycosylation site is marked by a star. (*e*, *f*) Two views of the SP2–SP3 heterohexamer composed of SP2 and SP3 trimers with the subdomains highlighted.

and *A* and an adjacent molecule *A*. Although the resolution of the data was 2.9 Å, the location of the glycan moieties within constricted clefts fixed their conformation and produced electron-density maps in which a large fragment of the modification was clearly visible (Fig. 7). Five sugar units (two *N*-acetylglucosamine, one β-mannose and two α-mannose moieties) in each protein molecule were built in the model. The sequence of the saccharides inserted in the model was based on the experiments reported by Kim *et al.* (2003). The molecular weight of the carbohydrates visible in the electron-density maps is about 1.0 kDa. The theoretical mass of the polypeptide chains (as calculated using *ProtParam*; http://web.expasy.org/protparam) is close to 82 kDa (SP2, 81.8 kDa; SP3, 81.6 kDa), whereas the MALDI-TOF MS analysis estimated the molecular weight of SPs at 84.2 kDa. It brings us to the conclusion that the crystallographic model of SP2–SP3 contains about half of the sugar residues of the N-glycan. The terminal sugar moieties could not be traced in the electron-density maps, most likely owing to their increased mobility.

3.5. Protein–protein interactions

The storage proteins of mulberry silkworm exist in the silkworm body as hexamers (Kanost *et al.*, 1990; Telfer & Kunkel, 1991). However, the two well characterized hexamerins SP1 and SP2 were considered to be homohexamers (Fujii *et al.*, 1989). Our studies show that mulberry silkworm arylphorins occur as heterohexamers composed of SP2 and SP3 trimers. The SP2–SP3 complex was obtained from a mixture of storage proteins and it appears that such a complex is formed more preferably than homohexamers.

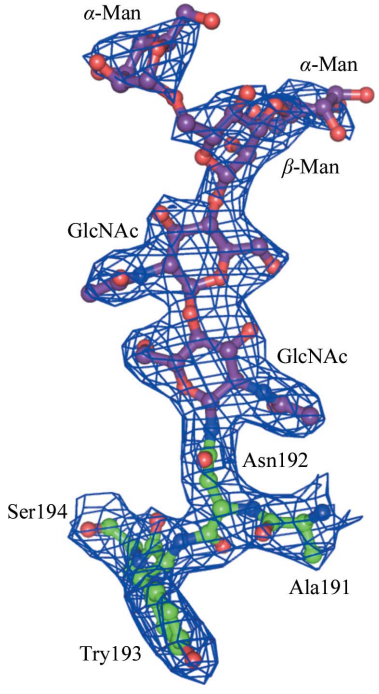

Figure 7
The glycosylation site of SP3. The $2F_o - F_c$ map is contoured at the 1.0σ level.

The interactions between the SP2 and SP3 molecules in the hexamer are strong and the total contact surface area between the six protomers calculated with the *PISA* server (Krissinel & Henrick, 2007) is 43 650 Å^2. The subdomains denoted as haemocyanin_M-like are mainly responsible for hexamer formation (Figs. 6*e* and 6*e*). Structural fragments of this subdomain occupy the central part of the hexamer. The haemocyanin_N-like and haemocyanin_C-like subdomains are located at the periphery of the hexamer.

The hexamer is stabilized by a large number of interactions between the side chains of residues from adjacent protein molecules. The hexamer is additionally stabilized by the N-glycan discussed above, which mediates the interactions between the SP2 and SP3 molecules. It has been reported that the presence of conserved glycosylation motifs improves the overall stability of arylphorin hexamers (Ryu *et al.*, 2009).

3.6. Structural comparisons: arylphorins *versus* haemocyanins

A search for structural homologues of SP2 and SP3 was performed for each of them separately using the *PDBeFold* (*SSM*) server (http://www.ebi.ac.uk/msd-srv/ssm/) and *DALI* (Holm & Rosenström, 2010; http://ekhidna.biocenter.helsinki.fi/dali_server/). Both servers returned the same results. The most similar structure to SP2 and SP3 is that of APA, which was in fact used for our MR calculations. The other proteins with similar folds are arthropodan haemocyanins and phenoloxidases. Detailed information about the arylphorin homologues is summarized in Table 4.

The overall fold of SP2 and SP3 resembles the fold of APA. More differences could be observed when the arylphorins were compared with the haemocyanins. Specifically, several additional β-strands (β1, β2 and β13) and one α-helix (α8) are present in arylphorins; they have been described in detail by Ryu *et al.* (2009). The content of aromatic amino acids in the sequences of arylphorins is significantly higher than in the sequences of haemocyanins. The fraction of aromatic amino acids in SP2, SP3 and APA is 19.0, 20.7 and 19.3% of their total amino-acid content, respectively, whereas the content of aromatic residues in haemocyanins is about 10%. Although both protein families share a similar fold, the level of sequence identity between haemocyanins and arylphorins ranges from 26 to 29%.

The biological roles of arylphorins and haemocyanins are very different and are closely related to the differences in their primary structures. Haemocyanins, as well as the phenoloxidases mentioned earlier, belong to metalloproteins with type 3 copper centres (Decker *et al.*, 2007). Haemocyanins are involved in oxygen transport throughout arthropod bodies (van Holde & Miller, 1995), whereas phenoloxidases catalyse two reactions, namely the *o*-hydroxylation of monophenols to catechols and the oxidization of catechols to *o*-quinones (Decker *et al.*, 2007). Both protein families contain di-copper centres in which two Cu atoms are coordinated by six histidine residues. In haemocyanins the di-copper centre serves as an oxygen-binding site (van Holde & Miller, 1995), whereas in

Table 4
R.m.s.d. values calculated for C^{α} atoms of mulberry silkworm storage proteins superposed on homologous proteins identified by their PDB codes.

The table presents the r.m.s.d. values in Å/the number of superposed C^{α} atoms for SP2 (chain *B*) and SP3 (chain *A*) with chain *A* of a structure representing each of the homologues, identified by their PDB codes: 3gwj (oak silkworm arylphorin; Ryu *et al.*, 2009), 1hc1 (California spiny lobster haemocyanin; Volbeda & Hol, 1989), 1lla (Atlantic horseshoe crab haemocyanin; Hazes *et al.*, 1993), 3ixv (Sahara scorpion haemocyanin; Cong *et al.*, 2009) and 3hhs (tobacco hornworm phenoloxidase; Li *et al.*, 2009).

	3gwj	1hc1	1lla	3ixv	3hhs
SP2	0.83/645	1.70/559	1.90/519	1.99/514	1.92/505
SP3	0.94/657	1.77/568	1.83/510	1.81/500	1.99/519

phenoloxidases it is the catalytic centre of the enzyme (Decker *et al.*, 2006). Arylphorins are not capable of oxygen binding or of catalyzing the above-mentioned reactions because the amino-acid residues present at the corresponding arylphorin site, located in the haemocyanin_M subdomain, are generally not capable of copper coordination. The corresponding site of SP2 and SP3 consists of three tyrosine residues (Tyr219, Tyr250 and Tyr412 in SP3; Tyr222, Tyr253 and Tyr415 in SP2), one arginine (Arg376 in SP3; Arg379 in SP2), one glutamate (Glu372 in SP3; Glu376 in SP2) and one histidine residue (His223 in SP3; His226 in SP2) (Fig. 8). The only histidine residue present at this site of SP2 and SP3 has a different conformation to the corresponding copper-coordinating histidine residue in haemocyanins. These residues belong to helices α9, α10, α13 and α14. Two additional residues (Phe and Glu), important for the enzymatic activity, are present in the active site of phenoloxidases. The phenoloxidase from tobacco hornworm is a heterodimer comprised of prophenoloxidases 1 (PPO1) and 2 (PPO2). The active sites of both proteins, besides the above-mentioned Cu atoms and histidine residues, also contain a phenylalanine residue that serves as a 'placeholder' for the phenolic substrates. This residue is conserved in arylphorins. Additionally, only the PPO2 molecule has *o*-hydroxylation activity, which is connected to the presence of Glu395 in the active site. This glutamate residue acts to deprotonate the hydroxyl group of monophenolic substrates, a role that is essential for the *o*-hydroxylation activity. The corresponding residue in arylphorins is a tyrosine.

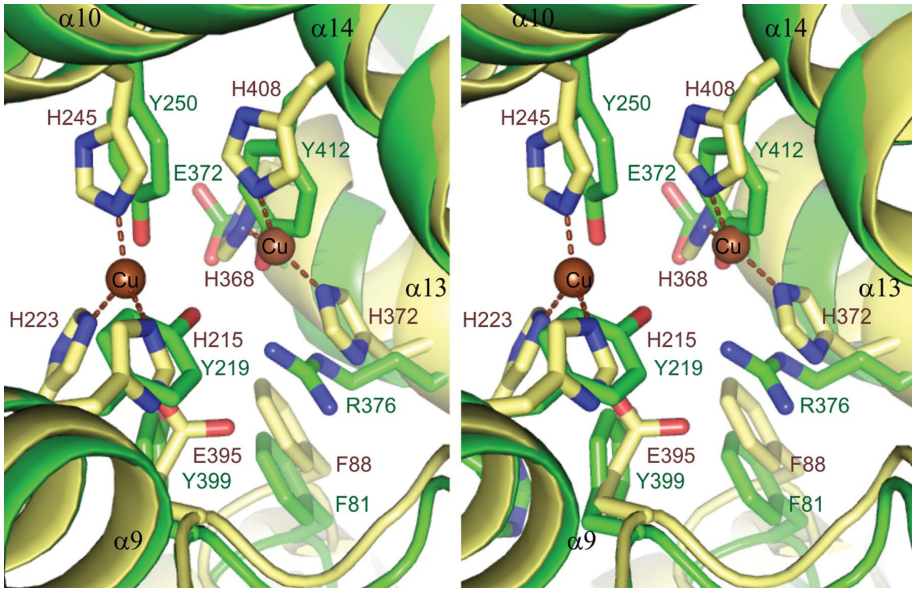

Figure 8
A comparison of the di-copper centre present in prophenoloxidase 2 (PPO2) from tobacco hornworm (PDB code 3hhs; Li *et al.*, 2009) and the corresponding site of arylphorins (stereoview). In order to compare the sites, the model of SP3 (green) has been superposed (C^{α} atoms) on that of PPO2 (yellow).

4. Conclusions

In this work, we have presented the first crystal structures of arylphorins from mulberry silkworm. The experimental electron-density maps were of good quality, allowing us to discover that the hexamerin structure unexpectedly consists of two trimers of two different storage proteins: SP2 and SP3. The crystal structure of the complex has been refined at 2.9 Å resolution. The overall fold of both arylphorins is characteristic of this group of proteins and consists of three haemocyanin-like subdomains. Additionally, the physiological N-glycosylation is clearly visible in the electron-density maps of both proteins: SP2 (at Asn195) and SP3 (Asn192).

The SP2–SP3 complex was isolated from its natural source, the haemolymph of the insect, and crystallized from a mixture of mulberry silkworm storage proteins. In this article, we have not only described the structure of the complex but also the complete methodological path of the discovery, leading from isolation of the complex, through its crystallization, its identification, which was mainly based on X-ray crystallographic data with corroboration from other methods (chemical sequencing, electrophoresis, mass spectrometry), to the final structural interpretation and comparisons.

Finally, our studies of the SP2–SP3 complex provide a significant contribution to the overall knowledge about mulberry silkworm storage proteins. To date, little has been known about SP3 and hexamerins have been assumed to always be homohexamers. Here, we have shown that silkworm arylphorins can exist as heterohexamers.

We would like to thank Dr Piotr Neumann for sharing his *XDS* scripts for efficient *XDS* data processing. We acknowledge BioCentrum, Krakow, Poland for N-terminal sequencing analysis. This work was supported in part by the European Union within the European Regional Development Fund and by grant 2011/03/B/NZ1/01238 from the National Science Centre to GB. The screening for initial crystallization conditions at the High Throughput

Crystallization (HTX) Facility, EMBL Hamburg, Germany was funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 227764 (P-CUBE). The use of the BESSY beamlines received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under BioStruct-X (grant agreement No. 283570).

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Bujacz, G., Wrzesniewska, B. & Bujacz, A. (2010). *Acta Cryst.* D**66**, 789–796.
Cong, Y., Zhang, Q., Woolford, D., Schweikardt, T., Khant, H., Dougherty, M., Ludtke, S. J., Chiu, W. & Decker, H. (2009). *Structure*, **17**, 749–758.
Decker, H., Schweikardt, T., Nillius, D., Salzbrunn, U., Jaenicke, E. & Tuczek, F. (2007). *Gene*, **398**, 183–191.
Decker, H., Schweikardt, T. & Tuczek, F. (2006). *Angew. Chem. Int. Ed. Engl.* **45**, 4546–4550.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Fujii, T., Sakurai, H., Izumi, S. & Tomino, S. (1989). *J. Biol. Chem.* **264**, 11020–11025.
Hazes, B., Magnus, K. A., Bonaventura, C., Bonaventura, J., Dauter, Z., Kalk, K. H. & Hol, W. G. J. (1993). *Protein Sci.* **2**, 597–619.
Holde, K. E. van & Miller, K. I. (1995). *Adv. Protein Chem.* **47**, 1–81.
Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
Hou, Y., Zou, Y., Wang, F., Gong, J., Zhong, X., Xia, Q. & Zhao, P. (2010). *Proteome Sci.* **8**, 45.
Kabsch, W. (2010*a*). *Acta Cryst.* D**66**, 125–132.
Kabsch, W. (2010*b*). *Acta Cryst.* D**66**, 133–144.
Kanost, M. R., Kawooya, J. K., Law, J. H., Ryan, R. O., Van Heusden, M. C. & Ziegler, R. (1990). *Adv. Insect Physiol.* **22**, 299–396.
Kim, S., Hwang, S. K., Dwek, R. A., Rudd, P. M., Ahn, Y. H., Kim, E.-H., Cheong, C., Kim, S. I., Park, N. S. & Lee, S. M. (2003). *Glycobiology*, **13**, 147–157.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Levenbook, L. (1985). *Comprehensive Insect Biochemistry, Physiology and Pharmacology*, edited by L. I. Gilbert & G. A. Kerkut, Vol. 10, pp. 307–346. New York: Pergamon Press.
Li, Y., Wang, Y., Jiang, H. & Deng, J. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 17002–17006.
McCormack, A. L., Schieltz, D. M., Goode, B., Yang, S., Barnes, G., Drubin, D. & Yates, J. R. (1997). *Anal. Chem.* **69**, 767–776.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Mueller, U., Darowski, N., Fuchs, M. R., Förster, R., Hellmig, M., Paithankar, K. S., Pühringer, S., Steffien, M., Zocher, G. & Weiss, M. S. (2012). *J. Synchrotron Rad.* **19**, 442–449.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Ogawa, K. & Tojo, S. (1981). *Appl. Entomol. Zool.* **16**, 288–296.
Painter, J. & Merritt, E. A. (2006). *Acta Cryst.* D**62**, 439–450.
Pietrzyk, A. J., Bujacz, A., Jaskolski, M. & Bujacz, G. (2013). *Biotechnologia*, **94**, 9–14.
Pietrzyk, A. J., Bujacz, A., Łochyńska, M., Jaskólski, M. & Bujacz, G. (2011). *Acta Cryst.* F**67**, 372–376.
Pietrzyk, A. J., Bujacz, A., Mueller-Dieckmann, J., Lochynska, M., Jaskolski, M. & Bujacz, G. (2013). *PLoS One*, **8**, e61303.
Pietrzyk, A. J., Panjikar, S., Bujacz, A., Mueller-Dieckmann, J., Lochynska, M., Jaskolski, M. & Bujacz, G. (2012). *Acta Cryst.* D**68**, 1140–1151.
Rhee, W. J., Lee, E. H., Park, J. H., Lee, J. E. & Park, T. H. (2007). *Biotechnol. Prog.* **23**, 1441–1446.
Riddiford, L. M. & Law, J. H. (1983). *The Larval Serum Proteins of Insects: Function, Biosynthesis, Genetics*, edited by K. Scheller, pp. 75–85. Stuttgart: Thieme.
Ryu, K.-S., Lee, J.-O., Kwon, T. H., Choi, H.-H., Park, H.-S., Hwang, S. K., Lee, Z.-W., Lee, K.-B., Han, Y. H., Choi, Y.-S., Jeon, Y. H., Cheong, C. & Kim, S. (2009). *Biochem. J.* **421**, 87–96.
Sakai, N., Mori, S., Izumi, S., Haino-Fukushima, K., Ogura, T., Maekawa, H. & Tomino, S. (1988). *Biochim. Biophys. Acta*, **949**, 224–232.
Sakurai, H., Fujii, T., Izumi, S. & Tomino, S. (1988). *J. Biol. Chem.* **263**, 7876–7880.
Schägger, H. & von Jagow, G. (1987). *Anal. Biochem.* **166**, 368–379.
Shimada, T., Kobayashi, M. & Yoshitake, N. (1985). *J. Seric. Sci. Jpn*, **54**, 464–469.
Telfer, W. H., Keim, P. S. & Law, J. H. (1983). *Insect Biochem.* **13**, 601–613.
Telfer, W. H. & Kunkel, J. G. (1991). *Annu. Rev. Entomol.* **36**, 205–228.
The International Silkworm Genome Consortium (2008). *Insect Biochem. Mol. Biol.* **38**, 1036–1045.
Tojo, S., Liu, Y. & Zheng, Y. (2012). *Hemolymph Proteins and Functional Peptides: Recent Advances in Insects and Other Arthropods*, edited by M. Tufail, pp. 32–61. Bentham eBooks. doi:10.2174/97816080540151120101.
Vanishree, V., Nirmala, X., Arul, E. & Krishnan, M. (2005). *Invertebr. Reprod. Dev.* **48**, 81–88.
Volbeda, A. & Hol, W. G. J. (1989). *J. Mol. Biol.* **209**, 249–279.
Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. (2009). *Bioinformatics*, **25**, 1189–1191.
Yu, W., Wang, M., Zhang, H., Quan, Y. & Zhang, Y. (2013). *Int. J. Genomics*, **2013**, 145450.