Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

High-resolution structure of *Bombyx mori* lipoprotein 7: crystallographic determination of the identity of the protein and its potential role in detoxification

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

[a]Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland, [b]European Molecular Biology Laboratory, Hamburg Outstation, c/o Deutsches Elektronen-Synchrotron, Notkestrasse 85, Hamburg, Germany, [c]Australian Synchrotron, 800 Blackburn Road, Clayton, Victoria 3168, Australia, [d]Institute of Technical Biochemistry, Faculty of Biotechnology and Food Sciences, Technical University of Lodz, Stefanowskiego 4/10, 90-924 Lodz, Poland, [e]Institute of Natural Fibres and Medicinal Plants, Wojska Polskiego 71B, 60-630 Poznan, Poland, and [f]Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Grunwaldzka 6, 60-780 Poznan, Poland

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

Received 4 April 2012
Accepted 11 May 2012

PDB References: *Bombyx mori* lipoprotein 7, 4efp; 4efq; 4efr.

Three crystal structures of a lipoprotein (Bmlp7) of unknown function, a member of the 30 kDa lipoprotein family from mulberry silkworm (*Bombyx mori* L.) haemolymph, have been determined. The 1.33 Å resolution structure is an excellent example of how a precise crystallographic study can contribute to protein identification. The correct sequence of this haemolymph-isolated protein was assigned thanks to superb-quality electron-density maps. Two unexpected cadmium cations were found in this crystal structure [Bmlp7-I(Cd)] and their presence may be connected to a detoxification mechanism in this insect. For a comparison of the metal-binding sites, the crystal structure of a platinum complex (Bmlp7-Pt) was also solved at 1.94 Å resolution. The third (2.50 Å resolution) structure, of the native protein harvested in a different season (Bmlp7-II), corresponds to a different polymorph with an altered pattern of intermolecular interactions and with a total absence of cadmium ions and highlights the possible involvement of Bmlp7 in the response to environmental pollution. The N-terminal domain of Bmlp7 has a fold resembling a clockwise spiral created by six helices and can be classified as a VHS domain. The C-terminal domain is folded as a β-trefoil. The biological function of Bmlp7 is unknown, but its structural homology to sugar-binding proteins suggests that, in analogy to other 30 kDa haemolymph lipoproteins, it could play a role as an anti-apoptotic factor or function in the immune response of the insect to fungal infections.

1. Introduction

30 kDa proteins are specific to the fifth-instar larvae of the mulberry silkworm *Bombyx mori* L. They are often referred to as major haemolymph proteins owing to their high content in the haemolymph. It has been confirmed that accumulation of these proteins in the haemolymph is developmentally regulated in a stage-dependent manner at the transcriptional level (Mori *et al.*, 1991). Synthesis of the 30 kDa proteins occurs in the so-called fat body and the proteins are subsequently secreted to the haemolymph (Gamo, 1978). Five cDNA clones with distinct mRNA sequences, annotated as pBmHP-6, pBmHP-12, pBmHPC-19, pBmHPC-21 and pBmHPC-23, were isolated by Sakai *et al.* (1988) from a cDNA library constructed from the fat-body mRNA of fifth-instar larvae. The proteins expressed from these constructs are denoted PBMHP-6, PBMHP-12, PBMHPC-19, PBMHP-21 and PBMHPC-23, respectively (Sakai *et al.*, 1988). The sequences of additional members of the 30 kDa lipoprotein family, 6G1,

© 2012 International Union of Crystallography
Printed in Singapore – all rights reserved

19G1 and 21G1, have been reported by Mori *et al.* (1991). Ten further 30 kDa lipoprotein sequences (Bmlp1–10) were found in a bioinformatics analysis of the genes coding these proteins (Sun *et al.*, 2007).

The biological function of many members of the 30 kDa lipoprotein family is still unknown. However, some of them are involved in the immune response to fungal infection. Two members of the lipoprotein family, 19G1 and 6G1, specifically bind glucose and glucans, which suggests that they might be involved in the antifungal defence system of the insect (Ujita *et al.*, 2002). The fungal cell wall consists of glycoproteins, glucans and chitin, which are present on the cell surface (Shepherd, 1987). The most important and common virulence determinants of pathogenic fungi are glucans (Klimpel & Goldman, 1988), especially β-1,3-glucans and β-1,6-glucans, which are the major component of the fungal cell wall (Jiang *et al.*, 2004). Thus, insect glucan-binding proteins could be essential components of the antifungal mechanism. In the course of the study of glucan-binding proteins from silkworm haemolymph, the two lipoproteins mentioned above (19G1 and 6G1) were identified using N-terminal sequencing. Other glucose-binding proteins could not be identified, possibly because of N-terminal modification. Moreover, analysis of the saccharide-binding properties of the lipoproteins 19G1 and 6G1 indicated that they interact specifically with glucose and glucans but do not interact with other monosaccharides such as *N*-acetylglucosamine, galactose, *N*-acetylgalactosamine, mannose or fucose (Ujita *et al.*, 2005). Further studies revealed that the 6G1 lipoprotein specifically recognizes laminaribose (Glcβ1–3Glc) and laminarin, a soluble β-glucan, but does not bind trehalose, an insect blood sugar, or glycoproteins. Moreover, it has been shown that the 6G1 lipoprotein bound to β-glucan present on the surface of invading fungi participates in the activation of a prophenoloxidase cascade, which involves a serine protease pathway (Ujita *et al.*, 2005). The phenoloxidase cascade generates quinones, which are intermediates in melanization. At a later stage, melanin forms a complex with lipophorin to encapsulate the invading microorganisms (Gupta, 1991). 6G1 and 19G1 are homologous to vitellogenins (Ujita *et al.*, 2005), which are identical to haemagglutinins, from the Colorado potato beetle *Leptinotarsa decemlineata* (Gupta, 1991).

The 30 kDa lipoproteins also have anti-apoptotic properties (Rhee & Park, 2000; Kim *et al.*, 2001; Choi *et al.*, 2002). The addition of silkworm haemolymph to a baculovirus-induced insect-cell culture inhibited apoptosis of the cells. When the medium was supplemented with 5% haemolymph, the viability of the host insect cells remained high for 6 d after baculovirus infection (Rhee & Park, 2000). The apoptosis-inhibiting component of the haemolymph was isolated and characterized as belonging to the 30 kDa lipoprotein family, with a molecular mass of 28 kDa (Kim *et al.*, 2001). The N-terminal amino-acid sequence of Bmlp7 indicates homology to the protein characterized by Kim *et al.* (2001). The haemolymph can also inhibit apoptosis of human cells. In this case, vaccinia virus was used to induce apoptosis of HeLa cells (Choi *et al.*, 2002). For further studies, the PBMHP-6 product was used as a representative 30 kDa protein with anti-apoptotic properties. Its effect on apoptosis was tested using human embryonic kidney cell culture (HEK293) and Chinese hamster ovary cell culture (CHOK1). The addition of PBMHP-6 decreased the activity of caspase 3, a protein that is involved in apoptosis (Kim *et al.*, 2004).

In this work, we determined three crystal structures of a 30 kDa protein purified from silkworm larvae haemolymph. Previously, based on mass-spectrometric analysis, the protein was tentatively identified as PBMHP-12 (Pietrzyk *et al.*, 2011), but the present detailed crystallographic analysis has revealed that the protein is in fact Bmlp7, annotated in the Silkworm Knowledgebase (http://silkdb.genomics.org.cn/silkworm/index.jsp; accession code Bmb021422) as *B. mori* lipoprotein 7. The same sequence (but incomplete) could also be found in UniProt (http://www.uniprot.org) without a signal peptide as the putative uncharacterized protein 19G1P (accession code D4QGC0) or with a fragment of the signal peptide as a 30 kDa protein (accession code E5EVW2). While the present manuscript was in preparation, the structure of recombinant Bmlp7 at 1.91 Å resolution was reported by Yang *et al.* (2011). The quality of the recombinant Bmlp7 structure (PDB entry 3pub) was good ($R_{\rm work}$ = 19.5%, $R_{\rm free}$ = 23.5%). The overall shape and fold of the recombinant and haemolymph-isolated Bmlp7 models are almost identical. The present work, which describes the structure of natural rather than recombinant Bmlp7, shows that the protein has an unexpected affinity for cadmium cations, which can be found in the haemolymph. The structure of the native Bmlp7–cadmium complex is denoted Bmlp7-I(Cd). In addition, the presence of an oxidized tryptophan residue has been discovered. Moreover, the structure of Bmlp7 in complex with platinum (Bmlp7-Pt) has also been solved in order to compare the heavy-metal binding sites with those in Bmlp7-I(Cd). Finally, the third structure reported here (Bmlp7-II) represents native Bmlp7 isolated from a batch of Cd-free haemolymph collected two years later. The present structural studies of the Bmlp7 protein are only the first step towards the elucidation of the mode of action of the 30 kDa lipoproteins in insect immune response and apoptosis inhibition. They are also the starting point for the elucidation of the biological function of the Bmlp7 protein and the identification of its potential binding partners.

2. Materials and methods

2.1. Haemolymph collection, protein purification and crystallization

B. mori haemolymph was collected from fifth-instar larvae two times: in 2008 and two seasons later in 2010. Bmlp7 was purified and crystallized as described by Pietrzyk *et al.* (2011). Briefly, a two-step purification protocol was applied. The first step consisted of gel filtration carried out using a Superdex 200 prep-grade column (XK 16/100; Amersham Biosciences). The collected fractions containing the 30 kDa proteins were concentrated and applied onto a Q Sepharose column (XK 16/10; Amersham Biosciences). For protein separation,

stepwise elution was used. Bmlp7 was eluted with 60 m*M* NaCl. The collected peak fractions containing Bmlp7 were concentrated to 10 mg ml^{-1}. Crystals were grown by the hanging-drop vapour-diffusion method at 293 K using 0.2 *M* KSCN, 0.1 *M* HEPES pH 7.5, 22% PEG 3350 as the precipitating buffer. In a test experiment aimed at reproducing the native Bmlp7-I(Cd) crystals, the same conditions were used for cocrystallization of Bmlp7 with cadmium, which was externally added to the crystallization drop as $CdCl_2$ at a final concentration of 0.2 m*M* (yielding Bmlp7-Cd). The Bmlp7-I(Cd) (without the addition of external Cd^{2+}), Bmlp7-Hg and Bmlp7-Pt crystals were grown using Bmlp7 protein isolated from the haemolymph collected in 2008. Bmlp7 purified from the haemolymph collected in 2010 was used for crystallization of Bmlp7-II and Bmlp7-Cd (cocrystallized with artificially added cadmium).

2.2. Data collection and processing

X-ray diffraction data for Bmlp7-I(Cd) crystals were collected on beamline BL14.1 of the BESSY synchrotron, Berlin, Germany (Mueller *et al.*, 2012). The data were of very good quality and extended to 1.33 Å resolution (Pietrzyk *et al.*, 2011).

At the time when the diffraction data for the presented Bmlp7-I(Cd) structure were collected, there was no suitable model for this protein in the Protein Data Bank for use in molecular replacement. Thus, other methods based on halide and heavy-metal derivatives were used to solve the phase problem. Many derivatives (substances including Hg, Pt, Ta, Os, Au, Ag, Pt, Br and I atoms) were used, but most of them turned out to be unsuccessful. Some of these agents had a damaging effect on the crystals, while others did not bind to the protein. It is likely that in several cases the strongly coordinated cadmium cations prevented the derivatizing agent from binding to the protein. Finally, the best results were obtained by soaking the crystals in 2 m*M* mercury(II) acetate or 2 m*M* K_2PtCl_6 for one week. Single-wavelength anomalous diffraction (SAD) data for the Hg and Pt derivatives (Bmlp7-Hg and Bmlp7-Pt) were collected to 2.22 Å resolution with a 1.0055 Å wavelength (selected according to a fluorescence scan) for Bmlp7-Hg and to 1.94 Å resolution with a 1.071 Å wavelength for Bmlp7-Pt on beamline X12 of the DESY synchrotron, Hamburg, Germany. The detector was a MAR CCD 225 and the data were recorded from a single crystal using the rotation method with an oscillation of 0.5° and exposure times of 6 s for the Hg derivative and 4 s for the Pt derivative at a temperature of 100 K. A solution consisting of 0.1 *M* HEPES pH 7.0, 50% PEG 400 and 200 m*M* KSCN mixed with the reservoir solution in a 1:1 ratio was used as a cryoprotectant for the derivative crystals. Before cryocooling in a stream of nitrogen gas, the crystals were transferred into the cryoprotectant solution for a few seconds.

X-ray diffraction data for Bmlp7-II and Bmlp7-Cd were also collected on beamline X12 of the DESY synchrotron. Crystals of Bmlp7-II diffracted X-rays to 2.50 Å resolution with a 0.949 Å wavelength, an exposure time of 20 s and a temperature of 100 K. The diffraction data for Bmlp7-Cd were collected to 1.79 Å resolution with a 0.900 Å wavelength, an exposure time of 15 s and a temperature of 100 K. The rotation method with an oscillation of 0.5° was used in both cases. A solution consisting of 0.1 *M* HEPES pH 7.0, 20% PEG 400, 200 m*M* KSCN and 20 m*M* glucose mixed with the reservoir solution in a 1:1 ratio was used as a cryoprotectant.

The previously collected diffraction images for Bmlp7-I(Cd) (Pietrzyk *et al.*, 2011) were reprocessed using *HKL*-2000 (Otwinowski & Minor, 1997). Although the data-collection statistics were similar to those reported previously (Pietrzyk *et al.*, 2011), structure refinement progressed much better using the reprocessed data. The data for Bmlp7-Hg and Bmlp7-Pt were also indexed, integrated and scaled using *HKL*-2000 (Otwinowski & Minor, 1997). Diffraction images of Bmlp7-II and Bmlp7-Cd were processed 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 crystal structure of natural-source Bmlp7 was determined using a combination of the SAD and native MRSAD protocols of *Auto-Rickshaw* (Panjikar *et al.*, 2005, 2009). Initial phases were determined using the Hg derivative. The input diffraction data were prepared and converted for input to *Auto-Rickshaw* using programs from the *CCP*4 suite (Winn *et al.*, 2011). Based on an initial analysis of the data, the maximum resolution for substructure determination and initial phase calculation was set to 3.0 Å. All of the five heavy atoms requested were found using the program *SHELXD* (Sheldrick, 2008). The correct space-group enantiomorph for the substructure was determined using the programs *ABS* (Hao, 2004) and *SHELXE* (Sheldrick, 2008). The heavy-atom refinement and initial phase calculation were performed in the program *BP*3 (Pannu *et al.*, 2003). Density modification was carried out with the program *RESOLVE* (Terwilliger, 2000). The resulting phases were used for automated model building in *ARP*/*wARP* (Perrakis *et al.*, 2001). At this stage, the resulting model contained very few meaningful secondary structures. This model, together with the high-resolution native and Hg-derivative data, as well as the initial heavy-atom sites, were used in the native MRSAD protocol of *Auto-Rickshaw* (to be published elsewhere). Within the protocol, the Hg-derivative data set was used for heavy-atom phasing and the high-resolution native data set was used for phase extension, density modification and automated model building. The procedure resulted in a total of 460 residues as a polyalanine model with excellent-quality electron-density maps. This allowed the identification of the protein sequence using the electron-density maps and the structure was completed in several manual model-building cycles carried out in *Coot* (Emsley & Cowtan, 2004) and was refined using *REFMAC*5 (Murshudov *et al.*, 2011) and *PHENIX* (Adams *et al.*, 2010). The Bmlp7-I(Cd) structure was refined with anisotropic ADP parameters to the full resolution of 1.33 Å. The final model consisted of two protein chains (*A* and *B*)

Table 1
Diffraction data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	Bmlp7-I(Cd)	Bmlp7-Hg	Bmlp7-Pt	Bmlp7-II	Bmlp7-Cd
Space group	*P*1	*P*1	*P*1	*P*1	*P*1
Unit-cell parameters					
a (Å)	42.2	41.7	41.8	35.1	42.9
b (Å)	50.0	49.7	49.7	71.7	50.7
c (Å)	55.2	54.8	54.9	105.1	55.9
α (°)	93.4	93.5	93.2	78.8	93.3
β (°)	94.6	94.7	94.8	89.9	94.7
γ (°)	102.7	103.0	102.8	75.8	102.6
Molecules in asymmetric unit	2	2	2	4	2
V_M (Å^3 Da^{-1})	2.10	2.02	2.03	2.31	2.07
Solvent content (%)	41.5	39.1	39.4	46.8	40.5
X-ray data collection					
Temperature (K)	100	100	100	100	100
Radiation source	BL14.1, BESSY	X12, DESY	X12, DESY	X12, DESY	X12, DESY
Wavelength (Å)	0.918	1.005	1.071	0.949	0.900
Detector	MAR CCD 225	MAR CCD 225	MAR CCD 225	MAR CCD 225	MAR CCD 225
Crystal-to-detector distance (mm)	140	230	180	200	200
Rotation range (°)	0.5	0.5	0.5	0.5	0.5
Total rotation (°)	360	360	360	180	360
Exposure per image (s)	5.2	6.0	4.0	20.0	15.0
Resolution (Å)	23–1.33 (1.38–1.33)	50–2.22 (2.30–2.22)	50–1.94 (2.01–1.94)	37.0–2.5 (2.56–2.50)	40.0–1.79 (1.90–1.79)
Mosaicity (°)	0.70–1.20	0.59–1.33	0.76–1.54	0.80	0.98
Intensities measured	309628	72420	110745	52068	138001
Unique reflections	94702	40082	29990	29885	38532
R_{merge}† (%)	10.9 (31.9)	8.1 (34.3)	6.8 (15.3)	4.3 (14.7)	5.2 (12.7)
Multiplicity	3.3 (1.6)	1.9 (1.6)	3.7 (2.8)	1.8 (1.6)	3.6 (3.0)
$\langle I/\sigma(I)\rangle$	11.66 (2.42)	8.98 (2.20)	25.2 (4.98)	13.9 (4.67)	18.2 (7.98)
Completeness (%)	94.0 (78.8)	94.7 (76.0)	94.2 (75.0)	88.8 (88.8)	93.9 (80.4)
Refinement					
R_{work}‡/R_{free}‡ (%)	18.36/23.06		18.53/22.67	20.11/28.53	
R_{free} test-set count	1223		1019	949	
No. of protein atoms	4006		3871	7702	
No. of solvent atoms	551		234	137	
No. of metal atoms	4		9	0	
R.m.s. deviations					
Bond lengths (Å)	0.020		0.019	0.020	
Bond angles (°)	1.94		1.69	1.98	
Average *B* factor (Å^2)	16.07		10.52	20.5	
Ramachandran plot					
Favoured φ/ψ (%)	97.0		95.86	95.45	
Disallowed φ/ψ (%)	0.00		0.22	0.32	
PDB code	4efp		4efq	4efr	

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

containing residues Val5–Phe239. The model included 551 water molecules, two potassium cations, two thiocyanate anions (SCN^-) and two heavy-metal sites identified as cadmium cations. Several polyethylene glycol fragments are also visible in the structure.

The structures of Bmlp7-Pt and Bmlp7-II were determined using the *Auto-Rickshaw* MR protocol (Panjikar *et al.*, 2005) with the coordinates of Bmlp7-I(Cd) as the starting model. The final models were produced after several cycles of manual rebuilding in *Coot* (Emsley & Cowtan, 2004) and TLS restrained refinement in *REFMAC*5 (Murshudov *et al.*, 2011). Water molecules were added manually during the iterative cycles of rebuilding and refinement. The TLS groups were defined according to the *TLSMD* server (Painter & Merritt, 2006). The final model of Bmlp7-Pt was highly similar to the Bmlp7-I(Cd) model. The main difference is the presence of six platinum-binding sites in Bmlp7-Pt. The Bmlp7-II crystals have a different crystal form. The arrangement of the four (instead of two) independent protein molecules in the Bmlp7-II structure differs significantly from the Bmlp7-I(Cd) and Bmlp7-Pt structures. The geometry of all models was assessed using *PROCHECK* (Laskowski *et al.*, 1993).

2.4. NMR measurements

Samples from the Bmlp7-I(Cd) crystallization experiment were analyzed by NMR for the presence of cadmium. 133.188 MHz ^{113}Cd NMR spectra were recorded using a Bruker Avance III 600 spectrometer at a magnetic field of 14.0954 T. 27 259 and 25 639 scans were accumulated for the crystallization drop and well solution samples, respectively. The spectral width was 200 kHz and the relaxation delay was 1 s for both samples.

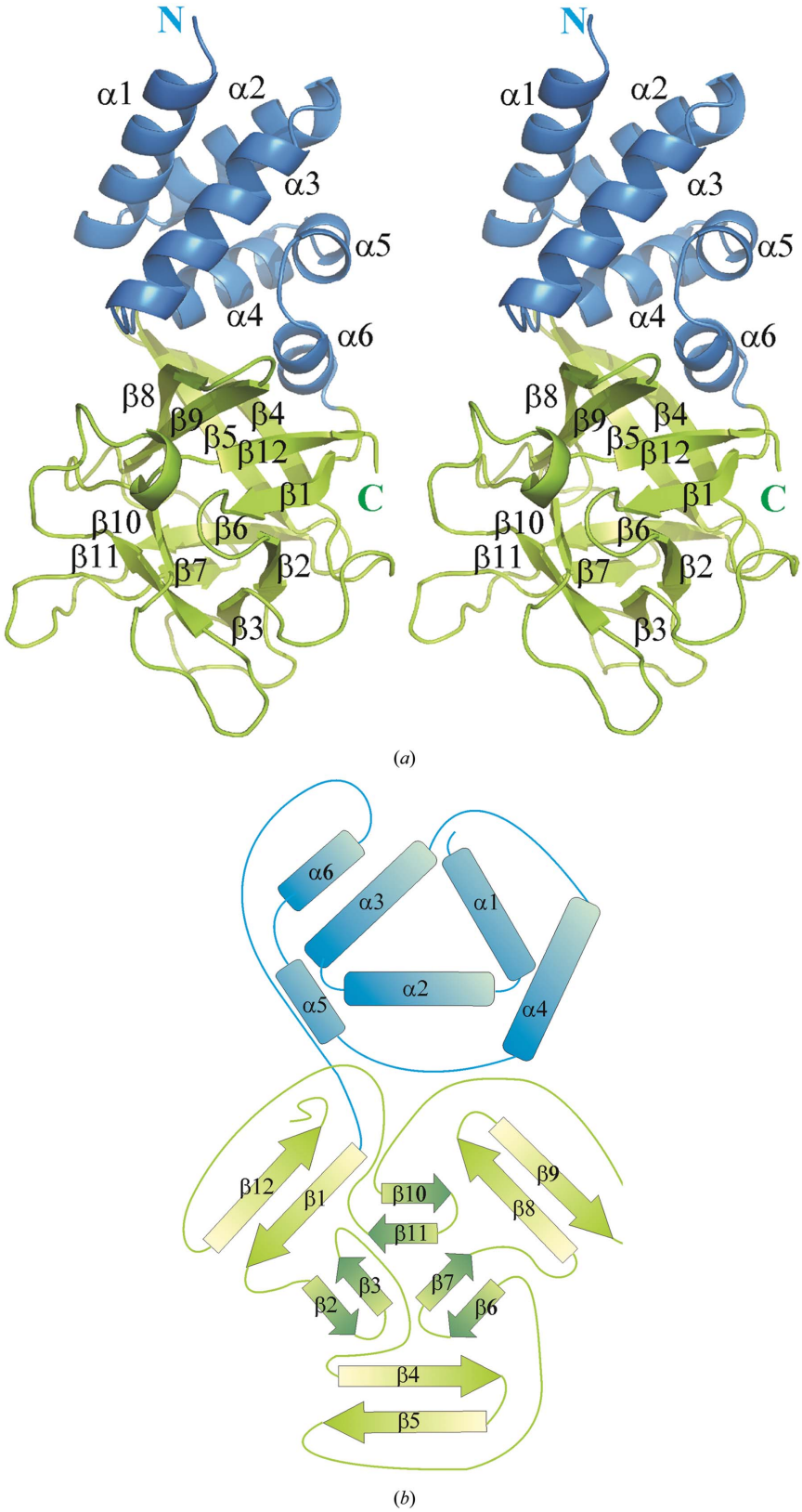

Figure 1
Overall structure of Bmlp7. (*a*) A cartoon representation of the Bmlp7 fold (stereoview). The N-terminal domain is shown in blue and the C-terminal β-trefoil domain is shown in green. Secondary-structure elements are numbered $\alpha1$–$\alpha6$ and $\beta1$–$\beta12$. (*b*) Topology diagram of Bmlp7. Cylinders and arrows represent α-helices and β-strands, respectively. The lengths of the linkers between secondary-structure elements are not commensurate with the numbers of amino-acid residues in the loops.

2.5. N-terminal sequencing analysis

The Bmlp7 sample for Edman degradation was first separated by gel electrophoresis according to Schägger & von Jagow (1987). The protein was transferred to a PVDF Immobilon membrane (PSQ 0.22 µm; Millipore). The single protein band corresponding to the molecular weight of 30 kDa was cut and subjected to Edman degradation cycles performed using a fully automated sequencer (Procise 491; Applied Biosystems).

3. Results and discussion

3.1. The overall fold of Bmlp7, identification of the protein and structural comparison

The *B. mori* haemolymph protein Bmlp7 used in this study was isolated from its natural source and purified using a two-step purification protocol. All crystals of Bmlp7 grown in this study were triclinic, but they belonged to two distinct polymorphic forms. X-ray diffraction data were collected from the natural cadmium-containing protein crystal [Bmlp7-I(Cd)] to 1.30 Å resolution (Pietrzyk *et al.*, 2011), but the data set was finally processed to 1.33 Å resolution. X-ray diffraction data collected from the Hg derivative at its absorption edge were sufficient to solve the phase problem by SAD and to obtain an initial model. Refinement was carried out using anisotropic ADP parameters. The final R factor was 18.36% ($R_{\rm free}$ = 23.06%). The N-terminal domain (residues 5–87) is helical, in contrast to the C-terminal domain (residues 88–239) which has the form of a highly deformed β-barrel with a β-trefoil fold (Fig. 1). The six α-helices of the N-terminal domain create a clockwise spiral assembly. This domain could be classified as a VHS domain (Schultz *et al.*, 1998), which is usually a right-handed superhelix of eight helices. To date, Bmlp7 is a unique example of a VHS domain that is composed of only six helices. The C-terminal domain can be divided into two parts, each containing six β-strands arranged as a three-blade β-propeller. Each blade consists of a single β-hairpin. The β-strands are longer in the subdomain closer to the helical domain and form an assembly reminiscent of a barrel. The β-strands in the inner propeller are much shorter. The C-terminal domain also

Table 2
Sequence homology among silkworm 30 kDa haemolymph proteins.

The table presents percentage sequence similarities (and identities) among different 30 kDa proteins from *B. mori* haemolymph calculated using *BLAST* (http://blast.ncbi.nlm.nih.gov/Blast.cgi; Altschul *et al.*, 1990).

	PBMHP-12	PBMHPC-19	PBMHPC-21	PBMHPC-23	Bmlp7
PBMHP-6	66 (40)	67 (45)	65 (44)	67 (42)	67 (45)
PBMHP-12		64 (50)	91 (79)	98 (93)	66 (49)
PBMHPC-19			64 (50)	68 (48)	97 (94)
PBMHPC-21				83 (71)	67 (52)
PBMHPC-23					67 (49)

contains a number of loop structures, which are well defined.

The first cycles of refinement performed with *REFMAC*5 (Murshudov *et al.*, 2011) and visual inspection of the electron-density maps (Fig. 2) in *Coot* (Emsley & Cowtan, 2004) clearly revealed that the protein was not PBMHP-12, as was previously suspected (Pietrzyk *et al.*, 2011). Subsequently, other sequences of proteins from the 30 kDa lipoprotein family were checked in order to match the correct sequence. Amino-acid sequence alignment of these proteins indicates a high level of homology (Table 2). The main differences were found in the N-terminal region and in the positions of aromatic residues (Fig. 3). The clearly visible side chains in the 1.33 Å electron-density maps allowed correction of the sequence of the model. For instance, residue 131, which is Leu in PBMHP-12, had to be replaced by phenylalanine in the Bmlp7-I(Cd) structure. Finally, the correct sequence of the crystallized protein was found to be identical to the sequence deposited in GenBank under accession code BAJ04788.1, corresponding to a *B. mori* hypothetical protein. The same sequence was also found in UniProt (http://www.uniprot.org) as putative uncharacterized protein 19G1P (accession code D4QGC0) or a 30 kDa protein (accession code E5EVW2) and in the Silkworm Knowledgebase (http://silkdb.genomics.org.cn/silkworm/index.jsp; accession code Bmb021422) as *B. mori* lipoprotein 7 (Bmlp7) supplemented by an additional signal peptide. However, none of these sequence depositions shows which amino acid is the N-terminal residue of the mature protein. The sequences deposited without signal peptide were assigned as incomplete and the cleavage sites are not indicated in the sequences containing signal peptides.

After the structure had been determined, a protein sample (corresponding to the 2010 isolation) was submitted to N-terminal sequencing by chemical degradation. Interestingly, the analysis returned two completely different N-terminal sequences: ADSDVPNDIL– and GVVELSADSM–. There was no ambiguity about which amino-acid residue belonged to which sequence because each step of Edman degradation produced two peaks on the chromatogram, one of which was significantly higher than the other. According to these results, it became obvious that the protein sample used for crystallization was nonhomogenous and consisted of two proteins in an approximate 3:1 ratio. The ADSDVPNDIL– sequence represents the major component and corresponds to the N-terminal fragment of Bmlp7. The minor-component sequence corresponds to PBMHP-12, which explains why the Bmlp7-dominated sample was previously identified as PBMHP-12 (Pietrzyk *et al.*, 2011).

It therefore seems that in this case crystallization was the final purification step. Only one protein (Bmlp7) crystallized from the nonhomogenous sample, while the other (PBMHP-12) remained in solution. The current high-resolution crystallographic model of the Bmlp7-I(Cd) protein is composed of two protein molecules, each containing residues 5–239 of the 239 residues of the mature protein. The residues in all three Bmlp7 structures were numbered according to the sequence of the mature protein.

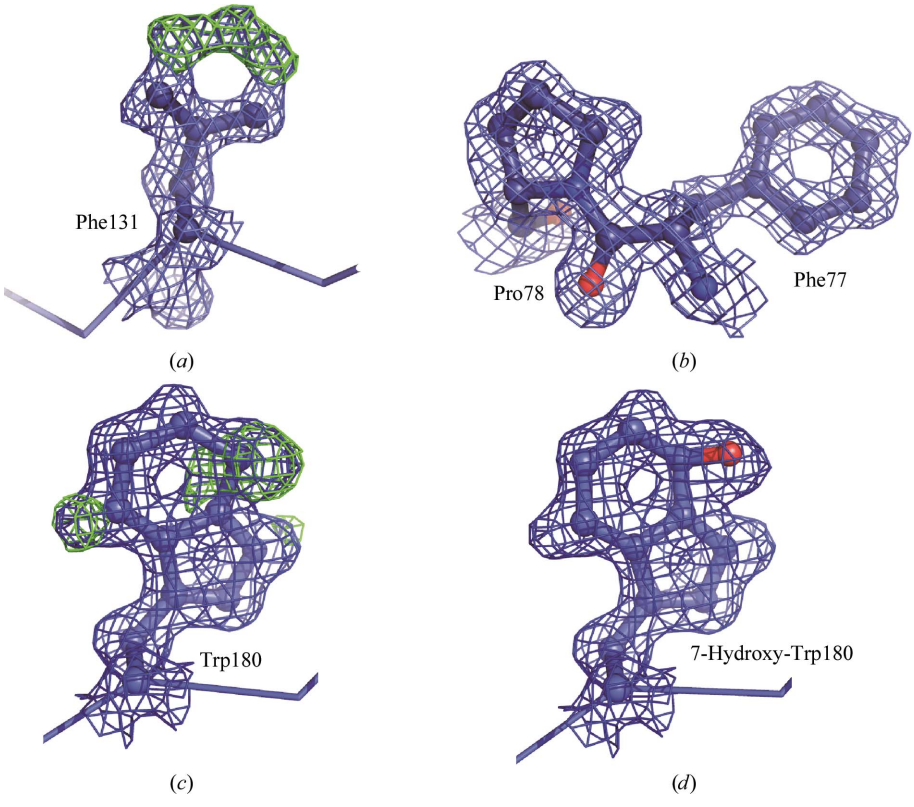

Figure 2
The electron-density maps are of superb quality, enabling verification and correction of the amino-acid sequence, as well as identification of the modified residues. (*a*) The shape of the electron density shows phenylalanine instead of Leu131. (*b*) All residues with annular side chains were easy to identify; Phe77 and Pro78 are presented as an example. (*c*) The geometry of the tryptophan residue (Trp180*A*/*B*) was distorted by a high positive peak in the 1.33 Å electron-density map of the Bmlp7-I(Cd) structure. (*d*) The most probable modification is oxidation of the tryptophan ring at C7. The $2F_o - F_c$ maps are displayed in blue at 1.3σ and the $F_o - F_c$ maps are displayed in green at 3.0σ.

The electron-density maps also revealed the presence of modified residues in the structures. Two modified tryptophan residues (Trp180*A*/*B*), one in each protein molecule in the asymmetric unit, are present in the Bmlp7-I(Cd) and Bmlp7-Pt structures and four modified residues are present in Bmlp7-II (Figs. 2*c* and 2*d*). This observation can be treated with confidence because of the high resolution of the diffraction data and the superb quality of the electron-density maps. One positive peak in difference electron-density maps ($F_o - F_c$, 10.0σ) was detected very close to the C7 atom of the tryptophan ring. The most probable modification is an oxidation of C7. The refinement statistics were improved by building a model with the Trp180 residue modified to 7-hydroxytryptophan. The hydroxyl group of this residue acts as a donor in a hydrogen bond formed to the main-chain O atom of Leu198. The presence of 7-hydroxytryptophan residues in protein structures has recently been reported by Jensen *et al.* (2010) and Shima *et al.* (2011). Tryptophan oxidation can occur in the silkworm haemolymph because of the presence of oxidases, which are usually inhibited by the addition of 1-phenylthiourea during the purification procedure (Kramer *et al.*, 1976; Kurata *et al.*, 1994).

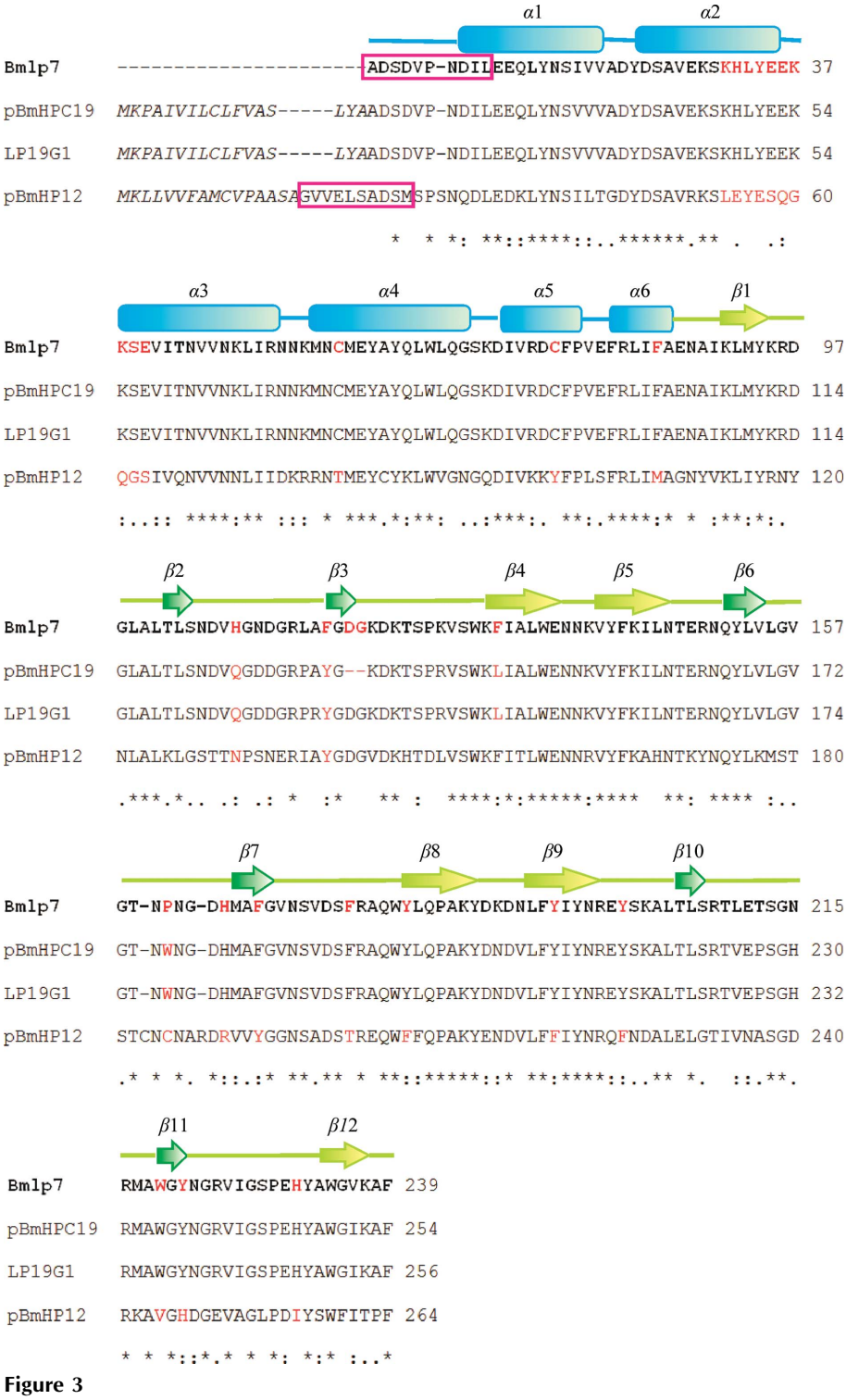

Figure 3
Sequence alignment of silkworm haemolymph 30 kDa lipoproteins: PBMHP-12, PBMHPC-19, lipoprotein 19G1 (LP19G1) and Bmlp7 (highlighted in bold). The N-terminal signal peptides are shown in italics. The main differences which helped to identify the correct sequence are highlighted in red. The N-terminal sequences compatible with the N-terminal sequencing results are shown in magenta boxes. The sequence alignment was calculated in *ClustalW* (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Pairwise sequence similarities with Bmlp7 in this alignment are as follows: 66% to PBMHP-12, 97% to PBMHPC-19 and 98% to LP19G1.

The overall shape and fold of the Bmlp7 models are almost identical to those of the structure of recombinant Bmlp7 determined at 1.9 Å resolution that was recently deposited in the PDB (PDB entry 3pub; Yang *et al.*, 2011). The 3pub structure is isomorphous with our crystal form Bmlp7-I(Cd). The recombinant protein consists of residues 46–284 (corresponding to residues 1–239 of the natural-source Bmlp7) and a C-terminal hexahistidine tag (–AAALEHHHHHH). In the recombinant Bmlp7 structure only residues 50–284 were modelled (Yang *et al.*, 2011) and no metal cations were reported to be bound to the protein molecules. The modification of Trp180 was not detected in the recombinant protein structure. The r.m.s.d. value for superposition of the C^{α} atoms of Bmlp7-I(Cd) with those of the recombinant Bmlp7 model was 0.220 Å. Our high-resolution data allowed a more detailed insight into the Bmlp7 structure. Two or more rotamers are observed for a number of residues with polar side chains, such as Asp, Glu, Lys and Arg. The protein model is also supplemented with additional water

Table 3
Cadmium coordination spheres in Bmlp7-I(Cd) and Bmlp7-Cd.

The table presents the distances between the cadmium cation and its coordinating ligands (Å).

	Bmlp7-I(Cd)		Bmlp7-Cd	
	Chain *A*	Chain *B*	Chain *A*	Chain *B*
Glu40 OE1	2.49	2.49	2.51	2.51
Glu40 OE2	2.37	2.40	2.33	2.35
Glu199† OE1	2.35	2.35	2.40	2.41
Glu199† OE2	2.49	2.43	2.48	2.49
Lys37 O	2.42	2.41	2.40	2.40
SCN^- N	2.30	2.29	2.31	2.32
H_2O O	2.34	2.50	2.52	2.54

† From a symmetric molecule.

molecules. Comparing the arrangement and interaction patterns of the protein molecules in the unit cells of the two crystal forms, one can observe that the contacts are almost identical. The conformations of amino-acid side chains are different in some cases. The Bmlp7-II structure also provides new information. It has very clear electron density for the N-terminal residue Asp4 of chain *C*, which was missing in all other protein-chain models of Bmlp7.

A search for structural homologues of Bmlp7 using *DALI* (Holm & Rosenström, 2010; http://ekhidna.biocenter.helsinki.fi/dali_server/) has already been published (Yang *et al.*, 2011). We carried out an additional analysis using the *PDBeFold* (*SSM*) server (http://www.ebi.ac.uk/msd-srv/ssm/) separately for the N-terminal (NTD) and C-terminal (CTD) domains of Bmlp7. All the top-ranked structures retrieved in the CTD comparison were proteins containing β-trefoil domains. The most similar domain was found in the mosquitocidal holotoxin (MTX_{holo}) from *Bacillus sphaericus* (PDB entry 2vsa; Treiber *et al.*, 2008), as reported previously (Yang *et al.*, 2011). Other top-ranked proteins were also the same as in the case of the search with the *DALI* server. Structural comparisons for the NTD carried out with the *PDBeFold* (*SSM*) server provided more interesting results that differed from those obtained with the *DALI* server. The top-ranked protein was the C-terminal MA3 domain of tumour suppressor Pdcd4 (LaRonde-LeBlanc *et al.*, 2007). Form 3 of the Pdcd4 MA3 domain (PDB entry 2ios) has the fold that is most similar to the NTD of Bmlp7, with an r.m.s.d. of 2.67 Å over 72 C^α atoms. The MA3 domain of Pdcd4 is a seven-helical assembly and the first four helices align with helices α1–α4 of Bmlp7 when the domains are superposed (Fig. 4). Pdcd4 is involved in transcription inhibition and interacts with transcription-initiating factors (Yang *et al.*, 2001). According to its structural similarity with Pdcd4, Bmlp7 might be able to interact with other proteins and this could be connected to its role in inhibition of apoptosis.

In order to find potential sugar ligands for Bmlp7, the crystals were soaked in 100 m*M* glucose or α-glucosamine. Data were collected and the structures were solved by molecular replacement, but no sugar molecules were present in the structures (data not shown). Thus, further studies aimed at ligand and protein-partner identification are needed.

3.2. Metal-binding sites in the Bmlp7 structures

Difference electron-density maps for Bmlp7-I(Cd) at 1.33 Å resolution revealed the presence of two very high peaks ($F_o - F_c$, 75.0σ), one in each protein molecule, that obviously corresponded to metal cations. These two cations are bound at the protein surface, with seven ligands forming a distorted pentagonal bipyramid. The pentagonal base is created by the side chains of Glu40*A* and Glu199*B* from a symmetry-related monomer and the main-chain carbonyl of Lys37*A* (Fig. 5*a*). The axial ligands are Wat400 and the N atom of SCN^-. An analogous coordination site is located in monomer *B*. The ligands creating the pentagonal base have distances of between 2.35 and 2.49 Å and the axial ligands are 2.29–2.50 Å from the central atom. Several different heavy-metal cations were tentatively placed at this binding site and analyzed using the *CheckMyMetal* server (http://csgid.org/csgid/metal_sites) to calculate the bond-valence test according to Müller *et al.* (2003). The best fitting heavy-metal cation was cadmium.

At the time of X-ray data collection, we were working with the assumption that the crystals of native Bmlp7 [later annotated Bmlp7-I(Cd)] were free of heavy-metal cations. Thus, no X-ray fluorescence scan was performed for these crystals. In any case, it would not have been possible to detect cadmium on the beamlines of the BESSY synchrotron in Berlin or the DORIS synchrotron in Hamburg because the absorption edges of cadmium are located at wavelengths of 0.4642 and 3.0857 Å (Kissel & Pratt, 1990). The purification buffers and

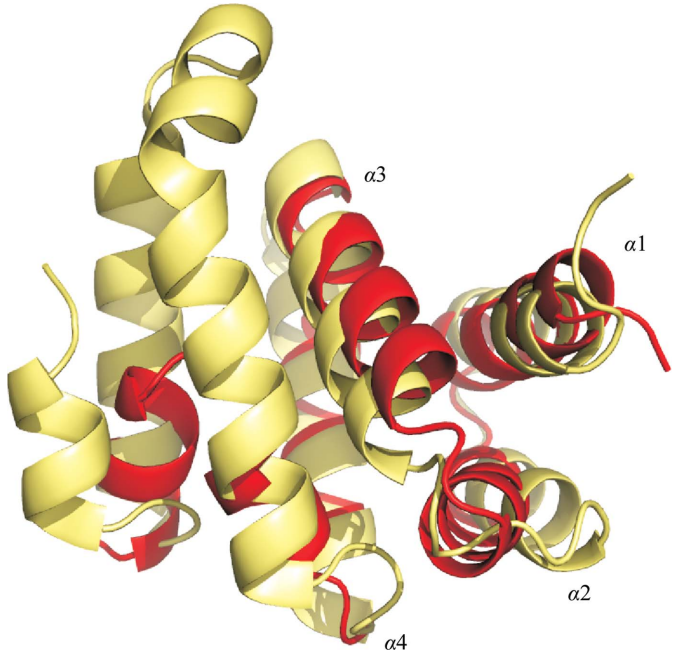

Figure 4
Structural superposition carried out with the *PDBeFold* (*SSM*) server (http://www.ebi.ac.uk/msd-srv/ssm/) shows that the fold of the Bmlp7 N-terminal domain (red) is similar to that of the C-terminal MA3 domain (yellow) of tumour suppressor Pdcd4 (PDB entry 2ios; LaRonde-LeBlanc *et al.*, 2007). MA3 of Pdcd4 is a seven-helical assembly and the first four helices superpose on helices of Bmlp7.

crystallization solutions for this particular sample did not contain any cadmium (Pietrzyk *et al.*, 2011), so the presence of these ions in the structure was very intriguing. Taking advantage of the high sensitivity of NMR for the detection of cadmium, we performed overnight measurements of the well solution and the crystallization-drop solution. There was no cadmium in the former sample and a trace of Cd in the latter, which indicated that the sole source of the cadmium ions must have been the protein sample. It thus must be concluded that the source of this heavy-metal cation was the insect material, or specifically the haemolymph collected in 2008. Interestingly, the Bmlp7-II crystals obtained using Bmlp7 isolated from another batch of larvae (haemolymph collected in 2010) did not contain any cadmium.

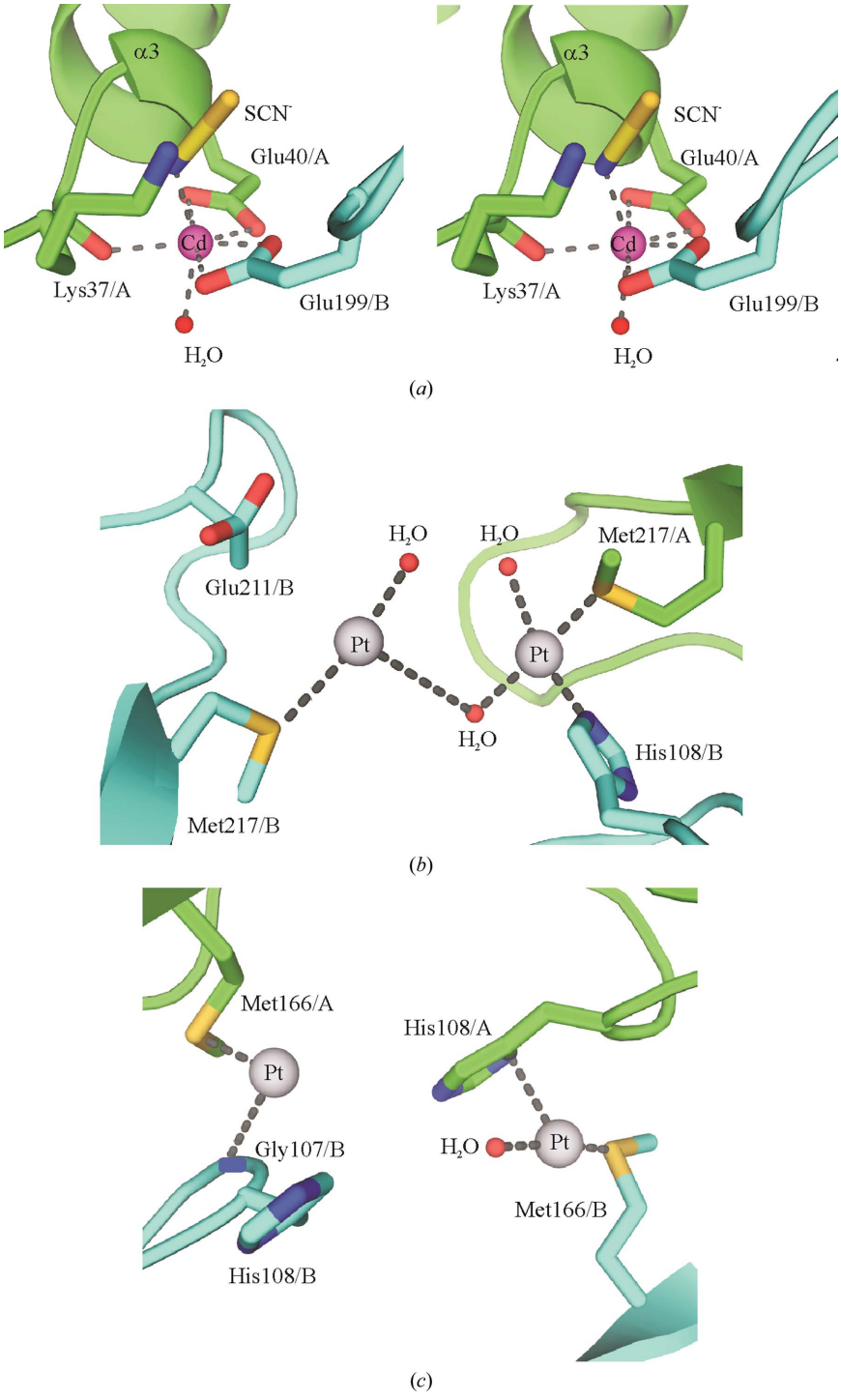

Figure 5
Metal-binding sites. (*a*) In the cadmium complex (stereoview), the pentagonal base of the coordination bipyramid is created by the main-chain carbonyl O atom of Lys37*A* and the side chains of Glu40*A* and Glu199*B*. The red ball represents a water molecule. The cadmium-binding site involves a small loop between helices $\alpha2$ and $\alpha3$ and a loop between strands $\beta11$ and $\beta12$ (see Fig. 1). (*b*, *c*) Four extra binding sites for platinum. The coordination spheres of platinum are created by methionine (Met166*A*/*B*, Met217*A*/*B*) and histidine (His108*A*/*B*) residues. Several water molecules participate in platinum coordination. These four platinum-binding sites are located in the loop between $\beta2$ and $\beta3$, in strand $\beta7$ and in the loop between $\beta10$ and $\beta11$ (see Fig. 1).

In order to confirm that the metal present in the Bmlp7-I(Cd) structure was indeed cadmium, the crystallization of cadmium-free Bmlp7 protein was repeated with the addition of a small amount (0.2 m*M*) of $CdCl_2$. The crystals grown in this experiment diffracted X-rays to 1.79 Å resolution. Two cadmium cations were present in this Bmlp7-Cd structure solved using molecular replacement (data not shown). After careful comparison of the metal-coordination geometry (Table 3), it can be unequivocally concluded that the metal present in the Bmlp7-I(Cd) structure is indeed cadmium. The two structures, Bmlp7-I(Cd) and Bmlp7-Cd, are of course isomorphous.

As stated, we propose that the cadmium cations in the Bmlp7-I(Cd) structure originated from the haemolymph. In this context, it is very interesting to note that the mulberry silkworm has been shown to bioaccumulate cadmium present in the mulberry leaves on which it feeds (Prince *et al.*, 2001; Wang *et al.*, 2004). In particular, fifth-instar larvae are able to ingest leaves with very high cadmium content (Wang *et al.*, 2004). We isolated haemolymph from silkworm larvae grown at the Institute of Natural Fibres and Medicinal Plants, Poznan, Poland. The mulberry trees that are the food source for the larvae grow in the grounds of the Institute. In 2008, heavy road construction took place nearby. We speculate that the road work resulted in micro-scale pollution that was the source of the cadmium found in the silkworm haemolymph isolated in 2008. It can be hypothesized that the Bmlp7 protein plays a decontaminating role or has another (unknown) physiological role that is sensitive to heavy-metal pollutants in the

environment. The unexpected presence of cadmium in the structure made structure solution more difficult and frustrated the fully automatic structure-determination methods.

The final model of Bmlp7-Pt is highly similar to the Bmlp7-I(Cd) model: there are also two protein molecules in the asymmetric unit with the same packing interactions. The structure of the Bmlp7-Pt crystal which was prepared for derivative data collection by soaking in K_2PtCl_6 contained six Pt atoms.

In the Bmlp7-Pt crystal obtained by prolonged soaking in a 2 m*M* concentration of platinum, six platinum-binding sites are present. Two of those six sites are almost the same as in the case of the cadmium complex; only the coordination distances between the ligands are slightly different, which suggests that Glu40 and Glu199 could bind different heavy atoms. In the Hg derivative heavy-metal cations were also present at these sites. The four other sites are typical for platinum binding (Figs. 5*b* and 5*c*) because of the presence of methionine (Met166*A*/*B* and Met217*A*/*B*) and histidine residues (His108*A*/*B*) in the coordination spheres.

3.3. Crystal packing and protein–protein interactions

The asymmetric unit of Bmlp7-I(Cd) is comprised of two protein molecules denoted *A* and *B* (Figs. 6*a* and 6*b*). The r.m.s.d. on superposition of their C^{α} atoms is 0.298 Å. The two protein molecules are related by a noncrystallographic twofold axis roughly parallel to the [110] face diagonal of the *P*1 unit cell choice. The interaction interface between protein molecules *A* and *B* is predominantly hydrophilic. The interactions are not fully symmetric because the interacting side chains of the two molecules are not the same. The total contact surface area calculated with the *PISA* server (Krissinel & Henrick, 2007) is 1610 $Å^2$. The space between the protein chains is filled by a number of water molecules. According to the *PISA* analysis, there is no strong indication that molecules *A* and *B* could form a complex in solution. This prediction agrees with the results of our experiments. In solution the protein exists as a monomer, as confirmed by dynamic light scattering. Also, during gel filtration the protein elutes as a peak with a molecular weight of about 30 kDa. We speculate that the cadmium cations bound to the protein during purification, but that the coordination sphere was slightly different at this stage. At a lower protein concentration, cadmium could be bound to only one protein molecule with two additional coordination sites (provided in the crystal structure by Glu199) filled by water molecules.

The structure of Bmlp7-II differs from those of Bmlp7-I(Cd) and Bmlp7-Pt. No heavy-metal cations are bound to the protein. The arrangement of four (instead of two) protein molecules in the asymmetric unit (Fig. 6*c*) is different from that in Bmlp7-I(Cd). Residues 105–110, 158–161 and 208–212 involved in the interactions between the C-terminal domains of chains *C* and *D* could be traced in all three structures, but all other interactions in the Bmlp7-II structure are totally different. An example is Lys70*A*, which interacts with Glu87*B* in Bmlp7-II, while in Bmlp7-I(Cd) and Bmlp7-Pt the same residue forms a contact with the carbonyl group of Gln148*B*. Glu40*A*/*B* is involved in heavy-metal binding in Bmlp7-I(Cd)

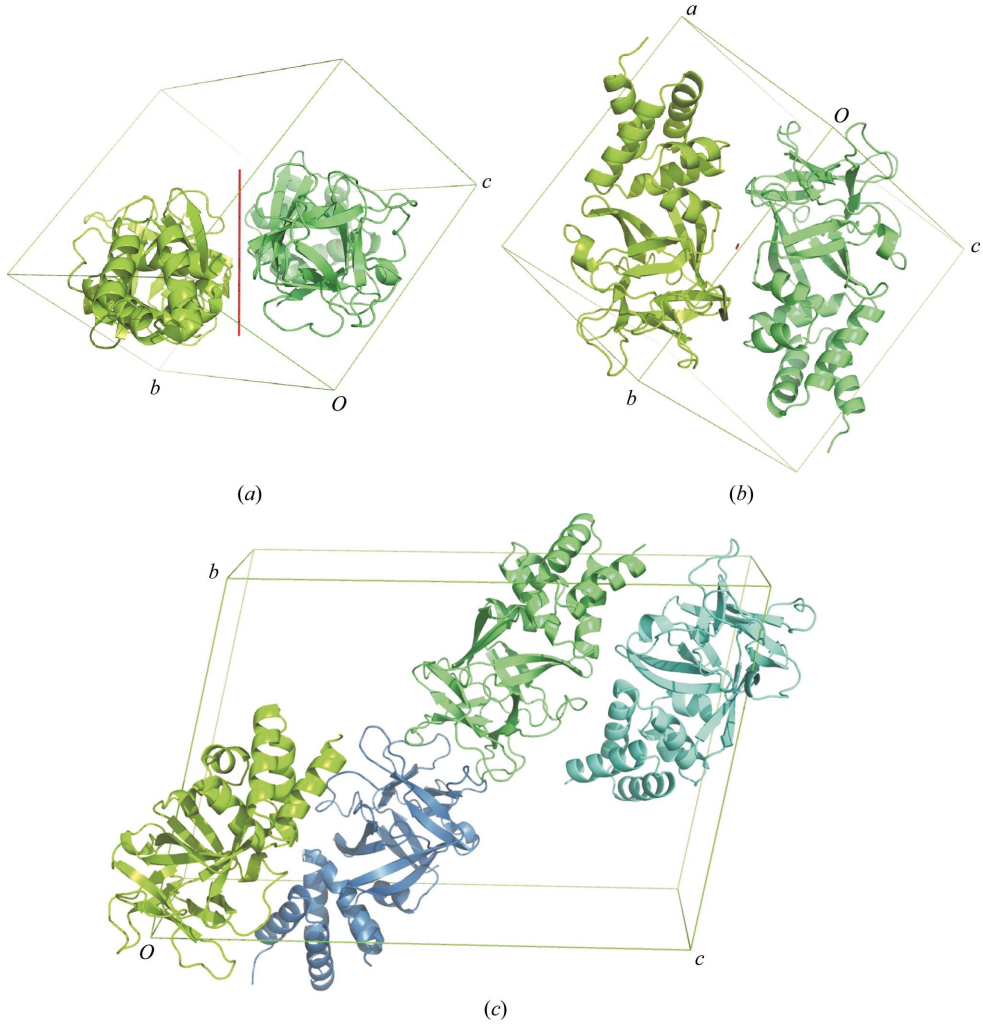

Figure 6
Unit-cell contents of Bmlp7 crystal forms. (*a*) The noncrystallographic twofold axis is shown in red in the unit cell of Bmlp7-I(Cd). (*b*) The unit cell of Bmlp7-I(Cd) after a 90° rotation about *x* (horizontal line). (*c*) The unit cell of Bmlp7-II presents a different interaction pattern between the four protein molecules.

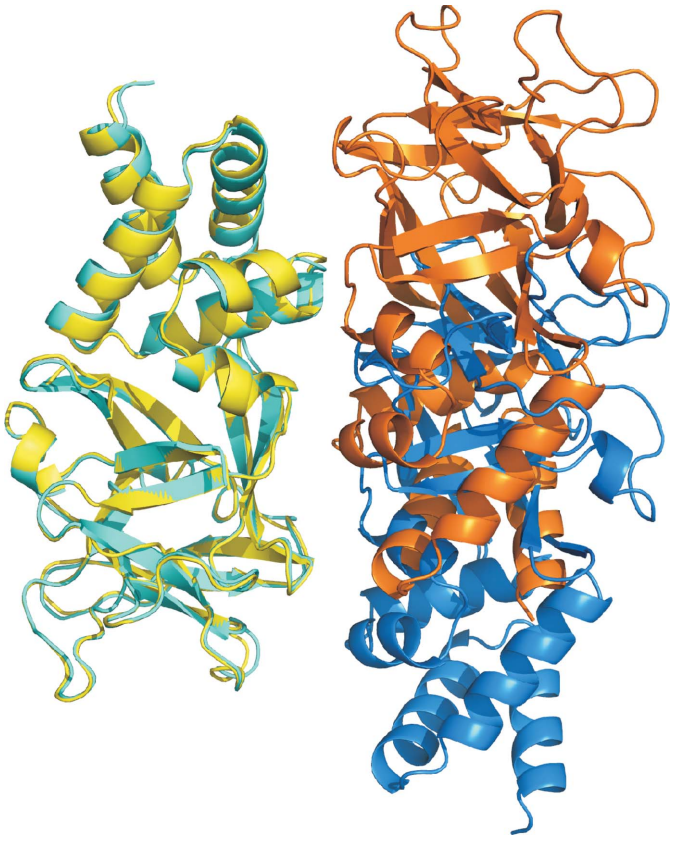

Figure 7
Superposition of the Bmlp7-I(Cd) structure (chain *A* in cyan, chain *B* in blue) with chain *A* (yellow) and *B* (orange) of Bmlp7-II shows that the protein dimer is formed in a different manner in the two crystal forms.

and Bmlp7-Pt. In Bmlp7-II Glu40*A*/*B*/*C*/*D* does not form any intermolecular contacts and is located quite distant from the side chains of the other protein molecules. Among the four protein molecules (*A*, *B*, *C* and *D*) in the Bmlp7-II structure, there are two discernible 'dimers' that are somewhat similar to that present in Bmlp-I(Cd). However, the detailed interactions are different, mainly as a consequence of a mutual shift (Fig. 7) of the molecules with respect to the noncrystallographic dyad.

The SCN^- ions had a beneficial influence on the crystallization experiments (they improved the morphology of the Bmlp7 crystals) and were present in all crystallization conditions. Potassium and SCN^- ions have been detected in the crystal structures of Bmlp7-I(Cd) and Bmlp7-Pt. Thiocyanate ions are also present in Bmlp7-II. The crystallization solution that produced the best crystals contained 0.2 *M* KSCN (Pietrzyk *et al.*, 2011); therefore, the presence of these ions is not surprising.

4. Conclusions

We have described the high-resolution crystal structure of Bmlp7, a protein of unknown function belonging to the 30 kDa lipoprotein family from mulberry silkworm haemolymph, and its interactions with heavy-metal cations. The protein is comprised of two domains, each possessing a distinct structure. The N-terminal domain has a right-handed six-helical spiral fold. The C-terminal domain has a β-trefoil fold, suggesting a role of Bmlp7 in sugar binding. This role is consistent with a hypothetical involvement of Bmlp7 in the immune response of the insect to fungal invasion. Bmlp7 could also have anti-apoptotic properties, which might be related to the N-terminal domain or to its sugar-binding properties. For instance, some proteins bind to sugar molecules present on the cell surface in a process that influences apoptosis in cell cultures.

The unexpected identification of cadmium cations in the Bmlp7-I(Cd) crystal structure suggests a possible involvement of Bmlp7 in insect detoxification mechanisms related to environmental pollution.

The present case is an excellent example of how a precise crystallographic study can contribute to protein identification. The protein sequence was deduced by combining the results of amino-acid sequencing by mass spectrometry with the superb electron-density maps at 1.33 Å resolution and with the holdings of global sequence databases. Although previous sequence analyses from LC/MS/MS mass spectrometry were available, their indications were neither conclusive nor correct and final protein identification was only possible after crystal structure determination. When the resolution of the X-ray diffraction data is sufficiently high ($d_{\rm min}$ < 2.0 Å), electron-density maps allow the identification of the amino-acid side-chain sequence and can resolve sequence-related identification ambiguities.

Another interesting observation is that the transcription level of the Bmlp7 gene is obviously very high in *B. mori* L. The amount of Bmlp7 in the 30 kDa fraction of the haemolymph proteins is higher than the aggregated amounts of the remaining proteins from this group. This observation suggests that Bmlp7 is a very important component of silkworm haemolymph.

We would like to thank Professor M. Potrzebowski from the Centre of Molecular and Macromolecular Studies, Polish Academy of Sciences, Lodz, Poland for NMR measurements. 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.

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.
Choi, S. S., Rhee, W. J. & Park, T. H. (2002). *Biotechnol. Prog.* **18**, 874–878.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Gamo, T. (1978). *Insect Biochem.* **8**, 457–470.
Gupta, A. P. (1991). Editor. *Immunology of Insects and Other Arthropods*, pp. 19–118. London: CRC Press.
Hao, Q. (2004). *J. Appl. Cryst.* **37**, 498–499.
Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.

Jensen, L. M., Sanishvili, R., Davidson, V. L. & Wilmot, C. M. (2010). *Science*, **327**, 1392–1394.
Jiang, H., Ma, C., Lu, Z.-Q. & Kanost, M. R. (2004). *Insect Biochem. Biol. Mol.* **34**, 89–100.
Kabsch, W. (2010*a*). *Acta Cryst.* D**66**, 125–132.
Kabsch, W. (2010*b*). *Acta Cryst.* D**66**, 133–144.
Kim, E. J., Rhee, W. J. & Park, T. H. (2001). *Biochem. Biophys. Res. Commun.* **285**, 224–228.
Kim, E. J., Rhee, W. J. & Park, T. H. (2004). *Biotechnol. Prog.* **20**, 324–329.
Kissel, L. & Pratt, R. H. (1990). *Acta Cryst.* A**46**, 170–175.
Klimpel, K. R. & Goldman, W. E. (1988). *Infect. Immun.* **56**, 2997–3000.
Kramer, K. J., Dunn, P. E., Peterson, R. C., Seballos, H. L., Sanburg, L. L. & Law, J. H. (1976). *J. Biol. Chem.* **251**, 4979–4985.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Kurata, K., Nakamura, M., Okuda, T., Hirano, H. & Shinbo, H. (1994). *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **109**, 105–114.
LaRonde-LeBlanc, N., Santhanam, A. N., Baker, A. R., Wlodawer, A. & Colburn, N. H. (2007). *Mol. Cell. Biol.* **27**, 147–156.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Mori, S., Izumi, S. & Tomino, S. (1991). *Biochim. Biophys. Acta*, **1090**, 129–132.
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.
Müller, P., Köpke, S. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 32–37.
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.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Painter, J. & Merritt, E. A. (2006). *Acta Cryst.* D**62**, 439–450.
Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2005). *Acta Cryst.* D**61**, 449–457.
Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2009). *Acta Cryst.* D**65**, 1089–1097.
Pannu, N. S., McCoy, A. J. & Read, R. J. (2003). *Acta Cryst.* D**59**, 1801–1808.
Perrakis, A., Harkiolaki, M., Wilson, K. S. & Lamzin, V. S. (2001). *Acta Cryst.* D**57**, 1445–1450.
Pietrzyk, A. J., Bujacz, A., Łochyńska, M., Jaskólski, M. & Bujacz, G. (2011). *Acta Cryst.* F**67**, 372–376.
Prince, S. P., Senthilkumar, P. & Subburam, V. (2001). *Environ. Monit. Assess.* **69**, 231–238.
Rhee, W. J. & Park, T. H. (2000). *Biochem. Biophys. Res. Commun.* **271**, 186–190.
Sakai, N., Mori, S., Izumi, S., Haino-Fukushima, K., Ogura, T., Maekawa, H. & Tomino, S. (1988). *Biochim. Biophys. Acta*, **949**, 224–232.
Schägger, H. & von Jagow, G. (1987). *Anal. Biochem.* **166**, 368–379.
Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 5857–5864.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Shepherd, M. G. (1987). *Crit. Rev. Microbiol.* **15**, 7–25.
Shima, S., Krueger, M., Weinert, T., Demmer, U., Kahnt, J., Thauer, R. K. & Ermler, U. (2011). *Nature (London)*, **481**, 98–101.
Sun, Q., Zhao, P., Lin, Y., Hou, Y., Xia, Q.-Y. & Xiang, Z.-H. (2007). *Insect Sci.* **14**, 5–14.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Treiber, N., Reinert, D. J., Carpusca, I., Aktories, K. & Schulz, G. E. (2008). *J. Mol. Biol.* **381**, 150–159.
Ujita, M., Katsuno, Y., Kawachi, I., Ueno, Y., Banno, Y., Fujii, H. & Hara, A. (2005). *Biosci. Biotechnol. Biochem.* **69**, 1178–1185.
Ujita, M., Kimura, A., Nishino, D., Yokoyama, E., Banno, Y., Fujii, H. & Hara, A. (2002). *Biosci. Biotechnol. Biochem.* **66**, 2264–2266.
Wang, K. R., Gong, H., Wang, Y. & van der Zee, S. E. A. T. M. (2004). *Plant Soil*, **261**, 171–180.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Yang, H.-S., Jansen, A. P., Nair, R., Shibahara, K., Verma, A. K., Cmarik, J. L. & Colburn, N. H. (2001). *Oncogene*, **20**, 669–676.
Yang, J.-P., Ma, X.-X., He, Y.-X., Li, W.-F., Kang, Y., Bao, R., Chen, Y. & Zhou, C.-Z. (2011). *J. Struct. Biol.* **175**, 97–103.