The Structure and Function of a Bovine Salivary Protein - BSP30b

A thesis submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Biological Sciences

at

The University of Waikato

by

Heng Zhang

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2018

Abstract

The Bovine Salivary Protein 30b (BSP30b) is a member of the palate, lung, and nasal epithelium clone (PLUNC) family which includes human bactericidal/permeability-increasing protein (BPI) and lipopolysaccaride binding protein (LBP). BSP30b is predominantly found in bovine saliva but its function is not understood. Although PLUNC family members are proposed to be lipid binding proteins, in most cases their lipid ligands are also unknown.

Here, the X-ray crystal structure of BSP30b is solved at 2 Å resolution using a double methionine mutant and Se-Met MAD phasing. Each molecule adopts a slightly curved cylindrical structure with a hydrophobic channel formed by an α/β wrap, which is highly conserved in the PLUNC family. The structure of BSP30b complexed with oleic acid is also presented where the ligand is accommodated within the hydrophobic channel. The electron density for oleic acid suggests that the ligand is only partially occupied in the binding site implying that oleic acid might not be the preferred ligand of BSP30b.

Nonetheless, GFP-tagged BSP30b binds to the surface of olive oil droplets, as observed by fluorescent microscopy. Bacteria extracted directly from rumen samples indicate that the GFP_BSP30b fusion protein binds to a small number of selected bacterial strains *in vivo*. These results suggest that BSP30b could have a specific range of lipid ligands and may play important roles via its interaction with select rumen bacteria.

Acknowledgements

First and foremost, I would like to thank my primary supervisor, Professor Vic Arcus for the opportunities you have provided. Your guidance, expertise and enthusiasm have been invaluable in this journey. Also to Dr Ray Cursons, thank you for all your expertise and help as a secondary supervisor.

Thank you to all the members in the Waikato university Proteins and Microbes lab - Dr Emma Andrews, Dr Judith Burrows, Dr Emma Summer, Dr Jo Hicks, Dr Erica Prentice, Claire Mulholland, Brooke Dillon, Dr Vikas Chonira, Kirsty Mayall, Dr Chelsea Vickers, Kirsty Kraakman, Emily Grout, Mitchell Murray, Daniel Schipper, and Chris Batterton. Thank you all for your support during my PhD study, especially those brilliant suggestions from lab meetings with you guys. A special thank you to Judith who gave me a load of help.

Dr Greg Jacobson and Emily Grout, thank you for taking the time to read chapters of this thesis.

To Dr Graeme Attwood at AgResearch in Palmerston North, thank you for giving me the opportunity to learn from you and your group.

Thank you to DairyNZ for providing the funding to support me through this research.

To friends and family, thank you for your continued support throughout my studies. To my parents, family, and friends, thank you for providing the work life balance and your encouragement.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Aa	amino acid
ASL	airways surface liquid
bp	base pair(s)
BPIF	Bactericidal/Permeability-Increasing Protein Fold–Containing Family
BSP30b	bovine salivary protein 30b
°C	degrees Celsius
CA6	carbonic anhydrase 6
CETP	the cholesteryl ester transfer protein
CF	cystic fibrosis
CSN1S1	casein 1S1
CSTB	cystatin B or stefin B
C-terminal	carboxyl terminus of peptide chain
3D	three dimensional
Da	Daltons
DNA	deoxyribonucleic acid

EDTA	ethylene diamine tetraacetic acid
ENaC	epithelial Na^+ channel
E-Syts	extended synaptotagmins
FPLC	fast performance liquid chromatography
G	times the force of gravity
HBB	haemoglobin subunit beta
HDL	high-density lipoproteins
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His tag	poly histidine tag
IMAC	immobilised metal affinity chromatography
IPTG	isopropylthio-β-D-galactosidase
JHBPs	juvenile hormone-binding proteins
Kb	kilobase
kDa	kiloDaltons
KRT1	keratin 1
L	litre
LB	Luria Bertani broth
LBP	LPS-binding protein

LDL	low-density lipoproteins
LPO	lactoperoxidase
MBP	maltose binding protein
Mdm12	mitochondrial distribution and morphology protein 12
MQ	milli Q water – ion exchanged purified water
Native PAGE	non denaturing polyacrylamide gel electrophoresis
nm	nanometer
Obp1f	odorant binding protein 1f
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
pI	isoelectric point
PIP	prolactin-inducible protein
PLTP	the phospholipid transfer protein
PLUNC	palate, lung, and nasal epithelium clone
PSP	parotid secretory protein

RMSD	root mean square deviation
rpm	revolutions per minute
SEC	size exclusion chromatography
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SMP	synaptotagmin-like mitochondrial-lipid-binding protein
SPLUNC2a	BSP30a
TAE	tris acetate ethylene diamine tetraacetic acid
TE	tris ethylene diamine tetraacetic acid
TEMED	tetramethylethylenediamine
Tm	melting temperature
TULIP	tubular lipid binding protein
Mm	micrometer
UV	ultraviolet light (280 nm)
V	volts
VEMSGP	von Ebner minor salivary gland protein
VLDL	very-low-density lipoproteins

v/v	volume per volume
w/v	weight per volume
X-gal	5-bromo-4-chloro-indolyl-β-D-galactopyranoside

Chapter One. Introduction

1.1 Ruminant salivary glands and saliva composition

In mammals, saliva serves as a lubricant in the oral cavity, assisting mastication and deglutition (McDougall, 1948). However, in ruminants, saliva has more extensive roles. Firstly, their saliva forms a fluid medium which transports the "chewing cud" back to the mouth for rumination and carries the ingesta onwards through the stomachs to the small intestine. Secondly, their saliva also forms a buffered medium (bicarbonate phosphate) for the survival and growth of ruminal micro-organisms (McDougall, 1948).

1.1.1 Salivary glands in ruminants

Ruminant salivary glands can be categorised into two types – major and minor glands (Fig. 1.1). The minor glands are distributed in the oral cavity mucosa or submucosa and they are named according to their location: buccal, labial, palatine and lingual (Elewa, Ichii, Otsuka, Hashimoto, & Kon, 2014).

In ruminants, the major salivary glands consist of parotid, mandibular and sublingual glands (Dyce, 1996). The parotid glands are the major and largest salivary gland in the ruminant. Both parotid glands are located ventral to the ear and its secreted saliva will drain into the mouth through a single large duct (parotid duct), which has a final opening in the caudal aspect of the mouth at the level of the second to last cheek tooth (Dyce, 1996; Tadjalli, Dehghani, & Ghadiri, 2002). The left and right submandibular glands are more axial than the parotid glands， located at the back of the jaw on the floor of the oral cavity. Each submandibular gland has its own duct, which extends along the floor of the mouth and opens on its respective

sublingual caruncle (Dyce, 1996; Tadjalli et al., 2002). Each sublingual salivary gland contains two parts: monostomatic and polystomatic parts. The left or right monostomatic gland is located in front of its respective polystomatic gland and both monostomatic glands secret saliva into the mouth through a single duct. The polystomatic glands lie on either side of the tongue and drain saliva on the floor of the mouth (Dyce, 1996).

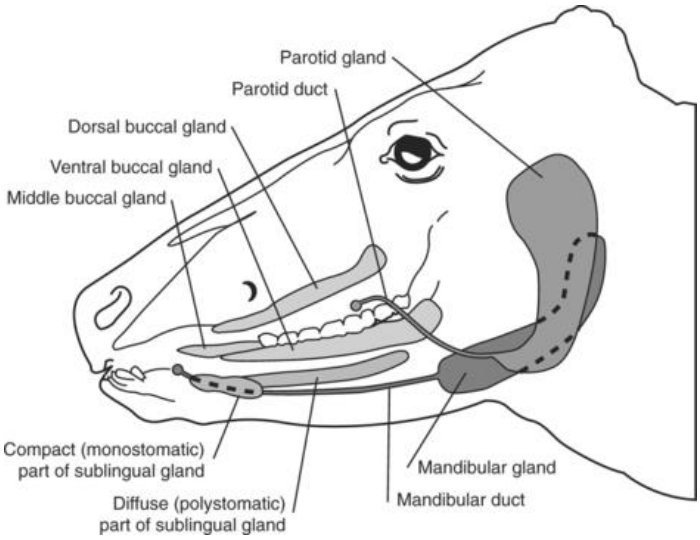

Figure 1.1 Schematic diagram of salivary glands in cattle. Used with permission from (Dyce, 1996).

The major salivary glands consist of three types of cells: acinar, duct and myoepithelial cells. The acinar cells make up the largest part of the salivary glands and they are further classified into two types: serous cells and mucous cells. Serous cells are large granular cells with round nuclei that secrete a watery fluid. In contrast, mucous cells have a clear cytoplasm and flattened nucleus and secrete a mucus-rich saliva, which is high in carbohydrates and proteins (Kay, 1966). The parotid glands can continuously produce a serous and isotonic saliva as they are only made up of

serous acini (Elewa et al., 2010). The saliva secreted from sublingual glands is mucus-rich and hypotonic as they are mostly composed of mucous acini. The submandibular gland is made up of a mixture of serous and mucous cells and hence its secreted saliva is variably mucus and hypotonic (Kay, 1966).

1.1.2 Ruminant saliva

Ruminant saliva is an isotonic, bicarbonate phosphate buffer with high pH, which helps to maintain the acid-base equilibrium of the ruminal contents by neutralizing acids formed by fermentation in the rumen (Singh & Singh, 2017). Compared with other mammals, ruminants can produce large amounts of saliva daily due to the large salivary glands. It is estimated that 100-190 litres of saliva can be produced by a single adult cattle daily.

Due to the different compositions of cells within each major salivary gland, different responses can be accomplished dependent upon the presence or absence of stimulation. The sublingual and submandibular glands, when stimulated by feeding, secrete a saliva which is hypotonic and weakly buffered. It functions as a lubricant to the bolus of food (McDougall, 1948). In contrast, the parotid saliva is released in large amounts during eating and rumination and its secretion continues at a reduced rate during the intervals of rest. It has a high alkalinity and buffering power and hence is crucial for the maintenance of normal rumen function (Kay, 1966; McDougall, 1948).

1.1.3 Ruminant saliva proteins and their functions

Although saliva is predominantly a watery fluid, it also contains a complex mixture of proteins mostly secreted by the salivary glands with a small portion originating

from the blood (Pedersen, Bardow, Jensen, & Nauntofte, 2002). The growing interest in health issues and productivity of ruminants has stimulated studies on their salivary proteins, especially with advances in proteomic techniques (Lamy & Mau, 2012; Shibata et al., 2009).

Recently, an in-depth systematic proteomic analysis of saliva from different mammal species, including cattle, was carried out using GeLC-MS/MS (SDS-PAGE-LC coupled to MS/MS) methods (de Sousa-Pereira et al., 2015). According to their analysis, 74 proteins were identified in cattle's saliva, of which nine proteins were higher than 0.9 % in abundance, namely CA6 (carbonic anhydrase 6), CSN1S1 (casein 1S1), CSTB (cystatin B or stefin B), HBB (haemoglobin subunit beta), KRT1 (keratin 1), LPO (lactoperoxidase), Obp1f (odorant binding protein 1f), PIP (prolactin-inducible protein), and SPLUNC2a (BSP30a).

Carbonic anhydrases are a group of zinc metalloenzymes which are responsible for the interconversion of carbon dioxide to bicarbonate (Tashian, 1989). They are known to be involved in the maintenance of the buffer capacity of saliva. Of which, CA6 is a 42 kDa secreted protein which has the ability to bind to the enamel surface and maintain pH homeostasis (Gilmour, 2010; Kivela, Parkkila, Parkkila, Leinonen, & Rajaniemi, 1999).

Casein is the name for a group of secretory calcium-binding phosphoproteins. It has been reported that casein acts in calcium homeostasis in saliva and appears to prevent adherence of salivary components and bacteria to enamel, which protects the hard oral tissues (Johansson & Lif Holgerson, 2011; Rijnkels, Elnitski, Miller, & Rosen, 2003).

CSTB (cystatin B or stefin B) is a cystatin family 1 protein of the cystatin superfamily. It was initially found to be an intracellular and reversible inhibitor of papain-like cysteine proteinases, which has a unique free Cys in the N-terminal segment of the proteinase binding region (Pol & Bjork, 2001; Turk, Turk, & Turk, 1997). It has been proposed to have a general protective role in mammals and most likely prevents inappropriate proteolysis caused by the action of lysosomal cysteine proteinases (Turk, Turk, & Turk, 2000; Turk et al., 1997). More recently, its presence in saliva as S-modified derivatives was detected, but its function is still unclear (Cabras et al., 2012).

Haemoglobin is a highly conserved oxygen transport protein found in mammalian red blood cells, its presence in saliva is linked to a range of diseases, particularly periodontitis (Maeng et al., 2016; Marti, Friedman, Cabrini, & Costa, 2002).

Similar to haemoglobin, the increased concentration of keratin in saliva is also linked to gingivitis and periodontitis (McLaughlin, Kirkham, Kowolik, & Robinson, 1996).

Lactoperoxidase is a heme containing protein which is released into several body fluids including the saliva. Within saliva, it is secreted by acinus cells of the salivary gland and helps to maintain the oral bacterial homeostasis by catalysing the oxidation of thiocyanate (SCN^-) to hypo-thiocyanite (^-OSCN) (Chandler & Day, 2012; Ihalin, Loimaranta, & Tenovuo, 2006).

Odorant binding proteins are mainly found in the nasal mucus of vertebrates and in the sensillar lymph of insects. They act as olfactory perception because of their high affinity to odors and pheromones, but their function in saliva has not been clearly defined (Pelosi, 1994).

Prolactin-inducible protein in saliva is considered to have an important role in host defence by promoting aggregation of oral bacteria (Nistor, Bowden, Blanchard, & Myal, 2009).

SPLUNC2a (BSP30a) belongs to PLUNC family of proteins and the structure and function of this family of bovine salivary proteins is the subject of this thesis (Section 1.2.2).

1.2 PLUNC (palate, lung, and nasal epithelium clone) proteins

PLUNC proteins are a family of proteins which are predominantly expressed in the upper respiratory tract, nasal mucosa, and oral cavity. The first gene belonging to this group was identified in the developing palate of the mouse embryo (Weston et al., 1999). It was noticed that its RNA was expressed strongly in the mouse embryo on day 14 after gestation at discrete regions of the nasal epithelium. In addition, significant expression in the adult mouse was seen in the tracheal and bronchial epithelium (Weston et al., 1999).

The sequence of murine *plunc* contains an ORF (open reading frame) of 834 bp (base pairs), giving a putative protein product of molecular mass 28,618 Da. Analysis of the predicted amino acid sequence indicates that it is a secreted protein which has homology with two salivary gland proteins: the von Ebner minor salivary gland protein and the parotid secretory protein (PSP) precursor (Weston et al., 1999).

Systematic study of this gene revealed that it is located in a gene cluster on chromosome 2 (Fig. 1.2) together with seven other *plunc* genes and two genes coding for parotid secretory protein (PSP) and von Ebner minor salivary gland protein (VEMSGP), respectively (LeClair, 2003). This cluster is also known as *BPI (bactericidal/permeability-increasing protein) fold containing family (BPIF)* gene locus because the protein structures of those genes in this locus share high sequence similarity with BPI, which is an important host defence molecule in mammals and has potent antimicrobial activity against Gram negative bacteria through its binding to lipopolysaccharide (LPS) (Haigh et al., 2008; LeClair, 2003).

1.2.1 Human PLUNC family

The human *PLUNC* gene locus, located on chromosome 20, spans a region of approximately 300 kb and contains nine *plunc* genes, and two pseudogenes (Fig. 1.2) (Haigh et al., 2008). Among those genes, *splunc1* (also known as *BPIFA1*) has attracted more attention and its protein has been intensively studied because of its continuous secretion in the airways, which may have multiple functions in combating respiratory infection induced by Gram-negative bacteria (Garland et al., 2013).

The antimicrobial activity of SPLUNC1 was first reported in nasopharyngeal carcinoma (NPC) epithelial HNE1 cells (Zhou et al., 2006). Co-localization of GFP-tagged SPLUNC1 with Gram-negative nanobacteria (NB) within NPC epithelial HNE1 cells indicated the potential binding between SPLUNC1 and the LPS of NB, which can play an important role in the host defence of nasopharyngeal epithelium (Zhou et al., 2006). Further studies have shown its growth inhibition to *Pseudomonas aeruginosa* and killing activity against *Haemophilus influenza*

(McGillivary & Bakaletz, 2010; Zhou et al., 2008). Moreover, SPLUNC1 may act as a novel airway surfactant with antibiofilm activity which disrupts the biofilm formation and growth of *P. aeruginosa* in airways (J. Bartlett et al., 2011; Gakhar et al., 2010a). Based on the sequence similarity between SPLUNC1 and BPI, it was predicted that SPLUNC1 could also bind to LPS in a similar way to BPI. Indeed, several studies have reported direct binding of SPLUNC1 to LPS (Di, 2011; Ghafouri, Kihlstrom, Tagesson, & Lindahl, 2004; Sayeed, Nistico, St Croix, & Di, 2013a; Zhou et al., 2008). However, failure to bind has also reported elsewhere (Campos et al., 2004; Ning et al., 2014).

More recently, SPLUNC1 is proposed to be a pH-sensitive regulator of epithelial Na^+ channel (ENaC), which in normal airways can preserve the ASL (airway surface liquid) volume through its inhibition of the epithelial Na^+ channel (ENaC)-dependent Na^+ absorption (Garland et al., 2013). However, in the acidic cystic fibrosis (CF) airway environment it is unable to inhibit ENaC which leads to the airways surface liquid (ASL) dehydration, a major syndrome of CF. SPLUNC1 regains its ability to prevent CF ASL hyperabsorption following alkalinisation of CF airway cultures. Structural studies reveal that its pH-sensitive regulation comes from two pH-sensitive salt bridges in the structure. Removal of these two salt bridges will render SPLUNC1 pH-insensitive and unable to regulate ASL volume in acidic ASL (Garland et al., 2013).

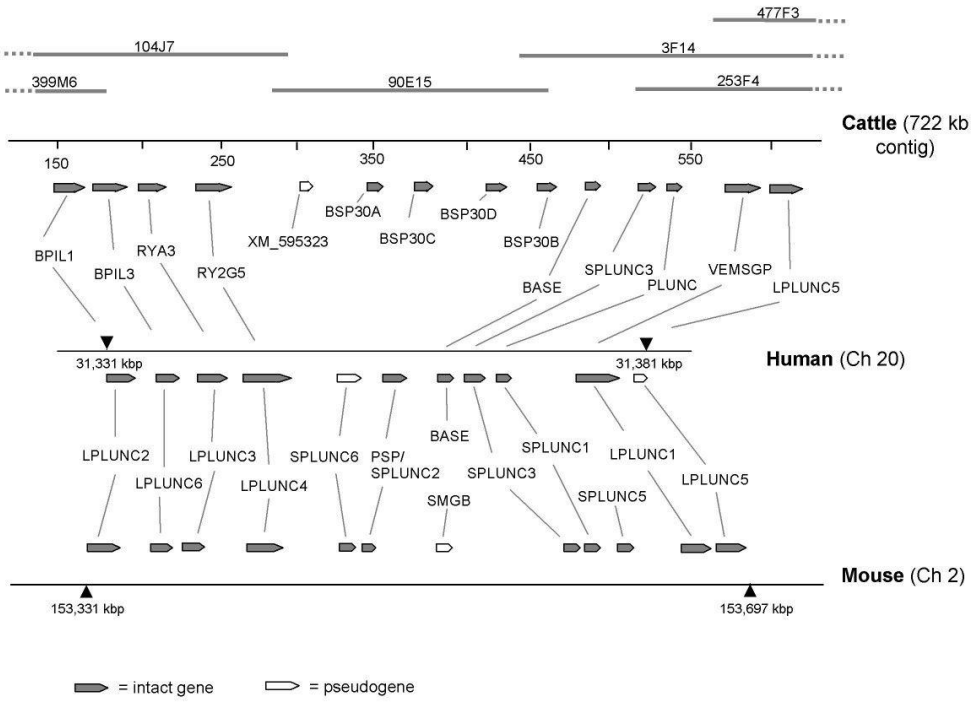

Figure 1.2 The PLUNC protein family in cattle, human and mouse. The direction of the ORF is indicated by the arrowed end of each gene. Used with permission from (Haigh et al., 2008).

1.2.2 Bovine PLUNC family

Cattle have an expanded *PLUNC* gene locus, which contains thirteen *plunc* genes and one pseudogene on chromosome 13 (Fig. 1.2) (Wheeler et al., 2007). Nine of the thirteen *plunc* genes are orthologous to genes in the human and mouse locus, and are thought to have similar roles in host defence. The remaining four genes, namely *BSP30a*, *BSP30b*, *BSP30c*, and *BSP30d*, have arisen in cattle through a series of duplications which may be a characteristic of ruminants and suggests that their proteins may contribute to the functions of ruminant physiology (Wheeler et al., 2007).

The first described PLUNC protein in cattle was named "BSP30", and had a mass of around 30 kDa on a SDS-PAGE gel following electrophoresis of bovine saliva

(Rajan, Morris, Carruthers, Wilkins, & Wheeler, 1996). This "BSP30" protein is one of the most abundant proteins in bovine saliva, with a concentration of 0.1-0.5 mg/ml. Its abundance was proposed to be associated with the susceptibility of the cattle to pasture bloat, a metabolic disease characterized by build-up of stable foam in the rumen and hence impairment of the eructation mechanism, causing rumen distension and respiratory distress (Clarke & Reid, 1974). Further investigation of this protein revealed that it is a mixture of two closely related proteins - BSP30a and BSP30b, sharing 83% amino acid identity, with expression restricted to salivary tissue (Wheeler et al., 2002). A gene expression study indicated that the mRNAs of the four *BSP30* genes are most abundant in tissues associated with the oral cavity and airways. Notably, *BSP30a* and *BSP30b* are abundantly expressed in the salivary gland independently of one another and their proteins are secreted into saliva, accounting for 15-30% of the total protein in bovine saliva (Wheeler, Hood, Oden, McCracken, & Morris, 2003).

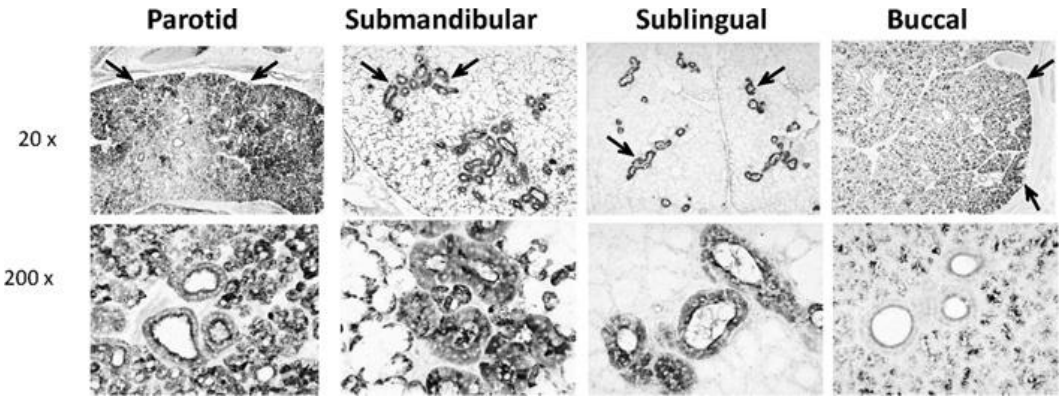

Figure 1.3 Immunohistochemical analysis of BSP30 proteins in bovine salivary tissues. Antibody raised in rabbit recognizes both BSP30a and BSP30b. The immunoreactive signal (dark areas, and examples arrowed) was visualized with diaminobenzidine at x20 and x200 magnification. Used with permission from (Wheeler, Haigh, Broadhurst, Hood, & Maqbool, 2011).

Immunohistochemical analysis of bovine salivary glands with antibodies specific for both BSP30a and BSP30b suggested that they are localized to the serous secretory cells (Fig. 1.3) (Wheeler et al., 2011). Further analysis with Western blot revealed that BSP30b is present in the parotid, submandibular and buccal salivary glands; whilst BSP30a is only present in the submandibular gland, although its mRNA was shown to be expressed abundantly in the parotid (Fig. 1.4) (Haigh et al., 2008; Wheeler et al., 2003). Functional studies suggest that both BSP30a and BSP30b can suppress the growth of *Pseudomonas aeruginosa* with moderate potency but suppression was not observed for other pathogens tested, indicating that (unlike BPI proteins) the BSP30 proteins may not have a primary function as bactericidal proteins (Haigh et al., 2008; Wheeler et al., 2011). It has also been observed that both BSP30a and BSP30b have no significant LPS binding activity, suggesting that the mechanism of their antibacterial activity is independent of LPS opsonisation (Haigh et al., 2008).

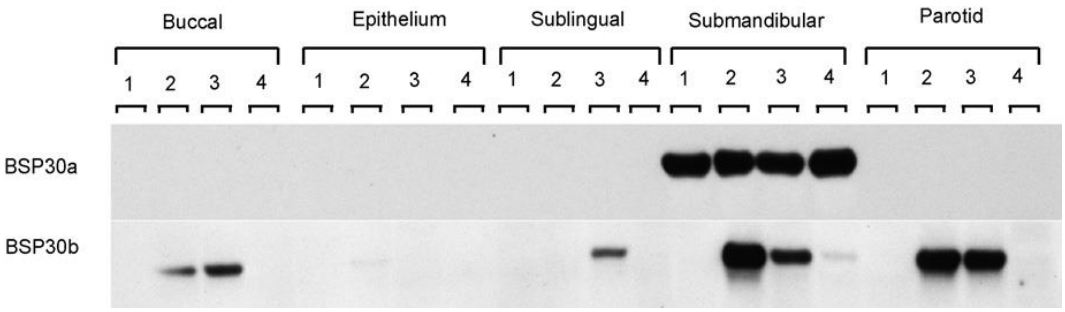

Figure 1.4 Western analysis of BSP30a and BSP30b expression in the major bovine salivary glands. Used with permission from (Haigh et al., 2008)

1.3 TULIP (tubular lipid-binding) superfamily

The TULIP superfamily consists of three protein families – SMP-like, BPI-like, and Takeout-like protein families (Kopec, Alva, & Lupas, 2011). Proteins in this

superfamily all have the N-terminal domain of BPI, also known as a TULIP domain. The tertiary structure of this domain has a long bent helix wrapped in a highly curved anti-parallel β-sheet, forming a hydrophobic channel in the central part of the protein (Alva & Lupas, 2016).

Despite the structural conservation among proteins within this superfamily, their amino acid sequences are not well conserved. The Hidden Markov Model (HMM) analysis (Schuster-Bockler, Schultz, & Rahmann, 2004) of the LBP_BPI_CETP family indicates limited amino acid sequence conservation in the alignment (Fig. 1.5). The HMM logo reveals a low level of conservation involving just a few hydrophobic amino acids such as leucine at positions 2, 149, and 161, isoleucine at position 46. Only the two cystines at positions 115 and 154 are highly conserved across this family. These two cystines form a disulphide bond, which is proposed to stabilize the structure for proteins in this family such as SPLUNC1 (Ning et al., 2014).

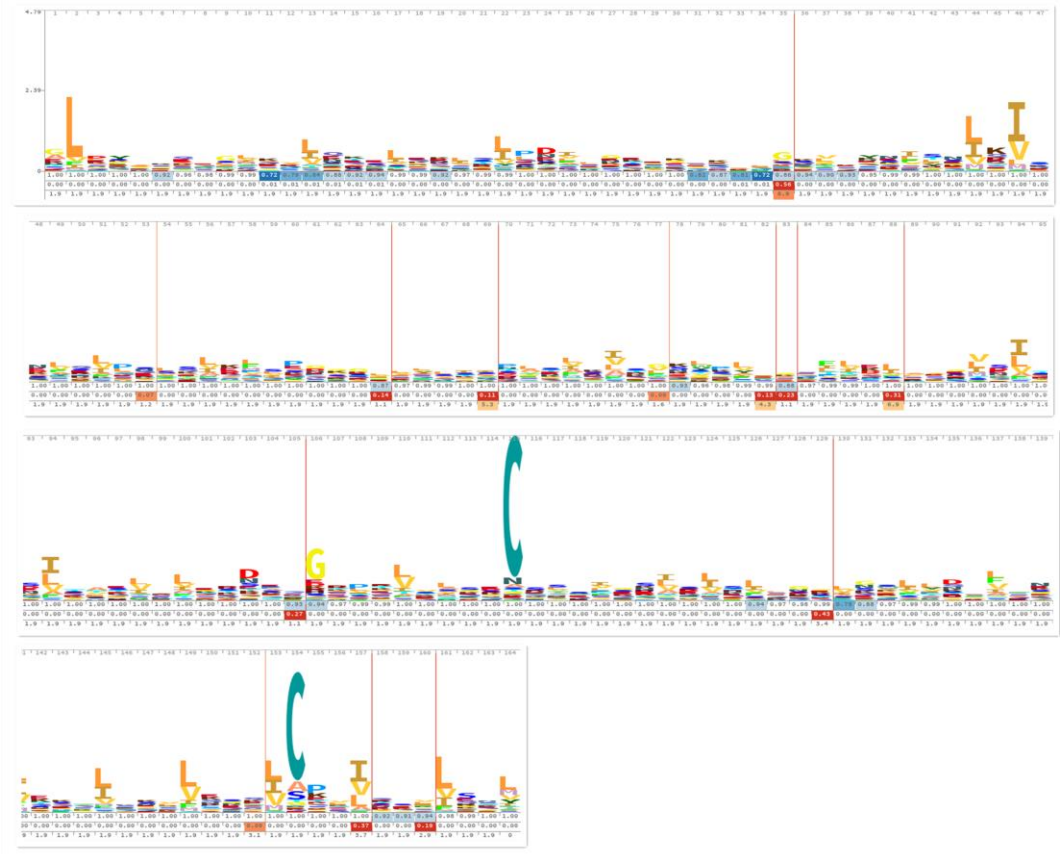

Figure 1.5 Hidden Markov Model (HMM) logo of the LBP_BPI_CETP family shows the distribution of amino acids from a multiple alignment (https://pfam.xfam.org/family/PF01273#tabview=tab3). The height and width of each letter at each position is proportional to its occurrence and relative role, respectively. The red bars represent regions of residue insertion in the alignment. Amino acid colours reflect their biological properties (red = charged; blues = polar, uncharged; yellows = aliphatic; greens = aromatic).

1.3.1 BPI-like Protein Family

The BPI-like family includes BPI (the bactericidal/permeability-increasing protein), LBP (LPS-binding protein), CETP (the cholesteryl ester transfer protein), PLTP (the phospholipid transfer protein), and PLUNC (Alva & Lupas, 2016). They are either involved in innate immunity against bacteria through their ability to bind LPS, or in lipid exchange between lipoprotein particles.

BPI is a predominantly intracellular protein found in the primary granules of polymorphonuclear neutrophils (Elsbach & Weiss, 1995). BPI binds to LPS with high affinity and it has potent bactericidal activity against Gram-negative bacteria in the phagocytic compartment of the neutrophil to kill engulfed bacteria (Weiss, Elsbach, Olsson, & Odeberg, 1978). The three dimensional structure of BPI is initially solved at 2.4 Å (Beamer, Carroll, & Eisenberg, 1999) and later solved at 1.7 Å (Kleiger, Beamer, Grothe, Mallick, & Eisenberg, 2000). Unlike other soluble proteins, BPI has an interesting "boomerang" shape (Fig. 1.6). Its N-terminal domain and C-terminal domain are connected by a proline-rich linker. Although the sequence similarity between the N- and C-terminal domains is very low (<20%), they both have a tubular shape formed by two α helices and a highly twisted antiparallel β sheet. Electrospray mass spectrometry confirmed that each tubular cavity has a phosphatidylcholine sitting in it. Based on the similarities between LPS and phosphatidylcholine, further research showed that BPI binds to the lipid A region of LPS with high affinity. Combined with antibacterial studies, it was suggested that BPI has bactericidal activity against Gram negative bacteria via LPS binding (Beamer et al., 1999).

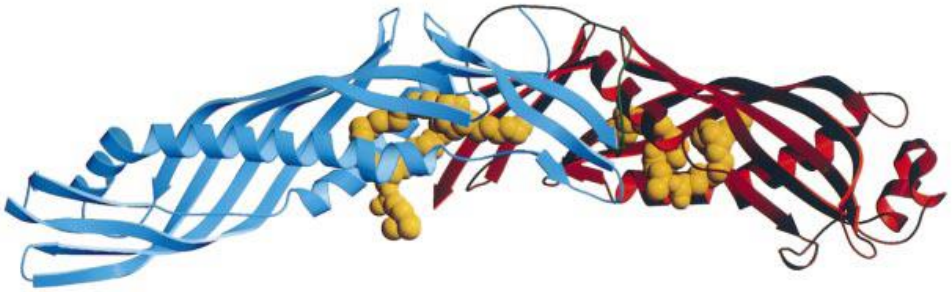

Figure 1.6 Cartoon representation of human BPI. Its N-terminal domain is blue and C-terminal domain is red. The phospholipid is shown as orange spheres. Used with permission from (Kleiger et al., 2000).

Another protein belonging to this family is LBP, which is constitutively produced by hepatocytes and secreted into the bloodstream (Tobias, Soldau, & Ulevitch, 1986). It also binds to LPS with high affinity, and plays a critical role in the activation of immune cells by LPS (Hailman et al., 1994; Tobias, Soldau, Gegner, Mintz, & Ulevitch, 1995). After binding LPS, LBP delivers it to a cell surface protein (CD14) of macrophages and monocytes, activating the inflammatory cascade (Fig. 1.7a). Interestingly, *in vitro* studies show that, in contrast to LBP, BPI can neutralize the inflammatory properties of LPS (von der Mohlen et al., 1995). This observation suggested that the inflammatory response of LBP to LPS in the blood stream could be alleviated by addition of BPI to the bloodstream. The three dimensional structure of LPS (Fig. 1.7b) is similar to BPI, which also has a boomerang-like structure (Eckert et al., 2013). Unlike BPI, only the N-terminal domain of LBP binds LPS; whereas its C-terminal domain is believed to mediate the delivery of LPS to other host molecules.

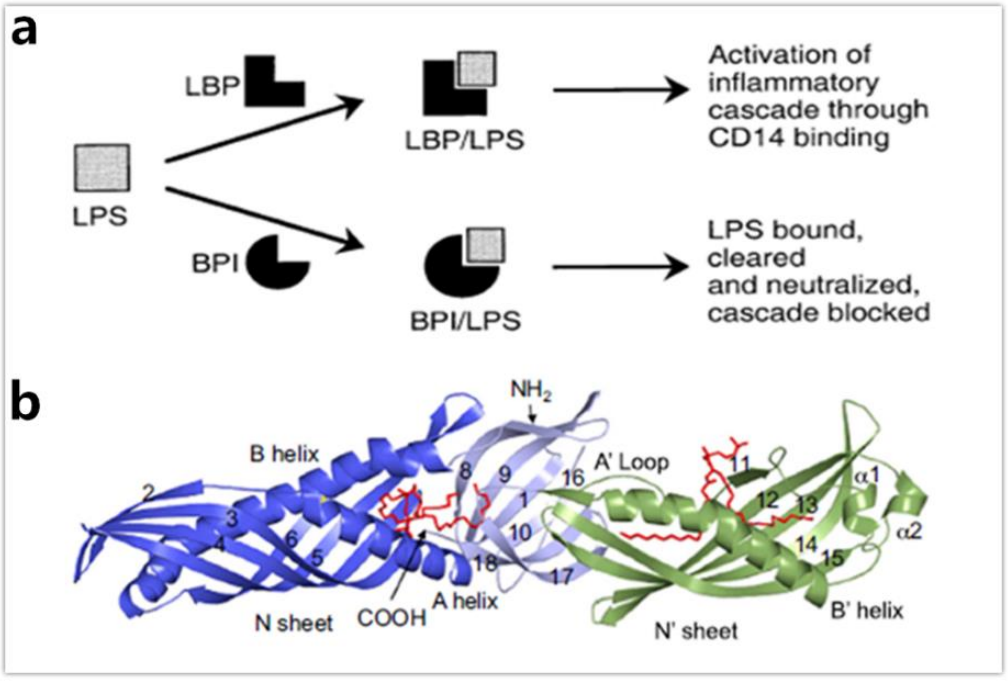

Figure 1.7 Structure of LBP and its effects on the LPS induced inflammatory cascade. 3a, schematic illustration of the different effects of LBP and BPI on the LPS induced inflammatory cascade. 3b, Cartoon representation of the crystal structure of LBP. Its N-terminal domain is in blue and C-terminal domain is in green. Used with permission from (Beamer et al., 1999; Eckert et al., 2013).

CETP is a hydrophobic glycoprotein, mediating the transfer of neutral lipids, including cholesteryl esters and triglycerides, between high-density lipoproteins (HDL), low-density lipoproteins (LDL) and very-low-density lipoproteins (VLDL) (Barter et al., 2003; Tall, 1993). The structure of CETP at 2.2 Å resolution (Fig. 1.8a) revealed that it also has a boomerang-shaped structure with two cholesteryl esters and two phosphatidylcholine bound in its hydrophobic tunnel (Qiu et al., 2007). Its potential interaction with lipoproteins is illustrated from an electron microscopy study (Zhang et al., 2012). It is proposed that the N-terminal β-barrel domain of CETP penetrates the HDL surface, and its distal end interacts with the cholesteryl ester core of HDL. Then, the C-terminal β-barrel domain interacts with LDL or VLDL. Both β-barrel domains are twisted because of the binding to form a tunnel, which serves as a conduit, resulting in a transfer of cholesteryl esters to LDL or VLDL and a reduction in HDL (Fig. 1.8b).

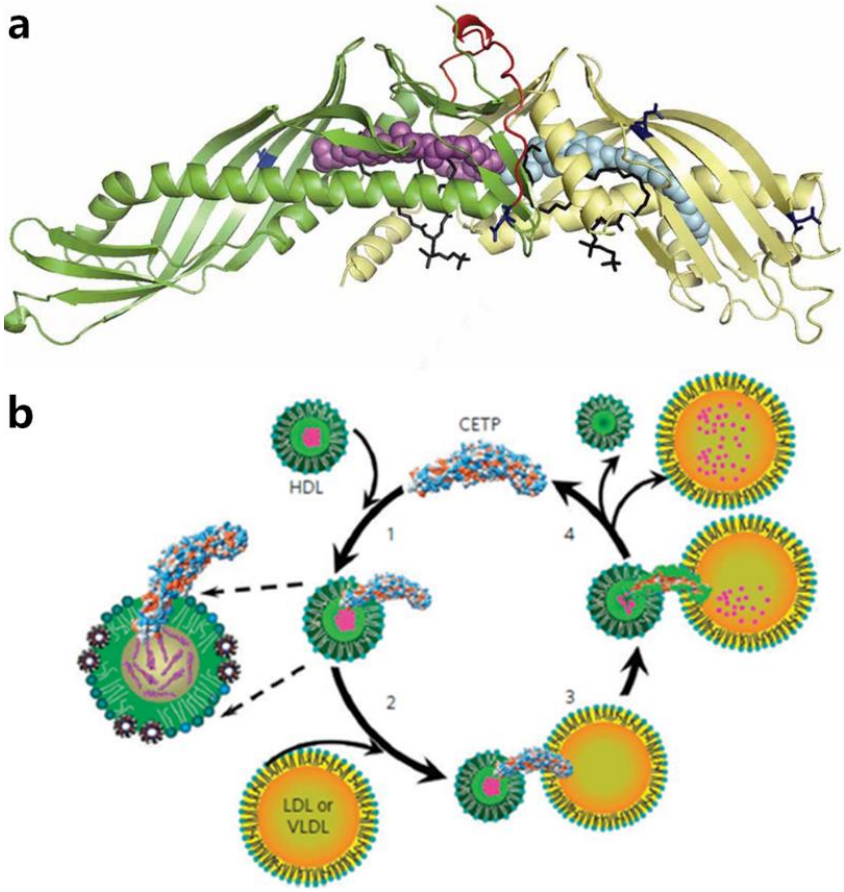

Figure 1.8 Structure and function of CETP. 4a, cartoon representation of CETP. N-terminal domain is in green and C-terminal domain is in yellow. 4b, proposed mechanism for cholesteryl ester transfer by CETP. The N-terminal domain of CETP firstly penetrates the HDL surface and then its C-terminal domain interacts with LDL or VLDL. Those interactions result in the transfer of cholesteryl esters from HDL to LDL or VLDL. Used with permissions from (Qiu et al., 2007; Zhang et al., 2012).

PLTP (the phospholipid transfer protein) is a main factor which regulates the size and composition of HDL in the bood circulation and also plays an important role in controlling plasma HDL levels (Huuskonen, Olkkonen, Jauhiainen, & Ehnholm, 2001). Its regulator function is fulfiled through its phospholipid transfer activity and its HDL conversion ability.

The PLUNC proteins also belong to this family. Proteins in this large group are predominantly expressed in the upper respiratory tract, nasal mucosa, and oral cavity. PLUNC proteins can be divided into two subgroups: LPLUNC (long

PLUNC) proteins and SPLUNC (short PLUNC) proteins (Canny & Levy, 2008). LPLUNC proteins have homology to both the N- and C-terminal domains of BPI and LBP, while SPLUNC proteins are only homologous to the N-terminal domain of BPI (Canny & Levy, 2008). Although several studies have characterized the expression of PLUNC proteins, their functions are still not well understood. So far, only the structures of SPLUNC1 from human and Latherin from horse have been solved (Fig. 1.9) (Ning et al., 2014; Vance, McDonald, Cooper, Smith, & Kennedy, 2013). Both structures adopt a slightly curved cylindrical structure with a hydrophobic channel formed by an α/β wrap. However, neither of their ligands are found in their structure, although dipalmitoylphosphatidylcholine (DPPC) has shown to have high affinity to SPLUNC1 *in vitro (Ning et al., 2014).*

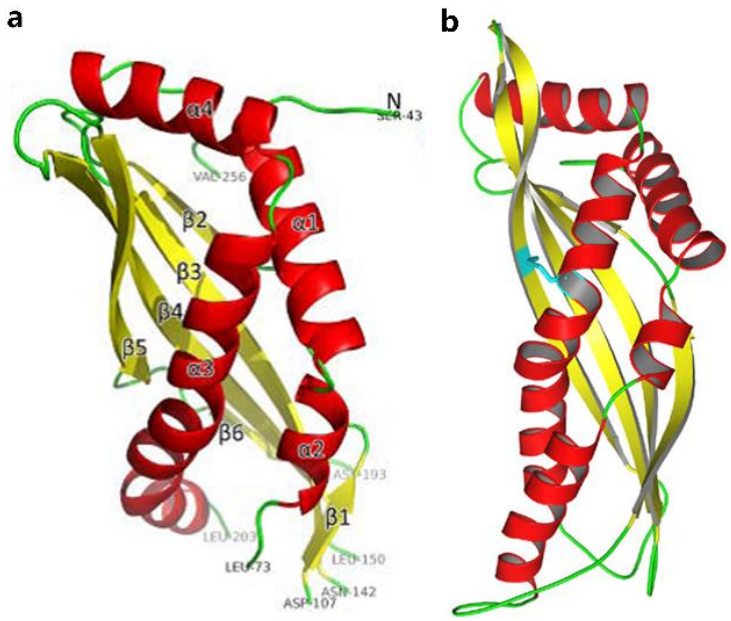

Figure 1.9 Crystal structure of SPLUNC1 and latherin. a, cartoon representation of SPLUNC1 from human. Secondary structure elements are labeled. b, cartoon representation of latherin from horse. Used with permissions from (Ning et al., 2014; Vance et al., 2013).

1.3.2 Takeout-like protein family

This family consists of insect juvenile hormone-binding proteins and arthropod allergens, transporting lipid hormones to target tissues during insect development.

JHBPs (juvenile hormone-binding proteins) from insects are responsible for the JH (juvenile hormone) delivery process (Kramer, Sanburg, Kezdy, & Law, 1974). They have high affinity and enantioselectivity for naturally occurring JHs. Recently the structure of JHBP from silkworm (*Bombyx mori*) was reported and reveals a possible gate and latch mechanism for hormone binding and delivery (Suzuki et al., 2011). Comparison of the apo- and JH-bound JHBP structures (Fig. 1.10a and b) shows significant differences in helix α1, which is suggested to be the gate and controls the binding and releasing of JH. At the *corpora allata* cell membrane, JH binds to the hormone binding pocket of the apo-JHBP in the gate-open form followed by the closure of gate helix α1, leading to the formation of hydrogen bonds between the N-terminal arm and the C-terminal tail. This fully gate-closed JHBP-JH complex will protect the bound JH from unfavourable nonspecific absorption and enzymatic degradation. When the complex reaches the target tissues, the bound JH molecule will be released after the open of gate helix α1 (Fig. 1.10c).

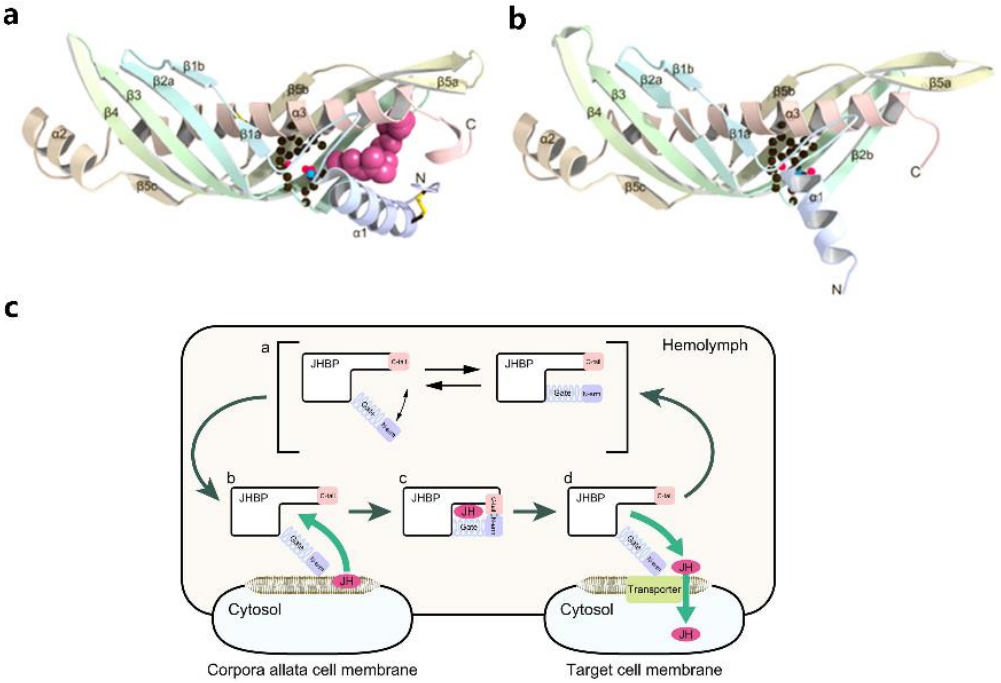

Figure 1.10 Structure and function of JHBP. a, the crystal structure of JHBP in complex with JH. b, the crystal structure of apo-JHBP. c, the proposed mechanism for JH transport mediated by JHBP in the hemolymph. Used with permission from (Suzuki et al., 2011)

Takeout proteins are found across a diverse range of insect species and act as ligand carriers. Its structure was solved at 1.3 Å initially (Fig. 1.11a), showing its TULIP domain consisting of a long α helix wrapped into a highly curved anti-parallel β sheet (Hamiaux, Stanley, Greenwood, Baker, & Newcomb, 2009). The formed internal cavity is 45 Å long and a ubiquinone-8 molecule is found inside, which might be derived from *Escherichia coli*. Recently, crystallography combined with nanoelectrospray mass spectrometry revealed a number of fatty acids (C14, myristic; C15, pentadecanoic; C16, palmitic; traces of C17, heptadecanoic; and C18, stearic) present in the internal cavity of EpTo1 when it is expressed in *sf9* insect cells (Fig. 1.11b) (Hamiaux, Basten, Greenwood, Baker, & Newcomb, 2013).

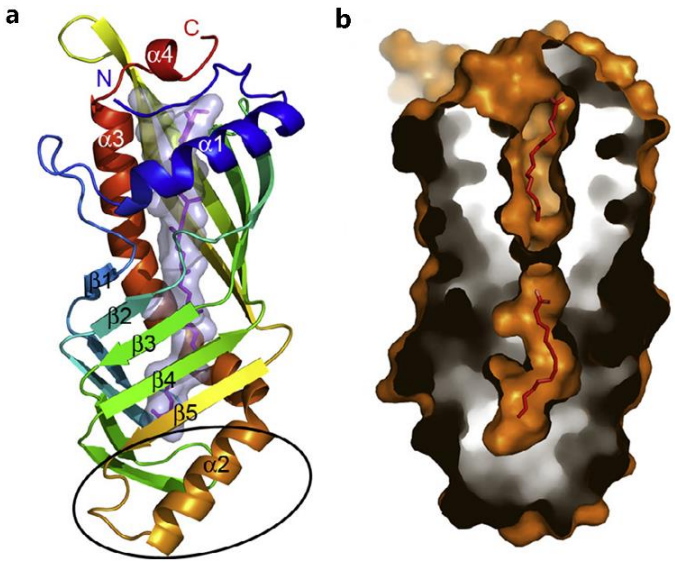

Figure 1.11 Crystal structure of EpTo1. a, crystal structure of EpTo1 with ubiquinone-8. b, slab surface representation of EpTo1 with myristic acids inside the internal cavity. Used with permission from (Hamiaux et al., 2013).

Until now, the structures of only two closely related arthropod allergens – Der f 7 and Der p 7 – have been solved (Mueller et al., 2010; Tan et al., 2012). Those two proteins share 86% similarity at amino acid level and their structures all contain two anti-parallel β sheets which wrap around a long C-terminal helix (Fig. 1.12). The key difference between their structures is in the β1-β2 loop region, proximal to the IgE epitope in Der f 7 but not in Der p 7. Therefore, monoclonal antibody binding to Leu48 and Phe50 can inhibit IgE binding to Asp159 in Der f 7 due to their close proximity. Both allergens can weakly bind to polymyxin B via a similar binding site and both of their stabilities are highly depend on pH (Tan et al., 2012).

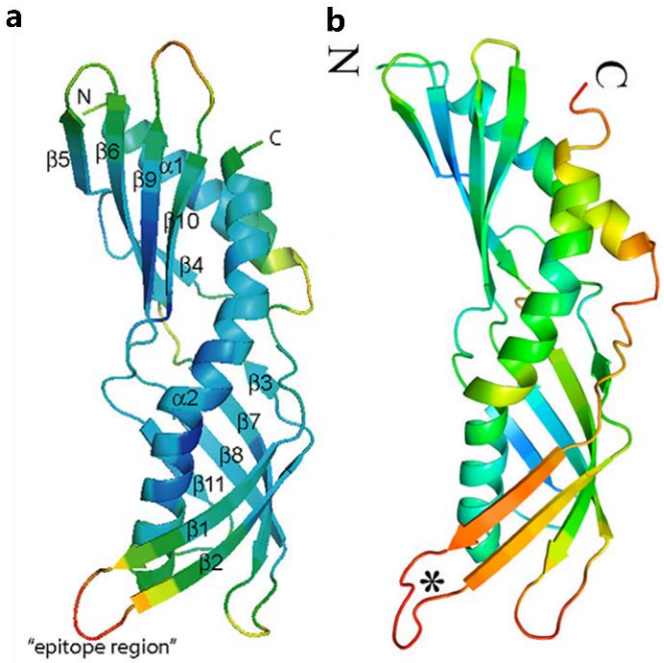

Figure 1.12 Crystal structures of Der f 7 and Der p 7. a, cartoon representation of Der f 7. b, cartoon representation of Der p 7. Used with permissionsfrom (Mueller et al., 2010; Tan et al., 2012).

1.3.3 SMP-like protein family

This family includes subunits of the ERMES complex and the extended synaptotagmins (E-Syts), which appear to be mainly located at membrane contact sites between organelles, mediating inter-organelle lipid exchange.

The endoplasmic reticulum-mitochondria encounter structure (ERMES) complex was first identified as a molecular tethering factor in the formation of ER-mitochondrial junctions in *S. cerevisiae* (Kornmann et al., 2009). The ERMES includes four proteins – Mdm12, Mmm1, Mdm34, and Mdm10 – with different subcellular localizations. Mdm12 (mitochondrial distribution and morphology protein 12) is a cytosolic protein, while Mmm1 (maintenance of mitochondrial morphology protein 1) and Mdm34/10 are integral membrane proteins anchoring on the ER and mitochondrial outer membranes, separately. The primary role of the

ERMES complex is to maintain a close proximity (10-30 nm) between two membranes; it also has functions in lipid trafficking between ER and mitochondria (Kopec, Alva, & Lupas, 2010; Kornmann et al., 2009; Toulmay & Prinz, 2012). A primary structure analysis of an ERMEs complex indicates that Mdm12, Mmm1, and Mdm34 all contain a synaptotagmin-like mitochondrial-lipid-binding protein (SMP) domain (Kopec et al., 2010). More recently, the crystal structure of Mdm12 revealed that two Mdm12 molecules interact with each other, forming a hydrophobic channel to fulfil its lipid trafficking function (Jeong, Park, & Lee, 2016). Electron density together with denaturing quantitative APCI-MS analysis confirmed the lipid sitting in the hydrophobic channel is PE, consistent with a previous study. It is proposed that Mdm12 interacts with Mdm34 in a head to head manner and that Mdm12 interacts with Mmm1 in a tail to tail formation. This hexamer is proposed to mediate the lipid transfer between ER and mitochondria (Fig. 1.13).

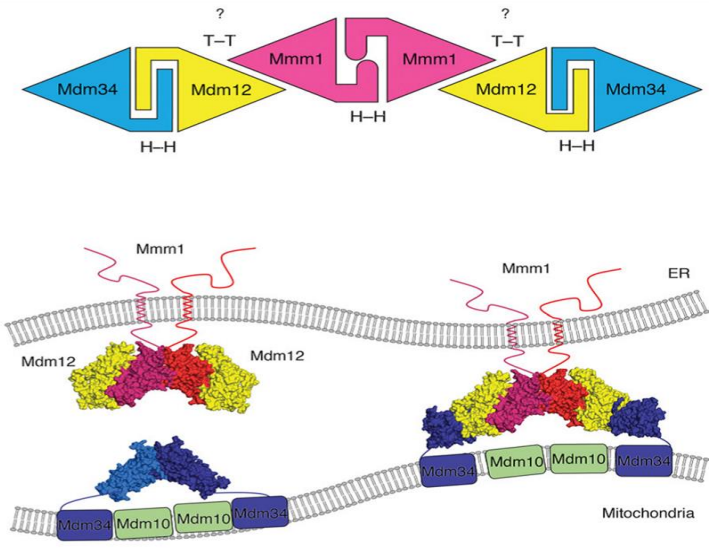

Figure 1.13 Putative architecture of the Mdm12-Mdm1-Mdm34 complex. Used with permissionfrom (Jeong et al., 2016).

E-Syts are ER-resident proteins which bring ER and PM membranes into close apposition (Manford, Stefan, Yuan, Macgurn, & Emr, 2012; Toulmay & Prinz, 2012). There are three E-Syts (E-Syts1, 2, and 3) in mammalian cells, which are either homo- or heterodimerized (Giordano et al., 2013). Partial structures of human E-Syts2 (residues 163-634) have been solved recently (Schauder et al., 2014). Its core part, synaptotagmin-like mitochondrial-lipid-binding protein (SMP) domain, together with SMP domain from another E-syts2, form a 90 Å long cylinder which is similar to BPI and CETP. Within each SMP domain, two lipid-like molecules – diacylglycerol lipid and Triton X-100 – were found. Denaturing MS analysis of E-Syts2 expressed in mammalian cells reveals the presence of glycerophospholipids phosphatidylcholine (PC; 67%), PE (21%), phosphatidylinositol (7%) and phosphatidylserine (3%), indicating a lack of specificity with respect to particular glycerophospholipid ligands. Based on E-Syt2's location within the cell and its lipid ligands, two models were proposed to illustrate its potential function (Fig. 1.14). The first one is a tunnel model, which assumes the E-Syt2 dimer bridges the ER and PM to transfer lipids between them. E-Syt2 being anchored to the ER through its N-terminal region and to the PM through its C2C domain. C2A-C2B interacts with the membrane through loops at the ends of the arch and could interact with the ER, PM or either, depending on the inter-membrane distance. The second model, the shuttle model, may present at larger ER-PM distances compared to the tunnel model. In this model, the SMP dimer may shuttle lipids between the ER and PM, and partner proteins might be involved to accomplish its job (Schauder et al., 2014).

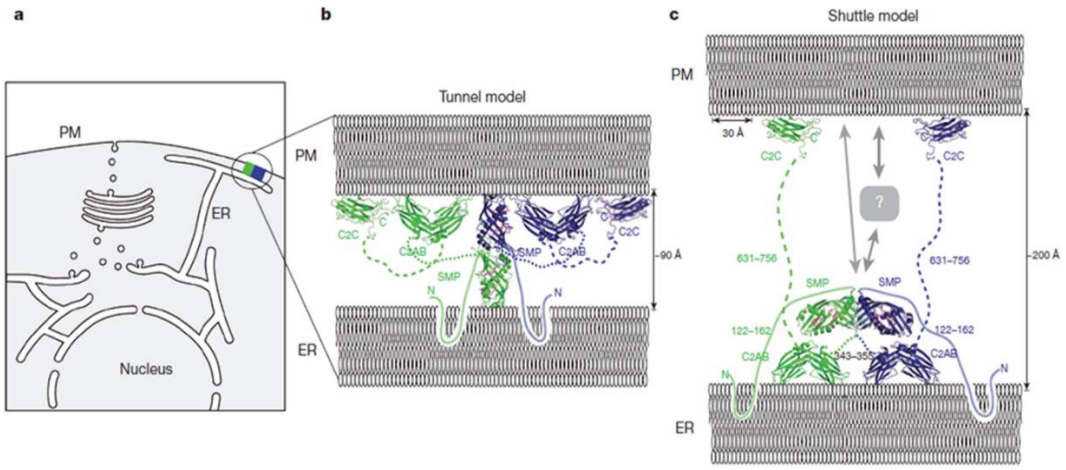

Figure 1.14 ER-PM contact site and proposed models of E-SYT2. a, schematic diagram of ER-PM contact site. b, the proposed tunnel model for lipid transfer. c, the proposed shuttle model for lipid transfer. Used with permissionfrom (Schauder et al., 2014).

1.3.4 "Global" overview of TULIP superfamily

Most proteins within the BPI-like and Takeout-like families are comprised of only the TULIP domain. In contrast, proteins of the SMP-like family frequently contain a single TULIP domain in combination with other domians that generally have membrane-associated activities, like PH, C2, and PDZ, and form elongated homo- and hetero- oligomers by end-to-end association (Alva & Lupas, 2016). Proteins from the Takeout-like family and SPLUNCs from the BPI-like family have a single TULIP domain, whilst the other proteins within BPI-like family consist of tandem TULIP domains forming a single chain pseudodimer.

Overall, proteins of TULIP superfamily have four different roles based on their ligands (Alva & Lupas, 2016). CETP and PLTP from the BPI-like family can facilitate the exchange of neutral lipids (e.g., fatty acids, triglycerides, and cholesteryl esters) and phospholipids between lipoproteins. Proteins of SMP-like protein family mainly function in exchanging lipids at membrane contact sites, although the impairment of ERMES in yeast does not completely disrupt lipid exchange and does not yield a lethal phenotype. This suggests that there are other redundant mechanisms for lipid trafficking between the ER and mitochondria. It is clear that many proteins of the BPI-like family are important components of the

innate immune system. Proteins of the Takeout-like family are only found in arthropods and they act as hormonal regulation and development.

1.4 BSP30b proteins in cattle

BSP30b has been identified as the orthologue of the human and rodent parotid secretory protein (PSP) (Wheeler et al., 2011). Phylogenetic analysis indicates that *BSP30b* appears to have arisen in cattle through gene duplications (Haigh et al., 2008). *BSP30b* has a similar expression pattern as that of *PSP* from other species. Northern blots of *BSP30b* RNA indicated that expression of BSP30b is restricted to the major salivary glands like parotid, submandibular, sublingual and buccal (Wheeler et al., 2011).

At the protein level, cattle saliva has a much higher abundance of BSP30b compared to PSP in human saliva (Wheeler et al., 2011). BSP30b, together with BSP30a, account for 15-30% of total bovine salivary proteins and can be clearly seen on SDS PAGE gels after Coomassie Blue staining. In contrast, the PSP in human saliva has a much lower concentration and is just visible on SDS PAGE gels after Coomassie Blue staining. Carbohydrate-specific stain (Pro-Q Emerald, Invitrogen) assay indicated that BSP30b, but not BSP30a, is glycosylated and it is predicted that BSP30b contains a mucin-type O-glycosylation site at serine 239 by using the NetOGlyc 3.1 predictor method (Haigh et al., 2008). Although BSP30b cannot interact with bacterial LPS, it still has significant antimicrobial activity against two out of three *P. aeruginosa* strains, suggesting the mechanism of its antibacterial activity is independent of LPS opsonisation (Haigh et al., 2008).

The initial attempt to crystalize native BSP30b in the Arcus lab was carried out by Dr Judith Burrows and Dr Graeme Card in 2005. The purified BSP30b protein made by Judith Burrows was sent to Graeme Card at the Stanford Synchrotron for crystallization. BSP30b was crystallized in the optimised 24 condition matrix (Appendix 6), especially condition B3. Those crystals had the best diffraction to 2.5 Å resolution, but the structure of BSP30b couldn't be solved by the molecular replacement (no structure of sufficient similarity had been solved) or using SAD/MAD methods (no heavy atoms were introduced to the protein).

Dr Judith Burrows also purchased genes encoding one double L-to-Met mutant and one triple L-to-Met mutant of BSP30b from GeneArt (Thermo Fisher Scientific, US), aiming to solve the structure of BSP30b using selenomethionine and SAD/MAD methods. Both mutants were ligated into *pProEx Htb* with a 6x histidine tag but the expression level of both proteins was poor and no crystals were produced after repeated experiments.

1.5 Hypothesis and aims

Although the gene expression of BSP30b has been well-studied on both mRNA level and protein level, its biochemical function(s) are not understood. No three dimensional structures of bovine PLUNC proteins have been reported. The lack of structural information or knowledge about their lipid substrates of the cattle PLUNC family members impedes our understanding of their functions.

1.5.1 Hypothesis

That BSP30b is a lipid binding protein based on similar observations with homologues in the TULIP superfamily. More specifically, as BSP30b is abundantly expressed in bovine saliva and will enter the rumen together with food swallowed, it is hypothesised that it can bind to specific lipids within the rumen, especially the outer membrane lipid profile of rumen bacteria.

1.5.2 Aims

1. To express BSP30b in *E. coli* and purify the expressed protein using different chromatography methods.
2. To crystallise BSP30b and determine its structure using X-ray diffraction.
3. To determine the substrate(s) of BSP30b and its function(s) in rumen.

Chapter Two. Materials and Methods

2.1 Materials and Methods Related to Chapter 3

2.1.1 Expression and Purification of His Tagged BSP30b

Plasmid (*pTriEx 4*) containing *BSP30b* was kindly provided by Tom Wheeler from AgResearch (Hamilton, New Zealand) and the plasmid was transformed into *E. coli BL21* by a former researcher (Dr Judith Burrows) in our lab. Small scale expression was tested and optimised by Dr Judith Burrows; hence, only large scale expression of BSP30b was investigated in this thesis.

2.1.1.1 Large Scale Expression of BSP30b with LB media

A colony taken from an LB plate containing ampicillin was inoculated into 10 mL LB culture with appropriate antibiotic. The culture was grown overnight at 37 °C with shaking at 200 rpm. The following morning, this 10 mL culture was used to seed a 1 L LB culture with appropriate antibiotic in a 2 L baffled conical flask, which was then grown at 37 °C with shaking at 200 rpm until the OD_{600} reached ~ 0.6. Protein expression was induced by adding 1 mM IPTG and growth was continued overnight under the same conditions. Induced cell culture was centrifuged at 6,000 rpm for 20 min at 4 °C and the pellet was either frozen at -80 °C or re-suspended in 25 mL of lysis buffer (20 mM HEPES, 150 mM NaCl, 1 mM mercaptoethanol, 20 mM imidazole, pH 8) for sonication.

2.1.1.2 Large Scale Expression of BSP30b with Auto-induction Media

Auto-induction media was also used for BSP30b expression. A colony taken from an LB plate containing ampicillin was inoculated into auto-induction starter culture

(PA-0.5G, Appendix A2) and was grown overnight at 37 °C with shaking at 200 rpm. This 10 ml starter culture was used to seed into 1 L auto-induction media (ZYP-5052, Appendix A2), which was grown overnight under the same conditions. Induced cell culture was centrifuged at 6,000 rpm for 20 min at 4 °C and the pellet was either frozen at -80 °C or re-suspended in 25 mL of lysis buffer for sonication.

2.1.1.3 Purification of His Tagged BSP30b by Immobilised Metal Affinity Chromatography (IMAC)

Cells from section 2.1.1.1 or 2.1.1.2 were lysed by using a Misonix sonicator (USA) at a setting 4. 20 second bursts were followed by 30 second breaks for 8 times or until sufficient lysis. The lysed suspension was centrifuged at 13,000 rpm for 20 min at 4 °C. The supernatant was filtered through 1 μm, 0.4 μm and 0.2 μm Minisart filters (Sartorius AG, Germany) consecutively.

IMAC (Immobilised Metal Affinity Chromatoghraphy) was performed by using a 5 mL HiTrap™ Chelating HP column (GE Healthcare, Sweden). The filtered supernatant was loaded onto the IMAC column, which was primed with 5 mL 100 mM $NiCl_2$ and equilibrated with appropriate lysis buffer (20 mM HEPES, 150 mM NaCl, 1 mM mercaptoethanol, 20 mM imidazole, pH 8). The column was then attached to an ÄKTA Basic™ FPLC system which monitored protein elution by absorbance at 280 nm. The FPLC system was washed with MQ H_2O while lines A and B were equilibrated in lysis buffer and elution buffer respectively before column attachment. The column was washed with 15 to 25 mL of lysis buffer at a flow rate of 1 mL/min to remove unbound proteins. Bound proteins were removed from column by running a gradient of 0% to 100% elution buffer (lysis buffer + 500mM imidazole) over 80 mL at a flow rate of 1 mL/min. Fractions containing the desired protein were then analysed by SDS-PAGE gel. After each use, the column

was washed with either 20 ml of MQ H_2O or 20 mL of 20% ethanol and stored at 4 °C.

2.1.1.4 SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE) of Purified Protein

12% SDS-PAGE gels were cast in a Hoeffer® gel casting system SE275 (Hoefer® Inc, USA). SDS-PAGE gels consisted of a 5 % acrylamide stacking gel overlaid on a 12 % acrylamide resolving gel. All gels were made with 30% acrylamide with an acrylamide/bis ratio of 37.5:1 (BioRad Laboratories, USA) and were polymerised by the addition of 0.05% (w/v) ammonium persulphate (APS) and 0.05% (v/v) TEMED.

Protein samples were mixed in a 3:1 ratio with 4x SDS loading buffer (250mM Tris HCl pH 6.8, 20% glycerol, 4% SDS, 10% β-mercaptoethanol, 0.025% w/v bromophenol blue) followed by heating at 95 °C for 5 min to denature the protein. 10 μL samples were then loaded on a SDS-PAGE gel, which was running at 50 V until the dye front entered the resolving gel and then running at 150 V until the dye front reached the bottom of the gel.

Gel staining was following the quick coomassie blue staining method (Wong et al., 2000). Briefly, the gel was soaked in 50 mL Fairbanks staining solution A (0.05% coomassie blue, 25% isopropanol, 10% acetic acid) followed by microwaving for 30 s on full power, then the gel was shaken on a plate shaker for 5 min. After removing the stain, 50 mL of Fairbanks staining solution B (0.005% coomassie blue, 10% isopropanol, 10% acetic acid) was added and then microwaving as above. This was repeated for Fairbanks staining solution C (0.002% coomassie blue, 10% acetic acid). After decanting solution C, Fairbanks destaining solution D (10%

acetic acid) was poured into the box containing the gel and then shaken on a horizontal shaker overnight. The SDS PAGE gel was imaged by using OMEGA LUM^{TM} G imaging system (Aplegen).

2.1.1.5 Concentrating BSP30b Samples

Fractions containing BSP30b indicated from SDS-PAGE gel results combined and were loaded to the upper reservoir of a 20 mL Vivaspin concentrator (Sartorius AG, Germany) with a molecular weight cut-off of 10,000 Dalton. The Vivaspin concentrator was spun at 3,000 g at 4 °C until the desired volume or concentration was reached.

2.1.1.6 Concentration Measurement of BSP30b

BSP30b has no tryptophans, and hence it has poor absorbance at 280 nm. Therefore, a Bradford Assay was used to measure its concentration. The Pierce^{TM} 660nm Protein Assay reagent (Thermo Scientific, USA) was used to measure BSP30b's concentration according to the manufacturer's instructions. Briefly, 10 μL of each replicate of standard (Bovine Serum Albumin Standards, Thermo Scientific) and serially diluted BSP30b was pipetted into a microplate well, followed by adding 150 μL of the protein assay reagent to each well. The plate was covered with tin foil and mixed on a plate shaker at medium speed for 1 min and then rested at room temperature for 5 mins. Absorbance at 660 nm was measured by using a plate reader (Thermo Scientific, USA). The plotted standard curve was used to determine the protein concentration in Excel 2010 (Microsoft, USA).

2.1.1.7 Purification of BSP30b by Using Size Exclusion Chromatography

Protein was further purified by size exclusion chromatography (SEC) using a Superdex^{TM} 200 10/300 (S200 10/300) column (GE Healthcare, Sweden), which

has a separation range of 10-600 kDa. SEC was performed on an ÄKTA Basic™ FPLC system, which was washed with MQ H_2O prior to column attachment. The superdex column was equilibrated with SEC buffer (10 mM HEPES pH 8.0, 150 mM NaCl) overnight on an ÄKTA Basic™ according to the set program. Proteins obtained from section 2.1.1.4 were concentrated to 2 mg/mL and filtered with a 0.2 μm filter before loading onto the superdex column. Proteins were eluted with SEC buffer according to their size. Fractions containing the desired protein were then analysed by SDS-PAGE and concentrated using a 20 mL Vivaspin concentrator for further purpose.

Calibration of Superdex™ 75 16/60 GL SEC column was performed using the high and low molecular weight gel filtration calibration kits (GE Healthcare, Sweden) according to the manufacturer's instructions. The molecular weights of unknown proteins were calculated from calibration curves based on proteins of known molecular weights. The curves were prepared by plotting retention volume data (Kav) against the logarithm of the molecular weights of the calibration kit proteins. The molecular weights of unknown proteins can be determined from the curve together with calculated correlation and variance.

2.1.2 Generation, Expression and Purification of BSP30b_F52M_F106M

2.1.2.1 Synthesis of *BSP30b_F52M_F106M* gene

As BSP30b does not contain any methionine, two methionine sites (F52M, F106M) were introduced into the protein sequence for SAD (single-wavelength anomalous dispersion) studies. DNA sequence coding BSP30b_F52M_F106M with restriction sites (*BamHI* at N terminal and *XhoI* at C terminal) was codon optimized and

synthesized by GeneArt (Life Technologies, Carslbad CA). The synthesized sequence was initially inserted into plasmid *pMA-T*.

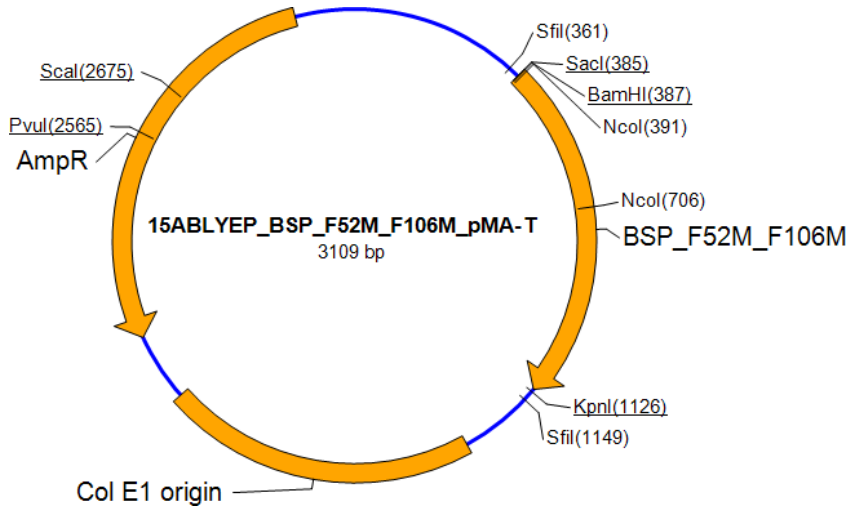

Figure 2.1 Plasmid map of pMA-T with the synthetic gene BSP30b_F52M_F106M inserted. Restriction sites BamHI and XhoI were designed at its N and C terminal, respectively.

2.1.2.2 Preparation of Electrocompetent *E. coli BL21* Cells

A glycerol stock of *E.coli BL21* cells was streaked onto a low salt LB plate and incubated at 37 °C overnight. On the second afternoon, a single colony was inoculated into a 10 mL low salt LB starter culture and grown overnight at 37 °C with shaking at 200 rpm. One microliter of this starter culture was then used to inoculate 1 L of low salt LB, which was grown at 37 °C with shaking at 200 rpm to an OD_{600} of around 0.7. The cells were then chilled on ice for 1.5 hours followed by centrifugation at 4,000 g for 20 mins at 4 °C. After discarding the supernatant immediately, the cell pellet was gently re-suspended in 100 mL of ice cold sterile 10% glycerol and centrifuged again at the same conditions as above. The supernatant was removed again and the pellet was re-suspended in 2 mL of ice cold

sterile 10% glycerol. 50 µL aliquots were flash frozen in liquid nitrogen and stored at -80 °C.

2.1.2.3 Electroporation of Customised Plasmid into *E. coli BL21*

Lyophilized plasmid from GeneArt was initially re-suspended in 50 µL distilled water, 1 µL of which was added into 50 µL of freshly thawed electrocompetent *E. coli BL21* cells on ice. The mixture was then transferred into a 0.2 cm electroporation cuvette (Bio-Rad Laboratories, USA). Electroporation was undertaken with a Bio-Rad Gene Pulser™ (Bio-Rad Laboratories, USA) with settings 2.5 kV, 25 µF capacitance, and 200 Ω resistance. After electroporation, 1 mL of SOC medium was immediately added into transformed mixture and transferred into a 1.5 mL Eppendorf tube followed by shaking incubation at 37 °C for 1 hr. After incubation, 100 µL of the culture was spread on a LB agar plate containing ampicillin (100 µg/mL). The remaining culture was centrifuged at 13,000 rpm for 1 min and the pellet was re-suspended in 100 µL LB medium, which was then spread on another LB agar plate containing ampicillin. The plates were incubated at 37 °C overnight.

2.1.2.4 Primer Design for PCR

In order to confirm the success of transformation of plasmid *pMA-T* into *E. coli BL21*, primers for amplification of the *BSP30b_F52M_F106M* were designed and analysed using Geneious 7.1.5 (Biomatters Ltd, Auckland). Primers were supplied by IDT (Coralville, IA) and were re-suspended in 1x TE buffer (10 mM Tris-HCl, 1 mM EDTA; pH 8.0) to a concentration of 100 µM as stock kept in -20 °C freezer and to a concentration of 10 µM as working stock.

2.1.2.5 Colony PCR (polymerase chain reaction) Amplification of *BSP30b_F52M_F106M*

To confirm the success of cloning in section 2.1.2.3, five randomly selected colonies were re-suspended in 10 µL of MQ H_2O, respectively. 1 µL from each dilution was used as template for PCR using gene specific primers (Table 2.1). HOT FIREPol® DNA polymerase (SOLIS BIODYNE, Estonia) was used for the amplification of gene insert. PCR reactions were carried out in 25 µL volumes. The reagents and their concentrations for PCR were list in table 2.2 and the cycling conditions were described in table 2.3.

Table 2.1 primers used for transformation confirmation of BSP30b_F52M_F106M into E. coli BL21

	Primer sequence	Tm
BSP30b_DM_F	CTGGTTCTGCTGTGTGGTCT	64.4 °C
BSP30b_DM_R	ACTTCTGAGTTCGGCATGGG	64.5 °C

Table 2.2 PCR reaction components for amplification

Components	Concentration
DNA Template	~100 ng
HOT FIREPol® polymerase	2.5 U
PCR Buffer	1X
$MgCl_2$	1.5 mM
dNTPs	0.2 mM
Primers	0.1 µM

Table 2.3 PCR conditions for amplification

Hot start	95 °C	15 min
30 cycles	95 °C 55 °C 72 °C	20s 20s 1kb/min
Final extension	72 °C	10 min

2.1.2.6 Agarose Gel Electrophoresis

PCR amplified DNA from section 2.1.2.5 was separated via agarose gel electrophoresis for visualisation. Samples mixed with 5 x DNA loading dye were running on a 1% agarose gel (made with SYBR Safe^{TM} DNA dye) in 1 x TAE buffer at 90 V for 20 min. PCR amplicons were visualised under UV light and their sizes were determined by comparison against 1 kb+ DNA ladder (Invitrogen).

2.1.2.7 Plasmid DNA Extraction

Colony confirmed containing gene insert was sub-cloned on a LB agar plate containing ampicillin. A 10 mL culture of the sub-cloned cells was grown overnight at 37 °C with shaking at 200 rpm. For plasmid DNA extraction, Qiagen Miniprep Kit (Qiagen, Netherlands) was used for plasmid extraction following the manufacturer's instructions and eluted in 30 – 50 µL of elution buffer depending on preferred final concentration.

2.1.2.8 DNA Quantification

DNA concentration was measured by using a Nanodrop ND-1000 spectrophotometer (Nanodrop technologies, USA). In short, the spectrophotometer was first blanked with 2 µL of elution buffer and then 2 µL of DNA sample was loaded on the machine. The concentration of DNA was determined by its absorbance at 260 nm.

2.1.2.9 Cloning of BSP30b_F52M_F106M into *pProEx Htb*

pProEx Htb is a bacterial expression vector with ampicillin resistance. Plasmid *pProEx Htb* was minipreped as described in section 2.1.2.7.

2.1.2.9.1 Restriction Enzyme Digestion

Restriction enzymes were purchased from Invitrogen (USA). *pProEx Htb* and purchased plasmid were digested separately with restriction enzymes *BamHI* and *XhoI*. In short, restriction enzyme digestion was performed in a final volume of 30 µL using the appropriate buffer as recommended by the manufacturer, at 37 °C for 4 hours with shaking at 400 rpm on a thermomixer (Eppendorf, Germany). The reaction was performed with 20 U of each enzyme and 1000 ng of DNA.

Digested DNA was separated on a 1% agarose gel as described in section 2.1.2.6. The desired bands were extracted from agarose gel using a sterile scalpel blade and DNA was purified using a High Pure PCR Product Purification Kit (Roche Applied Science, Switzerland) or QIAquick Gel Extraction Kit (Qiagen, Netherlands) following the manufacturer's recommendations. DNA was eluted in 30 - 50 µL of elution buffer and its concentration was determined by using NanoDrop as described in section 2.1.2.8.

2.1.2.9.2 DNA Ligation

DNA ligation was performed with 0.1 unit of T4 DNA Ligase (Invitrogen, USA) in 20 µL reaction according to the manufacture's protocol with slight changes. The ligation was carried out as illustrated in table 2.3.

Table 2.4 Ligation reaction components and conditions

5x Ligase Reaction Buffer	4 µL
Insert: Vector Molar Ratio	3:1
Vector Ends	30 fmol
Insert Ends	90 fmol
T4 DNA Ligase	0.1 unit
Autoclaved distilled water	To 20 µL
Temperature	4 °C
Time	overnight

On the second day, 2 µL of this ligated reaction was transformed into *E. coli BL21* following the method as per 2.1.2.3. The transformation success of new plasmid construct was assessed by PCR amplification using primers in table 2.1 and agarose gel electrophoresis as described in section 2.1.2.5.

2.1.2.10 DNA Sequencing

The new plasmid construct was also sequenced by Massey Genome Service (MGS) using Sanger sequencing on an ABI3730 DNA Analyzer (Applied Biosystems, Foster City CA) at Massey University. Sequencing results were visualized and

analysed by using Geneious 7.1.5 (Biomatters, Auckland) to confirm correct insertion.

2.1.2.11 Glycerol Stock Preparation

Glycerol stocks were made for long term storage of transformed strains. In short, 0.5 mL of fresh bacterial culture was mixed with 0.5 mL of 50% glycerol in a 2 mL storage tube, which was then placed into -80°C freezer.

2.1.2.12 BSP30b_F52M_F106M Expression Trial in M9 Minimal Media

Selenomethionine (SeMet) incorporated protein was expressed in *E. coli* BL21 using the methionine pathway inhibition method (Hendrickson et al., 1990). Methionine biosynthesis was blocked by inhibiting aspartokinases in the presence of a high concentration of isoleucine, lysine, and threonine. This allows the incorporation of SeMet when it is added into the growth media.

The expression ability of BSP30b_F52M_F106M in M9 minimal media (Appendix A2) was tested before the incorporation of selenomethionine (SeMet). A single colony of the transformed *E. coli BL21* was inoculated into 10 mL of LB media with appropriate antibiotic and grown overnight at 37 °C with shaking at 200 rpm. On the second day morning, the culture was centrifuged at 4,600 rpm for 10 min and the cell pellet was re-suspended in 1 mL of M9 minimal media. This culture was used to inoculate 1 L of M9 minimal media with appropriate antibiotic and grown at 37°C with shaking at 200 rpm. When the OD_{600} of the culture reached 0.4-0.6, 100 mg each of lysine, phenylalanine and threonine, 50 mg each of isoleucine and valine, 60 mg of methionine were added and the growth was continued for a further 15 min. 0.5 mM IPTG was added for protein induction and the culture was grown overnight. Induced cell culture was centrifuged at 6,000 rpm for 20 min at

4 °C and the pellet was either frozen at -80 °C or re-suspended in 25 mL of lysis buffer (20 mM HEPES, 150 mM NaCl, 1 mM mercaptoethanol, 20 mM imidazole, pH 8) for sonication.

BSP30b_F52M_F106M was first purified by IMAC using a 5 mL HiTrap™ Chelating HP column as per 2.1.1.3. Fractions containing the protein of interest were analysed by SDS PAGE and the protein expression level was determined from the SDS PAGE gel result.

BSP30b_F52M_F106M was further purified with SEC using a Superdex™ 200 10/300 GL column as per 2.1.1.7. Fractions of interest were analysed by SDS PAGE and protein was concentrated using a 20 mL Vivaspin concentrator for further purpose.

2.1.2.13 Incorporation of Selenomethionine (SeMet) into BSP30b_F52M_F106M

The procedure to express SeMet incorporated protein is similar to the expression trial in minimal media as per 2.1.2.12. The only difference is that instead of using 60 mg of methionine, 60 mg of selenomethionine was added to the media. Induced cell culture was centrifuged at 6,000 rpm for 20 min at 4 °C and the pellet was either frozen at -80 °C or re-suspended in 25 mL of lysis buffer (20 mM HEPES, 150 mM NaCl, 1 mM mercaptoethanol, 20 mM imidazole, pH 8) for sonication.

2.1.2.14 Purification of BSP30b SeMet derivative

Protein was first purified by IMAC using a 5 ml HiTrap™ Chelating HP column as per 2.1.1.3. Fractions containing protein of interest were analysed by 12% SDS PAGE.

An anion exchange chromatography step was added after IMAC to improve the purity of BSP30b SeMet derivative. Anion exchange chromatography was performed using a Mono Q anion exchange column (GE Healthcare) on ÄKTA Basic™ HPLC system. Briefly, the Mono Q anion exchange column was equilibrated with 10 ml of buffer A (20 mM HEPES pH 7.5, 50 mM NaCl, 0.1 mM EDTA) and then the concentrated BSP30b_F52M_F106M SeMet derivative was loaded into the sample loop. Proteins were eluted from column by running a gradient of 0% to 100% buffer B (20 mM HEPES pH 7.5, 0.1 mM EDTA, 600 mM NaCl) over 80 mL at a flow rate of 1 mL/min. Fractions containing the desired protein were then analysed by 12% SDS-PAGE.

Protein was further purified with SEC using a Superdex™ 200 10/300 GL column as per 2.1.1.7. Purified protein was analysed by SDS PAGE and concentrated using a 20 mL Vivaspin concentrator for further purpose.

2.1.3 Expression and Purification of MBP Fused BSP30b_F52M_F106M

2.1.3.1 Cloning of *BSP30b_F52M_F106M* into *pMAL-c2X*

The *BSP30b_F52M_F106M* gene was cloned into the maltose binding protein fusion tag *E. coli* expression vector *pMAL-c2X* (New England Biolabs, USA) as follows. The *BSP30b_F52M_F106M* gene was amplified from the *pMA-T* vector via PCR using forward and reverse primers (Table 2.5) containing the restriction enzyme sites *BamHI* and *PstI*, respectively.

Table 2.5 Primers used for cloning of BSP30b_F52M_F106M into pMAL-c2X

	Primer sequence	Tm
BSP30b_DM_F	CTGGTTCTGCTGTGTGGTCT	64.4 °C
pBAD_R	GATTTAATCTGTATCAGG	49.7 °C

Purified PCR product and miniprepped *pMAL-x2X* plasmid were digested with restriction enzymes BamHI and PstI, respectively. After gel electrophoresis and purification, both digested products were ligated and then transformed into *E. coli BL21* as per 2.1.2.9. The success of cloning was confirmed by colony PCR using primers in table 2.5 and agarose gel electrophoresis. Glycerol stocks of transformed strains were made for long term storage.

2.1.3.2 Large Scale Expression and Purification of MBP Tagged BSP30b_F52M_F106M

A 10 mL culture of *BSP30b_F52M_F106M-pMAL-c2X* in *E. coli BL21* in LB media with ampicillin (100 μg/mL) and glucose (2 g/L) was grown overnight at 37 °C with shaking at 180 rpm. This culture was used to inoculate 1 L of fresh pre-warmed LB media with ampicillin (100 μg/mL) and glucose (2g/L), which was then grown at 37 °C, shaking at 180 rpm until it reached an OD_{600} of 0.4 – 0.6. Protein expression was induced by adding IPTG to a final concentration of 1 mM and growth was continued at 37 °C overnight with shaking at 180 rpm. The culture was pelleted by centrifugation at 6,000 rpm for 20 minutes at 4 °C, and re-suspended in 20 mL of lysis buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 1 mM EDTA) and a

Complete Mini EDTA-free protease inhibitor cocktail tablet added (Roche Diagnostics, USA) prior to cell lysis. Cells were lysed on ice using a sonicator® ultrasonic processor XL fitted with a flat probe at level 4 for 8 x 20 seconds (Misonix Incorporated, USA). After sonication, the sample was centrifuged for 20 minutes at 13,000 rpm at 4 °C to pellet the cell debris. The supernatant was filtered through 1.2 µm, 0.45 µm, and 0.2 µm Minisart syringe filters (Sartorius AG, Germany) prior to protein purification.

A 5 mL MBP Trap™ HP column (GE Healthcare, Sweden) was prepared by washing with 10 mL MQ H_2O and equilibrating with 15 mL of lysis buffer. The filtered supernatant was then loaded onto the column using a syringe and the column was attached to an ÄKTA Basic™ FPLC system which monitors protein elution by absorbance at 280 nm. The FPLC system was washed with MQ H_2O and lines A and B equilibrated in lysis buffer and elution buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 1 mM EDTA, 10 mM maltose) respectively prior to column attachment. At a flow rate of 1 mL/min, the column was washed with 50 mL of lysis buffer to remove unbound proteins and then bound proteins were eluted with 25 mL of 100% elution buffer. Fractions containing protein of interest were analysed by 12 % SDS-PAGE. After use, the MBP Trap™ HP column was regenerated by washing with 10 mL of 0.1% SDS, and 20 mL of MQ H_2O and rinsed with 10 mL of 20% ethanol for storage at 4 °C.

Protein was further purified with SEC using a Superdex^{TM} 200 10/300 GL column as described in 2.1.1.7. The purified protein was analysed by SDS PAGE and concentrated to the desired concentration for further purpose.

2.2 Materials and Methods Related to Chapter 4

2.2.1 Crystallization Trials for BSP30b

2.2.1.1 Robotic Crystallization Screens of BSP30b

Concentrated protein after SEC purification was subjected to robotic crystallisation screens, which were performed using the high throughput Mosquito® crystal robot screening system (TTP Labtech Ltd, UK). Sitting drop vapour diffusion method was performed with PEGRx HT, Index HT, Crystal Screen HT, and SaltRx HT crystallisation screens (Hampton Research, Aliso Viejo CA). 100 μL of each condition was dispensed into 96 well Intelli-plate (Hampton Research) reservoirs manually and then 100 nL of protein and reservoir solutions were robotically loaded into the crystallisation well respectively. All Intelli plates were sealed with ClearSeal™ film (Hampton Research) and stored at 18 °C for crystal growth. All of the crystallisation wells were checked routinely for crystal "hits" using a bench top microscope.

2.2.1.2 Fine Screen of Native BSP30b

Conditions showing crystal "hits" from robotic screening were chosen for fine screens, which were performed using hanging drop vapour diffusion method in 24 well plates (Hampton Research). 500 μL of robotic screen condition solution, which is also called 'mother liquor', was pipetted into a well. 1 μL of protein and 1 μL of 'mother liquor' were mixed on a 22 mm square siliconized coverslip (Hampton Research), which was then inverted over the same well and sealed by grease (Glisseal N). The plate was left on a steady shelf at 18 °C and was checked routinely for crystal "hits" using a bench top microscope.

Conditions optimised by modifying the pH, crystallisation component concentrations were also set up for fine screens. Fine screens using conditions (Appendix A6) from former study of BSP30b in Vic Arcus' lab were also set for crystallization.

2.2.1.3 Additive Screen

Additive screens based on the most promising conditions ware performed using the high throughput Mosquito® crystal robot screening system as mentioned above. Commercial reagent (Additive Screen HT) purchased from Hampton research was used. Conditions having crystal "hits" were reproduced with fine screen.

2.2.1.4 Silver Bullet Screen

Silver bullet screen based on the most promising conditions ware performed using the high throughput Mosquito® crystal robot screening system as mentioned above. Commercial reagent (Silver Bullets) purchased from Hampton research was used. Conditions having crystal "hits" were reproduced with fine screen.

2.2.1.5 Trypsin Digestion

In situ proteolysis using trypsin for crystallization of BSP30b was set up, intending to improve crystal quality. Trypsin digestion of BSP30b was tested at room temperature within a 1.5 mL Eppendorf tube. Trypsin from bovine pancreas (Sigma Aldrich, St. Louis MO) was added to BSP30b at ratio of 1 mg:100 mg and 20 μL of this mixture was pipetted into a new Eppendorf tube containing 5 μL of SDS PAGE loading dye after 15s, 30s, 1 min, 5 min, 10min, 1 hr, and 4 hr, respectively. Notably, those new Eppendorf tubes containing digested BSP30b were immediately heated at 95 °C for 5 mins on a thermomixer after each incubation in order to inactivate trypsin and denature BSP30b. All of the digested BSP30b was run on a

12 % SDS PAGE gel and stained with coomassine (detailed protocol in section 2.1.1.4). The best digestion time was determined from this gel result.

To set up the crystallization trial after digestion, the proper amount of trypsin was added to the protein. After the optimised digestion time, fine screen was immediately set for crystallization. Crystallization progress was monitored and recorded routinely.

2.2.1.6 Streak Seeding

Streak seeding is a method to introduce pre-formed crystal nuclei into a fine screen drop with a cat whisker. The fine screen was set normally as described in section 2.2.1.2 and pre-equilibrated for 2-3 hr before seeding. A cat whisker was streaked through an existing crystal drop and then the cat whisker was run through the pre-equilibrated drop starting from outside the drop and ending outside the drop in a smooth motion. The cat whisker was rinsed with 70% ethanol before seeding a different drop. Crystallization progress was monitored and recorded routinely.

2.2.1.7 Batch Seeding

To prepare the seed stock, 2 μL of the crystal drop was added into 18 μL of mother liquor in an Eppendorf tube and crystals were crushed with vortex. Those crushed crystals were serially diluted 10,000 times and the final diluted one was used as the seed stock. The hanging drop of fine screen was set up with 1 μL of BSP30b and 1 μL of the seed stock. Crystallization progress was monitored and recorded routinely.

2.2.1.8 Heavy Metal Soaks

In order to solve the phase problem, selected crystals from fine screens were soaked in heavy metal solutions in order to obtain anomalous signal during X-ray diffraction experiments. The binding ability of several heavy metals to BSP30b was determined by using a native PAGE gel shift method (Boggon & Shapiro, 2000). The casting of native PAGE gel is similar to SDS-PAGE gel (section 2.1.1.4). The only difference is that SDS is not added to make native PAGE gel. To perform the gel shift assay, 5 µL of each heavy metal compound (HgAC, $HgBr_2$, $Hg(CN)_2$, $KAu(CN)_2$, K_2PtCl_4, K_2PtCl_6, $HgCl_2$) at concentrations of 10 mM was added to 20 µL of BSP30b solution, respectively. 10 µL of each mixture or BSP30b itself together with 2 µL of 5x native gel loading dye (300 mM Tris-HCl pH 6.8, 50% (v/v) glycerol, and 0.05% (w/v) bromophenol blue) was run on a 12 % native PAGE gel. Those heavy metal compounds whose mixture with BSP30b having shifted bands compared to BSP30b itself on native PAGE gel will be used for crystal soaking experiment.

Crystal soaking was performed with a 24 well fine screen plate. 5 µL of mother liquor containing 2 mM of heavy metal was pipetted on a cover slip and BSP30b crystals were looped into the drop. The cover slip was inverted on a well having 500 µL of mother liquor and sealed with grease for 2 days. Those crystals were frozen in liquid nitrogen without cryo-protectant and taken to Australian Synchrotron Center for X-ray diffraction experiment.

2.2.2 X-ray Diffraction and Data Analysis of BSP30b Crystals

2.2.2.1 Crystal X- ray Diffraction Testing

X-Ray diffraction testing of crystals was performed on an Agilent Supernova X-Ray Diffractometer (Agilent Technologies, Liberty Lake WA) at the University of Waikato using a Titan S2 CCD detector. Home source X-Rays were generated using CuKα radiation from a sealed source generator and reflections were detected using a CCD detector. In order to reduce radiation damage to crystals, crystals were soaked into a cryo-protectant (crystallization conditions + 30% PEG 400) before mounting on the X-Ray diffractometer. The cryo-protectant itself was tested first for X-Ray diffraction to detect if any ice rings were formed which would interfere with the diffraction results of protein crystals.

2.2.2.2 X-ray Diffraction Data Collection for BSP30b at the Australian Synchrotron

Crystals of BSP30b from fine screens were initially flash-frozen in liquid nitrogen with cryo-protectant (crystallization conditions + 30% PEG 400) prior to data collection at the Australian Synchrotron. BSP30b crystals tended to crack when transferred from mother liquor to cryo-protectant, and hence crystals were flash-frozen in liquid nitrogen without cryo-protectant during all of the following experiments.

Data were collected at the macromolecular crystallography beamline (MX1 or MX2) at the Australian Synchrotron using an ADSC Quantum 210r detector.

Crystals having diffraction better than 3 Å at initial shot were chosen for further data collection. 360° of data were collected using individual images with an oscillation of 1 degree.

2.2.2.3 General Data Processing for Native BSP30b

The dataset was indexed and integrated in iMosfilm (Leslie & Powell, 2007) to integrate the reflections and to determine the unit cell parameters in the CCP4 platform suite 6.4.0 (Collaborative Computational Project, 1994). Briefly, two images that were 90 degrees apart in Phi were used to search for differaction spots. Autoindexing was then carried out to determine space group and cell paramaters. The cell parameters were further refined by Cell Refinement to obtain more accurate cell parameters. The distribution of the intensities of partially recorded reflections over the images on which they occur was used to refine the unit cell, crystal orientation and mosaic spread. All of the images were integrated into an mtz file to produce a list of reflection intensities.

The output mtz file from iMosfilm was then scaled and truncated using Aimless (Evans & Murshudov, 2013). The programme Pointless was used first to determine the point-group symmetry the likely space group. The programme Aimless was then used to scale symmetry related intensities and produce statistics on data quality. Ctruncate was executed to calculate the structure amplitudes |F| from the intensities and to output various diagnositic statistics, for example to detect twinning. The newly created file was examined for unfavoured R_{merge}, $I/\sigma(I)$ and data completeness and rescaled to improve these parameters if necessary.

In order to determine the number of BSP30b molecules in the ASU and the solvent content of the unit cell, Matthews_coeff in the CCP4 platform was used. The scaled mtz file from Aimless together with the molecular mass of BSP30b were used to calculate the molecule number of ASU and the solvent content.

2.2.2.4 X-ray Diffraction Data Collection for Heavy Metal Soaked BSP30b

Heavy metal soaked BSP30b crystals were flash-frozen in liquid nitrogen with or without cryo-protectant. Data were collected at the macromolecular crystallography beamline MX2 at the Australian Synchrotron using an ADSC Quantum 210r detector. Crystals having diffraction better than 3.5 Å resolution at the initial shot were chosen for further data collection. 360 diffraction images were collected with oscillation of 1 degree.

2.2.2.5 General Data Processing for Heavy Metal Soaked BSP30b

The dataset was indexed and integrated in iMosfilm to integrate the reflections and to determine the unit cell parameters in the CCP4 platform suite 6.4.0. The output mtz file from iMosfilm was then scaled and truncated using Aimless. This new created file would be examined by eye for unfavoured R_{merge}, $I/\sigma(I)$ and data completeness and rescaled to improve these parameters if necessary. Anomalous signal was checked from the Aimless log file.

2.2.3 Crystallization Trials for His Tagged BSP30b_F52M_F106M

Concentrated BSP30b_F52M_F106M after SEC purification was subjected to robotic crystallisation screens followed the instructions in 2.2.1.1. All of the crystallisation wells were checked for crystal "hits" routinely. Fine screen will be set with those conditions having crystal "hits".

Fine screens of BSP30b_F52M_F106M were also laid down with the old 24 conditions as per 2.2.1.2. Crystallization progress was monitored and recorded routinely. Crystals from fine screen were frozen in liquid nitrogen and taken to Australian Synchrotron for X-ray diffraction experiment.

2.2.4 X-ray Diffraction and Data Analysis of BSP30b_F52M_F106M Crystals

2.2.4.1 His Tagged BSP30b_F52M_F106M X-ray Diffraction Experiment at Australian Synchrotron

Crystals were flash frozen in liquid nitrogen without cryo-protectant and data were collected at the macromolecular crystallography beamline MX2 at the Australian Synchrotron using an ADSC Quantum 210r detector.

In order to capture the anomalous signal generated from selenium, a special wavelength of X-ray (0.9537 Å, just below the K shell absorption edge of selenium) was chosen for data collection. Crystals having diffraction better than 2.5 Å at initial shot were chosen for further data collection. 360 diffraction images were collected with oscillation of 1 degree.

2.2.4.2 General Data Processing

The dataset obtained from Australian Synchrotron was processed as per 2.2.2.3.

2.2.4.3 Phase Determination

The re-scaled mtz files were input into AutoSol (Terwilliger et al., 2009) within the PHENIX platform (Adams et al., 2011) for searching heavy atom sites, experimental phases, and model-building. Heavy atom selenium was added into the input file interface of AutoSol to help the phase processing.

2.2.4.4 Model Building and Refinement

Initial structure was built using AutoBuild (Terwilliger et al., 2008) with phases from AutoSol. The structure was then refined using *phenix.refine* (Afonine, Grosse-Kunstleve, & Adams, 2005) in the PHENIX platform. Subsequent cycles of manual

model building using COOT 0.7 (Emsley & Cowtan, 2004) and refinement in *phenix.refine* were performed until the R free could not be improved.

2.2.5 Crystallization of MBP Tagged BSP30b_F52M_F106M

2.2.5.1 Robotic Crystallization Screens

Concentrated protein after SEC purification was subjected to robotic crystallisation screens, which were performed using the high throughput Mosquito® crystal robot screening system (TTP Labtech Ltd, UK). Sitting drop vapour diffusion method was performed with PEGRx HT, Index HT, Crystal Screen HT, and SaltRx HT crystallisation screens (Hampton Research, Aliso Viejo CA) as per 2.2.1.1. Crystallization progress was monitored routinely and the conditions showing crystal "hits" were reproduced with fine screen.

2.2.5.2 Fine Screen

Fine screens of MBP tagged BSP30b_F52M_F106M was also performed with the old 24 conditions as in 2.2.1.2. Crystallization progress was monitored and recorded routinely. Crystals were frozen in liquid nitrogen and taken to Australian Synchrotron for X-ray diffraction experiment.

2.2.6 X-ray Diffraction and Data Analysis of MBP Tagged BSP30b_F52M_F106M

2.2.6.1 MBP Tagged BSP30b_F52M_F106M X-ray Diffraction Experiment at Australian Synchrotron

Data was collected at the macromolecular crystallography beamline MX2 at the Australian Synchrotron using an ADSC Quantum 210r detector. Crystals having

diffraction better than 3 Å at initial shot were chosen for further data collection. 360 diffraction images were collected with oscillation of 1 degree.

2.2.6.2 General Data Processing for MBP Fused BSP30b_F52M_F106M

The dataset was indexed and integrated in iMosfilm to integrate the reflections and to determine the unit cell parameters in the CCP4 platform suite 6.4.0. The output mtz file from iMosfilm was then scaled and truncated using Aimless. This new created file would be examined by eye for unfavoured R_{merge}, $I/\sigma(I)$ and data completeness and re-scaled to improve these parameters if necessary.

Phase information was derived using molecular replacement method. Specifically, the scaled file was input into Phaser-MR (McCoy et al., 2007) within PHENIX platform. The pdb files of MBP (PDB code: 4MBP or 1ANF) or the pdb file of the solved BSP30b in this study was used as the search model for molecular replacement. The generated structure was then refined using *phenix.refine* in PHENIX platform. Subsequent cycles of manual refinement using COOT 0.7 and refine in *phenix.refine* were performed until the R free dropped to the desired value.

2.3 Materials and Methods Related to Chapter 5

2.3.1 Expression and Purification of BSP30b_F52M_F106M with Its Potential Ligands

Since better quality crystals were produced by BSP30b_F52M_F106M, this mutant was used for crystallization with potential ligands. Protein expression was conducted in LB media.

A colony taken from an LB plate containing ampicillin was inoculated into 10 mL LB culture with appropriate antibiotic. The culture was grown overnight at 37 °C

with shaking at 200 rpm. The following morning, this 10 mL culture was used to seed a 1 L LB culture with appropriate antibiotic in a 2 L baffled conical flask, which was then grown at 37 °C with shaking at 200 rpm until the OD_{600} reached ~ 0.6. Protein expression was induced by adding 1 mM IPTG and growth was continued overnight under the same conditions. Induced cell culture was centrifuged at 6,000 rpm for 20 min at 4 °C and the pellet was either frozen at -80 °C or re-suspended in 25 mL of lysis buffer (20 mM HEPES, 150 mM NaCl, 1 mM mercaptoethanol, 20 mM imidazole, pH 8) for sonication.

Protein was first purified by IMAC using a 5 ml HiTrap™ Chelating HP column as per 2.1.1.3. Fractions containing protein of interest were analysed by 12% SDS PAGE.

Protein was further purified by anion exchange chromatography and SEC as per 2.1.2.14. Purified protein was then concentrated to 2 mg/mL 20 μL of either linolenic acid, ethyl linoleate, glycerol monolinoleate, or oleic acid (Fig. 2.2) was added to protein solution and this mixture was incubated at 4 °C overnight. On the second day, this mixture was purified again with SEC on ÄKTA Basic™ using Superdex™ 200 10/300 column to remove the extra lipid. Fractions from SEC were analysed by SDS-PAGE and protein was concentrated to desired concentration for crystallization.

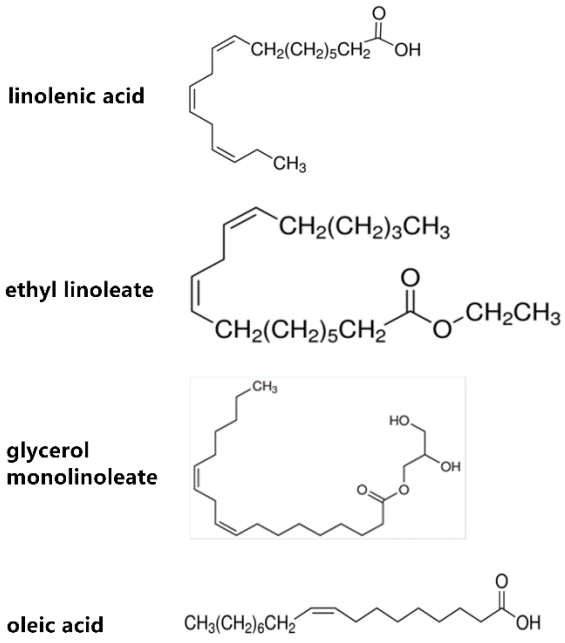

Figure 2.2 Fatty acids used for co-crystallization. Linolenic acid, ethyl linoleate, glycerol monolinoleate, and oleic acid were used to co-crystallize with BSP30b mutant separately.

2.3.2 Crystallization of BSP30b mutant with Potential Ligands

BSP30b mutant purified from 2.3.1 with different fatty acids was concentrated and fine crystallization screens were set up separately using the old 24 crystallization conditions (Appendix A6). Crystallization progress was monitored and recorded routinely. Crystals were flash-frozen in liquid nitrogen without cryo-protectant and taken to Australian Synchrotron for X-ray diffraction experiment.

2.3.3 X-ray Diffraction and Data Analysis of BSP30b Mutant with Potential Ligands

2.3.3.1 X-ray Diffraction Experiment at Australian Synchrotron of BSP30b_F52M_F106M with Potential Ligands

Crystals were flash-frozen in liquid nitrogen without cryo-protectant. Data was collected at the macromolecular crystallography beamline MX2 at the Australian

Synchrotron using an Eiger 16M detector. Crystals having diffraction better than 2.5 Å at initial shot were chosen for further data collection. 3600 diffraction images were collected with 0% attenuation, oscillation of 0.1 degree in 36 seconds.

2.3.3.2 Structure Determination

The scaled file obtained from Australian Synchrotron was examined by eye for unfavoured R_{merge}, I/σ(I) and data completeness, and rescaled to improve these parameters if necessary.

The rescaled mtz file was subjected to molecular replacement. Molecular replacement was performed using Phaser-MR in PHENIX platform with the solved BSP30b structure as the reference model.

The structure model was manually refined in COOT with the 2mFo-dFc electron density map contoured at 1 sigma. Ligand searched from the monomer library was manually fitted into the empty positive electron density contour. Several rounds of refinement using *phenix.refine* in PHENIX and manual refinement in COOT was performed until the R free dropped to the desired value.

2.3.4 Cloning of *BSP30b* and *GFP* into Plasmid *pProEx Htb*

In order to create fused GFP_BSP30b construct, sf*GFP* and *BSP30b* were cloned into *pProEx Htb* sequentially.

sfGFP in plasmid *pIM1773* (kindly given by Dr Jo Hicks) was used for this experiment. *sfGFP* sequence was amplified with primers containing *BamHI* and *PstI* restriction digestion sites (Table 2.6). Purified PCR product and *pProEx Htb* plasmid were digested with restriction enzymes BamHI and PstI, respectively. After gel electrophoresis and purification, both digested products were ligated and then

transformed into *E. coli BL21* as per 2.1.2.9. The success of cloning was confirmed by colony PCR. Glycerol stocks of transformed strains were made for long term storage.

Table 2.6 primers used for cloning of GFP into pProEx Htb

	Primer sequence	Tm
GFP_F	GGTGGTGGATCCGGCTCCAAAGGCGAGGAA	72.7 °C
GFP_R	GGTGGTCTGCAGCTTGTATAATTCATCCAT	60.3 °C

The cloning of *BSP30b* into *pProEx Htb_GFP* will follow the same procedures. *BSP30b* sequence was amplified with primers containing *XhoI* and *hindIII* restriction sites (Table 2.7). Purified PCR amplicon and *pProEx Htb_sfGFP* were digested with XhoI and hindIII, respectively. After gel electrophoresis and purification, both digested products were ligated and then transformed into *E. coli BL21*.The success of cloning was confirmed by colony PCR. Glycerol stocks of transformed strains were made for long term storage.

Table 2.7 primers used for cloning of BSP30b_F52M_F106M into pProEx Htb

	Primer sequence	Tm
BSP_XhoI	GGTGGTCTCGAGGGAGCCTGCCGGATATTCGTG	70.4 °C
BSP_HindIII	GGTGGTAAGCTTTTAACGGCTACCTTCCGG	64.7 °C

2.3.5 Large Scale Expression and Purification of His Tagged GFP_BSP30b

2.3.5.1 Large Scale Expression of GFP_BSP30b

The expression of GFP_BSP30b in 1 L LB media was carried out in large scale as described in 2.1.1.1. Induced cell culture was centrifuged at 6,000 rpm for 20 min at 4 °C and the pellet was either frozen at -80 °C or re-suspended in 25 mL of lysis buffer (20 mM HEPES, 150 mM NaCl, 1 mM mercaptoethanol, 20 mM imidazole, pH 8) for sonication.

2.3.5.2 Purification of GFP_BSP30B with IMAC and SEC

The lysed suspension was centrifuged at 13,000 rpm for 20 min at 4 °C. The supernatant was filtered through 1 μm, 0.4 μm and 0.2 μm Minisart filters (Sartorius AG, Germany) consecutively before loading on the 5 ml HiTrap™ Chelating HP column.

IMAC was performed as described in 2.1.1.3. Fractions containing GFP_BSP30b were then analysed by 12% SDS-PAGE gel.

GFP_BSP30b was further purified with SEC as per 2.1.1.7 using Superdex™ 200 10/300 (S200 10/300) column (GE Healthcare, Sweden). Fractions containing GFP_BSP30b were then analysed by 12% SDS-PAGE gel and concentrated for further purpose.

2.3.5.3 Concentration Measurement of GFP_BSP30b

Since GFP_BSP30b has decent light absorbance at 280 nm, protein was quantified using a Nanodrop ND-2000 spectrophotometer (Nanodrop Technologies, USA). Protein concentration was calculated using the Beer's Law equation and the

theoretical molar extinction coefficient was calculated using the online ProtParam tool (https://web.expasy.org/protparam/).

2.3.6 Fluorescent Microscopy of GFP_BSP30b Mixed with Olive Oil

20 µL of olive oil was mixed with 500 µL of GFP_BSP30b at concentration of 2 mg/ml. This mixture was vortexed for 2 mins and left at room temperature for 20 mins for equilibration. 20 µL of this mixture was pipetted on a glass slide and then covered with a cover slip. The slide having the mixture was viewed by using a ZEISS AXIOSTAR (LLC, US) bench top microscope. Images were captured at 40 x magnification with filter 1 (blue light) or filter 2 (green light) using a Nikon Coolpix 4500 camera (Tokyo, Japan).

The mixture was further viewed under a fluorescent microscope at Waikato Biological Imaging Facility (the University of Waikato, New Zealand). Images were captured at 100 x magnification using a Leica BFP/GFP filter block and Olympus DP70 camera.

2.3.7 Transmission Electron Microscopy (TEM) of Protein Samples

Emulsion samples for negative stain electron microscopy were prepared as follows: 5 µL of freshly prepared emulsions were applied to a 300 mesh copper EM grid (GCU300 ProSciTech, Australia) covered with carbon-film supported by plastic that had been previously rendered hydrophilic by glow discharge. After 90s, the emulsion sample was blotted from the edge of the drop with filter paper (Whatman No 1) and washed three times with filtered MQ H_2O. The specimen was now stained with 5 µL of 1% uranyl acetate for 45s followed by blotting and restaining for 25s.

The sample was then blotted again and air dried. Images were captured at magnification of 40,000 using either a Phillips CM12 or a FEI TecNai G_2 electron microscope. Both electron microscopes were equipped with a LaB_6 filament and operated at 120 kV. Images with scale bars of 0.2 µm, 0.5 µm and 1 µm were recorded at 52,000 x, 40,000 x, and 11,500 x magnification respectively.

2.3.8 Binding Ability of GFP_BSP30b to Rumen Samples

2.3.8.1 Rumen Fluid Collection and Sample Preparation

Rumen fluid was collected from a fistulated cow at AgResearch (Palmerston North, New Zealand). After filtration through cheesecloth to remove large feed particles, the fluid was kept at 39 °C in a water bath with CO_2 supplied. This filtered fluid will be used directly or separated into bacterial and protozoal fractions.

2.3.8.2 Binding of GFP_BSP30b to Rumen Fluid

Purified GFP_BSP30b protein was taken to Agresearch and serially diluted 1:100, $1:10^4$, and $1:10^6$ with sterile MilliQ water before mixing with rumen fluid. The filtered rumen fluid was then mixed with GFP_BSP30b dilutions separately. 20 µL of each mixture was loaded on a slide and viewed under fluorescent microscope at AgResearch. Images were captured at 40 x or 100x magnification with blue light or bright field.

2.3.8.3 Binding Assay of GFP_BSP30b to Rumen Bacterial Fraction or Protozoal Fraction

To obtain the bacterial and protozoal fractions, rumen fluid was centrifuged at 300 g for 5 min. The pellet, which contained the protozoa and some bacteria, was diluted with Saliva Buffer (142 mM Na_2HPO_4, 84 mM $NaHCO_3$, 100.1 mM $KHCO_3$, 60.1

mM Urea, and 58.4 mM NaCl) to the same volume as the supernatant. The supernatant, which then contained the bulk of the bacteria, was centrifuged at 15,000 g for 15 min to pellet the bacteria. The bacteria pellet was then re-suspended with Saliva Buffer to the same volume as the supernatant.

Serially diluted protein was mixed with protozoal or bacterial fractions at a ratio of 1:1. 20 μL of each mixture was pipetted on a slide and viewed under fluorescent microscope. Images were captured at 100 x magnification with blue light or bright field at AgResearch.

Chapter Three. Expression and Purification of BSP30b and its derivatives

3.1 Introduction

This chapter focuses on the expression and purification of BSP30b in the *pTriex-4 Ek/LIC* vector, BSP30b_F52M_F106M with or without SeMet incorporation in the *pProEx Htb* plasmid, and MBP_BSP30b_F52M_F106M in the *pMAL-c2X* vector.

pTriex-4 Ek/LIC containing BSP30b was kindly given by Tom Wheeler at AgResearch and the protein had been previously expressed in *E. coli Origami (DE3) pLacI* cells. This plasmid was later transformed into *E. coli* BL21 in the Arcus lab by Dr Judith Burrows and expression trials had been performed using LB media. During this study, only large scale expression has been conducted using both LB media and auto-induction media.

My strategy was to take a synthesized *BSP30b_F52M_F106M* gene and ligate this into *pProEx Htb* for protein expressed in M9 minimal media. This would then allow the protein to be further expressed with SeMet incorporation. This protein would then be purified and crystallised for SAD experiments to solve the phase problem.

A second strategy was to use MBP as a fusion protein to facilitate crystallization. A previous study has shown that SPLUNC1 was successfully crystallized as MBP-SPLUNC1 fusion protein (Ning et al., 2014). Hence, MBP_BSP30b_F52M_F106M was constructed for this study.

All of the expressed proteins above were purified using ÄKTA Basic™ FPLC system. The purified proteins were subjected to crystallization trials for X-ray diffraction experiment.

3.2 BSP30b Expression in Both LB and Auto-induction Media

Cloning and small scale expression trials were previously conducted and optimized by Dr Judith Burrows in the Arcus lab. Protein expression cultures then were scaled up and BSP30b was expressed in *E. coli BL21* cells as a 29.6 kDa recombinant protein fused to an N-terminal 6x histidine tag. Large scale expression was conducted with both LB media and auto-induction media.

BSP30b was first purified with a 5 ml Ni Trap column on ÄKTA Basic™ and analysed by SDS-PAGE. Protein expressed in LB media has two small peaks (A and B) on chromatogram for 280 nm (Fig. 3.1a). SDS PAGE gel results indicated that most of the weak binding impurities were eluted between fractions A5 and A13 (Fig. 3.1a), corresponding to peak A on the chromatogram. BSP30b was eluted between fractions B5 and C2, corresponding to peak B on chromatogram.

BSP30b was expressed at a much higher levels in auto-induction media compared to LB media. Instead of having two small peaks, there was only one large peak shown on this chromatogram. SDS PAGE results suggested that BSP30b and impurities were eluted at the same time between B8 and C4, corresponding to peak A on chromatogram (Fig. 3.1b).

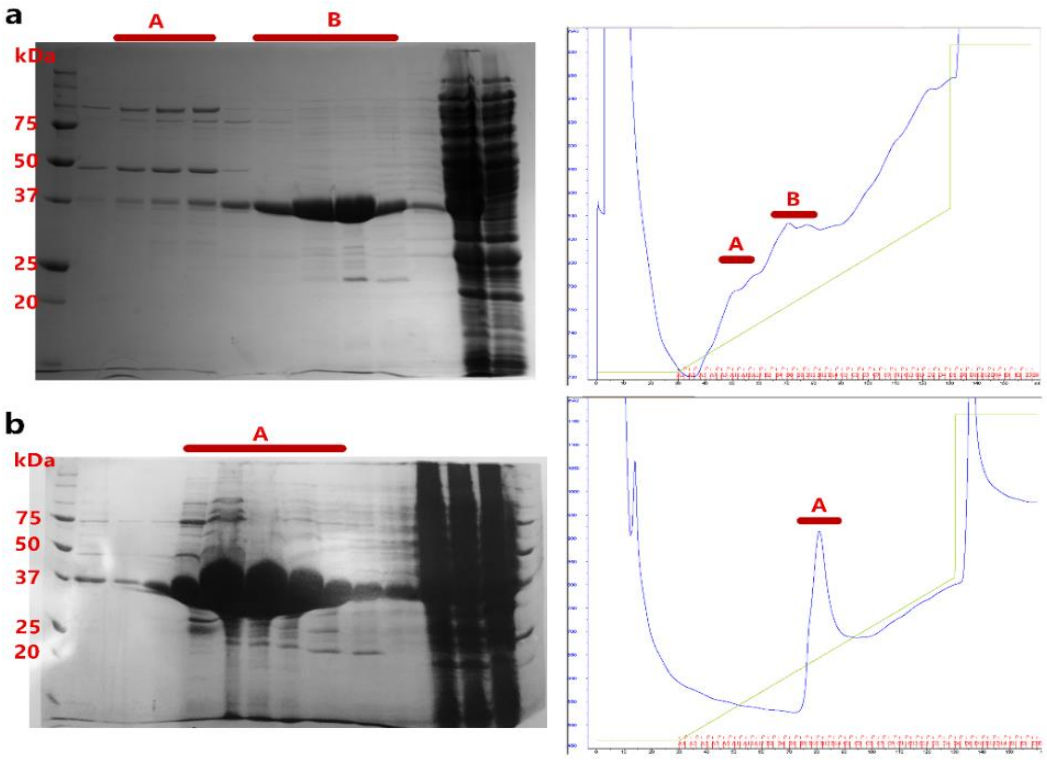

Figure 3.1 IMAC purification and corresponding SDS-PAGE analysis of BSP30b. Chromatogram of UV absorbance profile and its corresponding fractions on SDS PAGE when BSP30b was expressed in LB (a) and auto-induction (b) media, respectively. Peaks on chromatograms and their corresponding protein bands on SDS PAGE gel were highlighted.

The protein was further purified with a size exclusion chromatography column and analysed with SDS PAGE. Protein expressed in LB media had three peaks on the chromatogram of SEC (Fig. 3.2a). Peak A corresponds to high molecular weight impurities on SDS PAGE. BSP30b was eluted during peak B and peak C. However, only BSP30b from peak C was concentrated for crystallization trials as fractions of peak B also contained other protein impurities as shown on SDS PAGE (Fig. 3.2a), which may prevent the crystallization of BSP30b.

Protein expressed in auto-induction media had two peaks on the chromatogram of SEC. SDS PAGE results suggested that BSP30b was eluted during peak A and peak B (Fig. 3.2b). Although there was more absorbance at 280 nm at peak A compared to peak B, more protein impurities and less BSP30b were present at peak A as

shown on SDS PAGE gel. Only BSP30b eluted at peak B was used for crystallization trials because of its higher purity compared to fractions of peak A.

Both chromatograms of SEC and their corresponding SDS PAGE gels suggested that BSP30b was eluted during 50 – 65 mL elution buffer, within the size range of 17 – 44 kDa, suggesting it is in a monomeric state (29.6 kDa).

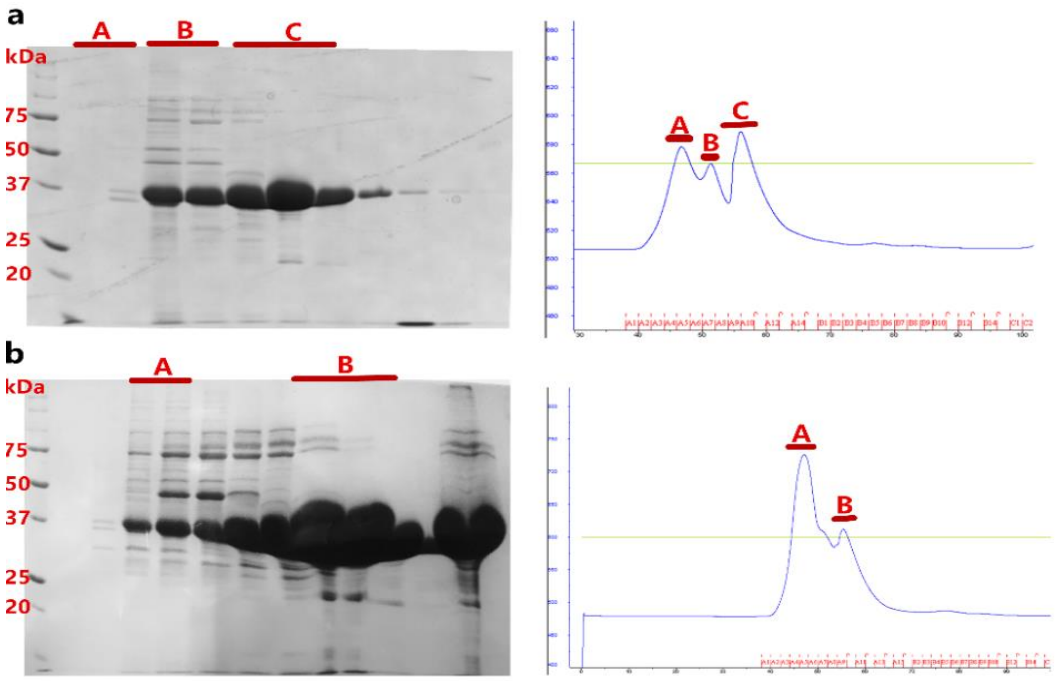

Figure 3.2. SEC of BSP30b and corresponding SDS PAGE analysis. The UV absorbance profile during elution and its corresponding fractions on SDS PAGE when BSP30b was expressed in LB (a) or auto-induction (b) media. Peaks on chromatograms and their corresponding protein bands on SDS PAGE gel were highlighted.

3.3 *BSP30b_F52M_F106M* Was Successfully Cloned into *pProEx Htb*

3.3.1 Determination of the Mutation Sites of BSP30b

The two methionine mutation sites were selected carefully as this was crucial so as not to disrupt the structure and to ensure crystallization. SeMet derivatives could then be produced to solve the phase problem using anomalous diffraction methods.

Since only two protein structures (SPLUNC1 and latherin) of the PLUNC family were available, the mutation sites were determined according to the structural alignment of SPLUNC1 and latherin, combined with the sequence alignment of BSP30b, SPLUNC1, and Latherin (Fig. 3.3 and 3.4).

The mutation sites were selected based on three criteria. The first being that the amino acid is hydrophobic and conserved in latherin and SPLUNC1 but varied in BSP30b. The second being that the side chain of the amino acid is buried in the structure so that the introduced methionine would be ordered and provide an anomalous signal. The third criteria is that the amino acids are separated spatially to provide a phasing signal across the molecule. After inspection of both the structural alignment and the sequence alignment, two phenylalanine residues (residues 52 and 106) for BSP30b were chosen for mutagenesis.

Figure 3.3 Sequence alignment of BSP30b, latherin, and SPLUNC1 using Geneious 7.1.5 (Biomatters Ltd, Auckland). The determined mutation sites were highlighted with red box.

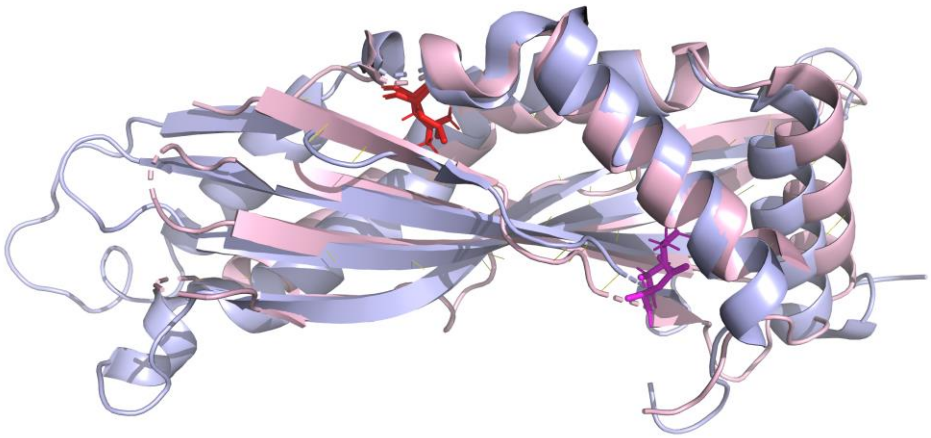

Figure 3.4 Structural alignment of SPLUNC1 and latherin using PyMOL (Delano Scientific). SPLUNC1 and latherin was coloured in light pink and light blue, respectively. The two corresponding mutation sties were shown as sticks, which were coloured in red and magenta, separately.

3.3.2 Cloning of BSP30b_F52M_F106M into *pProEx Htb*

A commercially synthesized gene encoding *BSP30b_F52M_F106M* in the plasmid *pMA-T* was first transformed into *E. coli DH5α* cells for plasmid enrichment. The transformation success was confirmed by colony PCR (Fig. 3.5a).

BSP30b_F52M_F106M was then digested out from plasmid *pMA-T*, and ligated into *pProEx Htb*. This ligated reaction was transformed into *E. coli BL21* for protein expression. The success of transformation was confirmed by colony PCR and the predicted 777 bp product was correctly shown on an agarose gel after electrophoresis of the PCR amplicon (Fig. 3.5b). Analysis of the Sanger Sequencing result of this PCR product with Geneious indicated that the BSP30b mutant gene was successfully inserted into *pProEx Htb* and no unwanted mutations were introduced into this gene during cloning or PCR (Appendix A4).

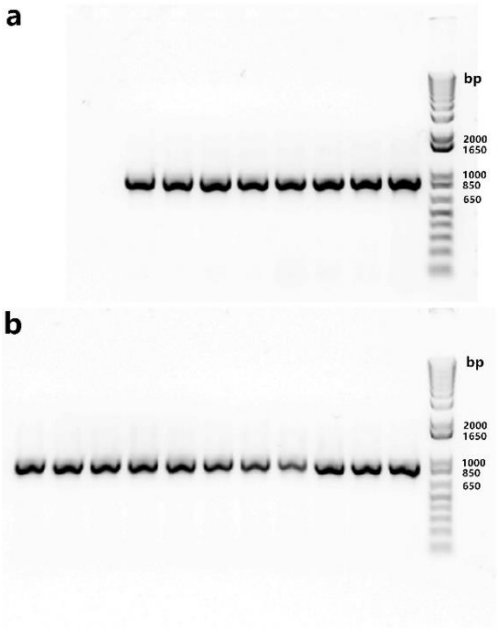

Figure 3.5 Colony PCR results. a, colony PCR using gene specific primers had an amplicon ~ 723 bp. b, colony PCR using primers flanking BSP30b_F52M_106M had an amplicon ~ 777 bp.

3.4 BSP30b_F52M_F106M Expressed in Minimal Media at Low Level

BSP30b_F52M_F106M was first expressed in minimal media without SeMet incorporation to examine its expression level. It was expressed as a 27.8 kDa recombinant protein fused to a N-terminal 6 X histidine tag. Large scale expressed BSP30b_F52M_F106M was first purified using IMAC. The chromatogram and corresponding SDS-PAGE gel results indicated that most of BSP30b_F52M_F106M was eluted between B14 and C5, corresponding to peak B on chromatogram (Fig. 3.6). It was noticed that the BSP30b mutant was expressed at a much lower level compared to native BSP30b, suggested by those light bands on SDS PAGE gel.

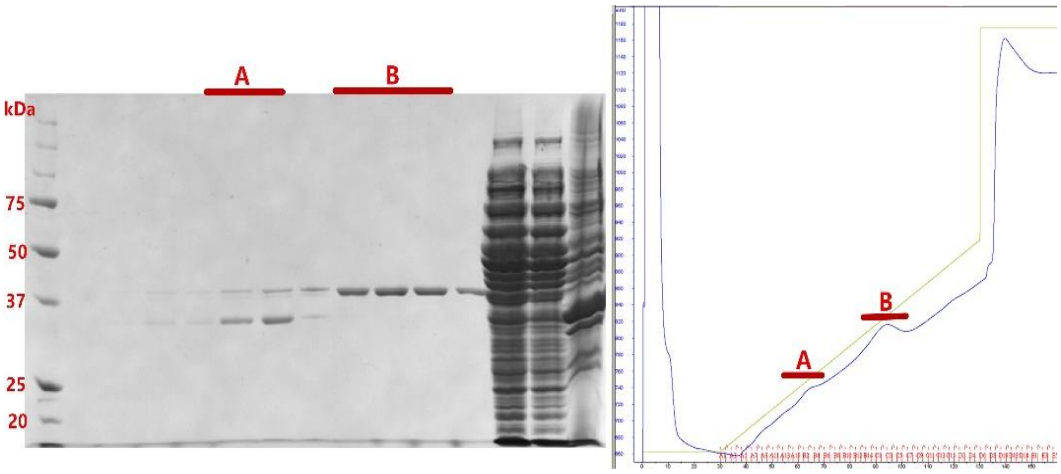

Figure 3.6 IMAC purification and corresponding SDS PAGE analysis of BSP30b_F52M_F106M. The chromatogram depicts the UV absorbance during elution and the corresponding SDS PAGE gel depicts the fractions containing BSP30b_F52M_F106M. Peaks on chromatograms and their corresponding protein bands on SDS PAGE gel were highlighted.

BSP30b_F52M_F106M was further purified with SEC to remove impurities. The chromatogram and corresponding SDS PAGE results suggested that BSP30b_F52M_F106M was eluted off the column as a single peak between fractions A8 and B3 (Fig. 3.7). Those fractions corresponded to 50 – 55 mL of elution buffer, within the size range of 17 – 44 kDa, suggesting it is in a monomeric state (27.8 kDa).

Based on these results, the BSP30b double mutant was further expressed and purified as selenomethionine incorporated protein.

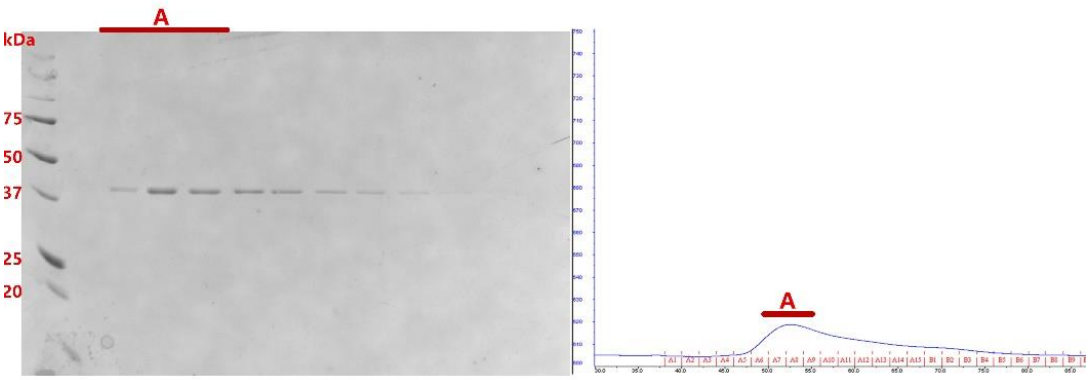

Figure 3.7 SEC purification and corresponding SDS PAGE analysis of BSP30b_F52M_F106M. The chromatogram depicts the UV absorbance during elution and the corresponding SDS PAGE gel depicts the fractions containing BSP30b_F52M_F106M. Peaks on chromatograms and their corresponding protein bands on SDS PAGE gel were highlighted.

3.5 Expression and Purification of SeMet Incorporated BSP30b_F52M_F106M

The BSP30b double mutant was further expressed as selenomethionine (SeMet) incorporated protein in order to solve the phase problem. The expressed protein was first purified with IMAC. There was only a single peak on the chromatogram but this peak represented a mixture of different proteins including SeMet incorporated BSP30b mutant as shown on SDS PAGE gel (Fig 3.8a). This protein mixture was mainly composed of two different proteins, which have similar molecular weight. Both proteins were expressed at a relatively low levels and it was hard to determine which band represented the BSP30b double mutant SeMet derivative. Moreover, separation of these two proteins by SEC was impractical because of their similar size.

In order to solve this problem, an anion exchange chromatography step was inserted before SEC in order to separate these two proteins based on their charge. BSP30b has a relatively low pI of 4.4, which is unusual for many proteins. The

chromatogram and corresponding SDS PAGE results showed that these two proteins were well separated after anion exchange chromatography (Fig. 3.8b). The unknown protein was eluted first between fractions B12 and C2, shown as peak A on chromatogram and section A on SDS PAGE gel. While BSP30b double mutant SeMet derivative was eluted later, between fractions C14 and D6.

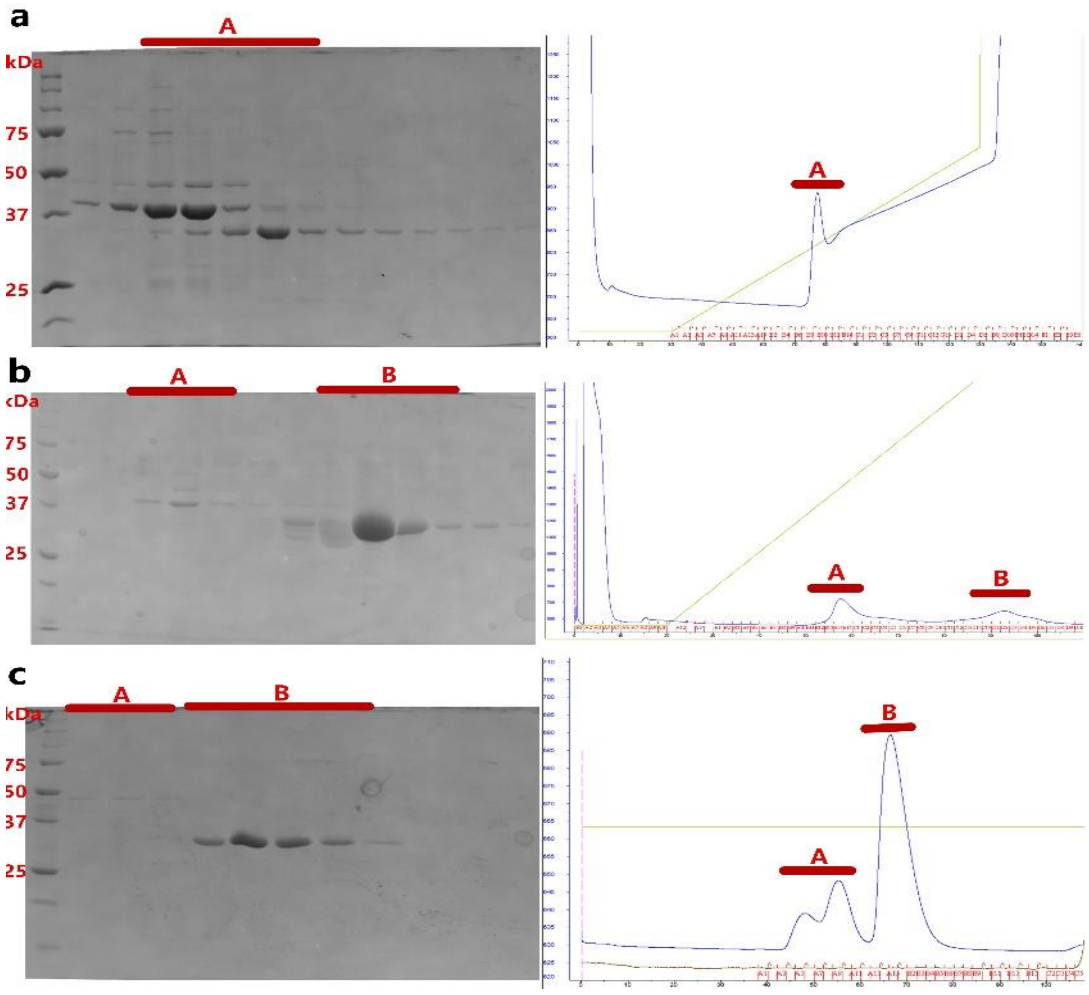

Figure 3.8 Purification of BSP30b_F52M_F106M SeMet derivative. a, chromatogram and corresponding SDS PAGE result of IMAC. b, chromatogram and corresponding SDS PAGE result of anion exchange chromatography. c, chromatogram and corresponding SDS PAGE result of SEC. Peaks on chromatograms and their corresponding protein bands on SDS PAGE gel were highlighted.

Size Exclusion Chromatography was still used for further purification despite the relatively clean band on the SDS PAGE gel. One reason was to separate the BSP30b

mutant from other protein contaminants at low concentration; another reason was to remove the excess salt from anion exchange chromatography, which may inhibit the crystallization of SeMet incorporated BSP30b mutant. The chromatogram and corresponding SDS PAGE results suggested that BSP30b double mutant SeMet derivative was eluted between fractions A13 and B5 (peak B in figure 3.8c). Protein from these fractions was concentrated and used for crystallization trials later.

3.6 Generation of MBP fused BSP30b_F52M_F106M

MBP fused BSP30b mutant was constructed due to the fact that the structure of SPLUNC1 (also a PLUNC protein) was solved with a MBP tag (Ning et al., 2014), which was supposed to improve the stability of SPLUNC1 and also help to solve the phase problem using MR (molecular replacement).

The *BSP30b_F52M_F106M* gene was amplified by PCR from the *pMA-T_BSP30b mutant* plasmid and cloned into the cytoplasmic expression vector *pMAL-c2X* using the *BamHI/PstI* restriction sites, positioning an MBP tag at the N-terminus of the protein. The insertion of *BSP30b double mutant* was confirmed by colony PCR amplification, resulting in an 890 bp band as expected on the agarose gel after electrophoresis (Fig. 3.9).

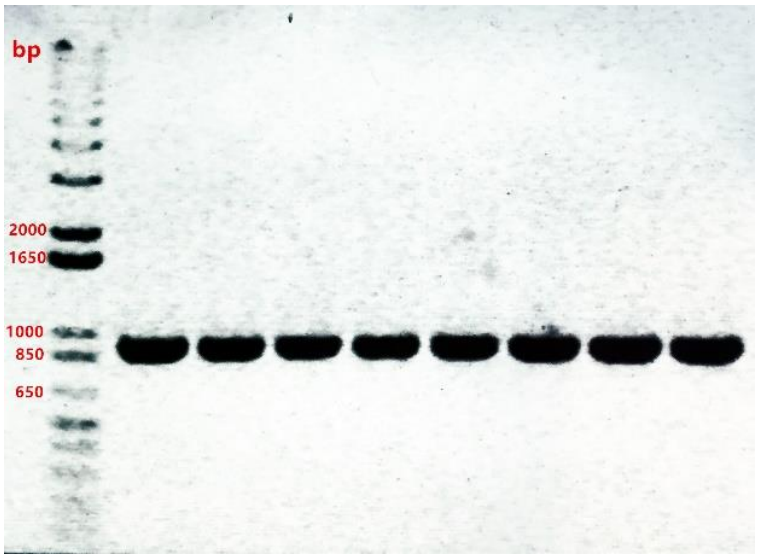

Figure 3.9 Colony PCR results. PCR using primers flanking BSP30b_F52M_F106M, resulting in an ~ 890 bp band on agarose gel after electrophoresis.

3.7 Expression of BSP30b Mutant with MBP Tag Improves Protein Yield

MBP fused BSP30b mutant was expressed in LB media as a 75 kDa protein. The protein was first purified with amylose resin chromatography. The chromatogram and corresponding SDS PAGE gel results suggested that the protein was tightly bound to the column and was well eluted between fractions A7 and A12 (Fig. 3.10a). Noticeably, compared to the expression level of BSP30b double mutant in *E. coli BL21* cells, there was a large improvement on the yield of the expressed protein, indicated by the significant UV absorbance on chromatogram and thick bands on SDS PAGE gel. It is noted that a few protein impurities were also eluted at the same time as MBP_BSP30b_F52M_F106M.

The protein was further purified with SEC. The chromatogram and corresponding SDS PAGE results suggested that the fused protein was eluted off the column with

two peaks (peak B and C, Fig. 3.10b). Protein eluted during peak B had a molecular weight similar to the model protein IgG (molecular weight of 158 kDa), indicating the eluted MBP_BSP30b_F52M_F106M was a dimer. Protein eluted during peak C had a similar elution profile to the model protein albumin, which has a molecular weight of 66 kDa, suggesting the protein is a monomer from this peak. The dimer was about two times the quantity of the monomer, indicated from their UV_{280} absorbance and SDS PAGE gel results. Both monomeric and dimeric proteins were concentrated separately for crystallization trials.

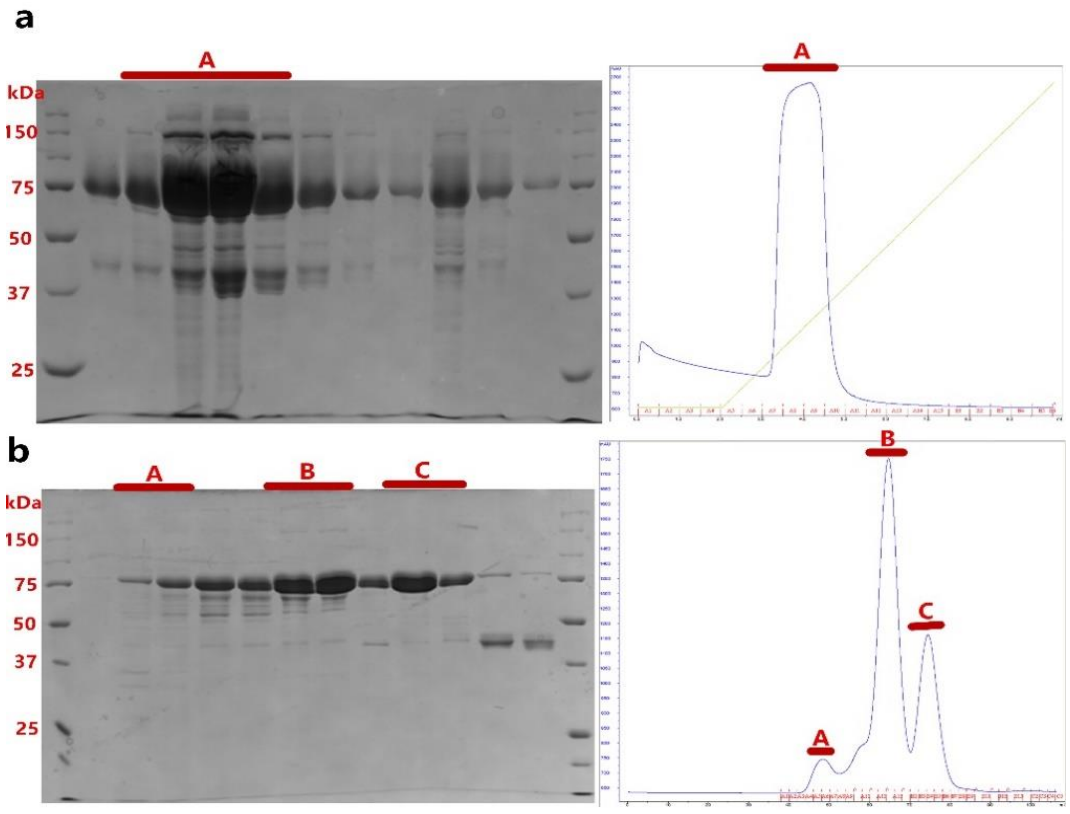

Figure 3.10 Purification of MBP_BSP30b mutant. a, chromatogram and corresponding SDS PAGE result of IMAC. b, chromatogram and corresponding SDS PAGE result of SEC. Peaks on chromatograms and their corresponding protein bands on SDS PAGE gel were highlighted.

3.8 Discussion

Native BSP30b was successfully expressed at high level in *E. coli BL21* cells using both LB and auto-induction media. Unexpectedly, the expression level was dramatically increased in auto-induction media compared to its expression in LB media. Proteins expressed in both media were successfully purified using IMAC and SEC. Analysis of the SEC chromatogram suggested that BSP30b was in a monomeric state.

Although BSP30b was well expressed in both media indicated by their corresponding SDS PAGE gel results, only small peaks were shown on its chromatograms during purifications. This is because BSP30b has no tryptophan amino acids, and hence it has poor UV light absorbance at 280 nm. It was noticed that there was only one single peak with decent UV absorbance in the chromatogram of IMAC of auto-induction expressed BSP30b; however, a few non-BSP30b proteins were also purified together with BSP30b, which may contribute to most of the light absorbance at 280 nm. This could be true as, after SEC, the light absorbance of those non-BSP30b proteins is much higher compared to that of BSP30b, although they were at much lower quantities.

As BSP30b doesn't contain any methionine, two phenylalanine to methionine mutations (F52M, F106M) were introduced into BSP30b for phasing. The synthesized gene coding for BSP30b_F52M_F106M was successfully inserted into protein expression vector *pProEx Htb*, downstream of the trc promoter.

BSP30b_F52M_F106M was first expressed in *E. coli* using M9 minimal media. However, it was expressed at much lower level compared to native BSP30b. Protein was first purified with IMAC. Although there were two bands of similar size present

on the SDS PAGE gel, they did not trigger special attention as those two proteins were eluted separately during IMAC and native BSP30b was well eluted as a single band on SDS PAGE gel when using SEC. However, when BSP30b_F52M_F106M was expressed as SeMet incorporated protein in *E. coli* using M9 minimal media, these two bands were overlapped on SDS PAGE gel after IMAC. Because of the overlapping of these two proteins and their similar size, further purification with SEC was impractical. Another question from this situation was which band on SDS PAGE represents the SeMet incorporated protein. Based on the fact that BSP30b has a relatively low pI of 4.4, an anion exchange chromatography step was inserted before SEC for protein separation. These two proteins were well separated after anion exchange chromatography. The unknown protein was eluted out first and BSP30b_F52M_F106M SeMet derivative was salted out later because of its tight binding to the column. The SeMet incorporated BSP30b mutant was further purified with SEC to remove extra salt, which may inhibit the crystallization of the protein.

The gene coding BSP30b_F52M_F106M was also cloned into *pMAL-c2X*, which will leave an MBP tag at the N terminus of BSP30b mutant. There have been a growing number of proteins which have been crystalized as MBP fusion proteins. Indeed, about 100 MBP fusion protein structures has been deposited in the Protein Data Bank (PDB) since 2003 (Waugh, 2016). The MBP tag was believed to improve the expression, solubility, stability, or crystallization of the target protein. Moreover, the initial phases can be simply obtained by molecular replacement using MBP as the reference model. Another advantage for this study is that the quantity of the expressed MBP fusion protein can be easily examined using Nanodrop instead of Bradford Assay.

In this study, the expression level of BSP30b_F52M_F106M was increased dramatically after fused to MBP. The MBP fusion protein was successfully eluted using MBP Trap column at first, and then purified by SEC. The SEC purified MBP_BSP30b_F52M_F106M has both monomeric and dimeric states. Both monomer and dimer of MBP-BSP30b_F52M_F106M were concentrated separately for crystallization trails later.

Chapter Four. The Three-dimensional Structure of BSP30b

4.1 Introduction

One of the central objectives of this PhD research was to determine the three dimensional structure of BSP30b as a means of narrowing down its possible biological function. In order to solve the structure of BSP30b, the purified BSP30b, BSP30b_F52M_F106M SeMet derivative, and MBP-BSP30b_F52M_F106M were all screened using crystallization trials and the resulting crystals were then subjected to X-ray diffraction experiments.

Previously, an X-ray diffraction dataset of native BSP30b was collected at 2.5 Å from a colleague of the Arcus group. In the present study, in-depth optimization of the previous crystallization conditions were performed to yield crystals with improved diffraction so the structure of BSP30b might be solved by sulfur-SAD or direct methods. To solve the phase problem, the traditional MIR (multiple isomorphous replacement) approach was also tested by soaking BSP30b crystals in heavy atom solutions.

Crystallization trials were also conducted for SeMet incorporated BSP30b_F52M_F106M in order to solve the structure of BSP30b using the anomalous signal of selenium.

The MBP-BSP30b_F52M_F106M fusion protein was also set up for crystallization trials so that the structure of BSP30b could be solved by molecular replacement, using MBP as the reference model.

4.2 Crystallization of BSP30b

4.2.1 Robotic Crystallization Screens of BSP30b

Size exclusion chromatography purified BSP30b was subjected to robotic crystallisation screens (96 x 4 different conditions) at concentration of 10 mg/ml and needle shaped crystals were observed in seven conditions (Appendix A5). Fine screens based on these condidtions were then trialed using the hanging drop method, however this resulted in limited crystal development.

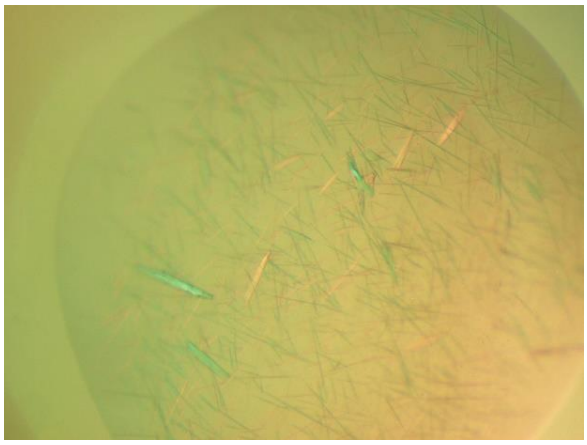

Figure 4.1 Representation of needle shaped crystals from robotic screen.

4.2.2 Fine Screen of BSP30b

Previously determined crystallization conditions from the Arcus lab were also set up for fine screens with BSP30b at a concentration of 10 mg/ml. After repeating the fine screen a few times, condition "B3" (0.1 M HEPES pH 7.5, 0.2 M $CaCl_2$, 9% w/v PEG 3350, 5% v/v IPA) repeatedly produced needle-shaped or plate-shaped crystal "hits" (Fig. 4.2). Crystals were taken to the Australian Synchrotron for X-ray diffraction. Those needle-shaped crystals generally had poor diffraction at ~8 Å resolution. The plate-shaped crystals had relatively improved diffraction, however

only diffracted to ~ 6 Å resolution. Extensive optimisation of this crystalization condition were carried out to obtain crystals with better diffraction quality.

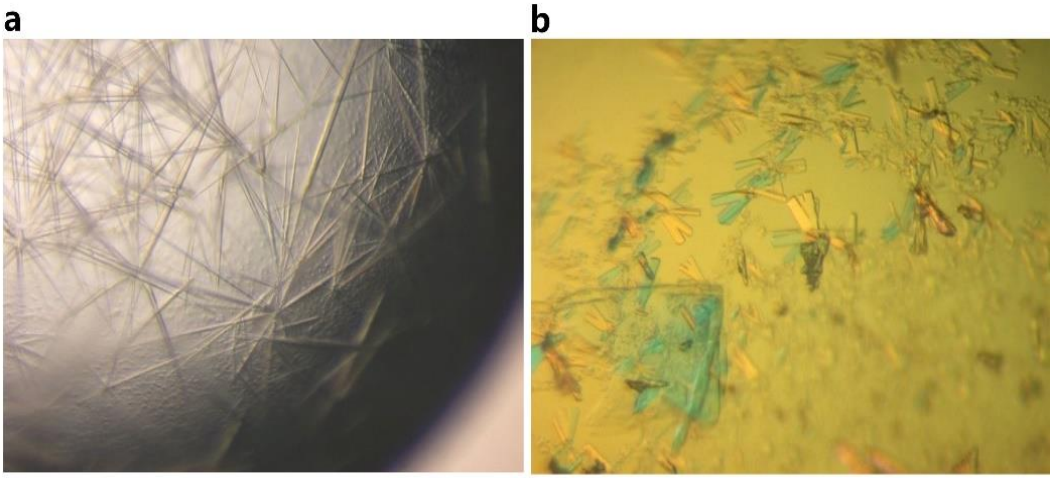

Figure 4.2 Crystals of BSP30b generated from condition "B3". a, needle shaped crystals grown in condition "B3" with poor diffraction. b, thin plate shaped crystals grown in condition "B3" with diffraction ~ 6 Å.

4.2.3 Optimisation of Condition "B3"

4.2.3.1 Additive Screen

The Hampton additive screen was set up with BSP30b at 10 mg/ml using condition "B3" as its basis. Among the 96 conditions, 55 conditions had crystals grown after 19 days, and 20 of them resulted in thick-rod or thick-plate shaped crystals. Fine screens of 15 conditions were performed and protein crystals were grown in six conditions after ~ 2 weeks (Fig. 4.3). Thick rods were grown in fine screens with the "additives" glucose monohydrate or galactose. Single small rods were grown in fine screens with the "additives" glycine or sodium iodide. Both fine and robotic additive screens with manganese chloride tetrahydrate had irregular shaped crystals. There wasn't much improvement on crystal quality with the additive ethanol.

Crystals from those conditions were tested on our home source X-ray beam and at the Australian Synchrotron: the thick rod-shaped crystals had relatively better diffraction to ~ 5 Å resolution.

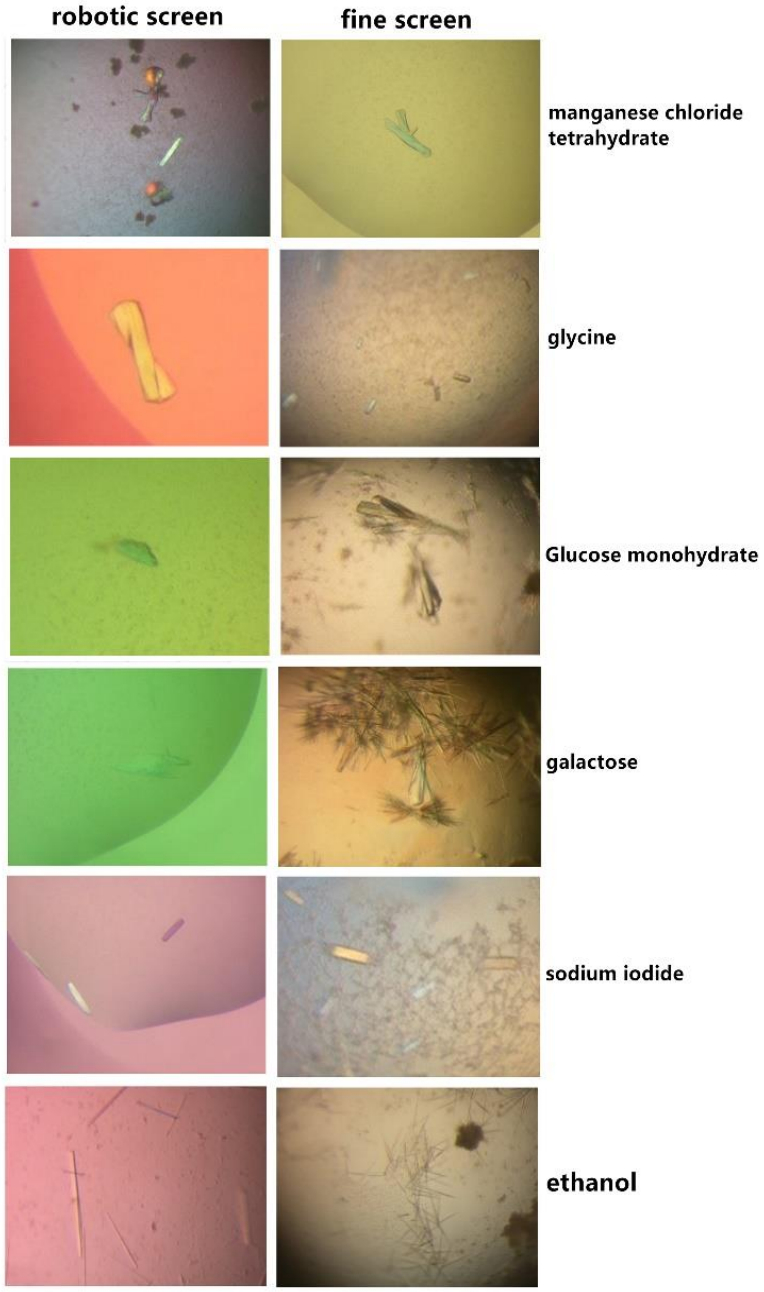

Figure 4.3 Crystals grown in 6 conditions of additive screen in both robotic and fine screen. Thick rods were grown in fine screens with "additives" glucose monohydrate or galactose. Single small rods were grown in fine screens with "additives" glycine or sodium iodide. Irregular shaped crystals were grown with additive manganese chloride tetrahydrate. Only needle shaped crystals were grown in condition with ethanol.

4.2.3.2 Trypsin digestion of BSP30b

Pre-incubation of BSP30b at 10 mg/ml with trypsin was also performed for crystallization. The pre-incubation time was optimised by running an SDS PAGE gel with mixtures of trypsin and BSP30b at a ratio of 1:100 at different time intervals (Fig. 4.4). Nearly all BSP30b was digested into a single band (~32 kDa) after pre-incubation of 1 minute. After incubation of 1 hour, most of the BSP30b was digested into a smaller bands (~27 kDa).

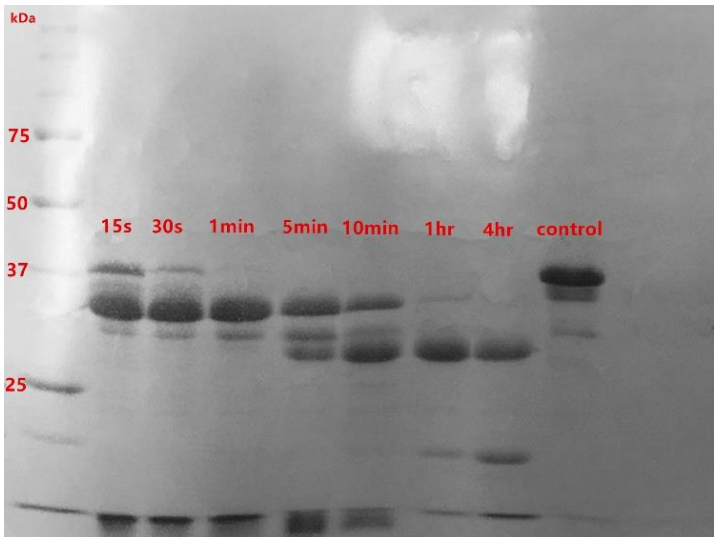

Figure 4.4 Trypsin digestion assay of BSP30b. BSP30b was digested during different time intervals and then electrophoresed on SDS PAGE gel. Different band patterns were seen on the gel. BSP30b without digestion was used as control.

Fine screens of BSP30b pre-incubated with trypsin for 1 min was performed and tiny crystals appeared in the drop after a week (Fig. 4.5). Those crystals were taken to Australian Synchrotron for X-ray diffraction and very poor diffraction was observed.

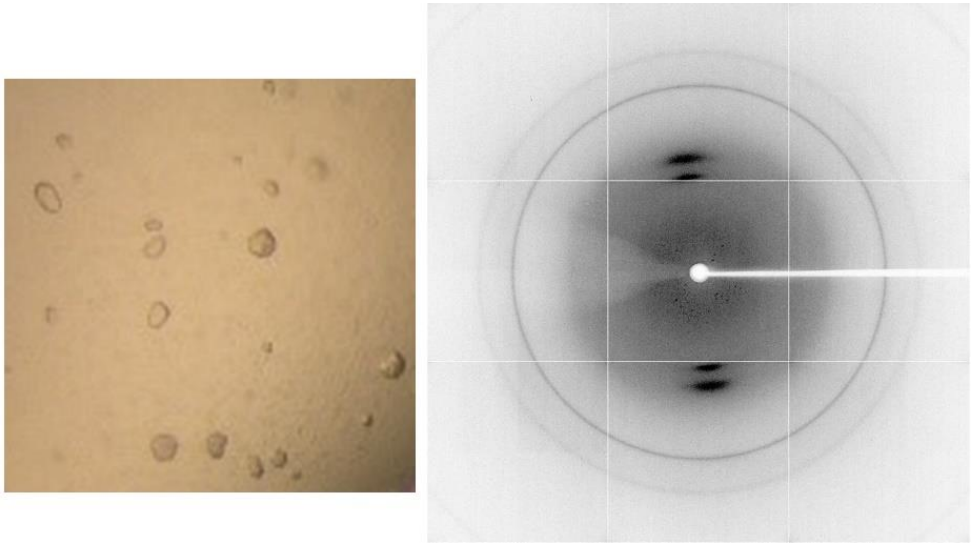

Figure 4.5 Crystals of trypsin digested protein having a poor diffraction. Tiny crystals with uneven surface were grown after trypsin digestion. Only a few spots appeared on the detector after shooting with X-ray, suggesting its poor diffraction.

4.2.3.3 Silver bullet screen of BSP30b

The Hampton silver bullet screen was also trialled with BSP30b at 10 mg/ml using condition "B3". 13 conditions had better needle-shaped crystals formed after ~2 weeks (Fig. 4.6). However, the crystals from these fine screens had no or very poor diffraction at the Australian Synchrotron.

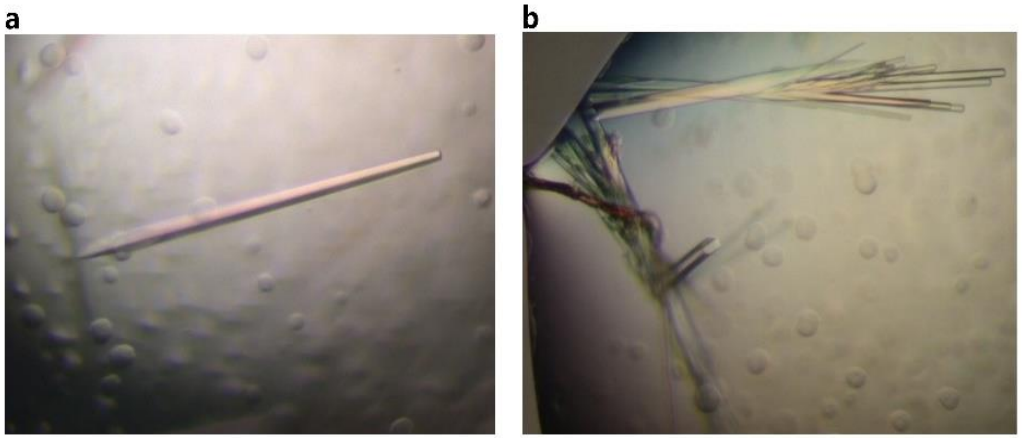

Figure 4.6 Silver bullet screen. Needle (a) or thin-rod (b) shaped crystals were formed during silver bullet screen. However, those crystals had no or very poor X-ray diffraction at Australian Synchrotron.

4.2.3.4 Seeding of BSP30b

4.2.3.4.1 Crystals Grown in Streak Seeding Dissolved Again

Streaking seeding yielded some chunky and thick rod-shaped crystals along the streak as well as some thin needles within the drop after 2 days. Unfortunately, those crystals dissolved after 2 weeks (Fig. 4.7).

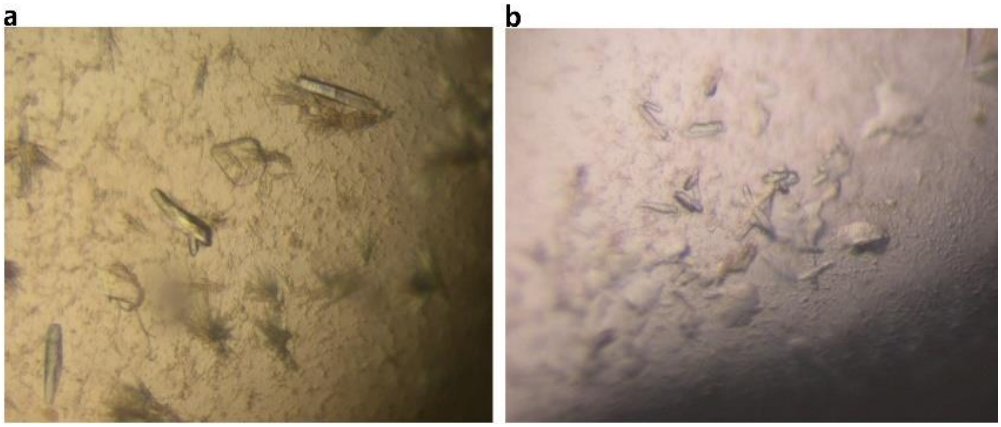

Figure 4.7 Streak seeding. Chunky and fatty rod-shaped crystals (a) were produced with streak seeding. However, those crystals were dissolved later (b) before shooting by X-ray.

4.2.3.4.2 Batch Seeding Yielded Crystals with Improved Diffraction

Crystallization using batch seeding resulted in more success. Most of the batch seeding drops had crystal "hits" in three different morphologies after 5 days (Fig. 4.8). All of the drops had needle-like and plate-shaped crystals with medium thickness. Several drops also had crystals with a third boulder-like morphology. Those plate-shaped and boulder-shaped crystals were taken to the Australian Synchrotron for diffraction. Those boulder-shaped crystals had poor diffraction ~ 5 Å resolution. Those plate-shaped crystals had relatively better diffraction ~ 3 Å resolution and were later used for heavy metal soaks to try to solve the phase problem.

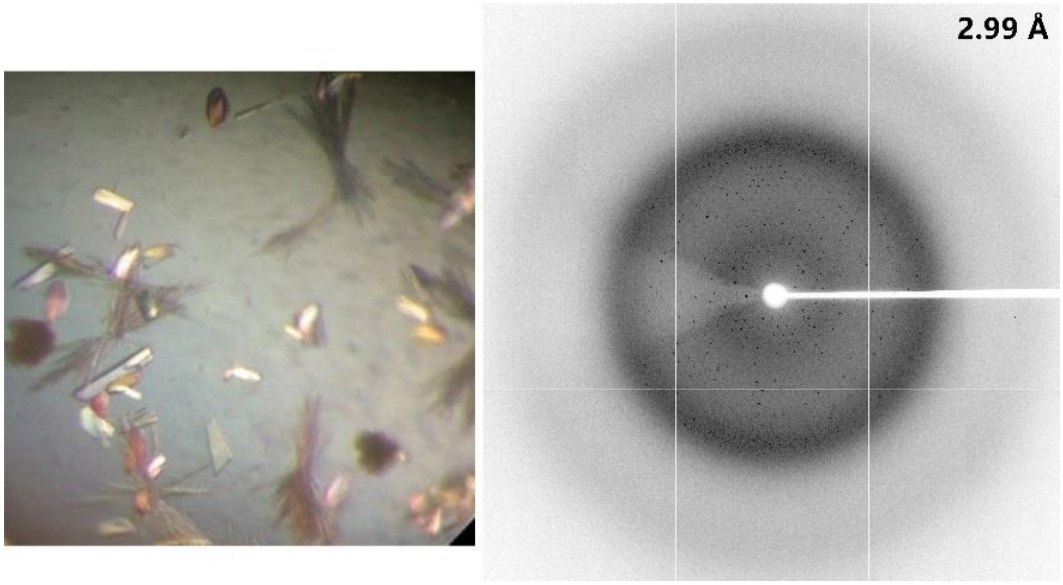

Figure 4.8 Batch seeding of BSP30b. Different morphological crystals were grown in batch seeding. The boulder shaped ones had diffractions ~ 5 Å. Plate shaped crystals had diffractions ~ 3 Å.

4.2.3.5 Gel Shift Assay and Heavy Metal Soak of BSP30b

Native gel shift assays are a simple method for searching for heavy metal derivatives. Eight heavy metal compounds were chosen to mix with BSP30b separately followed by running native PAGE. Obvious band shifts were observed when BSP30b was mixed with $KAu(CN)_2$, K_2PtCl_4, and K_2PtCl_6, respectively. For heavy metal soaking procedure, $KAu(CN)_2$ and K_2PtCl_4 were selected.

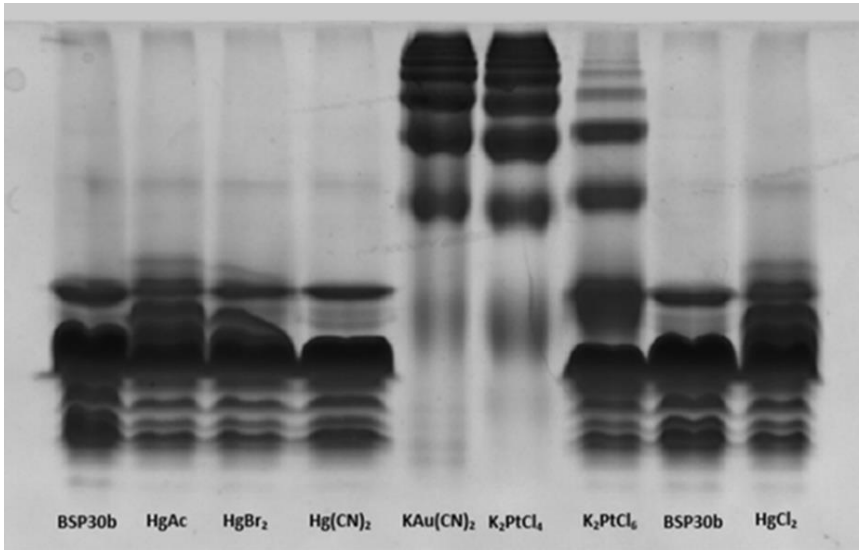

Figure 4.9 Native gel shift assay results. All of the heavy atom compounds used at a final concentration of 2mM. BSP30b itself without heavy atom compound was used as control. Band shifting was seen with heavy atom compounds $KAu(CN)_2$, K_2PtCl_4, and K_2PtCl_6, respectively.

$KAu(CN)_2$ soaked crystals had a pink/red colour, indicating the potential binding of heavy atoms to the crystals (Fig. 4.10b). Those crystals diffracted to ~ 3 Å resolution at the Australian Synchrotron; however, the anomalous signal was too weak after scaling in CCP4. This could because there was only transient binding between the gold ion and BSP30b, and hence no obvious anomalous signal was observed.

After soaking in K_2PtCl_4, there was no colour change for the crystals (Fig. 4.10a). Crystals were taken to Australian Synchrotron and most of them diffracted to ~ 5 Å resolution. After integration and scaling in CCP4, no anomalous signal was detected.

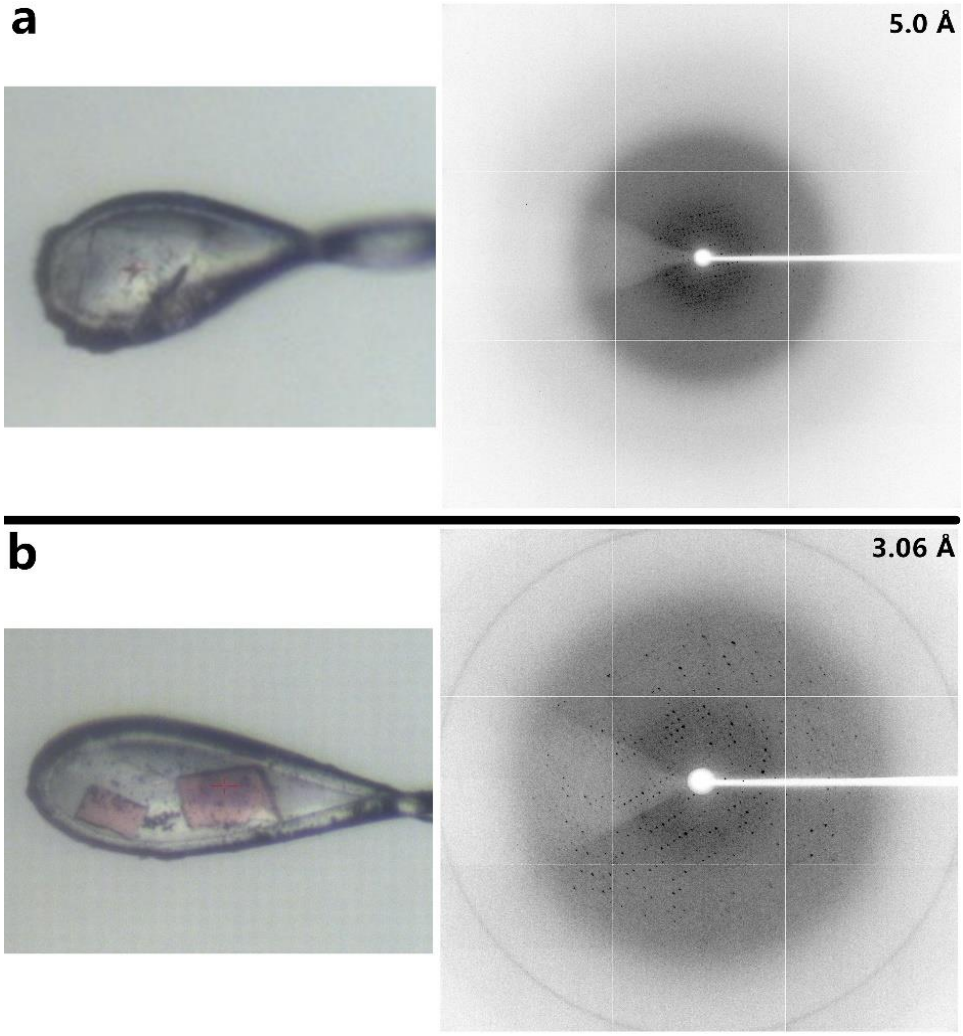

Figure 4.10 Heavy metal soaked crystals and their diffraction. a, K_2PtCl_4 soaked BSP30b crystal with diffraction of 5 Å. b, crystal soaked in $KAu(CN)_2$ had a pink/red colour and diffracted to 3.06 Å.

4.3 Crystallization of BSP30b_F52M_F106M SeMet Derivative

4.3.1 Robotic Screen

Because of the low expression of SeMet-incorporated BSP30b double mutant, only crystal screens and index screens were set up for crystallization at 8 mg/ml. After 10 days, tiny crystals appeared in two conditions. Hanging drop fine screens of these conditions were prepared but resulted in limited crystal development.

4.3.2 Fine screen

BSP30b double mutant SeMet derivative at 10 mg/ml was set up for fine screens with the old 24 conditions (Table 4.1). The conditions comprise a combination of HEPES (pH 7.5), $CaCl_2$, PEG 3350, and isoproponal at various concentrations. Those conditons have been proven to produce crystals using native BSP30b. Here, they were tested for crystallization of BSP30b double mutant SeMet derivative.

Rod/cluster-shaped crystals appeared in conditions B3 – B6, C3 – C6 and D4 – D6 after 2 weeks (Fig. 4.11). Those crystals were broken into single crystals, which were then frozen and taken to the Australian Synchrotron for X-ray diffraction. Most of the crystals had good diffraction ranging from 1.9 Å to 2.5 Å resolution.

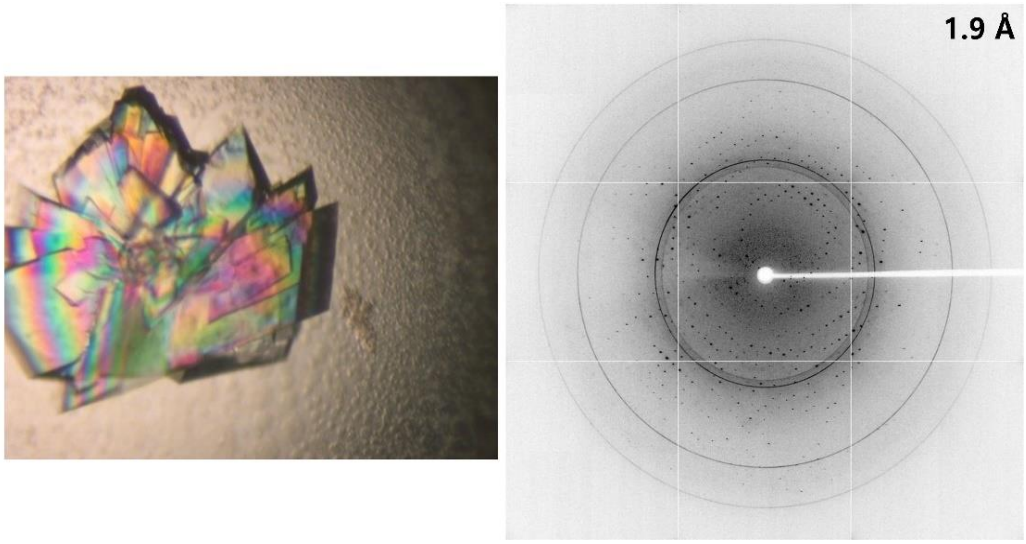

Figure 4.11 Rod-cluster shaped crystals of BSP30b mutant SeMet derivative with diffraction of 1.9 Å. Rod-cluster shaped crystals was grown after 2 weeks. Crystals were broken into single ones with the best diffraction of 1.9 Å.

Table 4.1 The previously determined 24 conditons

	Final conc.		Final conc.		Final conc.		Final Conc.		Final conc.		Final conc.
A1		**A2**		**A3**		**A4**		**A5**		**A6**	
HEPES	0.1	HEPES	0.1	HEPES	0.1	HEPES	0.1	HEPES	0.1	HEPES	0.1
CaCl2	0.2	CaCl2	0.2	CaCl2	0.2	CaCl2	0.2	CaCL2	0.2	CaCL2	0.2
PEG 3350	5	PEG 3350	7	PEG 3350	9	PEG 3350	11	PEG 3350	13	PEG 3350	15
IPA	0	IPA	0	IPA	0	IPA	0	IPA	0	IPA	0
MQ H2O		MQ H2O		MQ H2O		MQ H2O		MQ H2O		MQ H2O	
B1		**B2**		**B3**		**B4**		**B5**		**B6**	
HEPES	0.1	HEPES	0.1	HEPES	0.1	HEPES	0.1	HEPES	0.1	HEPES	0.1
CaCl2	0.2	CaCl2	0.2	CaCl2	0.2	CaCl2	0.2	CaCL2	0.2	CaCL2	0.2
PEG 3350	5	PEG 3350	7	PEG 3350	9	PEG 3350	11	PEG 3350	13	PEG 3350	15
IPA	5	IPA	5	IPA	5	IPA	5	IPA	5	IPA	5
MQ H2O		MQ H2O		MQ H2O		MQ H2O		MQ H2O		MQ H2O	
C1		**C2**		**C3**		**C4**		**C5**		**C6**	
HEPES	0.1	HEPES	0.1	HEPES	0.1	HEPES	0.1	HEPES	0.1	HEPES	0.1
CaCl2	0.2	CaCl2	0.2	CaCl2	0.2	CaCl2	0.2	CaCL2	0.2	CaCL2	0.2
PEG 3350	5	PEG 3350	7	PEG 3350	9	PEG 3350	11	PEG 3350	13	PEG 3350	15
IPA	10	IPA	10	IPA	10	IPA	10	IPA	10	IPA	10
MQ H2O		MQ H2O		MQ H2O		MQ H2O		MQ H2O		MQ H2O	
D1		**D2**		**D3**		**D4**		**D5**		**D6**	
HEPES	0.1	HEPES	0.1	HEPES	0.1	HEPES	0.1	HEPES	0.1	HEPES	0.1
CaCl2	0.2	CaCl2	0.2	CaCl2	0.2	CaCl2	0.2	CaCL2	0.2	CaCL2	0.2
PEG 3350	5	PEG 3350	7	PEG 3350	9	PEG 3350	11	PEG 3350	13	PEG 3350	15
IPA	15	IPA	15	IPA	15	IPA	15	IPA	15	IPA	15
MQ H2O		MQ H2O		MQ H2O		MQ H2O		MQ H2O		MQ H2O	

4.3.3 Initial Data Processing for BSP30b Double Mutant SeMet Derivative

Data was collected to 1.9 Å with 98.9% completeness at a wavelength of 0.9537 Å. The dataset was indexed as the C2 space group and integrated using iMosflm (Leslie & Powell, 2007). Matthews_Coeff analysis indicated that the asymmetric unit contains two BSP30b molecules and the solvent content of the unit cell is 42.7% (Table 4.2). The data were scaled with a resolution limit of 2 Å to reach appropriate R_{merge} and CC (1/2) values (see table 4.5). Significant anomalous signal was detected after scaling.

Table 4.2 Matthew Coeff results of BSP30b mutant

Nmol/asym	Matthews Coeff	%solvent	P(2.00)	P(tot)
1	4.29	71.35	0.01	0.03
2	2.15	42.70	0.99	0.97
3	1.43	14.05	0.00	0.00

4.3.4 Structure Solution and Refinement

Phasing was achieved using the single-wavelength anomalous diffraction (SAD) technique to solve the structure of BSP30b. Six initial selenium sites were identified by running *AutoSol* in Phenix (Table 4.3). Phases from these six sites gave a low figure of merit (FOM) of 0.273. Nonetheless, using these phases as a starting point, 260 residues were automatically built resulting in an R_{free} of 0.38 and R_{work} of 0.35.

Table 4.3 AutoSol results. a, results from heavy atom search and phasing. b, results from model building

a

Number	Space group	# of refined sites	Figure of merit	Overall score
3	C 1 2 1	6	0.273	30.86 +/- 13.78
1	C 1 2 1	6	0.289	30.85 +/- 13.77
4	C 1 2 1	6	0.273	15.56 +/- 13.83

b

Solution number	# of residues	Map-model CC	R(work)	R(free)
1	260	0.59	0.3520	0.3791
3	312	0.57	0.3970	0.4365

The structure of BSP30b was further built with *AutoBuild* in Phenix. The protein sequence data and four files created by *AutoSol* (containing experimental data, initial map, starting model and heavy-atom sites, respectively) were used for this purpose. The resulting model from *AutoBuild* gave 337 protein residues and 251 water molecules (Table 4.4).

Table 4.4 AutoBuild results

Final model

R-work:	0.2838	**R-free:**	0.3134	**Residues:**	337
Fragments:	9	**Waters:**	251		

The structure of BSP30b from AutoBuild was then refined with *phenix.refine* in Phenix. Subsequent cycles of manual model building using COOT 0.7 and refinement using *phenix.refine* were performed until no improvement in R_{free} values could be achieved (Table 4.5).

Table 4.5 Data collection and refinement statistics

Property	Value
Wavelength	0.9537 Å
Space group	C121
Cell constants a, b, c, α, β, γ	81.65Å 59.57Å 89.94Å 90.00° 106.25° 90.00°
Resolution (Å)	47.43 – 2.00 (2.11 – 2.00)
R_{merge}	0.100 (0.622)
CC (1/2)	0.998 (0.922)
Completeness (%)	99.9 (100)
Redundancy	7.2 (7.3)
No. of observations	201779 (29708)
No. of unique reflections	28136 (4080)
Mean I/σ(I)	11.8 (3.5)
Anomalous completeness (%)	98.5 (98.1)
Anomalous multiplicity	3.7(3.8)
DelAnom correlation between half sets	0.130 (0.020)
R_{work}	0.2051
R_{free}	0.2514
Total number of atoms	3064
Average B, all atoms ($Å^2$)	36.0
Ramachandran favoured (outliers)	97.05% (0.00%)
Rotamer outliers	0.00%
Clash score	4.93

4.3.5 Structural analysis

Crystals of a double Phe-to-Met mutant of BSP30b diffracted to 2.0 Å resolution with a space group of C2 and unit cell dimensions and angles: a = 81.65Å, b = 59.57Å, c = 89.94Å, α = γ = 90°, β = 106.25°. The structure was determined using SeMet incorporated protein and Single-wavelength anomalous dispersion (SAD) methods.

BSP30b crystallized as two molecules in ASU (asymmetry unit) shown in figure 4.12a. Each molecule of BSP30b consists of 7 α helices and 6 β sheets as well as connecting turns and loops which envelope a hydrophobic channel conserved within the TULIP superfamily (Fig. 4.12b). It is clear that the first 39 amino acids of the recombinated BSP30b are disordered. Of which, 29 amino acids come from the vector while the remaining 10 amino acids come from BSP30b. The structure of BSP30b begins with a helix (α1, residues 40-48), followed by a short helix (α2, residues 50-53). Two antiparallel α helices (α3, residues 55-69; α4 residues 81-100) follow and noticeably, no continuous density was found to connect these two helices in either the initial electron density map or the final refined model. Hence, residues 74-79 are missing from the model. This is followed by antiparallel β strands from β1 to β6 (β1 residues 103-106; β2 residues 114-117; β3 residues 122-137; β4 140-158; β5 167-169; β6 179-185) and these form a long highly twisted β-strand. There is a short α helix (α5 residues 176-178) between β5 and β6. A long bent α helix (α6 residues 193-224) follows β6 and the structure ends with a unique C-terminal helix (α7 residues 227-235).

The hydrophobic channel of BSP30b (well-conserved across the TULIP superfamily) is enclosed by helices α3, α4, α6, α7, and sheets β1 – β6. The

calculated internal cavity using PyMOL (DeLano Scientific) suggests that its potential opening is at the C-terminal end along the long axis of the structure (Fig 4.12d). The length of the internal cavity extends ~ 22.6 Å. We expect that the flexible loop between helices α3 and α4 at the opening region will move away upon ligand binding. BSP30b also has a conserved disulfide bridge (of the PLUNC family) formed between cysteine 172 and 225 and joining the long-curved helix α6 and the loop region between helix α5 and sheet β5. This disulfide bond has been proposed to maintain the stability of the entire structure of SPLUNC1 (Ning et al., 2014). Since this disulphide bond of BSP30b is located at the same postion as in SPLUNC1, we believe that this disulphide bond also stabilises the whole structure of BSP30b.. Moreover, two metal ion sites are predicted from the electron density map, and both have regular octahedral geometry. The most likely metal ion is calcium ion from the crystallization “mother liquor”, however it is not a conclusive finding.

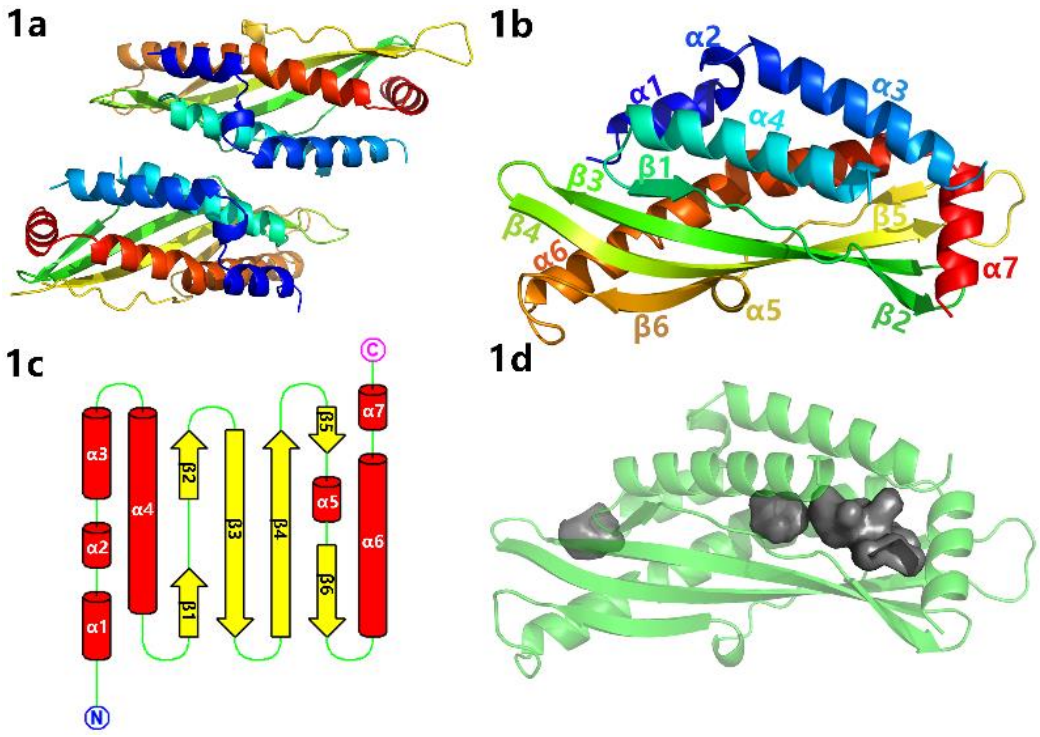

Figure 4.12 Overall structure of BSP30b. 1a, asymmetric unit of the BSP30b crystal, composed of two copies of the BSP30b molecule. 1b, the tertiary structure of apo-BSP30b, coloured in rainbow mode ranging from blue (N terminus) to red (C terminus). 1c, topology diagram of BSP30b. All seven α helices are represented by red cylinders; six β strands by yellow arrows; green lines represent the loops and turns within the structure. 1d, representation of the internal cavity, calculated using PyMOL (Delano Scientific).

4.3.6 Comparisons between BSP30b and SPLUNC1 Structures

The BSP30b structure was compared with structures in PDB using the Dali server. From the Dali alignment, SPLUNC1 (Fig. 4.13a, PDB code 5I7K) from human has the closest structure to BSP30b, with a Z score of 10.5 and a root mean squared deviation (RMSD) for Cα of 3.8 Å (total of 204 residues included). Both proteins are members of the PLUNC group and have a tubular structure with the conserved hydrophobic channel. The conserved disulfide bond is also conserved between the two structures (Fig. 4.13a and b) and thought to maintain structural stability.

There are two significant differences between the two structures. The first difference results from the presence of a second unique disulfide bond (Fig. 4.13b,

coloured in cyan) in BSP30b. This disulfide bond, formed between Cys72 and Cys229, connects the flexible loop just after helix α3 to the C-terminal helix α7. Intriguingly, electron density signifying this bond is only clear in one molecule in the asymmetric unit and is practically missing in the other molecule. This may indicate that this second disulfide bond is partially reduced.

The second obvious difference is that the N-terminal and C-terminal ends of SPLUNC1 are located at the same place on the long axis; whilst both ends of BSP30b are located at the opposite site of the long axis (Fig. 4.13c). In BSP30b, its N-terminal consists of 4 helices (α1-α4) whilst there are only two helices (α1 and α2) present in SPLUNC1. Helix α4 of BSP30b runs roughly parallel to α1/α2 of SPLUNC1, while helices α1-α3 of BSP30b run antiparallel to α1/α2 of SPLUNC1. As a consequence, the respective N-terminus and C-terminus of both proteins are located either on the same (SPLUNC1) or opposite (BSP30b) sides. This structural difference seems to be critical for their ligand binding. Helices α1 and α2 of SPLUNC1 are V-shaped with helix α1 blocking the ligand access to the cavity, and it is anticipated that helix α1 will move away to "open up" the binding site upon ligand binding. For BSP30b, its helix α4 runs straight and leaves the cavity open. This different topology of the N-terminal part of BSP30b seems to be responsible for making the cavity accessible to ligand.

The surface properties of BSP30b and SPLUNC1 are similar, containing negatively charged regions at their potential substrate binding sites (Fig4.14 c and d). This is different to BPI, whose substrate binding site is positively charged and it functions to neutralize the negative charge of LPS through formation of salt bridges. Since

SPLUNC1 has been proposed to be a lipid binding protein based on its ligand binding site charge and *in vitro* assay, we believe BSP30b could also be a lipid binding protein because its ligand binding site has similar electrostatic potential as SPLUNC1, although their binding sites are formed differently.

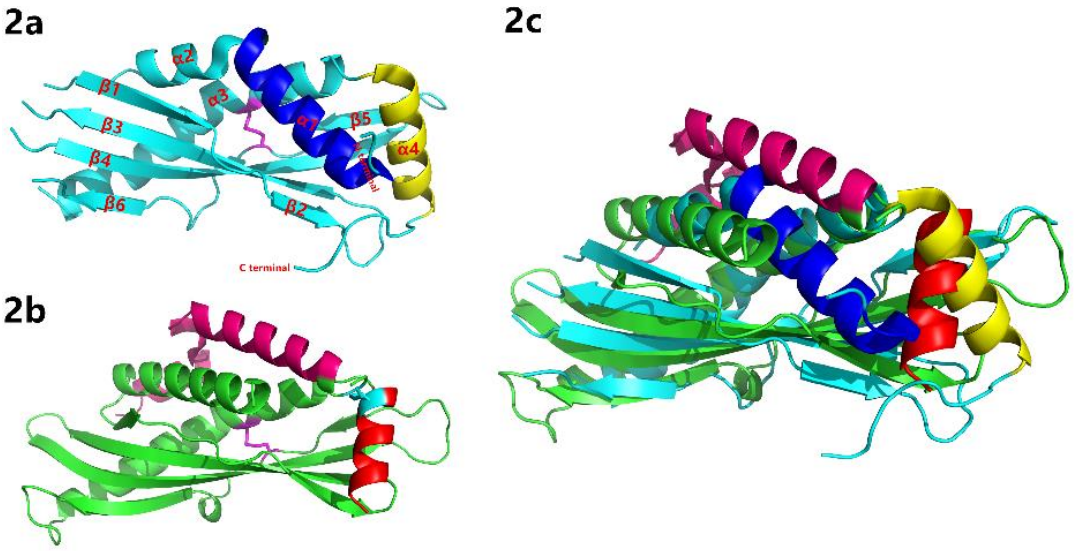

Figure 4.13 Structural comparison of BSP30b with hSPLUNC1. 2a, the structure of hSPLUNC1, with its disulfide bond shown in magenta. 2b, the structure of BSP30b. The universal disulfide bond is shown in magenta and its unique disulfide bond is shown is cyan. 2c, structural alignment of BSP30b to hSPLUNC1 using PyMOL. BSP30b is shown in cyan and hSPLUNC1 is shown in green. Their structural differences are marked with different colours. For BSP30b, helices α1, α2, α3 is in hotpink and helix α7 is in red. For hSPLUNC1, helix α1 is in blue and helix α4 is in yellow.

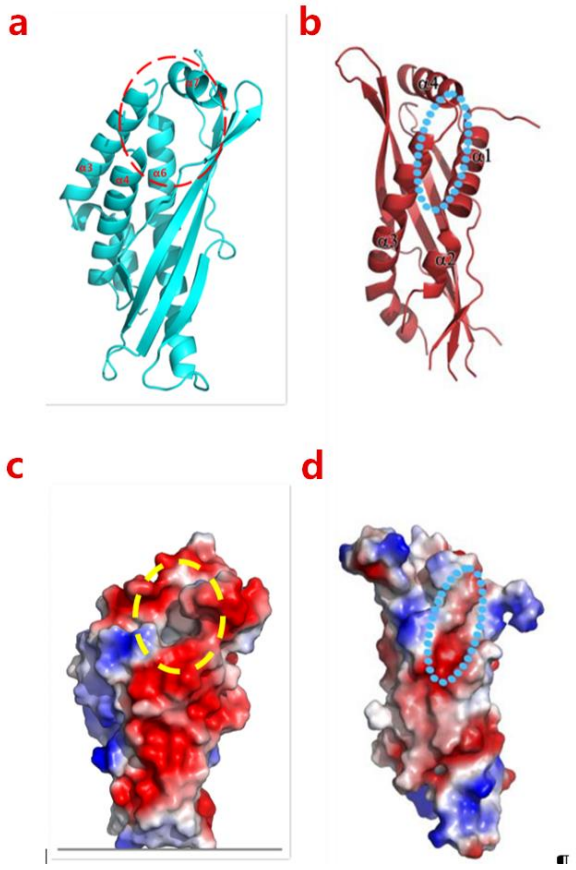

Figure 4.14 Comparison of the potential ligand binding sites of BSP30b and hSPLUNC1. The potential ligand binding sites were highlighted using dashed circles in each figure. a, cartoon representation of BSP30b. b, cartoon representation of hSPLUNC1. The electrostatic potential of BSP30b (c) and hSPLUNC1 (d) were depicted. Colours indicate the surface electrostatic potential: blue = positive (or above 0.1 Volt), white = neutral and red = negative (or below 0.1 Volt).

4.3.7 Structural comparison of BSP30b to TULIP proteins with known structures

The structure of BSP30b was also compared with the TULIP proteins whose structures have been deposited in the PDB. A sequence alignment of these proteins (13 in total) based on structure alignment was generated using PROMALS3D (Fig. 4.15, http://prodata.swmed.edu/promals3d/promals3d.php). Generally, the secondary structures constituting the conserved tubular channel are well conserved among all of the proteins, although their amino acids are poorly conserved. It is also

clear that the C-terminal domains of BPI (1ewf), LBP (4m4d), and CETP (2obd) are well aligned but are different from E-syt 2 (4p42). This is because that the C-terminal domain of E-syt 2 interacts with lipid bilayers, rather than lipid binding/transportation (Schauder et al., 2014).

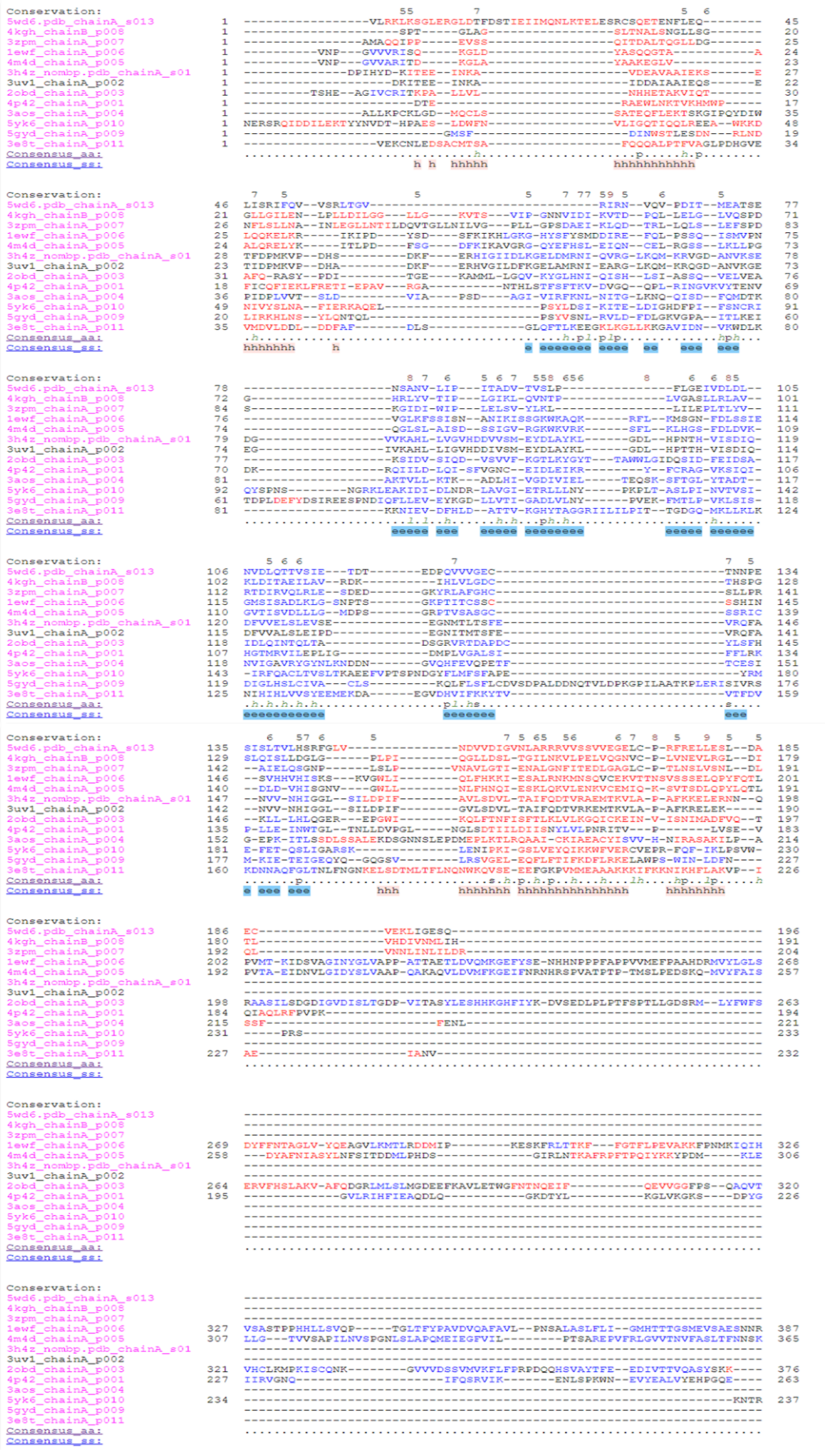

Figure 4.15 Sequence alignment of BSP30b with other 12 TULIP proteins based on structure alignment using PROMALS3D (http://prodata.swmed.edu/promals3d/promals3d.php). The corresponding proteins of these PDB codes in the figure are: BSP30b (5wd6), SPLUNC1 (4kgh), latherin (3zpm), BPI (1ewf), LBP (4m4d), Der p 7 (3h4z), Der f 7 (3uv1), CETP (2obd), E-syt 2 (4p42), JHBP (3aos), Mmm1 (5yk6), Mdm (5gyd), Takeout 1 (3e8t). The conservative property of each aligned position is valued between 0 and 9, with 9 corresponding to highest conservation. The consensus_ss represents the conserved secondary structure, with h standing for alpha-helix (the corresponding amino acids are coloured in red) and e standing for beta-strand (the corresponding amino acids are coloured in blue). The amino acid having same property at each position is represented in consensus_aa, with p standing for polar, h standing for hydrophobic, l standing for aliphatic, and s standing for small.

4.3.8 Sequence Alignment of BSP30a and b

A sequence alignment of BSP30a and b aiming to map the BSP30a sequence onto the structure of BSP30b was performed using PROMALS3D (Fig. 4.16). It is obvious that the two protein sequences are well aligned and are highly conserved (83% identity). However, some amino acid differences are also found in the alignment, which may affect the structure and function of BSP30a. The core differences were highlighted on the structure of BSP30b (Fig. 4.17). It is clear that the three hydrophobic amino acids (F85, L126, F188) on the surface of BSP30b are substituted to serine and proline. Since BSP30b is highly soluble and stable, it is hard to tell what the effects of these changes will be. It is also noticed that, one amino acid (F218) of BSP30b whose sidechain is buried in the hydrophobic channel is substituted to isoleucine in BSP30a. Compared to phenolalanine, isoleucine has a smaller sidechain, which might result in slightly larger internal space and different ligands of BSP30a.

```
Conservation:                                  999 9 9999999 999999999 99999999999
BSP30a              1  MVQLWKLVLLCGLLAGTSESLLDIRGNDVLRRLISGLERGLGTFDSTIEIIFQNLKTELESRCLNDVVEE   70
5wd6_chainA_p001    1  ---------------------------VLRKLKSGLERGLDTFDSTIEIIMQNLKTELESRC-------   35
Consensus_aa:              ........................VLR+LbSGLERGLsTFDSTIEIIhQNLKTELESRC.......
Consensus_ss:                   eeeeee       hhhhh    hhhhhhh             hhhhh    hhh    hhhhhhh

Conservation:                   9 999 99 999999999 9999999999999999 99999999 9 999999999999 99999 999
BSP30a             71  TQQTENSLEGLISRIFQVVNRLTGVRIRNVQVPDITFEATSENSADVSIPITADVTVSLPLLGEIVKLDL  140
5wd6_chainA_p001   36  SQETENFLEQLISRIFQVVSRLTGVRIRNVQVPDITMEATSENSANVLIPITADVTVSLPFLGEIVDLDL  105
Consensus_aa:              oQpTEN.LE.LISRIFQVVsRLTGVRIRNVQVPDIThEATSENSAsV.IPITADVTVSLPhLGEIVcLDL
Consensus_ss:              h                 eeeeeeeeeeeee  eeeeee    eeeeeeeeeeeeee      eeeeee

Conservation:                   999999 999999 9  9  999999 999999999999 9 99 99999 9999  9  99999  999
BSP30a             141  NVDLQTSVSIETDAETGDSRVVVGECPNNPESISLTVLHRRPGLLNDVVDFGVNLV-RQLVSSVVQHELC  209
5wd6_chainA_p001   106  NVDLQTTVSIETDTE--DPQVVVGECTNNPESISLTVLHSRFGLVNDVVDIGVNLARRRVVSSVVEGELC  173
Consensus_aa:              NVDLQToVSIETDhE..DspVVVGECsNNPESISLTVLHpR.GLlNDVVDhGVNLh.RplVSSVVp.ELC
Consensus_ss:              eeeeeeeeeeee      eeeee         eeeeee     hhhhhhhhhhhhh hhhhhhhhhhhh

Conservation:                   99 99999999 99  99999 9
BSP30a             210  PRIRELLESLDTECIKKLIGEPQVTTQQEI  239
5wd6_chainA_p001   174  PRFRELLESLDAECVEKLIGESQ-------  196
Consensus_aa:              PRhRELLESLDhEClcKLIGEsQ.......
Consensus_ss:              hhhhhhhhh  hhhhhhhhhhhhhh
```

Figure 4.16 Sequence alignment of BSP30a and b using PROMALS3D. The predicted secondary structures of BSP30a are highlighted in blue (beta-strand) and red (alpha-helix) on amino acid sequence. The conserved secondary structures are also represented in consensus_ss, with e standing for beta-strand and h standing for alpha-helix.

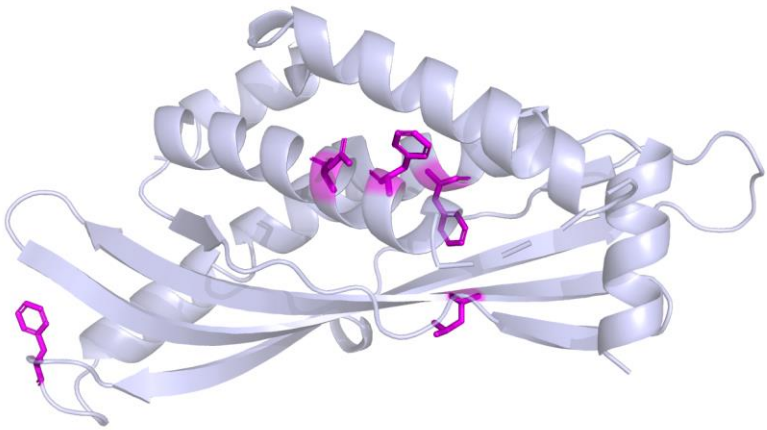

Figure 4.17 The cartoon representation of BSP30b generated by using PYMOL. The major amino acid differences between BSP30a and b were highlighted in magenta, with sidechains shown as sticks. The highlighted amino acids are F85, L126, F188, and F218 of BSP30b.

4.4 BSP30b Structure Solved Using MBP_BSP30b_F52M_F106M

4.4.1 Crystallization of MBP_BSP30b Mutant

Purified MBP_BSP30b double mutant was subjected to fine screen at a concentration of 10 mg/ml using the previously determined 24 conditions for BSP30b. Single protein crystals were grown in conditions B3-B6, C3-C6, and D3-D6 after 2 weeks (Fig. 4.18). Those crystals were taken to the Australian Synchrotron and most diffracted to ~ 2.5 Å. It was noticed that strong ice rings were seen during X-ray diffraction. This could be caused by a film layer formed over crystallization drops which may have attached to the surface of the protein crystals during looping. Nevertheless, data was collected for structure determination.

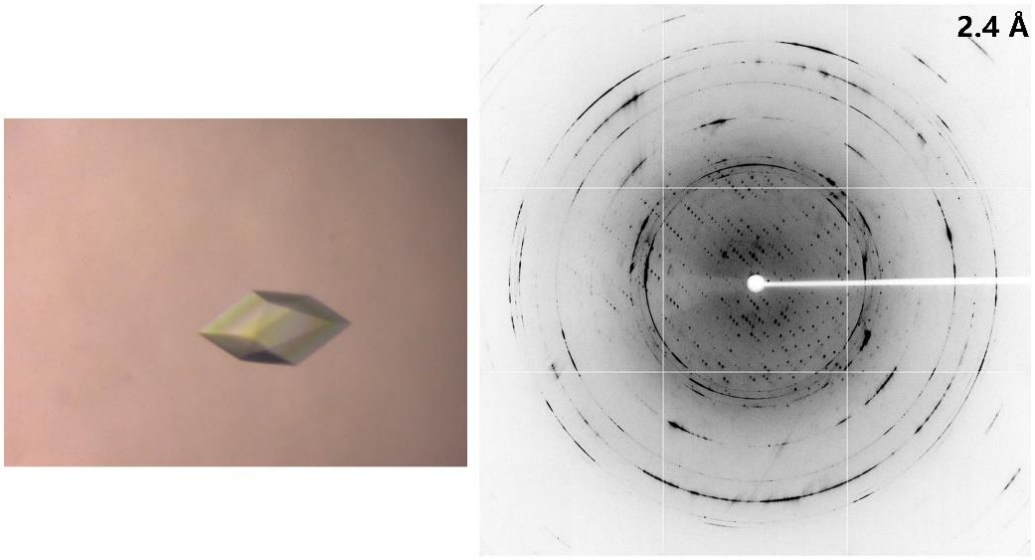

Figure 4.18 Crystal of MBP_BSP30b_F52M_F106M with diffraction of 2.4 Å. Strong ice rings were shown during data collection.

4.4.2 Data Collection and Initial Analysis

Data was collected to 2.4 Å resolution with 94.5% completeness at a wavelength of 0.9537 Å. The dataset was indexed as C2 space group and integrated into an mtz file using iMosflm (Leslie & Powell, 2007). Initial Matthews Probability analysis indicated that the asymmetric unit has only one protein molecule and the solvent content is 16.5%, surprisingly low for a protein crystal (Table 4.6).

Table 4.6 Matthews_Coeff results

Nmol/asym	Matthews Coeff	%solvent	P(2.40)	P(tot)
1	1.47	16.50	1.00	1.00

Initially, molecular replacement method was chosen to solve the structure of MBP_BSP30b mutant using the known structure of MBP on PDB (PDB code: 4MBP or 1ANF). However, molecular replacement using phaser-MR failed to generate any solutions. During this time, the structure of BSP30b was solved using its SeMet derivative. Hence, molecular replacement was performed using BSP30b as the reference model. When looking back to the integrated dataset, it had the same space group as BSP30b itself and similar unit cell parameters. It is suspicious that only BSP30b_F52M_F106M was crystallized and MBP was removed during crystallization for some reason.

After using BSP30b as a reference model and two molecules in the ASU, the structure was easily solved using *phase_MR* in Phenix. The resulting structure was first refined with *phenix.refine* in Phenix. Subsequent cycles of manual refinement using COOT 0.7 and *phenix.refine* were done until desirable R_{free} and R_{work} were reached (Table 4.7).

Table 4.7 Data collection and refinement parameters

Property	Value
Wavelength	0.9537 Å
Space group	C121
Cell constants a, b, c, α, β, γ	83.35Å 61.23Å 90.85Å 90.00° 107.73° 90.00°
Resolution (Å)	86.53 – 2.40 (2.53 – 2.40)
R_{merge}	0.177 (0.698)
CC (1/2)	0.998 (0.929)
Completeness (%)	94.5 (84.4)
Redundancy	6.0 (6.2)
No. of observations	97078 (13004)
No. of unique reflections	16290 (2113)
Mean I/σ(I)	5.5 (1.6)
R_{work}	0.2339
R_{free}	0.3062
Average B, all atoms ($Å^2$)	64.77
Ramachandran favoured (outliers)	96.23% (0.54%)
Rotamer outliers	0.00%
Clash score	14.67

4.4.3 Structural Analysis

The structure of BSP30b was also solved using MBP tagged BSP30b, although MBP was removed during crystallization. Crystal diffracted to 2.4 Å with space group of C2 and dimensions: a = 83.35Å, b = 61.23Å, c = 90.85Å, α = γ = 90°, β = 107.73°.

BSP30b was crystallized as two molecules in the ASU as shown in Fig. 4.19. The structure also contains the well conserved hydrophobic channel. Although it was solved at a relative low resolution, it still maintained all the features as the one solved by its SeMet derivative. Both structures can be well aligned using PyMOL with a RMSD of 0.25 (149 to 149 atoms, fig. 4.20). It is interesting that both structures were solved from Valine38. The first 38 amino acids including the 6x His tag were missing in both structures, suggesting that this region should be very flexible and not well packed in the crystal. The potential ligand binding region was also well formed and the electron density for the loop between helices α3 and α4 was also missing. Both structures ended with the C-terminal α helix and the last 13 amino acid residues were missing in both structures.

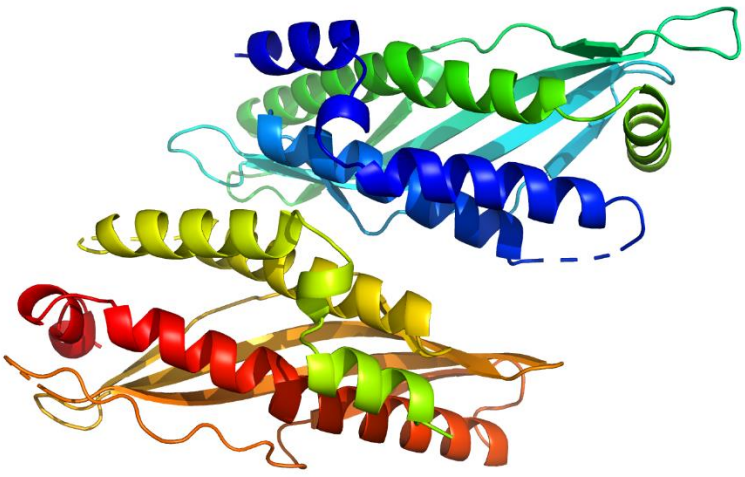

Figure 4.19 Cartoon representation of BSP30b structure. BSP30b was solved using MBP_BSP30b_F52M_F106M with two molecules in its ASU.

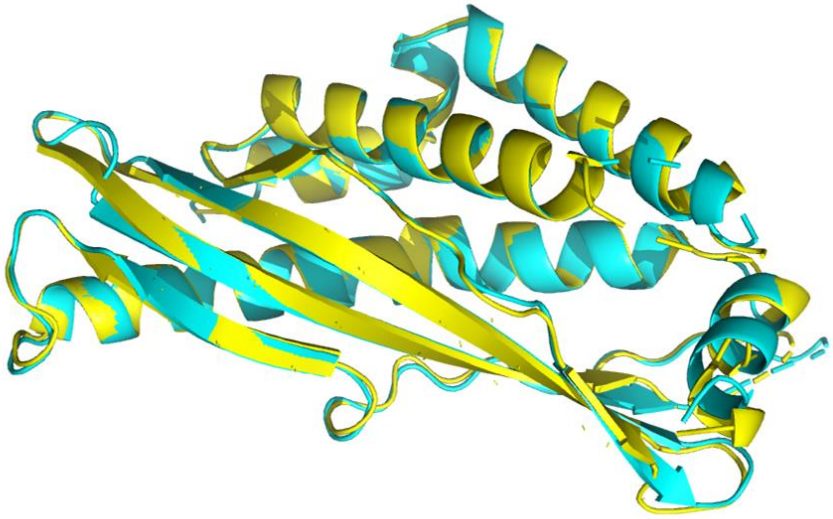

Figure 4.20. Structural alignment of BSP30b solved by using BSP30b_F52M_F106M SeMet derivative (colored in cyan) and MBP-BSP30b_F52M_F106M (colored in yellow).

4.5 Discussion

4.5.1 Multiple Optimizations of BSP30b Crystallization Conditions Failed to Yield High Quality Crystals

Multiple optimizations of crystallization conditions for BSP30b were conducted in this chapter aiming to produce protein crystals with high quality diffraction. Initial robot and fine screens only resulted in thin needles with very poor diffraction. Twenty four previously determined conditions were also used to set up fine screens and condition “B3” repeatedly had needle or plate shaped crystals which diffracted to around 6 Å resolution, insufficient to solve the structure. .

Multiple optimizations of condition “B3” in this chapter improved the crystal quality of BSP30b. Additive Screens from Hampton Research contains 96 unique reagents, which can affect the solubility and crystallization of biological macromolecules. These reagents can perturb the interactions between molecules

including the protein-protein interactions, or protein-solvent interacgions, as well as water structure. In this study, crystals were produced in 6 additive screen conditions, four of which produced better-quality crystals ("B3" + additive glucose monohydrate, galactose, glycine, or sodium iodide). Unfortunately, those crystals only marginally improved the diffraction to around 5 Å resolution.

The Silver Bullets Screen is a library of small molecules which can promote stability, lattice formation, and crystallization. These small molecules are able to establish stabilizing, inter molecular, hydrogen bonding, hydrophobic and electrostatic interactions. Although the silver bullet screen yielded better-shaped needle crystals in this study, only poor diffraction was detected, suggesting that needle shaped crystals of BSP30b were not ideal for X-ray diffraction.

In situ trypsin proteolysis is a well-developed method for improving protein crystallization, and aims to produce crystals after removing the protein residues in flexible regions. Dong *et al*. (Dong et al., 2007) reported that, after *in situ* proteolysis, 10 of 69 proteins which resisted structure determination were crystallized and their structures determined (14.5% success rate). Later research expanded to over 270 proteins that had failed in the past to produce crystals suitable for structure determination revealed that 34 yielded deposited crystal structures with an 12.6% success rate (Wernimont & Edwards, 2009). However, in this study, although the flexible regions were trimmed after *in situ* proteolysis, only tiny crystals were produced with uneven surface which had very poor diffraction.

Seeding is also a well-developed method for optimization, which involves adding homogeneous or heterogeneous crystals to a crystallising solution to nucleate and/or grow more crystals (Bergfors, 2003). It is a powerful tool for the separation

of nucleation and growth. In this study, streak seeding yielded crystals but desolved later before collecting diffraction data. Batch seeding resulted in crystals of BSP30b with best quality, which had diffraction of ~ 3 Å. Although this is not enough to solve the structure of BSP30b using either direct methods or S-SAD, heavy atom soaks were performed using those crystals for multiple isomorphous replacement (MIR). Gel shift assays had clearly shown binding of gold and platinum atoms to BSP30b, however, crystals soaked in their compound solutions failed to generate any anomalous signals after scaling the X-ray diffraction dataset.

4.5.2 Structure of BSP30b Solved Using SeMet Derivative

The heavy atom selenium has been successfully used for phase determination via single-wavelength anomalous dispersion (SAD) after its incorporation into protein as the modified residue selenomethionine (SeMet). In this study, we firstly designed a double methionine mutant to allow incorporation of SeMet into the protein. Although BSP30b_F52M_F106M SeMet derivative was only expressed at low levels, it crystallized as rod-cluster shaped crystals with sharp and defined edges using the previously determined crystallization conditions. Native BSP30b was expressed in pTriEx-4, which leaves a longer 6x His tag at the N-terminal end of BSP30b compared to BSP30b_F52M_F106M expressed in pProEx Htb. This longer His tag may have increased the flexibility of native BSP30b, which may increase the difficulty of producing high quality crystals of native BSP30b.

Here is the first reported crystal structure of BSP30b at 2.0 Å resolution. The structure shows the conserved tubular-like structure of the TULIP superfamily, with its hydrophobic channel formed by α/β wrap. Structural comparison to the closely related structure in the PDB (human SPLUNC1) has revealed both conserved and

unique features which might be important for function. The unique disulfide bond present in BSP30b but not in other PLUNC proteins may be important for its ligand binding given it proximity to the entrance to the ligand binding cavity. Further evidence for this includes the fact that a similar disulfide bond is also found in the structure of Takeout 1 of *Epiphyas postvittana* and JHBP proteins of silkworm, which are also members of the TULIP superfamily. Their disulfide bonds are reported to have a critical role in ligand binding rather than maintaining the stability of the entire structure (Hamiaux et al., 2013; Suzuki et al., 2011). Additional evidence is provided by the fact that this disulfide bond is only partially intact in the structure of apo-BSP30b and is broken after binding of oleic acid (see Chapter 5).

The N-terminal helical part of BSP30b is different from SPLUNC1. For SPLUNC1, its N-terminal helices α1 and α2 are V-shaped and the access of potential ligands is blocked by helix α1, unless major conformational change occurs to "open-up" the binding site. While the re-arrangement of N-terminal helices of BSP30b leaves the cavity open.

4.5.3 MBP was Removed During Crystallization of MBP Fused BSP30b_F52M_F106M

It is interesting that crystallization trials using MBP-BSP30b_F52M_F106M produced the best-looking crystals (singular and diamond-like, with sharp edges). However, those crystals had relative worse diffractions (> 2.4 Å) than BSP30b_F52M_F106M. The processed dataset also had worse parameters but still capable of solving the structure of BSP30b using molecular replacement.

The solved structure using MBP-BSP30b_F52M_F106M still maintains the core structure of BSP30b. It is interesting that, in both solved structures, the first 38 and the last 13 amino acids were missing. This may suggest that those amino acids are highly flexible and not well packed in the structure, or they were partially removed by the proteases expressed from *E. coli*.

Chapter Five. Ligand Binding and Biological Function of BSP30b

5.1 Introduction

The BSP30b structure maintains the tubular structure of the TULIP superfamily, and this suggests that its biological ligand may be a fatty acid similar to those which have been proposed to be the ligands of proteins in the TULIP superfamily. Indeed, fatty acids were found in the hydrophobic channels of several proteins in the TULIP superfamily. Since BSP30b is a salivary protein and the primary lipid source for cattle is from pasture grass, the binding of the dominant fatty acids in pasture grass to BSP30b was examined by co-crystallization.

Green Fluorescent Protein (GFP) was discovered as a companion protein to aequorin from Aequorea jellyfish. It is widely used in cell biological applications as a tag or as an indicator. In this study, BSP30b was fused to GFP so its interaction with lipids could be visualized under fluorescent microscope. Olive oil was used instead of lipids extracted from pasture grass due to its similar lipid composition and availability.

Because BSP30b secreted in cattle's saliva will enter the rumen, its interaction with rumen microorganisms was also examined. It was thought that BSP30b might bind to specific components on certain bacterial surfaces, which may give hints to its biological role within the rumen. GFP tagged BSP30b was used and mixed with rumen microorganisms for this purpose, followed by visualization under fluorescent microscope.

5.2 Purification and Crystallization of BSP30b_F52M_F106M with Potential Ligands

5.2.1 Expression and Purification of BSP30b_F52M_F106M with its Potential Ligands

BSP30b_F52M_F106M was expressed in 1 L LB media instead of M9 minimal media for crystallization. However, its expression level was not improved by using LB media. The sonicated culture was then purified with IMAC, Anion Exchange Chromatography, and SEC. After the final purification step, those fractions containing BSP30b were concentrated to 2 mg/mL and incubated with 20 µL of either linolenic acid, ethyl linoleate, glycerol monolinoleate, or oleic acid.

The mixture of BSP30b mutant protein with individual fatty acids was incubated overnight and then purified with SEC. Chromatography and SDS PAGE indicated that BSP30b mutant showed no change after mixing with fatty acids (Fig. 5.1). It eluted as a single peak on chromatography and single band on SDS PAGE.

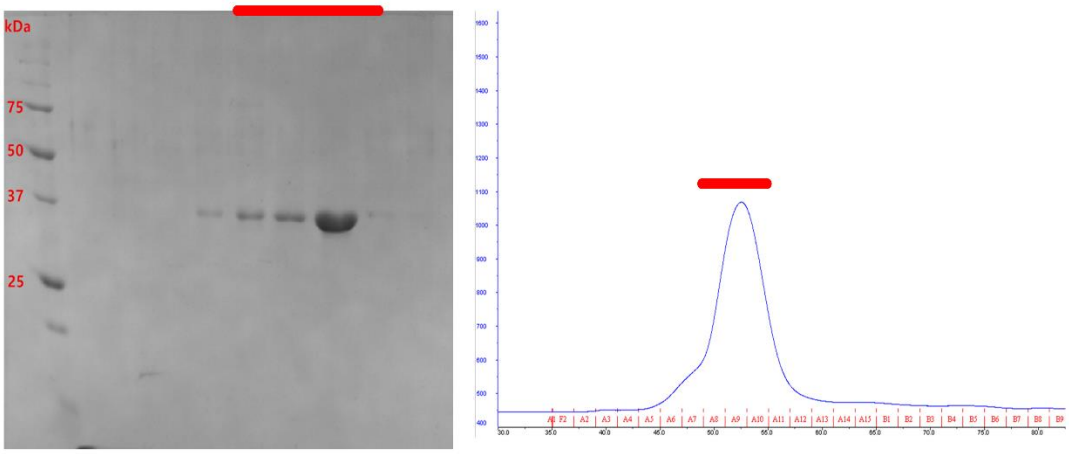

Figure 5.1 IMAC purification and corresponding SDS-PAGE analysis of BSP30b_F52M_F106M with fatty acids. Peaks on chromatograms and their corresponding protein bands on SDS PAGE gel were highlighted.

5.2.2 Crystallization of BSP30b Mutant with Potential Ligands

Co-purified BSP30b mutant with different fatty acids were subjected to fine screens at a concentration of 8 mg/ml, separately. After setting up fine screens, most of the protein was precipitated in the drop immediately. However, singular crystals with sharp edges appeared after about 2 weeks in conditions B3-B6, C3-C6, D3-D6 (Fig. 5.2). Those crystals were taken to the Australian Synchrotron and most diffracted to ~ 2 Å. Data was collected for structure determination *in silico*.

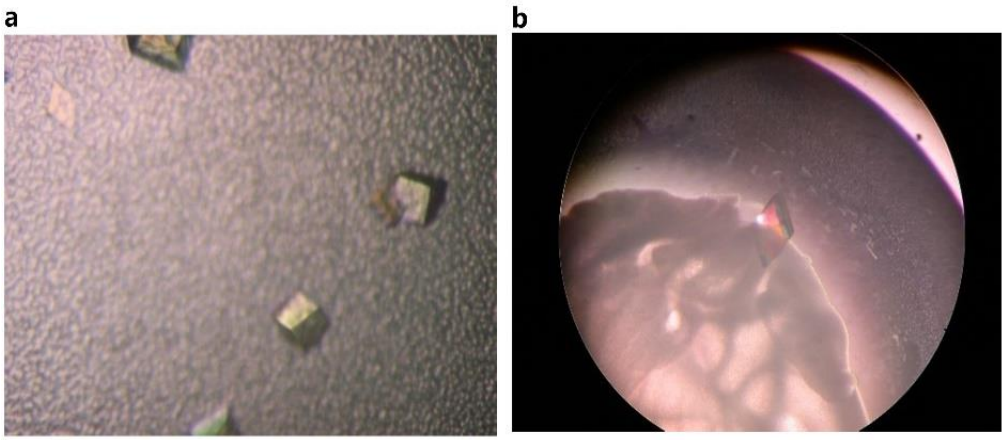

Figure 5.2 Crystals of BSP30b_F52M_F106M mxied with fatty acids. Singular crystals were appeared after 2 weeks with sharp edges. It is noted that protein was partially precipitated in the drops as shown in both a and b.

5.2.3 Initial Data Analysis

All the datasets collected for BSP30b mutant with different fatty acids were automatically processed and scaled at the Australian Synchrotron. The structures of BSP30b mutant with fatty acids were all solved using molecular replacement (*Phaser-MR* in PHENIX suite). The solved structure of BSP30b mutant was used as the reference model and all the structures solved by Phaser-MR with high LLG and TFZ values, indicating successful molecular replacement as expected.

Initial inspection of those structures in Coot revealed that only BSP30b mutant co-crystallized with oleic acid had unoccupied electron densities within its hydrophobic channel. Subsequent refinement was carried out for BSP30b mutant-oleic acid complex only.

Oleic acid searched from monomer library in Coot was placed into this positive contoured electron density and the structure was refined with *phenix.refine*. Subsequent cycles of manual refinement using COOT and automatic refinement with *phenix.refine* were done until R_{free} did not reduce further (Table 5.1).

Table 5.1 crystallographic data collection and refinement statistics

Property	Value
Wavelength	0.95369 Å
Space group	C121
Cell constants a, b, c, α, β, γ	83.75Å 61.40Å 91.09Å 90.00° 107.74° 90.00°
Resolution (Å)	86.76 – 1.97 (2.02 – 1.97)
R_{merge}	0.067 (0.683)
CC (1/2)	0.998 (0.912)
Completeness (%)	99.7 (99.7)
Redundancy	5.7 (6.1)
No. of observations	177824 (13279)
No. of unique reflections	31005 (2186)
Mean I/σ(I)	9.9 (2.3)
R_{work}	0.2326

R_{free}	0.2878
Total number of atoms	3064
Average B, all atoms (ligands) ($Å^2$)	55.88 (63.48)
Ramachandran favoured (outliers)	94.29% (0.54%)
Rotamer outliers	0.32%
Clash score	10.36

5.2.4 Structural Analysis of BSP30b Mutant – Oleic Acid Complex

The protein – oleic acid complex was also solved in C2 space group with two protein molecules in its ASU (Fig. 5.3a). The hydrophobic end of oleic acid is buried in the tunnel and its carboxyl group stretches out from the tunnel's opening, possibly forming hydrogen bonds with the carboxyl group of ASP112 and the carbonyl group of ILE113 (Fig. 5.3b). We suggest that the hydrophobic tail of oleic acid first "glides" into the internal cavity, driven by strong hydrophobic interactions, while its carboxyl group stretches out and is stabilised by forming hydrogen bonds with proximate amino acids. A similar hypothesis has been proposed for *Epiphyas postvittana* Takeout 1 protein (Hamiaux, Basten, Greenwood, Baker, & Newcomb, 2013), where both fatty acid ligands (including the head groups) "slide down" into the hydrophobic tunnel.

We also observe that the overall structure of BSP30b and its internal cavity remain unchanged regardless of the binding of lipid ligand, suggesting that the internal cavity of BSP30b also forms a rigid scaffold for its lipid ligands. Only the helix α3 moves away very slightly, which may result from the breaking of the unique disulfide bond upon oleic acid binding. When comparing Fig. 5.3 and 4.12, it seems

that the internal cavity is narrowed towards its end in the apo structure, while it has expanded around the ligand in the ligand bound structure. This is due to the side chain movement of L68, L86, F218, and L221, which will optimize the fit around the bound ligand.

It is also clear that upon ligand binding, the overall B-factor of BSP30b increased from 36.0 $Å^2$ to 55.9 $Å^2$, suggesting the ligand bound protein has greater molecular motions. For both ligand bound and apo BSP30b, the amino acids forming the hydrophobic channel have relatively low B-factors. While the amino acids close to the lipid binding opening have much higher B-factors, such as the amino acids of helices α3 and α7.

It is noteworthy that the electron density of oleic acid is relatively clear in one molecule of the asymmetric unit but incomplete for oleic acid in the other molecule (Fig. 5.3c). The occupancies of both oleic acids in the asymmetric unit were refined in PHENIX to 0.81 and 0.86, respectively, suggesting that both of the hydrophobic channels are only partially occupied. This may also indicate that oleic acid is not the preferred ligand of BSP30b.

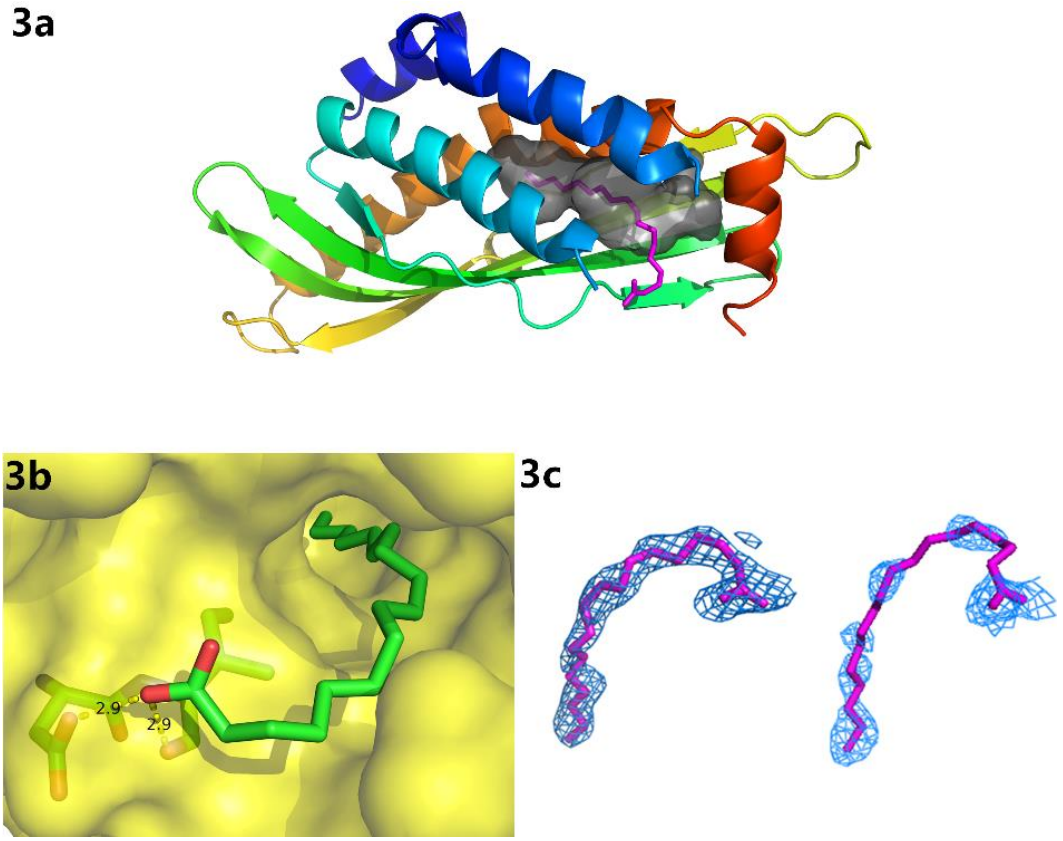

Figure 5.3 Structure of BSP30b with bound oleic acid. 3a, the structure of BSP30b with bound oleic acid. Oleic acid is coloured in magenta and well-positioned in the internal cavity, which is shown in grey. 3b, surface representation of the binding pocket with oleic acid. The potential hydrogen bonds between the carboxyl group of oleic acid and carboxyl group of ASP112, and hydroxyl group of ILE113 are shown as yellow dashed lines. 3c, electron density maps around both oleic acids in the asymmetric unit. The final σa-weighted 2mFo-DFc map contoured at 0.8 σ (blue). All structural illustrations were prepared with PyMOL.

5.3 Generation, Expression and Purification of GFP (Green Fluorescent Protein) Fused BSP30b

BSP30b is hypothesised to interact with lipid. To visualize the interaction, N terminal GFP tagged BSP30b was created for this purpose. Both genes were cloned into *pProEx Htb* with a N terminal 6x His tag for protein purification.

5.3.1 Generation of *pProEx Htb_sfGFP* construct in *E. coli BL21*

The *sfGFP* gene was first amplified from plasmid *PIM1773* and ligated into *pProEx Htb* after restriction digestion of both *GFP* amplicon and *pProEx Htb* plasmid

followed by transformation into *E. coli BL21*. After agarose gel electrophoresis of colony PCR amplicons, a band with size of ~720 bp under UV light indicated the success of cloning (Fig. 5.4).

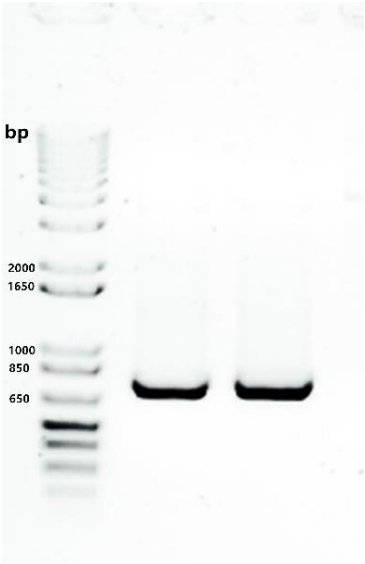

Figure 5.4 Colony PCR results. Colony PCR of GFP using gene specific primers, resulting in an ~ 720 bp band on agarose gel after electrophoresis.

5.3.2 Generation of *pProEx Htb_sfGFP_BSP30b* Construct

The gene encoding BSP30b was ligated into *pProEx Htb* downstream of the *sfGFP* gene. The *BSP30b* gene was first amplified from the *pTrix4* and ligated into *pProEx Htb_sfGFP* after restriction digestion of both BSP30b amplicon and *pProEx Htb sfGFP* plasmid. The ligated plasmid was transformed into *E. coli BL21* cells. Colony PCR with gene specific primers gave the PCR amplicon of ~680 bp as expected after agarose gel electrophoresis; when using primers targeting both *sfGFP* and *BSP30b* genes, the PCR amplicon had a size of ~1400 bp on agarose gel after electrophoresis (Fig. 5.5). Those results together suggested the success of constructing *pProEx Htb_sfGFP_BSP30b*.

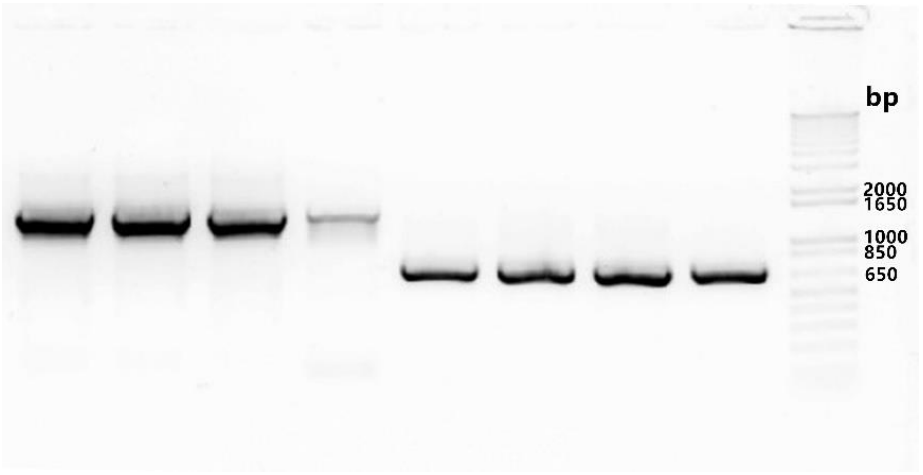

Figure 5.5 Colony PCR results. Colony PCR using gene specific primers for BSP30b resulted a 680 bp band on agarose gel after electrophoresis; while colony PCR using primers targeting both sfGFP and BSP30b genes resulted a ~ 1400 bp band on agarose gel after electrophoresis.

5.3.3 GFP Tagged BSP30b Expression and Purification

GFP tagged BSP30b was expressed in LB media as a 55 kDa recombinant protein fused to a N terminal 6x histidine tag. The colour change of the media from yellow to green after induction immediately indicated the expression of GFP tagged BSP30b.

The expressed protein was first purified with IMAC. The chromatogram and corresponding SDS PAGE demonstrate that GFP_BSP30b was well expressed in *E. coli BL21* and it was successfully eluted off the HisTrap HP column between fractions B2 and B8 (Fig. 5.6a). There were some protein impurities which were also eluted at the same time, especially those having molecular weights around 30 kDa.

The protein was further purified with SEC. The chromatogram and corresponding SDS PAGE suggested that GFP_BSP30b was eluted off the column from fraction

A5 together with few protein impurities, indicated by two close peaks on chromatogram and additional bands on SDS-PAGE (Fig. 5.6b). Most of the protein impurities were eluted after the desired protein. Proteins from fraction A5 to A10 were concentrated for future study.

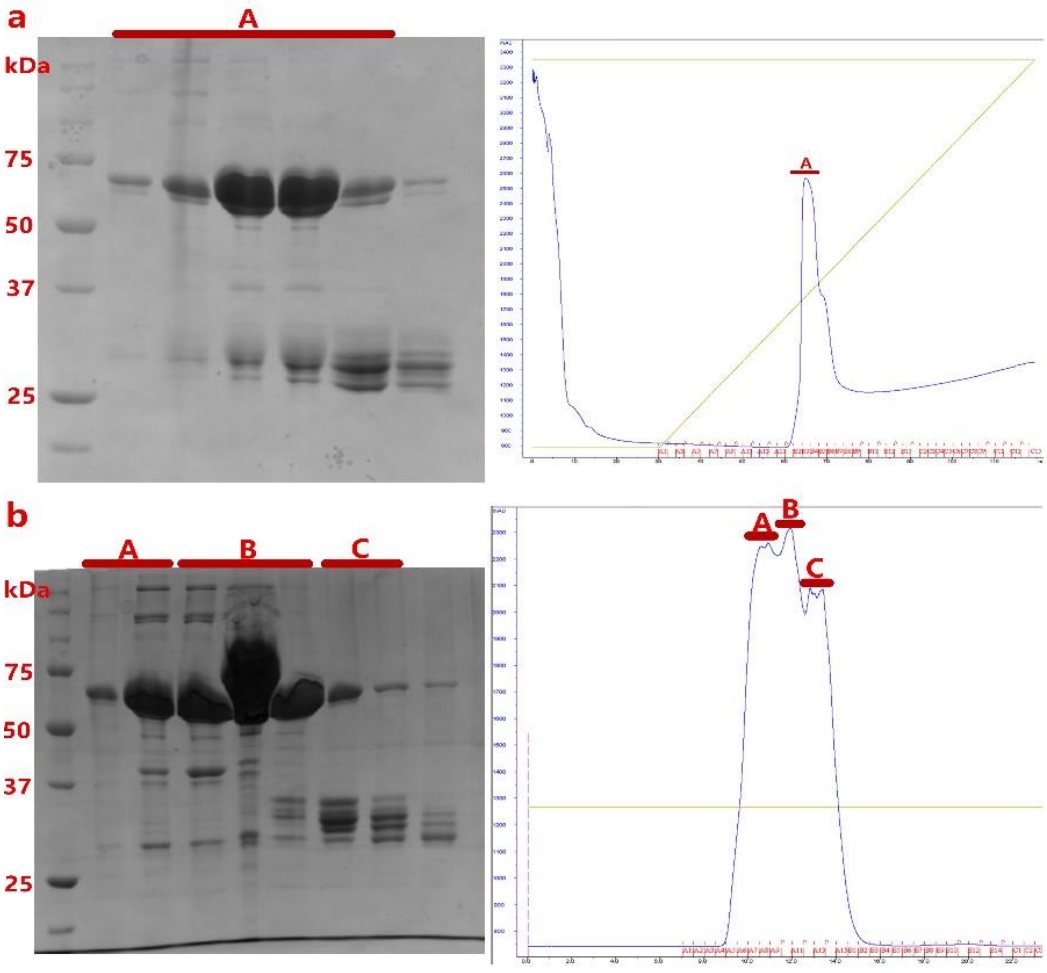

Figure 5.6 Purification of GFP_BSP30b complex. a, chromatogram and corresponding SDS PAGE result of IMAC. b, chromatogram and corresponding SDS PAGE result of SEC. Peaks on chromatograms and their corresponding protein bands on SDS PAGE gel were highlighted.

5.4 Micro emulsions were Formed after Mixing of GFP_BSP30b with Olive Oil

20 μL of olive oil was mixed with 500 μL of GFP_BSP30b at concentration of 2 mg/ml. After 20 min's settling of the mixture, two layers were formed. The top layer was cloudy and bubbly; the bottom layer was slightly cloudy. When viewed under blue light using a fluorescent microscope, the top layer was full of small spheres squeezed against each other (Fig. 5.7a). Protein GFP_BSP30b was located at the surface of each sphere indicated by the green fluorescence colour under blue light. The spheres were made of lipids indicated by the red colour under green light (Fig. 5.7b, lipid was stained by Nile Red). The bottom layer also formed spheres within the solution, but they were apart from each other. When viewing the emulsions under the bench top fluorescent microscope, they also had proteins on the surface and lipids inside of micelles (lipid was stained by Nile Red), suggesting the similar constitution of the emulsions formed in both layers (Fig.5.7c and d).

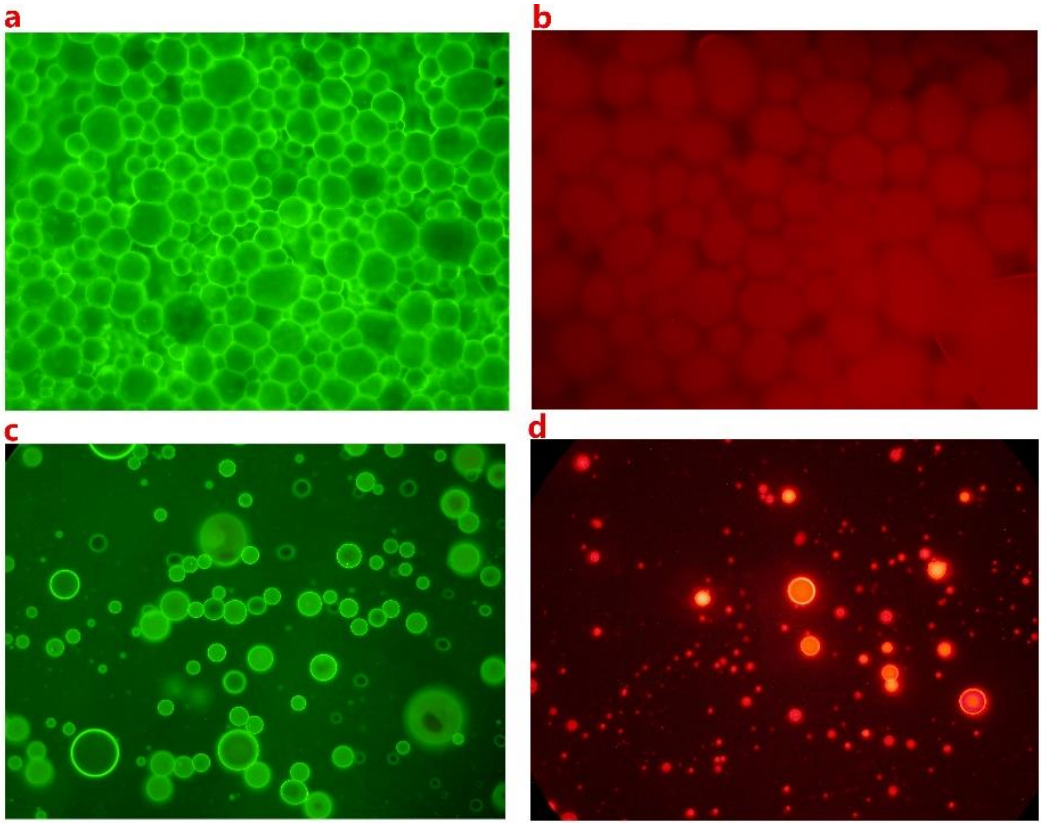

Figure 5.7 Visualization of emulsions formed by mixing GFP_BSP30b and olive oil under fluorescent microscope. When viewed under blue light (a and c), both layer had protein attached to the surface of each micelles. Under green light, lipid located inside of each micelles indicated by red colour (b and d, lipid was stained by Nile Red).

5.5 Size Distribution of Emulsions

In order to determine the maximum emulsification ability of BSP30b, the mixture of BSP30b and olive oil was vigorously shaked for 10 min followed by settling on the bench for 20 min. The size of micelles from the bottom layer was measured using the fluorescent microscope at Waikato Biological Imaging Facility. The bottom layer was diluted 10 times and the size of each micelle was indicated by scale bar on the captured images. It is clear that most of the micelles had a size range between sub-micron to a few microns (Fig. 5.8 a and b). However, it is hard to determine their sizes more accurate because of the resolution limitation of fluorescent microscopy.

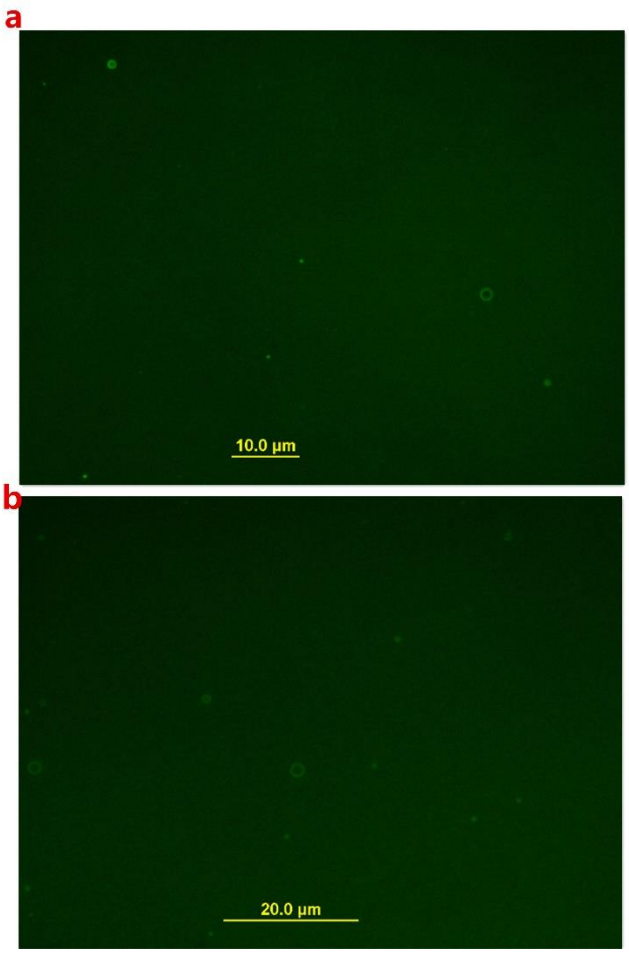

Figure 5.8 Visualization of micelles under fluorescent microscope at the Waikato Biological Imaging Facility. The size of each micelle was indicated by the scale bar on captured images.

To better determine the size of micelles, the emulsion was further visualized using TEM at Auckland University. Diluted emulsions were visualized using negative stain TEM (Fig. 5.9). The protein coated micelles were negatively stained as white spheres with sizes less than 0.5 µm, suggesting that the formed emulsion is micro-emulsions.

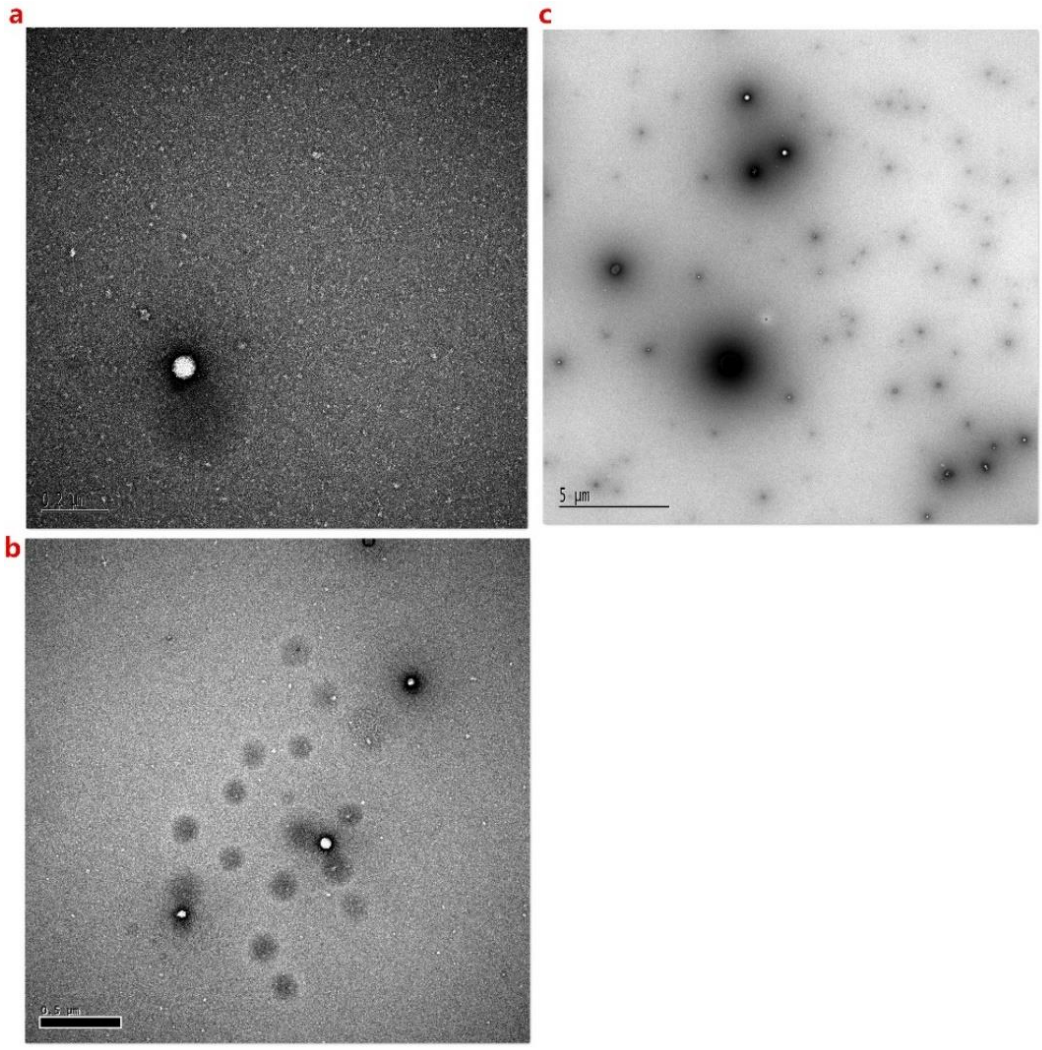

Figure 5.9 visualization of micelles under TEM. The size of each micelle was indicated by the scale bar on each photo.

5.6 GFP_BSP30b Selectively Binds to Specific Rumen Bacteria

5.6.1 Binding of GFP_BSP30b to Selected Rumen Fluid Micro-organisms

The binding ability of GFP_BSP30b to rumen micro-organisms was first examined with cheese cloth filtered rumen fluid, which contains all the micro-organisms within rumen including protozoa and bacteria. The concentration of GFP_BSP30b

at 0.1 mg/ml was tested to be the optimal concentration for viewing under fluorescent microscope when mixed with rumen fluid at a ratio of 1:1.

When viewing the mixture under bright field, both protozoa and bacteria were clearly seen together with dry matter residues (Fig. 5.10a). After switching to blue light, only bacteria were fluorescent but it is hard to determine their morphologies and whether all bacteria were bound by GFP_BSP30b (Fig. 5.10b). To further investigate the binding specificity of GFP_BSP30b, protozoa and bacteria were separated and their binding to GFP_BSP30b was investigated respectively.

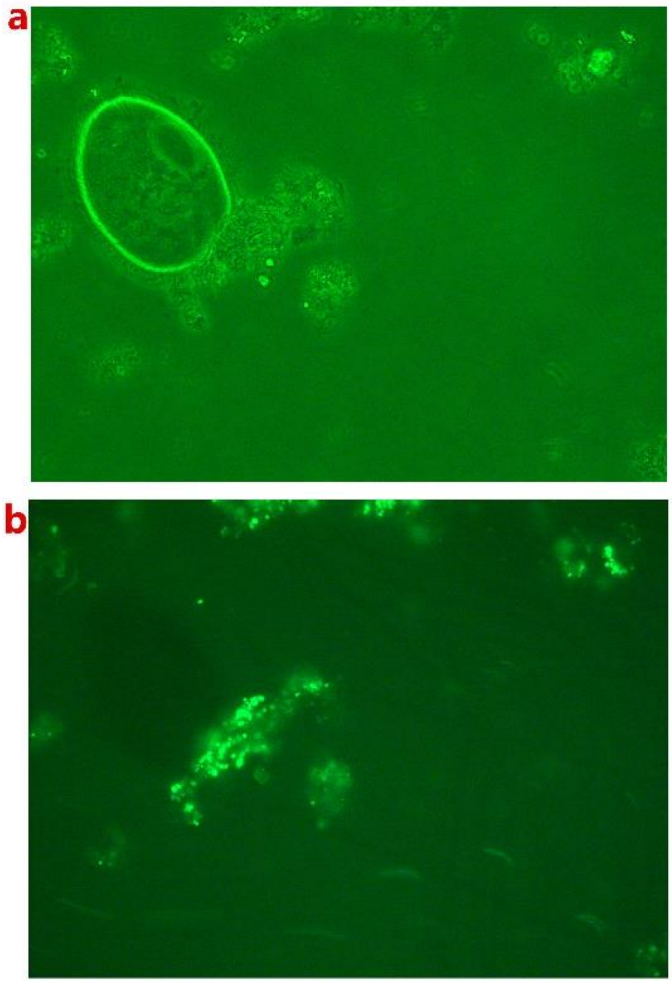

Figure 5.10 Binding assay of GFP_BSP30b to rumen fluid. a, photo was taken under bright field. Protozoa, bacteria, and drymatter residues were seen. b, photo was taken under blue light. Only bacteria was fluorescent but it is hard to determine their morphology.

5.6.2 GFP_BSP30b does not Bind to Protozoa from Rumen

When viewing the protozoal fraction mixture with GFP_BSP30b under bright field, protozoa were clearly seen as in figure 5.11a. Whilst viewing the mixture under blue light, there were no fluorescent signals for protozoa, suggesting that there is no interaction between protozoa and BSP30b. It was noticed that there was one glowing spot on the left of figure 5.11b, which could be bacterial contamination in the protozoal fraction.

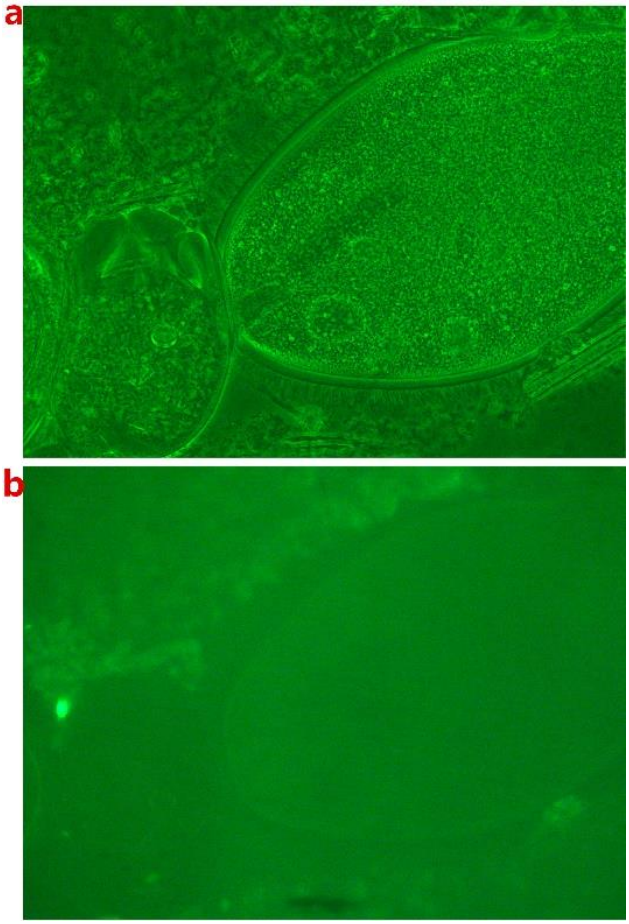

Figure 5.11 Binding assay of GFP_BSP30b to protozoa fraction of rumen fluid. a, photo was captured under bright field. Protozoa was seen in this fraction. b, photo was taken under blue light. Only the outline shade of protozoa was seen, suggesting no binding of protein to protozoa. The single bright spot could be bacteria contamination during fractionation.

5.6.3 GFP_BSP30b Selectively Binds to Certain Rumen Bacteria

When viewing the mixture of bacteria and protein under bright field, a range of bacterial species were seen (Fig. 5.12a). The most surprising result to us is the observation that the mixture of GFP_BSP30b with ruminal bacteria observed under blue light shows that BSP30b binds to just a small number of rumen bacterial species (Fig. 5.12b). It is clearly seen that BSP30b has specific binding to a short and curved rod-like bacterial species, which is clustered as a triad or tetrad (Fig. 5.12b). We observed that BSP30b can also bind to a few cocci-like species but the binding is relatively weak. We suggest that the binding specificity of BSP30b to certain bacterial surface lipids may result in its binding selectivity to a small subset of ruminal bacteria strains.

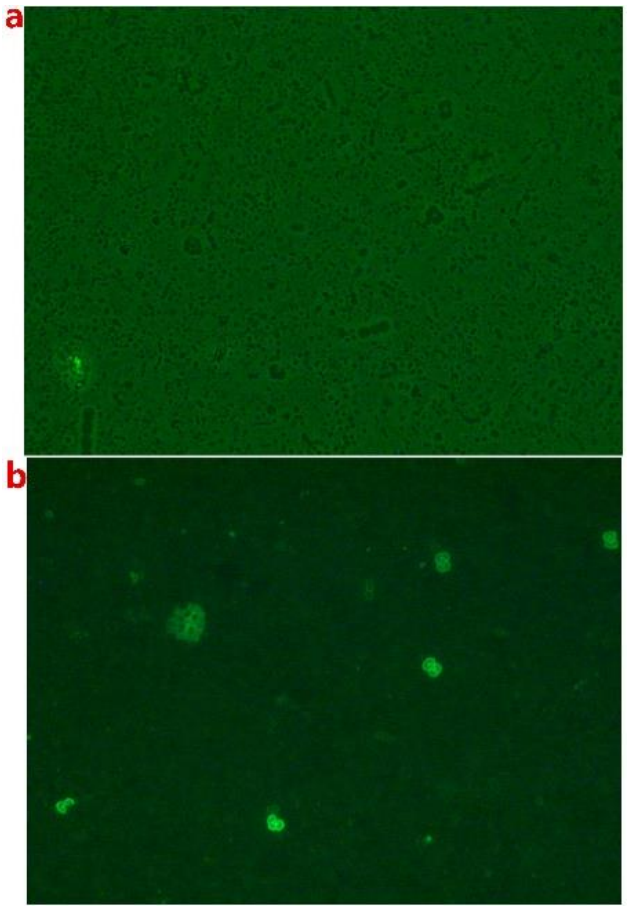

Figure 5.12 Binding assay of GFP_BSP30b to bacterial fraction of rumen fluid. a, under bright field, rumen bacteria with different morphologies were seen. b, under blue light, only a small fraction of rumen bacteria was fluorescent.

5.7 Discussion

5.7.1 BSP30b is A Lipid Binding Protein

Although PLUNC proteins belong to the TULIP lipid binding protein superfamily, there was no structure of PLUNC proteins that has been solved with its lipid ligands. Although SPLUNC1 from human has been demonstrated to bind to DPPC using lipid binding assay on solid surface, its structure did not show any lipid binding to its internal cavity.

This is the first time that a PLUNC family protein has been crystallized with a ligand (oleic acid), although the electron density map of the ligand is not complete. We have also tried to crystalize BSP30b with linolenic acid and linoleic acid, but in the structural data showed no electron density for either fatty acids. Linolenic acid and linoleic acid have more rigid structures compared to oleic acid as they have three and two double bonds on their carbon backbones, respectively. In contrast, oleic acid has only one double bond, which gives it more flexibility and make the "sliding" of oleic acid into the relatively narrow hydrophobic channel possible.

It was noticed that there is no major conformational change of BSP30b after oleic acid binding. Its overall structure and hydrophobic channel were well maintained. This is in contrast to the lipid binding hypothesis for SPLUNC1. It was hypothesized that upon lipid ligand binding, a major conformational change was needed to open-up the hydrophobic channel of SPLUNC1. Since SPLUNC1 has the closest structure to BSP30b among existing solved structures, we suspect that it has

a similar lipid binding mechanism. Possibly, upon lipid binding of SPLUNC1, its helix α4 will move aside but no major conformational change will be needed. After helix α4 moves away, its hydrophobic channel will be exposed to its lipid ligands and ready to accommodate them. This also could be the case for latherin, a PLUNC protein from horse sweat and saliva. It has similar structure to SPLUNC1 and its C-terminal helix could move away upon lipid binding.

We believe this rigid structure of BSP30b constrains its lipid ligand profile. Although BSP30b was crystallized with oleic acid, it is not the preferred lipid ligand indicated by its poor electron density profile. We have shown that fatty acids with two or three double bonds on their carbon back bones weren't co-crystallized with BSP30b. This is presumably because they have more rigid structures compared with those lipids with no or just one double bond on their carbon back bones. All of these may suggest that the lipid ligands of BSP30b should have no more than one carbon-carbon double bond.

Moreover, its hydrophobic channel can only accommodate one carbon chain. This is similar to JHBP and Takeout proteins, whose substrate has only one carbon chain (Hamiaux et al., 2013; Suzuki et al., 2011). In contrast, the substrates of BPI or CETP have two carbon chains and both of these are accommodated in their hydrophobic channel (Kleiger et al., 2000; Qiu et al., 2007). This may suggest that BSP30b has similar functions to JHBP and Takeout proteins but different to BPI or CETP.

5.7.2 Function Studies Using GFP Fused BSP30b

Green Fluorescent protein (GFP) was discovered in Aequorea jellyfish, which is a companion protein to the famous chemiluminescent protein – aequorin (Shimomura, Johnson, & Saiga, 1962). For cell biological application, GFP was first used to detect gene expression *in vivo*. In the nematode Caenorhabditis elegans, GFP successfully confirmed the expression pattern of the mec-7 promotor, which drives the formation of β-tubulin in a limited number of mechanosensory neurons (Chalfie, Tu, Euskirchen, Ward, & Prasher, 1994). GFP is also particularly promising in intact transgenic embryos and animals as well as for monitoring the effectiveness of gene transfer because of its independence from enzymatic substrates (Ikawa et al., 1995; Levy, Muldoon, Zolotukhin, & Link, 1996). Nevertheless, GFP needs strong promoters to drive sufficient expression for detection, especially in mammalian cells (Tsien, 1998).

The most successful application of GFP has been as a genetic fusion partner to host proteins, which helps to monitor their location and fate (Tsien, 1998). The fused protein will be expressed in the cell or organism of interest, which ideally maintains the normal functions and localizations of the host protein but is now fluorescent. Practically, all of the major organelles of the cell have been targeted by GFP, including plasma membrane, nucleus, endoplasmic reticulum, Golgi apparatus and *et. al* (Tsien, 1998).

In this study, BSP30b was fused to GFP to visualise its possible functions. Instead of monitoring the location of GFP fused protein *in vivo*, we investigated the location of the fused protein within emulsions after mixing with olive oil.

5.7.3 BSP30b is A Bio-emulsifier

Emulsion is a fine dispersion of a droplet phase stabilized by a thin film, usually a monomolecular film of an emulsifier. There are basically two different types of emulsions: oil in water (o/w emulsion) or water in oil (w/o emulsion). The emulsion system is thermodynamically unstable, which tends to separate into an oil and an aqueous phase with time (Nursakinah, Ismail, Rahimi, & Idris, 2013).

Many proteins are surface active molecules, which form monolayers on aqueous solutions or the interlayer between oil and water (McClements, 2004). They also have the property of forming clusters and micelles in the bulk. Those proteins can be used as emulsifiers based on their ability to facilitate this formation, improve the stability and produce desirable physicochemical properties in oil-in-water emulsions. Proteins will adsorb to the surface of oil droplets created by homogenization of oil-water-protein mixtures, where proteins will facilitate further droplet disruption by lowering the interfacial tension and prevent droplet coalescence by forming a protective membrane around the droplets (McClements, 2004).

Proteins in the PLUNC family are proven to have potent surfactant properties at water/liquid interface (J. A. Bartlett et al., 2011; McDonald et al., 2009). Latherin from the sweat and saliva of horses is believed to be a highly surface-active allergen protein. It probably functions as a wetting agent in evaporative cooling in horses and may also assist in mastication of fibrous food as well as inhibition of microbial biofilms. The surfactant activity of SPLUNC1 from human has also been investigated by using a variety of *in vitro* assays. Contact angle studies showed that SPLUNC1-containing solution significantly reduced the contact angles compared

to drops containing buffer alone or control proteins. It was also found that recombinant SPLUNC1 reduced surface tension in a dose-dependent manner by using a pulsating bubble surfactometer.

Here, we demonstrated that BSP30b behaves like an emulsifier, dispersing olive oil into small droplets in aqueous solution. We also suggest that its emulsification ability comes from its binding to the fatty acid chains of glycerides, with small spheres forming in the solution. The internal core of each sphere is full of lipids while the surface is coated by BSP30b, the polar side chains of whose amino acids are exposed to the solution, keeping the micelles soluble.

5.7.4 BSP30b Selectively Binds to Certain Rumen Bacterial Species

The most interesting finding in this study is that BSP30b only binds to a small portion of bacteria species within rumen. "Bacteria coating" by proteins from the PLUNC group has been reported before (Robinson et al., 1997; Sayeed, Nistico, St Croix, & Di, 2013b). The binding of murine PSP to *E. coli*, *S. mutans*, *A. actinomycetemcomitans*, and *L. monocytogenes* has been demonstrated to be protein-protein interactions dependent on the presence of Zn^{2+}. However, we suggest that BSP30b binds to the lipids on the bacterial surface. The "bacterial cell coating" by human SPLUNC1 of *P. aeruginosa*-pMF230 has been investigated using confocal microscopy, and the protein is proposed to display a bacteriostatic property in preventing bacterial growth. But other studies also report that this anti-biofilm formation of human SPLUNC1 could due to its surfactant property (Gakhar

et al., 2010b). Here, we propose that BSP30b could have important roles in the rumen via its binding specificity to certain ruminal bacteria species.

Chapter Six. Conclusion

6.1 General Background

BSP30b from cattle is an acidic protein (pI = 4.4) with a molecular weight of ~30 kDa, which is proposed to have arisen through a series of gene duplications (Haigh et al., 2008). BSP30b was first identified from SDS PAGE analysis of bovine salivary proteins. Its abundance in saliva was related to the bloat susceptibility of cattle. Although BSP30b was suggested to be a BPI-like protein, it failed to bind LPS and only had moderate antimicrobial activity against two *P. aeruginosa* strains (Haigh et al., 2008).

Despite the high expression of BSP30b in the saliva of cattle, its function is still unclear. One difficulty is that the functions of most PLUNC proteins are largely unknown. Although PLUNC proteins from human have been intensively investigated, especially hSPLUNC1, their functions are still inconclusive (Ning et al., 2014; Sayeed et al., 2013a; Zhou et al., 2008). Another difficulty is that until now, the structures of only two PLUNC proteins have been solved and both without ligands. All of these difficulties impede our understanding of the biological functions of PLUNC proteins.

This project was undertaken to investigate the structural characteristics and functions of the cattle PLUNC protein BSP30b.

6.2 BSP30b is a TULIP protein

Although BSP30b has been demonstrated to be a TULIP protein from sequence analysis, our structural result further confirms this. Each molecule of BSP30b adopts a slightly curved cylindrical structure with a hydrophobic channel formed by an α/β wrap, which is highly conserved in the TULIP superfamily. Moreover, BSP30b also has the conserved disulfide bond seen across the PLUNC family, which connects the α helix and β sheet region and stabilizes the whole structure.

Compared to other TULIP superfamily proteins, BSP30b also has its own unique features. Firstly, its N terminal consists of four helices (α1-α4) while there are only two helices (α1 and α2) present in SPLUNC1/Latherin. The respective N-terminus and C-terminus of BSP30b are located on the opposite ends but they are on the same side for SPLUNC1/Latherin. .Moreover, its internal cavity opening is covered by a flexible loop. Compared to the solved structures of hSPLUNC1 and Latherin, BSP30b has an enlarged internal cavity. These features enable the binding of ligand without any major conformational changes. However, there are still some local adjustments of side chains around the bound ligand. An extra disulfide bond has also been observed in the structure of BSP30b. The electron density is clear in one molecule of its ASU but is missing in another molecule. Although the electron densities are completely missing in the ligand bound structure, it is still not conclusive regarding the role of the second disulphide bond.

We co-crystallised BSP30b with oleic acid bound in its hydrophobic channel. This is the first time that a ligand has been seen in the structure of a PLUNC group protein. Despite oleic acid not being the ideal ligand for BSP30b (as previously mentioned), we believe that its biological ligand should be closely related to oleic

acid, having a long hydrophobic carbon chain with reasonable flexibility. Since all the TULIP proteins have been proposed to be lipid binding proteins, the ligand of BSP30b should have these properties.

6.3 BSP30b is a Surfactant Protein

In this study, BSP30b has been shown to have surfactant properties. This is consistent with the studies on SPLUNC1 and latherin. The surfactant activity of SPLUNC1 from human has been investigated by using a variety of *in vitro* assays. Latherin from horse saliva and sweat has been proposed to be first localized to the non-polar interface via its flexible loops at one end of the protein molecule, which then unfolds and flattens to expose its hydrophobic interior to the air or non-polar surface.

However, our finding suggests that BSP30b fulfils its surfactant property without unfolding or flattening to expose its hydrophobic interior. From our emulsification study, we propose that the fatty acid moiety of triglyceride will enter the hydrophobic pocket forming a protein layer around the lipid droplet. These results could give us some indication as to the functions of BSP30b pertaining to lipolysis or biohydrogenation within rumen. Its interaction with lipids could facilitate the lipolysis or biohydrogenation.

6.4 The interaction between BSP30b and rumen bacteria

The rumen is a complex ecosystem, in which nutrients consumed by ruminant are fermented by microorganisms living in the rumen. The diet influence on rumen fermentation has been well studied. However, the contribution of ruminant salivary proteins to the fermentation is not well known.

Here, we investigated the potential roles of salivary protein BSP30b within rumen. We used a GFP tag to indicate the location of BSP30b within the rumen samples. Our results showed that BSP30b specifically binds to certain rumen bacterial species. This is the first time to look at the interaction between ruminant salivary protein and rumen bacteria. It would be interesting to determine which rumen bacteria species are targeted by BSP30b using FACS (fluorescence-activated cell sorting) and 16S RNA gene sequencing from these bacteria. Furthermore, this binding may have crucial effects on rumen microbial metabolism. This hypothesis can be addressed by culturing the microbe in isolation (if possible) and using RNAseq methods to analyse genes that are up- or down- regulated in response to BSP30b protein bidning. It will be also interesting to use metabolomics at the level of rumen contents with and without BSP30b protein to identify community metabolic differences in the rumen. This could provide us with insights into the biological role(s) of the BSP30 proteins in the rumen at the scale of individual bacteria and at the community scale.

6.5 Future Research

Research presented in this thesis provides the structure of BSP30b and its ligand binding. It would be of great interest to further determine the ligand of BSP30b by exploring BSP30b bound bacterial surface lipids. This could include using methods such as co-crystallization of the protein and the ligand under investigation for binding. This can further our understanding on the binding mechanism of BSP30b to specific rumen bacteria.

In this study, we have described the morphologies of the rumen bacteria bound to BSP30b, further characterization of the bound bacteria can be undertaken by using FACS (fluorescence-activated cell sorting) and NGS (next generation sequencing). The bacteria bound by GFP_BSP30b can be separated from un-bound bacteria by FACS, which can purify the GFP_BSP30b bound bacteria based upon GFP fluorescence. The genus of these enriched bacteria can be then determined by NGS targeting the 16s rRNA gene region. Once the bacteria species are identified, the potential regulatory roles of BSP30b on the associated metabolic and fermentation pathways can be analysed using a combination of specialised culture methods, metabolomics and RNA sequencing.

It may be the case that the main role for BSP30b is to simply inhibit growth of particular species within rumen. However, it is much more likely that BSP30b is regulating metabolism in some way. To prove this, metabolomics approaches can be undertaken on microbial cultures (with and without BSP30b protein).

Overall, undertanding more fully the binding characteristics of BSP30b will have broader implications for understanding nutrient availability and rumen metabolism within ruminants. The predicted regulation roles of BSP30b can potentially offer improvements in productivity in the dairy and beef industries through identifying new ways to optimize nutrient digestion and utilization.

Appendices

A1: Buffers and solutions

Electrophoresis buffers were made with distilled water whereas chromatography buffers were made with MQ water and were then filtered with a 0.2 μm filter prior to use. Loading dye was made using MQ water.

10 x DNA loading dye	0.4% (w/v) bromophenol blue, 0.4% (w/v) xylene cyanol, 50% (v/v) glycerol
Fairbanks A stain	0.05% (w/v) coomassie blue R-250, 25% (v/v) isopropanol, 10% (v/v) acetic acid
Fairbanks B stain	0.005% (w/v) coomassie blue R-250, 10% (v/v) isopropanol, 10% (v/v) acetic acid
Fairbanks C stain	0.002% (w/v) coomassie blue, 10% (v/v) acetic acid
Fairbanks D stain	10% (v/v) acetic acid
5 x native loading dye	300 mM Tris-HCl pH 6.8, 50% (v/v) glycerol and 0.05% (w/v) bromophenol blue
4 x SDS loading dye	200 mM Tris-HCl pH 6.8, 8% (w/v) SDS, 40% (v/v) glycerol, 0.4% (w/v) bromophenol blue, 400 mM β-mercaptoethanol
50 x TAE	242 g Tris base, 57.1 ml glacial acetic acid, 100 ml 0.5 M EDTA (pH 8.0)

10 x TAE	400 mM Tris-acetate, 10 mM EDTA
1 x TAE	100 ml 10 x TAE + 900 ml H2O
TE	10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0

A2: Growth media

Dehydrated Difco™, BDL™ and Bacto™ media and ADC enrichment media was supplied by Becton, Dickinson and Company, USA. Growth media was made up in distilled water and sterilised by autoclaving at 121°C for 15 minutes.

LB	1% (w/v) bactotryptone or bactopeptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, pH 8.0, autoclave
LB agar	1% (w/v) bactotryptone or bactopeptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, 15 g/L agar, pH 8.0, autoclave
M9 salts	42 mM Na2HPO4, 22 mM KH2PO4, 19 mM NH4Cl, 8.5 mM NaCl, pH to 7.4 with 10 M NaOH, autoclave
M9 minimal	42 mM Na2HPO4, 22 mM KH2PO4, 19 mM NH4Cl, 8.5 mM NaCl, pH to 7.4 with 10 M NaOH, autoclave, then add sterile 2 mM MgSO4, 0.1 mM CaCl2 and 0.4% (w/v) glucose
PA-0.5G	50 mM Na2HPO4, 50 mM KH2PO4, 25 mM (NH4)2SO4, 1 mM MgSO4, 0.5% (w/v) glucose, 0.1 x metals mix***, 100 μg/ml each of 17 amino acids (no Cys, Tyr or Met). For

	methionine auxotroph DL41 add 100 μg/ml methionine. Individual components autoclaved or sterile filtered before adding to sterile dH2O
PASM-5052	50 mM Na2HPO4, 50 mM KH2PO4, 25 mM (NH4)2SO4, 1 mM MgSO4, 0.5% (w/v) glycerol, 0.05% (w/v) glucose, 0.2% (w/v) α-lactose, 1 x metals mix***, 100 nM vitamin B12, 200 μg/ml each of 17 amino acids (no Cys, Tyr or Met), 10 μg/ml methionine, 125 μg/ml selenomethionine
Rich media + glucose	1% (w/v) bactotryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 0.2% (w/v) glucose
SOC	2% (w/v) bactotryptone or bactopeptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgSO4, 10 mM MgCl2, autoclave. Once cooled add sterile glucose to 20 mM

*100 x trace metals stock for HdB media: EDTA [0.1 g], MgCl2.6H2O [1 g], CaCl2.2H2O [10 mg], NaMoO4.2H2O [2 mg], CoCl2.6H2O [4 mg], MnCl2.2H2O [10 mg], ZnSO4.7H2O [20 mg], FeSO4.7H2O [50 mg], CuSO4.5H2O to 100 ml with dH2O.

**100 x phosphate stock for HdB media: K2HPO4 [15.5g], NaH2PO4.2H2O [11.1g] to 100 ml with dH2O.

***1000 x metals mix made up from the sterile stock solutions of each component to give the following concentrations: 50 μM FeCl3 in 0.12 M HCl (filter sterile), 20

µM CaCl2, 10 µM MnCl2, 10 µM ZnSO4, 2 µM CoCl2, 2 µM CuCl2, 2 µM NiCl2, 2 µM Na2MoO4, 2 µM Na2SeO3, 2 µM H3BO3.

A3: Primers used in this study

Table A.1 Primers used in this study

primer	Sequence	Tm
BSP30b_DM_F1	CTGGTTCTGCTGTGTGGTCT	64.4 °C
pProEx_Rev_BSP30	ACTTCTGAGTTCGGCATGGG	64.5 °C
F	GGATCCAGCCTGCCGGATAT	62.5 °C
R	ACTTCTGAGTTCGGCATGGG	60.4 °C
Bsp30b_F	GGTGGTGGATCCAGCCTGCCGGATATT CGTGGT	69.5
BSP30b_R	GGTGGTCTGCAGCTTGTATAATTCATC CAT	60.3
GFP_forward	GGATCCGGCTCCAAAGGCGAGGAA	68.6
GFP_reverse	CTGCAGCTTGTATAATTCATCCAT	57.4

A4: Protein Information

BSP30b_F52M_F106M

MVQLWKLVLLCGLLAGTSASLPDIRGNDVLRKLKSGLERGLDTFDSTIEII
MQNLKTELESRCSDEVVEQQETENFLEQLISRIFQVVSRLTGVRIRNVQVP
DITMEATSENSANVLIPITADVTVSLPFLGEIVDLDLNVDLQTTVSIETDTE
DPQVVVGECTNNPESISLTVLHSRFGLVNDVVDIGVNLARRVVSSVVEGE
LCPRFRELLESLDAECVEKLIGESQDTTQQEPEGSR

MBP-BSP30b_F52M_F106M

MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQ
VAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVR
YNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMF
NLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDL
IKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVL
PTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDK
PLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTA
VINAASGRQTVDEALKDAQTNSSSNNNNNNNNNNLGIEGRISEFGSSLPDI
RGNDVLRKLKSGLERGLDTFDSTIEIIMQNLKTELESRCSDEVVEQQETEN
FLEQLISRIFQVVSRLTGVRIRNVQVPDITMEATSENSANVLIPITADVTVSL
PFLGEIVDLDLNVDLQTTVSIETDTEDPQVVVGECTNNPESISLTVLHSRFG
LVNDVVDIGVNLARRVVSSVVEGELCPRFRELLESLDAECVEKLIGESQDT
TQQEPEGSR

GFP_BSP30b

MSYYHHHHHHDYDIPTTENLYFQGAMGSGSKGEELFTGVVPILVELDGD

VNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCF

SRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLV

NRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNV

EDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVL

LEFVTAAGITHGMDELYKLQSRGSLPDIRGNDVLRKLKSGLERGLDTFDS

TIEIIMQNLKTELESRCSDEVVEQQETENFLEQLISRIFQVVSRLTGVRIRNV

QVPDITMEATSENSANVLIPITADVTVSLPFLGEIVDLDLNVDLQTTVSIET

DTEDPQVVVGECTNNPESISLTVLHSRFGLVNDVVDIGVNLARRVVSSVV

EGELCPRFRELLESLDAECVEKLIGESQDTTQQEPEGSR

A5: Robotic Screen Conditions with Crystal "Hits" for BSP30b

Table A.1 Robotic screen conditions with crystal "hits" for BSP30b

Index™ – HR2-144	82. 0.2 M Magnesium chloride hexahydrate 0.1 M BIS-TRIS pH 5.5 25% w/v Polyethylene glycol 3,350
	84. 0.2 M Magnesium chloride hexahydrate 0.1 M HEPES pH 7.5 25% w/v Polyethylene glycol 3,350
	85. 0.2 M Magnesium chloride hexahydrate 0.1 Tris pH 8.5 25% w/v Polyethylene glycol 3,350
PEGRx HT™ – HR2-086	41. 0.1 M Sodium citrate tribasic dehydrate pH 5.5 16% w/v Polyethylene glycol 8,000
Crystal Screen HT™ – HR2-130	14. 0.2 M Calcium chloride dehydrate 0.1 M HEPES pH 7.5 30% v/v Polyethylene glycol 400
	23. 0.2 M Magnesium chloride hexahydrate 0.1 M HEPES pH 7.5 30% v/v Polyethylene glycol 400
	82. 0.05 M Cadmium sulfate hydrate 0.1 M HEPES pH 7.5 1.0 M Sodium acetate trihydrate

A6: Bacterial strains used in this study

Table A.2 Bacterial strains used in this study

Strain	Description
DH5α	High efficiency transformation and high plasmid copy number. Inactivated intracellular endonucleases for high plasmid DNA yield.
BL21 (DE3)	DE3 lysogen contains T7 polymerase upon IPTG induction. This strain is deficient of lon and omp-t proteases and is therefore suitable for expression of non-toxic genes.

A7: Plasmids used in this study

Table A.4 Plasmids used in this study

Plasmid	Description
pPROExHTb	E. coli expression vector. Trc promoter. Encodes an N-terminal 6 x histidine tag with an rTEV protease cleavage site. Ampicillin resistance (Invitrogn, USA).
pMAL-c2x	E. coli expression vector. Tac promoter. Encodes malE for expression of an N-terminal MBP fusion tag with a Factor Xa protease cleavage site. Ampicillin resistance (NEB, USA).
pMA-T	E. coli expression vector. Carrier of synthetic genes. Ampicillin resistance (GeneArt Life Technologies, USA)
pTriEx-4	Uniquely designed to allow rapid characterization of target genes in multiple expression systems.

	Expression in E. coli is regulated by T7 promoter. Ampicillin resistance (EMD Millipore, USA)

A8: Plasmid constructs used in this study.

Table A5 Plasmid constructs used in this study

Plasmid	Description	Reference
BSP30b-pTriEx-4	pTriEx-4 with BSP30b inserted into LIC site.	(Haigh et al., 2008)
BSP30b_F52M_F106M-pMA-T	pMA-T with synthesized BSP30b_F52M_F106M inserted into BamHI/XhoI restriction sites.	This study
BSP30b_F52M_F106M-pProEx Htb	pProEx Htb with BSP30b_F52M_F106M inserted into BamHI/XhoI restriction sites.	This study
BSP30b_F52M_F106M-pMAL-c2X	pMAL-c2X with BSP30b_F52M_F106M inserted into BamHI/PstI restriction sites.	This study
GFP-BSP30b-pProEx Htb	pProEx Htb with GFP and BSP30b inserted. GFP was cloned into BamHI/PstI restriction sites. While BSP30b was cloned into XhoI/hindIII sites.	This study

References

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Echols, N., Headd, J. J., . . . Zwart, P. H. (2011). The Phenix software for automated determination of macromolecular structures. *Methods, 55*(1), 94-106. doi: 10.1016/j.ymeth.2011.07.005

Afonine, P. V., Grosse-Kunstleve, R. W., & Adams, P. D. (2005). A robust bulk-solvent correction and anisotropic scaling procedure. *Acta Crystallographica Section D-Biological Crystallography, 61*, 850-855. doi: 10.1107/S0907444905007894

Alva, V., & Lupas, A. N. (2016). The TULIP superfamily of eukaryotic lipid-binding proteins as a mediator of lipid sensing and transport. *Biochim Biophys Acta, 1861*(8 Pt B), 913-923. doi: 10.1016/j.bbalip.2016.01.016

Barter, P. J., Brewer, H. B., Jr., Chapman, M. J., Hennekens, C. H., Rader, D. J., & Tall, A. R. (2003). Cholesteryl ester transfer protein: a novel target for raising HDL and inhibiting atherosclerosis. *Arterioscler Thromb Vasc Biol, 23*(2), 160-167.

Bartlett, J. A., Gakhar, L., Penterman, J., Singh, P. K., Mallampalli, R. K., Porter, E., & McCray, P. B., Jr. (2011). PLUNC: a multifunctional surfactant of the airways. *Biochem Soc Trans, 39*(4), 1012-1016. doi: 10.1042/BST0391012

Bartlett, J., Gakhar, L., Penterman, J., Singh, P., Mallampalli, R. K., Porter, E., & McCray, P. B. (2011). PLUNC: a multifunctional surfactant of the airways. *Biochemical Society Transactions, 39*, 1012-1016. doi: 10.1042/Bst0391012

Beamer, L. J., Carroll, S. F., & Eisenberg, D. (1999). The three-dimensional structure of human bactericidal/permeability-increasing protein: implications for understanding protein-lipopolysaccharide interactions. *Biochem Pharmacol, 57*(3), 225-229.

Bergfors, T. (2003). Seeds to crystals. *J Struct Biol, 142*(1), 66-76.

Cabras, T., Manconi, B., Iavarone, F., Fanali, C., Nemolato, S., Fiorita, A., . . . Castagnola, M. (2012). RP-HPLC-ESI-MS evidenced that salivary cystatin B is detectable in adult human whole saliva mostly as S-modified derivatives: S-Glutathionyl, S-cysteinyl and S-S 2-mer. *J Proteomics, 75*(3), 908-913. doi: 10.1016/j.jprot.2011.10.006

Campos, M. A., Abreu, A. R., Nlend, M. C., Cobas, M. A., Conner, G. E., & Whitney, P. L. (2004). Purification and characterization of PLUNC from human - Tracheobronchial secretions. *American Journal of Respiratory Cell and Molecular Biology, 30*(2), 184-192. doi: 10.1165/rcmb.2003-01420C

Canny, G., & Levy, O. (2008). Bactericidal/permeability-increasing protein (BPI) and BPI homologs at mucosal sites. *Trends Immunol, 29*(11), 541-547. doi: 10.1016/j.it.2008.07.012

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., & Prasher, D. C. (1994). Green Fluorescent Protein as a Marker for Gene-Expression. *Science, 263*(5148), 802-805. doi: DOI 10.1126/science.8303295

Chandler, J. D., & Day, B. J. (2012). Thiocyanate: a potentially useful therapeutic agent with host defense and antioxidant properties. *Biochem Pharmacol, 84*(11), 1381-1387. doi: 10.1016/j.bcp.2012.07.029

Clarke, R. T., & Reid, C. S. (1974). Foamy bloat of cattle. A review. *J Dairy Sci, 57*(7), 753-785. doi: 10.3168/jds.S0022-0302(74)84964-7

Collaborative Computational Project, Number. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr, 50*(Pt 5), 760-763. doi: 10.1107/S0907444994003112

de Sousa-Pereira, P., Cova, M., Abrantes, J., Ferreira, R., Trindade, F., Barros, A., . . . Vitorino, R. (2015). Cross-species comparison of mammalian saliva using an LC-MALDI based proteomic approach. *Proteomics, 15*(9), 1598-1607. doi: 10.1002/pmic.201400083

Di, Y. P. (2011). Functional roles of SPLUNC1 in the innate immune response against Gram-negative bacteria. *Biochemical Society Transactions, 39*, 1051-1055. doi: 10.1042/Bst0391051

Dong, A., Xu, X., Edwards, A. M., Midwest Center for Structural, Genomics, Structural Genomics, Consortium, Chang, C., . . . Zhu, H. (2007). In situ proteolysis for protein crystallization and structure determination. *Nat Methods, 4*(12), 1019-1021. doi: 10.1038/nmeth1118

Dyce, K.M.; Sack, W.O.; Wensing, C.J.G. (1996). *Veterinary anatomy*. Philadelphia: WB Saunders.

Eckert, J. K., Kim, Y. J., Kim, J. I., Gurtler, K., Oh, D. Y., Sur, S., . . . Schumann, R. R. (2013). The crystal structure of lipopolysaccharide binding protein reveals the location of a frequent mutation that impairs innate immunity. *Immunity, 39*(4), 647-660. doi: 10.1016/j.immuni.2013.09.005

Elewa, Y. H., Bareedy, M. H., Abuel-Atta, A. A., Ichii, O., Otsuka, S., Kanazawa, T., . . . Kon, Y. (2010). Structural characteristics of goat (Capra hircus) parotid salivary glands. *Jpn J Vet Res, 58*(2), 121-135.

Elewa, Y. H., Ichii, O., Otsuka, S., Hashimoto, Y., & Kon, Y. (2014). Structural changes of goat parotid salivary gland: pre- and post-weaning periods. *Anat Histol Embryol, 43*(4), 265-272. doi: 10.1111/ahe.12071

Elsbach, P., & Weiss, J. (1995). Prospects for use of recombinant BPI in the treatment of gram-negative bacterial infections. *Infect Agents Dis, 4*(2), 102-109.

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography, 60*, 2126-2132. doi: 10.1107/S0907444904019158

Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallographica Section D-Biological Crystallography, 69*, 1204-1214. doi: 10.1107/S0907444913000061

Gakhar, L., Bartlett, J. A., Penterman, J., Mizrachi, D., Singh, P. K., Mallampalli, R. K., . . . McCray, P. B. (2010a). PLUNC Is a Novel Airway Surfactant Protein with Anti-Biofilm Activity. *Plos One, 5*(2). doi: ARTN e909810.1371/journal.pone.0009098

Gakhar, L., Bartlett, J. A., Penterman, J., Mizrachi, D., Singh, P. K., Mallampalli, R. K., . . . McCray, P. B., Jr. (2010b). PLUNC is a novel airway surfactant protein with anti-biofilm activity. *PLoS One, 5*(2), e9098. doi: 10.1371/journal.pone.0009098

Garland, A. L., Walton, W. G., Coakley, R. D., Tan, C. D., Gilmore, R. C., Hobbs, C. A., . . . Tarran, R. (2013). Molecular basis for pH-dependent mucosal dehydration in cystic fibrosis airways. *Proceedings of the National Academy of Sciences of the United States of America, 110*(40), 15973-15978. doi: 10.1073/pnas.1311999110

Ghafouri, B., Kihlstrom, E., Tagesson, C., & Lindahl, M. (2004). PLUNC in human nasal lavage fluid: multiple isoforms that bind to lipopolysaccharide. *Biochimica Et Biophysica Acta-Proteins and Proteomics, 1699*(1-2), 57-63. doi: 10.1016/j.bbapap.2004.01.001

Gilmour, K. M. (2010). Perspectives on carbonic anhydrase. *Comp Biochem Physiol A Mol Integr Physiol, 157*(3), 193-197. doi: 10.1016/j.cbpa.2010.06.161

Giordano, F., Saheki, Y., Idevall-Hagren, O., Colombo, S. F., Pirruccello, M., Milosevic, I., . . . De Camilli, P. (2013). PI(4,5)P(2)-dependent and Ca(2+)-regulated ER-PM interactions mediated by the extended synaptotagmins. *Cell, 153*(7), 1494-1509. doi: 10.1016/j.cell.2013.05.026

Haigh, B., Hood, K., Broadhurst, M., Medele, S., Callaghan, M., Smolenski, G., . . . Wheeler, T. (2008). The bovine salivary proteins BSP30a and BSP30b are independently expressed BPI-like proteins with anti-Pseudomonas activity. *Molecular Immunology, 45*(7), 1944-1951. doi: 10.1016/j.molimm.2007.10.032

Hailman, E., Lichenstein, H. S., Wurfel, M. M., Miller, D. S., Johnson, D. A., Kelley, M., . . . Wright, S. D. (1994). Lipopolysaccharide (LPS)-binding protein accelerates the binding of LPS to CD14. *J Exp Med, 179*(1), 269-277.

Hamiaux, C., Basten, L., Greenwood, D. R., Baker, E. N., & Newcomb, R. D. (2013). Ligand promiscuity within the internal cavity of Epiphyas postvittana Takeout 1 protein. *J Struct Biol, 182*(3), 259-263. doi: 10.1016/j.jsb.2013.03.013

Hamiaux, C., Stanley, D., Greenwood, D. R., Baker, E. N., & Newcomb, R. D. (2009). Crystal structure of Epiphyas postvittana takeout 1 with bound ubiquinone supports a role as ligand carriers for takeout proteins in insects. *J Biol Chem, 284*(6), 3496-3503. doi: 10.1074/jbc.M807467200

Huuskonen, J., Olkkonen, V. M., Jauhiainen, M., & Ehnholm, C. (2001). The impact of phospholipid transfer protein (PLTP) on HDL metabolism. *Atherosclerosis, 155*(2), 269-281.

Ihalin, R., Loimaranta, V., & Tenovuo, J. (2006). Origin, structure, and biological activities of peroxidases in human saliva. *Arch Biochem Biophys, 445*(2), 261-268. doi: 10.1016/j.abb.2005.07.004

Ikawa, M., Kominami, K., Yoshimura, Y., Tanaka, K., Nishimune, Y., & Okabe, M. (1995). A rapid and non-invasive selection of transgenic embryos before implantation using green fluorescent protein (GFP). *FEBS Lett, 375*(1-2), 125-128.

Jeong, H., Park, J., & Lee, C. (2016). Crystal structure of Mdm12 reveals the architecture and dynamic organization of the ERMES complex. *EMBO Rep, 17*(12), 1857-1871. doi: 10.15252/embr.201642706

Johansson, I., & Lif Holgerson, P. (2011). Milk and oral health. *Nestle Nutr Workshop Ser Pediatr Program, 67*, 55-66. doi: 10.1159/000325575

Kay, R. N. (1966). The influence of saliva on digestion in ruminants. *World Rev Nutr Diet, 6*, 292-325.

Kivela, J., Parkkila, S., Parkkila, A. K., Leinonen, J., & Rajaniemi, H. (1999). Salivary carbonic anhydrase isoenzyme VI. *J Physiol, 520 Pt 2*, 315-320.

Kleiger, G., Beamer, L. J., Grothe, R., Mallick, P., & Eisenberg, D. (2000). The 1.7 A crystal structure of BPI: a study of how two dissimilar amino acid

sequences can adopt the same fold. *J Mol Biol, 299*(4), 1019-1034. doi: 10.1006/jmbi.2000.3805

Kopec, K. O., Alva, V., & Lupas, A. N. (2010). Homology of SMP domains to the TULIP superfamily of lipid-binding proteins provides a structural basis for lipid exchange between ER and mitochondria. *Bioinformatics, 26*(16), 1927-1931. doi: 10.1093/bioinformatics/btq326

Kopec, K. O., Alva, V., & Lupas, A. N. (2011). Bioinformatics of the TULIP domain superfamily. *Biochem Soc Trans, 39*(4), 1033-1038. doi: 10.1042/BST0391033

Kornmann, B., Currie, E., Collins, S. R., Schuldiner, M., Nunnari, J., Weissman, J. S., & Walter, P. (2009). An ER-mitochondria tethering complex revealed by a synthetic biology screen. *Science, 325*(5939), 477-481. doi: 10.1126/science.1175088

Kramer, K. J., Sanburg, L. L., Kezdy, F. J., & Law, J. H. (1974). The Juvenile Hormone Binding Protein in the Hemolymph of Manduca sexta Johannson (Lepidoptera: Sphingidae). *Proc Natl Acad Sci U S A, 71*(2), 493-497.

Lamy, E., & Mau, M. (2012). Saliva proteomics as an emerging, non-invasive tool to study livestock physiology, nutrition and diseases. *Journal of Proteomics, 75*(14), 4251-4258. doi: 10.1016/j.jprot.2012.05.007

LeClair, E. E. (2003). Four BPI (bactericidal/permeability-increasing protein)-like genes expressed in the mouse nasal, oral, airway and digestive epithelia. *Biochemical Society Transactions, 31*, 801-805. doi: Doi 10.1042/Bst0310801

Leslie, A. G. W., & Powell, H. R. (2007). Processing diffraction data with MOSFLM. *Evolving Methods for Macromolecular Crystallography*, *245*, 41-+.

Levy, J. P., Muldoon, R. R., Zolotukhin, S., & Link, C. J., Jr. (1996). Retroviral transfer and expression of a humanized, red-shifted green fluorescent protein gene into human tumor cells. *Nat Biotechnol, 14*(5), 610-614. doi: 10.1038/nbt0596-610

Maeng, Y. J., Kim, B. R., Jung, H. I., Jung, U. W., Kim, H. E., & Kim, B. I. (2016). Diagnostic accuracy of a combination of salivary hemoglobin levels, self-report questionnaires, and age in periodontitis screening. *J Periodontal Implant Sci, 46*(1), 10-21. doi: 10.5051/jpis.2016.46.1.10

Manford, A. G., Stefan, C. J., Yuan, H. L., Macgurn, J. A., & Emr, S. D. (2012). ER-to-plasma membrane tethering proteins regulate cell signaling and ER morphology. *Dev Cell, 23*(6), 1129-1140. doi: 10.1016/j.devcel.2012.11.004

Marti, G. A., Friedman, S. M., Cabrini, R. L., & Costa, O. R. (2002). Role of crevice in the occurrence of hemoglobin in saliva. *Acta Odontol Latinoam, 15*(1-2), 11-13.

McClements, D. J. (2004). Protein-stabilized emulsions. *Current Opinion in Colloid & Interface Science, 9*(5), 305-313. doi: 10.1016/j.cocis.2004.09.003

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 Crystallogr, 40*(Pt 4), 658-674. doi: 10.1107/S0021889807021206

McDonald, R. E., Fleming, R. I., Beeley, J. G., Bovell, D. L., Lu, J. R., Zhao, X., . . . Kennedy, M. W. (2009). Latherin: a surfactant protein of horse sweat and saliva. *PLoS One, 4*(5), e5726. doi: 10.1371/journal.pone.0005726

McDougall, E. I. (1948). Studies on ruminant saliva. 1. The composition and output of sheep's saliva. *Biochem J, 43*(1), 99-109.

McGillivary, G., & Bakaletz, L. O. (2010). The Multifunctional Host Defense Peptide SPLUNC1 Is Critical for Homeostasis of the Mammalian Upper Airway. *Plos One, 5*(10). doi: ARTN e13224

10.1371/journal.pone.0013224

McLaughlin, W. S., Kirkham, J., Kowolik, M. J., & Robinson, C. (1996). Human gingival crevicular fluid keratin at healthy, chronic gingivitis and chronic adult periodontitis sites. *J Clin Periodontol, 23*(4), 331-335.

Mueller, G. A., Edwards, L. L., Aloor, J. J., Fessler, M. B., Glesner, J., Pomes, A., . . . Pedersen, L. C. (2010). The structure of the dust mite allergen Der p 7 reveals similarities to innate immune proteins. *J Allergy Clin Immunol, 125*(4), 909-917 e904. doi: 10.1016/j.jaci.2009.12.016

Ning, F. K., Wang, C., Berry, K. Z., Kandasamy, P., Liu, H. L., Murphy, R. C., . . . Zhang, G. Y. (2014). Structural characterization of the pulmonary innate immune protein SPLUNC1 and identification of lipid ligands. *Faseb Journal, 28*(12), 5349-5360. doi: 10.1096/fj.14-259291

Nistor, A., Bowden, G., Blanchard, A., & Myal, Y. (2009). Influence of mouse prolactin-inducible protein in saliva on the aggregation of oral bacteria. *Oral Microbiol Immunol, 24*(6), 510-513. doi: 10.1111/j.1399-302X.2009.00543.x

Nursakinah, I., Ismail, A. R., Rahimi, M. Y., & Idris, A. B. (2013). Evaluation of HLB Values of Mixed Non-ionic Surfactants on the Stability of Oil-in-water Emulsion System. *2013 Ukm Fst Postgraduate Colloquium, 1571*, 850-856.

Pedersen, A. M., Bardow, A., Jensen, S. B., & Nauntofte, B. (2002). Saliva and gastrointestinal functions of taste, mastication, swallowing and digestion. *Oral Diseases, 8*(3), 117-129. doi: DOI 10.1034/j.1601-0825.2002.02851.x

Pelosi, P. (1994). Odorant-binding proteins. *Crit Rev Biochem Mol Biol, 29*(3), 199-228. doi: 10.3109/10409239409086801

Pol, E., & Bjork, I. (2001). Role of the single cysteine residue, Cys 3, of human and bovine cystatin B (stefin B) in the inhibition of cysteine proteinases. *Protein Sci, 10*(9), 1729-1738. doi: 10.1110/ps.11901

Qiu, X., Mistry, A., Ammirati, M. J., Chrunyk, B. A., Clark, R. W., Cong, Y., . . . Seddon, A. P. (2007). Crystal structure of cholesteryl ester transfer protein reveals a long tunnel and four bound lipid molecules. *Nat Struct Mol Biol, 14*(2), 106-113. doi: 10.1038/nsmb1197

Rajan, G. H., Morris, C. A., Carruthers, V. R., Wilkins, R. J., & Wheeler, T. T. (1996). The relative abundance of a salivary protein, bSP30, is correlated with susceptibility to bloat in cattle herds selected for high or low bloat susceptibility. *Animal Genetics, 27*(6), 407-414.

Rijnkels, M., Elnitski, L., Miller, W., & Rosen, J. M. (2003). Multispecies comparative analysis of a mammalian-specific genomic domain encoding secretory proteins. *Genomics, 82*(4), 417-432.

Robinson, C. P., Bounous, D. I., Alford, C. E., Nguyen, K. H., Nanni, J. M., Peck, A. B., & Humphreys-Beher, M. G. (1997). PSP expression in murine lacrimal glands and function as a bacteria binding protein in exocrine

secretions. *Am J Physiol, 272*(4 Pt 1), G863-871. doi: 10.1152/ajpgi.1997.272.4.G863

Sayeed, S., Nistico, L., St Croix, C., & Di, Y. P. (2013a). Multifunctional Role of Human SPLUNC1 in Pseudomonas aeruginosa Infection. *Infection and Immunity, 81*(1), 285-291. doi: 10.1128/Iai.00500-12

Sayeed, S., Nistico, L., St Croix, C., & Di, Y. P. (2013b). Multifunctional role of human SPLUNC1 in Pseudomonas aeruginosa infection. *Infect Immun, 81*(1), 285-291. doi: 10.1128/IAI.00500-12

Schauder, C. M., Wu, X., Saheki, Y., Narayanaswamy, P., Torta, F., Wenk, M. R., . . . Reinisch, K. M. (2014). Structure of a lipid-bound extended synaptotagmin indicates a role in lipid transfer. *Nature, 510*(7506), 552-555. doi: 10.1038/nature13269

Schuster-Bockler, B., Schultz, J., & Rahmann, S. (2004). HMM Logos for visualization of protein families. *BMC Bioinformatics, 5*, 7. doi: 10.1186/1471-2105-5-7

Shibata, M., Matsumoto, K., Oe, M., Ohnishi-Kameyama, M., Ojima, K., Nakajima, I., . . . Chikuni, K. (2009). Differential expression of the skeletal muscle proteome in grazed cattle. *Journal of Animal Science, 87*(8), 2700-2708. doi: 10.2527/jas.2008-1486

Shimomura, O., Johnson, F. H., & Saiga, Y. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *J Cell Comp Physiol, 59*, 223-239.

Singh, A. D., & Singh, O. (2017). Prenatal and neonatal development of mandibular salivary gland of Indian buffalo. *Journal of Applied Animal Research, 45*(1), 373-383. doi: 10.1080/09712119.2016.1195392

Suzuki, R., Fujimoto, Z., Shiotsuki, T., Tsuchiya, W., Momma, M., Tase, A., . . . Yamazaki, T. (2011). Structural mechanism of JH delivery in hemolymph by JHBP of silkworm, Bombyx mori. *Sci Rep, 1*, 133. doi: 10.1038/srep00133

Tadjalli, M., Dehghani, S. N., & Ghadiri, M. (2002). Sialography of the goat parotid, mandibular and sublingual salivary glands. *Small Ruminant Research, 44*(3), 179-185. doi: Pii S0921-4488(02)00024-XDoi 10.1016/S0921-4488(02)00024-X

Tall, A. R. (1993). Plasma cholesteryl ester transfer protein. *J Lipid Res, 34*(8), 1255-1274.

Tan, K. W., Jobichen, C., Ong, T. C., Gao, Y. F., Tiong, Y. S., Wong, K. N., . . . Mok, Y. K. (2012). Crystal structure of Der f 7, a dust mite allergen from Dermatophagoides farinae. *PLoS One, 7*(9), e44850. doi: 10.1371/journal.pone.0044850

Tashian, R. E. (1989). The carbonic anhydrases: widening perspectives on their evolution, expression and function. *Bioessays, 10*(6), 186-192. doi: 10.1002/bies.950100603

Terwilliger, T. C., Adams, P. D., Read, R. J., Mccoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., . . . Hung, L. W. (2009). Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallographica Section D-Biological Crystallography, 65*, 582-601. doi: 10.1107/S0907444909012098

Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., . . . Adams, P. D. (2008). Iterative model building, structure refinement and density modification with the PHENIX

AutoBuild wizard. *Acta Crystallographica Section D-Biological Crystallography, 64*, 61-69. doi: 10.1107/S090744490705024x

Tobias, P. S., Soldau, K., Gegner, J. A., Mintz, D., & Ulevitch, R. J. (1995). Lipopolysaccharide binding protein-mediated complexation of lipopolysaccharide with soluble CD14. *J Biol Chem, 270*(18), 10482-10488.

Tobias, P. S., Soldau, K., & Ulevitch, R. J. (1986). Isolation of a lipopolysaccharide-binding acute phase reactant from rabbit serum. *J Exp Med, 164*(3), 777-793.

Toulmay, A., & Prinz, W. A. (2012). A conserved membrane-binding domain targets proteins to organelle contact sites. *J Cell Sci, 125*(Pt 1), 49-58. doi: 10.1242/jcs.085118

Tsien, R. Y. (1998). The green fluorescent protein. *Annu Rev Biochem, 67*, 509-544. doi: 10.1146/annurev.biochem.67.1.509

Turk, B., Turk, D., & Turk, V. (2000). Lysosomal cysteine proteases: more than scavengers. *Biochim Biophys Acta, 1477*(1-2), 98-111.

Turk, B., Turk, V., & Turk, D. (1997). Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors. *Biol Chem, 378*(3-4), 141-150.

Vance, S. J., McDonald, R. E., Cooper, A., Smith, B. O., & Kennedy, M. W. (2013). The structure of latherin, a surfactant allergen protein from horse sweat and saliva. *J R Soc Interface, 10*(85), 20130453. doi: 10.1098/rsif.2013.0453

von der Mohlen, M. A., Kimmings, A. N., Wedel, N. I., Mevissen, M. L., Jansen, J., Friedmann, N., . . . et al. (1995). Inhibition of endotoxin-induced cytokine release and neutrophil activation in humans by use of recombinant bactericidal/permeability-increasing protein. *J Infect Dis, 172*(1), 144-151.

Waugh, D. S. (2016). Crystal structures of MBP fusion proteins. *Protein Sci, 25*(3), 559-571. doi: 10.1002/pro.2863

Weiss, J., Elsbach, P., Olsson, I., & Odeberg, H. (1978). Purification and characterization of a potent bactericidal and membrane active protein from the granules of human polymorphonuclear leukocytes. *J Biol Chem, 253*(8), 2664-2672.

Wernimont, A., & Edwards, A. (2009). In situ proteolysis to generate crystals for structure determination: an update. *PLoS One, 4*(4), e5094. doi: 10.1371/journal.pone.0005094

Weston, W. M., LeClair, E. E., Trzyna, W., McHugh, K. M., Nugent, P., Lafferty, C. M., . . . Greene, R. M. (1999). Differential display identification of plunc, a novel gene expressed in embryonic palate, nasal epithelium, and adult lung. *Journal of Biological Chemistry*, *274*(19), 13698-13703. doi: DOI 10.1074/jbc.274.19.13698

Wheeler, T. T., Haigh, B. J., Broadhurst, M. K., Hood, K. A., & Maqbool, N. J. (2011). The BPI-like/PLUNC family proteins in cattle. *Biochemical Society Transactions, 39*, 1006-1011. doi: 10.1042/Bst0391006

Wheeler, T. T., Haigh, B. J., McCracken, J. Y., Wilkins, R. J., Morris, C. A., & Grigor, M. R. (2002). The BSP30 salivary proteins from cattle, LUNX/PLUNC and von Ebner's minor salivary gland protein are members of the PSP/LBP superfamily of proteins. *Biochim Biophys Acta, 1579*(2-3), 92-100.

Wheeler, T. T., Hood, K. A., Maqbool, N. J., McEwan, J. C., Bingle, C. D., & Zhao, S. Y. (2007). Expansion of the bactericidal/permeability increasing-like

(BPI-like) protein locus in cattle. *Bmc Genomics, 8*. doi: Artn 7510.1186/1471-2164-8-75

Wheeler, T. T., Hood, K., Oden, K., McCracken, J., & Morris, C. A. (2003). Bovine parotid secretory protein: structure, expression and relatedness to other BPI (bactericidal/permeability-increasing protein)-like proteins. *Biochem Soc Trans, 31*(Pt 4), 781-784. doi: 10.1042/

Zhang, L., Yan, F., Zhang, S., Lei, D., Charles, M. A., Cavigiolio, G., . . . Ren, G. (2012). Structural basis of transfer between lipoproteins by cholesteryl ester transfer protein. *Nat Chem Biol, 8*(4), 342-349. doi: 10.1038/nchembio.796

Zhou, H. D., Li, G. Y., Yang, Y. X., Li, X. L., Sheng, S. R., Zhang, W. L., & Zhao, J. (2006). Intracellular co-localization of SPLUNC1 protein with nanobacteria in nasopharyngeal carcinoma epithelia HNE1 cells depended on the bactericidal permeability increasing protein domain. *Molecular Immunology, 43*(11), 1864-1871. doi: 10.1016/j.molimm.2005.10.021

Zhou, H. D., Li, X. L., Li, G. Y., Zhou, M., Liu, H. Y., Yang, Y. X., . . . Sheng, S. R. (2008). Effect of SPLUNC1 protein on the Pseudomonas aeruginosa and Epstein-Barr virus. *Molecular and Cellular Biochemistry, 309*(1-2), 191-197. doi: 10.1007/s11010-007-9659-3