



Crystal Structure of *Bombyx mori* Lipoprotein 6: Comparative Structural Analysis of the 30-kDa Lipoprotein Family

Agnieszka J. Pietrzyk¹, Anna Bujacz², Malgorzata Łochynska³, Mariusz Jaskolski^{1,4}, Grzegorz Bujacz^{1,2*}

1 Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland, **2** Institute of Technical Biochemistry, Faculty of Biotechnology and Food Sciences, Technical University of Lodz, Lodz, Poland, **3** Institute of Natural Fibers and Medicinal Plants, Poznan, Poland, **4** Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Poznan, Poland

Abstract

The 30-kDa lipoprotein (LP) family of mulberry silkworm comprises major hemolymph proteins specific to the fifth instar larvae. The family consists of 46 members, 24 of which are referred to as typical 30-kDa LPs. To date, two crystal structures of 30-kDa LPs from *Bombyx mori* have been described (Bmlp3 and Bmlp7). Here, we present the crystal structure of Bmlp6, another 30-kDa LP member. Bmlp6 is comprised of two domains characteristic of this family, the VHS-type N-terminal domain and β -trefoil C-terminal domain. The structures of the three 30-kDa LPs have been compared and a number of differences are noted, including loop conformation, the surface electrostatic potential, and the potential binding cavities. We discuss the observed structural differences in the light of the potential different roles of the particular 30-kDa LP members in silkworm physiology.

Citation: Pietrzyk AJ, Bujacz A, Łochynska M, Jaskolski M, Bujacz G (2014) Crystal Structure of *Bombyx mori* Lipoprotein 6: Comparative Structural Analysis of the 30-kDa Lipoprotein Family. PLoS ONE 9(11): e108761. doi:10.1371/journal.pone.0108761

Editor: Alexander Wlodawer, NCI-Frederick, United States of America

Received: July 19, 2014; **Accepted:** September 4, 2014; **Published:** November 7, 2014

Copyright: © 2014 Pietrzyk et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. The crystal structure files are available from the PDB database accession number 4PC4.

Funding: 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 use of the BESSY and the DESY beamlines received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under BioStruct-X (grant agreement No. 283570). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: grzegorz.bujacz@p.lodz.pl

Introduction

The mulberry silkworm, *Bombyx mori*, is an insect of high economical importance due to its ability to produce natural silk fibers. It was the first insect within the order *Lepidoptera* for which draft genomic sequence was completed in 2004, independently by a Chinese [1] and Japanese [2] groups. In 2008, The International Silkworm Genome Consortium was formed by these groups and all existing data were merged and supplemented with newly obtained sequences [3]. The genome information is accessible via the knowledgebase for silkworm biology and genomics, SilkDB [4,5]. Despite these efforts, the information about silkworm genes and proteins is still not well ordered in international databases. For example, there are significant differences in the annotations and even in the sequences of particular proteins. This is especially true of 30-kDa lipoproteins (30-kDa LPs), the most abundant group of proteins found in the hemolymph of the silkworm fifth instar larvae [6–9]. The first cDNA clones of 30-kDa LPs with distinct mRNA were isolated by Sakai et al. [8] and annotated as PBMHP-6, PBMHP-12, PBMHPC-19, PBMHPC-21 and PBMHPC 23 or LP(1–5), respectively [8]. Later, according to the draft of silkworm genomic sequence [1,2], ten genes of 30-kDa LPs were predicted and denoted as Bmlp1–10 [10]. A few years later, a more detailed analysis based on the completed silkworm genomic sequence revealed the presence of 46 genes of 30-kDa LPs in the silkworm

genome [11]. They were annotated as Bmlp1–46 according to the previous terminology rules [10]. The 30-kDa LPs can be divided into three subfamilies: typical 30-kDa LPs (Bmlp1–24), serine/threonine-rich LPs (Bmlp25–36) and ENF peptide-binding proteins (Bmlp37–46) [11].

An analysis of expressed sequence tags (ESTs) indicated that the level of 30-kDa LPs expression is variable for the different genes. The highest expression level was observed for *Bmlp7*, *Bmlp3*, *Bmlp2*, *Bmlp1* and *Bmlp9*, listed in the decreasing order of the number of ESTs [10]. Moreover, proteomic studies of silkworm hemolymph [12–15] and of the fat bodies [14], the site of 30-kDa LP synthesis, revealed that the expression of particular genes varies from the end of the fourth to the end of the fifth instar. Although 30-kDa LPs are usually described as proteins specific to the fifth instar larvae [7], some members of this family, e.g. Bmlp20, are expressed only at the end of the fourth instar [15].

The 30-kDa LPs are involved in many physiological processes in the silkworm body. They transport lipids in the hemolymph [16], tryptophan metabolites and pigments in diapause silkworm eggs [17], are involved in an immune response pathway in an antifungal defense system [18,19], serve as the source of nutrients during pupation and adult development [13,20], and are the second major yolk protein component [21–23].

The 30-kDa LPs can be also very useful laboratory tools due to their anti-apoptotic properties. An addition of silkworm hemo-

lymph to different cell cultures inhibits apoptosis and improves the viability of the cells. The anti-apoptotic properties of the hemolymph were demonstrated for a number of cell cultures, including insect Sf9 [24–27], mammalian CHO [28–29], and human HeLa [30] and HEK293 [29] cell lines. 30-kDa LP members were identified as the hemolymph component responsible for those properties [29,31]. Finally, it was shown that a 30-kDa LP protein is able to penetrate various cell types and can also deliver a cargo protein [32].

Here we present the crystal structure of *Bombyx mori* lipoprotein 6 (Bmlp6). To-date, only two 30-kDa LPs were characterized structurally, Bmlp7 [33,34] and Bmlp3 [35]. In this paper we also present a detailed comparative structural analysis of the three 30-kDa LPs, and discuss how the structural differences could be

related to diverse physiological roles of the members of this large protein family.

Materials and Methods

Hemolymph collection, protein purification and crystallization

B. mori hemolymph was collected from fifth instar larvae as described previously [36]. Bmlp6 was isolated from the hemolymph using a three-step purification protocol. The first two steps have been already described [36]. Briefly, initial protein separation was carried out using size exclusion chromatography with a Superdex 200 prep grade column (XK 16/100, Amersham Biosciences). The collected fractions containing the 30-kDa

Table 1. Diffraction data collection and refinement statistics.

Crystal data	Bmlp6
Space group	$P2_1$
Unit cell parameters a, b, c [Å]; α , β , γ [°]	83.2, 85.7, 104.5; 90.0, 104.8, 90.0
Protein molecules/ASU	5
VM [Å ³ /Da]	2.62
Solvent content [%]	53.1
X-ray data collection	
Temperature [K]	100
Radiation source	PETRA III, P14 (MX2)
Wavelength [Å]	0.976
Detector	DECTRIS PILATUS 6M
Crystal-detector distance [mm]	388
Rotation range [°]	0.2
Total rotation [°]	360
Exposure/image [s]	0.2
Resolution [Å]	47–1.8 (1.9–1.8) ^a
Mosaicity [°]	0.30
Intensities measured	803 831
Unique reflections	124 752
R_{merge}^b [%]	4.8 (55.1)
Redundancy	6.4 (5.2)
$\langle I/\sigma \rangle$	22.3 (3.5)
Completeness [%]	94.9 (73.0)
Refinement	
Resolution [Å]	47–1.8
R_{work}/R_{free}^c [%]	16.1/18.3
R_{work} set count	123 486
R_{free} test set count	1265
Protein/solvent atoms	9885/1131
R.m.s. deviations (bond lengths) [Å]	0.013
R.m.s. deviations (bond angles) [°]	1.53
$\langle B \rangle$ for protein/solvent [Å ²]	25.8/43.6
Favored/disallowed Ramachandran ϕ/ψ angles [%]	92.8/0.0
PDB code	4PC4

^aValues in parentheses are for the highest resolution shell.

^b $R_{merge} = \sum_h \sum_j |I_{hj} - \langle I_h \rangle| / \sum_h \sum_j I_{hj}$, where I_{hj} is the intensity of observation j of reflection h .

^c $R_{work} = \sum_h ||F_o| - |F_c|| / \sum_h |F_o|$ for all reflections, where F_o and F_c are observed and calculated structure factors, respectively. R_{free} is calculated analogously for the test reflections, randomly selected and excluded from the refinement.

doi:10.1371/journal.pone.0108761.t001

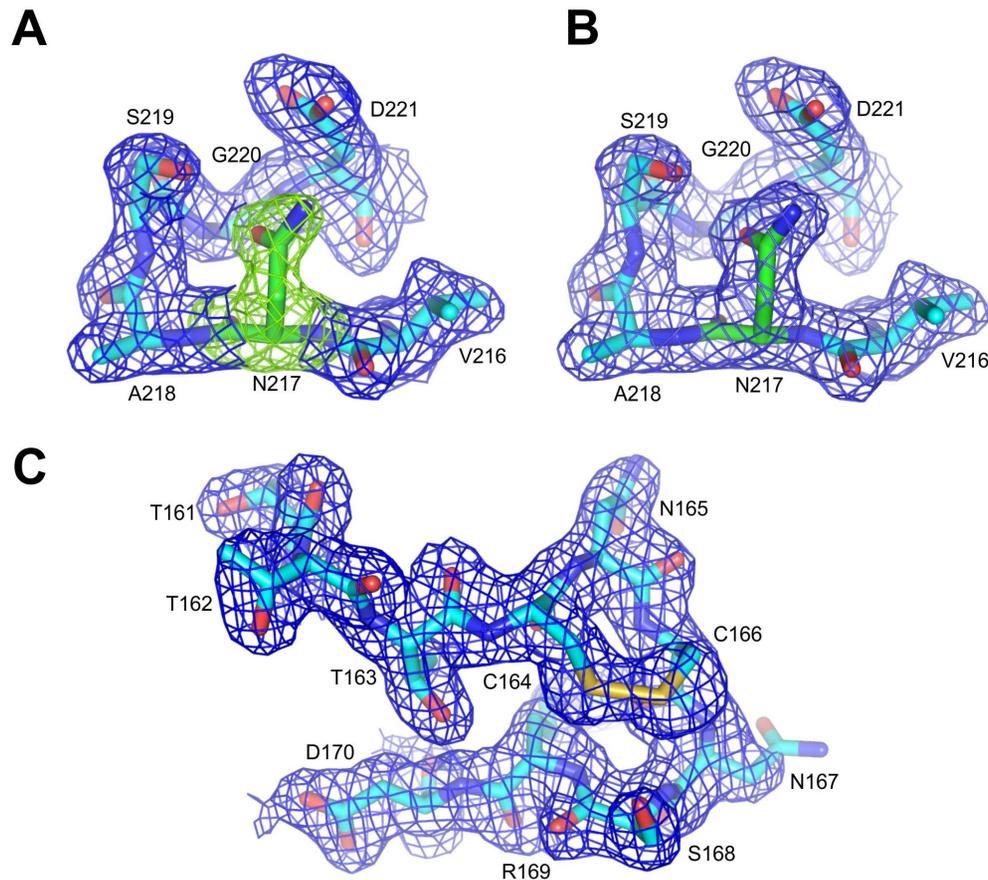


Figure 1. Electron density maps of Bmlp6. (A, B) *Fo-Fc* and *2Fo-Fc* electron density maps (contour 5.0σ and 1.0σ , respectively), unequivocally demonstrate that residue 217 of Bmlp6 (green) is not Tyr but Asn. (C) *2Fo-Fc* electron density map (contour 1.0σ) of a loop (residues 161–170) containing disulfide bridge is presented to illustrate that electron density maps were of a very good quality also at the loop regions. For both fragments the molecule A was arbitrary chosen.
doi:10.1371/journal.pone.0108761.g001

proteins were concentrated and applied on a Q Sepharose column (XK 16/10, Amersham Biosciences). Stepwise-elution ion exchange chromatography allowed further separation. The fraction containing Bmlp6 was eluted with 30 mM NaCl. The third purification step was chromatofocusing performed on a MONO P 5/200 GL (GE Healthcare Life Sciences) column equilibrated with 0.025 M Bis-Tris/CH₃COOH pH 7.0. Prior to sample application, the column was washed with ~10 ml of diluted mixture of Polybuffer 96 and Polybuffer 74 (in 19 1 ratio) pH 6.0. The same mixture was used as elution buffer. The fractions containing Bmlp6 were concentrated to ~3 mg/ml.

Crystals of Bmlp6 were grown at 20°C using the hanging drop vapor diffusion method and 10% isopropanol, 10% PEG 3350 and 0.1 M MES pH 6.0 as the precipitating buffer. The crystals were tiny plates and grew in clusters. Initial crystals were grown using the sample obtained after two-step purification. However, the third purification step was crucial to obtain the crystals of good quality and good diffraction properties.

Data collection and processing

X-ray diffraction data for initial crystals were collected at beamlines BL14.1 and BL14.2 of the BESSY synchrotron, Berlin, Germany [37]. The crystals diffracted only to 7.0 Å resolution. X-Ray diffraction data for Bmlp6 crystals, obtained after the third purification step, were collected at the EMBL beamline P14 of the

PETRA III synchrotron at DESY, Hamburg, Germany, using a PILATUS 6M detector. The diffraction images were recorded at 100 K for one single crystal using the oscillation method with a rotation of 0.2° per image. No cryoprotection was necessary due to the presence of isopropanol in the crystallization buffer. The crystal was mounted in a nylon fiber loop and vitrified in a stream of cold nitrogen gas. The diffraction images were indexed, integrated and scaled using XDS [38,39]. The data collection and processing statistics are presented in Table 1.

Structure determination and refinement

The crystal structure of Bmlp6 was determined by molecular replacement (MR) using Phaser-MR [40] and the coordinates of chain A of Bmlp7 (PDB: 4EFP) [34] as a starting model. The final model was completed after several cycles of manual rebuilding in COOT [41] and refinement in REFMAC [42] with the inclusion of TLS parameters [43]. The progress of the refinement was monitored and validated using 1265 reflections set aside for R_{free} testing [44]. The geometry of the model was assessed in PROCHECK [45]. The refinement statistics are summarized in Table 1.

N-terminal sequencing

Preparation of a Bmlp6 sample for Edman degradation included tricine-SDS gel electrophoretic separation [46] followed

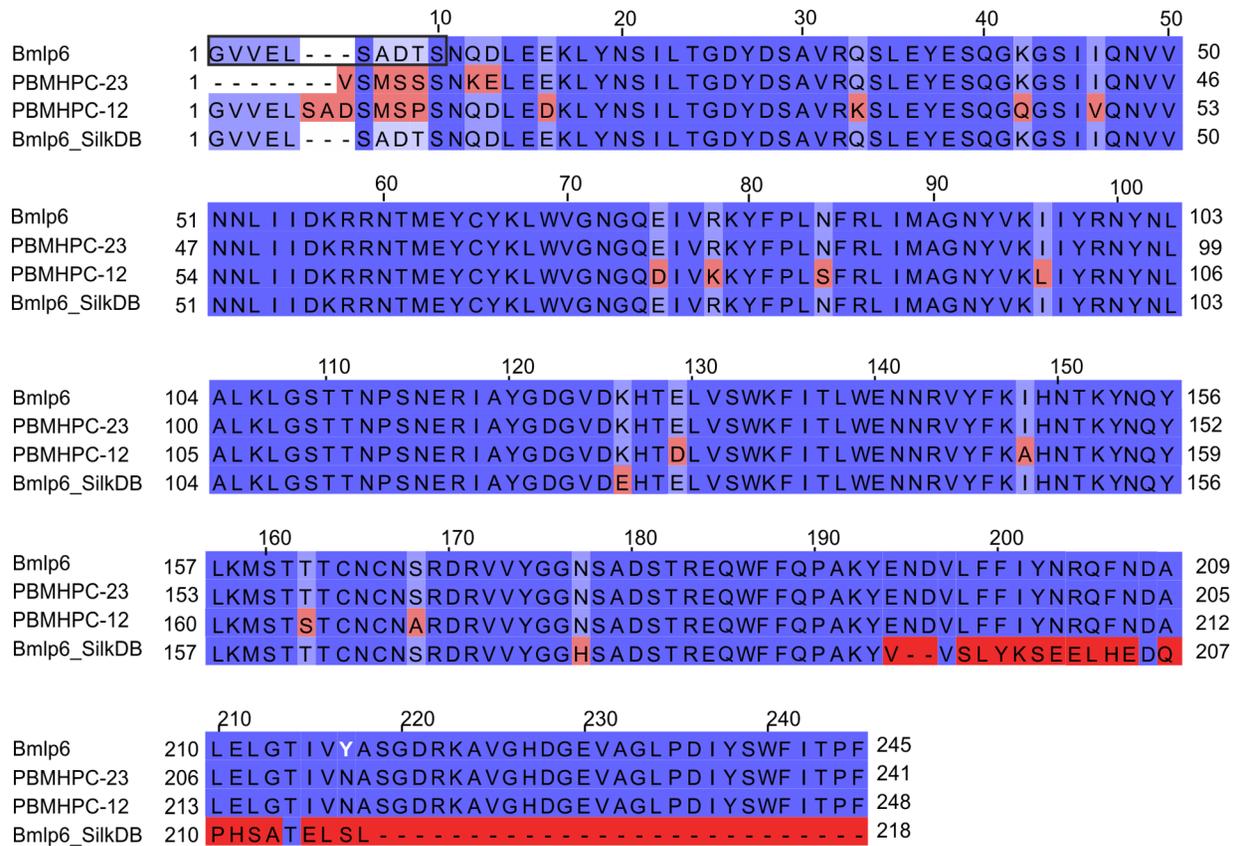


Figure 2. Alignment of sequences corresponding to Bmlp6. Amino acid sequence alignment of Bmlp6 (UniProt: A7LIK7; NCBI-Protein: NP_001095198), PBMHPC-23 (UniProt: P09338), PBMHPC-12 (UniProt: P09335) and Bmlp6_SilkDB (SilkDB: BGIBMGA004457) calculated in ClustalW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). The alignment is colored according to identity (dark blue) and similarity (light blue) using Jalview (<http://www.jalview.org/>) [60]. The evident sequencing error at the C-terminus of Bmlp6_SilkDB (highlighted in dark red, starting from position 194) has been disregarded in sequence similarity calculations. The only discrepancy between Bmlp6 (UniProt: A7LIK7) and the amino acid sequence determined by X-ray crystallography is indicated by white font, at position 217. The N-terminal sequence of Bmlp6 established by Edman degradation is boxed. The displayed sequences correspond to mature proteins without signal peptides.
doi:10.1371/journal.pone.0108761.g002

by protein transfer to a PVDF Immobilon membrane, PSQ 0.22 μm (Millipore). The single protein band corresponding to the molecular weight of 28 kDa was cut and subjected to Edman degradation cycles, performed using a fully automated Procise 491 (Applied Biosystems) sequencer.

Results and Discussion

Protein identification and database-related ambiguities

Bmlp6 was isolated from mulberry silkworm hemolymph as an unknown protein and successfully purified by FPLC using a three-step purification protocol. To facilitate identification, ten N-terminal amino acid residues were sequenced by Edman degradation. The analysis returned the following N-terminal sequence: $_1\text{GVVELSADTS}_{10}$. The final identification of the protein was complicated because of the high level of amino acid sequence similarity among the members of the 30-kDa LP family. The problem was aggravated by the confusing information about silkworm proteins in the international sequence databases. Ultimately, the X-ray crystallographic electron density maps were instrumental for the final identification of the investigated protein. This study is part of a wider project, aimed at structural analysis of the most abundant proteins in the hemolymph of silkworm fifth instar larvae. The investigated proteins are typically isolated as

unknown proteins, requiring definitive identification. Previously, we have described our results for Bmlp7 [34], Bmlp3 [35] and for a complex of two arylphorins, SP2-SP3 [47]. In all of those cases, the successful amino acid sequence identification was largely based on electron density maps. An overview of the benefits of sequencing from electron density maps has been presented by Pietrzyk et al. [48].

The crystal structure of Bmlp6 was solved at 1.8 \AA resolution by molecular replacement using Bmlp7 as a model (PDB ID: 4EFP) [34]. The initial electron density maps were of a good quality. A plausible amino acid sequence of the investigated protein was found, according to the experimental information from the N-terminal sequencing and electron density maps, in two databases, namely in UniProt under the accession code A7LIK7 and annotated as 30K lipoprotein; and in NCBI-Protein under the accession code NP_001095198 and annotated as low-molecular 30-kDa lipoprotein PBMHPC-23 precursor. The only discrepancy between the two sequences and the amino acid sequence deduced from electron density maps is at position 217, which is Asn in the crystal structure (Fig. 1), while according to the sequences in the databases it should be Tyr. A distinction between Asn and Asp at this position was based on the analysis of H-bonding network, the analysis of temperature factors and the fact that Asn residue is present at this position in a number of other 30-kDa LPs. Small

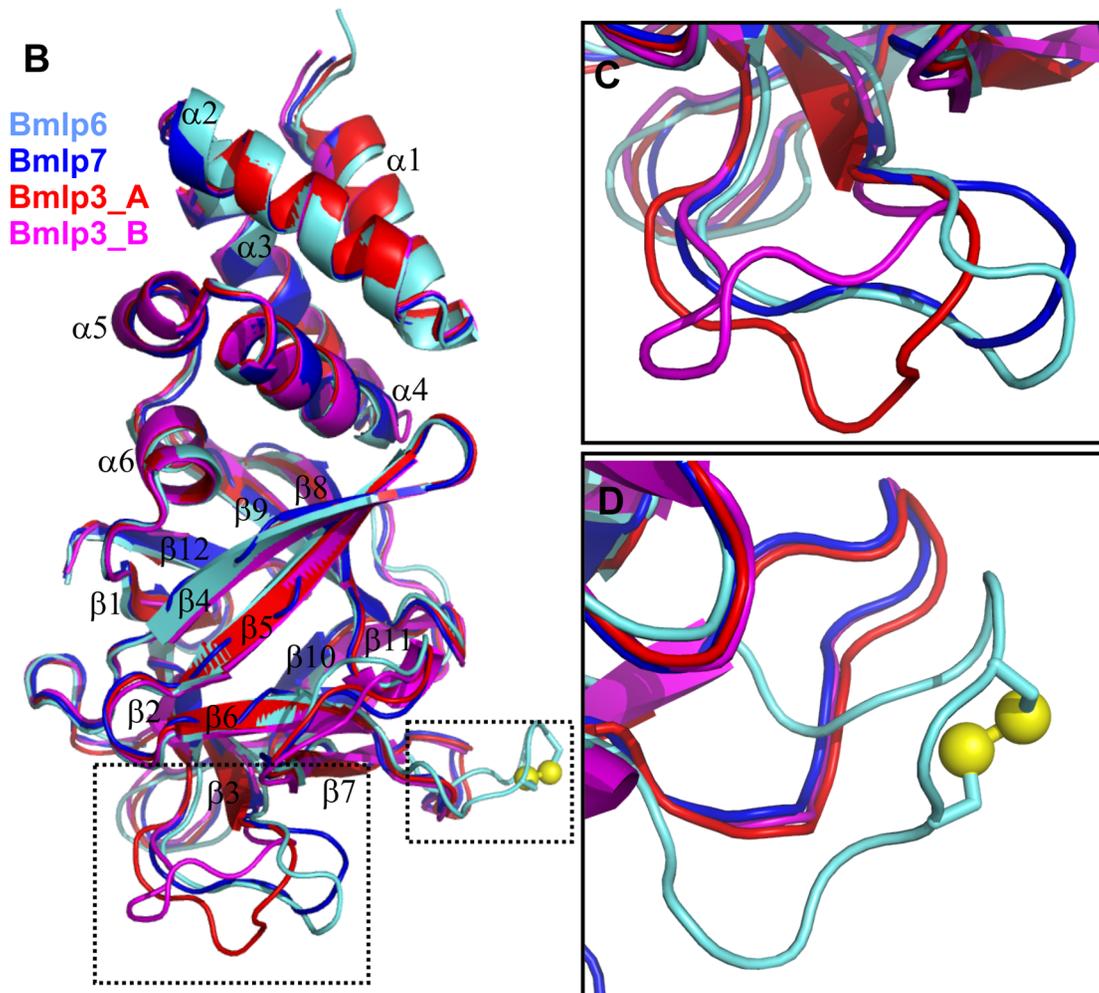
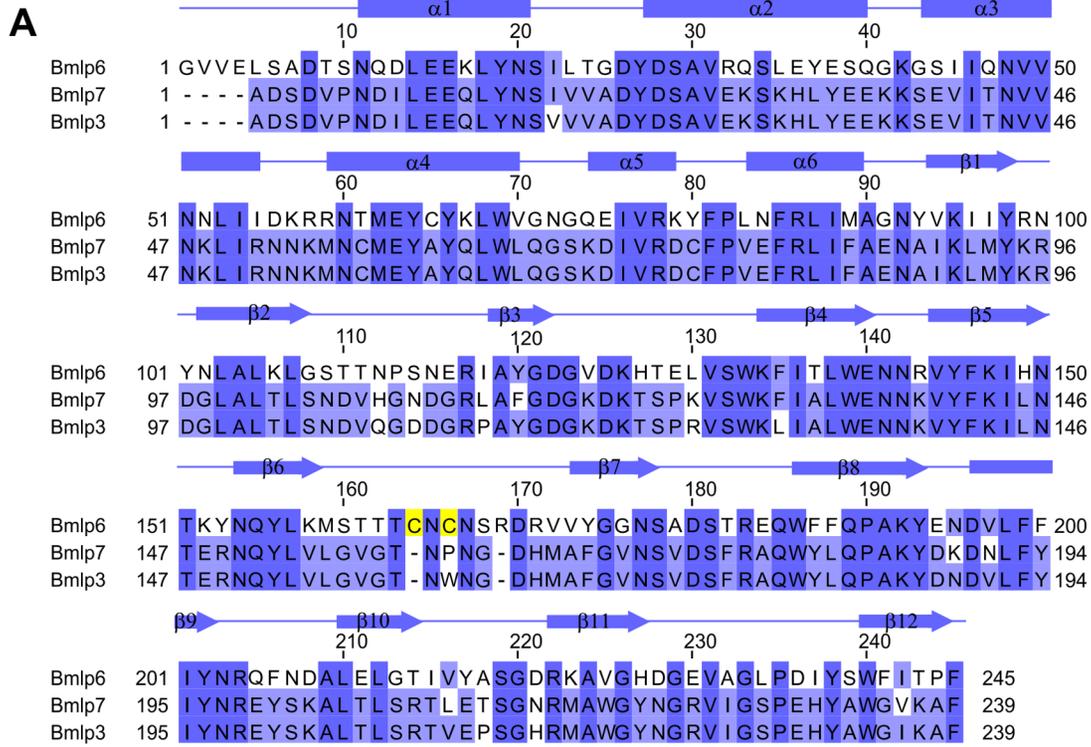


Figure 3. Structural comparison of Bmlp6, Bmlp3 and Bmlp7. (A) Amino acid sequence alignment of Bmlp6 (UniProt: A7LIK7), Bmlp7 (UniProt: E5EVW2) and Bmlp3 (UniProt: H9J4F6) calculated in ClustalW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). The alignment is coloured according to identity (dark blue) and similarity (light blue) using Jalview (<http://www.jalview.org/>) [60]. The cysteine residues forming a disulfide bridge are highlighted in yellow. The secondary structure elements are assigned as α -helices (cylinders) and β -strands (arrows). (B) C α atom superpositions of the crystallographic models of Bmlp6 (chain A, this work), Bmlp7 (PDB: 4EFP, chain A) and Bmlp3 (PDB: 4IY9, chain A and B). (C) The most flexible loop of the CTD among the three compared proteins, located between strands β 2 and β 3. (D) The Bmlp6 loop containing a disulfide bridge (yellow spheres), located between strands β 6 and β 7, is compared with the corresponding loops in Bmlp7 and Bmlp3. The color code in panels B–D is light blue for Bmlp6, dark blue for Bmlp7, and shades of red for chains A and B of Bmlp3. doi:10.1371/journal.pone.0108761.g003

sequence variations (e.g. Tyr/Asn) are not unusual when the same protein is isolated from different silkworm strains [14] and may represent genuine genetic polymorphism. In this particular case, a single nucleotide codon mutation can cause the Tyr/Asn substitution. Interestingly, the amino acid sequence of PBMHPC-23 available in UniProt (Accession code: P09338) is incomplete at the N-terminus (Fig. 2). The search for sequences corresponding to PBMHPC-23 was also performed against the SilkDB database [4,5]. A major part of PBMHPC-23 (starting at the N-terminal residue) corresponds to residues 1-193 of the Bmlp6 sequence found in the SilkDB, but the C-terminal part of the Bmlp6 sequence is incorrect and incomplete, starting at residues 194 and 218, respectively (Fig. 2). However, to be consistent with the current terminology rules introduced by Sun et al. [10], the name of the investigated protein has been assigned as Bmlp6 (despite of the errors in SilkDB) instead of PBMHPC-23, which refers to the nomenclature introduced in 1980s [8]. The highest sequence similarity to Bmlp6 is observed for Bmlp2 (UniProt PBMHP-12, accession code: P09335). The Bmlp6/Bmlp2 amino acid sequences alignment (Fig. 2) indicates 97% similar and 93% identical residues.

Our results indicate that Bmlp6 is the third most abundant 30-kDa LP protein in the hemolymph of the fifth instar larvae. The first and the second most abundant proteins are Bmlp7 and Bmlp3, respectively. The first two steps of purification protocol (size exclusion and ion exchange chromatography) were common for three mentioned 30-kDa LPs and they always provided three samples (corresponding to three different peaks on chromatograms), each containing one of the three 30-kDa LPs as the main

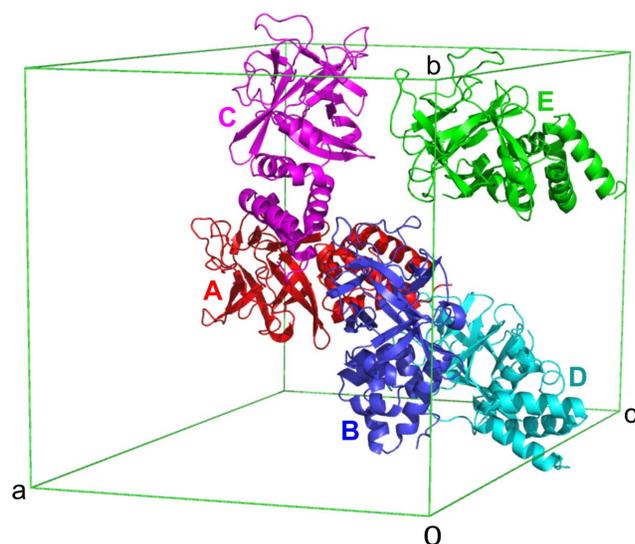


Figure 4. Crystal packing of Bmlp6. The asymmetric unit of Bmlp6 (shown with unit cell outline) is composed of five protein molecules, A–E, represented by different colors. doi:10.1371/journal.pone.0108761.g004

component. According to all chromatograms obtained using a number of hemolymph samples, the quantities of the proteins were always the highest for Bmlp7, then for Bmlp3 and Bmlp6. The first

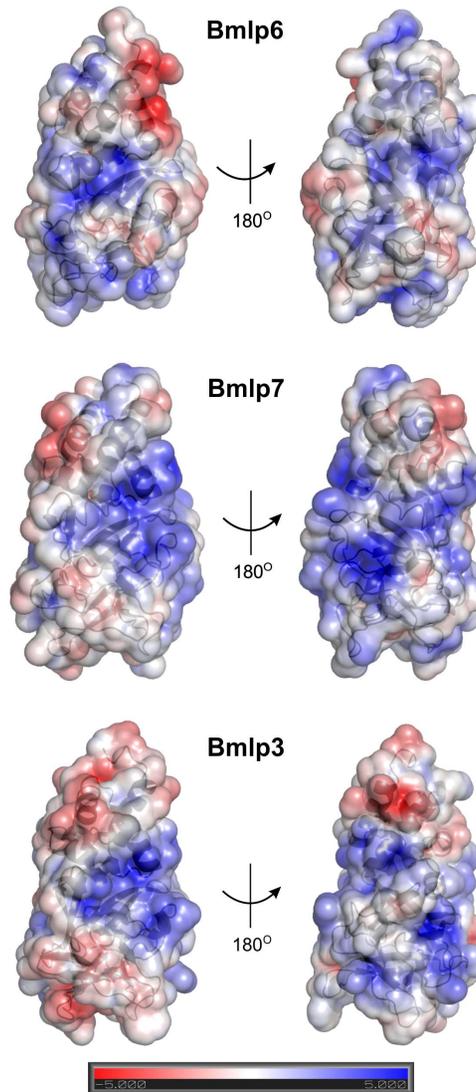


Figure 5. Electrostatic surface potential of Bmlp6, Bmlp7 and Bmlp3 at pH 6.5. Electrostatic surface potential of Bmlp6, Bmlp7 and Bmlp3 was calculated using the APBS algorithm [52] and the PDB2PQR program [53,54] at pH 6.5, which is the physiological pH of silkworm hemolymph. The protein surfaces are shown in the same orientation, in two different views for each protein. The positive and negative charges are colored blue and red, respectively, according to the scale. The APBS writes out the electrostatic potential in dimensionless units of $k_b T e_c^{-1}$ where k_b is Boltzmann's constant, T is the temperature of calculation and e_c is the charge of an electron [52]. doi:10.1371/journal.pone.0108761.g005

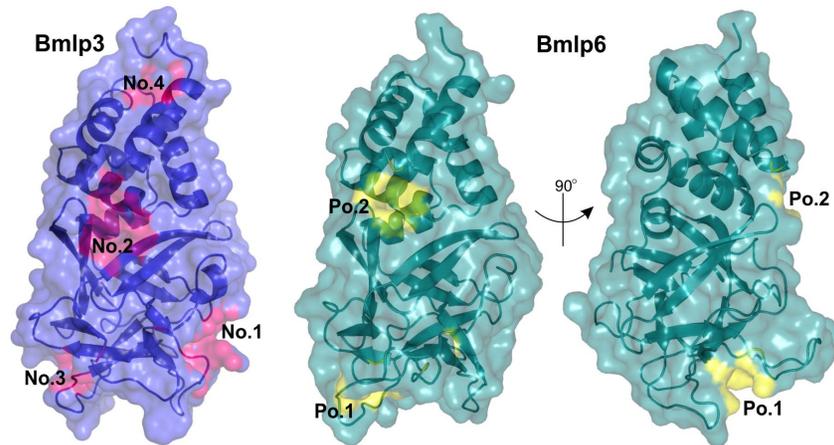


Figure 6. Potential binding sites of Bmlp3 and Bmlp6. Potential binding sites of Bmlp3 (blue/pink) and of Bmlp6 (green/yellow) were marked as No. 1–4 and Po.1–2, respectively. Two different views of Bmlp6 are shown to present both cavities.
doi:10.1371/journal.pone.0108761.g006

(Bmlp7) and the second position (Bmlp3) is in agreement with the EST analysis [10]. However, it was also reported that Bmlp2 is at the third most abundant 30-kDa LP with 413 ESTs and the number of ESTs for Bmlp6 was estimated at only 15 [10]. However, this discrepancy with our data is most likely due to the errors in the sequence databases mentioned above. It is of note that EST analysis is based on short sequence fragments only, whereas our results reflect the intact proteins isolated from their natural source.

Crystal structure, overall fold and crystal packing of Bmlp6

Bmlp6 crystallized in space group $P2_1$. The molecular replacement procedure found five copies of the polypeptide chain in the asymmetric unit. The final model was refined to R/R_{free} factors of 16.1/18.3% for diffraction data between 47.0 and 1.8 Å. According to PROCHECK [45], 92.8% of all residues were in the most favored regions of the Ramachandran plot and no residues had disallowed conformation.

Bmlp6, as a 30-kDa LP member, has a fold (Fig. 3B) characteristic of this protein family, classified as Lipoprotein_11 family (Pfam: PF03260). Thus, Bmlp6 is composed of two domains, the N-terminal VHS domain (Pfam: PF00790) and the C-terminal β -trefoil domain (Pfam: PF14200). The N-terminal domain (NTD) of Bmlp6 (residues 1–91) consists of six helices forming a right-handed superhelix and packed in a globular form. The C-terminal domain (CTD) of Bmlp6 (residues 92–245) is folded as a ricin-type β -trefoil lectin-like domain. The β -strands forming the trefoil are connected by an intricate system of loops, all of which are well-defined in electron density. A detailed comparison of the NTD and CTD domains of the 30-kDa LPs with other structures available in the PDB was presented previously [33,34].

The asymmetric unit of Bmlp6 is comprised of five protein molecules denoted A, B, C, D and E (Fig. 4). The r.m.s.d. values for all pairwise $C\alpha$ superpositions of the monomers are very small (ranging from 0.18 to 0.27 Å) indicating an essentially identical folding pattern. The interactions between the monomers in the crystal lattice are weak with no indication of possible quaternary structure, as established by the PISA server [49]. This observation is in agreement with the results of gel filtration and dynamic light scattering (DLS) experiments (not shown), which confirmed the monomeric state of Bmlp6 in solution.

According to the Matthews coefficient and solvent-content analysis [50], the asymmetric unit of the Bmlp6 crystal could contain from 4 to 6 protein molecules. The final solution of the phase problem corresponds to a model comprising five Bmlp6 monomers in the asymmetric unit, with a Matthews coefficient of 2.62 Å³/Da and solvent content of 53.1%. Such a composition of the asymmetric unit is quite rare in protein crystallography, and especially in the space group $P2_1$. Four of the Bmlp6 molecules (A, B, C, D) are arranged in such a manner that the NTD of one monomer (e.g. C) interacts with residues at the NTD/CTD junction of the following (e.g. A) monomer (Fig. 4), forming a chain along the crystallographic **b** axis. The chain is composed of only four protein molecules (C, A, B, D) and is not continuous, meaning that the terminal monomers (D and C) do not interact with their symmetry-related copies. Monomer E is located at the side of the A–D chain and shows even weaker intermolecular interactions than the chain-forming monomers. It is of note that the electron density maps for molecule E were of lower quality, probably because of a higher propensity for disorder allowed by less tight crystal packing. This is also reflected in the average B-factor values calculated in the program BAVEGAGE from the CCP4 suite [51] for each monomer separately. The values are

Table 2. Potential binding sites of the Bmlp6 protein.

Cavity	Area [Å ²]	Volume [Å ³]	Location in the structure	Residues forming the cavity
Po.1	86	97	CTD	N112, S114, E116, S160, T161, T162, D170, R171, Y174
Po.2	101	84	NTD-CTD	G25, Y27, R59, N60, E63, N142, R143

doi:10.1371/journal.pone.0108761.t002

23.4, 21.3, 21.9, 27.5 and 36.0 Å² for monomers A, B, C, D and E, respectively.

At 1.8 Å resolution, we were able to identify in the solvent area of the crystal a number of small molecules from the crystallization buffer. Specifically, molecules of MES, isopropanol and fragments of PEG were mainly located at the protein surface.

Structural comparison of Bmlp6, Bmlp7 and Bmlp3

Bmlp6 shares with Bmlp7 and Bmlp3 ~65% similar and ~49% identical residues (Fig. 3A), which is significantly lower than for the Bmlp3/Bmlp7 pair (97% similarity, 94% identity). When chain A of the Bmlp6 crystal structure was superposed (Fig. 3B) with chain A of Bmlp7 (PDB ID: 4EFP), and with chains A and B of Bmlp3 (PDB ID: 4IY9), the r.m.s.d. values for the C α atoms were 0.8, 0.8 and 1.0 Å, respectively. In terms of secondary structure elements, the main differences are observed for the loops present in the CTD. As reported previously [35], the most flexible loop (Fig. 3C) is formed by residues 106–116 in Bmlp3 and Bmlp7 (110–120 in Bmlp6). Another significant difference is the conformation of loop 159–171 in Bmlp6 (155–165 in Bmlp3 and Bmlp7), Fig. 3D. It is of note that two cysteine residues separated by only one residue (Cys164, Cys166) present in this loop in Bmlp6, form a disulfide bridge, rendering this loop much more rigid. Bmlp3 and Bmlp7 do not contain any disulfide bonds, and the corresponding loop does not contain any cysteine residues and is slightly shorter in those two proteins. All Bmlp6-CTD loops have the same conformation in each of the five monomers present in the asymmetric unit.

Additionally, the protein surface potential of Bmlp6, Bmlp3 and Bmlp7 was compared (Fig. 5). The Poisson-Boltzmann electrostatic potential on the molecular surface of chains A of the crystallographic models of these proteins was calculated using the APBS algorithm [52] and the PDB2PQR program [53,54]. The electrostatic potential calculations were performed at pH 6.5, which is the physiological pH of silkworm hemolymph [55]. Prior to the calculations, the proper side-chain protonation states of the analyzed proteins were determined in PropKa [56]. The patterns of the electrostatic potential on the surfaces of Bmlp3 and Bmlp7 are very similar, whereas significant differences are observed for Bmlp6 (Fig. 5). For instance, the area between helices α 1 and α 3 is negatively charged in Bmlp6, while not in Bmlp3 and Bmlp7.

Chain A of Bmlp6 was also analyzed for potential binding cavities using the CASTp [57] and metaPocket 2.0 [58] servers. There are only two potential cavities in Bmlp6, denoted Po.1 and Po.2 (Fig. 6), in contrast to Bmlp3 and Bmlp7, where three cavities (No. 1–3) are present. Since the analysis produced identical results for the latter two proteins, Bmlp3 [35] will be used for comparison. The cavity Po.2 of Bmlp6 is found at a similar location to pocket No. 2 of Bmlp3, between the NTD and CTD. Moreover, it is formed by corresponding residues in both proteins. However, the cavity No. 2 of Bmlp3 is larger and involves more CTD residues. The pocket Po.1 of Bmlp6 is formed by different residues than cavity No. 3 of Bmlp3, although both pockets are located in a similar region (in the CTD) of the structure. Moreover, the two potential pockets of Bmlp6 (Table 2) are small, with a volume of ~90 Å³, while the largest cavity of Bmlp3 has a volume of ~400 Å³. The pocket Po.1 has a small surface area of 86 Å², although it has a slightly larger volume. It is formed mainly by hydrophilic residues. Interestingly, these residues belong to the two loops discussed above, the flexible loop 110–120 and the Bmlp6-characteristic loop 159–171. It is plausible that this cavity might

undergo some conformational changes upon ligand binding. The cavity Po.2 covers an area of 101 Å² and is also hydrophilic. Both Bmlp6 cavities are located in positively charged areas.

Structural differences of 30-kDa LPs as potential basis of functional diversity

As mentioned above, the silkworm genome contains 46 genes of 30-kDa LPs, 24 of which are classified as typical 30-kDa LPs [11]. 30-kDa LPs have been demonstrated to bind physiological ligands belonging to at least two groups, namely lipids and carbohydrates [16,18,19]. The 30-kDa proteins from silkworm hemolymph have been classified as lipoproteins according to specific lipid staining [16]. On the other hand, the 30-kDa LPs are involved in the immune response pathway: they recognize and bind β -glucans present on the surface of fungal cells [18,19]. Moreover, the insect hemolymph contains a number of chemical compounds [59], and some of them might be binding partners of the 30-kDa LPs as well. Bmlp3, Bmlp6 and Bmlp7 share a similar fold and it is very likely that all 24 typical LPs have the same fold. However, the high number of “isoforms” of these proteins in one organism is suggestive of different ligand specificity of each particular 30-kDa LP member. Moreover, it might be speculated that the proteins play in the insect immune response a role similar to human antibodies. For example, each of the 30-kDa LPs could be specific to a different carbohydrate pattern.

It is of note that even with the crystal structures of only three members of the 30-kDa LP family determined and compared, many differences are already obvious, such as: (i) different conformation of CTD loops, (ii) different electrostatic surface potential, and (iii) different potential binding cavities. It is also possible that the 30-kDa LPs can interact with other silkworm proteins and that the structural differences enable them to recognize different binding partners.

Taken together, the high number of 30-kDa LP genes, their diverse expression patterns and structural differences, all seem to indicate that members of the 30-kDa LP family might represent a wide range of specificities for different physiological ligands present in the hemolymph, as well as different specificity for protein partners, implicating different roles in silkworm physiology.

Conclusions

Bmlp6 is the third most abundant 30-kDa LP in the hemolymph of the silkworm fifth instar larvae. Its identification could be accomplished mainly by sequencing according to electron density maps. The crystal structure of 1.8 Å resolution shows five copies of a two-domain molecule with a typical 30-kDa LP fold. The structure demonstrates that Bmlp6 differs from Bmlp3 and Bmlp7 in the conformation of loops, the electrostatic surface potential, and potential binding cavities. Taking into account the large number of 30-kDa LP genes and the structural differences between the proteins, the question about the exact physiological role of each particular 30-kDa LP member remains open. Our studies suggest functional diversity, fine-tuned to the structural differences, which exist despite the overall similarity.

Acknowledgments

We thank Dr. Piotr Neumann for sharing his scripts for efficient XDS data processing. We acknowledge BioCentrum, Krakow, Poland, for N-terminal sequencing.

Author Contributions

Conceived and designed the experiments: AJP AB MJ GB. Performed the experiments: AJP AB GB. Analyzed the data: AJP AB MJ GB. Contributed reagents/materials/analysis tools: ML. Wrote the paper: AJP AB MJ GB.

References

- Xia Q, Zhou Z, Lu C, Cheng D, Dai F, et al. (2004) A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science* 306: 1937–1940.
- Mita K, Kasahara M, Sasaki S, Nagayasu Y, Yamada T, et al. (2004) The genome sequence of silkworm, *Bombyx mori*. *DNA Res* 11: 27–35.
- The International Silkworm Genome Consortium (2008) The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. *Insect Biochem Mol Biol* 38: 1036–1045.
- Wang J, Xia Q, He X, Dai M, Ruan J, et al. (2005) SilkDB: a knowledgebase for silkworm biology and genomics. *Nucleic Acids Res* 33: D399–D402.
- Duan J, Li R, Cheng D, Fan W, Zha X, et al. (2010) SilkDB v2.0: a platform for silkworm (*Bombyx mori*) genome biology. *Nucleic Acids Res* 38: D453–D456.
- Gamo T (1978) Low molecular weight lipoproteins in the haemolymph of the silkworm, *Bombyx mori*: inheritance, isolation and some properties. *Insect Biochem* 8: 457–470.
- Izumi S, Fuji J, Yamada S, Tomino S (1981) Molecular properties and biosynthesis of major plasma proteins in *Bombyx mori*. *Biochem et Biophys Acta* 660: 222–229.
- Sakai N, Mori S, Izumi S, Haino-Fukushima K, Ogura T, et al. (1988) Structures and expression of mRNAs coding for major plasma proteins of *Bombyx mori*. *Biochem et Biophys Acta* 949: 224–232.
- Mori S, Izumi S, Tomino S (1991) Complete nucleotide sequences of major plasma protein genes of *Bombyx mori*. *Biochem et Biophys Acta* 1090: 129–132.
- Sun Q, Zhao P, Lin Y, Hou Y, Xia QY, et al. (2007) Analysis of the structure and expression of the 30K protein genes in silkworm, *Bombyx mori*. *Insect Sci* 14: 5–14.
- Zhang Y, Dong Z, Liu S, Yang Q, Zhao P, et al. (2012) Identification of novel members reveals the structural and functional divergence of lepidopteran-specific Lipoprotein_11 family. *Funct Integr Genomics* 2: 705–715.
- Li XH, Wu XF, Yue WF, Liu JM, Li GL, et al. (2006) Proteomic analysis of the silkworm (*Bombyx mori* L.) hemolymph during developmental stage. *J Proteome Res* 5: 2809–2814.
- Hou Y, Zou Y, Wang F, Gong J, Zhong X, et al. (2010) Comparative analysis of proteome maps of silkworm hemolymph during different developmental stages. *Proteome Sci* 8: 45.
- Pakkianathan BC, Singh NK, Krishnan M, König S (2012) A proteomic view on the developmental transfer of homologous 30 kDa lipoproteins from peripheral fat body to perivisceral fat body via hemolymph in silkworm, *Bombyx mori*. *BMC Biochem* 13: 5.
- Zhang Y, Dong Z, Wang D, Wu Y, Song Q, et al. (2014) Proteomics of larval hemolymph in *Bombyx mori* reveals various nutrient-storage and immunity-related proteins. *Amino Acids* 46: 1021–1031.
- Whitmore E, Gilbert LI (1974) Hemolymph proteins and lipoproteins in *Lepidoptera*: a comparative electrophoretic study. *Comp Biochem Physiol* 47B: 63–78.
- Sawada H, Yamahama Y, Mase K, Hirakawa H, Iino T (2007) Molecular properties and tissue distribution of 30K proteins as ommin-binding proteins from diapause eggs of the silkworm, *Bombyx mori*. *Comp Biochem Physiol B Biochem Mol Biol* 146: 172e179.
- Ujita M, Kimura A, Nishino D, Yokoyama E, Banno Y, et al. (2002) Specific binding of silkworm *Bombyx mori* 30-kDa lipoproteins to carbohydrates containing glucose. *Biosci Biotechnol Biochem* 66: 2264e2266.
- Ujita M, Katsuno Y, Kawachi I, Ueno Y, Banno Y, et al. (2005) Glucanbinding activity of silkworm 30-kDa apolipoprotein and its involvement in defense against fungal infection. *Biosci Biotechnol Biochem* 69: 1178e1185.
- Vanishree V, Nirmala X, Arul E, Krishnan M (2005) Differential sequestration of storage proteins by various fat body tissues during post larval development in silkworm, *Bombyx mori* L. *Invertebrate Reproduction and Development* 48: 81–88.
- Zhu J, Indrasith LS, Yamashita O (1986) Characterization of vitellin, egg-specific protein and 30 kDa protein from *Bombyx* eggs, and their fates during oogenesis and embryogenesis. *Biochim Biophys Acta* 882: 427–436.
- Chen YL, Yamashita O (1990) Nonselective uptake of different 30 kDa plasma proteins by developing ovaries of the silkworm, *Bombyx mori*. *J Seric Sci Japan* 59: 202–209.
- Maki N, Yamashita O (2001) The 30kP protease A responsible for 30-kDa yolk protein degradation of the silkworm, *Bombyx mori*: cDNA structure, developmental change and regulation by feeding. *Insect Biochem Mol Biol* 31: 407–413.
- Rhee WJ, Kim EJ, Park TH (1999) Kinetic effect of silkworm hemolymph on the delayed host cell death in an insect cell-baculovirus system. *Biotechnol Prog* 15: 1028–1032.
- Rhee WJ, Park TH (2000) Silkworm hemolymph inhibits baculovirus-induced insect cell apoptosis. *Biochem Biophys Res Commun* 271: 186–190.
- Rhee WJ, Park TH (2001) Flow cytometric analysis of the effect of silkworm hemolymph on the baculovirus-induced insect cell apoptosis. *J Microbiol Biotechnol* 11: 853–857.
- Rhee WJ, Kim EJ, Park TH (2002) Silkworm hemolymph as a potent inhibitor of apoptosis in Sf9 cells. *Biochem Biophys Res Commun* 295: 779–783.
- Choi SS, Rhee WJ, Park TH (2005) Beneficial effect of silkworm hemolymph on a CHO cell system: Inhibition of apoptosis and increase of EPO production. *Biotechnol Bioeng* 91: 793–800.
- Kim EJ, Rhee WJ, Park TH (2004) Inhibition of apoptosis by a *Bombyx mori* gene. *Biotechnol Prog* 20: 324–329.
- Choi SS, Rhee WJ, Park TH (2002) Inhibition of human cell apoptosis by silkworm hemolymph. *Biotechnol Prog* 18: 874–878.
- Kim EJ, Rhee WJ, Park TH (2001) Isolation and characterization of an apoptosis-inhibiting component from the hemolymph of *Bombyx mori*. *Biochem Biophys Res Commun* 285: 224–228.
- Park JH, Lee JH, Park HH, Rhee WJ, Choi SS, et al. (2012) A protein delivery system using 30Kc19 cell-penetrating protein originating from silkworm. *Biomaterials* 33: 9127–9134.
- Yang J-P, Ma X-X, He Y-X, Li W-F, Kang Y, et al. (2011) Crystal structure of the 30 K protein from the silkworm *Bombyx mori* reveals a new member of the β -trefoil superfamily. *J Struct Biol* 175: 97–103.
- Pietrzyk AJ, Panjikar S, Bujacz A, Mueller-Dieckmann J, Lochynska M, et al. (2012) High-resolution crystal structure of *Bombyx mori* lipoprotein 7: crystallographic determination of the identity of the protein and its potential role in detoxification. *Acta Cryst D* 68: 1140–1151.
- Pietrzyk AJ, Bujacz A, Mueller-Dieckmann J, Lochynska M, Jaskolski M, et al. (2013) Two crystal structures of *Bombyx mori* lipoprotein 3 - structural characterization of a new 30-kDa lipoprotein family member. *PLoS ONE* 8: e61303.
- Pietrzyk AJ, Bujacz A, Lochynska M, Jaskolski M, Bujacz G (2011) Isolation, purification, crystallization and preliminary X-ray studies of two 30 kDa proteins from silkworm hemolymph. *Acta Cryst F* 67: 372–376.
- Mueller U, Darowski N, Fuchs MR, Förster R, Hellmig M, et al. (2012) Facilities for macromolecular crystallography at the Helmholtz-Zentrum Berlin. *J Synch Rad* 19: 442–449.
- Kabsch W (2010) XDS. *Acta Cryst D* 66: 125–132.
- Kabsch W (2010b). Integration, scaling, space-group assignment and post refinement. *Acta Cryst D* 66: 133–144.
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. *J Appl Cryst* 40: 658–674.
- Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Cryst D* 66: 486–501.
- Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, et al. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst D* 67: 355–367.
- Painter J, Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Cryst D* 62: 439–450.
- Brünger AT (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature (London)* 355, 472–475.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 26: 283–291.
- Schaeffer H, Jagow G (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem* 166: 368–379.
- Pietrzyk AJ, Bujacz A, Mueller-Dieckmann J, Lochynska M, Jaskolski M, et al. (2013) Crystallographic identification of an unexpected protein complex in silkworm hemolymph. *Acta Cryst D* 69: 2353–2364.
- Pietrzyk AJ, Bujacz A, Jaskolski M, Bujacz G (2013) Identification of amino acid sequence by X-ray crystallography: a mini review of case studies. *BioTechnology - J Biotech Comp Biol Bionanotech* 94: 9–14.
- Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372: 774–797.
- Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* 33: 491–497.
- Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Cryst D* 67: 235–242.
- Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98: 10037–10041.
- Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* 32: W665–667.
- Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, et al. (2007) PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res* 35: W522–525.

55. Kaissling KE, Thorson J (1980) Receptors for neurotransmitters, hormones and pheromones in insects, eds. Sattelle D B, Hall L M, Hildebrand J G, Elsevier, Amsterdam, 261–282.
56. Li H, Robertson AD, Jensen JH (2005) Very fast empirical prediction and rationalization of protein pKa values. *Proteins* 61: 704–721.
57. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, et al. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acid Res* 34: W116–W118.
58. Huang B (2009) metaPocket: a meta approach to improve protein ligand binding site prediction. *Omics* 13: 325–330.
59. Wyatt GR, Loughheed TC, Wyatt SS (1956) The chemistry of insect hemolymph. *J Gen Physiol* 39: 853–868.
60. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25: 1189–1191.