



INNOVATIVE ECONOMY
NATIONAL COHESION STRATEGY



INSTITUTE OF BIOORGANIC CHEMISTRY
POLISH ACADEMY OF SCIENCES
POZNAN



EUROPEAN UNION
EUROPEAN REGIONAL
DEVELOPMENT FUND

**Polish Academy of Sciences
Institute of Bioorganic Chemistry
Center for Biocrystallographic Research**

**STRUCTURAL STUDIES
OF MAJOR HEMOLYMPH PROTEINS
FROM MULBERRY SILKWORM
Bombyx mori L.**

AGNIESZKA J. PIETRZYK

*Thesis presented to the Scientific Council
of the Institute of Bioorganic Chemistry
Polish Academy of Sciences in Poznan
as a Ph.D. dissertation.*

Thesis supervisor: Prof. Grzegorz Bujacz

Poznan 2014



**Partial support was provided by the European Union
within the European Regional Development Fund.**

***To my parents
for their love and endless support***

*I wish to thank, first and foremost, my supervisor, Professor Grzegorz Bujacz,
for introducing me to the secrets of crystallography,
for his guidance, encouragement and immense knowledge.
I have been extremely lucky to have a supervisor who cared so much about my work.*

*My special thanks go to Dr. Anna Bujacz for her support, enthusiasm, advice
and for introducing me to the world of crystallographic software.*

*It gives me great pleasure to acknowledge Professor Mariusz Jaskólski
for his valuable comments and help in preparation
of manuscripts describing obtained results.*

*My sincere thanks also go to Dr. Santosh Panjikar and Dr. Jochen Müller-Dieckmann
for leading me during my internship at EMBL-Hamburg.*

*This thesis would not have been possible without the starting material,
silkworm larvae, obtained thanks to the courtesy of Dr. Małgorzata Łochyńska.*

*I would like to thank Dr. Piotr Neumann
for sharing his XDS scripts for efficient XDS data processing.*

I acknowledge MSc. Wiesław Majzner for technical support.

*I wish to thank my fellow labmates,
Dr. Izabela Szymczak, MSc. Bartosz Sekuła ,
MSc. Paweł Strzelczyk, MSc. Kamil Zieliński and MSc. Julita Tałaj,
for the stimulating discussions, for the sleepless nights
we were collecting together X-ray diffraction data,
and for all the fun we have had learning crystallography.*

*Last but not least, I am deeply grateful to my fiancé, Marek Brzeziński,
who has been always by my side throughout my PhD studies.*

*"Why water boils at 100° and methane at -161°, why blood is red and grass is green,
why diamond is hard and wax is soft, why glaciers flow and iron gets hard
when you hammer it, how muscles contract, how sunlight makes plants grow
and how living organisms have been able to evolve into ever more complex forms...
the answers to all these problems have come from structural analysis."*

Max Perutz, July 1996

TABLE OF CONTENTS

1. ABBREVIATIONS	1
2. ABSTRACT	3
3. INTRODUCTION	4
3.1. MULBERRY SILKWORM	4
3.1.1. Life cycle of mulberry silkworm	5
3.1.2. The silkworm genome and transcriptome	7
3.1.3. Mulberry silkworm as an important insect in industry and science	8
3.2. HEMOLYMPH PROTEINS	9
3.2.1. High molecular weight storage proteins	11
3.2.2. 30-kDa Lipoproteins	12
3.2.2.1. The fate and the role of 30-kDa LPs in silkworm body	12
3.2.2.2. The nucleotide and amino acid sequences of 30-kDa LPs	14
3.2.3. Anti-apoptotic properties of hemolymph proteins	17
3.2.4. Cell-penetrating properties of hemolymph proteins	19
4. THE AIM OF THE WORK	20
5. MATERIALS & METHODS	21
5.1. HEMOLYMPH COLLECTION	21
5.2. PROTEIN PREPARATION	21
5.2.1. Gel filtration	21
5.2.2. Ion exchange chromatography	22
5.2.3. Hydrophobic interaction chromatography	22
5.2.4. Protein desalting	22
5.2.5. Protein sample characterization	23
5.3. PROTEIN CRYSTALLIZATION	24
5.4. PROTEIN CRYSTALS DERIVATIZATION	25
5.5. X-RAY DIFFRACTION DATA COLLECTION	26
5.6. X-RAY DIFFRACTION DATA PROCESSING	27
5.7. STRUCTURE DETERMINATION	28
5.7.1. Single- and multiple-wavelength anomalous dispersion	29
5.7.2. Molecular replacement	31
5.8. STRUCTURE REFINEMENT	33
5.8.1. Reciprocal space refinement	33
5.8.2. Real space refinement	34
5.9. STRUCTURE VALIDATION	35

Table of contents

5.10. STRUCTURE ANALYSIS	36
6. RESULTS & DISCUSSION	37
6.1. PROTEIN PURIFICATION AND INITIAL CHARACTERIZATION OF THE SAMPLES	37
6.2. PROTEIN CRYSTALLIZATION, X-RAY DIFFRACTION DATA COLLECTION AND DATA PROCESSING	42
6.3. STRUCTURE DETERMINATION	47
6.3.1. The initial model of Bmlp7 was obtained by SAD	47
6.3.2. MR as the most common method for phase problem solution	47
6.4. SEQUENCING FROM ELECTRON DENSITY MAPS	48
6.4.1. Crystallographic identification of Bmlp7	49
6.4.2. Crystallographic identification of Bmlp3	51
6.4.3. Crystallographic identification of an unexpected complex of SP2 and SP3	51
6.5. STRUCTURE REFINEMENT AND VALIDATION	57
6.6. OVERALL STRUCTURE	57
6.6.1. Overall structure of SP2-SP3 complex	57
6.6.2. Overall structure of 30-kDa LPs	59
6.7. STRUCTURAL COMPARISON	62
6.7.1. Primary structure determines the role of a protein: arylphorins vs. hemocyanins	62
6.7.2. The putative role of 30-kDa LPs domains deduced from structural comparison	64
6.7.2.1. The putative role of the 30-kDa LPs NTD	64
6.7.2.2. The putative role of the 30-kDa LPs CTD	66
6.8. PROTEIN-PROTEIN INTERACTIONS AND CRYSTAL PACKING	66
6.8.1. The heterohexamer of SP2 and SP3	66
6.8.2. Crystal packing of 30-kDa LPs	67
6.9. MODIFIED AMINO ACID RESIDUES	70
6.9.1. N-glycosylation motif of arylphorins	71
6.9.2. Modified Trp residue of Bmlp7	72
6.10. THE PUTATIVE ROLE OF Bmlp7 IN DETOXIFICATION MECHANISM OF SILKWORM	73
6.11. THE PUTATIVE BINDING CAVITIES OF 30-kDa LPs	76
6.12. PROTEIN SURFACE OF Bmlp3 vs. CELL-PENETRATING PROPERTIES	78
7. SUMMARY	80

8. REFERENCES	82
9. LIST OF TABLES	94
10. LIST OF FIGURES	95
11. SCIENTIFIC OUTPUT	97

APPENDIX – PUBLICATIONS

- I **Pietrzyk A.J.**, 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.* **F67**, 372-376.

- II **Pietrzyk A.J.**, Panjekar S., Bujacz A., Mueller-Dieckmann J., Lochynska M., Jaskolski M., Bujacz G. (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.* **D68**, 1140-1151.

- III **Pietrzyk A.J.**, Bujacz A., Mueller-Dieckmann J., Lochynska M., Jaskolski M., Bujacz G. (2013). Two crystal structures of *Bombyx mori* lipoprotein 3 - structural characterization of a new 30-kDa lipoprotein family member. *PLoS ONE* **8**, e61303.

- IV **Pietrzyk A.J.**, Bujacz A., Jaskolski M., Bujacz G. (2013). Identification of amino acid sequence by X-ray crystallography: a mini review of case studies. *BioTechnologia - J. Biotech. Comp. Biol. Bionanotech.* **94**, 9-14.

- V **Pietrzyk A.J.**, Bujacz A., Mueller-Dieckmann J., Lochynska M., Jaskolski M., Bujacz G. (2013). Crystallographic identification of an unexpected protein complex in silkworm hemolymph. *Acta Cryst.* **D69**, 2353-2364.

20E	20-hydroxyecdysone
30-kDa LPs	30-kDa lipoproteins
A ₂₈₀	absorbance value at 280 nm
AmS	ammonium sulfate
APA	<i>Antheraea pernyi</i> arylphorin
AU	asymmetric unit
Bis-Tris	bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane
BmJHBP	<i>Bombyx mori</i> juvenile hormone binding protein
Bmlp(1-46)	<i>Bombyx mori</i> lipoprotein (1-46)
CCD	charge-coupled device
CTD	C-terminal domain
DLS	dynamic light scattering
EST	expressed sequence tag
FB	fat body
FPLC	fast performance liquid chromatography
GF	gel filtration
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIC	hydrophobic interaction chromatography
IEC	ion exchange chromatography
JH	juvenile hormone and its isoforms
LC-MS/MS	liquid chromatography coupled tandem mass spectrometry
MAD	multiple-wavelength anomalous dispersion
MALDI-TOF MS	matrix-assisted laser desorption ionization time-of-flight mass spectrometry
MIR	multiple isomorphous replacement
MR	molecular replacement
NTD	N-terminal domain
PDB	Protein Data Bank
PEG	polyethylene glycol
PEG MME	polyethylene glycol monomethyl ether
PPFB	peripheral fat body
PVFB	perivisceral fat body
R _{merge}	merging R-factor

1. Abbreviations

r.m.s.d.	root-mean-square deviation
SAD	single-wavelength anomalous dispersion
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SIR	single-wavelength anomalous dispersion
SP(1-3)	storage protein (1 - 3)
TAT	trans-activator of transcription
Tris	tris(hydroxymethyl)aminomethane

Insects are a class of invertebrates belonging to arthropods and one of the most diverse groups of animals on Earth. It is easy to find common features of insects, they are tiny six-legged animals with a three-part body structure (comprising of head, thorax and abdomen) and a chitinous exoskeleton. However, further comparison becomes more difficult. Morphology, physiology, food habitats and development of a particular insect reveal differences between insects' species and are strictly connected to the adaptation of the insect to its environment (Klowden, 2007). At the molecular level these differences are determined by genes and consequently, by proteins, the products of the genes' expression. Proteins are involved in metabolic pathways, signal transduction, transport of specific ligands, immune response, etc. Proteins participate in almost all processes occurring in the living body and, what is noteworthy, their function is strictly dependent on their structure, which is unique for each protein molecule. Nowadays, the final characterization of particular species should consist not only of the description of their morphology and physiology, but it should also contain the information about their genes and proteins.

My PhD project concerns the purification, crystallization and crystal structure determination of silkworm hemolymph proteins. The first part of this chapter is dedicated to the mulberry silkworm, its development and its significance to industry, medicine and molecular biology. In the second part, all available information about the major hemolymph proteins is summarized.

3.1. MULBERRY SILKWORM

Among more than a million described insect species (Chapman, 2006) only two are domesticated: the honey bee (*Apis mellifera*) and the mulberry silkworm (*Bombyx mori*). The name of the latter one refers to its preferred source of food, namely leaves of white mulberry (*Morus alba*; Hiratsuka, 1999). The mulberry silkworm belongs to the family *Bombycidae* of insect order *Lepidoptera*. The breeding of silkworms for the production of silk dates back to the third millennium B.C.E. In consequence of over 5 000 years of breeding, the insect became incapable of living in a natural environment. The silkmoth lost its ability to fly and is now strictly dependent on humans (Goldsmith *et al.*, 2005). Its closest relative is the wild silkworm (*Bombyx mandarina*), with which domesticated silkworm is still able to hybridize (Arunkumar *et al.*, 2006). To this date, the mulberry silkworm remains the main producer of natural silk fiber, but more and more often this insect serves as a model organism in science (Goldsmith *et al.*, 2005) and as a "factory" for production of recombinant

eukaryotic proteins (Mathavan *et al.*, 1995; Kadono-Okuda *et al.*, 1995; Sumathy *et al.*, 1996; Tomita *et al.*, 2003).

3.1.1. Life cycle of mulberry silkworm

The mulberry silkworm, like all insects belonging to the *Lepidoptera* order, undergoes a complete metamorphosis (Grzimek *et al.*, 2004), which is a highly successful mechanism of biological adaptation. A larva of a particular insect differs significantly from its adult form in morphology, physiology and diet. Such differences allow the adaptation of one developmental stage for a particular role (Truman and Riddiford, 1999).

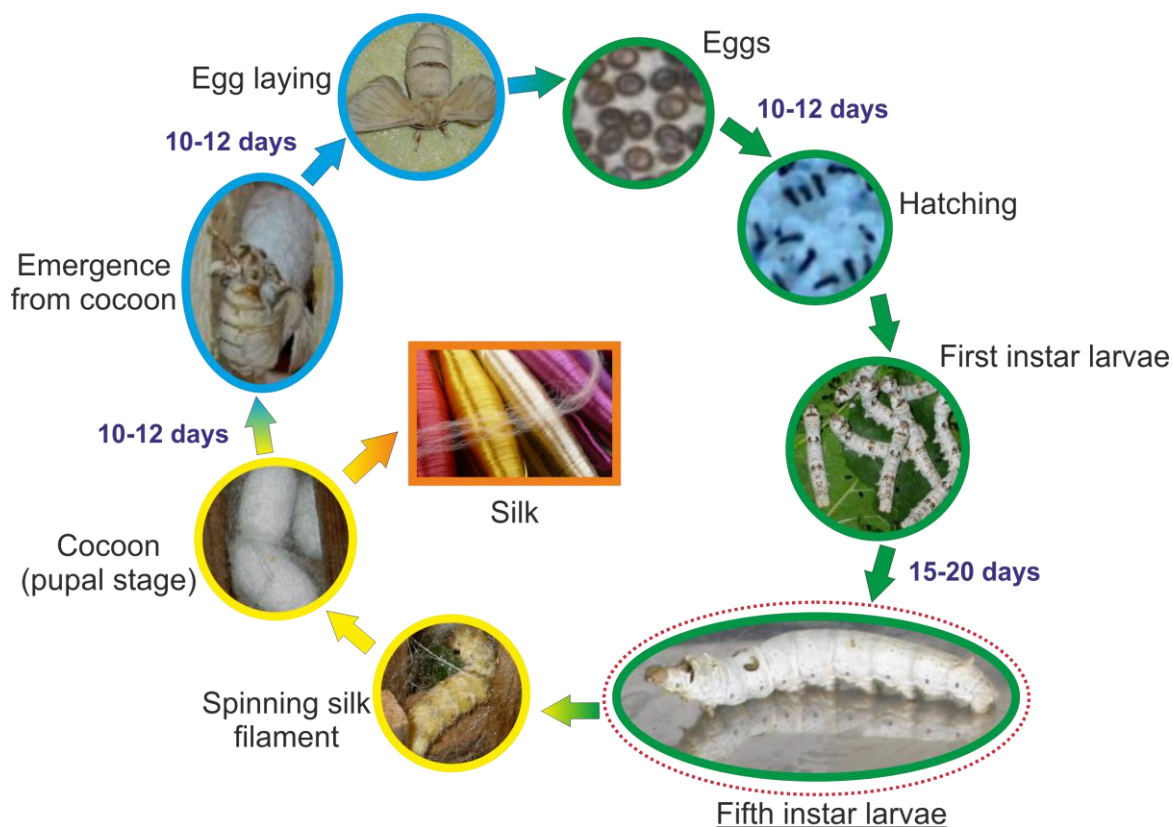


Fig. 3.1 The full life cycle of the mulberry silkworm.

The scheme presents the development of the mulberry silkworm divided into three stages: the larval (green), the pupa (yellow) and the adult form (blue). The proteins of interest, the object of this dissertation, were isolated from fifth instar larvae, marked by a red dotted ellipse. The scheme was prepared according to Hiratsuka, 1999; Grzimek *et al.*, 2004. Photos were obtained thanks to the courtesy of Dr. Małgorzata Łochyńska from the Institute of Natural Fibres and Medicinal Plants in Poznań.

The full cycle of development (Fig. 3.1), starting from a caterpillar hatched from an egg to the adult form, is controlled by two families of hormones, the juvenile hormones (JH) and the ecdysteroids (Fig. 3.2) (Riddiford and Truman, 1978). Interestingly,

3. Introduction

the ecdysone was isolated for the first time from the mulberry silkworm pupae (Butenandt and Karlson, 1954). Silkworm's development can be divided into three stages: the larva (caterpillar), the pupa and the adult form (butterfly). The presence of JH in the hemolymph maintains the larval stage (Riddiford and Truman, 1978), a caterpillar feeds itself with mulberry leaves and is growing through five instars (growth steps with molting). At the end of the fifth instar, the level of JH decreases in the hemolymph and an increase of ecdysteroids, especially 20-hydroxyecdysone (20E), occurs and triggers the metamorphosis (Riddiford, 1993). The larva spins itself into a cocoon of raw silk produced in the silk glands. Fibroins, one of the main constituents of raw silk, are synthesized in the posterior silk glands, then extruded by peristaltic motion to the ducts of the middle silk glands and finally they reach the lumen of the anterior silk glands (Magoschi *et al.*, 1994). The reduction of pH from the posterior to the anterior part of the silk glands causes gelation of the condensed fibroin (Terry *et al.*, 2004). At the end of the larval and the beginning of the pupal stage the high titre of 20E in the hemolymph induces the programmed cell-death of the silk glands (Terashima *et al.*, 2000). The caterpillar enclosed in the cocoon digests itself. During metamorphosis all tissues are disintegrated except for the imaginal discs which are responsible for cellular differentiation and morphogenesis (Andres and Thummel, 1992). An adult form of mulberry silkworm is a white butterfly, which has reduced mouth parts and does not feed (Grzimek *et al.*, 2004).

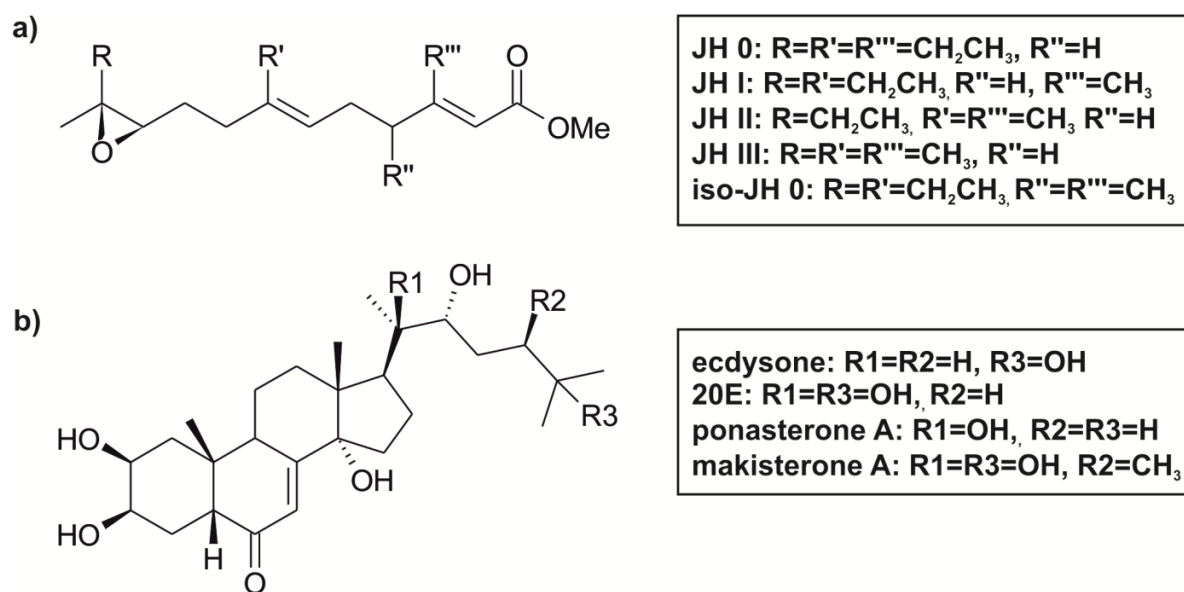


Fig. 3.2 Insect hormones.

Chemical structures of **a)** juvenile hormone isoforms (Kochman and Wieczorek, 1994) and **b)** ecdysteroids (Cymborowski, 1984).

3.1.2. The silkworm genome and transcriptome

The mulberry silkworm was the first insect within the order *Lepidoptera* for which a draft of the genomic sequence was completed in 2004, independently by Chinese (Xia *et al.*, 2004) and Japanese (Mita *et al.*, 2004) groups. Four years later, the two data sets from whole-genome shotgun sequencing were merged, supplemented with newly obtained sequences and made available by The International Silkworm Genome Consortium formed by the two mentioned research teams. 14 623 gene models were predicted and for about three thousand of them no homologs in other insect or vertebrate genomes were found. These genes are related to silk production and the silkworm's food habitats. For instance, the presence of new-type sucrase genes is an adaptation to feeding on mulberry leaves containing toxic alkaloids (The International Silkworm Genome Consortium, 2008). The genome information obtained in 2004 by the Chinese group was collected and in 2005 the first knowledgebase for silkworm biology and genomics, SilkDB, was created (Wang *et al.*, 2005). The updated version of this database, SilkDB v2.0, has been accessible since 2010 (Duan *et al.*, 2010).

In 2012 the silkworm transcriptome was studied. High-throughput RNA sequencing was used to explore the transcriptome of silkworm at different developmental stages. 11 884 of 14 623 predicted genes were found to be expressed in the silkworm body. The transcriptome analysis indicated that 3 247 genes undergo alternative splicing. The authors concluded that the silkworm transcriptome has not yet been fully elucidated due to its functional complexity. A database with the transcriptome data (SilkTransDB) is accessible online (Li *et al.*, 2012).

Although detailed information about the silkworm genome and transcriptome is now available, further studies on the silkworm proteome are necessary for the exact description of the biological processes occurring in its body. Still few silkworm proteins have been well-characterized and the number of silkworm protein structures in the Protein Data Bank (PDB; Berman *et al.*, 2000) is very low. In September 2009, at the beginning of the project presented in this dissertation, there were 34 structures of silkworm proteins deposited in the PDB. As of December 2013 this number increased to 61 structures, among them six are the results of the research described in this thesis.

3.1.3. Mulberry silkworm as an important insect in industry and science

The art of rearing silkworms for the production of raw silk is called sericulture and has become one of the most important cottage industries in many countries. The two main silk producers are China and India, manufacturing 60 % of the world's silk. Sericulture has been taken up to prevent poverty and unemployment of people in rural areas of a number of countries, i.e. Brazil, Thailand, Vietnam, Indonesia, Egypt, Iran, Sri Lanka, the Philippines, Bangladesh, Nepal, Myanmar, Turkey, Papua New Guinea, Mexico, Uzbekistan. The global silk production is around 70, 000 to 90, 000 tonnes per year (Sohn, 2003).

Furthermore, the chrysalides, rejected from silk industry after fiber extraction, can be used as adsorbents for lead, nickel and copper cations from wastewaters. The chrysalides are rich in chitin and have good metal-adsorbent capacity (Paulino *et al.*, 2006). Chitin can also be extracted from chrysalides and deacetylated to obtain chitosan which adsorbs metal cations more efficiently. The procedure of chitosan production from silkworm chrysalides was developed by Paulino *et al.*, (2007).

The silkworm also became an attractive object for science due to low costs of its feeding and the ease of its rearing. It is a good model organism for pest control studies (Goldsmith *et al.*, 2005) and a bioreactor for recombinant eukaryotic protein production (Mathavan *et al.*, 1995; Kadono-Okuda *et al.*, 1995; Sumathy *et al.*, 1996; Tomita *et al.*, 2003). Viruses were employed as expression vectors and the *Bombyx mori* nuclear polyhedrosis virus (NPV) was the most popular. The first expressed recombinant proteins in silkworm expression system were: human parathyroid hormone (Mathavan *et al.*, 1995) and human growth hormone (Kadono-Okuda *et al.*, 1995; Sumathy *et al.*, 1996). The recombinant proteins are usually secreted into the hemolymph; however, the generation of transgenic silkworms that produce cocoons containing recombinant human III procollagen mini-chain was also reported (Tomita *et al.* 2003). Recently, a new baculovirus-silkworm multigene expression system (*Bombyx mori* MultiBac) was developed enabling an effective simultaneous expression of several heterologous genes in silkworm larvae or pupae (Yao *et al.*, 2012).

Mulberry silkworm is also used in medical tests. The researchers were looking for a non-vertebrate model to test drug efficacy in the early stages of its development and to eliminate drug candidates with a high toxicity. It was established that the mulberry silkworm is a good model, consistent with the murine model, for development of drugs

against *Staphylococcus aureus* infection (Fujiyuki *et al.*, 2010), as well as for the evaluation of therapeutic drugs for hyperuricemia and gout (also known as podagra) treatment (Zhang *et al.*, 2012a).

Finally, the genome and proteome comparison of domesticated silkworm (*B. mori*) and wild silkworm (*B. mandarina*) is interesting from an evolutionary point of view. The comparative studies of both species enable us to observe the impact of selective breeding and to describe the evolution related gene mutations. Molecular phylogenetic analysis indicates that the mulberry silkworm probably originated from *B. mandarina* (Arunkumar *et al.*, 2006).

3.2. HEMOLYMPH PROTEINS

The silkworm, as all insects, has an open circulatory system and hemolymph is the arthropods' analog of blood. The most important functions of hemolymph are: the transport of nutrient substrates to utilization sites and delivery of metabolic wastes to the excretory system. Hemolymph also serves as a reservoir of water and as a medium for cellular and humoral defense. The immune response is triggered in the hemolymph when bacteria or fungi enter a silkworm's body (Klowden, 2007). The percentage of hemolymph as a proportion of body weight in the silkworm is especially high after eating, thus the body of the extensively feeding silkworm larvae contains a lot of hemolymph (Hiratsuka, 1999).

The most abundant group of hemolymph proteins are lipoproteins, the proteins involved in lipid transport, which was established by specific lipid staining following hemolymph electrophoresis (Whitmore and Gilbert, 1974). Silkworm hemolymph contains two main lipoprotein fractions, high molecular weight storage proteins (Whitmore and Gilbert, 1974) and low molecular weight 30-kDa lipoproteins (30-kDa LPs) (Gamo, 1978; Izumi *et al.*, 1981), both groups will be discussed in the following sections. Hemolymph lipoproteins transport lipids released from the silkworm fat body (FB). The main form in which lipids are transported in insects are diglycerides, although they are stored in the FB as triglycerides. The FB is an analog of the vertebrate liver and adipose tissue (Gilbert and Chino, 1974). There are two types of FB: larval (peripheral) FB (PPFB) located at the epidermis, and pupal (perivisceral) FB (PVFB) situated in the visceral compartment (Anderson, 1972). Hemolymph lipoproteins are synthesized in large quantities in the PPFB (Fig. 3.3a) during the feeding period, mainly during the fifth instar, and secreted (Fig. 3.3b)

to hemolymph (Vanishree *et al.*, 2005). At the end of the fifth instar, when the spinning starts, the lipoproteins are taken up into the PVFB and stored as granules (Fig. 3.3c). They are storage proteins for adult development when the moth cannot feed anymore (Locke and Collins, 1968).

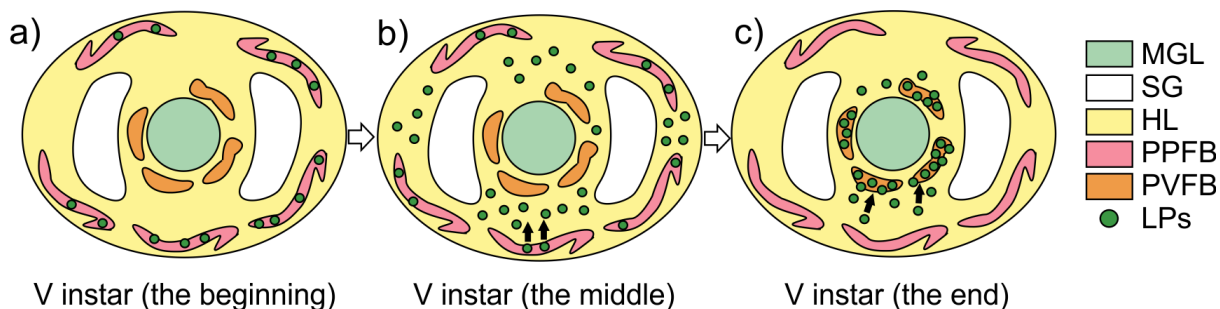


Fig. 3.3 The fate of lipoproteins in a silkworm's body.

The cross section scheme of the silkworm larva **a)** at the beginning, **b)** the middle and **c)** the end of the fifth instar and the localization of the lipoproteins. The expansions of the abbreviations in the legend are: MGL - midgut lumen, SG - silk glands, HL - hemolymph, PPFB - peripheral fat body, PVFB - perivisceral fat body, LPs - lipoproteins. Only the elements important for this scheme are shown. The scheme was prepared according to Locke and Collins, 1968; Vanishree *et al.*, 2005.

In the last few years proteomic studies of silkworm were focused on the quantitative identification of the proteins present in the silkworm body. Three publications reported the analysis of hemolymph proteome (Li *et al.*, 2006; Hou *et al.*, 2010; Pakkianathan *et al.*, 2012). The proteome of PPFB and PVFB was also described (Pakkianathan *et al.*, 2012). Additionally, one report concerned the embryonic proteome (Li *et al.*, 2009a), one the endocrine organs (Li *et al.*, 2009b) and one was about the proteome of silkworm head (Li *et al.*, 2010). All these studies were performed using 2D electrophoresis, liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) or matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS). The identification of proteins was done according to LC-MS/MS data and its analysis using the public databases, i.e. UniProt, SilkDB. The hemolymph proteome studies indicated that the protein content varies greatly at different stages, the analysis was carried out for hemolymph collected from the first day of the fifth instar larvae to the matured fifth instar larvae (Li *et al.*, 2006), to the sixth day of pupal stage (Pakkianathan *et al.*, 2012) or to the emergence of adult moth from cocoon (Hou *et al.*, 2010). The number of proteins present in the hemolymph ranges from 241 to 298. The proteins are concentrated mainly in pI

3.5-6.5 (Li *et al.*, 2006) and usually have a molecular weight (of monomers) between 20 and 97 kDa (Hou *et al.*, 2010).

3.2.1. High molecular weight storage proteins

Two forms of high molecular weight storage proteins, which are present in the mulberry silkworm were widely described to this date. They are referred to as *Bombyx mori* storage protein 1 (SP1; Sakurai *et al.*, 1988) and 2 (SP2; Fujii *et al.*, 1989). Both proteins are often also denoted as sex-specific proteins. Until the end of the fourth larval instar their amount in the hemolymph is relatively low and equal in both sexes. They are synthesized predominantly during the last instar, parallel with the disappearance of JH from hemolymph (Tojo *et al.*, 1980) and SP1 is then expressed only in females (Sakurai *et al.*, 1988). The level of SP2 expression is two times higher in females, but the protein is present also in males (Fujii *et al.*, 1989). Both proteins are transported to PVFB and the protein granules are formed (Locke and Collins, 1968). The proteomic studies revealed that SP1 and SP2 constitute 60% of total FB protein in females, whereas only 20% in males (Hou *et al.*, 2010).

Both proteins, SP1 and SP2, have a similar molecular size of around 500 kDa and are hexamerins. Each contains six subunits of a molecular weight of about 85 kDa (Tojo *et al.*, 1980). However, their amino acid sequences (Table 3.1) and biochemical characteristics are different and they belong to two different protein families. SP1 is a methionine-rich hexamerin with a content of methionine of 11.1% (Sakurai *et al.*, 1988). The amino acids composition of SP2 is rich in aromatic amino acids (mainly phenylalanine and tyrosine) which constitute 19.0%. Thus, SP2 is classified as an arylphorin-type storage protein (Fujii *et al.*, 1989). The term "arylphorin" was proposed by Telfer *et al.* (1983) when the first protein of this type was isolated from the tobacco hornworm, *Manduca sexta*. The content of aromatic amino acids in arylphorins isolated from lepidopteran species is usually between 15 and 20 % mol. (Tojo *et al.*, 2012). Both types of hexamerines are evolutionarily related to arthropod hemocyanin family (Telfer and Kunkel, 1991). The solution of insect high molecular weight lipoproteins is usually yellow, due to their content of 3-carotene and lutein which are not covalently bound to the protein molecules (Chino *et al.*, 1969). The main physiological role of SP1 and SP2 is the storage of amino acids for the development of adult tissues at the final developmental stages (Levenbook and Bauer, 1984). The more detailed review about storage proteins in *Lepidoptera*, including SP1 and SP2, was prepared by Tojo *et al.* (2012).

3. Introduction

Table 3.1 Sequence homology among silkworm storage proteins.

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

The table presents % sequence similarities (and identities) among different sequences of oak silkworm (APA) and mulberry silkworm storage proteins (SP1, SP2, SP3) calculated using BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>; Altschul *et al.*, 1990). UniProt_1 and UniProt_2 refer to two different sequences of SP2, accession codes: P20613 and Q1HPP4, respectively.

According to recent reports, SP1 and SP2 are not the only storage proteins present in mulberry silkworm hemolymph. The genomic (The International Silkworm Genome Consortium, 2008) and proteomic (Hou *et al.*, 2010) studies revealed that there is at least one more silkworm storage protein, SP3. The only available information about this uncharacterized protein is that it can be classified as an arylphorin (aromatic amino acid residues constitute 20.7% of the sequence) and shares a similar expression pattern with SP2 (Hou *et al.*, 2010).

To this date, only one crystal structure of arylphorin (from wild oak silkworm, *Antheraea pernyi*) was determined (PDB code: 3GWJ; Ryu *et al.*, 2009) and no structures of methionine-rich hexamerins are available in the PDB.

3.2.2. 30-kDa Lipoproteins

Hemolymph low molecular weight lipoproteins are usually referred to as major hemolymph proteins or 30-kDa lipoproteins (30-kDa LPs) due to their molecular weight of about 30 kDa (Gamo, 1978; Izumi *et al.*, 1981; Sakai *et al.*, 1988; Mori *et al.*, 1991). In 1981 it was reported that 30-kDa LPs are synthesized in the silkworm body in a stage-dependent manner. The accumulation of these proteins in hemolymph occurs after the third day of the fifth larval instar (Izumi *et al.*, 1981) and thus 30-kDa LPs are also defined as proteins specific to the final larval stage.

3.2.2.1. The fate and the role of 30-kDa LPs in silkworm body

The biosynthesis of 30-kDa LPs is regulated at the transcriptional level (Mori *et al.*, 1991). The 30-kDa LPs are synthesized in PPFB of both sexes (Vanishree *et al.*, 2005), mainly during the fifth instar - from the very beginning of this developmental stage

to the pupation. The start of the synthesis is connected to the disappearance of the JH from the hemolymph (Izumi *et al.*, 1981). It was established that surgical removal of *corpora allata*, which is the site of JHs synthesis, leads to an accumulation of 30-kDa LPs in the hemolymph of fourth instar larvae (Izumi *et al.*, 1984). Inversely, cultured PPFb cells treated with methoprene, a JH analog, do not synthesize 30-kDa LPs. The JH response element was identified in the promoter regions of the 30-kDa genes (Ogawa *et al.*, 2005). Additionally, it was found that PBX proteins, a superfamily of atypical homeodomain-containing transcription factors (Burglin, 1994), suppress the 30-kDa genes' expression (Ogawa *et al.*, 2005).

As well as high molecular weight lipoproteins, 30-kDa LPs are secreted into the hemolymph after synthesis and moved into PVFB (Vanishree *et al.*, 2005), where they accumulate forming high molecular weight aggregates (Pakkianathan *et al.*, 2012). They serve as storage proteins during pupation and adult development (Vanishree *et al.*, 2005). The 30-kDa LPs are used as the source of nutrition after enzymatic digestion (Hou *et al.*, 2010).

In female moths, 30-kDa LPs are relocated from PVFB into yolk granules and after vitellin they are the second major yolk protein (Zhu *et al.*, 1986; Chen and Yamashita, 1990; Maki and Yamashita, 2001). The 30-kDa LPs constitute about 35 % of total yolk proteins (Zhu *et al.*, 1986). During embryonic development 30-kDa LPs are still present in the silkworm embryo body (Hou *et al.*, 2010). It was reported that in diapause silkworm eggs a complex transporting tryptophan metabolites and ommochrome (a natural organic pigment) for the creation of serosa, membranes which line and enclose several body cavities, is formed by 30-kDa LPs (Sawada *et al.*, 2007). The 30-kDa LPs disappear after hatching (Hou *et al.*, 2010).

Beside being a source of nutrition, 30-kDa LPs are probably also involved in immune response pathway, namely antifungal defense system (Ujita *et al.*, 2002; 2005). It was established that members of the 30-kDa LPs family are able to specifically bind glucose and glucans (Ujita *et al.*, 2002) which are the main components of fungal cell walls (Jiang *et al.*, 2004). Members of the 30-kDa LPs family specifically recognize laminaribiose (Glc β 1-3Glc) and laminarin, a soluble β -glucan, but do not interact with glycoproteins and trehalose, an insect blood sugar. Glucose is the only monosaccharide recognized by the investigated 30-kDa LPs, which do not bind N-acetylglucosamine, galactose, N-acetylgalactosamine, mannose or fucose. Finally, it was reported that a member

of the 30-kDa LP family activates a prophenoloxidase cascade, the immune response pathway (Ujita *et al.*, 2005). The cascade activation results in the generation of quinones, intermediates for melanization. Later, melanin forms a complex with lipophorin and the invading fungi are encapsulated by this complex (Gupta, 1991).

3.2.2.2. The nucleotide and amino acid sequences of 30-kDa LPs

There are a number of available nucleotide and amino acid sequences of 30-kDa LPs in several international databases, such as: GeneBank, UniProt or SilkDB. Unfortunately, the information about 30-kDa LPs is not well-ordered and often the same sequence can be found with several different accession codes. Sometimes, among these few sequences referring to the same protein, slight differences are present. The sequence variability is connected to the use of different strains of *B. mori* during sequencing experiments. It was found that amino acid composition of 30-kDa LPs is different among mulberry silkworm strains and races (Chen and Yamashita, 1990).

The first cDNA clones of 30-kDa LPs with distinct mRNA were isolated by Sakai *et al.* (1988) from a cDNA library constructed from the FB mRNA of the fifth instar larvae. Five amino acid sequences, annotated PBMHP-6, PBMHP-12, PBMHPC-19, PBMHP-21 and PBMHPC-23 or LP(1-5), were determined (Sakai *et al.*, 1988). Three isoforms of PBMHP-6, PBMHPC-19 and PBMHP-21, denoted as 6G1, 19G1 and 21G1, respectively, were reported by Mori *et al.* (1991).

Ten genes of 30-kDa LPs, *Bmlp1-10*, were predicted (Sun *et al.*, 2007) based on the analysis of the completed silkworm genomic sequence (Mita *et al.*, 2004; Xia *et al.*, 2004). Furthermore, phylogenetic analysis indicated that it is most likely that the reported 30-kDa LPs genes have a common ancestor which underwent differentiation and divergent evolution. Additionally, expressed sequence tags (ESTs) derived from different silkworm tissues were also analyzed which revealed that the level of expression of the 30-kDa LPs varies. The highest number of ESTs, indicating a high expression level of the protein, was found for *Bombyx mori* lipoprotein 3 (*Bmlp3*) and 7 (*Bmlp7*), with values of 607 and 864, respectively (Sun *et al.*, 2007). It is noteworthy that the amino acid sequence alignment of 30-kDa LPs indicates a high level of homology (Table 3.2).

A more comprehensive genomic analysis was performed a few years later (Zhang *et al.*, 2012b) and 46 genes of proteins belonging to the 30-kDa lipoprotein family were found in the silkworm genome. They were denoted *Bmlp1-46* according to the previous terminology

3. Introduction

rules (Sun *et al.*, 2007). The 30-kDa LPs can be divided into three subfamilies (Fig. 3.4): typical 30-kDa LPs, serine/threonine-rich LPs and ENF peptide-binding proteins (Zhang *et al.*, 2012b). The name of ENF is the consensus N-terminal amino acid sequence of the family peptides belonging to multifunctional cytokines (Kamimura *et al.*, 2001). A comparison of the gene structure of the three subfamilies indicated that the C-terminal regions are common, but the N-terminal regions are highly variable. The *Bmlp1-24* genes encode typical 30-kDa LPs, which are LPs with the N-terminal region composed of the signal peptide and an unconserved region. The second group of genes, *Bmlp25-36*, have an additional domain rich in serine and threonine residues located after the signal peptide. The remaining genes, *Bmlp37-46*, encode proteins of molecular weight of 50 kDa, there is no signal peptide in their structure, but instead they have a so-called *Pseudomonas putida* homologous domain in the N-termini (Zhang *et al.*, 2012b).

Table 3.2 Sequence homology among silkworm 30-kDa LPs.

	PBMHP-12	PBMHPC-19	PBMHPC-21	PBMHPC-23	Bmlp3	Bmlp7
PBMHP-6	66 (40)	67 (45)	65 (44)	67 (42)	66 (45)	67 (45)
PBMHP-12		64 (50)	91 (79)	98 (93)	64 (49)	66 (49)
PBMHPC-19			64 (50)	68 (48)	99 (98)	97 (94)
PBMHPC-21				83 (71)	60 (44)	67 (52)
PBMHPC-23					65 (49)	67 (49)
Bmlp3						97 (94)

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

The genomic studies were followed by the analysis of tissue expression profiles at different developmental stages. Interestingly, it turned out that there are some exceptions and not all 30-kDa LPs are fifth instar-specific (Zhang *et al.*, 2012b). For instance, Bmlp20, a *Bombyx mori* larval serum protein described earlier, was detectable in the silkworm hemolymph from the first to the fourth instar (Fujiwara and Yamashita 1992) with an accumulation in the fourth instar (Zhang *et al.*, 2012b).

The mulberry silkworm was the first lepidopteran in whose body the 30-kDa LPs were discovered (Gamo, 1978). Only microvitellogenin isolated from *Manduca sexta* was reported to be homologous to 30-kDa proteins (Wang *et al.*, 1989) until 2012, when 27 genes of the 30-kDa LPs were found in 11 lepidopteran species, not including *Bombyx mori*.

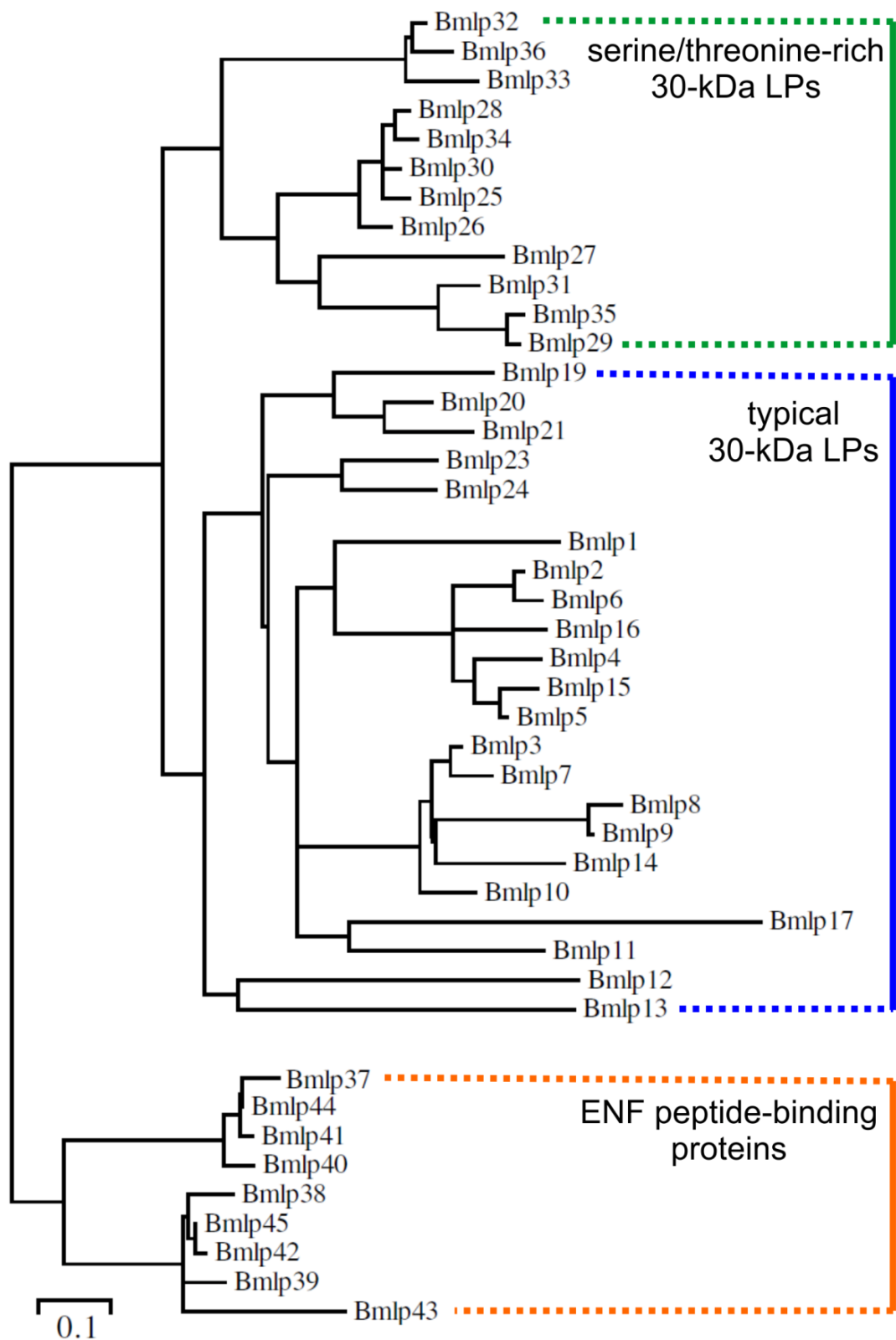


Fig. 3.4 Phylogenetic tree of 30-kDa lipoprotein family.

The phylogenetic tree of *Bombyx mori* 30-kDa LPs shows the relations between three subfamilies, adapted from Zhang et al., 2012b.

The nucleotide sequences of typical 30-kDa LPs were discovered in *Samia cynthia ricini*, *Antheraea assama*, *Antheraea mylitta* and *Manduca sexta*. The C-terminal sequence of *B. mori* paralytic peptide binding proteins and of growth-blocking peptide binding protein of *Pseudaletia separata* were also found to be homologous to 30-kDa LPs (Zhang *et al.*, 2012b). Nevertheless, 30-kDa LPs seem to be specific to the *Lepidoptera* order.

3.2.3. Anti-apoptotic properties of hemolymph proteins

Apoptosis is the process of programmed cell death under genetic control and accompanied by characteristic cell changes, such as the convolution of plasma membranes, nuclear and cell segmentation, etc. (Walker *et al.*, 1988; Cotter *et al.*, 1992). Although this process involves cellular self-destruction, it is essential for the development and homeostasis of all multicellular organisms, including humans. Alterations in cell death lead to a number of human diseases. Apoptosis inhibition causes an excessive accumulation of cells, which can result in cancer, autoimmune or inflammatory diseases. On the other hand, hyperactive apoptosis is also dangerous and leads to neurodegenerative or haematologic diseases (Ramirez-Chamond *et al.*, 1999). Thus the regulation of the apoptosis process has been extensively investigated in recent years.

Although apoptosis is an essential process in living organisms, it becomes a problem when it comes to maintaining high viable cell concentration in insect and mammalian cell cultures (Goswami *et al.*, 1999) used in laboratories for the production of recombinant eukaryotic proteins (Yoon *et al.*, 2006). Moreover, some of the additives which increase protein biosynthesis at the same time induce apoptosis. An example of such a compound is sodium butyrate, which is usually added to Chinese hamster ovary (CHO) cell cultures producing recombinant erythropoietin (Chang *et al.*, 1999; Kim and Lee, 2001). Silkworm hemolymph was demonstrated to possess beneficial properties for the solution of this problem. The addition of hemolymph to cell cultures inhibits apoptosis and improves viability of the cells. To this date, anti-apoptotic properties of hemolymph were proven for insect, Sf9 (Rhee *et al.*, 1999; 2002; Rhee and Park, 2000; 2001), mammalian, CHO (Kim *et al.*, 2004; Choi *et al.*, 2005) and human cell lines, HeLa (Choi *et al.*, 2002) and HEK293 (Kim *et al.*, 2004). Additionally, apoptosis inhibition by hemolymph resulted in a five-fold increase of erythropoietin production in the CHO cell culture (Choi *et al.*, 2005).

During the mentioned experiments apoptosis was induced by viruses (Rhee *et al.*, 1999; 2002; Rhee and Park, 2000; 2001; Choi *et al.*, 2002) or chemicals such as sodium

butyrate (Choi *et al.*, 2005), actinomycine D, camptothecin (Rhee *et al.*, 2002) and staurosporine (Kim *et al.*, 2004). In all cases silkworm hemolymph overcame the apoptosis problem. Regardless of the type of apoptosis inducer, cell death was inhibited in cell cultures. Hemolymph did not interact with the inducer, its inhibitory effect was manifested by an decrease of cell detachment from an adhering surface (Choi *et al.*, 2002). More detailed studies revealed that hemolymph also prevented the release of cytochrome *c* from mitochondria and inhibited the activation of caspase-3. The integrity of mitochondrial membrane was maintained and the ATP generation was not interrupted (Choi *et al.*, 2005).

In 2001, the protein responsible for anti-apoptotic properties of hemolymph was isolated and the first 24 residues of its N-terminus were identified. It was established that the protein belonged to the 30-kDa LP family; however, due to previously mentioned high level of amino acid sequence homology between the 30-kDa LPs, the final assignment of the proper sequence was not possible (Kim *et al.*, 2001). The researchers chose 30Kc6 (mature PBMHP-6) for further analysis. The recombinant 30Kc6 was expressed in CHO and HEK293 cells. Its apoptosis inhibitory effect was comparable to hemolymph what indicated that both intracellular expression and external supplementation inhibited apoptosis efficiently (Kim *et al.*, 2004).

Later on, another apoptosis inhibitor was isolated from the high molecular weight fraction of hemolymph. The protein was identified as SP2 and found to inhibit the programmed cell death induced by staurosporine in HeLa cells. The detailed studies revealed that SP2 suppressed nuclear fragmentation, apoptotic body formation and the inhibition of the reactive oxygen species generation (Rhee *et al.*, 2007).

In conclusion, silkworm hemolymph or its components could be efficiently used for apoptosis inhibition in commercial cell cultures and improve their productivity (Choi *et al.*, 2002; Kim *et al.*, 2004). Silkworm hemolymph could serve as a medium supplement and replace commonly used fetal bovine serum (Ha *et al.*, 1996). The other important factor is that hemolymph does not interact with viruses (Choi *et al.*, 2002) which are often used as infective expression vectors for the production of recombinant proteins in cell cultures (Moss and Flexner, 1987). Finally, particular hemolymph proteins could be used in the future to treat diseases related to hyperactive apoptosis (Kim *et al.*, 2004).

3.2.4. Cell-penetrating properties of hemolymph proteins

Cell-penetrating proteins are a group of proteins able to penetrate into various types of living cells via a receptor-independent endocytosis. They could be used for the delivery of biologically active proteins, DNA and other compounds into cell cultures and animal models *in vivo* (Schwarze *et al.*, 2000; Wadia and Dowdy, 2002). The HIV-1 trans-activator of transcription (TAT) protein is the first reported representative of this protein class (Frankel and Pabo, 1988; Green and Loewenstein, 1988). The mechanism of the HIV-TAT cell-penetration was described as a receptor-independent endocytosis, known as macropinocytosis. Briefly, the process is initiated by an interaction between positively charged HIV-TAT amino acid side chains and heparan sulfate present in the cell membrane. The plasma membrane invaginates and the penetrating protein is internalized in the cell (Frankel and Pabo, 1988; Kaplan *et al.*, 2005).

The first cell-penetrating protein found in insects is a member of the 30-kDa LP family, 30Kc19 (mature PBMHPC-19). After the addition of the recombinant 30Kc19 into the mammalian cell cultures the protein penetrated into various types of cells and localized mainly in mitochondria and the cytoplasm. The 30Kc19 protein was also able to deliver into cells a cargo protein, green fluorescence protein, which does not possess cell-penetrating properties. What is more, 30Kc19 exhibited cell-penetrating properties and could deliver a cargo also *in vivo*. After injection of 30Kc19 into mice, the protein penetrated into different cell types. It was found in heart, lung, kidney and liver cells. Therefore, 30Kc19 and probably also other members of the 30-kDa LP family are potential medicinal tools for cargo molecule delivery into target tissues (Park *et al.*, 2012).

The main goal of this work is the determination and analysis of crystal structures of major hemolymph proteins isolated from the mulberry silkworm. Six crystal structures of four hemolymph proteins were solved and characterized within this project, three of them are structures of Bmlp7, two of Bmlp3 and one is a complex of two high molecular weight proteins, SP2 and SP3. In view of the limited structural information about silkworm proteins, these studies contribute to the overall knowledge about hemolymph proteins and are important for elucidation of their structure-function relationship.

The principal research method within this work was X-ray crystallography. However, the experimental work could be divided into four parts reflecting different research tasks:

1. The isolation of hemolymph from the fifth instar larvae of mulberry silkworm and preparation of protein samples which were sufficiently pure for crystallization.
2. Protein crystallization, X-ray diffraction data collection of the obtained crystals and determination of protein crystal structures using computational methods.
3. Establishing of the unknown proteins identity according to biochemical analysis and electron density maps obtained from the diffraction experiments.
4. Detailed analysis of the obtained protein structures including protein-protein interactions, structural comparison with other structures from the PDB, description of metal binding sites and glycosylation sites, search for putative ligand-binding cavities, and computation of protein surface charge, all taken together, allowed to draw conclusions about the protein function.

5.1. HEMOLYMPH COLLECTION

The fifth instar larvae of the mulberry silkworm, *Bombyx mori* L., were obtained thanks to the courtesy of Dr. Małgorzata Łochyńska from the Institute of Natural Fibres and Medicinal Plants in Poznań.

Silkworm hemolymph was collected from three batches of obtained larvae: in 2008, 2010 and 2012. The procedure of hemolymph collection (Fig. 5.1) involved cutting off two abdominal legs of each larva. The larvae were not separated according to sex. Routinely, 0.3-0.4 ml of hemolymph per insect was collected and hemolymph was stored in 2 ml aliquots at 193 K.



Fig. 5.1 Hemolymph collection.

Photos were made in 2010 and present the hemolymph collection procedure: **a)** a batch of fifth instar larvae (320 insects) was brought to the laboratory, **b)** two abdominal legs of each larva were cut off with nail scissors **c)** and hemolymph was collected to Eppendorf tubes. The larvae were immediately frozen in liquid nitrogen.

5.2. PROTEIN PREPARATION

The pooled hemolymph samples were thawed at 277 K with 0.025 mM 1-phenyl-2-thiourea; no protease inhibitors were added. The samples were centrifuged at 18 000 g for 30 min at 277 K.

Purification protocols developed for particular proteins consisted of several different types of column chromatography. All separations were performed using Äkta FPLC (Amersham Biosciences).

5.2.1. Gel filtration

The first step of the hemolymph proteins' separation was gel filtration (GF) carried out using a Superdex 200 pg column (XK 16/100, Amersham Biosciences) equilibrated

with 100 mM NaCl, 10 mM Tris pH 7.3 and 0.025 mM 1-phenyl-2-thiourea. Two main fractions were collected, one containing proteins of high molecular weight (~500 kDa) and the other comprising of 30-kDa proteins. Each fraction was concentrated to 1.0 ml using Vivaspin® 15R centrifugal concentrators.

5.2.2. Ion exchange chromatography

The concentrated 500-kDa fraction was applied onto a Q Sepharose column (XK 16/10, Amersham Biosciences) equilibrated with 30 mM NaCl, 5 mM Tris pH 7.3. Stepwise elution was the method used for protein separation in ion exchange chromatography (IEC). The fraction containing three different storage proteins (SPs), including the SP2-SP3 complex, was eluted with 450 mM NaCl. The collected peak fractions were concentrated to 100 mg/ml using Vivaspin® 6R centrifugal concentrators. The buffer exchange, with 10 mM Tris pH 7.3, was performed during concentration. The same concentration method was used for all fractions described below.

The concentrated fraction containing 30-kDa proteins was also applied onto a Q Sepharose column (XK 16/10, Amersham Biosciences) equilibrated with 30 mM NaCl, 5 mM Tris pH 7.3. In this case, stepwise elution was also chosen for protein separation, but lower NaCl concentration was used to elute the 30-kDa proteins: 60 mM and 90 mM NaCl, for the fractions containing Bmlp7 and Bmlp3, respectively. The collected peak fractions of Bmlp7 were concentrated to 10 mg/ml and the final volume of sample containing Bmlp3 was 0.2 ml.

5.2.3. Hydrophobic interaction chromatography

The sample containing Bmlp3 was prepared for hydrophobic interaction chromatography (HIC) by mixing it with 0.55 ml of 10 mM Tris pH 7.3 and 0.25 ml of 4 M ammonium sulfate. This step was carried out on a HiTrap Phenyl Sepharose™ HP column (Amersham Biosciences) equilibrated with 1.0 M ammonium sulfate, 10 mM Tris pH 7.3. A linear elution gradient was applied and Bmlp3 was eluted with 0.8 M ammonium sulfate. The collected fractions were concentrated to 0.5 ml.

5.2.4. Protein desalting

The sample of Bmlp3 was desalted after HIC due to a high concentration of ammonium sulfate. Desalting was performed on a HiTrap Desalting column (Amersham

Biosciences) equilibrated with 10 mM Tris pH 7.3. The collected fractions of Bmlp3 were concentrated to 10 mg/ml.

5.2.5. Protein sample characterization

Routinely, the quantity and quality assessment of the produced protein preparations was done as described briefly below, using:

➤ The absorption measurements at 280 nm

The concentration of the purified proteins was determined by measuring the absorption at 280 nm (using the theoretical extinction coefficients; Table 5.1).

➤ Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

All purification steps were monitored by SDS-PAGE performed in 12.5% polyacrylamide resolving gels (Laemli, 1970). The gels were visualized with Coomassie dye.

➤ Dynamic light scattering (DLS)

DLS measurements providing information about the protein size and monodispersity were carried out on a Zetasizer Nano (Malvern Instruments Ltd.) at 291 K using a scattering angle of 90° and a laser working at 830 nm. A ZMV1002 quartz cuvette was used, with 2 µl of concentrated protein sample, for each measurement. All samples were centrifuged at 18 000 g for 30 min at 291 K prior to the measurements.

Table 5.1 Biophysical parameters calculated for the studied proteins.

Protein	Molecular mass [kDa]	pI	ϵ_{280} [M ⁻¹ cm ⁻¹]
Bmlp3	27.57	6.14	62340
Bmlp7	27.53	7.11	55350
SP2	81.88	5.63	117930
SP3	81.58	5.81	131590

Presented parameters were calculated using ProtParam (<http://web.expasy.org/protparam/>).

At the very beginning of this work, all proteins were isolated from hemolymph as unknown proteins. Several analyses were necessary to enable the identification of proper amino acid sequences of the proteins. The mentioned analyses are the only investigations

in which the author of this dissertation was not directly involved. The measurements were outsourced and performed in different laboratories, as listed below:

➤ N-terminal sequencing analysis

The amino acid sequence of the first ten N-terminal residues of the investigated proteins was determined by Edman degradation in BioCentrum (Krakow, Poland).

➤ LC-MS/MS

The initial identification of Bmlp7 and SP2 with LC-MS/MS was performed in The Central Laboratory of the Medical University (Lodz, Poland) and at EMBL (Heidelberg, Germany), respectively.

➤ MALDI-TOF MS

The determination of the proteins' molecular weight was carried out in The Centre of Molecular and Macromolecular Studies, Polish Academy of Sciences (Lodz, Poland) and at EMBL (Hamburg, Germany).

➤ NMR measurements

This analysis was not directly connected with protein identification; however, it is listed in this section, because these measurements were also outsourced. The ^{113}Cd NMR spectra were recorded and analyzed in The Centre of Molecular and Macromolecular Studies, Polish Academy of Sciences (Lodz, Poland) in order to establish the presence of cadmium in Bmlp7 samples.

5.3. PROTEIN CRYSTALLIZATION

All crystals were grown using the vapor diffusion method, which can be described as controlled protein precipitation from an aqueous solution. A crystallization drop contains the protein of interest and a small amount of reservoir solution consisting of a precipitant, a buffer and additives. The drop is equilibrated against a reservoir containing the same precipitant at a higher concentration in comparison to the drop. Vapor diffusion in this closed system leads to a net transfer of water from the drop to the reservoir. The system comes to equilibrium when the precipitant concentration is the same in both solutions. The decrease of the amount of water in the drop causes an increase of the protein concentration. All these factors together favor crystal growth (Rhodes, 2000).

The great majority of the crystallization experiments were set manually applying the hanging drop technique. The proteins were crystallized in 24-well hanging drop plates (Hampton Research), the volume of the reservoir solution in each well was 0.5 ml

and the crystallization drops were made up of equal volumes of the concentrated protein solution and the reservoir solution, usually 1.0 or 1.5 μ l.

Optimization of the crystallization conditions of Bmlp3 and the initial screening of crystallization conditions for SP2-SP3 were carried out using the High Throughput Crystallization (HTX) Facility at the EMBL (Hamburg, Germany). Crystallization experiments were set up automatically using a nanodispenser and 96-well sitting drop plates.

The initial crystallization conditions were found using commercially available screening solutions (PEG/Ion, PEG/Ion2, Index and Additive Screen, Hampton Research). The solution compositions for optimization of the crystallization conditions were developed individually for each protein. All crystallization trials were performed at 291 K.

5.4. PROTEIN CRYSTALS DERIVATIZATION

At the beginning of my work on the hemolymph proteins, there were no crystal structures of 30-kDa lipoproteins available in the PDB. Thus the crystal structure of Bmlp7 could not be determined using molecular replacement. Other methods, based on heavy atom derivatization, were chosen for the phase problem solution.

The use of these methods requires an incorporation of an anomalous scatterer (see Section 5.7.1.) into the protein crystal. Heavy atoms were introduced into protein crystals for the first time by Max Perutz and John Kendrew (Perutz, 1956; Kendrew *et al.*, 1958). Since then, this technique has been widely used for the determination of protein crystal structures. It involves the protein crystals being soaked in solutions of heavy ions, containing for instance Hg, Pt, or Au. Alternatively, a protein could be crystallized with heavy atom additives (Rhodes, 2000). Recently, the use of $\text{Ta}_6\text{Br}_{12}^{2+}$ cluster for soaking is often used. The cluster of tantalum and bromine is a powerful derivatization agent due to its compact, almost spherical shape and the presence of six tantalum and twelve bromine atoms which display a significant anomalous diffraction signal at their absorption edges (Banumathi *et al.*, 2003). Nowadays “dauterization”, a soaking of protein crystals with halides, has also become a common method; its great advantage being that it does not require work with toxic heavy ions (Dauter *et al.*, 2000). Finally, an alternative method to heavy atom derivatization is the incorporation of selenomethionine residues into the protein (Hendrickson *et al.*, 1990). However, the introduction of modified methionine residues is possible only during protein expression, thus this method cannot be used for natural source proteins.

The Bmlp7 crystals were soaked with agents which could provide anomalous scatterers. All of the mentioned methods, excluding selenomethionine incorporation, were tested in the case of Bmlp7. Potassium bromide and iodide were added at the final concentration of 0.8 M to the drop containing crystals and cryoprotectant solution. The soaking time was 1.0 min. The solid $\text{Ta}_6\text{Br}_{12}^{2+}$ cluster was added with a micro spatula to the crystallization drop containing Bmlp7 crystals. Finally, soaking with the heavy ions solution (Heavy Atom Screen, Jena BioScience) was carried out. Twenty different solutions, containing Pt, Hg, Au, Pb, Ir or Os, were tested. The final concentration of heavy ions in the crystallization drop was 2.0 mM and the soaking time was seven days.

5.5. X-RAY DIFFRACTION DATA COLLECTION

X-ray diffraction data for the obtained crystals were collected at the European synchrotrons, on beamlines BL14.1, BL14.2 and BL14.3 of the BESSY synchrotron (Berlin, Germany; Mueller *et al.*, 2012) and on beamlines X11, X12 and X13 of the DESY synchrotron (Hamburg, Germany). The detectors on all beamlines were charge-coupled devices (CCDs). Data were recorded from single crystals mounted on a goniometer what allowed automated rotation of the crystal during data collection. All crystals were cryocooled in a stream of nitrogen gas and the measurements were carried out at a temperature of 100 K.

The low temperature during measurements increases the crystal stability and molecular order in the crystal. On the other hand, water present in the protein crystal becomes ice and damages the crystal. Thus, agents called cryoprotectants are used to prevent ice formation (Rhodes, 2000). Crystals, which were obtained using precipitants with ice-preventing properties, for instance PEG MME 550, were fished out from their crystallization drops and cryocooled. When PEG 3350 was the main precipitant, a solution containing PEG 400 was mixed with the reservoir solution at a 1:1 ratio and the crystals were transferred into the cryoprotectant solution for a few seconds. After that they were cryocooled in a stream of nitrogen gas. A solution of Tacsimate, a mixture of organic acid salts, was chosen as the cryoprotectant for crystals obtained in the presence of ammonium sulfate.

At the beginning of each measurement several images were collected and X-ray diffraction data were analyzed. According to this analysis the experiment parameters (angle of rotation, oscillation range, crystal-to-detector distance and exposure time) were established.

5.6. X-RAY DIFFRACTION DATA PROCESSING

The collected images were processed using *HKL-2000* (Otwinowski and Minor, 1997) or *XDS* and *XSCALE* (Kabsch, 2010a,b). Regardless of the chosen program, several important steps of data processing can always be distinguished:

➤ Indexing of the reflections

At the beginning of data processing, reflections are indexed according to corresponding positions in the reciprocal space and in the image(s). Afterwards the unit cell parameters, symmetry of the crystal and its space group can be identified. Indexing is followed by initial refinement of the established unit cell parameters, crystal-to-detector distance and mosaicity.

➤ Integration of the diffraction peaks

The reflections present at all collected diffraction images are converted into a set of Miller indices and corresponding intensities, which are evaluated according to a spot profile and the pixel density.

➤ Merging of symmetry-related reflections

Partially recorded reflections are summed and merged together, the scale factor between images is calculated and the intensity of redundant measurements is averaged.

➤ Assessment of X-ray diffraction data quality

One of the most important and commonly used quality indicators is the merging R-factor (R_{merge}), which is the sum of the differences of all measurements from the average value of the measurement divided by the sum of all measurements. Other important parameters are the completeness of data and the average signal-to-noise ratio, $I/\sigma(I)$, which is a criterion for the resolution cutoff.

The intensities obtained after data processing are the first key compound essential for structure determination.

5.7. STRUCTURE DETERMINATION

A contour map of the electron density $\rho(xyz)$ throughout the crystal unit cell tells us about the arrangement of the atoms in the protein molecule. The electron density is a periodic function which can be represented by a Fourier series and connected to the diffraction pattern by the equation:

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l F(hkl) e^{-2\pi i(hx+ky+lz)} \quad (5.1)$$

where V - unit cell volume, hkl - the indices of reflection, $F(hkl)$ - the structure factor describing the reflection. The structure factor (Fig. 5.2) can be represented by the equation:

$$F(hkl) = |F(hkl)| e^{i\varphi(hkl)} \quad (5.2)$$

The amplitude $|F(hkl)|$ of the structure factor is proportional to $\sqrt{I(hkl)}$, the square root of the intensity $I(hkl)$ of reflection hkl . The reflection intensities are measured during the diffraction experiment, but the diffraction images lack the information about the phase $\varphi(hkl)$ of the reflection hkl . This issue is called the phase problem (Rhodes, 2000).

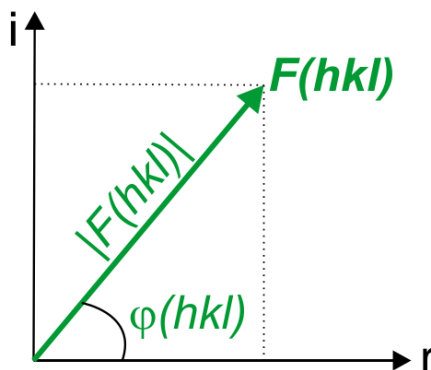


Fig. 5.2 The structure factor.

The structure factor $F(hkl)$ is represented as a vector in the complex plane (r - the real axis, i - the imaginary axis). The amplitude $|F(hkl)|$ is the length of the vector and $\varphi(hkl)$ is the angle between the vector and the real axis. The scheme was prepared according to Rhodes, 2000.

The phases of the reflections can be calculated using methods based on crystal derivatization or molecular replacement (MR). There are several methods belonging to the first group: single isomorphous replacement (SIR), multiple isomorphous replacement (MIR), single-wavelength anomalous dispersion (SAD) and multiple-wavelength anomalous dispersion (MAD). The first crystal structure of Bmlp7 was solved by a combination of SAD

and MR methods. All other structures presented in this dissertation were determined by MR. Thus, only methods based on anomalous dispersion and MR are described in next sections.

5.7.1. Single- and multiple-wavelength anomalous dispersion

The method using anomalous dispersion and synchrotron radiation for the solution of the phase problem was developed by Hendrickson *et al.* (1988). Prior to MAD or SAD experiment, the derivatives of protein crystals containing heavy atoms are prepared as described in Section 5.4. The heavy atoms are capable of absorbing X-rays of specified wavelengths. For instance, at wavelength of the absorption maximum of the particular heavy atom, the energy is sufficient to excite an electron in the strongly absorbing atom to a higher quantum state. This absorption results in a phase change for the X-rays scattered by that atom, relative to the phase of the X-rays scattered by the other atoms of the structure. The Friedel's law does not hold and the intensities of reflections I_{hkl} and I_{-h-k-l} are not equal. Each pair of such reflections is a Bijvoet pair. This phenomenon is called anomalous dispersion and the heavy atom is denoted often as the anomalous scatterer (Rhodes, 2000).

The atomic scattering factor $f_a(hkl)$ of the anomalous scatterer can be represented by the equation:

$$f_a = f^0 + f' + if'' \quad (5.3)$$

where f^0 is the normal scattering factor and $(f' + if'')$ represents the anomalous correction, which is a complex number consisting of a real (f') and an imaginary part (if''). The real part f' is called the dispersive correction and the component of the imaginary part f'' is denoted as the absorptive or Bijvoet correction. The value of f'' is significant only for heavy atoms and when the absorptive correction is taken into consideration, the Friedel's law does not hold. The vector corresponding to f'' is perpendicular to f^0 and f' , but its direction and orientation is different for positive and negative reflections. The phases of two reflections being a Bijvoet pair differ by sign (Fig. 5.3) (Rhodes, 2000; Drenth, 2007).

X-ray diffraction data are usually collected at three different wavelengths during a MAD experiment. The first chosen wavelength is the absorption maximum with the maximum value of f'' (absorption peak), the second one corresponds to the minimum value of f'' (so called inflection point or edge), and the last chosen wavelength (remote) is distant to absorption peak where the values of f' and f'' are small (Fig. 5.4). The data sets collected at the mentioned wavelengths have different values of the real and imaginary parts

of anomalous corrections what is used to solve the phase problem (Rhodes, 2000; Drenth, 2007).

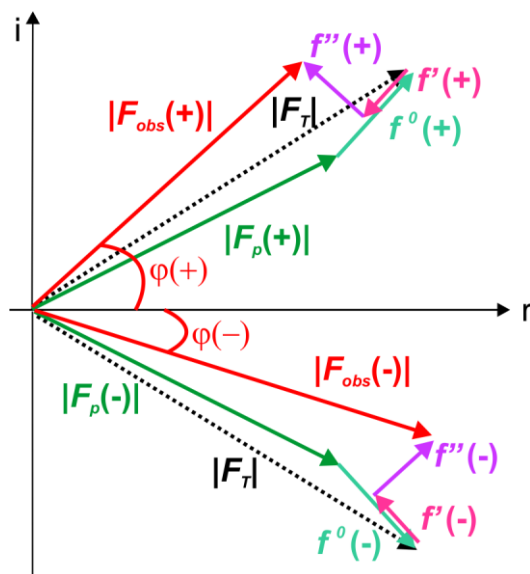


Fig. 5.3 Anomalous scattering.

$|F_{obs}(+)|$ and $|F_{obs}(-)|$ refers to the observed structure factor amplitudes for the reflection hkl , marked (+), and $-h-k-l$, marked (-), respectively. The amplitudes for normal scattering of protein atoms $|F_p|$ and of all the atoms, non-anomalous and anomalously scattering atoms, $|F_T|$ are also shown. The scheme also presents all the components of the atomic scattering factor: f^0 (the normal scattering factor), f' (the dispersive correction) and f'' (the absorptive correction). The diagram was prepared according to Rhodes, 2000.

The information about phases can be extracted in MAD by solving the following equation:

$$\begin{aligned} |^{\lambda}F(\pm hkl)|^2 = & |^{\circ}F_T|^2 + a(\lambda)|^{\circ}F_A|^2 \\ & + b(\lambda)|^{\circ}F_A||^{\circ}F_T|\cos(^{\circ}\varphi_T - ^{\circ}\varphi_A) \\ & + c(\lambda)|^{\circ}F_A||^{\circ}F_T|\sin(^{\circ}\varphi_T - ^{\circ}\varphi_A) \end{aligned} \quad (5.4)$$

where $a(\lambda) = [(f')^2 + (f'')^2]/(f^0)^2$; $b(\lambda) = 2f'/f^0$; $c(\lambda) = 2f''/f^0$; $|^{\lambda}F(\pm hkl)|$ - the observed structure factor amplitude of reflection $\pm hkl$; $|^{\circ}F_A|$ - the structure factor amplitude of the normal part of the anomalous scatterer; $|^{\circ}F_T|$ - the structure factor amplitude of the total structure (including all atoms); $(^{\circ}\varphi_T - ^{\circ}\varphi_A)$ - the phase difference; the superscript 0 refers to the wavelength-independent values which can be obtained by least-square fitting to the observational equation. Using $|^{\circ}F_A|$ and Patterson function the positions of heavy atoms can be determined and $^{\circ}\varphi_A$ can be calculated, which also enables the calculation of $^{\circ}\varphi_T$. Then electron density maps are obtained from normal scattering of all atoms (McRae, 1999).

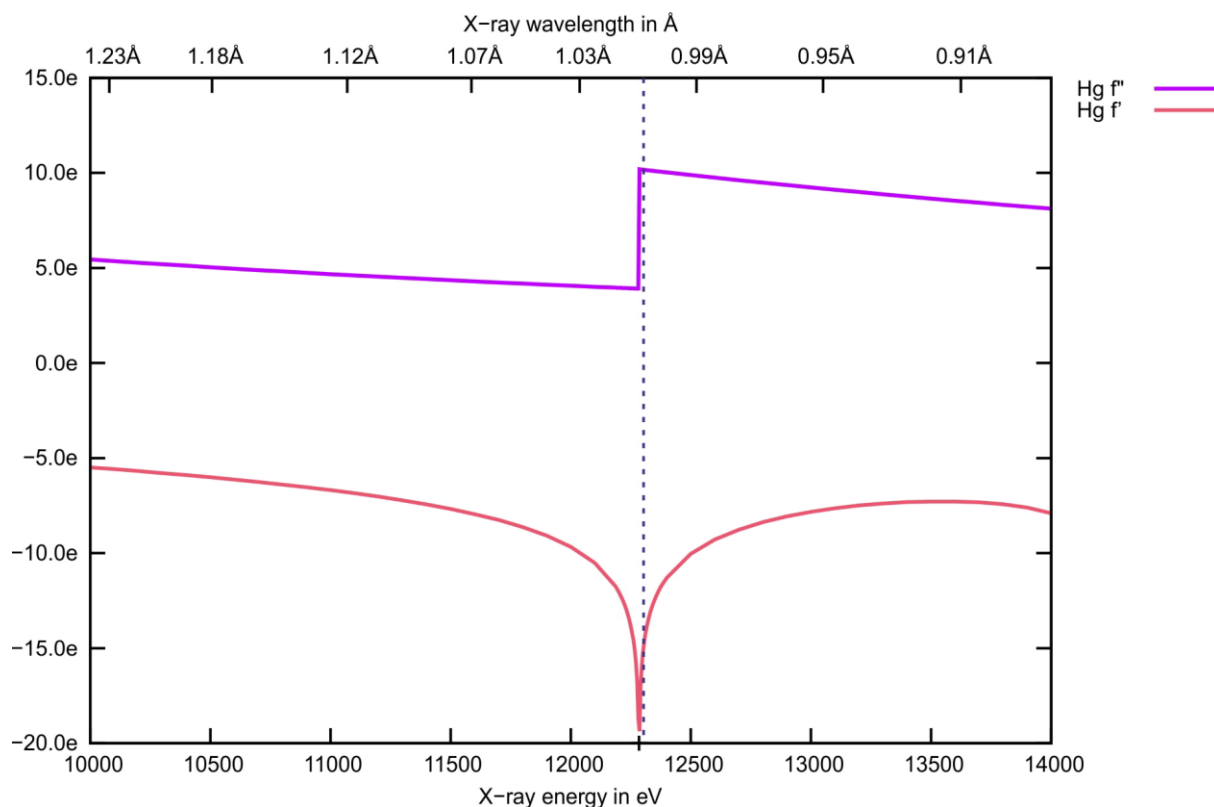


Fig. 5.4 The absorption edges of mercury.

The scattering factors f' and f'' of mercury are plotted as a function of energy. The plot was prepared using an online tool: <http://skuld.bmsc.washington.edu/scatter/> The blue dotted line indicates the absorption peak.

The main disadvantages of MAD are that the data have to be collected at three wavelengths, the exposure time is long and the crystals often suffer from radiation damage. Therefore, nowadays SAD is the method of choice. The anomalous X-ray diffraction data are collected at only one wavelength, usually at the wavelength corresponding to absorption peak. The experiment is successful when the crystal contains anomalous scatterer which provides a strong anomalous signal. It is of note that in SAD two solutions of the phase problem are possible. Thus the obtained electron density map contains the correct structure, but also noise. The density modification improves the map significantly producing the proper solution (Drenth, 2007).

The initial model of Bmlp7 was obtained using SAD.

5.7.2. Molecular replacement

Two main components are required to solve the structure by MR: a single native X-ray diffraction dataset and the protein structure which could be used as an initial model for phase

calculations. The initial model is usually a structure of a protein of which the amino acid sequence is similar in at least 30% with the amino acid sequence of the investigated protein. The higher the sequence homology, the more likely it is that the proteins will share a similar fold. In the MR method the initial model is fitted into the unit cell of the investigated protein. The known phases of the model are used to estimate the phases of the unknown structure. Two main steps can be distinguished in the MR procedure, the rotation and the translation search. The rotation search results in a proper orientation of the model, whereas the translation search provides information about the precise position of the model in the unit cell (McRee, 1999; Rhodes, 2000; Drenth, 2007).

The base of the MR calculations is the Patterson function which enables the calculation of a Patterson map (a vector map) using the equation:

$$P(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F|^2 \cos 2\pi(hx + ky + lz) \quad (5.5)$$

The experimental Patterson map (P_o) can be obtained using the measured intensities of the investigated protein for calculations. Subsequently, a set of interatomic vectors (P_m) is generated from the initial model. It is of note that the Patterson vectors can be divided into two groups: interatomic vectors (self-Patterson vectors) and intermolecular vectors (cross-Patterson vectors). The first group describes the pairs of atoms belonging to the same molecule and provides information about the molecule's orientation, whereas the intermolecular vectors supply the information about the correct position of the molecule in the unit cell (Drenth, 2007).

The rotation function could be represented by the equation:

$$R(C) = \int_U P_o(u) \cdot P_m(Cu) du \quad (5.6)$$

where P_o and P_m are Patterson maps of the unknown structure and the model, respectively. C is the 3D rotation matrix that relates the rotated and the stationary structure, u are the Patterson vectors. The $R(C)$ depends on the rotation angles and has a maximum value for correct overlap of P_o and P_m . When the model consists of a single molecule (the first step of MR calculations), the generated vectors are interatomic vectors (Drenth, 2007).

The translation function can be represented by the equation:

$$T(t) = \int_U P_m(u, t) \times P_o(u) du \quad (5.7)$$

where t represents the intermolecular Patterson vectors. $T(t)$ has a maximum when the model is in the best agreement with the unknown structure data. The cross-Patterson vectors are

in agreement with the experimental Patterson map when the molecules are in proper positions in respect to the symmetry elements (Drenth, 2007).

In more difficult cases of MR, it is possible to improve the search model. It can be accomplished by removal of the flexible regions, for instance N- or C-terminal residues or loop regions. Another possibility is the adaptation of the model to the sequence of the investigated protein. The terminal atoms of long amino acid side chains can be removed from the model to retain only the part of the side chains which is common for the sequence of the model and the sequence of the investigated protein. Alternatively, all residues can be mutated to alanine residues. Finally, when the protein consists of several domains, the search for a solution can be performed with single domains separately (McRee, 1999).

All the structures presented in this work, except the first Bmlp7 structure, were solved by MR. The high-resolution Bmlp7 structure served as an initial model in MR calculations for the other Bmlp7 and Bmlp3 structures. The structure of SP2-SP3 complex was solved using as an initial model the structure of arylphorin from oak silkworm (PDB code: 3GWJ; Ryu *et al.*, 2009). The MR phasing was carried out using different programs: *MolRep* (Vagin and Teplyakov, 1997), *Phaser-MR* (McCoy *et al.*, 2007) or the *Auto-Rickshaw MR* protocol (Panjikar *et al.*, 2005).

5.8. STRUCTURE REFINEMENT

Refinement of the solved structure is carried out in order to optimize the obtained model agreement with the native reflection intensities. A number of iterative cycles of manual model rebuilding (real space refinement) interspersed with structure factor refinement (reciprocal space refinement) is necessary to complete the model.

5.8.1. Reciprocal space refinement

Reciprocal space refinement can be based on the principle of least squares or maximum likelihood. All investigated structures were refined using maximum likelihood where the parameters of the model are estimated on the basis of observations. Everything that is known about the crystal prior to refinement, including unit cell parameters, structure factor amplitudes, standardized stereochemistry and experimentally obtained phases, is denoted as observations. In this method the investigated model is adjusted to maximize the probability of a given observation (Tronrond, 2004).

The parameters of the model which are refined and written in the PDB file are coordinates of all atoms present in the structure (x,y,z) and their B factors (Tronrond, 2004). The B factor provides information about the vibration of an atom about its central point. This parameter could be refined as isotropic or anisotropic (only for high resolution structures). The isotropic refinement of the B factor is based on the assumption that the atom vibrates equally in all directions and lies within a sphere, whereas the anisotropic refinement takes into account the actual vibration of the atom within an ellipsoid described with six parameters and centered at the atomic coordinates (Stout and Jensen, 1989). For structures at the resolution of 2.0 Å or lower, the anisotropic motion of the atoms can be described by a TLS system which requires fewer parameters compared to an independent anisotropic B factor for each atom. The TLS system applies for a group of atoms which move as a rigid unit. The protein molecule is divided into such groups before the inclusion of TLS into the refinement. The motion of each group of atoms is described by three matrices. The first one concerns a translational (T) vibration of the group, the second applies to libration (L) of the group about a fixed point and the third describes screw-rotation (S) displacements of the group (Winn *et al.*, 2001).

The refinement can also be improved by increasing the number of observations. It can be accomplished by the inclusion of restraints. Bond lengths, bond angles, dihedral angles, chiral volume and planarity of chemical moieties as well as their standard uncertainties, are known from small molecule structures (Allen, 2002). Then during the refinement the deviations from ideal stereochemistry are penalized (Tronrond, 2004). However, the conformational ϕ and ψ angles of the polypeptide backbone are not restrained. They should continually improve during the progress of refinement and they can be used as a quality indicator in structure validation (McRee, 1999).

The investigated structures were refined using the programs: *REFMAC5* (Murshudov *et al.*, 2011) from *CCP4* package (Winn *et al.*, 2011) and *phenix.refine* package in *PHENIX* (Adams *et al.*, 2010).

5.8.2. Real space refinement

Two types of electron density maps, *2Fo-Fc* and *Fo-Fc*, were used during the manual model rebuilding in *Coot* (Emsley and Cowtan, 2004; Emsley *et al.*, 2010). *Fo* refers to observed structure factors, whereas *Fc* describes the structure factors calculated based on the model. The difference electron density *Fo-Fc* map exists in the form of positive

and negative peaks. A positive peak in a region of the map reveals that the contribution of the observed intensities is larger than the contribution of the model at this site. The model should be adjusted to increase the electron density at this position. Reversely, a negative peak implies that the model produces more electron density in this region than the unit cell actually contains and the atoms present in the model should be removed from this site (Rhodes, 2000).

The $2Fo-Fc$ map can be represented as the sum of the Fo map and the difference map:

$$Fo + (Fo - Fc) = 2Fo - Fc \quad (5.8)$$

When the model is properly fitted, the $2Fo-Fc$ map should cover the model (Rhodes, 2000).

The manual model rebuilding enables the correction of the model errors, such as improperly fitted backbone or side chains. Sequence related ambiguities can also be observed. Moreover, the presence of ligands and water molecules can be detected. These molecules can then be incorporated into the model.

5.9. STRUCTURE VALIDATION

The measured and the calculated structure factors should converge, when the model represents the correct structure. The most widely used measure of convergence is the R factor, which can be described by the equation:

$$R = \frac{\sum_{hkl} \|F_o(hkl) - F_c(hkl)\|}{\sum_{hkl} |F_o(hkl)|} \quad (5.9)$$

The R factor indicates how well the current model converges on the entire data set that was used for this model's calculation. The other value which applies to model correctness is R_{free} which is calculated by an equation analogous to 5.9, but only for a set of test reflections which were randomly selected and excluded from the refinement at the beginning. Both R and R_{free} factors should improve after each cycle of refinement (Rhodes, 2000).

A well-refined model should be correct from a chemical point of view. It should not contain inverted centers of chirality. Furthermore, peptide bonds should be nearly planar and *trans*-, the only exception is the occasionally occurring *cis*-proline. The conformational ϕ and ψ angles of the polypeptide backbone should fall in the allowed ranges of the Ramachandran plot. Finally, the torsion angles at single bonds in side chains should lie within a few degrees of stable conformations (Rhodes, 2000). All these things were monitored after each cycle of refinement of the silkworm protein structures in *Coot* (Emsley and Cowtan, 2004; Emsley *et al.*, 2010) using validation tools. The final models were assessed using *PROCHECK* (Laskowski *et al.*, 1993). The validation of the metal binding

sites was additionally carried out using the *CheckMyMetal* server (http://csgid.org/csgid/metal_sites).

5.10. STRUCTURE ANALYSIS

The structural analysis of the obtained crystal structures was performed. Several issues were examined, as listed below:

➤ Primary structure analysis

The alignments of amino acid sequences were calculated in *ClustalW* (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). The alignments were colored using *Jalview* (<http://www.jalview.org/>; Waterhouse *et al.*, 2009).

➤ Structural comparison

A search for structural homologues of the investigated proteins was performed using *PDBeFold* (SSM) server (<http://www.ebi.ac.uk/msd-srv/ssm/>) and *DALI* (Holm and Rosenstrom, 2010; http://ekhidna.biocenter.helsinki.fi/dali_server/).

➤ Protein-protein interactions

The total contact surface area of the neighboring protein molecules was calculated with the *PISA* server (Krissinel and Henrick, 2007) and/or in *Areaimol* from the *CCP4* package (Winn *et al.*, 2011).

➤ Potential binding cavities

The search for potential binding sites was carried out on the *CASTp* (Dundas *et al.*, 2006) and *metaPocket 2.0* (Huang, 2009; Zhang *et al.*, 2011) servers.

➤ Electrostatic surface potential

Poisson-Boltzmann electrostatic potential on the molecular surface of Bmlp3 was calculated using the *APBS* algorithm (Baker *et al.*, 2001) and the *PDB2PQR* program (Dolinsky *et al.*, 2004; 2007). The protonation states of the amino acid side chains were determined at pH of 7.4 using *PropKa* (Li *et al.*, 2005).

The art work, meaning all structural illustrations, were prepared in PyMol from W.L. DeLano, The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA, USA (2002), <http://www.pymol.org>. The great majority of the figures presented in Results & Discussion comes from the publications describing the results of my PhD project. All the mentioned figures were prepared by the author of this dissertation.

6.1. PROTEIN PURIFICATION AND INITIAL CHARACTERIZATION OF THE SAMPLES

The starting material, hemolymph isolated from the fifth instar larvae of mulberry silkworm, contains up to 300 different proteins (Li *et al.*, 2006). The first task of this project was to purify the most abundant proteins to a level sufficient for successful crystallization. The purification protocol of all the investigated hemolymph proteins was described in details in Materials & Methods (5.2). The complete procedure is presented in Fig. 6.1. The first step was the separation of the proteins according to their molecular weight using gel filtration (5.2.1; Fig. 6.2a). Two fractions were collected: high molecular weight (~500 kDa) and low molecular weight fraction (~30 kDa).

In the second step, which involved ion exchange chromatography (5.2.2), the proteins were separated according to their pI. Two separate protocols were developed, one for the high molecular weight fraction (IEC I; Fig. 6.2b) and the other for the low molecular weight fraction (IEC II; Fig. 6.2c). The first procedure allowed me to obtain three different samples, BmA1, BmA2 and SPs (Fig. 6.2b). At the very beginning, the content of the samples was unknown and only the proteins present in the highest peak (SPs) were later identified. The description “BmA1” and “BmA2” was introduced only by the author of this dissertation. The BmA1 and BmA2 samples were highly heterogeneous which precluded crystallization of the proteins present in the samples.

The ion exchange chromatography protocol for low molecular weight proteins also provided three samples (Fig. 6.2c). The main proteins present in the samples were later identified as Bmlp2, Bmlp3 and Bmlp7. The amount and the level of homogeneity of the Bmlp2 sample was lower than these of the Bmlp3 and Bmlp7 samples. At the time when this thesis was in preparation, Bmlp2 was crystallized, however no final results were available and therefore this case will not be discussed in this dissertation. The level of purity for the Bmlp7 sample was estimated to be high, whereas the Bmlp3 sample required further purification, because the initial crystallization trials provided only twinned crystals (for details see the next section). The Bmlp3 sample also contained the juvenile hormone binding protein (BmJHBP), as it was established by the juvenile hormone binding activity assay (Ożyhar and Kochmann, 1987) performed for the sample. Final purification of Bmlp3 was accomplished by hydrophobic interaction chromatography (5.2.3; Fig. 6.2d). The collected

6. Results & Discussion

protein fraction was then desalted (5.2.4) in order to remove the ammonium sulfate used in the previous protocol.

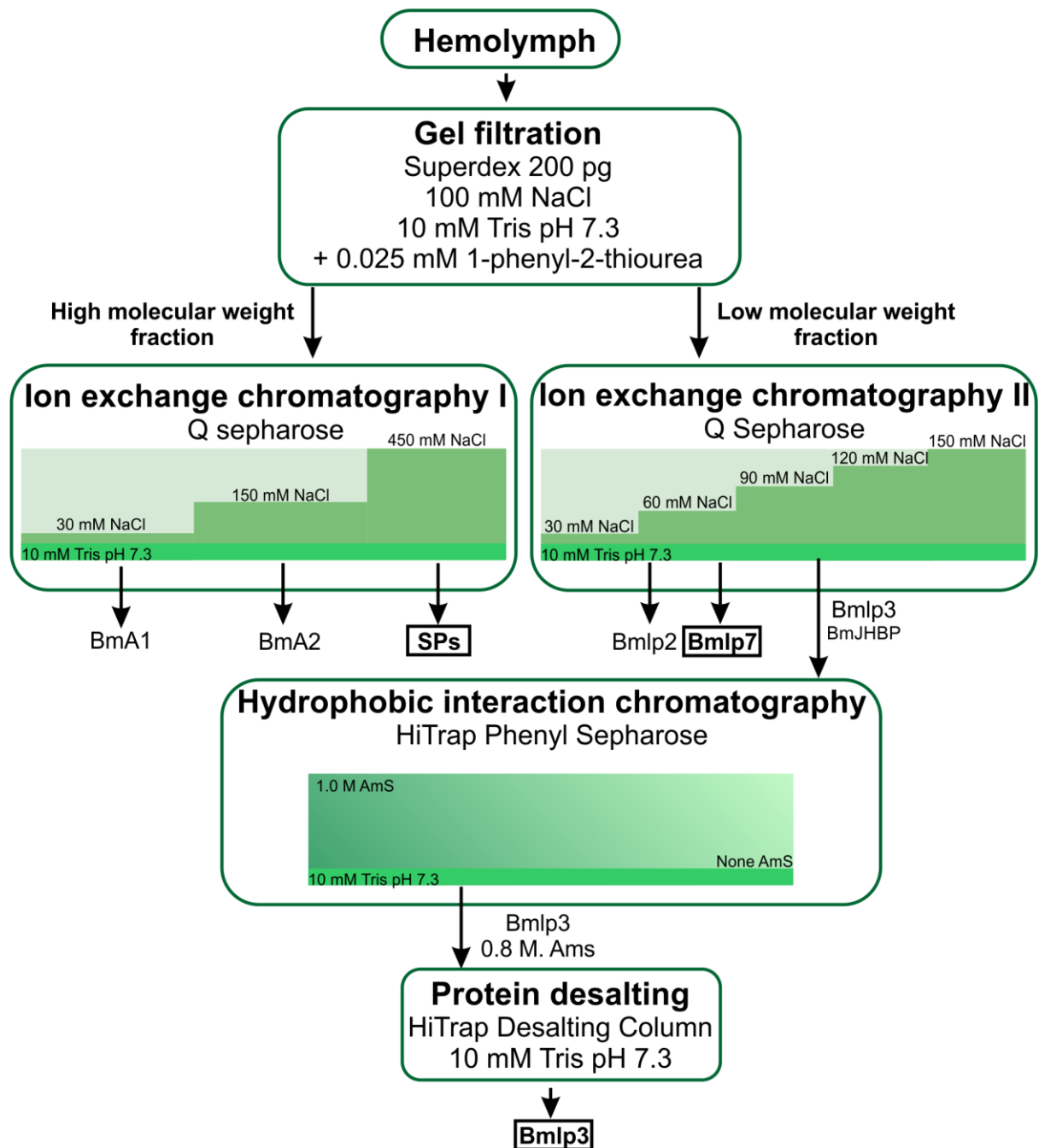


Fig. 6.1 The scheme of purification procedure of silkworm hemolymph proteins.

The hemolymph proteins, SPs and Bmlp7, were purified using a two-step purification protocol, including gel filtration and ion exchange chromatography, whereas two additional stages, hydrophobic interaction chromatography and protein desalting, were required to obtain Bmlp3.

6. Results & Discussion

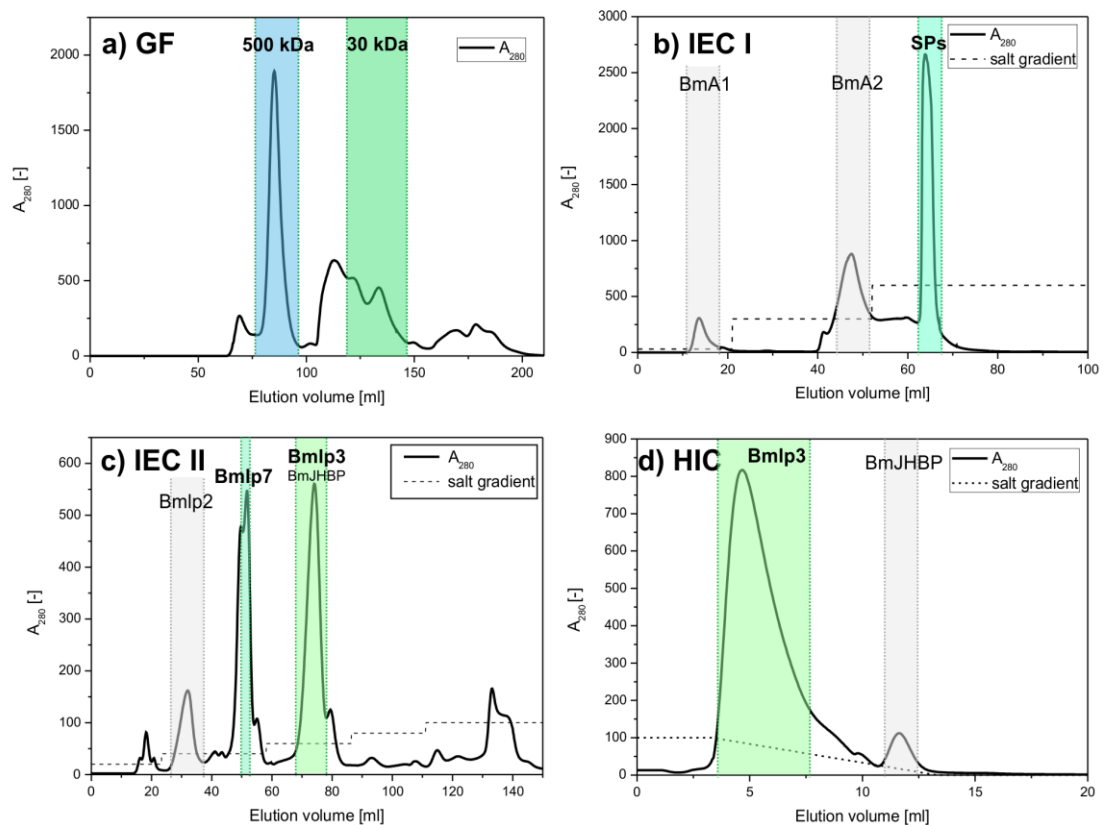


Fig. 6.2 Purification of the silkworm hemolymph proteins.

The chromatograms present elution peaks of all FPLC purification steps: **a)** gel filtration, **b)** ion exchange chromatography of high molecular weight fraction proteins, **c)** ion exchange chromatography of low molecular weight fraction proteins and **d)** hydrophobic interaction chromatography of Bmlp3 sample contaminated with BmJHBP.

The complete purification procedure provided three samples (SPs, Bmlp7, Bmlp3) which were used for structural studies. The yield of the purification for SPs, Bmlp7 and Bmlp3 from 3 ml hemolymph was approximately 40 mg, 2 mg and 0.3 mg of the respective protein. It is of note that some fluctuations of the protein amount and purity were observed when the procedure was repeated a number of times. Mulberry silkworm larvae are not a standardized material, they differ in size and body weight which is reflected in the amount of proteins in the hemolymph and affects the results of the purification. Additionally, the amount of some proteins is related to the sex of the larvae. It is not possible to distinguish female and male larvae in the fifth instar.

The purity of the samples was judged using SDS-PAGE electrophoresis (Fig. 6.3). Each of the purification steps improved the samples' purity. Although some additional bands were visible in the lane corresponding to the final Bmlp7 sample, they did not affect

6. Results & Discussion

the crystallization process. The case of SPs was more complicated. SDS-PAGE electrophoresis revealed the presence of at least two proteins. The molecular weight of these proteins, estimated according to the location of the bands in the gel, was approximately 80 kDa. It revealed that proteins of the high molecular weight fraction (~500 kDa) are composed of 80 kDa monomers. The amount of the protein present in the lower band (smaller molecular weight) was significantly higher. However, the presence of the other protein (“contamination”) could not be ignored. Moreover, the content of this protein in the sample was variable, depending on the composition of the initial hemolymph sample used for purification. In this case, crystallization was the final purification step, as will be explained in the next section.

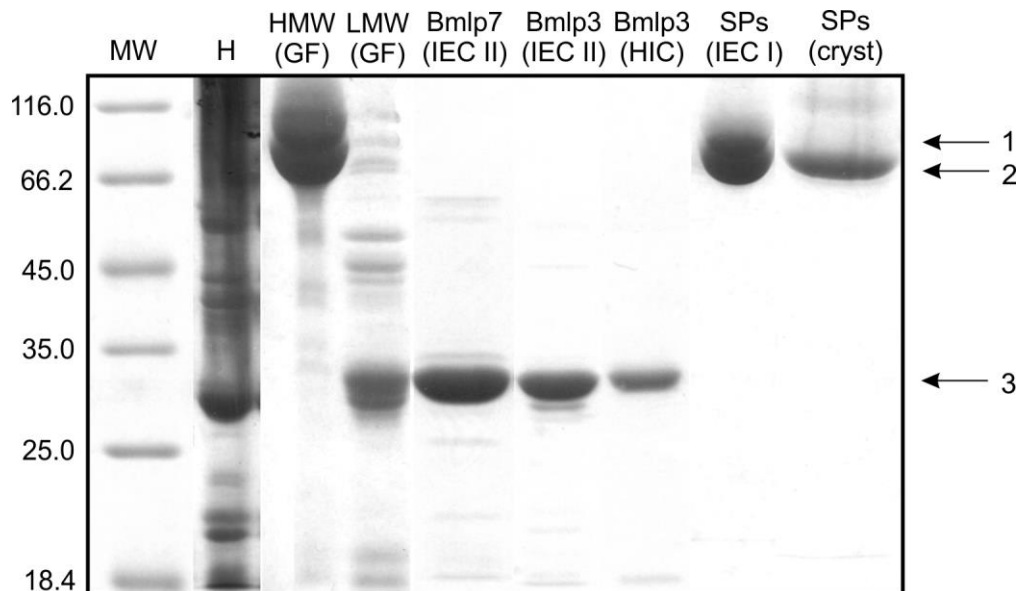


Fig. 6.3 SDS-PAGE electrophoregrams monitoring purification steps.

The results of SDS-PAGE analysis are presented for each purification step. The content of particular lanes is as follows: MW - Protein Molecular Weight Marker from Fermentas, H - hemolymph sample prior to the injection to GF column, HMW (GF) - high molecular weight fraction obtained after gel filtration, LMW (GF) - low molecular weight fraction obtained after gel filtration, Bmlp7 (IEC II) - the final Bmlp7 sample after ion exchange chromatography II, Bmlp3 (IEC II) - the Bmlp3 sample after ion exchange chromatography II, Bmlp3 (HIC) - the final Bmlp3 sample after hydrophobic interaction chromatography and desalting, SPs (IEC I) - the final SPs sample after ion exchange chromatography I, SPs (cryst) - the material collected from crystallization drops (dissolved crystals and solution surrounding the crystals, without the precipitate). The numbers 1, 2 and 3 indicate the band corresponding to contaminating protein of SPs sample, the band corresponding to the main component of SPs sample and the bands corresponding to 30 kDa lipoproteins, respectively.

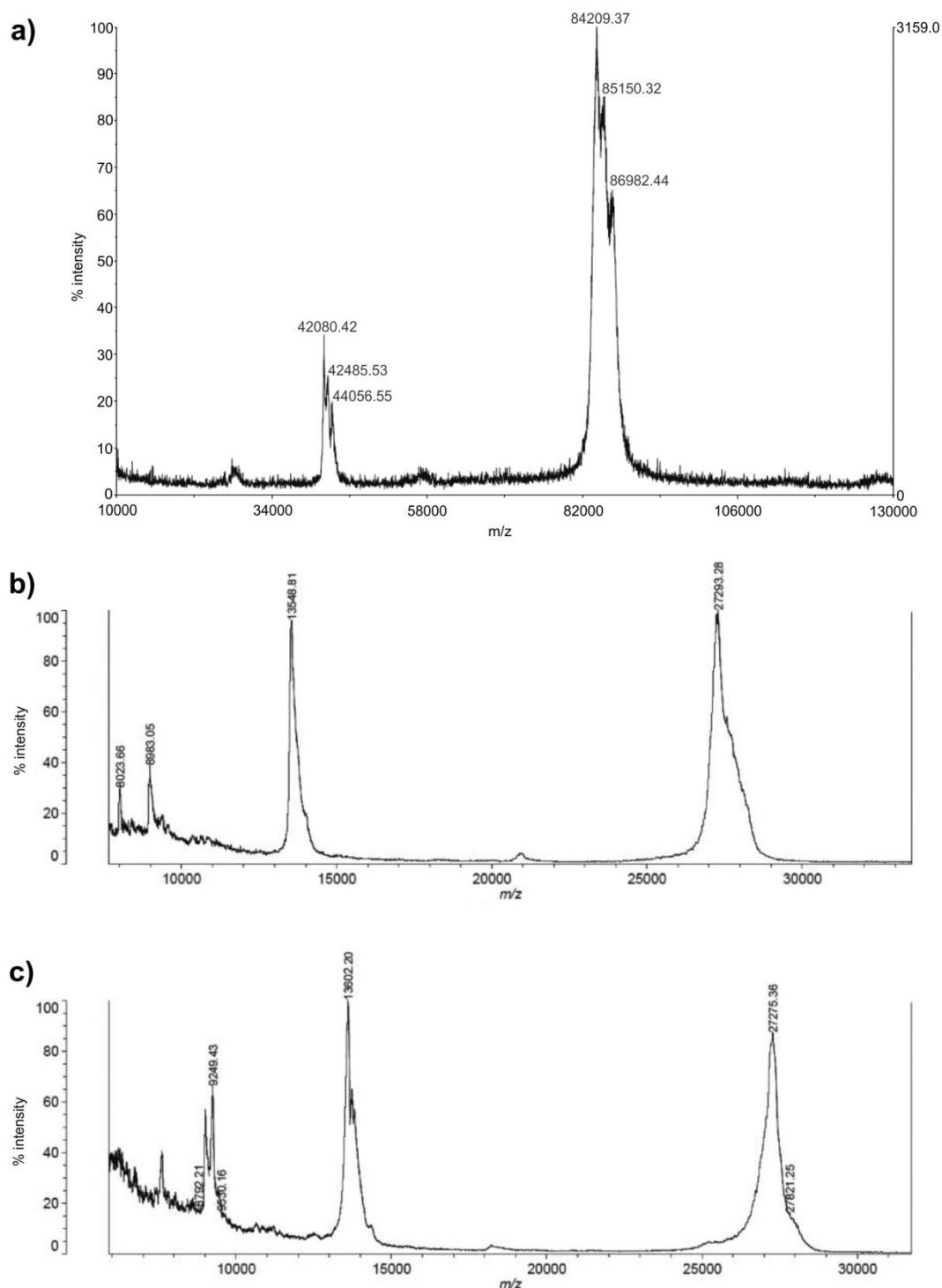


Fig. 6.4 MALDI-TOF MS spectra for investigated proteins.

MALDI-TOF MS spectra were recorded for **a)** SPs, **b)** Bmlp7 and **c)** Bmlp3.

MALDI-TOF MS analysis was carried out in order to establish the molecular weight of the investigated samples (Fig. 6.4). It was determined that the molecular weight of the main protein in the SPs sample was 84.2 kDa. However, two more peaks are visible in the MS

spectrum corresponding to the molecular masses of 85.2 and 86.9 kDa. The molecular weight of the Bmlp3 and Bmlp7 samples was established to be 27.3 kDa in both cases. The molecular mass calculated for these proteins according to their amino-acid sequence (Table 5.1) is slightly higher than that determined by MALDI-TOF MS, suggesting that probably partial degradation of the proteins occurred in these samples.

LC-MS/MS was used for the initial identification of the main, unknown at that time, components of the SPs and Bmlp7 samples. The particular bands were cut out from the SDS-PAGE gel and subjected to analysis. The analysis indicated that the main component of the SPs sample was SP2 at 75% level of confidence. The analysis of the Bmlp7 sample suggested tentatively that the investigated protein was PBMHP-12, the level of confidence in the identification of this protein was at 89%. Other top-ranking proteins were PBMHPC-23 and PBMHPC-19. All the listed proteins belong to the 30-kDa lipoprotein family, which indicated that the Bmlp7 protein is also a member of this protein family. In the case of the Bmlp3 sample, it was suspected that this protein also belongs to the 30-kDa LPs. In order to confirm this theory N-terminal sequencing was performed. The amino acid sequence of ten N-terminal residues was determined ($_1\text{ADSDVPNDIL}_{10}$). Several members of the 30-kDa LPs family have the same sequence at the N-terminus, thus it was not possible to identify a particular protein at this stage.

6.2. PROTEIN CRYSTALLIZATION, X-RAY DIFFRACTION DATA COLLECTION AND DATA PROCESSING

All crystals of the investigated proteins (Table 6.1; Fig. 6.5) were grown using the vapor diffusion method. A number of trials, including screening for initial crystallization conditions and further optimization of the conditions, were performed in order to obtain crystals of a good quality. Crystallization techniques were described in details in Materials & Methods (5.3).

Initial crystallization trials of SPs revealed that crystallization itself is the final purification step in this case. The use of 2.0 M ammonium sulfate as a precipitant caused the formation of a brown precipitate in the crystallization drops. Nevertheless, a month later, large crystals grew in the same drops containing the precipitate (Fig. 6.5a). In order to perform SDS-PAGE analysis the crystals were dissolved and collected with the solution surrounding them. The sample was then centrifuged to remove the precipitate. The supernatant was applied on gel. The lane corresponding to this sample (Fig. 6.3)

6. Results & Discussion

contained only the lower band (smaller molecular weight) indicating that the contaminating protein of a higher mass was the precipitating protein. The initial crystals had irregular prismatic shape (Fig. 6.5a) and diffracted X-rays to about 3.0 Å. The collected data could be indexed in the space group *C2* or *C222* and the calculated Mathews volume indicated the presence of 3 or 4 protein molecules in the asymmetric unit. The difficulties with indexing were caused by twinning, which was probably the result of two different ways of hexamers packing in space. The crystallization conditions were further optimized using Additive Screen (Hampton Research) and yielded crystals, identified later as SP2-SP3 complex crystals, with hexagonal plate morphology (Fig. 6.5b), belonging to space group *P6₃22*. The crystals grew in the presence of 0.02 M NaSCN and diffracted X-rays to 2.9 Å with no indication of twinning.

Table 6.1 Crystals of silkworm proteins.

	SP2-SP3	Bmlp7-I	Bmlp7-II	Bmlp3-p21	Bmlp3-c2
Protein concentration [mg/ml]	100	10	10	10	10
Crystallization conditions	2.0 M (NH ₄) ₂ SO ₄ , 0.02 M NaSCN, pH 5.5	22% PEG 3350, 0.2 M KSCN, 0.1 M HEPES, pH 7.5,	22% PEG 3350, 0.2 M KSCN, 0.1 M HEPES, pH 7.5,	30% PEG MME 550, 0.1 M CaCl ₂ , 0.1 M Bis-Tris, pH 6.5	2.0 M (NH ₄) ₂ SO ₄ , 0.1 M NH ₄ I, pH 5.5
Crystal size [mm]	0.15x0.15x0.05	0.35x0.2x0.1	0.45x0.3x0.15	0.2x0.1x0.05	0.15x0.1x0.05
Cryoprotection	70% Tacsimate, pH 5.5	50% PEG 400, 0.2 M KSCN, 0.1 M HEPES, pH 7.0, mixed with well solution in ratio 1:1	50% PEG 400, 0.02 M glucose, 0.2 M KSCN, 0.1 M HEPES, pH 7.0, mixed with well solution in ratio 1:1	no cryoprotection required	70% Tacsimate, pH 5.5

The table presents the information about the final crystals used for the X-ray diffraction experiment.

All crystals of Bmlp7 were grown using as the precipitating buffer 0.2 M KSCN, 0.1 M HEPES pH 7.5, and 22% PEG 3350. The Bmlp7 crystals can be divided into two groups: the crystals grown using Bmlp7 protein isolated from the hemolymph collected in 2008 (i) and in 2010 (ii). The first crystals, Bmlp7-I(Cd) (Fig. 6.5c), obtained from the 2008 batch of hemolymph belonged to space group *P1* and diffracted X-rays to 1.3 Å.

6. Results & Discussion

Unexpectedly, the presence of cadmium was detected later in these crystals. The heavy atom derivatives, Bmlp7-Hg and Bmlp7-Pt, were produced using the same crystals by soaking them in mercury (II) acetate (Bmlp7-Hg) or K_2PtCl_6 (Bmlp7-Pt) at the final concentration of 2.0 mM for one week. Bmlp7 purified from the 2010 batch of hemolymph provided crystals (Bmlp7-II; Fig. 6.5d) belonging to the second crystal form, space group $P1$ with different unit cell parameters (Table 6.2). These crystals were free of cadmium. The same protein sample was used for cocrystallization with cadmium ($CdCl_2$), which was externally added to the crystallization drop at a final concentration of 0.2 mM. The obtained crystals (Bmlp7-Cd) belonged to the same crystal form as Bmlp7-I(Cd).

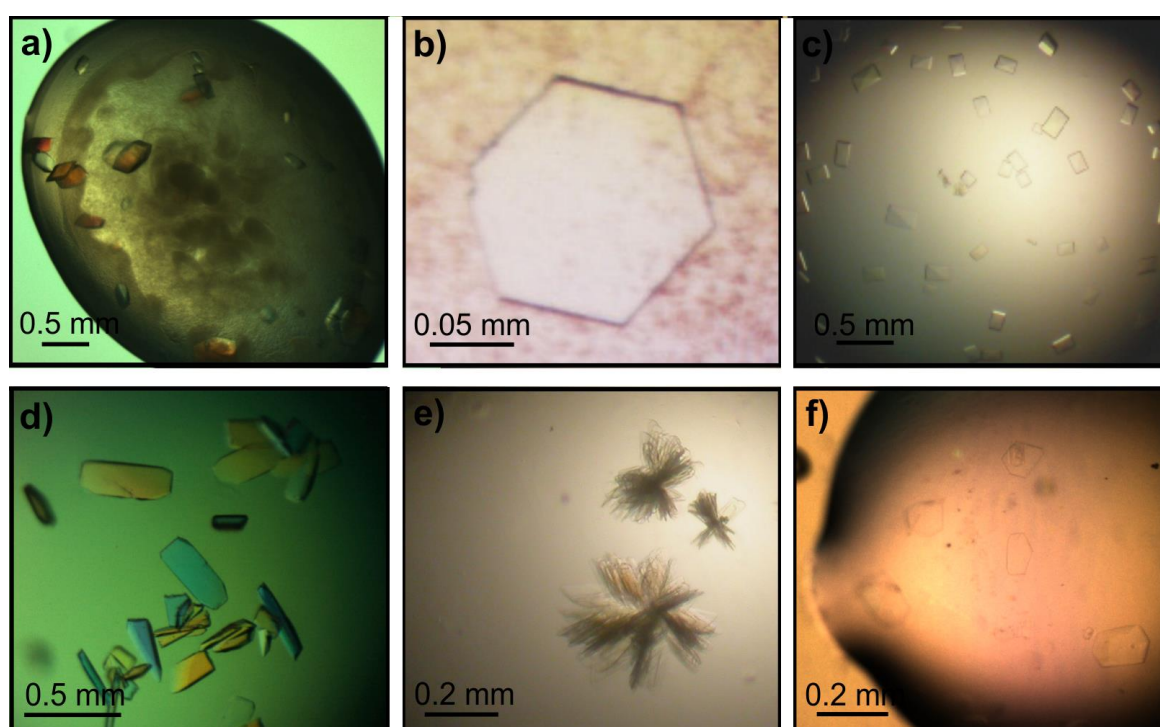


Fig. 6.5 Crystals of silkworm proteins.

Crystals of investigated proteins: **a)** SP2-SP3, the twinned crystals, **b)** SP2-SP3, hexagonal form, **c)** Bmlp7-I, **d)** Bmlp7-II, **e)** Bmlp3-p21, **f)** Bmlp3-c2.

Initial crystals of Bmlp3 were obtained using a sample after two steps of purification (GF and IEC II). The crystals belonged to space group $C222_1$ and diffracted X-rays to 2.9 Å, but intensity triage procedure indicated a twinning problem. Therefore additional purification step (HIC and protein desalting) were included in the purification protocol. Two crystal forms of Bmlp3 were obtained, each of them corresponds to different crystallization conditions. Crystals of the first crystal form (Bmlp3-p21; Fig. 6.5e) were grown using 0.1 M $CaCl_2$, 0.1 M Bis-Tris, pH 6.5 and 30% PEG MME 550 as the precipitating buffer. The obtained

6. Results & Discussion

crystals grew in clusters of tiny plates and belonged to space group $P2_1$. Crystals of the second crystal form (Bmlp3-c2; Fig. 6.5f) were grown using 2.0 M ammonium sulfate and 0.1 M ammonium iodide at pH 5.5 as the precipitating buffer. The crystals were tiny plates and belonged to space group $C2$.

The information about the crystals (their size, crystallization and cryoprotection conditions) is summarized in Table 6.1.

Table 6.2 X-ray diffraction data collection statistics of Bmlp7 crystals.

	Bmlp7-I(Cd)	Bmlp7-Hg	Bmlp7-Pt	Bmlp7-II	Bmlp7-Cd
Space group	$P1$	$P1$	$P1$	$P1$	$P1$
Unit cell parameters					
a, b, c [Å]	42.2, 50.0, 55.2	41.7, 49.7, 54.8	41.8, 49.7, 54.9	35.1, 71.7, 105.1	42.9, 50.7, 55.9
α, β, γ [°]	93.4, 94.6, 102.7	93.5, 94.7, 103.0	93.2, 94.8, 102.8	78.8, 89.9, 75.8	93.3, 94.7, 102.6
Molecules/AU	2	2	2	4	2
V_M [Å ³ /Da]	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	BESSY, BL114.1	DESY, X12	DESY, X12	DESY, X12	DESY, X12
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-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 / image [s]	5.2	6.0	4.0	20.0	15
Resolution [Å]	23-1.33 (1.38-1.33) ^a	50-2.22 (2.30-2.22)	50-1.94 (2.01-1.94)	37.0-2.5 (2.56-2.5)	40.0-1.79 (1.90-1.79)
Intensities measured	309 628	72 420	110 745	52 068	138 001
Unique reflections	94 702	40 082	29 990	29 885	38 532
R_{merge}^b [%]	10.9 (31.9)	8.1 (34.3)	6.8 (15.3)	4.3 (14.7)	5.2 (12.7)
Redundancy	3.3 (1.6)	1.9 (1.6)	3.7 (2.8)	1.8 (1.6)	3.6 (3.0)
$\langle I/\sigma I \rangle$	11.6 (2.4)	8.9 (2.2)	25.2 (4.9)	13.8 (4.7)	18.2 (7.9)
Completeness [%]	94.0 (78.8)	94.7 (76.0)	94.2 (75.0)	88.8 (88.8)	93.9 (80.4)

^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 .

6. Results & Discussion

X-ray diffraction data were collected and processed as was described in Materials & Methods (5.5; 5.6). The first and the most extensively studied protein was Bmlp7 and a number of data sets were collected for the Bmlp7 crystals (two crystal forms, heavy atom derivatives, crystals cocrystallized with cadmium). The data collection statistics for Bmlp7 are presented in Table 6.2. In the case of Bmlp3 the data representing two crystal forms were collected and for the SP2-SP3 complex only the hexagonal crystals were useful. The data collection statistics for Bmlp3 and SP2-SP3 are summarized in Table 6.3.

Table 6.3 X-ray diffraction data collection statistics of Bmlp3 and SP2-SP3 crystals.

	Bmlp3-p21	Bmlp3-c2	SP2-SP3
Space group	$P2_1$	$C2$	$P6_322$
Unit cell parameters			
a, b, c [Å]	56.9, 124.4, 67.9	154.1, 34.5, 93.1	192.8, 192.8, 180.8
α, β, γ [°]	90.0, 114.8, 90.0	90.0, 97.9, 90.0	90.0, 90.0, 120.0
Molecules/AU	4	2	2
V_M [Å ³ /Da]	2.01	2.25	2.89
Solvent content [%]	38.7	45.7	57.5
X-ray data collection			
Temperature [K]	100	100	100
Radiation source	BESSY, BL1 14.1	BESSY, BL1 14.1	BESSY, BL14.2
Wavelength [Å]	0.918	0.918	0.918
Detector	MAR CCD 225	MAR CCD 225	MAR CCD 225
Crystal-detector distance [mm]	250	250	290
Rotation range [°]	0.5	0.4	0.5
Total rotation [°]	180	180	75
Exposure / image [s]	20.0	17.0	35.0
Resolution [Å]	34.4-2.36 (2.46-2.36) ^a	38.1-2.10 (2.20-2.10)	48-2.9 (3.0-2.9)
Intensities measured	133 505	89 568	404 092
Unique reflections	35 189	26 553	44 232
R_{merge}^b [%]	8.2 (63.6)	6.0 (49.6)	10.8 (63.9)
Redundancy	3.8 (3.8)	3.7 (2.7)	10.2 (10.5)
$\langle I/\sigma I \rangle$	15.7 (2.7)	15.9 (2.6)	23.2 (5.0)
Completeness [%]	99.5 (99.7)	91.3 (86.3)	99.7 (100.0)

^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 .

6.3. STRUCTURE DETERMINATION

6.3.1. The initial model of Bmlp7 was obtained by SAD

The most challenging case of the phase problem was the structure of Bmlp7, Bmlp7-I(Cd). At the time when high-resolution X-ray diffraction data were collected for the Bmlp7 crystal, there was no model in the PDB suitable for MR calculations. Therefore, methods based on crystals' derivatization were chosen for the solution of the phase problem. Plenty of derivatives were prepared (5.4). Prior to the anomalous X-ray diffraction data collection, each derivatized crystal was checked by fluorescence scan for the presence of the heavy atom compounds. Only crystals containing heavy atoms were mounted and their diffraction properties were tested. Many crystals diffracted poorly or did not diffract after the derivatization procedure. Nevertheless, the anomalous X-ray diffraction data of a good quality were collected for mercury and platinum derivatives (Table 6.2). The use of Bmlp7-Hg data set was sufficient to solve the phase problem. The data set was collected at the 1.0055 Å wavelength which was selected according to a fluorescence scan. The subsequent steps of the structure solution were:

- the determination of the heavy atoms' positions by *SHELXD* (Schneider and Sheldrick, 2002);
- the determination of the correct space group enantiomorph for the substructure by *ABS* (Hao, 2004) and *SHELXE* (Sheldrick, 2002);
- the refinement of the heavy atom positions and initial calculation of the phases by *BP3* (Pannu *et al.*, 2003);
- density modification by *RESOLVE* (Terwilliger, 2000);
- the automated model building using the obtained phases by *ARP/wARP* (Perrakis *et al.*, 1999);
- the final structure solution by the *Auto-Rickshaw Native-MRSAD* protocol (unpublished) using as input: Hg-derivative data, the initial heavy atom sites, the resulting model (from *ARP/wARP*) that contained few meaningful secondary structures, and the high-resolution native data for phase extension.

6.3.2. MR as the most common method for phase problem solution

After the first crystal structure of Bmlp7 was solved, it served as an initial model for MR calculations of Bmlp3 and other Bmlp7 structures. Two more Bmlp7 structures, Bmlp7-Pt and Bmlp7-II, were solved by MR, refined and deposited in the PDB.

The Bmlp7-Pt structure was determined using anomalous data collected for a crystal derivative in order to compare its platinum binding sites with the unexpectedly detected cadmium binding sites in the Bmlp7-I(Cd) structure. The Bmlp7-II structure represented the second crystal form of Bmlp7. One more Bmlp7 structure (Bmlp7-Cd), obtained from a crystal with artificially added cadmium, was also solved for comparison of its cadmium binding sites with Bmlp7-I(Cd). The Bmlp7-Cd structure was not deposited in the PDB, because it was isomorphous with Bmlp7-I(Cd).

The Bmlp3-p21 structure was also solved by MR using the coordinates of chain A of Bmlp7-I(Cd) as the initial model. The starting model for Bmlp3-c2 was chain A of Bmlp3-p21, because this model already contained the correct Bmlp3 sequence.

Finally, the crystal structure of the SP2-SP3 complex was also determined by the MR method. The structure of arylphorin from oak silkworm (APA; PDB code: 3GWJ; Ryu *et al.*, 2009) was used as a model, because SP2 and APA share 85% similarities and 69% identities in their amino acid sequences.

The great majority of the structures described in this dissertation was determined by MR. When a suitable model is available, MR is the method of choice for solving the phase problem, because it does not require preparation of derivatives and additional X-ray diffraction measurements.

6.4. SEQUENCING FROM ELECTRON DENSITY MAPS

All investigated proteins were purified as unknown proteins. After characterization of the initial samples by LC-MS/MS or N-terminal sequencing, the information about the amino acid sequences of the studied proteins was incomplete or incorrect. It was thought that SP2-SP3 hexamerin consisted exclusively of SP2 subunits, Bmlp7 was tentatively identified as PBMHP-12 and the only information about the Bmlp3 sample was that the protein belonged to the 30-kDa LPs family.

The final identification of all proteins was possible only according to the electron density maps. X-ray crystallography had been used for resolving sequence-related ambiguities almost from its inception, and nowadays it still remains a powerful tool for protein identification, especially for unknown proteins isolated from their natural sources. Each protein 3D structure contains information on the protein primary structure. The most interesting case studies of sequencing from electron density maps were described in a review prepared by the author of this dissertation (Appendix: Publication IV).

6.4.1. Crystallographic identification of Bmlp7

After the first cycles of refinement in *REFMAC5* (Murshudov *et al.*, 2011) and visual inspection of the electron density maps in *Coot* (Emsley and Cowtan, 2004; Emsley *et al.*, 2010), it was possible to identify the correct amino acid sequence of the protein. It turned out that the protein was not PBMHP-12, but the amino acid sequence matched another protein belonging to 30-kDa LPs, Bmlp7. The alignment of the 30-kDa LPs amino acid sequences indicates a high level of homology (Table 3.2). Many differences between PBMHP-12 and Bmlp7 sequences were found in the N-terminal region and in the positions of aromatic residues (Fig. 6.6). The amino acid side chains were clearly visible in the high resolution, 1.33-Å, electron density maps (Fig. 6.7) which enabled the sequence identification and correction of the model.

As mentioned before (3.2.2.2) the information available in different databases about 30-kDa LPs sequences is not well-ordered, thus a number of entries corresponding to the Bmlp7 sequence were found (the SilkDB, accession code: Bmb021422; UniProt, accession code: E5EVW2, D4QGC0; GenBank, accession code: BAJ 04788.1). The protein was annotated as Bmlp7 only in the SilkDB, other databases denoted it as the putative uncharacterized protein 19G1P or as a 30 kDa protein. The Bmlp7 name is the most appropriate, because it refers to the gene name.

30-kDa LPs, as well as other silkworm storage proteins, contain an N-terminal signal peptide which is typical for proteins secreted from fat bodies into the hemolymph. It usually consists of a positively charged N-terminus followed by a path of hydrophobic residues (Shimada *et al.*, 1985). The investigated proteins were isolated from the hemolymph, thus they represented mature protein forms without the signal peptide. In the case of Bmlp7 it was unclear which residue is the first residue of the mature protein. The cleavage site of the signal peptide was not indicated in the databases. Therefore, the N-terminal sequencing was performed for Bmlp7. At the time of this analysis, the Bmlp7-I(Cd) structure was completed. The Bmlp7 sample was separated in SDS-PAGE electrophoresis, the protein was transferred to a PVDF membrane and a single band corresponding to Bmlp7 was subjected to N-terminal sequencing by chemical degradation. Unexpectedly, the results of the analysis were two completely different sequences of the N-terminal decapeptide: ${}^1\text{ADSDVPNDIL}_{10}$ and ${}^1\text{GVVELSADSM}_{10}$. The first sequence corresponded to Bmlp7, whereas the other one was the N-terminal sequence of PBMHP-12. Each step of Edman degradation returned two peaks on chromatograms and the ratio of their heights was 3:1. Therefore, it could be easily distinguished which amino acid belongs to which sequence. The main component

6. Results & Discussion

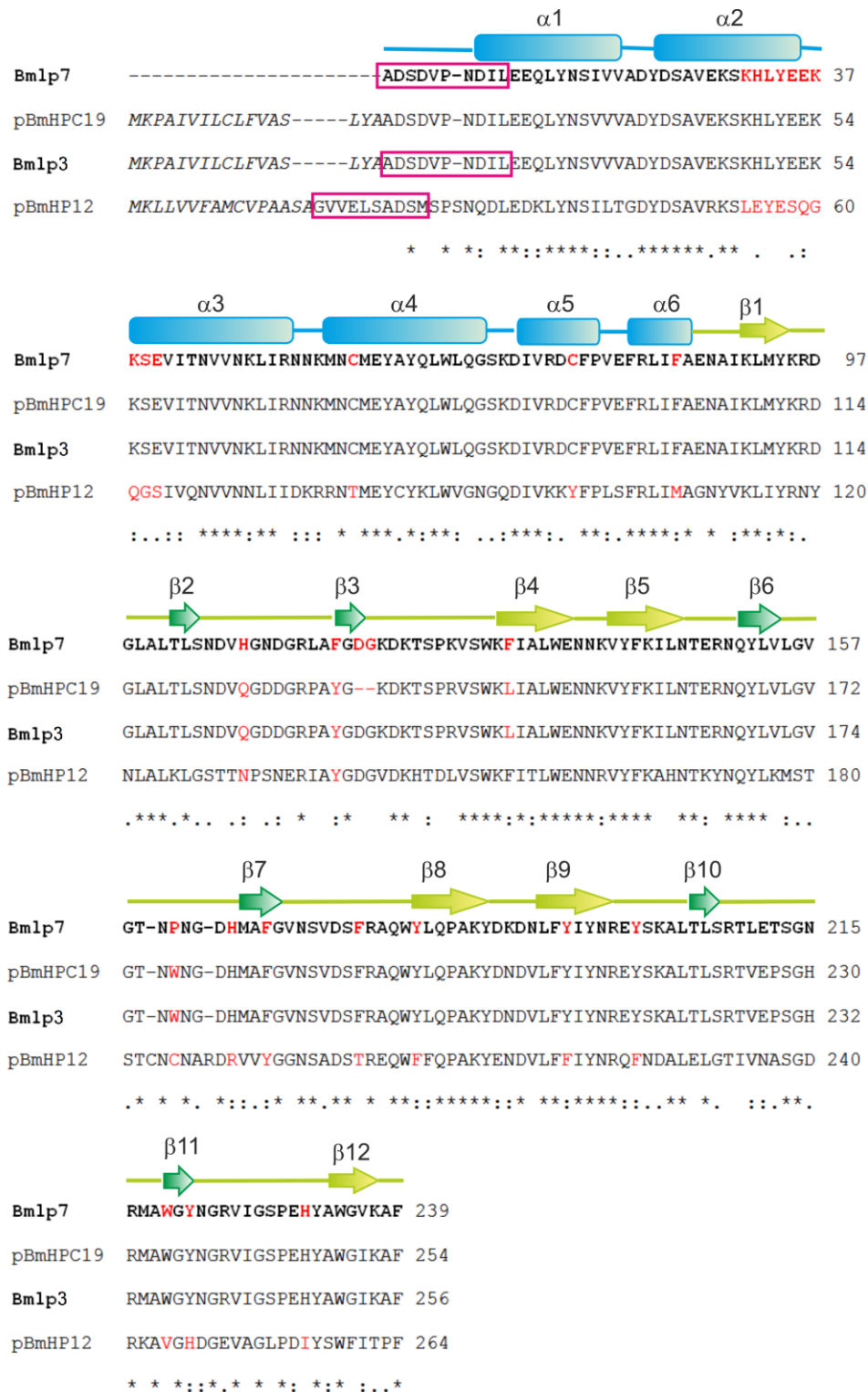


Fig. 6.6 Sequence alignment of 30-kDa LPs.

Amino acid sequence alignment of Bmlp7, pBmDP19 (30Kc19), Bmlp3 and pBmHP-12 (UniProt accession codes: E5EVW2, P09336, H9J4F6 and P09335, respectively). The N-terminal signal peptides are shown in italics. The main differences which helped to identify the correct sequences of Bmlp7 and Bmlp3 are in red lettering. The N-terminal sequences established by Edman degradation are shown in magenta boxes. The sequence alignment was calculated in ClustalW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). The secondary structure elements are assigned as α-helices (cylinders) and β-strands (arrows).

of the sample was Bmlp7. It turned out that the sample used for crystallization was heterogeneous and also in this case crystallization was the last purification step. These results explained the incorrect initial identification of Bmlp7 as PBMHP-12.

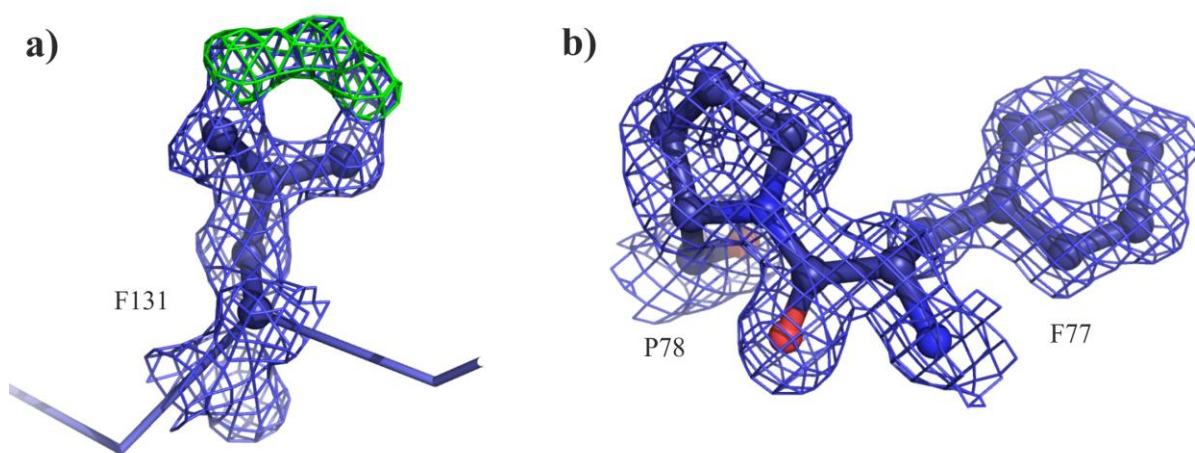


Fig. 6.7 The high resolution electron density maps of Bmlp7.

The electron density maps of Bmlp7 were of superb quality which enabled verification and correction of the amino acid sequence. **a)** The shape of the electron density clearly indicated that the amino acid residue is phenylalanine instead of leucine. **b)** All residues with annular side chains were easy to identify, for instance, Phe77 and Pro78 are shown. The $2Fo-Fc$ maps are displayed in blue at $1.3\ \sigma$ and the $Fo-Fc$ maps in green at $3.0\ \sigma$.

6.4.2. Crystallographic identification of Bmlp3

The other 30-kDa LP was identified as Bmlp3 also during the refinement in *REFMAC5* (Murshudov *et al.*, 2011) and *Coot* (Emsley and Cowtan, 2004; Emsley *et al.*, 2010). The amino acid sequences of Bmlp3 and Bmlp7 (Fig. 6.6) are similar in 97% and identical in 94%. Nevertheless, the inspection of the electron density (Fig. 6.8) allowed me to detect the proper sequence of Bmlp3. Two identical amino acid sequences corresponding to this protein were found in the databases (SilkDB, accession code: Bmb035159; UniProt, accession code: H9J4F6). The Bmlp3 name refers to the gene name.

6.4.3. Crystallographic identification of an unexpected complex of SP2 and SP3

Although SP2 was extensively studied during the last years, the amino acid sequence found in UniProt (accession code: P20613) and denoted as SP2 turned out to be incorrect in its C-terminal part (residues 634-688). This conclusion was based on the results of the first

refinement cycles in *REFMAC5* (Murshudov *et al.*, 2011) and *Coot* (Emsley and Cowtan, 2004; Emsley *et al.*, 2010). The correct sequence was found with the accession code Q1HPP4. Interestingly, both SP2 sequences deposited in the UniProt differ in 10% (Table 3.1; Fig. 6.9). Probably, some frame shift errors occurred during gene sequencing of the entry P20613. The C-terminal fragment of the model was improved after the sequence was changed for Q1HPP4 which was reflected in a decrease of *R* factors by about 3%.

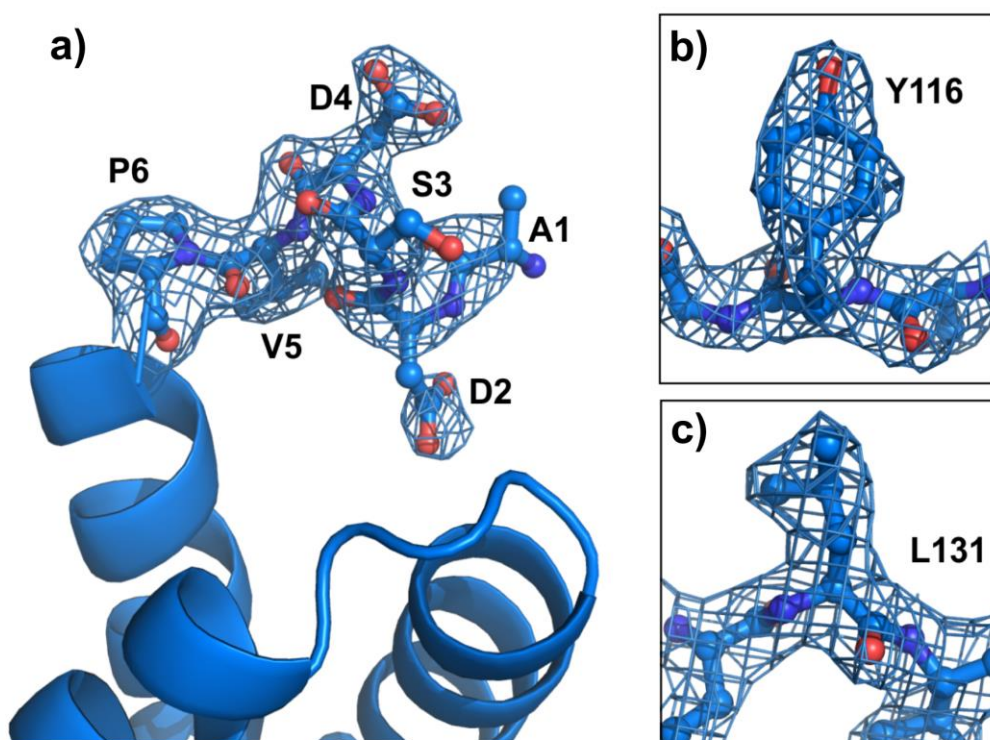


Fig. 6.8 The electron density maps of Bmlp3.

The resolution of the Bmlp3 data was lower by about 1 Å in comparison to the high resolution structure of Bmlp7. **a)** Bmlp3-p21 is the first and the only complete structure of a member of the 30-kDa LPs, deposited in the PDB. All amino acids residues of the N-terminus were clearly visible in the electron density maps. **b,c)** The good quality of the electron density maps allowed identification of the correct amino acid sequence of the protein. The $2Fo-Fc$ maps are displayed in blue at 1.0σ .

The asymmetric unit of the hexamerin was comprised of two protein molecules. The electron density maps obtained after subsequent rounds of refinement allowed me to detect further sequence-related ambiguities. Although $C\alpha$ atoms of the model could be easily traced in the maps, a number of positive and a few negative peaks at the 5σ level were present in the F_o-F_c difference electron density maps at amino acid side chains of subunit A (Fig. 6.10). No similar peaks were observed in corresponding sites of subunit B (Fig. 6.10).

6. Results & Discussion

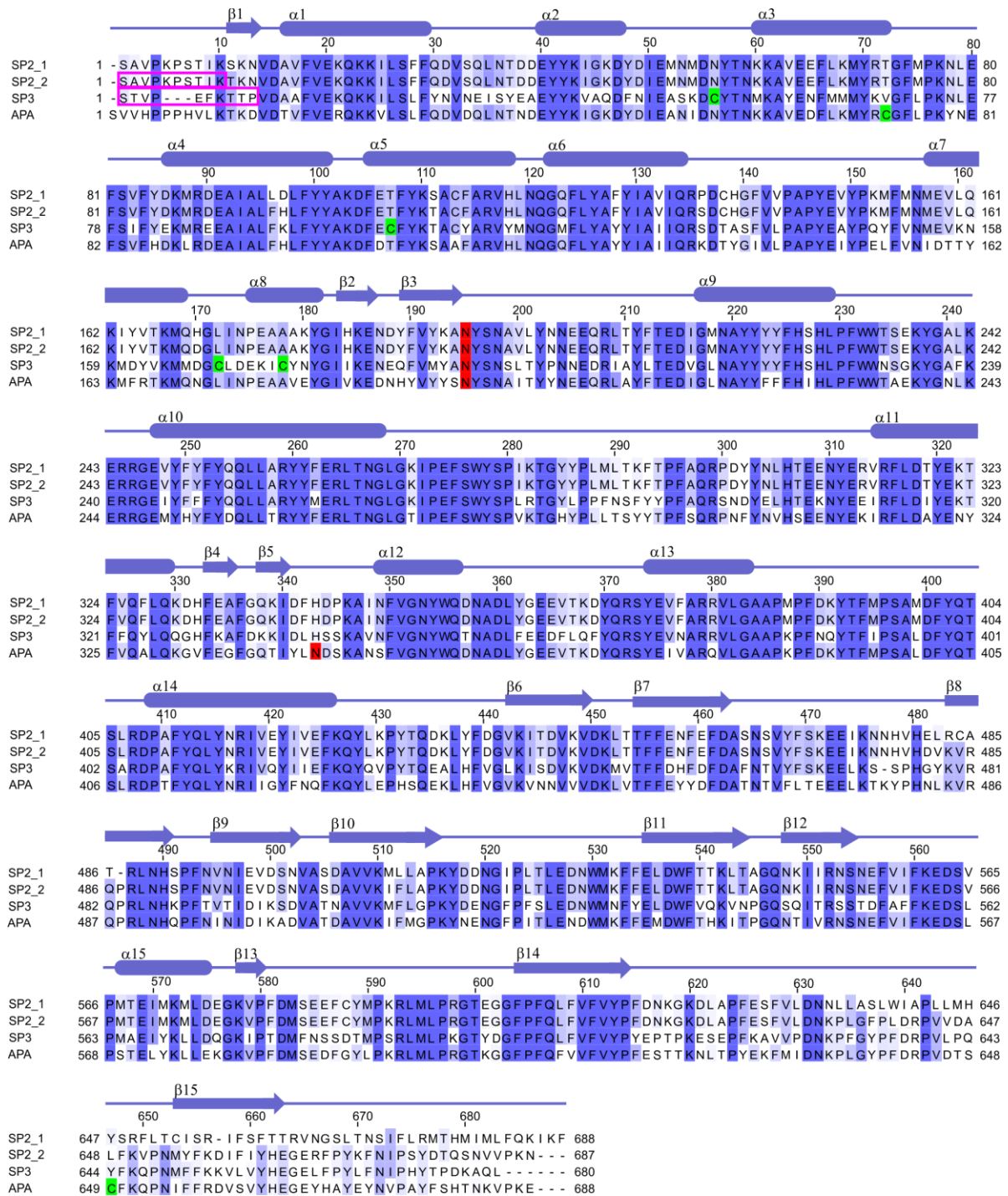


Fig. 6.9 Sequence alignment of arylphorins.

Amino acid sequence alignment of arylphorins, SP2_1, SP2_2, SP3 and APA (UniProt accession codes: P20613, Q1HPP4, H9JHM9 and Q7Z1F8, respectively), calculated in ClustalW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) and colored according to identities (dark blue) and similarities (light blue) using Jalview (<http://www.jalview.org/>; Waterhouse et al., 2009). The N-terminal sequences established by Edman degradation are shown in magenta boxes. The presented sequences correspond to mature proteins. The glycosylation sites are highlighted in red and cysteine residues which form disulfide bridges are highlighted in green. The secondary structure elements are assigned as α -helices (cylinders) and β -strands (arrows).

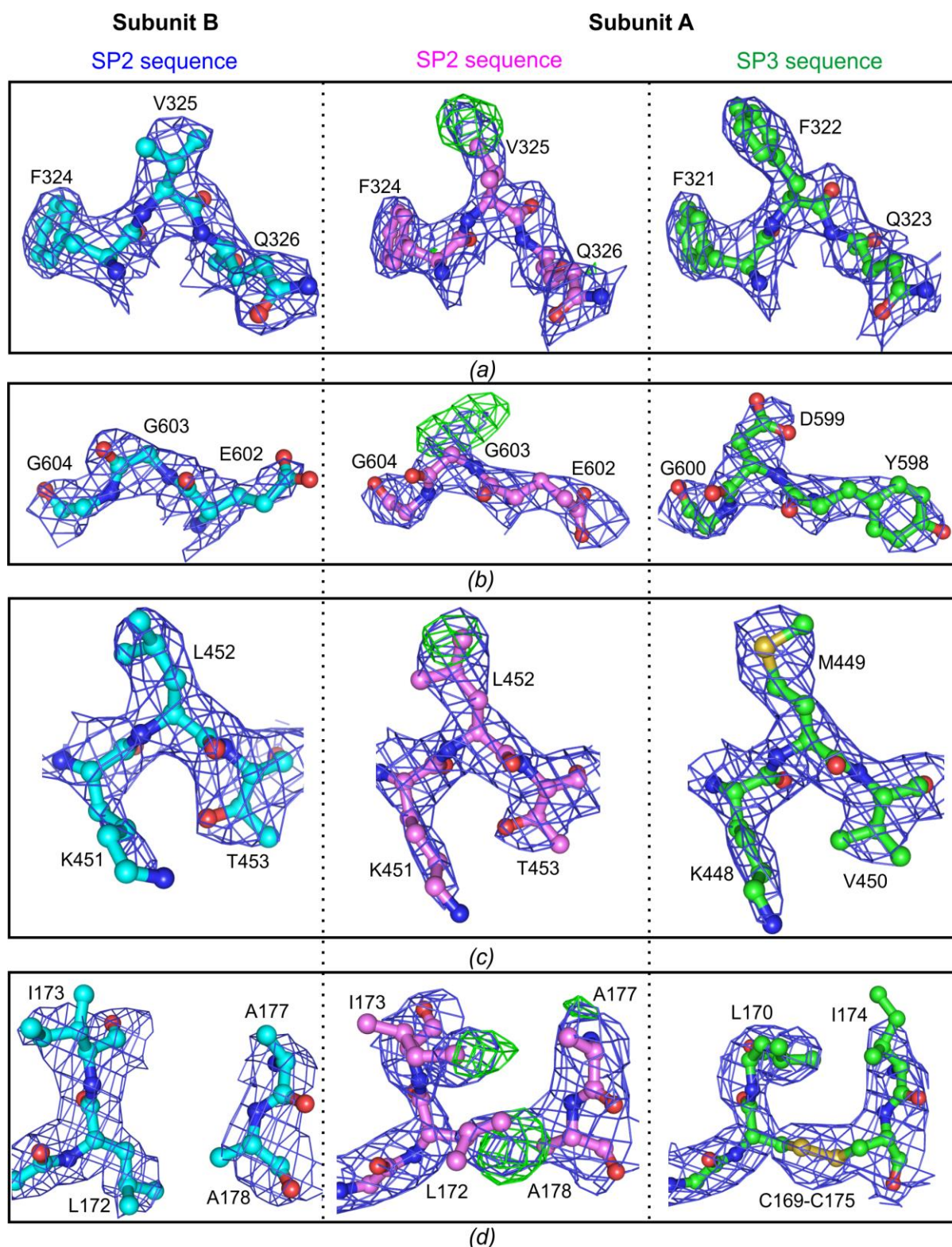


Fig. 6.10 The electron density maps of SP2-SP3.

The electron density maps enabled identification of the hexamerin as being composed of two different proteins. More than 20 peaks in the difference electron density maps of subunit A indicated that the SP2 sequence is not correct for this protein. Four most representative sites are shown in panels (a-d), presenting the same site in subunit B (cyan), for which the SP2 sequence is correct, the corresponding site in subunit A with the incorrect SP2 sequence (magenta) and after its correction for the SP3 sequence (green). The $2Fo-Fc$ maps (blue) are contoured at the 1.0σ level and the $Fo-Fc$ maps (green) at the 3.0σ level.

6. Results & Discussion

The search for other sequences of silkworm arylphorins was performed and the sequence of SP3, matching the electron density maps of subunit A perfectly, was found in the UniProt with accession code H9JHM9. The *R* factors decreased by about 2% after the SP3 sequence was assigned to subunit A. Crystallographic studies revealed that the hexamerin is a heterohexamer of two different arylphorins. Both storage proteins contain a significant amount of aromatic residues in their amino acid sequences, and what is also important, the positions of these residues are different in both proteins, which was essential for the sequencing from 2.9 Å electron density maps.

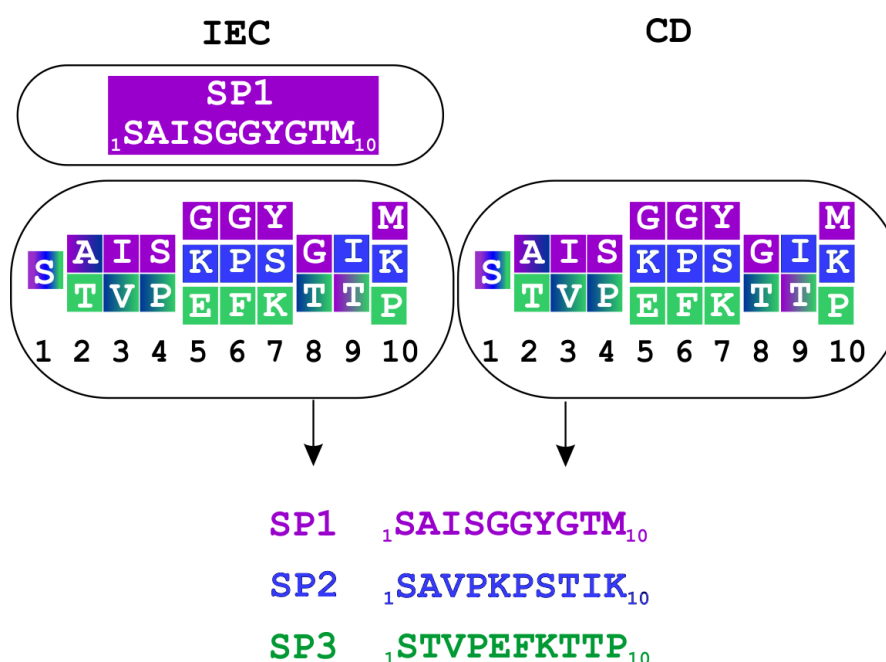
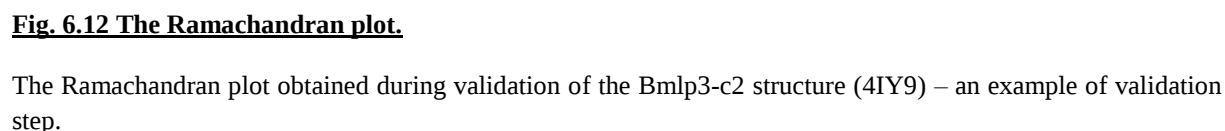


Fig. 6.11 The results of N-terminal sequencing of SPs samples.

The schematic representation of the SDS-PAGE bands (as in Fig. 6.3, trace SPs/IEC I and SPs/cryst) which were applied for the N-terminal sequencing analysis. The results of the analysis for a particular band are located within this band. The colors black, blue and green represent the N-terminal residues of SP1, SP2 and SP3, respectively. The SP1 protein is present in two bands which could be speculatively explained by different glycosylation forms of this protein. It can be connected to the results of MALDI-TOF MS (Fig. 6.4a), where two additional peaks (85.1 and 86.9 kDa) could be assigned to SP1.

The N-terminal sequencing by Edman degradation was carried out to confirm the information provided by X-ray crystallography. Two samples were analyzed: (1) the sample obtained after purification, also containing the contaminating protein and (2) the material collected from crystallization drops, i.e. dissolved crystals and solution surrounding them, with the precipitate removed by centrifugation of the sample. Firstly,

To this date, not much is known about SP3 and the N-terminal sequencing revealed the first residues of the mature protein (1STVP....). The storage proteins (in immature form, before secretion to hemolymph) also contain an N-terminal signal peptide (6.4.1).



6.5. STRUCTURE REFINEMENT AND VALIDATION

All six crystal structures of the major hemolymph proteins were refined to the final R/R_{free} factors below 22 / 29 %. The root-mean-square deviations (r.m.s.d.) from the ideal geometry for the bond lengths and for the angles were in the range of standard values. Excluding glycine and proline residues, the Ramachandran plot had more than 90 % of the residues in the most favored regions in each structure and no more than 0.4 % residues in the disallowed regions (Fig. 6.12). The detailed refinement statistics for all structures are summarized in Table 6.4. The structures were deposited in the PDB with the following codes: 4EFP [Bmlp7-I(Cd)], 4EFQ (Bmlp7-Pt), 4EFR (Bmlp7-II), 4IY8 (Bmlp3-p21), 4IY9 (Bmlp3-c2) and 4L37 (SP2-SP3).

Table 6.4 Refinement statistics of crystal structures of major hemolymph proteins.

	Bmlp7-I(Cd)	Bmlp7-Pt	Bmlp7-II	Bmlp3-p21	Bmlp3-c2	SP2-SP3
R_{work} / R_{free}^a [%]	18.4 / 23.1	18.5 / 22.7	20.1 / 28.5	17.3 / 25.9	21.6 / 27.1	16.6 / 22.3
R_{free} test set count	1223	1019	949	1106	1325	1 096
Protein /solvent atoms	4006/551	3871/234	7702/137	7817/220	3878/124	11 373/136
R.m.s. deviations (bond lengths) [Å]	0.020	0.019	0.020	0.015	0.019	0.013
R.m.s. deviations (bond angles) [°]	1.94	1.69	1.98	1.71	1.81	1.66
Average B factor [Å ²]	16.07	10.52	20.5	27.7	34.8	43.1
Favored/disallowed Ramachandran ϕ/ψ [%]	97.0/0.0	95.86/0.22	95.45/0.32	90.7/0.1	91.7/0.0	91.0/0.1
PDB code	4EFP	4EFQ	4EFR	4IY8	4IY9	4L37

^a $R_{work} = \sum_h ||F_o| - |F_c|| / \sum_h |F_o|$ 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.

6.6. OVERALL STRUCTURE

6.6.1. Overall structure of SP2-SP3 complex

The asymmetric unit of the SP2-SP3 complex contains two protein molecules, one molecule of SP2 (subunit B) and one of SP3 (subunit A). The biologically relevant oligomeric state of arylphorins is a hexamer, as indicated by the biochemical analysis (Burmester and Scheller, 1996). One trimer of SP2 and one trimer of SP3 form the biological

assembly which is in this case a heterohexamer. The trimers are located along the crystallographic 6_3 axis (6.8.1).

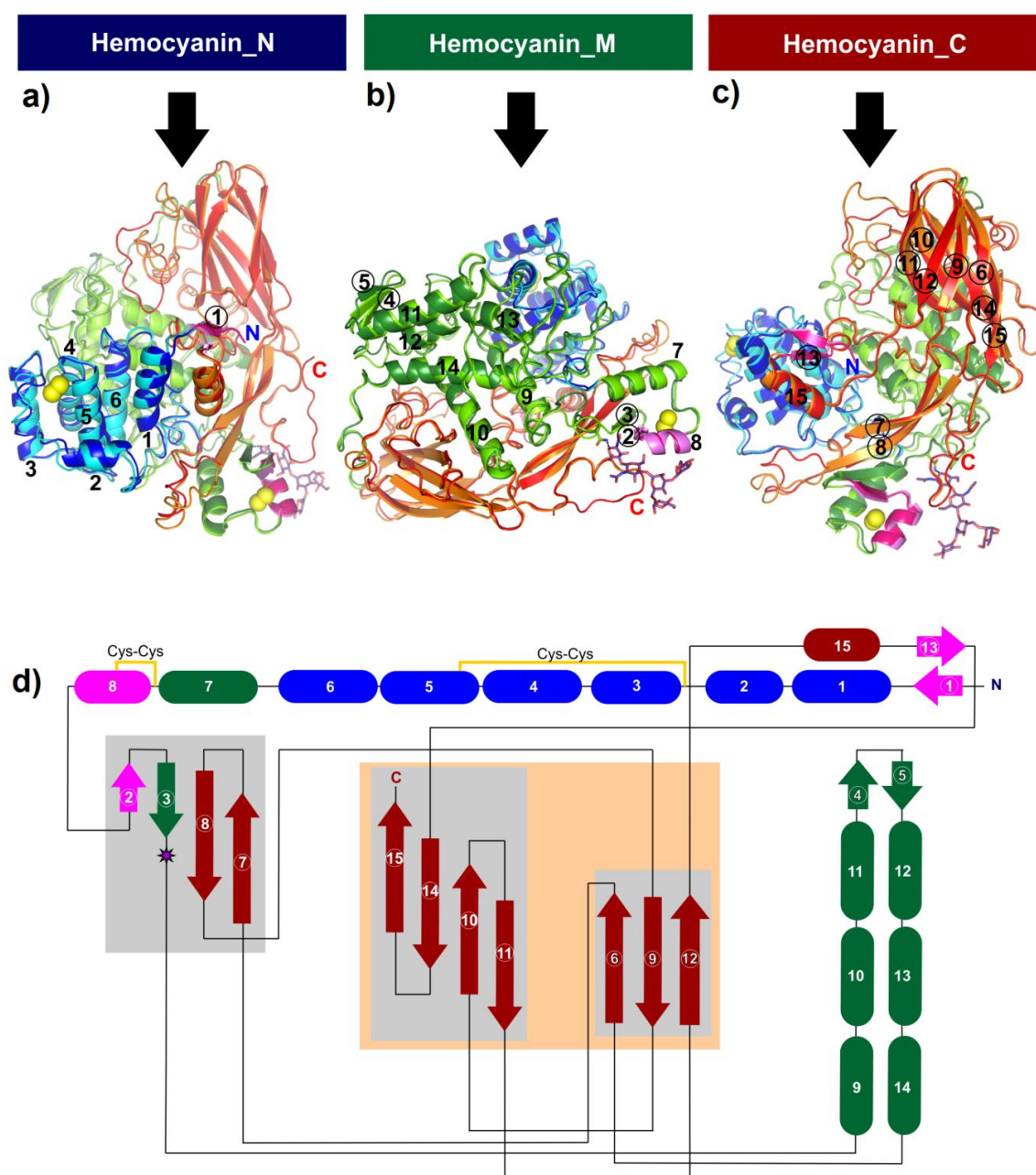


Fig. 6.13 Overall structure of SP2 and SP3.

a,b,c) A cartoon representation of the SP2 and SP3 fold. The SP2 (blue-green-red) and SP3 (cyan-light green-orange) chains are presented in three views to show each of the subdomains: **a)** hemocyanin_N, **b)** hemocyanin_M and **c)** hemocyanin_C. The secondary structure elements characteristic only of arylphorins are shown in magenta. The sulfur atoms of Cys participating in forming of disulfide bonds in SP3 structure are presented as yellow spheres. **d)** A topology diagram of SP2 and SP3. Cylinders and arrows represent α -helices and β -strands, respectively. The lengths of the α -helices, β -strands and of the loops are not commensurate with the numbers of amino-acid residues in these elements. The secondary structure elements forming β -sheets and an elongated β -barrel are highlighted in gray and in orange, respectively. The glycosylation site is marked with a star.

Chain A (SP3) and chain B (SP2) contain 674 of the 680 residues of mature SP3 (6-679) and 667 of the 687 residues of mature SP2 (9-675), respectively. The structure of SP2, as well as the structure of SP3, comprises three hemocyanin-like subdomains (Fig. 6.13): the hemocyanin_N domain (residues 17-157/SP2; 14-153/SP3), the hemocyanin_M domain (residues 157-441/SP2; 154-438/SP3), and the hemocyanin_C domain (residues 442-675/SP2; 439-679/SP3). The hemocyanin_N,M,C domains can be found in the Pfam database (<http://pfam.sanger.ac.uk>) with accession codes: PF03722, PF00372 and PF03723, respectively. This fold is consistent with the architecture of oak silkworm arylphorin (APA; Ryu *et al.*, 2009) and of arthropod hemocyanins. There is a hypothesis that arylphorins evolved from arthropod hemocyanins (Ryu *et al.*, 2009). The hemocyanin_N subdomain is an α -helical domain, formed exclusively of α -helices (α 1- α 6 in BmSP topology; Fig. 6.13d). The second, hemocyanin_M, subdomain contains six α -helices (α 7- α 14) and four β -strands (β 2- β 5). The interactions of the residues belonging to this subdomain dominate among the contacts forming the hexamer (6.8). The third subdomain, hemocyanin_C, comprises mainly of β -strands (β 6- β 15) which form an elongated β -barrel. Such a domain can be classified as an IG-like domain. Additionally, one, short α -helix (α 15) is also present in this subdomain.

The structures of SP2 and SP3 closely resemble each other, displaying a 0.83 Å r.m.s.d. over 645 aligned C α atoms with 65% sequence identity (Table 3.1). Both structures contain a number of flexible loops which adopt different conformations in SP2 and SP3, the most flexible is the loop between β 14 and β 15. The disulfide bonds are present only in SP3 and link Cys53-Cys104 and Cys169-Cys175. The N-glycosylation characteristic for arylphorins is present at Asn195 of SP2 and at Asn192 of SP3.

6.6.2. Overall structure of 30-kDa LPs

The fold of 30-kDa LPs (Fig. 6.14) is unique and characteristic only for this group of proteins. This type of architecture is classified as the Lipoprotein_11 family (Pfam, accession code: PF03260). 30-kDa LPs were comprised of two domains, N-terminal VHS domain (Pfam: PF00790) and C-terminal β -trefoil domain (Pfam: PF14200).

The N-terminal domain (NTD; residues 5-87/Bmlp7/Bmlp3-c2; 1-87/Bmlp3-p21) consists of six α -helices, which pack into a globular form of a right-handed superhelix. The name “VHS” corresponds to the first letters of the protein names (VPS-27, Hrs and STAM) which were classified into this group according to the SMART database (Schultz *et al.*, 1998).

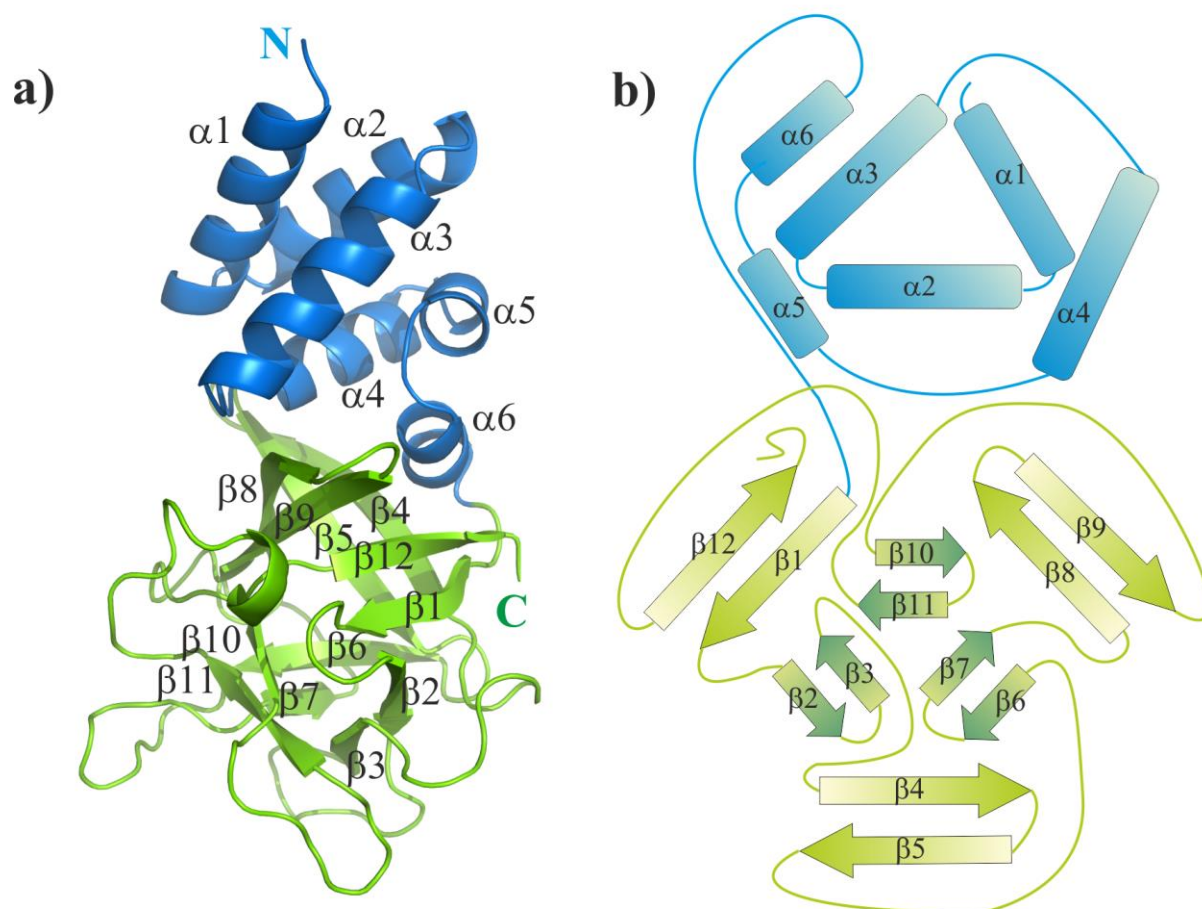


Fig. 6.14 Overall structure of 30-kDa LPs.

a) A cartoon representation of the Bmlp7 fold as an example of the overall architecture of 30-kDa LPs. The NTD and the CTD are shown in blue and green, respectively. **b)** A topology diagram of 30-kDa LPs. Cylinders and arrows represent α -helices and β -strands, respectively. The lengths of the secondary structure elements are not commensurate with the numbers of amino-acid residues in these elements.

The C-terminal domain (CTD; residues 88-239/Bmlp7/Bmlp3) is folded as a β -trefoil, or more precisely a ricin-type β -trefoil lectin-like domain. The CTD can be divided into two parts, each containing six β -strands paired in three β -hairpins, which are arranged as a three-blade β -propeller around a pseudo-3-fold axis. The β -strands closer to the NTD are longer. Both parts form a barrel-like assembly. The β -strands are connected by deeply intertwined loops which are well-defined.

The amino acid sequences of Bmlp3 and Bmlp7 are 97% similar and 94% identical which results in a high degree of structural similarity (Fig. 6.15). The chains A of Bmlp7-I(Cd) and of Bmlp3-p21 can be superposed with r.m.s.d. of 0.48 Å for C α pairs. The similar r.m.s.d. value (~ 0.5 Å) is obtained for the alignments of each chain of other Bmlp7 crystal structures (Bmlp7-Pt; Bmlp7-II) with chains A, B, C, D of Bmlp3-p21 or chain

A of Bmlp3-c2. The only exception is chain B of Bmlp3-c2 for which C α superposition with Bmlp7 chains produces r.m.s.d. values of 0.7 Å. It is connected to the fact that loop 106-116 of Bmlp3-c2/B has a completely different conformation from the corresponding loop in Bmlp7 and in other Bmlp3 molecules (Fig. 6.15). Such a conformational change enabled the formation of a dimer in Bmlp3-c2 (6.8.2). Differences between Bmlp3 and Bmlp7 can be observed mainly in loop regions. Moreover, only Bmlp7 contains in its structures a modified tryptophan residue.

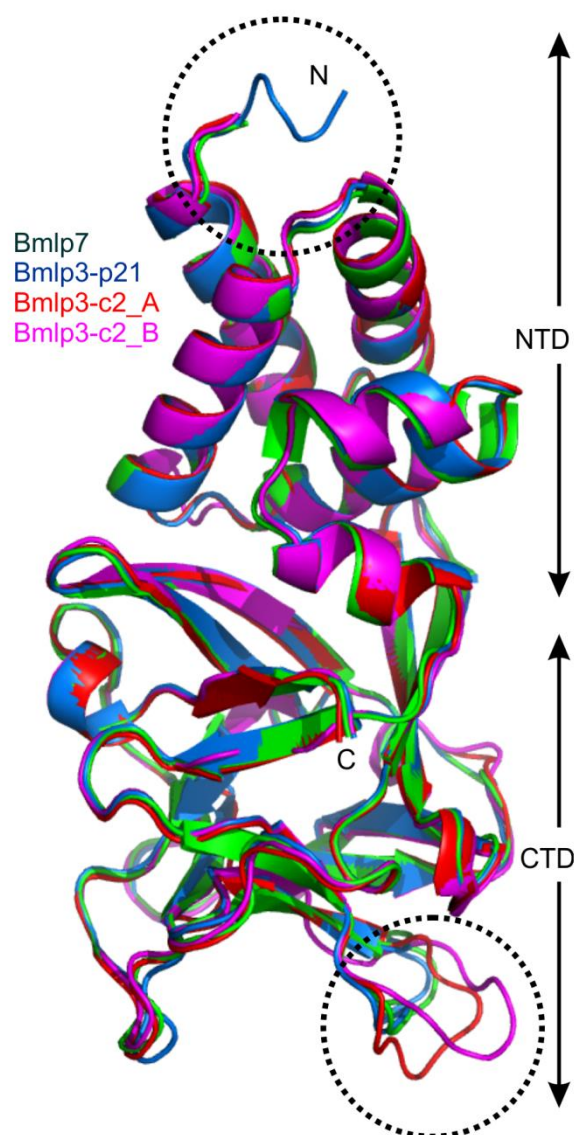


Fig. 6.15 Structural comparison of Bmlp3-p21, Bmlp3-c2 and Bmlp7.

The structures of Bmlp7-I(Cd) (chain A), Bmlp3-p21 (chain A), Bmlp3-c2 (chains A and B) were C α -superposed. The most flexible loop and the N-terminus are encircled.

6.7. STRUCTURAL COMPARISON

A search for structural homologues of major hemolymph proteins was performed using the *PDBeFold* (SSM) (<http://www.ebi.ac.uk/msd-srv/ssm/>) and *DALI* (Holm and Rosenstrom, 2010; http://ekhidna.biocenter.helsinki.fi/dali_server/) servers. In almost all cases, both servers returned identical results.

6.7.1. Primary structure determines the role of a protein: arylphorins vs. hemocyanins

Structural comparison was performed for both storage proteins separately. Nevertheless, the analysis returned exactly the same results for SP2 and SP3. At first the servers listed APA, which was not surprising, because this arylphorin structure was used for MR as an initial model. All other listed proteins, mainly arthropodan hemocyanins and phenoloxidases, belonged to the hemocyanin family. The results of C α superposition of SP2 or SP3 and their structural homologues is presented in Table 6.5.

Table 6.5 R.m.s.d. values calculated for C α atoms of mulberry silkworm storage proteins superposed on homologues proteins.

	3GWJ	1HC1	1LLA	3IXV	3HHS
SP2	0.83 / 645	1.70 / 559	1.90 / 519	1.99 / 514	1.92 / 505
SP3	0.94 / 657	1.77 / 568	1.83 / 510	1.81 / 500	1.99 / 519

The table presents the r.m.s.d. values in Å / the number of superposed C α atoms for SP2 (chain B) and SP3 (chain A) with chain A of a structure representing each of the homologues, represented by their PDB codes: 3GWJ (oak silkworm arylphorin; Ryu *et al.*, 2009), 1HC1 (California spiny lobster hemocyanin; Volbeda and Hol, 1989), 1LLA (Atlantic horseshoe crab hemocyanin; Hazes *et al.*, 1993), 3IXV (Sahara scorpion hemocyanin; Cong *et al.*, 2009) and 3HHS (tobacco hornworm phenoloxidase; Li *et al.*, 2009c).

The overall fold of SP2 and SP3 is consistent with the fold of APA. A comparison of arylphorins and hemocyanins reveals more differences. Arylphorins contain a number of additional secondary structure elements: one α -helix (α 8) and three β -strands (β 1, β 2 and β 13) forming β -sheets; β 1 is paired with β 13 and β 2 interacts with β 3. The aromatic amino acid residues of SP2, SP3 and APA constitute 19.0, 20.7 and 19.3%, respectively, of the total amino acid content, whereas their fraction in hemocyanins is about 10%. Although the tertiary structure of arylphorins and hemocyanins is very similar, their primary structure identity is only 26 to 29%.

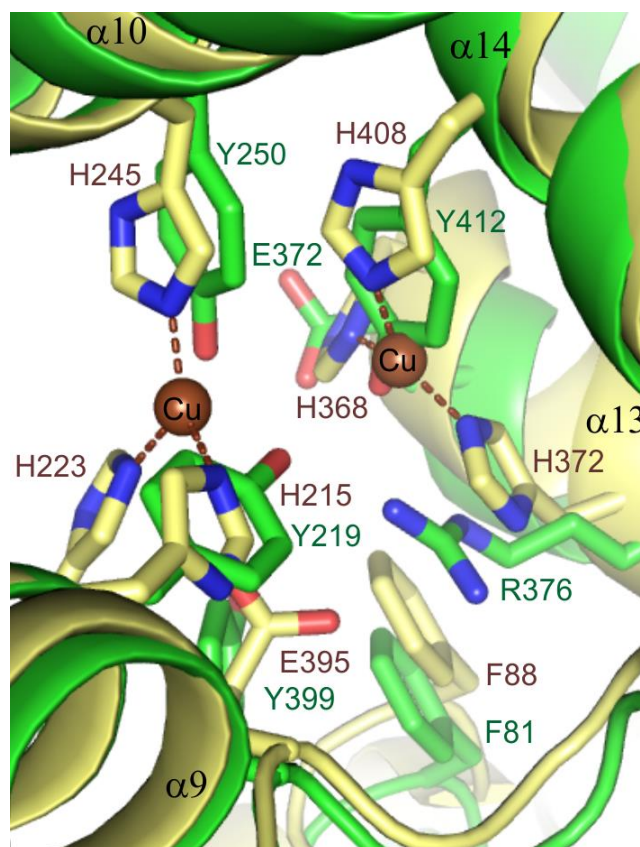


Fig. 6.16 A comparison of di-copper center of hemocyanins and phenoloxidases with the corresponding site of arylphorins.

In order to compare the di-copper site of hemocyanins and phenoloxidases with the corresponding site of arylphorins, the prophenoloxidase 2 (yellow; PDB code: 3HHS; Li et al., 2009) was Ca superposed with SP3 (green).

The biological role of proteins is strictly dependent on their primary structure. Arylphorins are silkworm storage proteins, they serve as the main supply of nitrogen during the pupal stage (Riddiford and Law, 1983), whereas hemocyanins are oxygen carriers in arthropods' bodies (van Holde and Miller, 1995) and phenoloxidases are enzymes catalyzing the *o*-hydroxylation of monophenols to catechols and oxidization of catechols to *o*-quinones (Decker *et al.*, 2006). Hemocyanins and phenoloxidases contain type 3 di-copper centers formed by two copper atoms coordinated by six histidine residues. The presence of this center is crucial in hemocyanins for oxygen binding (van Holde and Miller, 1995) and for the biocatalysis in phenoloxidases (Decker *et al.*, 2007). The di-copper site is located in the middle of the hemocyanin_M subdomain, between helices $\alpha 9$, $\alpha 10$, $\alpha 13$ and $\alpha 14$. The corresponding site of arylphorins contains residues incapable of copper binding: three tyrosine residues (Tyr219, Tyr250, Tyr412/SP3; Tyr222, Tyr253, Tyr415/SP2), one arginine (Arg376/SP3; Arg279/SP2), one glutamate (Glu372/SP3;

Glu376/SP2) and one histidine residue (His223/SP3; His226/SP2), which has completely different conformation than the corresponding residue of hemocyanins (Fig.6.16). Two additional residues are present in the active site of phenoloxidases: a phenylalanine and a glutamate, which act as a "place holder" for the phenolic substrates and the deprotonating agent of the hydroxyl group of monophenolic substrates, respectively. The corresponding residues of arylphorins are phenylalanine and tyrosine instead of glutamate.

6.7.2. The putative role of 30-kDa LPs domains deduced from structural comparison

As mentioned (6.6.2), the 30-kDa LPs fold, which is a combination of VHS and β -trefoil domains, is unique and to this date, there are no structural homologues in the PDB among other proteins beside this group. Only one 30-kDa LP crystal structure, which was not deposited by the author of this dissertation, is available in the PDB. It is a structure of recombinant Bmlp7 (PDB code: 3PUB; Yang *et al.*, 2011). The structure appeared in the PDB, when the manuscript describing the results of structural studies on Bmlp7-I(Cd), Bmlp7-Pt and Bmlp7-II was in preparation. The overall fold of the natural source Bmlp7 models and of recombinant Bmlp7 is almost identical. The 1.9 Å structure of recombinant Bmlp7 is of a good quality with R/R_{free} factors of 19.5 / 23.5% and displays 0.22 Å r.m.s.d. over 234 aligned C α atoms, when its chain A is superposed with chain A of Bmlp7-I(Cd). The recombinant Bmlp7 structure is isomorphous with the Bmlp7-I(Cd) and Bmlp7-Pt, the structures of the first Bmlp7 crystal form. Nevertheless, the structural studies on natural source Bmlp7 provided much more information, for instance the information about two Bmlp7 crystal forms, modified Trp residues and the cadmium binding site.

Further search for structural homologues was performed for the NTD and the CTD of 30-kDa LPs separately. The results are described in further sections.

6.7.2.1. The putative role of the 30-kDa LPs NTD

Interestingly, only in the case of the search carried out for the NTD, DALI server did not return any results, whereas the *PDBeFold* (SSM) server indicated that the most similar was the MA3-CTD of tumor suppressor Pdc4 (PDB code: 2ios; LaRonde-LeBlanc *et al.*, 2007). This domain is formed by seven α -helices and four of them could be superposed on helices α 1- α 4 of 30-kDa LPs NTD (Fig. 6.17a). This similarity might be important for the assignment of the biological role of the 30-kDa LPs NTD. Pdc4 interacts

6. Results & Discussion

with transcription-initiating factors (Yang *et al.*, 2001) and based on that it could be speculated that the NTD might be capable of interacting with other proteins, especially proteins involved in the apoptosis process. Other listed protein domains also belong to the VHS domain family. The summary of structural comparison for NTD, with r.m.s.d. below 3.0 Å, is shown in Table 6.6.

Table 6.6 R.m.s.d. values calculated for C α atoms of the 30-kDa LPs NTD and CTD superposed on homologues protein domains.

	2IOS	2DCP	3DWL	4FQ0	3VSF	2VSA	3AJ6
Bmlp7-I(Cd)- NTD	2.67 / 72	2.74 / 74	2.81 / 71	2.70 / 55	-	-	-
Bmlp7-I(Cd)- CTD	-	-	-	-	2.21 / 135	2.41 / 137	2.39 / 134

The table presents the r.m.s.d. values in Å / the number of superposed C α atoms for Bmlp7-I(Cd) NTD and CTD (chain A), as representative of 30-kDa LPs, with a structure representing each of the homologues, represented by their PDB codes: 2IOS (MA3-CTD of tumour suppressor Pdc4 from mouse; LaRonde-LeBlanc *et al.*, 2007), 2DCP (ENTH-VHS domain of hypothetical protein from *Arabidopsis thaliana*; Lopez-Mendez and Guntert, 2006), 3DWL (Actin-related protein 2/3 complex subunit 5 from yeast; Nolen and Pollard, 2008), 4FQ0 (flagellar motor switch protein from *Helicobacter pylori*; Lam *et al.*, 2013), 3VSF (exo- β -1,3-galactanase from *Clostridium thermocellum*; Jiang *et al.*, 2012), 2VSA (mosquitocidal holotoxin from *Bacillus sphaericus*; Treiber *et al.*, 2008) and 3AJ6 (main hemagglutinin component from *Clostridium botulinum*; Nakamura *et al.*, 2011).

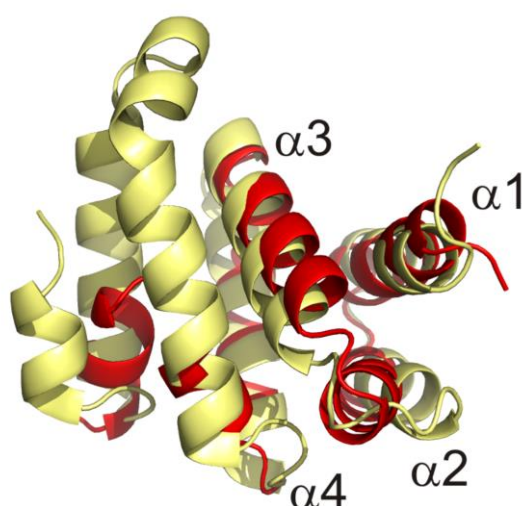


Fig. 6.17 A comparison of 30-kDa LPs NTD with MA3-CTD of tumour suppressor Pdc4.

The MA3-CTD of tumor suppressor Pdc4 (yellow; PDB code: 2IOS; LaRonde-LeBlanc *et al.*, 2007) was C α -superposed with the NTD of Bmlp7-I(Cd) (red).

6.7.2.2. The putative role of the 30-kDa LPs CTD

All homologues of the 30-kDa LPs CTD found by the servers belong to ricin-type β -trefoil lectin-like domain, the number of obtained results with r.m.s.d. below 2.0 Å was significant, but only the three top-ranked homologues are listed in Table 6.6. A detailed comparison of the 30-kDa LPs CTD with β -trefoil domains of other proteins was presented by Yang *et al.* (2011), therefore it would not be widely discussed in this dissertation. The most important conclusion coming from the analysis of the obtained results is that almost all β -trefoil homologues are able to bind carbohydrates. It is plausible that the 30-kDa LPs CTD is also capable of carbohydrate binding and this particular domain could be directly involved in immune response to fungal infections via β -glucan recognition.

6.8. PROTEIN-PROTEIN INTERACTIONS AND CRYSTAL PACKING

The biologically relevant oligomerization state of 30-kDa LPs and arylphorins was examined by gel filtration and DLS. All biochemical analyses indicated that 30-kDa LPs are monomeric in solutions. In contrast, arylphorins form hexamers which was confirmed biochemically and is consistent with other reports (Telfer *et al.*, 1991; Burmester and Scheller, 1996).

6.8.1. The heterohexamer of SP2 and SP3

The crystal structure of the SP2-SP3 complex confirmed that arylphorin exists in hexameric form (Fig. 6.18). However, SP2 forms a heterohexamer with SP3 and does not occur as a homohexamer, as it was previously believed (Fujii *et al.*, 1989). The complex was purified and crystallized from a mixture of three storage proteins in the form of a heterohexamer.

The buried surface area in the interface between all six subunits, three SP2 and three SP3 molecules, is 43 650 Å², as calculated using the PISA server (Krissinel and Henrick, 2007). The most extensive contacts between subunits are made by the hemocyanin_M subdomains (Fig. 6.18c,d), which are located in the central part of the hexamer, whereas the hemocyanin_N and hemocyanin_C subdomains are localized at the outside of the hexamer. Strong hydrophilic and hydrophobic interactions build the hexamer interface and the N-glycosylation characteristic to arylphorins additionally stabilizes the hexameric state.

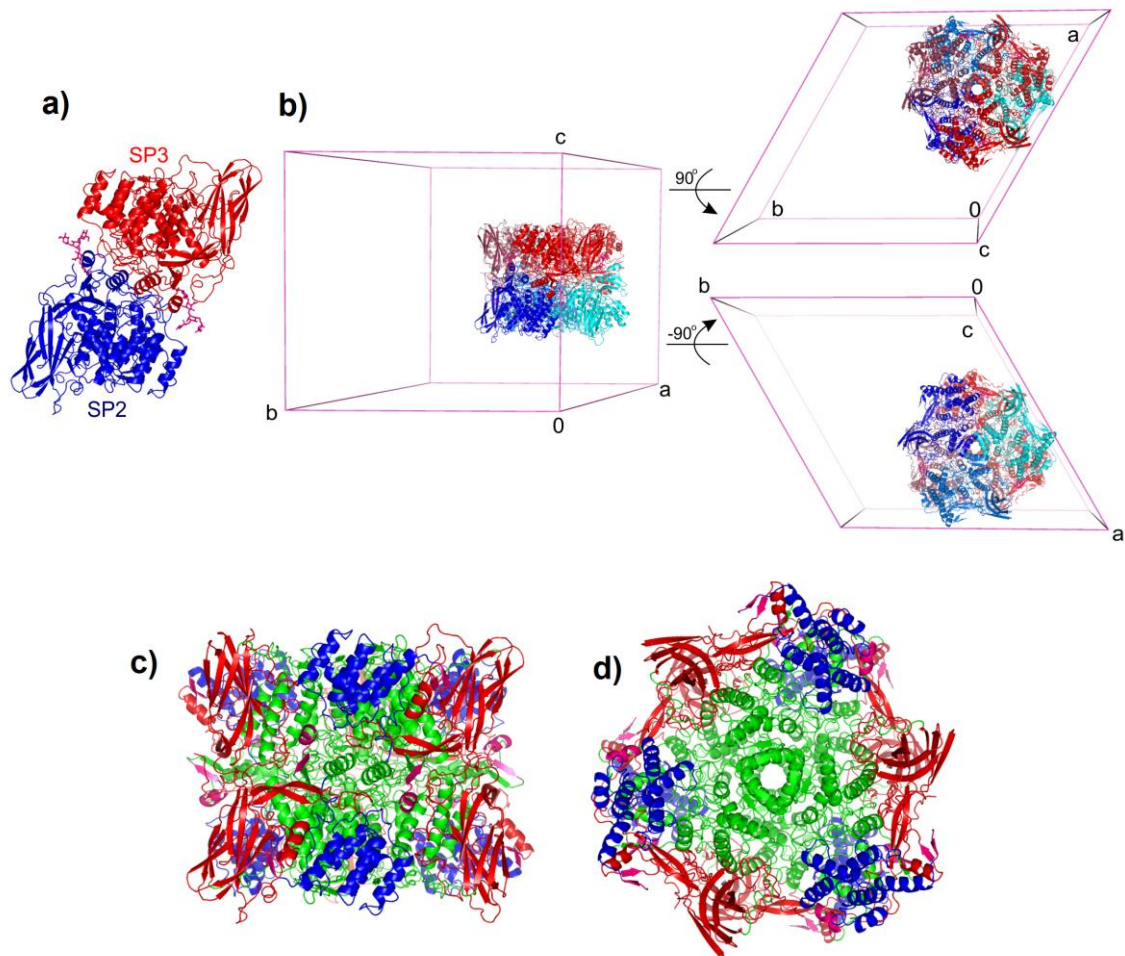


Fig. 6.18 A hexamer of SP2 and SP3.

a) The asymmetric unit of the SP2-SP3 complex, composed of two protein chains, SP2 (subunit B, blue) and SP3 (subunit A, red). **b)** A heterohexamer is presented in the unit cell; symmetry-related molecules of SP2 and SP3 are in the shades of blue and red, respectively. The rotation of the unit cell by 90° and -90° about the x axis is also shown. **c,d)** Two views of the hexamer with colored subdomains (hemocyanin_N - blue, hemocyanin_M - green, hemocyanin_C - red).

6.8.2. Crystal packing of 30-kDa LPs

Among the five investigated crystal structures of 30-kDa LPs, only two of them, Bmlp7-I(Cd) and Bmlp7-Pt, are isomorphous. Both structures contain two protein molecules in their asymmetric units (AU; Fig. 6.19a,b). The other structure with the same number of protein molecules in its AU is Bmlp3-c2 (Fig. 6.20b). However, the arrangement of the molecules in the AU of Bmlp3-c2 is different in comparison to the mentioned structures of Bmlp7. The molecules in these three structures are arranged in tail-to-tail fashion (Fig. 6.19a, 6.20b), where head and tail refer to NTD and CTD, respectively. The dimer interface is formed by the residues from loops between $\beta 2$ and $\beta 3$, $\beta 6$ and $\beta 7$, $\beta 10$ and $\beta 11$. The interactions are hydrophilic and the interface is also filled by a number of water

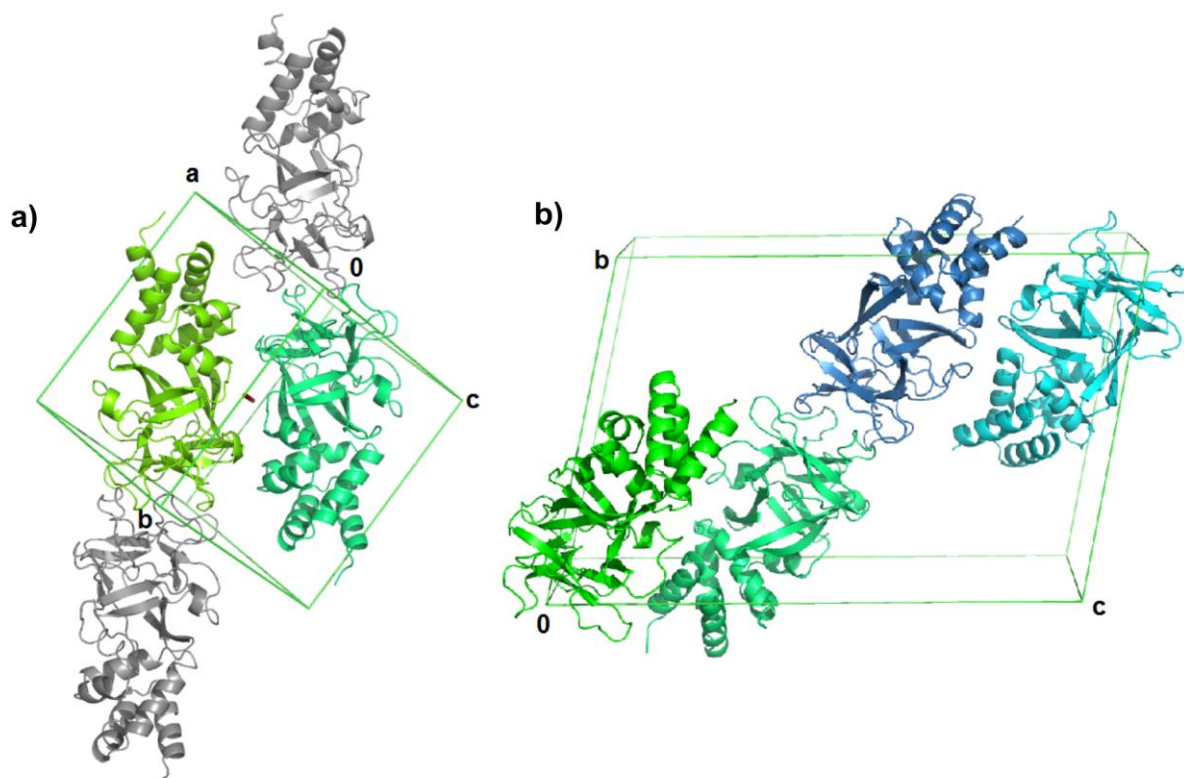


Fig. 6.19 A unit cell content of both Bmlp7 crystal forms.

a) A unit cell of Bmlp7-I(Cd). Due to the space group *P1*, two molecules filling the unit cell were chosen to represent the AU. However, the symmetry-related molecules (in gray) show that the molecules are packed in a tail-to-tail fashion. **b)** A unit cell of Bmlp7-II.

molecules. Nevertheless, the pattern of interactions in the Bmlp7 and the Bmlp3 dimers is different, for instance the loop containing residues 106–116 in molecule A of Bmlp3-c2 is shifted by about 5 Å compared to the same loop in Bmlp7-I(Cd) and the same loop in molecule B of Bmlp3-c2 is completely bent (Fig. 6.15). Such an arrangement of this loop in Bmlp3-c2 stabilizes the dimer more strongly than in the case of Bmlp7-I(Cd). Additionally, an iodide anion from the crystallization solution is located in the central part of dimerization interface on the non-crystallographic two-fold axis. The identity of the iodide anion was confirmed by the anomalous difference map (Fig. 6.21a) which was obtained after the X-ray diffractions were reprocessed with separation of the Bijvoet pairs. The data were collected at 0.918 Å wavelength where the f'' anomalous correction of iodine is 2.9e. The iodide anion interacts with two nitrogen atoms of the protein backbone (Gly158A/B) with the N...I distances of 3.7 and 3.9 Å and with three water molecules, with I...O distances of 3.2 to 3.7 Å (Fig. 21). The total contact surface area of Bmlp7-I(Cd) and Bmlp3-c2 is 1610 and 1620 Å²,

respectively, as calculated by the *PISA* server (Krissinel and Henrick, 2007). Although these values are close to each other, the server indicated that only the Bmlp3-c2 dimer could be stable in a solution. It is of note that *PISA* analyzes not only the area of surface contacts, but also the residue/atom composition, contacts between them, charge distribution, topological complementarity and other parameters; based on that the server assesses the complex stability from the point of view of thermodynamics (Krissinel and Henrick, 2007).

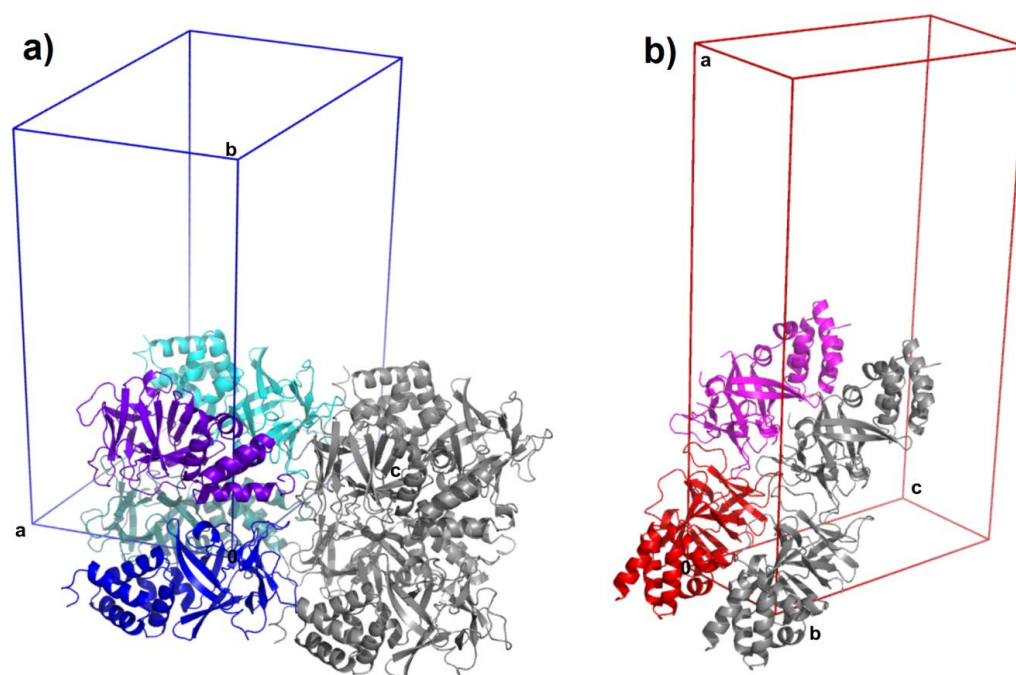


Fig. 6.20 A unit cell content of both Bmlp3 crystal forms.

a) A unit cell of Bmlp3-p21, four protein molecules (shades of blue) are shown in the AU. **b)** A unit cell of Bmlp3-c2, two protein molecules (red and pink) forming a dimer are shown in the AU. Symmetry-related molecules are shown in gray.

In contrast, the Bmlp7-II and Bmlp3-p21 structures contain four protein molecules in AU, but the arrangement of the molecules is different in both structures. In the Bmlp7-II structure, the molecules could be paired in dimers arranged in a tail-to-tail fashion (Fig. 6.19c), as in the above mentioned structures. The interaction pattern between CTDs is also completely different than this in Bmlp7-I(Cd) and Bmlp3-c2. The arrangement of the NTD and CTD fragments in Bmlp3-p21 could be described as head-to-tail, as the residues of the NTD have weak contacts with residues of the CTD (Fig. 6.20a).

Nevertheless, the calculations carried out using the *PISA* server (Krissinel and Henrick, 2007) indicated that the observed dimers or tetramers are only the result of crystal packing.

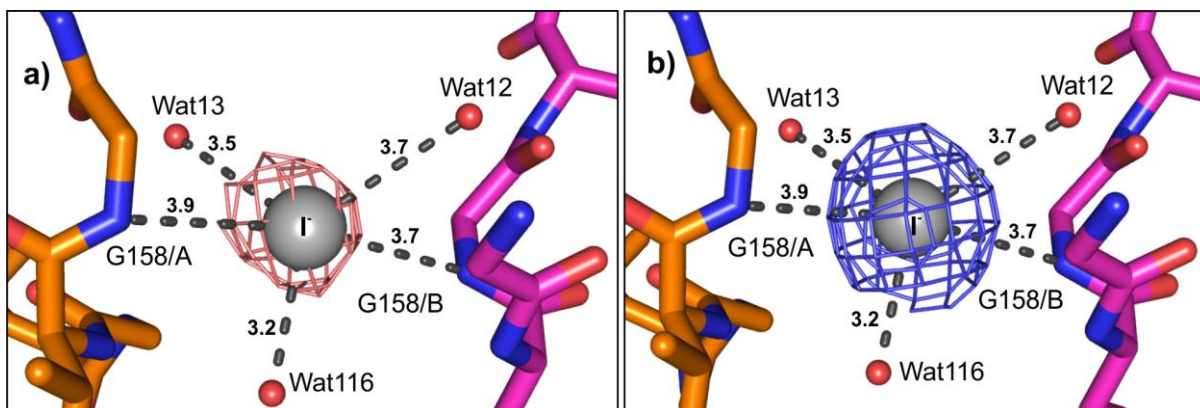


Fig. 6.21 The iodide anion binding site in Bmlp3-c2.

An iodide anion bound by the residues of chains A (orange) and B (magenta) of Bmlp3-c2. The N...I and O...I distances are in Å. **a)** The anomalous and **b)** the $2Fo-Fc$ maps are displayed at an iodine at 5.0σ and at 1.0σ , respectively.

The case of Bmlp3-c2 remains intriguing, because the biochemical analysis indicated that the protein is monomeric. However, the dimers formed in the Bmlp3-c2 crystals might also be stable in solution according to the *PISA* server (Krissinel and Henrick, 2007). It might be hypothesized that the 30-kDa LPs are monomeric in hemolymph, but they might dimerize at special, yet unknown conditions, for instance in silkworm fat body.

6.9. MODIFIED AMINO ACID RESIDUES

The proteins investigated within this project were isolated from their natural source. In the case of native macromolecules, X-ray crystallography reveals not only the 3D structure of proteins, but also provides information on post-translational modifications of the polypeptide chain. The electron density maps show the modified residues clearly. Additional phosphate, carbohydrate or other groups can be attached to amino acid residues, or the residues can undergo more radical modifications. The structural studies on hemolymph proteins provided the examples of two mentioned modifications, the molecules of SP2 and SP3 contain N-glycosylated residues, whereas one of the Trp residues of Bmlp7 was hydroxylated.

6.9.1. N-glycosylation motif of arylphorins

Although the resolution of the SP2-SP3 complex was not very high (2.9 Å), the electron density maps revealed the presence of a carbohydrate chain attached to Asn195 of SP2 and Asn192 of SP3. The localization of the N-glycan linked to chain A (SP3) in a deep cleft at the protomer interface, between molecules A, B and an adjacent molecule B from the same hexamer, caused five units of the N-glycan (two N-acetylglucosamine, one β -mannose and two α -mannose moieties) to be clearly visible in the electron density maps (Fig. 6.22). The N-glycan of chain B (SP2) is localized in a corresponding cleft, between molecules B, A and the adjacent A molecule. The sequence of the carbohydrate units built in the model was based on the experiments reported by Kim *et al.* (2003). The molecular weight of SP2 and SP3 determined by MALDI-TOF MS analysis, 84.2 kDa (Fig. 6.4a),

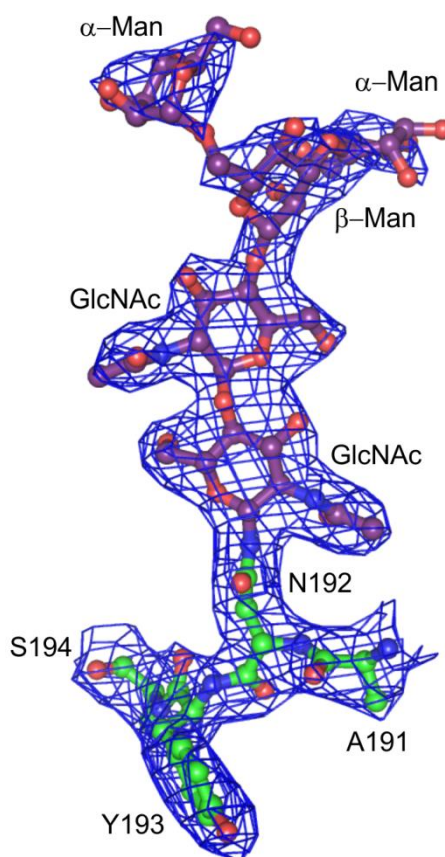


Fig. 6.22 The N-glycosylation of SP3.

The part of N-glucan attached to Asn192 and visible in the electron density maps is shown. The $2Fo-Fc$ map is displayed at 1.0σ .

is higher by about 2.0 kDa from the theoretical mass (Table 5.1) calculated for these proteins (SP2 - 81.8 kDa, SP3 - 81.6 kDa). The molecular weight of five saccharide units detected in the maps is about 1.0 kDa what indicates that the model of the complex contains half of the sugar residues of the N-glycan. The terminal sugar units were not visible in the maps due to their increased mobility. The described N-glycosylation motif is conserved in the arylphorins family and was also detected in APA, but not present in other types of storage proteins. Moreover, it increases the stability of the hexamer (Ryu *et al.*, 2009).

6.9.2. Modified Trp residue of Bmlp7

Thanks to the 1.3 Å high resolution electron density maps of a superb quality, a modification of Trp180/A/B in the Bmlp7-I(Cd) structure was detected (Fig. 6.23). A high positive peak in the difference electron density maps ($F_o - F_c$, 10.0σ) was visible close to the C7 atom of the Trp ring. The modification is also present in Bmlp7-Pt and Bmlp7-II. The most probable modification of the Trp residue was an oxidation of C7. Such a modified 7-hydroxy-Trp was also identified in other natural source proteins (Jensen *et al.*, 2010; Shima *et al.*, 2012). Moreover, the Trp oxidation could be a result of the action of oxidases present in the silkworm hemolymph; the oxidases are usually inhibited by the addition of 1-phenylthiourea during protein purification (Kramer *et al.*, 1976; Kurata *et al.*, 1994).

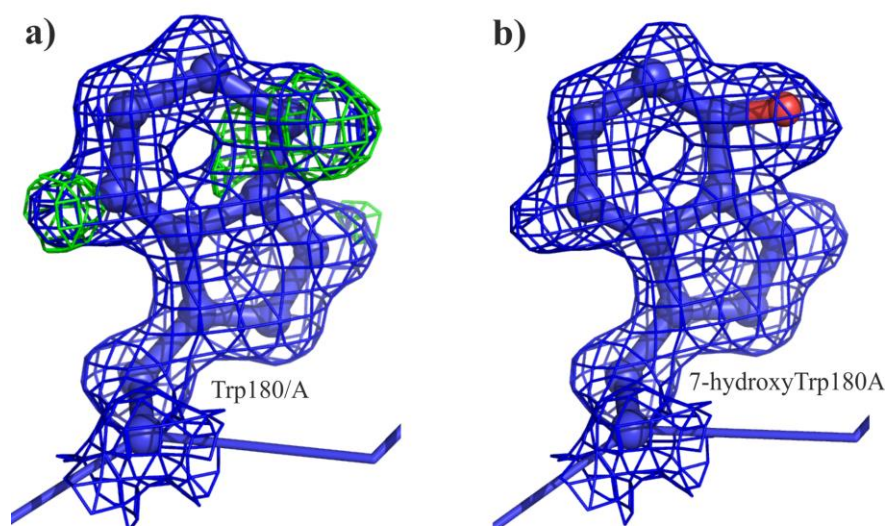


Fig. 6.23 The modified Trp residue of Bmlp7.

a) A positive peak in the electron density maps detected at C7 of Trp180 of Bmlp7. **b)** The most probable modification - oxidation of the Trp ring at C7. The $2F_o - F_c$ maps are displayed in blue at 1.3σ and the $F_o - F_c$ maps in green at 3.0σ .

The incorporation of the modified Trp improved the refinement statistics of the Bmlp7-I(Cd) structure, the R/R_{free} factors decreased by about 1.0 %. It seems that in Bmlp7 the loss of aromaticity of the six carbon ring of the indol ring accompanies the Trp oxidation. The hydroxyl group of 7-hydroxy-Trp180 is not planar which was indicated by the electron density maps. The hydrogen of the hydroxyl is a donor in a hydrogen bond formed with the main chain oxygen atom of Leu198. However, the shape of the electron density map for the indol ring of the modified residue is still flattened, what may suggest that the crystal consists of a mixture of Bmlp7 molecules containing 7-hydroxy-Trp180 and Bmlp7 molecules in which Trp180 was not modified.

6.10. THE PUTATIVE ROLE OF Bmlp7 IN DETOXIFICATION MECHANISM OF SILKWORM

The unexpected identification of a cadmium binding site in the Bmlp7-I(Cd) structure suggested that Bmlp7 could be involved in a detoxification mechanism related to heavy metal pollution. However, before this conclusion was made, several questions had to be answered, for instance: what kind of a heavy atom is present in the structure and where does it come from?

The presence of two heavy atom binding sites was revealed by two very high peaks (75.0σ) in the F_o-F_c maps. One of the peaks was located close to the Glu40 in molecule A and the other one was at the corresponding site in molecule B. It is of note that the crystals were not checked for the presence of cadmium during X-ray data collection, because it was not suspected that the native crystals could contain heavy metals. Nevertheless, the absorption edges of cadmium are located at the wavelengths of 0.4642 and 3.0857 Å (Kissel *et al.*, 1990), thus it would not be possible to detect this compound by a fluorescence scan at DESY or BESSY beam lines. Therefore, in order to identify the heavy atom present in the structure, several different heavy atoms were placed at the binding site and analyzed using the CheckMyMetal Server (http://csgid.org/csgid/metal_sites), and the bond valence test according to Mueller *et al.* (2003) was calculated. The results indicated that a cadmium atom fit very well into this site. Furthermore, the cadmium binding sites found in other structures from the PDB were also formed by glutamate residues, as it was in Bmlp7.

Each cadmium atom of Bmlp7-I(Cd) is bound by seven ligands and its coordination can be described as a distorted pentagonal bipyramid. The pentagonal base is formed by two glutamate side chains, Glu40A and Glu199B, from a symmetry-related monomer and by the main chain carbonyl of Lys37A (Fig. 6.24a). The axial ligands are a water

molecule and the nitrogen atom of the thiocyanate anion. An analogous site is located in monomer B.

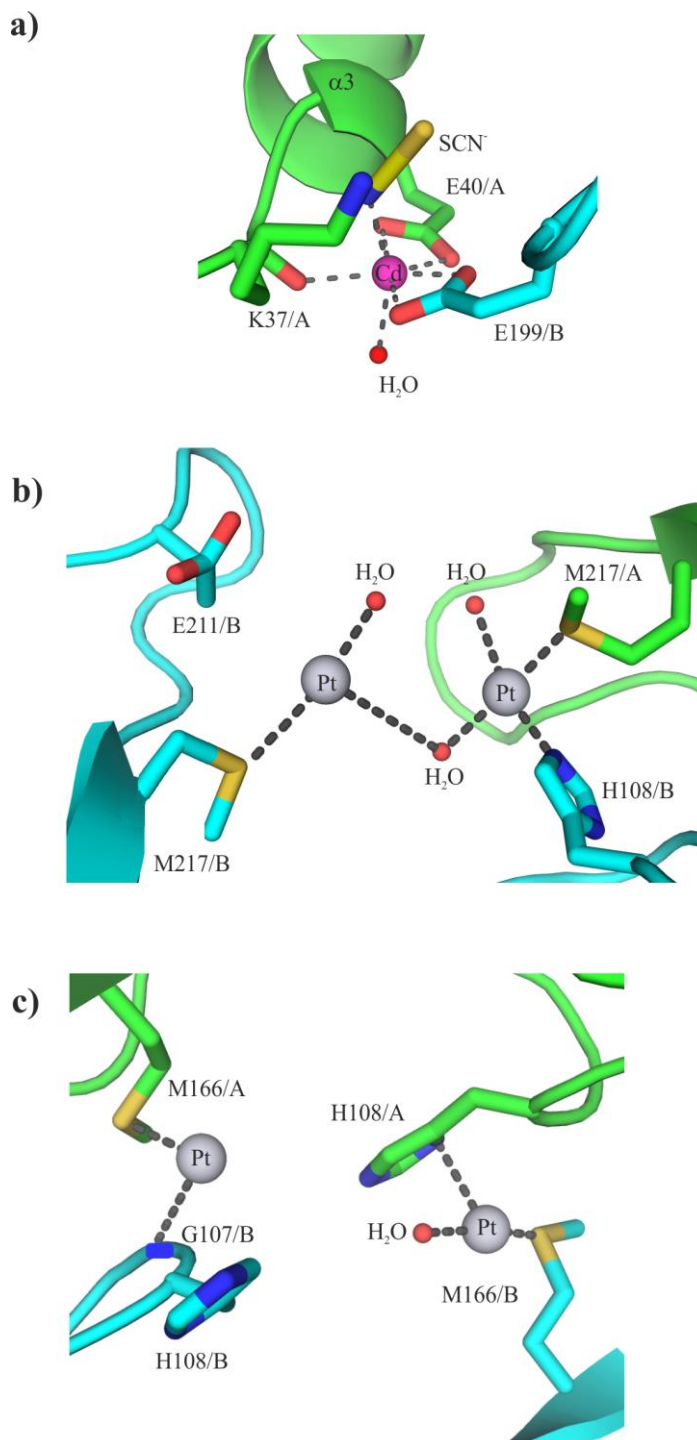


Fig. 6.24 The heavy metal binding sites of Bmlp7.

a) The cadmium binding site in the Bmlp7-I(Cd), the coordination sphere of Cd is formed by seven ligands. The cadmium binding site is formed by the residues belonging to a small loop between helices $\alpha 2$ and $\alpha 3$ and a loop between strands $\beta 11$ and $\beta 12$. **b,c)** Four additional platinum binding sites in the Bmlp7-Pt. These four sites are located at the loop between $\beta 2$ and $\beta 3$, at strand $\beta 7$ and at the loop between $\beta 10$ and $\beta 11$. The red balls represent water molecules.

6. Results & Discussion

The question of the cadmium's origin was the most intriguing. The purification buffers and crystallization solutions did not contain it. The well solution from crystallization plate and the crystallization drop of Bmlp7-I(Cd) were tested for the presence of cadmium using NMR measurements. Only in the sample from the crystallization drop a trace of cadmium was detected, what indicated that the source of cadmium must have been the insect material itself, namely the hemolymph. However, the Bmlp7-II crystals, which were obtained from Bmlp7 isolated from a batch of hemolymph collected in 2010, did not contain cadmium. Therefore, it seems that only the hemolymph collected in 2008 was contaminated with the heavy metal.

The final confirmation of cadmium identity was an experiment aimed at production of Bmlp7 crystals with artificially added cadmium. It was accomplished by the crystallization of Cd-free Bmlp7 protein with 0.2 mM of CdCl₂. The obtained crystals diffracted X-rays to 1.79 Å resolution and the Bmlp7-Cd structure was solved by MR. The Bmlp7-Cd structure was isomorphous with Bmlp7-I(Cd) and the coordination of cadmium atoms was identical in both structures (Table 6.7), which was the final confirmation of cadmium's identity in the Bmlp7-I(Cd).

Table 6.7 The comparison of cadmium coordination sphere in Bmlp7-I(Cd) and Bmlp7-Cd.

	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 ^a – OE1	2.35	2.35	2.40	2.41
Glu199 ^a – 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 ₂ O – O	2.34	2.50	2.52	2.54

^afrom symmetric molecule; the table presents distances between the cadmium cation and the coordinating ligand (Å).

The interaction of Bmlp7 with other heavy atoms was also an interesting issue and therefore the Bmlp7-Pt structure was solved. The platinum binding sites were compared with Cd sites of Bmlp7-I(Cd). Only two Pt sites corresponded to Cd sites. The Bmlp7-Pt structure contained four additional heavy atom sites, characteristic for platinum binding

due to the presence of methionine (Met166A/B and Met217A/B) and histidine residues (His108A/B) in the coordination spheres (Fig. 6.24b,c).

The link between the unexpected cadmium presence in the Bmlp7-I(Cd) and the biological sense of this discovery was established on the basis of reports about the silkworm ability to bioaccumulate heavy atoms (Prince *et al.*, 2001; Wang *et al.*, 2004). Especially, the fifth instar larvae is able to ingest mulberry leaves with very high cadmium content (Wang *et al.*, 2004). Cadmium was detected only in the crystals obtained from a batch of hemolymph collected in 2008 what suggested that in this year the silkworm larvae bioaccumulated cadmium. Probably, the initial source of this heavy metal were mulberry leaves. In 2008, there were heavy road construction works near the Institute of Natural Fibers and Medicinal Plants in Poznan, where the larvae were grown. The construction was close to the localization of the mulberry trees from which leaves were used for feeding the silkworms.

6.11. THE PUTATIVE BINDING CAVITIES OF 30-kDa LPs

30-kDa LPs are able to bind the physiological ligands belonging to at least two groups, lipids and carbohydrates. The 30-kDa proteins from silkworm hemolymph were classified as lipoproteins according to specific lipid staining (Withmore and Gilbert, 1974) which indicated their ability to bind lipids. On the other hand, 30-kDa LPs are involved in the immune response pathway; they recognize and bind β -glucans present on the surface of fungal cells (Ujita *et al.*, 2002; 2005). Nevertheless, hemolymph contains a number of different chemical compounds (Wyatt *et al.*, 1956) and it might be that 30-kDa LPs are able to interact with other, not mentioned, substances. Taking everything into account, it was concluded that the 30-kDa LPs should contain binding cavities for lipids, carbohydrates and maybe for other unknown compounds. The search for the putative binding sites was performed using the CASTp (Dundas *et al.*, 2006) and metaPocket 2.0 server (Huang, 2009; Zhang *et al.*, 2011). Initially, chain A of Bmlp3-p21 was used for the analysis, because no meaningful quaternary protein-protein interactions were indicated in this structure. Later on, a representative of Bmlp7, chain A of Bmlp7-I(Cd) was analyzed. The obtained results were identical for Bmlp3 and Bmlp7, therefore they will be discussed only for Bmlp3.

Four potential binding cavities were found by the servers and the information about them is summarized in Table 6.8. Two pockets are located in the CTD (No. 1 and No. 3), one in the NTD (No. 4) and one between the NTD and CTD (Fig. 6.25). The cavities were numbered according to their volume, from the largest to the smallest.

6. Results & Discussion

Table 6.8 The potential binding cavities of 30-kDa LPs.

Cavity number	Volume [$\text{\AA}^3$]	Location in the structure	Residues forming the cavity
1	399	CTD	Y94, R96, N160, N162, D164, T205, S206, R207, T208, W219, Y221, Y232
2	208	NTD-CTD	D21, D23, M55, N56, M58, E59, N138, K139, V140, Y141, F142, Y181, L182
3	110	CTD	D110, D111, V154, F168, V170, R177
4	60	NTD	I9, L10, Q13, L14, K29, L33

Cavity No. 1 might be either a lipid- or a sugar-binding site. It is built by residues from loops $\beta 1$ - $\beta 2$ and $\beta 6$ - $\beta 7$, and from the $\beta 11$ strand. The inner surface of the cavity is lined with three conserved tyrosine side chains (Tyr94, Tyr221, Tyr232). Tyrosine residues can be efficient in carbohydrate binding, but also their aromatic rings can interact with aliphatic fatty acid chains. In both structures, Bmlp7-I(Cd) and Bmlp3-p21, this pocket is occupied by a fragment of PEG from the crystallization solution (Fig. 6.25c,d). PEG is a good mimic of saccharides due to its ether/alcohol character. On the other hand, the elongated shape of the cavity and its large size indicate that it might also be a lipid-binding site.

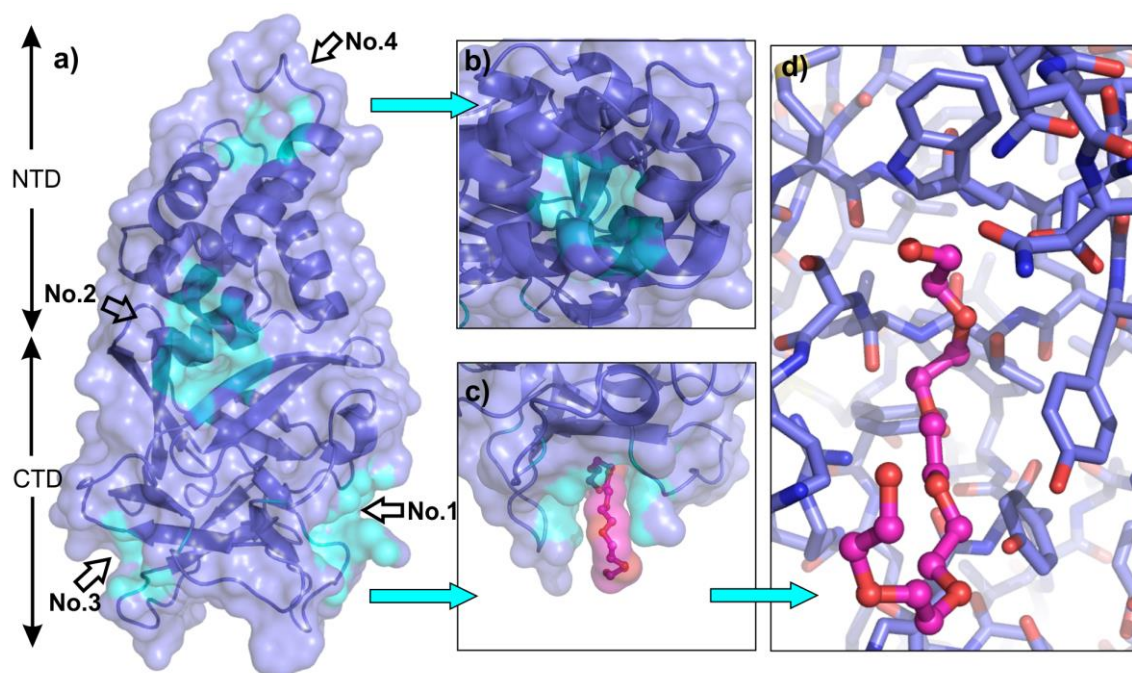


Fig. 6.25 The cavities of 30-kDa LPs.

a) The four cavities of Bmlp3 are marked by No. 1-4. **b)** Cavity No. 4 (at the top of molecule). **c,d)** Cavity No. 1 with a fragment of PEG bound.

Cavity No. 3, also situated in the CTD, has a hydrophobic interior (Val154, Phe168, Val170), but the entrance to this pocket is decorated with negatively charged aspartate residues (Asp110, Asp111). The volume of the cavity is rather small ($110 \text{ \AA}^3$), therefore it might be a binding site for small hydrophobic ligands with a hydrophilic terminal group.

The other pocket (No. 2), a potential carbohydrate binding site, is formed mainly by hydrophilic residues of loop $\alpha 1$ - $\alpha 2$, a helix $\alpha 4$ (NTD) and strands $\beta 5$ and $\beta 8$ (CTD). Few hydrophobic residues are also listed in Table 6.8, but only their backbone groups participate in cavity formation. The exceptions are two Tyr residues (Tyr141, Tyr181) which could interact with a sugar moiety. In the Bmlp3-p21 structure, water molecules are buried in this pocket.

The smallest cavity (No. 4) is localized in the NTD, between helices $\alpha 1$, $\alpha 2$ and $\alpha 3$, and built by hydrophobic residues. Although the volume of the pocket is only $60 \text{ \AA}^3$, it could be enlarged by a small movement of the helices. The cavity has an elongated shape and could play a role in lipid binding.

The analysis of the putative binding cavities gave indications about the potential complexes formed by 30-kDa LPs and their physiological ligands.

6.12. PROTEIN SURFACE OF Bmlp3 vs. CELL-PENETRATING PROPERTIES

The reports about cell-penetrating properties of a member of the 30-kDa LPs family (Park *et al.*, 2012) led to a question which structural elements of 30-kDa LPs could be responsible for these properties. In order to answer that question the structure of Bmlp3, namely Bmlp3-p21 as a representative, was further analyzed. Bmlp3 was chosen because it is an isoform of 30Kc19 for which the cell-penetrating properties were reported.

The analysis of the primary structure of Bmlp3 revealed that the protein contains a sequence motif ($_{116}\text{YGDGKDKTSPR}_{126}$) in consensus with the sequence fragments (YGRKKRRQRRR) responsible for cell-penetrating properties of HIV-TAT (Schwarze *et al.*, 2000). As the next step, the protein surface of Bmlp3 was analyzed. The Poisson-Boltzmann electrostatic potential on the molecular surface of chain A of Bmlp3-p21 was calculated using the APBS algorithm (Baker *et al.*, 2001) and the PDB2PQR program (Dolinsky *et al.*, 2004; 2007). All reported cell-penetration experiments were carried out in PBS buffer at pH 7.4 (Park *et al.*, 2012). Therefore, the electrostatic potential calculations were performed at this pH and prior to the analysis, the proper side chains protonation states of Bmlp3 were determined in PropKa (Li *et al.*, 2005).

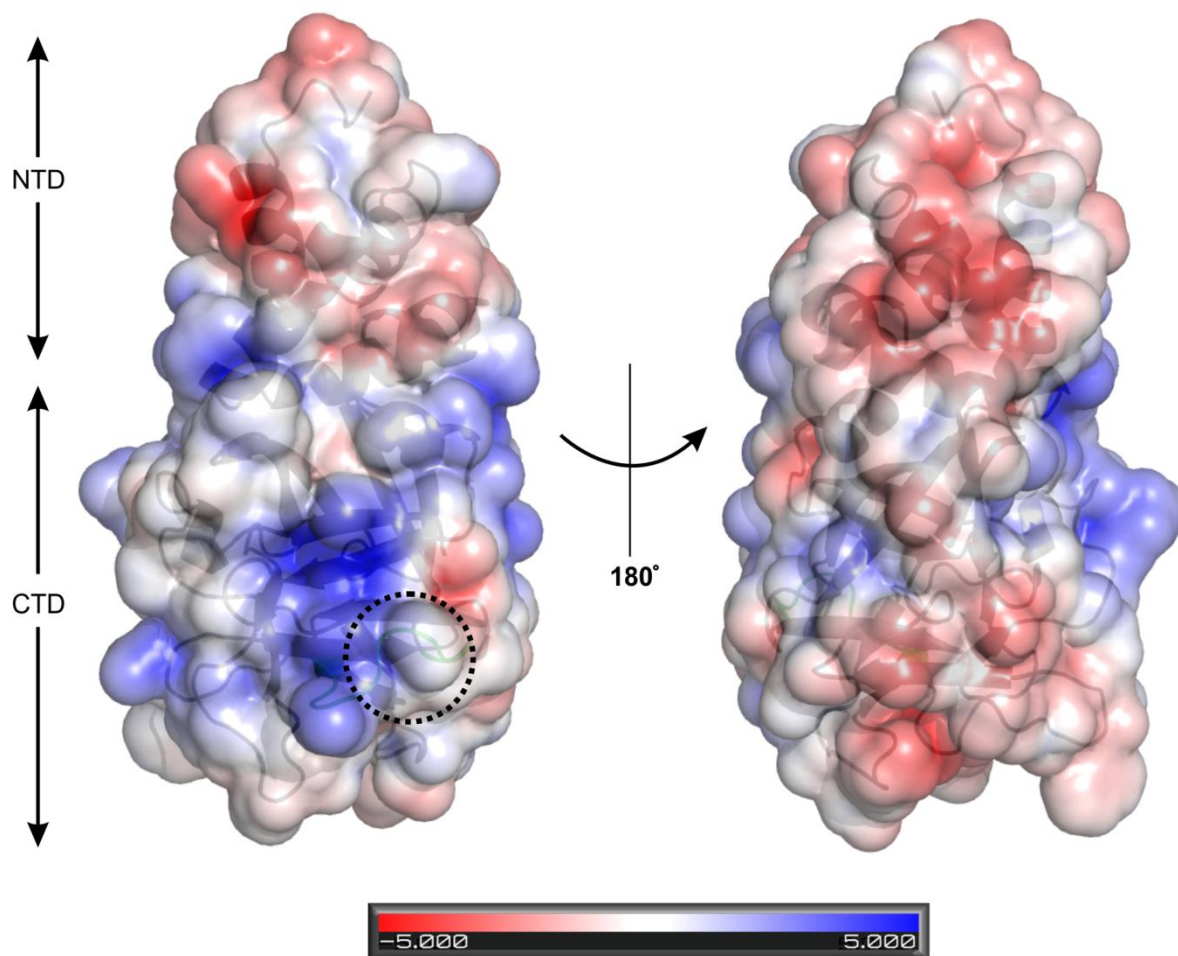


Fig. 6.26 The surface of Bmlp3.

The electrostatic potential calculated for Bmlp3-p21 at pH 7.4. The loop containing cell-penetrating signal motif was encircled.

The surface potential of the Bmlp3 NTD is mainly negative, whereas in the CTD positively and negatively charged areas cover two opposite sides of the domain (Fig. 6.26). The loop containing the cell-penetrating sequence motif is located at a positively charged patch of the CTD, between the $\beta 3$ and $\beta 4$ strands, and is exposed to the solvent. These results are consistent with the proposed mechanism of cell penetration, receptor-independent endocytosis (3.2.4). The residues of the positively charged loop $\beta 3$ - $\beta 4$ could interact with heparan sulfate of the cell membrane, initiating the penetration process.

The main achievements of this work are listed as follows:

- A purification procedure which allowed me to obtain samples containing four different hemolymph proteins (SP2, SP3, Bmlp7 and Bmlp3) was developed.
- The crystallization conditions which allowed me to obtain the crystals of a good quality were optimized.
- Six crystal structures of the four mentioned proteins were determined; the Bmlp7-I(Cd) structure was solved by SAD and five other structures (Bmlp7-Pt, Bmlp7-II, Bmlp3-p21, Bmlp3-c2, SP2-SP3) were solved by MR.
- The final identification of all investigated proteins which were isolated from hemolymph as unknown proteins was done according to the electron density maps obtained during an X-ray diffraction experiment. All cases show that X-ray crystallography could be extremely useful in solving sequence-related ambiguities.
- The analysis of silkworm arylphorins (SP2, SP3) structures and their comparison with hemocyanins enabled me to indicate secondary structure elements characteristic only for arylphorins and to demonstrate that the different roles of both protein groups are strictly connected with their primary structure.
- The unexpected discovery that SP2 and SP3 form a heterohexamer provided new information about silkworm hexamerins which were always considered to be homohexamers.
- The identification of SP3 as a component of the hexamerin complex confirmed that the protein exists in silkworm organism. Moreover, it is very abundant in silkworm hemolymph. The other reports concerning SP3 described only the gene of SP3 and its mRNA, identified during silkworm transcriptome studies.
- The analysis of Bmlp7 and Bmlp3 structures, or more precisely their NTD and CTD, followed by the search for their structural homologues allowed me to assign a potential biological role of the domains. The NTD is probably responsible for apoptosis inhibition, whereas the CTD plays a role in the immune response.
- The postranslational modifications were detected in SP2, SP3 and Bmlp7. Both arylphorins contain a characteristic N-glycosylation motif, whereas in the Bmlp7 one Trp residue is oxidized at the C7 atom of the indol ring.

7. Summary

- The unexpected cadmium identification in the Bmlp7-I(Cd) structure and further analysis related to this topic suggested that Bmlp7 might be involved in silkworm detoxification mechanism.
- Putative binding cavities of 30-kDa LPs were found, described and connected with potential physiological ligands.
- The analysis of Bmlp3 surface indicated how the protein can penetrate cells.

- Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.* D66, 213-221
- Allen, F.H. (2002). The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Cryst.* B58, 380-388.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403-410.
- Anderson, D.T. (1972). *Developmental System: Insects*. Vol. 1. London Academic Press, 96-242.
- Andres, A.J., Thummel, C.S. (1992). Hormones, puffs and flies: the molecular control of metamorphosis by ecdysone. *Trends Genet.* 8, 132-138.
- Arunkumar, K.P., Metta, M., Nagaraju, J. (2006). Molecular phylogeny of silkmoths reveals the origin of domesticated silkmoth, *Bombyx mori* from Chinese *Bombyx mandarina* and paternal inheritance of *Antheraea proylei* mitochondrial DNA. *Mol. Phyl. Evol.* 40, 419-427.
- Baker, N.A., Sept, D., Joseph, S., Holst, M.J., McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* 98, 10037-10041.
- Banumathi, S., Dauter, M., Dauter, Z. (2003). Phasing at high resolution using Ta6Br12 cluster. *Acta Cryst.* D59, 492-498
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Res.* 28, 235-242.
- Burmester, T., Scheller, K. (1996). Common origin of arthropod tyrosinase, arthropod hemocyanin, insect hexamerin, and dipteran arylphorin receptor. *J. Mol. Evol.* 42, 713-728.
- Burglin, T.R. (1994). A comprehensive classification of homeobox genes. *Guidebook to the homeobox genes*. Oxford University Press, 25-72.
- Butenandt, A., Karlson P. (1954). Über die isolierung eines metamorphose-hormons der insekten in kristallisierter form, *Z. Naturforsch.* B9, 389-391.
- Chang, K.H., Kim, K.S., Kim, J.H. (1999). N-acetylcysteine increases the biosynthesis of recombinant EPO in apoptotic Chinese hamster ovary cells. *Free Rad. Res.* 30, 85-91.
- Chapman, A.D. (2006). Numbers of living species in Australia and the World. *Australian Biological Resources Study* 60.
- Chen, Y.L., 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.

- Chino, H., Murakami, S., Harashima, K. (1969). Diglyceride-carrying lipoproteins in insect hemolymph: isolation, purification and properties. *Biochim. Biophys. Acta* 176, 1-26.
- Choi, S.S., Rhee, W.J., Park, T.H. (2002). Inhibition of human cell apoptosis by silkworm hemolymph. *Biotechnol. Prog.* 18, 874–878.
- Choi, S.S., Rhee, W.J., Park, T.H. (2005). Beneficial effect of silkworm hemolymph on a CHO cell system: Inhibition of apoptosis and increase of EPO production. *Biotechnol. Bioeng.* 91, 793-800.
- Cong, Y., Zhang, Q., Woolford, D., Schweikardt, T., Khant, H., Dougherty, M., Ludtke, S.J., Chiu, W., Decker, H. (2009). Structural mechanism of SDS-induced enzyme activity of scorpion hemocyanin revealed by electron cryomicroscopy. *Structure* 17, 749-758.
- Cotter, T.G., Lennons, S.V., Glynn, J.M., Green, D.R. (1992). Microfilament-disrupting agents prevent the formation of apoptotic bodies in tumor cells undergoing apoptosis. *Cancer Res.* 52, 997–1005.
- Cymborowski, B. (1984). Hormonalna regulacja wzrostu i metamorfozy owadów. *Endokrynologia owadów*. PWN.
- Dauter, Z., Dauter, M., Rajashankar, K.R. (2000). Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Cryst.* D56, 232-237.
- Decker, H., Schweikardt, T., Tuczek, F. (2006). The first crystal structure of tyrosinase: all questions answered? *Angew. Chem. Int. Ed.* 45, 4546–4550.
- Decker, H., Schweikardt, T., Nillius, D., Salzbrunn, U., Jaenicke, E., Tuczek, F. (2007). Similar activation process and catalysis in hemocyanins and tyrosinases. *Gene* 398, 183-191.
- Dolinsky, T.J., Nielsen, J.E., McCammon, J.A., Baker, N.A. (2004). PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* 32, W665–667.
- Dolinsky, T.J., Czodrowski, P., Li, H., Nielsen, J.E., Jensen, J.H., Klebe, G., Baker, N.A. (2007). PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* 35, W522–525.
- Drenth, J. (2007). *Principles of Protein X-Ray Crystallography*, Springer.
- Duan, J., Li, R., Cheng, D., Fan, W., Zha, X., Cheng, T., Wu, Y., Wang, J., Mita, K., Xiang, Z., Xia, Q. (2010). SilkDB v2.0: a platform for silkworm (*Bombyx mori*) genome biology. *Nucleic Acids Res.* 38, D453–D456.
- Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., Liang, J. (2006). CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acid Res.* 34, W116–W118.
- Emsley, P., Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Cryst.* D60, 2126–2132.

8. References

- Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K. (2010). Features and development of Coot. *Acta Cryst. D* 66, 486–501.
- Frankel, A.D., Pabo, C.O. (1988). Cellular uptake of the tat protein from human immunodeficiency virus. *Cell* 55, 1189e93.
- Fujii, T., Sakurai, H., Izumi, S., Tomino, S. (1989). Structure of the gene for the arylphorin-type storage protein SP 2 of *Bombyx mori*. *J. Biol. Chem.* 264, 11020-11025.
- Fujiwara, Y., Yamashita, O. (1992). Gene structure of *Bombyx mori* larval serum protein (BmLSP). *Insect Mol. Biol.* 1, 63–69.
- Fujiyuki, T., Imamura, K., Hamamoto, H., Sekimizu, K. (2010). Evaluation of therapeutic effects and pharmacokinetics of antibacterial chromogenic agents in a silkworm model of *Staphylococcus aureus* infection. *Drug Disc. Ther.* 4, 349-354.
- 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.
- Gilbert, L.I., Chino, H. (1974). Transport of lipids in insects. *J. Lipid. Res.* 15, 439-455.
- Grzimek, B., Kleiman, D.G., Geist, V., McDade, M.C. (2004). *Grzimek's Animal Life Encyclopedia*. Thomson-Gale.
- Goldsmith, M. R., Shimada, T., Abe, H. (2005). The genetics and genomics of the silkworm, *Bombyx mori*. *Annu. Rev. Ent.* 50, 71-100.
- Goswami, J., Sinskey, A.J., Steller, H., Stephanopoulos, G.N., Wang, D.I. (1999). Apoptosis in batch cultures of Chinese hamster ovary cells. *Biotechnol. Bioeng.* 62, 632-640.
- Green, M., Loewenstein, P.M. (1988). Autonomous functional domains of chemically synthesized human immunodeficiency virus tat trans-activator protein. *Cell* 55, 1179e88.
- Gupta, A.P. (1991). Insect immunocytes and other hemocytes: role in cellular and humoral immunity. *Immunology of Insects and Other Arthropods*, CRC Press, 19–118.
- Ha, S.H., Park, T.H., Kim, S.-E. (1996). Silkworm hemolymph as a substitute for fetal bovine serum in insect cell culture. *Biotechnol. Tech.* 10, 401-406.
- Hao, Q. (2004). ABS: a program to determine absolute configuration and evaluate anomalous scatterer substructure. *J. Appl. Cryst.* 37, 498-499.
- Hazes, B., Magnus, K.A., Bonaventura, C., Bonaventura, J., Dauter, Z., Kalk, K.H., Hol, W.G. (1993). Crystal structure of deoxygenated *Limulus polyphemus* subunit II hemocyanin at 2.18 Å resolution: clues for a mechanism for allosteric regulation. *Protein Sci.* 2, 597-619.
- Hendrickson, W.A., Smith J.L., Phizackerley, R.P., Merritt, E.A. (1988). Crystallographic structure analysis of lamprey hemoglobin from anomalous dispersion of synchrotron radiation. *Proteins* 4, 77–88.

8. References

- Hendrickson, W.A., Horton, J.R., LeMaster, D.M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* 9, 1665-1672.
- Hiratsuka, E. (1999). *Silkworm Breeding*, CRC Press.
- Holm, L., Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res.* 38, W545–W549.
- Hou, Y., Zou, Y., Wang, F., Gong, J., Zhong, X., Xia, Q., Zhao, P. (2010). Comparative analysis of proteome maps of silkworm hemolymph during different developmental stages. *Proteome Sci.* 8, 45.
- Huang, B. (2009). metaPocket: a meta approach to improve protein ligand binding site prediction. *Omics* 13, 325–330.
- Izumi, S., Fujie, J., Yamada, S., Tomino, S. (1981). Molecular properties and biosynthesis of major plasma proteins in *Bombyx mori*. *Biochem. et Biophys. Acta* 660, 222–229.
- Izumi, S., Kiguchi, K., Tomino, S. (1984). Hormonal regulation of biosynthesis of major plasma proteins in *Bombyx mori*. *Zoological Science* 1, 223–228.
- Jensen, L. M., Sanishvili, R., Davidson, V. L., Wilmot, C. M. (2010). In crystallo posttranslational modification within a MauG/pre-methylamine dehydrogenase complex. *Science* 327, 1392–1394.
- Jiang, H., Ma, C., Lu, Z.-Q., Kanost, M.R. (2004). β -1,3-glucan recognition protein-2 (β GRP-2) from *Manduca sexta*: an acute-phase protein that binds β -1,3-glucan and lipoteichoic acid to aggregate fungi and bacteria and stimulate prophenoloxidase activation. *Insect Biochem. Biol. Mol.* 34, 89–100.
- Jiang, D., Fan, J., Wang, X., Zhao, Y., Huang, B., Liu, J., Zhang, X.C. (2012). Crystal structure of 1,3-Gal43A, an exo- β -1,3-galactanase from *Clostridium thermocellum*. *J. Struct. Biol.* 180, 447-457.
- Kabsch, W. (2010a). XDS. *Acta Cryst.* D66, 125–132.
- Kabsch, W. (2010b). Integration, scaling, space-group assignment and post refinement. *Acta Cryst.* D66, 133–144.
- Kadono-Okuda, K., Yamamoto, M., Higashino, Y., Taniai, K., Kato, Y., Chowdhury, S., Xu, J., Choi, S.K., Sugiyama, M., Nakashima, K., et al. (1995). Baculovirus-mediated production of the human growth hormone in larvae of the silkworm, *Bombyx mori*. *Biochem. Biophys. Res. Commun.* 213, 389-396.
- Kamimura, M., Nakahara, Y., Kanamori, Y., Tsuzuki, S., Hayakawa, Y., Kiuchi, M. (2001). Molecular cloning of silkworm paralytic peptide and its developmental regulation. *Biochem. Biophys. Res. Commun.* 286, 67–73.
- Kaplan, I.M., Wadia, J.S., Dowdy, S.F. (2005). Cationic TAT peptide transduction domain enters cells by macropinocytosis. *J. Control. Release* 102, 247e53.

- Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., Phillips, D.C. (1958). A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature* 181, 662-666.
- Kim, N.S., Lee, G.M. (2001). Overexpression of bcl-2 inhibits sodium butyrate-induced apoptosis in Chinese hamster ovary cells resulting in enhanced humanized antibody production. *Biotechnol. Bioeng.* 71, 184–193.
- Kim, E.J., Rhee, W.J., Park, T.H. (2001). Isolation and characterization of an apoptosis-inhibiting component from the hemolymph of *Bombyx mori*. *Biochem. Biophys. Res. Commun.* 285, 224-228.
- Kim, S., Hwang, S.K., Dwek, R.A., Rudd, P.M., Ahn, Y.H., Kim, E.H., Cheong, C., Kim, S.I., Park, N.S., Lee, S.M. (2003). Structural determination of the N-glycans of a lepidopteran arylphorin reveals the presence of a monoglucosylated oligosaccharide in the storage protein. *Glycobiology* 13, 147-157.
- Kim, E.J., Rhee, W.J., Park, T.H. (2004). Inhibition of apoptosis by a *Bombyx mori* gene. *Biotechnol. Prog.* 20, 324–329.
- Kissel, L., Pratt, R. H. (1990). Corrections to tabulated anomalous-scattering factors. *Acta Cryst.* A46, 170–175.
- Klowden, M.J. (2007). *Physiological Systems in Insects*, Academic Press.
- Kochman, M.E., Wieczorek, E. (1994). Proteins involved in juvenile hormone signal transmission. *Insects - Chemical, physiological and environmental aspects*. University of Wrocław Press.
- Koolman, J. (1990). Ecdysteroids. *Zoological Science* 7, 563-580.
- Kramer, K. J., Dunn, P. E., Peterson, R. C., Seballos, H. L., Sanburg, L. L., Law, J. H. (1976). Purification and characterization of the carrier protein for juvenile hormone from the hemolymph of the tobacco hornworm *Manduca sexta* Johannson (*Lepidoptera: Sphingidae*). *J. Biol. Chem.* 251, 4979–4985.
- Krissinel, E., Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797.
- Kurata, K., Nakamura, M., Okuda, T., Hirano, H. Shinbo, H. (1994). Purification and characterization of a juvenile hormone binding protein from hemolymph of the silkworm, *Bombyx mori*. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 109, 105–114.
- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.
- Lam, K.H., Lam, W.W., Wong, J.Y., Chan, L.C., Kotaka, M., Ling, T.K., Jin, D.Y., Ottemann, K.M., Au, S.W. (2013). Structural basis of FliG-FliM interaction in *Helicobacter pylori*. *Mol. Microbiol.* 88, 798-812.

8. References

- LaRonde-LeBlanc, N., Santhanam, A. N., Baker, A. R., Wlodawer, A., Colburn, N. H. (2007). Structural basis for inhibition of translation by the tumor suppressor Pdc4. *Mol. Cell. Biol.* 27, 147–156.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26, 283–291.
- Levenbook, L., Bauer, A.C. (1984). The fate of the larval storage protein calliphorin during adult development of *Calliphora vicina*. *Insect Biochem.* 14, 77–86.
- Li, H., Robertson, A.D., Jensen, J.H. (2005). Very fast empirical prediction and rationalization of protein pKa values. *Proteins* 61, 704–721.
- Li, X.H., Wu, X.F., Yue, W.F., Liu, J.M., Li, G.L., Miao, Y.G. (2006). Proteomic analysis of the silkworm (*Bombyx mori* L.) hemolymph during developmental stage. *J. Proteome Res.* 5, 2809–2814.
- Li, J.Y., Chen, X., Moghaddam, S.H.H., Chen, M., Wei, H., Zhong, B.X. (2009a). Shotgun proteomics approach to characterizing the embryonic proteome of the silkworm, *Bombyx mori*, at labrum appearance stage. *Insect Mol. Biol.* 18, 649–660.
- Li, J.Y., Chen, X., Fan, W., Moghaddam, S.H., Chen, M., Zhou, Z.H., Yang, H.J., Chen, J.E., Zhong, B.X. (2009b). Proteomic and bioinformatic analysis on endocrine organs of domesticated silkworm, *Bombyx mori* L. for a comprehensive understanding of their roles and relations. *J. Proteome Res.* 8, 2620–2632.
- Li, Y., Wang, Y., Jiang, H., Deng, J. (2009c). Crystal structure of *Manduca sexta* prophenoloxidase provides insights into the mechanism of type 3 copper enzymes. *Proc. Natl. Acad. Sci. USA* 106, 17002–17006.
- Li, J., Hosseini, M.S.H., Chen, X., Chen, M., Zhong, B. (2010). Shotgun strategy-based proteome profiling analysis on the head of silkworm *Bombyx mori*. *Amino Acids* 39, 751–761.
- Li, Y., Wang, G., Tian, J., Liu, H., Yang, H., Yi, Y., Wang, J., Shi, X., Jiang, F., Yao, B., Zhang, Z. (2012). Transcriptome analysis of the silkworm (*Bombyx mori*) by high-throughput RNA sequencing. *PLoS ONE* 7, e43713.
- Locke, M., Collins, J.V. (1968). Protein uptake into multivesicullar bodies and storage granules in the fat body of insect. *J. Cell Biol.* 36, 453–483.
- Lopez-Mendez, B., Guntert, P. (2006). Automated protein structure determination from NMR spectra. *J. Am. Chem. Soc.* 128, 13112–13122.
- Magoshi, J., Magoshi, Y., Nakamura, S. (1994). Mechanism of fiber formation of silkworm. *Silk Polym.* 544, 292–310.
- 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.

- Mathavan, S., Gautvik, V.T., Rokkones, E., Olstad, O.K., Kareem, B.N., Maeda, S., Gautvik, K.M. (1995). High-level production of human parathyroid hormone in *Bombyx mori* larvae and BmN cells using recombinant baculovirus. *Gene* 167, 33-39.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J. (2007). Phaser crystallographic software. *J. Appl. Cryst.* 40, 658-674.
- McRee, D.E. (1999). *Practical Protein Crystallography*, Academic Press.
- Mita, K., Kasahara, M., Sasaki, S., Nagayasu, Y., Yamada, T., Kanamori, H., Namiki, N., Kitagawa, M., Yamashita, H., Yasukochi, Y., Kadono-Okuda, K., Yamamoto, K., Ajimura, M., Ravikumar, G., Shimomura, M., Nagamura, Y., Shin-I, T., Abe, H., Shimada, T., Morishita, S., Sasaki, T. (2004). The genome sequence of silkworm, *Bombyx mori*. *DNA Res.* 11, 27-35.
- 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.
- Moss, B., Flexner, C. (1987). Vaccinia virus expression vectors. *Annu. Rev. Immunol.* 5, 305-324.
- Mu ller, P., Kopke, S., Sheldrick, G. M. (2003). Is the bond-valence method able to identify metal atoms in protein structures? *Acta Cryst.* D59, 32-37.
- Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst.* D67, 355-367.
- Nakamura, T., Tono-zuka, T., Ito, S., Takeda, Y., Sato, R., Matsuo, I., Ito, Y., Oguma, K., Nishikawa, A. (2011). Molecular diversity of the two sugar-binding sites of the β -trefoil lectin HA33/C (HA1) from *Clostridium botulinum* type C neurotoxin. *Arch. Biochem. Biophys.* 512, 69-77.
- Nolen, B.J., Pollard, T.D. (2008). Structure and biochemical properties of fission yeast Arp2/3 complex lacking the Arp2 subunit. *J. Biol. Chem.* 283, 26490-26498.
- Ogawa, N., Kishimoto, A., Asano, T., Izumi, S. (2005). The homeodomain protein PBX participates in JH-related suppressive regulation on the expression of major plasma protein genes in the silkworm, *Bombyx mori*. *Insect Biochem. Mol. Biol.* 35, 217-29.
- Otwinowski, Z., Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307-326.
- Ozyhar, A., Kochman, M. (1987). Juvenile-hormone-binding protein from the hemolymph of *Galleria mellonella* (L). *Eur. J. Biochem.* 162, 675-682.
- Pakkianathan, B.C., Singh, N.K., 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.

8. References

- Panjikar, S., Parthasarathy, V., Lamzin, V.S., Weiss, M.S., Tucker, P. A. (2005). Auto-Rickshaw - An automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Cryst.* D61, 449–457.
- Pannu, N.S., McCoy A.J., Read, R.J. (2003). Application of the complex multivariate normal distribution to crystallographic methods with insights into multiple isomorphous replacement phasing. *Acta Cryst.* D59, 1801-1808.
- Park, J.H., Lee, J.H., Park, H.H., Rhee, W.J., Choi, S.S., Park, T.H. (2012). A protein delivery system using 30Kc19 cell-penetrating protein originating from silkworm. *Biomaterials* 33, 9127-9134.
- Paulino, A.T., Minasse, F.A.S., Guilherme, M.R., Reis, A.V., Muniz, E.C., Nozaki, J. (2006). Novel adsorbent based on silkworm chrysalides for removal of heavy metals from wastewaters. *J. Colloid. Interface Sci.* 301, 479-487.
- Paulino, A.T., Guilherme, M.R., Reis, A.V., Tambourgi, E.B., Nozaki, J., Muniz, E.C. (2007). Capacity of adsorption of Pb²⁺ and Ni²⁺ from aqueous solutions by chitosan produced from silkworm chrysalides in different degrees of deacetylation. *J. Hazard. Mat.* 147, 139–147.
- Perutz, M.F. (1956). Isomorphous replacement and phase determination in non-centrosymmetric space groups. *Acta Cryst.* 9, 867-873.
- Perrakis, A., Harkiolaki, M., Wilson, K.S., Lamzin, V.S. (2001). ARP/wARP and molecular replacement. *Acta Cryst.* D57, 1445-1450.
- Prince, S. P., Senthilkumar, P., Subburam, V. (2001). Mulberry-silkworm food chain: a templet to assess heavy metal mobility in terrestrial ecosystems. *Environ. Monit. Assess.* 69, 231–238.
- Ramirez-Chamond, R., Carracedo-Anon, J., Moreno-Aguilar, C., Guerra-Pasadas, F. (1999). Apoptosis and disease. *Alergol. Inmunol. Clin.* 14, 367-374.
- Rhee, W.J., Kim, E.J., Park, T.H. (1999). Kinetic effect of silkworm hemolymph on the delayed host cell death in an insect cell-baculovirus system. *Biotechnol. Prog.* 15, 1028-1032.
- Rhee, W.J., Park, T.H. (2000). Silkworm hemolymph inhibits baculovirus-induced insect cell apoptosis. *Biochem. Biophys. Res. Commun.* 271, 186-190.
- Rhee, W.J., Park, T.H. (2001). Flow cytometric analysis of the effect of silkworm hemolymph on the baculovirus-induced insect cell apoptosis. *J. Microbiol. Biotechnol.* 11, 853-857.
- Rhee, W.J., Kim, E.J., Park, T.H. (2002). Silkworm hemolymph as a potent inhibitor of apoptosis in Sf9 cells. *Biochem. Biophys. Res. Commun.* 295, 779-783.
- Rhee, W.J., Lee, E.H., Park, J.H., Lee, J.E., Park, T.H. (2007). Inhibition of HeLa Cell apoptosis by storage-protein 2. *Biotechnol. Prog.* 23, 1441-1446.
- Rhodes, G. (2000). *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models*, San Diego: Academic Press.

8. References

- Riddiford, L.M., Truman, J.W. (1978). *The Biochemistry of Insects*, New York: Academic Press.
- Riddiford, L.M., Law, J.H. (1983). *The Larval Serum Proteins of Insects: Function, Biosynthesis, Genetics*, ed. Scheller, K., Thieme, Stuttgart, 75-85.
- Riddiford, L.M. (1993). Hormone receptors and the regulation of insect metamorphosis. *Receptor* 3, 203-209.
- Ryu, K.S., Lee, J.O., Kwon, T.H., Choi, H.H., Park, H.S., Hwang, S.K., Lee, Z.W., Lee, K.B., Han, Y.H., Choi, Y.S., Jeon, Y.H., Cheong, C., Kim, S. (2009). The presence of monoglucosylated N196-glycan is important for the structural stability of storage protein, arylphorin. *Biochem. J.* 421, 87-96.
- Sakai, N., Mori, S., Izumi, S., Haino-Fukushima, K., Ogura, T., Maekawa, H., Tomino, S. (1988). Structures and expression of mRNAs coding for major plasma proteins of *Bombyx mori*. *Biochem. et Biophys. Acta* 949, 224–232.
- Sakurai, H., Fujii, T., Izumi, S., Tomino, S. (1988). Structure and expression of gene coding for sex-specific storage protein of *Bombyx mori*. *J. Biol. Chem.* 263, 7876-7880.
- 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.
- Schneider, T.R., Sheldrick, G. M. (2002). Substructure solution with SHELXD. *Acta Cryst.* D58, 1772-1779.
- Schultz, J., Milpetz, F., Bork, P., Ponting, C. P. (1998). SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl Acad. Sci. USA*, 95, 5857–5864.
- Schwarze, S.R., Hruska, K.A., Dowdy, S.F. (2000). Protein transduction: unrestricted delivery into all cells? *Trends Cell. Biol.* 10, 290e5.
- Sheldrick, G.M. (2002). Macromolecular phasing with SHELXE. *Z. Kristallogr.* 217, 644-650.
- Shima, S., Krueger, M., Weinert, T., Demmer, U., Kahnt, J., Thauer, R. K., Ermler, U. (2011). Structure of a methyl-coenzyme M reductase from Black Sea mats that oxidize methane anaerobically. *Nature* 481, 98–101.
- Shimada, T., Kobayashi, M., Yoshitake, N. (1985). Genetical analysis of storage proteins in the silkworm, *Bombyx mori*. *J. Seric. Sci. Jpn.* 54, 464-469.
- Sohn, K. W. (2003). Conservation Status of Sericulture Germplasm Resources in the World. *Food and Agriculture Organization of the United Nations*.
- Stout, G.H., Jensen, L.H. (1989). *X-ray Structure Determination: A Practical Guide*. New York: John Wiley & Sons, 424-426.

- Sumathy, S., Palhan, V.B., Gopinathan, K.P. (1996). Expression of human growth hormone in silkworm larvae through recombinant *Bombyx mori* nuclear polyhedrosis virus. *Protein Expr. Purif.* 7, 262-268.
- Sun, Q., Zhao, P., Lin, Y., Hou, Y., Xia, Q.-Y., Xiang, Z.-H. (2007). Analysis of the structure and expression of the 30 K protein genes in silkworm, *Bombyx mori*. *Insect Sci.* 14, 5-14.
- Telfer, W.H., Keim, P.S., Law, J.H. (1983). Arylphorin, a new protein from *Hyalophora cecropia*: comparison with calliphorin and manducin. *Insect Biochem.* 13, 601-613.
- Telfer, W.H., Kunkel, J.G. (1991). The function and evolution of insect storage hexamers. *Annu. Rev. Entomol.* 36, 205-228.
- Terashima, J., Yasuhara, N., Iwami, M., Sakurai, S., Sakurai, S. (2000). Programmed cell death triggered by insect steroid hormone, 20-hydroxyecdysone, in the anterior silk gland of the silkworm, *Bombyx mori*. *Dev. Genes Evol.* 210, 545-558.
- Terry, A.E., Knight, D.P., Porter, D., Vollrath, F. (2004). pH induced changes in the rheology of silk fibroin solution from the middle division of *Bombyx mori* silkworm. *Biomacromolecules*, 5, 768-772.
- Terwilliger, T.C. (2000). Maximum likelihood density modification. *Acta Cryst.* D56, 965-972.
- The International Silkworm Genome Consortium (2008). The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. *Insect Biochem. Mol. Biol.* 38, 1036-1045.
- Tojo, S., Nagata, M., Kobayashi, M. (1980). Storage proteins in the silkworm, *Bombyx mori*. *Insect Biochem.* 10, 289-613.
- Tojo, S., Liu, Y., Zheng, Y. (2012). Dynamics of storage proteins in Lepidoptera. *Hemolymph Proteins and Functional Peptides: Recent Advances in Insects and Other Arthropods*. Ed. Tufail, M., Bentham eBooks, 32-61.
- Tomita, M., Munetsuna, H., Sato, T., Adachi, T., Hino, R., Hayashi, M., Shimizu, K., Nakamura, N., Tamura, T., Yoshizato, K. (2003). Transgenic silkworms produce recombinant human type III procollagen in cocoons. *Nature Biotechnol.* 21, 52-56.
- Treiber, N., Reinert, D.J., Carpusca, I., Aktories, K., Schulz, G.E. (2008). Structure and mode of action of a mosquitocidal holotoxin. *J. Mol. Biol.* 381, 150-159.
- Tronrud, D.E. (2004). Introduction to macromolecular refinement. *Acta Cryst.* D60, 2156-2168.
- Truman, J.W., Riddiford, L.M. (1999). The origins of insect metamorphosis. *Nature* 401, 447-452.
- Ujita, M., Kimura, A., Nishino, D., Yokoyama, E., Banno, Y., Fujii, H., Hara, A. (2002). Specific binding of silkworm *Bombyx mori* 30-kDa lipoproteins to carbohydrates containing glucose. *Biosci. Biotechnol. Biochem.* 66, 2264e2266.

8. References

- Ujita, M., Katsuno, Y., Kawachi, I., Ueno, Y., Banno, Y., Fujii, H., Hara, A. (2005). Glucanbinding activity of silkworm 30-kDa apolipoprotein and its involvement in defense against fungal infection. *Biosci. Biotechnol. Biochem.* 69, 1178e1185.
- Vagin, A., Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* 30, 1022-1025.
- van Holde, K.E., Miller, K.I. (1995). Hemocyanins. *Adv. Protein Chem.* 47, 1–81.
- 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.
- Volbeda, A., Hol, W.G. (1989). Crystal structure of hexameric haemocyanin from *Panulirus interruptus* refined at 3.2 Å resolution. *J. Mol. Biol.* 209, 249-279.
- Wadia, J.S., Dowdy, S.F. (2002). Protein transduction technology. *Curr. Opin. Biotechnol.* 13, 52e6.
- Walker, N.I., Harmon, B.V., Gobe, G.C., Kerr, J.F.R. (1988). Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Met. Arch. Exp. Pathol.* 13, 18–54.
- Wang, X.Y., Cole, K.D., Law, J.H. (1989). The nucleotide sequence of a microvitellogenin encoding gene from the tobacco hornworm, *Manduca sexta*. *Gene* 80, 259-268.
- Wang, K.R., Gong, H., Wang, Y., van der Zee, S.E.A.T.M. (2004). Toxic effects of cadmium on *Morus alba* L. & *Bombyx mori* L. *Plant Soil*, 261, 171–180.
- Wang, J., Xia, Q., He, X., Dai, M., Ruan, J., Chen, J., Yu, G., Yuan, H., Hu, J., Li, R., Feng, T., Ye, C., Lu, C., Wang, J., Li, S., Wong, G.K., Yang, H., Wang, J., Xiang, Z., Zhou, Z., Yu, J. (2005). SilkDB: a knowledgebase for silkworm biology and genomics. *Nucleic Acids Res.* 33, D399–D402.
- Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J. (2009). Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189-1191.
- Whitmore, E., Gilbert, L.I. (1974). Hemolymph proteins and lipoproteins in *Lepidoptera*: a comparative electrophoretic study. *Comp. Biochem. Physiol.* 47B, 63-78.
- Winn, M.D., Isupov, M.N., Murshudov, G.N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Cryst.* D57, 122-33.
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A., Wilson, K.S. (2011). Overview of the CCP4 suite and current developments. *Acta Cryst.* D67, 235-242.

- Wyatt, G.R., Loughheed, T.C., Wyatt, S.S. (1956). The chemistry of insect hemolymph. *J. Gen. Physiol.* 39, 853-868.
- Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., Li, B., Zhao, P., Zha, X., Cheng, T., Chai, C., Pan, G., Xu, J., Liu, C., Lin, Y., Qian, J., Hou, Y., Wu, Z., Li, G., Pan, M., Li, C., Shen, Y., Lan, X., Yuan, L., Li, T., Xu, H., Yang, G., Wan, Y., Zhu, Y., Yu, M., Shen, W., Wu, D., Xiang, Z., Yu, J., Wang, J., Li, R., Shi, J., Li, H., Li, G., Su, J., Wang, X., Li, G., Zhang, Z., Wu, Q., Li, J., Zhang, Q., Wei, N., Xu, J., Sun, H., Dong, L., Liu, D., Zhao, S., Zhao, X., Meng, Q., Lan, F., Huang, X., Li, Y., Fang, L., Li, C., Li, D., Sun, Y., Zhang, Z., Yang, Z., Huang, Y., Xi, Y., Qi, Q., He, D., Huang, H., Zhang, X., Wang, Z., Li, W., Cao, Y., Yu, Y., Yu, H., Li, J., Ye, J., Chen, H., Zhou, Y., Liu, B., Wang, J., Ye, J., Ji, H., Li, S., Ni, P., Zhang, J., Zhang, Y., Zheng, H., Mao, B., Wang, W., Ye, C., Li, S., Wang, J., Wong, G.K., Yang, H. (2004). A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science* 306, 1937–1940.
- Yang, H.-S., Jansen, A.P., Nair, R., Shibahara, K., Verma, A.K., Cmarik, J.L., Colburn, N.H. (2001). A novel transformation suppressor, Pdcd4, inhibits AP-1 transactivation but not NF- κ B or ODC transactivation. *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). 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.
- Yao, L., Wang, S., Su, S., Yao, N., He, J., Peng, L., Sun, J. (2012). Construction of a Baculovirus-Silkworm Multigene Expression System and Its Application on Producing Virus-Like Particles. *PLoS ONE* 7, e32510.
- Yoon, S.K., Kim, S.H., Song, J.Y., Lee, G.M. (2006). Biphasic culture strategy for enhancing volumetric erythropoietin productivity of Chinese hamster ovary cells. *Enzyme Microb. Technol.* 39, 362-365.
- Zhang, Z., Li, Y., Lin, B., Schroeder, M., Huang, B. (2011). Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. *Bioinformatics* 27, 2083–2088.
- Zhang, X., Xue, R., Cao, G., Pan, Z., Zheng, X., Gong, C. (2012a). Silkworms can be used as an animal model to screen and evaluate gouty therapeutic drugs. *J. Insect Sci.* 12, 4.
- Zhang, Y., Dong, Z., Liu, S., Yang, Q., Zhao, P., Xia, Q. (2012b). Identification of novel members reveals the structural and functional divergence of lepidopteran-specific Lipoprotein_11 family. *Funct. Integr Genomics* 2, 705–715.
- Zhu, J., Indrasith, L.S., 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.

Table 3.1	Sequence homology among silkworm storage proteins.	12
Table 3.2	Sequence homology among silkworm 30-kDa LPs.	15
Table 5.1	Biophysical parameters calculated for the studied proteins.	23
Table 6.1	Crystals of silkworm proteins.	43
Table 6.2	X-ray diffraction data collection statistics of Bmlp7 crystals.	45
Table 6.3	X-ray diffraction data collection statistics of Bmlp3 and SP2-SP3 crystals.	46
Table 6.4	Refinement statistics of crystal structures of major hemolymph proteins.	57
Table 6.5	R.m.s.d. values calculated for C α atoms of mulberry silkworm storage proteins superposed on homologues proteins.	62
Table 6.6	R.m.s.d. values calculated for C α atoms of the 30-kDa LPs NTD and CTD superposed on homologues protein domains.	65
Table 6.7	The comparison of cadmium coordination sphere in Bmlp7-I(Cd) and Bmlp7-Cd.	75
Table 6.8	The potential binding cavities of 30-kDa LPs.	77

INTRODUCTION

Fig.3.1	The full life cycle of mulberry silkworm.	5
Fig.3.2	Insect hormones.	6
Fig.3.3	The fate of lipoproteins in silkworm body.	10
Fig.3.4	Phylogenetic tree of 30-kDa lipoprotein family.	16

MATERIALS & METHODS

Fig.5.1	Hemolymph collection.	21
Fig.5.2	The structure factor.	28
Fig.5.3	Anomalous scattering.	30
Fig.5.4	The absorption edges of mercury.	31

RESULTS & DISCUSSION

Fig.6.1	The scheme of purification procedure of silkworm hemolymph proteins.	38
Fig.6.2	Purification of the silkworm hemolymph proteins.	39
Fig.6.3	SDS-PAGE electrophoregrams monitoring purification steps.	40
Fig.6.4	MALDI-TOF MS spectra for investigated proteins.	41
Fig.6.5	Crystals of silkworm proteins.	44
Fig.6.6	Sequence alignment of 30-kDa LPs.	50
Fig.6.7	The high resolution electron density maps of Bmlp7.	51
Fig.6.8	The electron density maps of Bmlp3.	52
Fig.6.9	Sequence alignment of arylphorins.	53
Fig.6.10	The electron density maps of SP2-SP3.	54
Fig.6.11	The results of N-terminal sequencing of SPs samples.	55
Fig.6.12	The Ramachandran plot.	56
Fig.6.13	Overall structure of SP2 and SP3.	58
Fig.6.14	Overall structure of 30-kDa LPs.	60
Fig.6.15	Structural comparison of Bmlp3-p21, Bmlp3-c2 and Bmlp7.	61
Fig.6.16	A comparison of di-copper center of hemocyanins and phenoloxidases with the corresponding site of arylphorins.	63
Fig.6.17	A comparison of 30-kDa LPs NTD with MA3-CTD of tumour suppressor Pdcd4.	65
Fig.6.18	A hexamer of SP2 and SP3.	67

10. List of Figures

Fig.6.19	A unit cell content of both Bmlp7 crystal forms.	68
Fig.6.20	A unit cell content of both Bmlp3 crystal forms.	69
Fig.6.21	The iodide anion binding site in Bmlp3-c2.	70
Fig.6.22	The N-glycosylation of SP3.	71
Fig.6.23	The modified Trp residue of Bmlp7.	72
Fig.6.24	The heavy metal binding sites of Bmlp7.	74
Fig.6.25	The cavities of 30-kDa LPs.	77
Fig.6.26	The surface of Bmlp3.	79

The results obtained within this work have been published in five scientific articles and have been presented at a number of congresses, as listed below.

PUBLICATIONS

- **Pietrzyk A.J.**, 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.* **F67**, 372-376.
- **Pietrzyk A.J.**, Panjekar S., Bujacz A., Mueller-Dieckmann J., Lochynska M., Jaskolski M., Bujacz G. (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.* **D68**, 1140-1151.
- **Pietrzyk A.J.**, Bujacz A., Mueller-Dieckmann J., Lochynska M., Jaskolski M., Bujacz G. (2013). Two crystal structures of *Bombyx mori* lipoprotein 3 - structural characterization of a new 30-kDa lipoprotein family member. *PLoS ONE* **8**, e61303.
- **Pietrzyk A.J.**, Bujacz A., Jaskolski M., Bujacz G. (2013). Identification of amino acid sequence by X-ray crystallography: a mini review of case studies. *BioTechnologia - J. Biotech. Comp. Biol. Bionanotech.* **94**, 9-14.
- **Pietrzyk A.J.**, Bujacz A., Mueller-Dieckmann J., Lochynska M., Jaskolski M., Bujacz G. (2013). Crystallographic identification of an unexpected protein complex in silkworm hemolymph. *Acta Cryst.* **D69**, 2353-2364.

CONGRESSES

- **Pietrzyk, A.J.**, Bujacz, A., Lochynska, M., Jaskolski, M. and Bujacz, G. (2010). Structural analysis of hemolymph proteins from mulberry silkworm *Bombyx mori*. 52nd Crystallographic Meeting (poster presentation).
- **Pietrzyk, A.J.**, Bujacz, A., Lochynska, M., Jaskolski, M. and Bujacz G., (2010). Isolation, purification and preliminary crystallographic studies of two major proteins from mulberry silkworm (*Bombyx mori* L.) hemolymph. *Acta Biochim. Polonica* **57** Nr 4, Supplement 4, P1.27. 45th Polish Biochemical Society Meeting (poster presentation).
- Bujacz, G., **Pietrzyk, A.J.**, Bujacz, A. and Jaskólski, M. (2011). Badania strukturalne białek hemolimfy jedwabnika morwowego. 53rd Crystallographic Meeting.
- **Pietrzyk, A.J.**, Bujacz, A., Panjekar, S., Lochynska, M., Jaskolski, M. and Bujacz, G. (2011). Novel fold in 30 kDa protein from silkworm hemolymph. 53rd Crystallographic Meeting (poster presentation).
- Bujacz, G., **Pietrzyk, A.J.**, Bujacz, A., Panjekar, S. and Jaskólski, M. (2011). Structural investigation of major *Bombyx mori* hemolymph proteins. *Acta Biochim. Polonica* **58** Nr 2, Supplement 2, L14.2. 46th Polish Biochemical Society Meeting.
- **Pietrzyk, A.J.**, Bujacz, A., Mueller-Dieckmann, J., Lochynska, M., Jaskolski, M. and Bujacz G. (2011). Large-scale screening for crystallization conditions of major

hemolymph proteins from mulberry silkworm *Bombyx mori* L. *Acta Biochim. Polonica* **58** Nr 2, Supplement 2, P14.1. 46th Polish Biochemical Society Meeting (poster presentation).

- **Pietrzyk, A.J.**, Bujacz, A., Panjekar, S., Mueller-Dieckmann, J., Jaskolski, M. and Bujacz G. (2011). Three crystal structures of *Bombyx mori* lipoprotein 7 (Bmlp7) – interactions with different metal cations. 14th Heart of Europe bio-Crystallography Meeting (oral presentation).
- **Pietrzyk, A.J.**, Bujacz, A., Panjekar, S., Mueller-Dieckmann, J., Jaskolski, M. and Bujacz G. (2011). Biocrystallography as a tool in protein identification – the case of *Bombyx mori* lipoprotein 7 (Bmlp7) structure. Multi-Pole Approach to Structural Biology (poster presentation).
- **Pietrzyk, A.J.**, Jaskolski, M. and Bujacz G. (2012). High-resolution crystal structure of *Bombyx mori* lipoprotein 7 (Bmlp7): its metal binding sites and modified residues. New Frontiers in Structural Biology (oral presentation).
- **Pietrzyk, A.J.**, Bujacz, A., Panjekar, S., Mueller-Dieckmann, J., Lochynska, M., Jaskolski, M. and Bujacz G. (2012). *Bombyx mori* lipoprotein 7 as an example of successful use of biocrystallography for protein identification. Present and Future Methods for Biomolecular Crystallography (poster presentation).
- **Pietrzyk, A.J.**, Bujacz, A., Mueller-Dieckmann, J., Lochynska, M., Jaskolski, M. and Bujacz G. (2012). Crystallographic studies of 30 kDa lipoproteins from silkworm hemolymph. 54rd Crystallographic Meeting (oral presentation, selected from abstracts).
- **Pietrzyk, A.J.**, Bujacz, A., Mueller-Dieckmann, J., Lochynska, M., Jaskolski, M. and Bujacz G. (2012). Screening for physiological ligands of *Bombyx mori* lipoprotein 7. *Acta Biochim. Polonica* **59** Supplement 3, P2.56. Polish-German Biochemical Societies Joint Meeting / 47th Congress of the Polish Biochemical Society (poster presentation).
- **Pietrzyk, A.J.** (2012). HTP crystallization of silkworm proteins. 3rd P-CUBE User Meeting (poster and oral presentation).
- **Pietrzyk, A.J.**, Panjekar, S., Bujacz, A., Mueller-Dieckmann, J., Lochynska, M., Jaskolski, M., Bujacz, G. (2012). High-resolution crystal structure of *Bombyx mori* lipoprotein 7. 4th Joint BER II and BESSY II User Meeting (poster presentation).
- **Pietrzyk, A.J.** (2013). Structural studies on 30-kDa lipoproteins from mulberry silkworm. Biomolecules and Nanostructures 4 (oral presentation, selected from abstracts).
- **Pietrzyk, A.J.** (2013). Sequencing from electron density maps - crystal structures of silkworm proteins as case studies. International Conference Structural Biology of Plants and Microbes (oral presentation).
- **Pietrzyk, A.J.** (2013). An unexpected complex of arylphorins in silkworm hemolymph. 55rd Crystallographic Meeting (oral presentation, selected from abstracts).
- **Pietrzyk, A.J.**, Bujacz, G. (2013). A complex of two arylphorins – an example of sequencing from electron density maps. 16th Heart of Europe bio-Crystallography Meeting (oral presentation).

APPENDIX - PUBLICATIONS

- I **Pietrzyk A.J.**, 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.* **F67**, 372-376.
- II **Pietrzyk A.J.**, Panjekar S., Bujacz A., Mueller-Dieckmann J., Lochynska M., Jaskolski M., Bujacz G. (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.* **D68**, 1140-1151.
- III **Pietrzyk A.J.**, Bujacz A., Mueller-Dieckmann J., Lochynska M., Jaskolski M., Bujacz G. (2013). Two crystal structures of *Bombyx mori* lipoprotein 3 - structural characterization of a new 30-kDa lipoprotein family member. *PLoS ONE* **8**, e61303.
- IV **Pietrzyk A.J.**, Bujacz A., Jaskolski M., Bujacz G. (2013). Identification of amino acid sequence by X-ray crystallography: a mini review of case studies. *BioTechnologia - J. Biotech. Comp. Biol. Bionanotech.* **94**, 9-14.
- V **Pietrzyk A.J.**, Bujacz A., Mueller-Dieckmann J., Lochynska M., Jaskolski M., Bujacz G. (2013). Crystallographic identification of an unexpected protein complex in silkworm hemolymph. *Acta Cryst.* **D69**, 2353-2364.