Structural and functional studies of Gram-Negative Bacteria Phospholipid Transport Proteins and Human SHP2 Protein

James William Coleman
100173709

Thesis Submitted for the degree of Doctor of Philosophy

16th October 2020

ABSTRACT

Gram-Negative bacteria have an envelope composed of an Inner Membrane (IM) and Outer Membrane (OM) with a periplasmic space between the two. The OM is an asymmetrical membrane which forms an important protective layer for the cell and is a contributing factor to resistances to toxic compounds and antibiotic agents. The OM consists of an outer leaflet made up of Lipopolysaccharides (LPS), and an inner leaflet constructed from Phospholipids (PLs), predominantly Phosphatidylethanolamine (PE), Phosphatidylglycerol (PG) and Cardiolipin (CA) in *Escherichia coli.* The manner by which LPS is transported between the IM and the OM has been well studied, with the Lipopolysaccharide transport (Lpt) protein complex responsible for their trafficking from IM to OM. In contrast, the pathways by which Phospholipids are transported between the two membranes remain unclear. The Maintenance of lipid asymmetry (Mla) pathway has been partially characterised and is thought to shuttle PLs from the OM to the IM (the retrograde direction), although this directionality is a point of contention. The YebST and PqiABC pathways have also been implicated in PL trafficking, but the structures of these proteins and the associated mechanisms by which they potentially transport PLs are unknown. The work in this thesis uses the technique of Protein X-ray Crystallography to gain structural and functional insights into the putative PL transport pathways YebST and PqiABC. In the following work, a novel structure for PqiC is determined, with preliminary *in vitro* and *in vivo* assays suggesting functionally and structurally important residues, as well as the protein's capability of binding PLs. Successful purification and crystallisation of the protein YebT is also outlined, in addition to a purification technique and initial crystallisation trials for the IM-localised YebS protein, highlighting it as a target for future structural studies.

Further to this work, Protein X-ray Crystallography is also here used to characterise the binding between the *Homo sapiens* SHP2 protein and the compound Methylene Blue (MB). SHP2 is the product of a protooncogene and appears to allow cancer cells to evade the immune response through cooperation with the regulatory protein PD-1. The SHP2-MB interaction was identified by the work of collaborators and was shown to prevent

SHP2 binding to PD-1. A concurrent investigation into the interaction between SHP2 and PD-1 was also carried out. While SHP2 was successfully purified and crystallised, the work here shows that MB does not bind SHP2 directly, and is instead likely binds to PD-1, an insight which will inform further study into the potential therapeutic uses of MB in PD-1-mediated immune evasion.

Table of Contents

Acknowledgments

I would like to extend my gratitude to my supervisors, Professors Changjiang Dong and Tom Wileman for their unwavering support throughout my PhD. Changjiang's knowledge and supervision were invaluable to my research, and Tom's advice and guidance were vital during turbulent times.

I also wish to extend my thanks to all those who have assisted me during my time at the UEA. Particular thanks to Saannya Sequeira, Gareth Ashworth and Yinghong Gu for their help day-to-day in the lab, particularly when I was finding my feet.

I must also thank my friends and family for their support during my PhD; they have been a constant source of reassurance and happiness during my studies.

Finally, I wish to express my deepest appreciation to my wife Kelly, without whom I would have felt lost on this journey; thank you.

List of Abbreviations

ABC	ATP Binding Cassette
BAM	Beta Barrel Assembly Machinery
bp	Base Pairs
CA	Cardiolipin
CDK	Cyclin-dependent Kinase
CDP-DAG	cytidine diphosphate-diacylglycerol
Cryo-EM	Cryogenic electron microscopy
DDM	n-dodecyl-β-D-maltopyranoside
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EMD	Electron Microscopy Data Bank
EtBr	Ethidium Bromide
Hep	Heptose
IM	Inner Membrane
IMAC	Immobilised Metal Affinity Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
ITSM	immunoreceptor tyrosine-based switch motif
Kdo	3-deoxy-D-manno-oct-2-ulosonic acid
LB	Luria broth
LDAO	n-dodecyl-N,N-dimethylamine-N-Oxide
Lol	Localisation of lipoproteins
LPS	Lipopolysaccharide
Lpt	Lipopolysaccharide transport
MAD	Multiple -wavelength anomalous diffraction
MAPK	the Mitogen-activated protein kinase
MB	Methylene Blue
MBP	Maltose Binding Protein
MCE	Mammalian Cell-entry
MKP	MAPK phosphatase
Mla	Maintenance of Lipid Asymmetry
MR	Molecular Replacement
MWCO	Molecular Weight Cut-off
OD_{600}	Optical Density at 600nm
OM	Outer Membrane
OMPs	Outer Membrane Proteins
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
PDB	Protein Data bank
PE	Phosphatidylethanolamine
PEG	Polyethylene Glycol
PG	Phosphatidylglycerol
PL	Phospholipid

PTP	Protein tyrosine phosphatase
SAD	Single-wavelength anomalous diffraction
SDS-PAGE	Sodium dodecyl sulphate Polyacrylamide gel electrophoresis
SEC	Size-exclusion Chromatography
Sec	Secretion machinery
SeMet	Selenomethionine
SH2	Src homology region 2
TAE	Tris acetate EDTA
TCEP	Tris(2-carboxyethyl) phosphine
TEV	Tobacco Etch Virus nuclear-inclusion-a endopeptidase
TLC	Thin Layer Chromatography
TMDs	Transmembrane Domains
WT	Wild Type

List of Figures

List of Tables

CHAPTER 1 – The Outer Membrane of Gram-Negative Bacteria

1.1 Gram-Negative and Gram-Positive Bacteria

The domain of bacteria can be broadly divided into gram-positive and gram-negative organisms; this classification is dependent on the structure of their outer envelopes. Gram-negative bacteria have an inner and an outer membrane, with a periplasmic space containing a small amount of peptidoglycan, a large macromolecule consisting of strands of 1,4-linked N-acetylglucosamine (NAG) and N-acetyl muramic acid (NAM) sugar residues cross-linked by a short peptide moiety (1). Gram positive bacteria have no outer membrane, but a thick peptidoglycan layer forms a cell wall outside the plasma membrane (2). The unique asymmetric outer membrane of gram-negative bacteria acts as a significant barrier to the uptake of antibiotics and hydrophobic drugs, increasing drug resistance when compared to gram-positive bacteria (3).

1.2 The Envelope of Gram-Negative Bacteria

The envelope of gram-negative bacteria consists of the outer membrane (OM), inner membrane (IM) and the periplasm. A distinctive characteristic of gram-negative bacteria, the OM is an asymmetric bilayer made up of glycolipids and phospholipids; its outer envelope has high levels of the glycolipid Lipopolysaccharide (LPS), forming a thick, permeability-reducing barrier (4). Enclosed between the two membranes is the periplasm; this is the aqueous region found between the inner and outer membranes **(Fig 1.1)**. The periplasm is a viscous compartment, containing enzymes which would be toxic if located in the cytoplasm; it is also home to important components involved in drug efflux (5). The sequestration of RNases and proteases in this area bears significant similarity to the eukaryotic lysosome, which is a membrane-bound organelle containing otherwise harmful lytic enzymes (6).

Proteins embedded in the OM generally fall into two classes: Outer Membrane Proteins (OMPs) or Lipoproteins (4). OMPs take the form of β-barrel proteins, and often span both leaflets of the outer membrane, commonly forming channel proteins or porins (7). These porins take the form of water-containing pores which

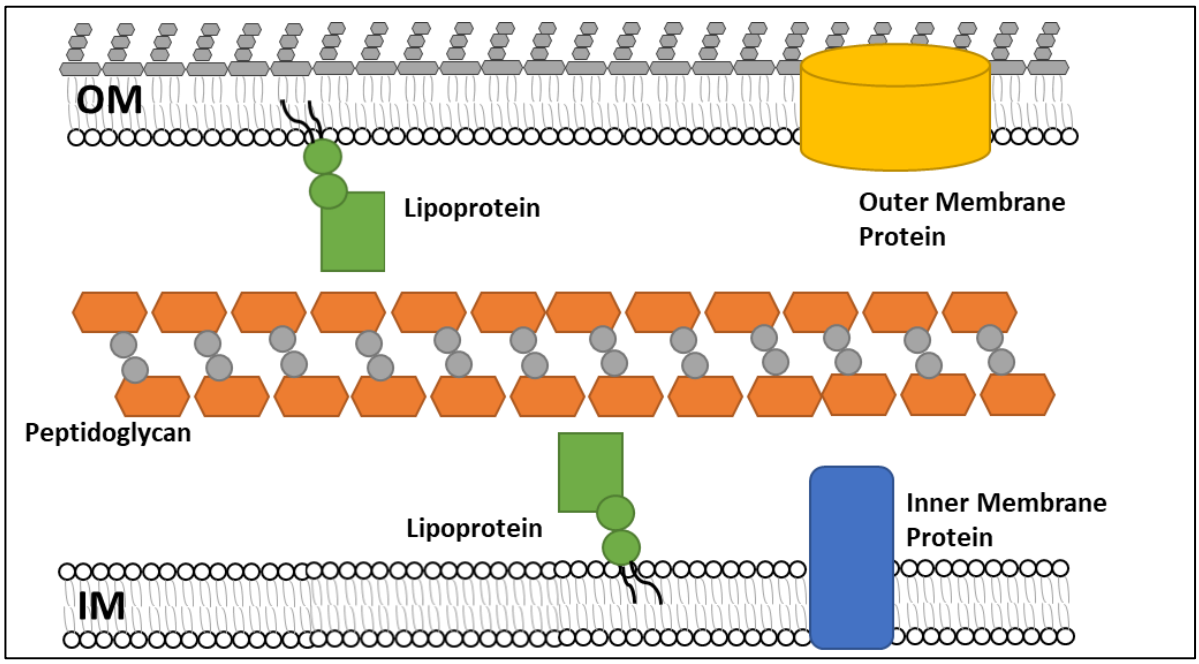


Figure 1.1 – Diagram of the Gram-negative cell membrane. Gram-negative bacteria possess two distinct membranes, separated by a layer of peptidoglycan. The inner membrane and inner leaflet of the outer membrane are composed of Phospholipids (PLs), with Lipopolysaccharides (LPS) making up the outer leaflet of the outer membrane. Proteins are found crossing both membrane leaflets (OMPs and IMPs), as well as anchored between PLs in the IM and inner leaflet of the OM (lipoproteins) (4).

allow the inward diffusion of small nutrient molecules into gram-negative cells (8). β-barrels are composed of 8-36 strands, arranged in a twisted-cylindrical conformation with hydrophobic residues exposed to the surrounding membrane in which it is embedded, enabling their OM anchoring (9, 10). OMPs are manufactured in the cytoplasm as nascent proteins with N-terminal signal peptides, which are not present in mature proteins (11). The proteins destined for OM incorporation are chaperoned by SecA and SecB, before their movement across the inner membrane by the Sec translocon, with unfolded nascent proteins passing through the SecYEG IM channel (12). Chaperoned by SurA, Skp and FkpA, OMPs remain in unfolded form until they reach the OM (13, 14). At the outer membrane, unfolded OMPs interact with the β-barrel assembly machinery (BAM complex), which facilitates their folding and insertion into the OM; recent studies have indicated that OMPs migrate into the OM through a lateral opening in BamA (15, 16) **(Fig 1.2)**.

In bacteria, lipoproteins are hydrophilic proteins which are anchored to the cell's membrane through an N-terminal fatty acid moiety (17). Outer membrane lipoproteins are attached to the inner envelope of the OM, and while *E. coli* contains at least 90 lipoproteins, their functions remain largely unknown (18). Braun's Lipoprotein is a notable exception; it is known to 'staple' peptidoglycan to the outer membrane, and is the protein of greatest abundance in *E. coli* (4, 19). Like OMPs, OM-destined lipoproteins are produced in the cytosol of gram-negative cells and translocated to the outer leaflet of the inner membrane by machinery of the Sec translocon. Here, lipoproteins are modified with a membrane anchor by Lgt, their helical signal peptides are removed by LspA and are then acylated on their N-terminal cysteine residues by Lnt (20). Then, the ABC transporter LolCDE transfers

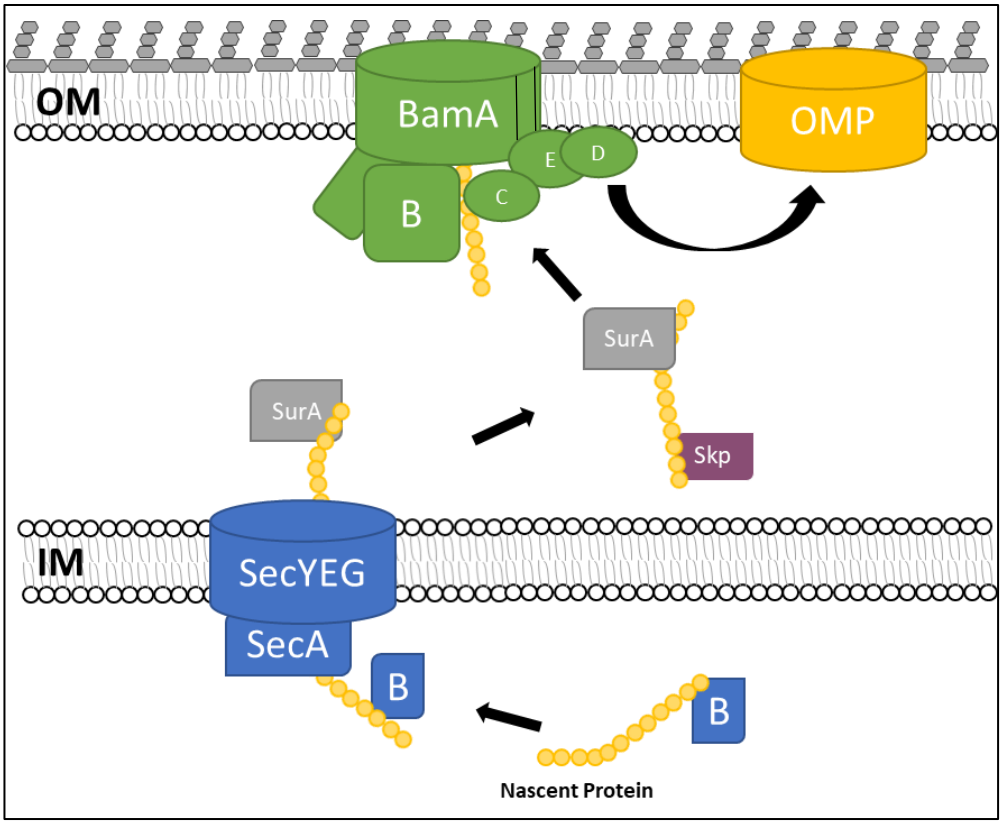


Figure 1.2 - Schematic of the transport of *E. coli* outer membrane proteins (OMPs). Nascent OMPs are synthesised in the cytosol and move to the SecYEG channel, chaperoned by SecB and SecA**.** With assistance of the chaperones SurA, Skp and FkpA, the OMP is transported to the BamABCDE complex in the outer membrane. This protein complex folds and inserts the OMP into the OM through a lateral gate in BamA (9, 22).

the mature protein to the LolA chaperone, mediating its transport to LolB which assists in its final insertion into the OM (21).

In the absence of the membrane-bound organelles existent in eukaryotic cells, the inner membrane of Gram-negative bacteria houses molecular machinery, including that pertaining to ATP synthesis, lipid biosynthesis and protein transport. These three processes have largely conserved components in eukaryotes but are instead carried out in organelles. Unlike the OM, the inner membrane contains no LPS, and is a symmetrical phospholipid bilayer composed of phosphatidyl ethanolamine and phosphatidyl glycerol, as well as small amounts of phosphatidyl serine and cardiolipin (4).

While transmembrane proteins in the outer membrane are almost exclusively formed from β-sheets, proteins spanning the inner membrane are formed of α-helical transmembrane domains (22). Insertion of these α-helical transmembrane proteins is mediated by the SecYEG channel and the YidC protein, in many cases after recruitment of the Signal Recognition Particle (SRP) to hydrophobic N-terminal regions of nascent protein (23). Inner membrane lipoproteins are targeted in a similar manner to their OM counterparts, with protein reaching the periplasmic face of the IM through the SecYEG translocon. Here, proteins are again processed by Lgt, LspA and Lnt, but instead of export into the periplasm through the Lol pathway, lipoproteins with aspartate as their second residue after their modified cysteine are retained in the IM (24).

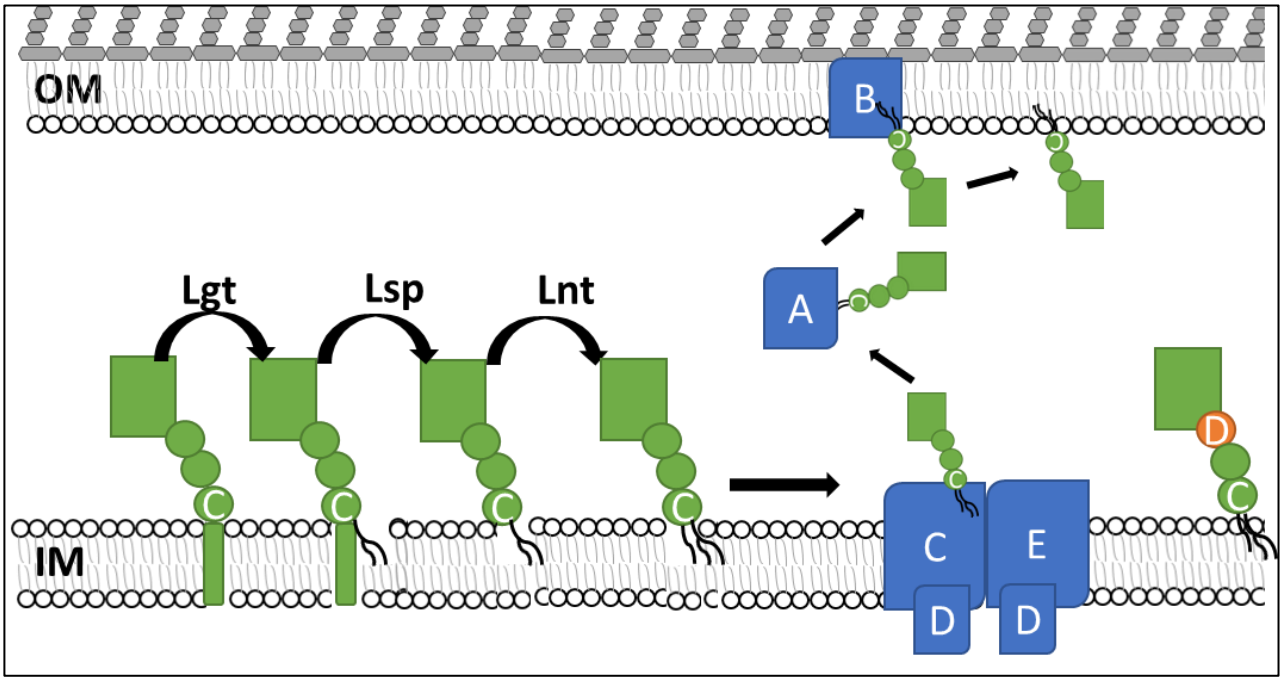


Figure 1.3 - Mechanism of lipoprotein transport from inner membrane to outer membrane using the Lol system. Lipoproteins (green) are anchored in the membrane with a diacyl-glyceryl moiety by Lgt, before cleavage of their N-terminal signal peptide by Lsp. The N-terminal cysteine residue is acylated by Lnt, before potential recognition by the LolCDE complex (blue). Proteins lacking a +2-aspartate residue are extracted into the periplasm by LolCDE and chaperoned by LolA before reaching the outer membrane, with their insertion mediated by LolB. Lipoproteins with a +2 aspartate are retained in the IM, without periplasmic expulsion by LolCDE (25).

1.3 Lipopolysaccharides (LPS)

Lipopolysaccharides (LPS) are a primary component of the outer membrane in gram-negative bacteria. They are vital to the cell's structural integrity and stability, as well as protecting the cell from chemicals in its environment (26). LPS is an archetypal pathogen associated molecular pattern (PAMP), allowing organisms to recognise bacterial infection and elicit an immune response. The polysaccharide constituent of LPS plays a protective role for a bacterial cell, both mimicking host carbohydrates and subverting the effects of the complement system (27). LPS has three distinct regions: Lipid A, Core Oligosaccharide and the O-antigen **(Fig 1.4)** (28).

Lipid A is recognised by the TLR4/MD2 receptor in mammalian immune systems, and is the region of LPS most responsible for the toxicity of LPS, as its presence can cause the overexpression of inflammatory signalling molecules, leading to apoptotic cell death (29) (30) (31). Lipid A is composed of a phosphorylated glucosamine dimer, with up to four acyl chains attached through ester or amide linkages. While Lipid A is the least variable region of LPS, these four acyl chains can themselves be substituted with fatty acid chains. In *E. coli* this is commonly with two further saturated chains; unsaturated fatty acid chains are uncommon in Lipid A (28).

The O-antigen is a highly variable portion of LPS which extends out from the cell into the environment. It consists of up to 40 repeating subunits of 3-5 sugar oligosaccharides and is the hydrophilic region of LPS. The O-antigen is a key part of a Gram-negative bacterium's recognition by a host immune system, as it is the site to which antibodies bind (32). The number of unique O-antigens is vast, with around 170 seen in *E. coli* alone (33).

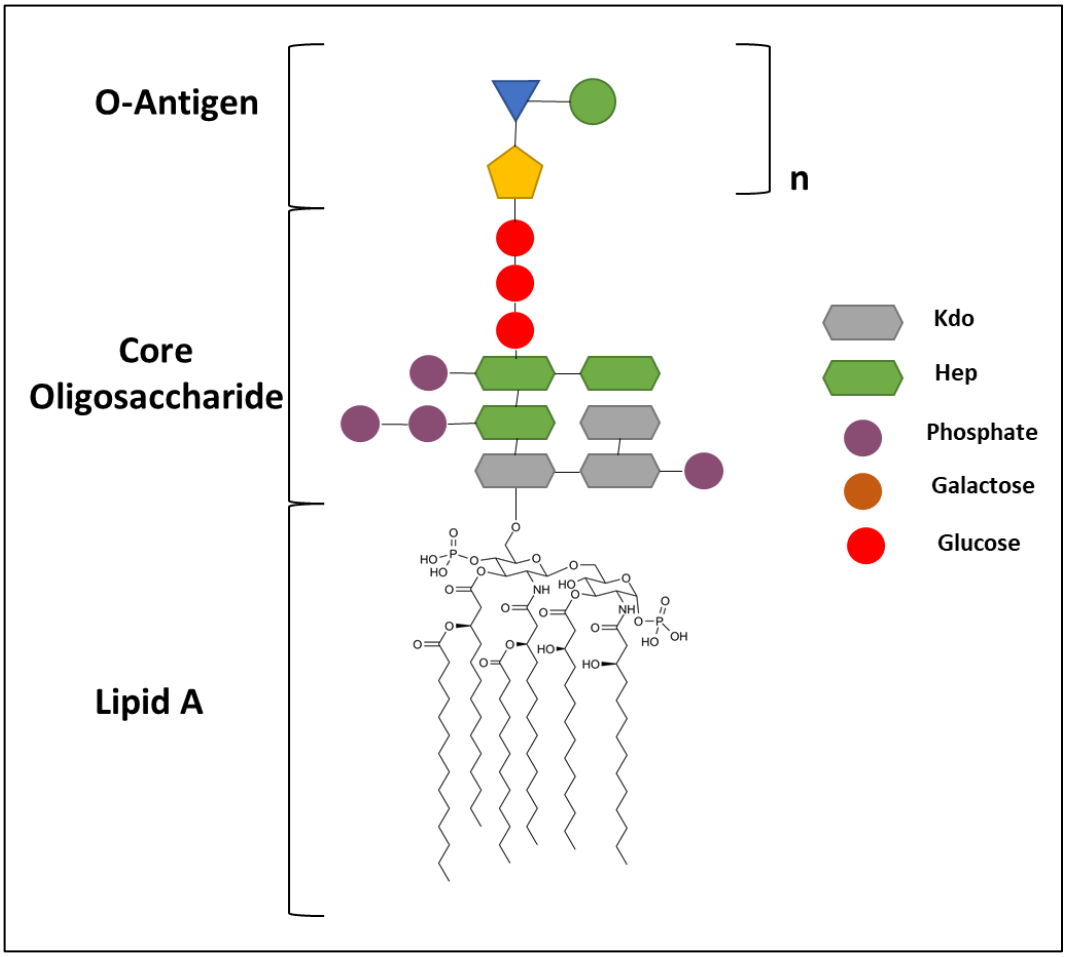


Figure 1.4 LPS structure of *E. coli* K-12. The three major regions of LPS are indicated. Lipid A is largely conserved, and the core oligosaccharide can vary by substitution or branching of different sugar residues. The O-antigen is by far the most variable region of LPS, forming the molecule's hydrophilic region, which is extended into the organism's environment (34).

The Core Oligosaccharide region of LPS is a short chain of sugar residues which connect the O-antigen to the lipid A anchor in the outer membrane. The Core can be considered as two separate regions: inner (the Lipid A adjoining side) and outer (the O-antigen adjoining side). The inner core region is composed of the monosaccharides 2-keto-3-deoxyoctonoic acid (Kdo) and l-glycero-d-manno-heptose (Hep), which can be modified by phosphorylation. A single Kdo residue is vital in connecting the inner core to lipid A, with the rest of the inner core comprising up to 6 Kdo or Hep residues (32). The inner and outer cores are connected through a Hep residue, with the outer core itself comprising of a backbone of three Hexose (Hex) residues, which themselves can be branched with further Hex residues (35). The core region of LPS can vary significantly between species, but also within the same species. In *E. coli* for example, there are five distinct core types, termed K-12, R1, R2, R3 and R4. These core types are defined by the structure of the outer core, all of which comprise of a backbone of 4 hexose sugars with two branched residues; the cores differ in whether residues are galactose or glucose (36).

LPS synthesis has been best defined in *E.* coli; the first step in LPS synthesis is the construction of Lipid A-Kdo, which takes place in the cytoplasm and inner face of the inner membrane. Nine distinct enzymes are involved in this synthetic pathway. Following this, the core oligosaccharide is rapidly assembled onto the Lipid A-Kdo molecule which is dependent on a number of membrane-bound glycosyltransferases, such as WaaC (37). The precursor to the O-antigen is also synthesised on the cytoplasmic side of the inner membrane, by glycosyltransferases encoded by the *rfb* gene cluster. Lipid A-core and the O-antigen precursor are both 'flipped' from the cytoplasmic to the periplasmic face of

the inner membrane by the MsbA and Wzx membrane proteins respectively. Once across the membrane, the precursor to the O-antigen is polymerised by the enzymes Wzy and Wzz to yield the completed O-antigen (37). Before any part of LPS is moved across the periplasmic space to the outer membrane, the O-antigen is ligated to the Lipid A-Core region complex. This is achieved by WaaL ligase (the product of the *waaL* gene*)*, a poorly understood enzyme found on the *waaQ* operon (33).

Once fully assembled by its ligation with WaaL, LPS is shuttled across the periplasmic space by the Lpt complex to the outer membrane; the Lpt complex contains 7 separate proteins, termed LptA-G (37). This molecular LPS transporting machine has IM ($LptB_2CFG$) and OM (LptDE) portions, with LptA bridging the gap across the periplasmic space. The first step in this pathway is the extraction of LPS from the inner leaflet of the IM by the ABC (ATP-binding-cassette) protein $LptB_2FG$ - powered by the hydrolysis of two ATP molecules - which distributes LPS to LptC (38). It should be noted, however, that while $LptB_2FG$ is an ABC transporter, it behaves atypically for a protein in its class – rather than moving its substrate across the membrane, it passes LPS from the IM inner leaflet directly to LptC (39). LPS is moved from LptC to LptA – again requiring ATP hydrolysis – which completes a bridge across the periplasm of the cell (38, 40). From LptA, LPS is transferred to the LptDE complex. These two proteins form a 'plug and barrel' structure, with the β-barrel LptD and its associated lipoprotein 'plug' LptE. While LptE and its location within the LptD barrel has been shown to be important for LPS insertion in the OM, its mechanistic involvement is unclear (41). Once in the lumen of LptD, lipid A from LPS is released into the OM through a lateral gate, and the translocation across the membrane is complete (37, 42-45) **(Fig1.5)**.

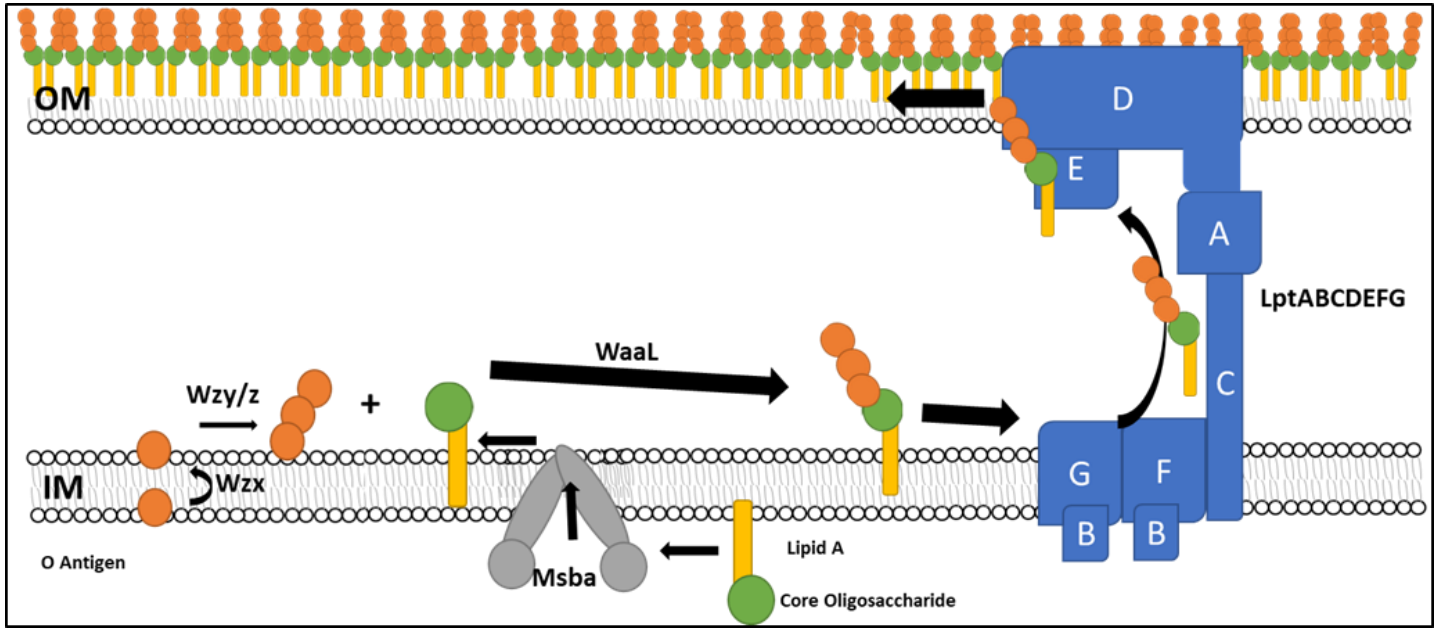


Figure 1.5 - The formation and export of LPS in *E. coli*. Lipid A, the Core oligosaccharide and O-antigen are all initially synthesised on the inner face of the inner membrane. Lipid A and the Core region are ligated and moved across the inner membrane by the ABC transporter MsbA; the O-antigen is moved across the same membrane by the Wzx transporter. The Core/Lipid-A molecule is attached to the O-antigen by WaaL, and nascent LPS is shuttled to the inner leaflet of the outer membrane by the LptBFG complex (Assisted by LptC/A). Finally, LPS is inserted into the outer leaflet of the outer membrane through the lateral gate of the LptDE complex of proteins (34).

1.4 Phospholipids (PLs)

While the outer leaflet of the outer membrane of gram-negative bacteria predominantly contains LPS, the inner membrane is formed of phospholipids (PLs). In *E. coli*, around 75% of the PLs are Phosphatidylethanolamines (PE), around 20% are Phosphatidylglycerols (PG), with a smaller portion of PLs in the form of cardiolipin (CA) (46). In other bacteria, monomethyl PE (MMPE), dimethyl PE (DMPE) and phosphatidylcholine (PC) are commonly found, all of which are derived from Phosphatidylethanolamines **(Fig 1.6)** (47). Phospholipids are evidently important in the function of the OM as a protective barrier for the cell, given that they govern membrane permeability and viscosity (48).While the exact roles of PLs in membrane formation and stability are not clear, it has been shown that changes in cellular phospholipid proportions can effect cell size and stress response (49). In addition, the phospholipid makeup of the OM has been shown to be important for the function of outer membrane proteins, further highlighting their importance to the cell (50).

E. coli phospholipids are synthesised on the cytoplasmic side of the inner membrane (51) **(Fig 1.7)**. The initial step of this pathway is the addition of two acyl groups to glycerol 3-phosphate - carried out by PlsB and PlsC - producing phosphatidic acid (52). Phosphatidic acid is subsequently modified by the enzyme CdsA to produce the shared precursor cytidine diphosphate-diacylglycerol (CDP-DAG) (47).

PE is formed by a subsequent two-step reaction, beginning with that between CDP-DAG and serine – catalysed by PssA - with Phosphatidylserine as an intermediate. The decarboxylation of Phosphatidyl serine by Psd completes the process and

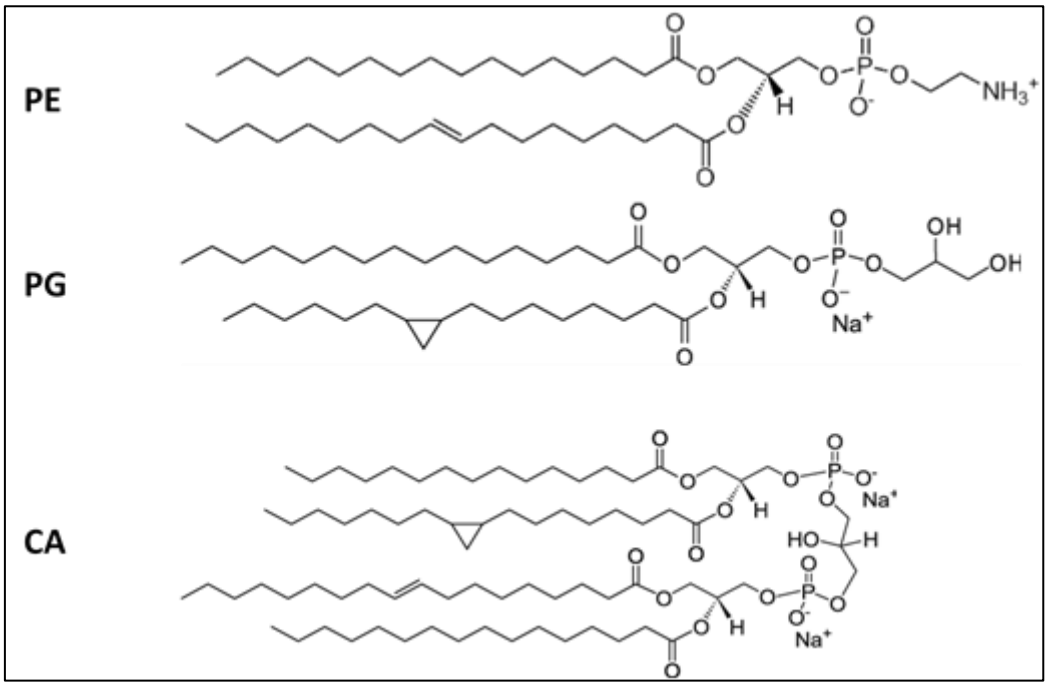


Figure 1.6 - Representative structures of *E. coli* Phospholipids. The three predominant phospholipids in the outer membrane of *E. coli* are Phosphatidylethanolamine (PE), Phosphatidylglycerol (PG), and Cardiolipin (CA). Shown are 'average' PL structures, which vary by fatty acid moieties. Each PL is composed of a long, hydrophobic fatty acid tail region, and a hydrophilic head region with a phosphate group (two in the case of CA) and an ethanolamine (PE) or glycerol group (PG, CA). PE is zwitterionic, while PG and CA are negatively charged at physiological pH.

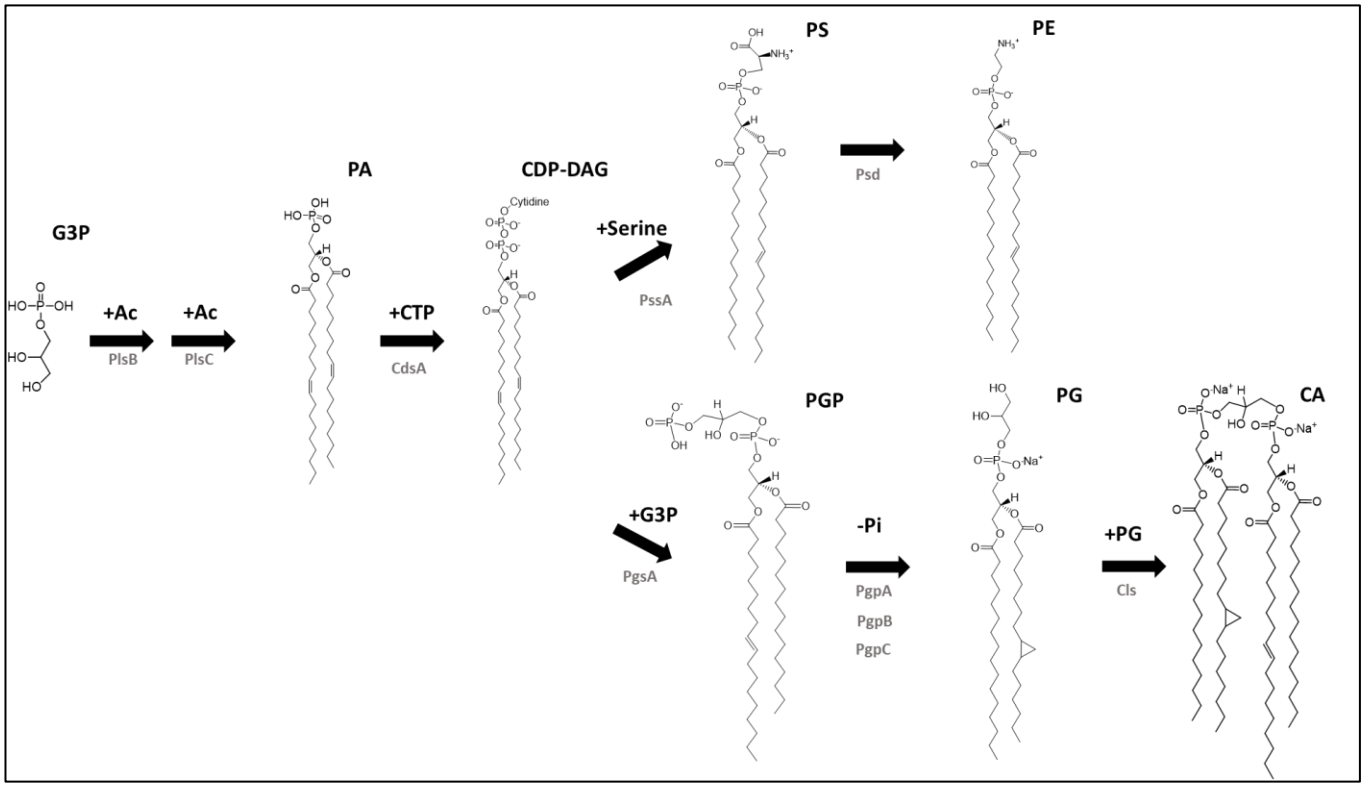


Figure 1.7 – Pathway for the formation of phospholipids Phosphatidylethanolamine (PE), Phosphatidylglycerol (PG) and Cardiolipin (CA) in E. *coli.* These reactions take place on the inner leaflet of the inner membrane in Gram -negative bacteria. Each phospholipid is ultimately derived from glycerol-3-phosphate (G3P), with the intermediate Phosphatidylserine (PS) in PE synthesis, and Phosphatidylglycerol phosphate (PGP) in PG and CA production. CA is produced from the fusion of two PG molecules. Enzymes involved in this pathway are listed in grey – notably, only the dephosphorylation of PGP has been attributed to more than one enzyme (PgpA-C).

yields PE molecules (53). PG is formed from a PgsA-mediated reaction between CDP-DAG and glycerol-3-phosphate, resulting in the intermediary molecule phosphatidylglycerol phosphate. PGP is subsequently dephosphorylated by enzyme PgpA, PgpB or PgpC to produce PG (52). In *E.coli,* CA is formed through the joining of two PG molecules by cardiolipin synthase, with the release of glycerol (54). The method by which mature phospholipids are 'flipped' across the IM remains uncharacterised – a mechanistic homologue for this role is MsbA in LPS transport.

1.5 Phospholipid transport

Whist the nature of the transport of LPS, OMPs and Lipoproteins across the periplasmic space in Gram-negative bacteria has been well-characterised, the mechanisms of Phospholipid transport remain unclear. While the synthetic pathways of PLs in Gram-negative bacteria are relatively well understood, even the pathway for phospholipid movement across the inner membrane is unknown, let alone pathway across the periplasmic space to their insertion in the outer membrane. Despite the overall uncertainty regarding PL transport, it was demonstrated as far back as 1977 that phospholipids could be transported from outer membrane to inner membrane (the retrograde direction), as well as from IM to the OM (anterograde) in *Salmonella* (55). This bidirectionality of transport mechanistically distinguishes PL transport from the trafficking of LPS, and at the time was attributed to physical contacts between the two membranes, rather than a membrane spanning macromolecule in the ilk of the Lpt complex. It was not until 2009 that the first transmembrane phospholipid transport pathway was as much as identified; this ABC pathway was henceforth known as the Mla system – so called as it was observed to influence the maintenance of lipid asymmetry (56). Interruption of components of this pathway resulted in the sequestration of PLs in the OM,

thereby indicating its importance in retrograde phospholipid transport (56). While the Mla pathway was initially thought to comprise the 6 Mla proteins (A-F), it was later noted that the β-barrel protein OmpC was also a key part of the system, which may be responsible for the translocation of aberrant PLs from the outer leaflet to the inner leaflet of the OM (57). The Mla pathway has been determined to three distinct sub-sections; there is the inner membrane portion (MlaFEDB), the outer membrane portion (MlaA/OmpC) and a soluble, periplasmic protein (MlaC) (58-60). MlaA has also been shown to interact with the trimeric β-barrel protein OmpF (which is structurally very similar to OmpC), but as yet there is no evidence that this interaction is of mechanistic significance to phospholipid transport (57). Mechanically, the Mla machinery draws parallels with the previously described pathways for lipoprotein and OMP export to the OM, with inner (Sec or Lol complexes) and outer (Bam/LolB) membrane components, alongside a periplasmic transporter (SurA/LolA).

Structural characterisation of the Mla pathway has been performed **(Fig 1.8)**, with the basis of a mechanism outlined. The β-barrel protein OmpC is thought to extract phospholipids from the outer membrane, before transferring them to interacting partner MlaA. Next, MlaA transfers these proteins to periplasmic transporter MlaC, which has been determined to interact with MlaD, in an interaction potentially pertaining to PL transfer to the IM (60). The inner membrane complex MlaFEDB takes on the structure of an ABC transporter, with transmembrane region (E), alongside a cytoplasmic nucleotide binding and accessory domain (F, B) (58). The MlaD protein is a hexamer, formed of a N-terminal helix in the IM and C-terminal mammalian cell entry (MCE) domain in the periplasm (58). These MCE domains have been previously shown to be involved in lipid binding, and may be key players in inter-membrane PL

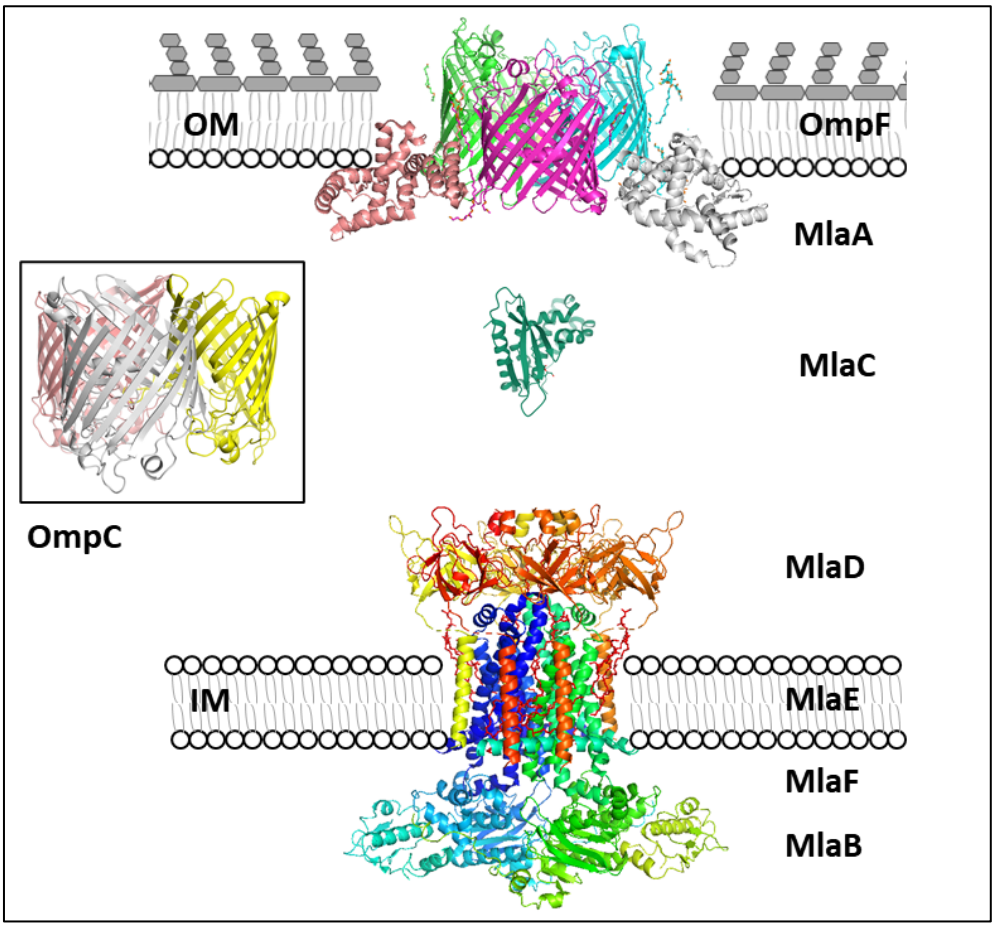


Figure 1.8 - Diagram of the Mla pathway, with structures for MlaFEDB, MlaC and OmpC from *E. coli,* and MlaA/OmpF from *K. pneumoniae* (Not to scale). Here the distinction between inner and outer membrane proteins can be seen, with MlaE forming of a bundle of α-helices, with OmpF taking the form of a trimeric β-barrel. MlaFEB follows the archetype of an ABC transporter, with two transmembrane (E) domains, as well as two nucleotide binding domains(F) and two cytoplasmic accessory proteins (B). MlaD takes the form of a mammalian cell entry (MCE) domain in the periplasm and is a hexamer of α-helical subunits. It should be noted that while MlaA appears to interact with the structurally similar OmpF and OmpC, only OmpC has been determined to be mechanistically important to the Mla pathway. OmpF is shown to illustrate the likely interaction between OmpC and MlaA, as a structure of MlaA in complex with OmpC is yet to be obtained. MlaC has been shown to transfer Phospholipids to and from MlaD, but the directionality of this movement is hotly debated. Structures from PDB files 7CGE (58), 5UWA (60), 5NUP (59) and 2XG6 (61).

movement (62). Once accepted by MlaD, PLs are thought to move into MlaE, where they are 'squeezed' into the Inner membrane in an ATP-dependent manner (58).

Although this retrograde PL transport mechanism explains many observations made about the Mla system, there is much debate as to whether this pathway is also involved in the anterograde movement of PLs. *In vitro* studies have shown that PL transfer can occur from MlaD to MlaC, as well as from C to D (63). With the nature of the key MlaA and MlaC interaction yet to be determined, the directionality of Mla PL transport remains uncertain (64). Clearly, more investigation into the Mla pathway is needed to confirm the true nature of its interaction with phospholipids. Even if the Mla system *was* an anterograde PL transport pathway, it is not essential for cell viability, and as such, other pathways must be present for anterograde PL transport (64).

While the directionality of Mla transport is debated, the nature of other putative PL transport pathways in *E. coli* remain far less clear. Proteins arising from the *Tol-Pal*, *YebST* and *PqiABC* gene clusters have been all been implicated in phospholipid transport, but their roles (and indeed the direction in which they *may* transport PLs) remain unclear (65) (60). The Tol-Pal system has been shown to be important for OM stability in Gram-negative bacteria and has been proposed as a transmembrane retrograde PL transporter. However, subsequent investigation has indicated that Tol-pal interacts with several OM lipoproteins involved in cell wall modifications and may be involved with cell division, as opposed to purely PL transport (65, 66).

YebST and *PqiABC* operons were first proposed as gene clusters encoding phospholipid transport machinery due to their possession of proteins with predicated MCE domains – YebS and PqiB were both determined to contain this

domain, which has been implicated in lipid binding (60, 67, 68). *PqiABC* and *YebST* operons were both subsequently shown to be significant in the stability of the outer membrane. The SDS-EDTA sensitivity of Cells deficient in the Mla pathway was heightened by the removal of both *Pqi* and *Yeb* operons, with complementation by PqiABC successfully reducing cell sensitivity. Interestingly, YebT overexpression was discovered to be toxic to cells, suggesting that its expression at the correct levels is of importance to membrane stability (68). To provide further evidence for the involvement of Pqi and Yeb systems in PL transport, Ekiert *et al.* performed a chloroform-methanol extraction on purified periplasmic regions of PqiB and YebT. Using thin layer chromatography, PLs were detected in the resulting extracts, which indicates that there is an interaction between PLs and the Pqi and Yeb systems (60).

Analysis has shown that *YebT* contain 7 MCE domains, with 3 of said domains apparent in *PqiB* (60). Negative-stain reconstructions of PqiB and YebT from *E. coli* **(Fig 1.9)** indicated that both of these proteins form membrane-spanning assemblies. YebT was formed of 7 flat, ring like structures formed from a hexamer of YebT polypeptides (60). This near 600 kDa structure was anchored into the IM by 6 α-helices (one per YebT monomer), drawing parallels to the MCE component of the Mla system – MlaD - with its periplasmic hexamer ring and transmembrane helix domains (60). Similar to YebT, PqiB was shown to associate as a hexamer, and formed a 350 kDa stack of 3 MCE rings, with a needle-like projection toward the OM formed from α-helices. Both of these protein structures were around 230 Å in length (60), which suggested that the two could form membrane-spanning PL transport macromolecules, given the intermembrane distance in *E. coli* is thought to be around 220 Å (69). In further support of this, PqiB - anchored into the inner membrane – has

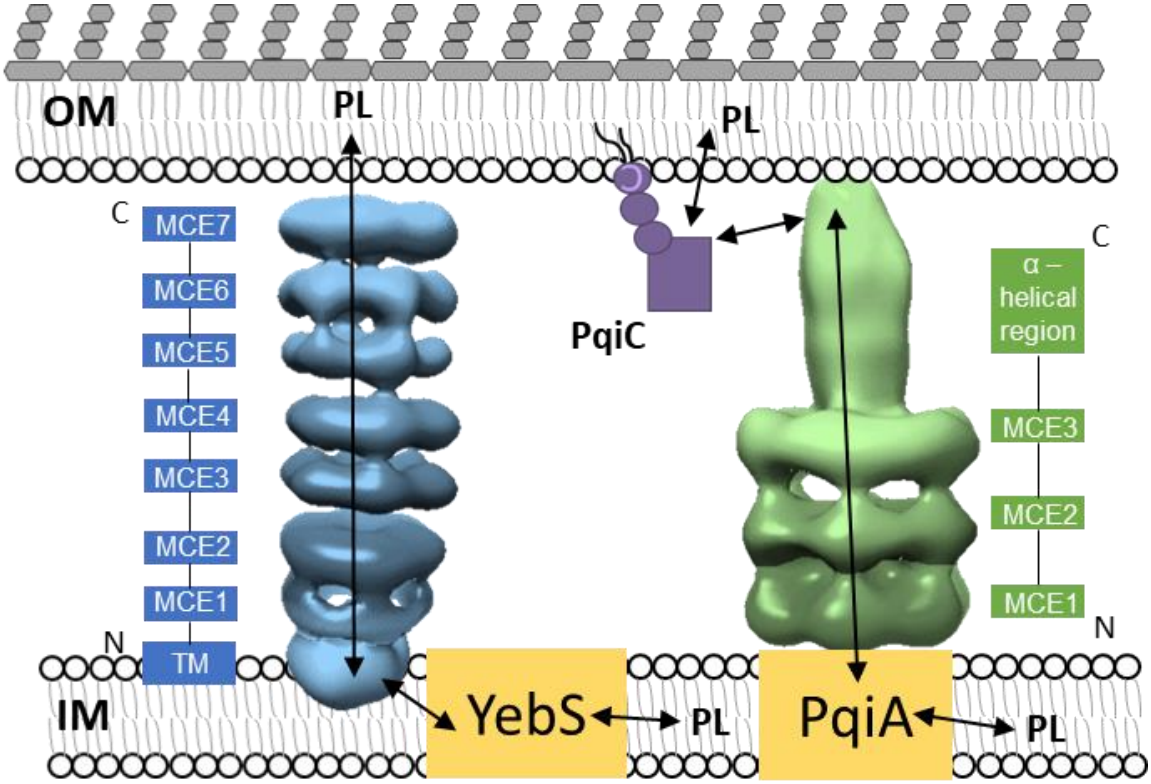


Figure 1.9 - Schematic view of putative phospholipid transport pathways YebST and PqiABC. YebT and PqiB are shown as negative – stain reconstructions to a resolution of 25Å. YebT (blue) is formed of 7 stacked ring-shaped MCE domains as well as a transmembrane helix in each subunit – in the reconstruction these TM helices are attached to a detergent micelle. PqiB (green) takes the form of 3 stacked MCE rings with a C-terminal needle-like region formed from six helices per subunit; PqiB is also attached to the IM through N-terminal helices (not shown). Each of these protein assemblies was determined to span around 230 Å and is likely to cross the periplasm. YebS and PqiA have been shown to be helical inner membrane proteins with 6 TMDs. PqiC (purple) is an outer membrane lipoprotein which has been shown to face the periplasmic space and has been demonstrated to interact with PqiB (9)**.** Phospholipids may be transported across the periplasmic space through the cental cavities of the YebT 'tube' or the PqiB 'needle' (likely route indicated by arrows). The directionality PL trnasport remains unclear. Reconstructions from (60) , with PDBE codes EMD-8611 (YebT) and EMD-8612 (PqiB).

been demonstrated to interact with the OM lipoprotein PqiC, further suggesting that the PqiABC complex can bridge the gap between membranes (68). Transmembrane protein complexes have previously been shown to be involved with hydrophobic molecule transport across the periplasm. The Lpt system also spans the two membranes, so there is clear precedence for the movement of lipid-containing molecules through periplasm-spanning protein complexes. Given their morphologies, it is generally assumed that transported PLs are shielded from the periplasm by transit through the centre of the MCE domains – the YebT 'tube' and the PqiB 'syringe' (60).

Between the presence of MCE domains, the demonstration of phospholipid binding and reduced OM membrane integrity in their absence, PqiABC and YebST have been heavily implicated in the transport of phospholipids between the inner and outer membranes of Gram-negative bacteria. While low-resolution (~25 Å) EM maps exist for YebT and PqiB (alongside higher-resolution EM data (~4 Å) on its MCE domains), the specific structures of YebS, PqiA and PqiC are less established.

PqiA and YebS appear to be similar topologically and are both inner membrane proteins. Both have been shown to consist of 6 helical transmembrane domains, with an additional cytoplasmic helix **(Fig 1.10)** (68). Other than this general topological model though, structural information for YebS and PqiA is unestablished. The proteins appear homologous to one another, with conserved transmembrane structures and comparable molecular weights – PqiA is a 46 kDa protein, and YebS is a 48 kDa protein (70). It is likely that the functions of the two proteins are similarly as linked as their putative structures. Interactions between YebS / YebT and PqiA / PqiB are likely but yet to be validated, not least due to cryo-EM studies of YebT/PqiB

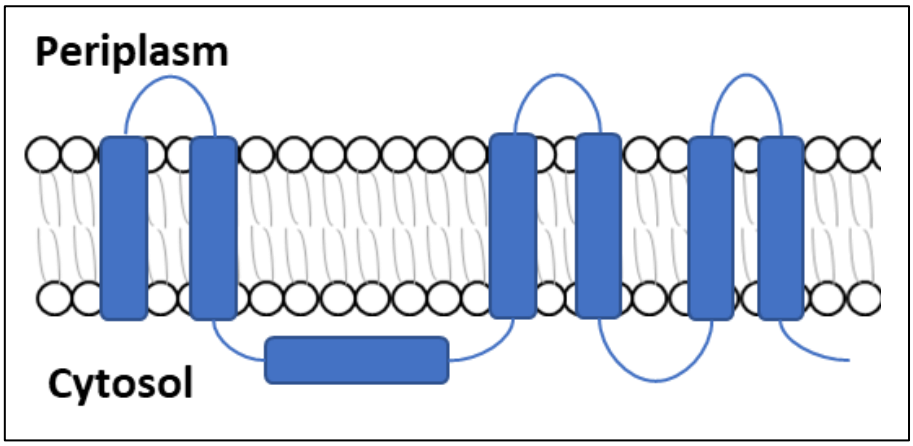


Figure 1.10 – Diagram of shared YebS and PqiA topology from initial predication in (68). YebS and PqiA have been shown to be inner membrane proteins, potentially forming PL transport pathways with YebT or PqiAB. Both proteins have been suggested to take the same general form, with a pair of transmembrane helices, with a single cytosolic helix followed by 4 more helical TMDs.

failing to resolve their N-terminal transmembrane regions which would likely form contacts with the TMDs of interacting partners.

Unlike *YebST*, the *PqiABC* operon contains a gene coding an outer-membrane lipoprotein in the form of PqiC. The *PqiC* gene codes for a 187-residue 20 kDa protein, with the first 15 of these likely to be a signal peptide absent in the mature protein, given the presence of the N-terminal LAGC lipobox motif (71). With no aspartate at the second amino acid residue position, the protein is most likely transported to (and inserted into) the OM through the Lol system (25). Protease accessibility assays have indicated that PqiC is present in the *E. coli* periplasm, and is therefore likely an inward-facing OM lipoprotein (68). As previously mentioned, a yeast two-hybrid assay from the same study have evidenced that PqiC and PqiB interact, and thus PqiC

could be involved with potential PL transfer from PqiB to the OM (68). Other than this, functional and structural information on PqiC remains to be elucidated.

While much remains unknown about the possible PL transport pathways PqiABC and YebST, it is not beyond the realms of possibility that there may be other interacting proteins within the pathway which have yet to be identified. While the needle and tube structures of PqiB & YebT are striking, as mentioned previously the Lpt complex has also been shown to cross this space. The Lpt pathway has an OM component in the form of the LptDE plug and barrel architecture (43) which facilitates the insertion of LPS into the OM. The Lol and Bam pathways of protein translocation likewise contain an OM receptor, in the form of BamABCDE and LolB (11, 72). Indeed, the known PL transport machinery Mla also has an outer membrane portion consisting of the OmpC barrel and MlaA (57). While it is possible that YebST acts without an OM partner, and PqiABC doesn't have an associated β-barrel OMP, it is entirely feasible that such interacting partners exist and are yet to be identified, in the mould of OmpC.

1.6 X-ray Crystallography

Generally, the solution of protein structures is carried out by the techniques of Cryo-Electron Microscopy (Cryo-EM), Nuclear Magnetic resonance (NMR) spectroscopy or X-ray crystallography (73). While Cryo-EM is becoming increasingly popular, X-ray crystallography is still much more widely used in the field of structural biology for the determination of protein structures, and X-ray crystallography is the technique used in the work outlined here (74).

For crystallography to be a viable technique, a large volume of pure protein sample at high concentration is required. From here, the first step in the process of determining a protein structure is – as the name suggests – the formation of crystals. Despite modern high-throughput techniques using commercial screens and liquid-handling robots, this can be a very slow process (75), and the quality of generated protein crystals is difficult to determine visually (76). Once protein crystals have been produced, they are used to diffract X-rays.

The scattering of X-rays by protein crystals is measured and can be used to determine the protein's three-dimensional structure. As protein molecules are packed in the same orientation in a lattice, X-rays are scattered as if from a single molecule, but with much increased intensity (67). Both the intensities and the positions of the X-ray waves scattered by crystals are collected in the form of X-ray diffraction patterns (76). These recorded intensities (reflections) are proportional to the square of the amplitude of the scattered X-rays, known as structure factor (77). To determine the electron density at each point in the crystal structure, the sum of structure factors across the crystal is needed - each recorded reflection is reliant on each atom in the unit cell. However, due to interference between scattered X-ray waves, this electron density is also dependent on the phase of the observed reflections, which cannot be determined from spot locations or intensities (77). Herein lies the phase problem in X-ray crystallography – to determine the structure of the crystallised protein, the phases of the diffracted X-rays are needed, but these cannot be directly measured.

In modern X-ray crystallography, there are three common phasing methods are used for diffraction data: Molecular Replacement (MR), Multi-wavelength anomalous diffraction (MAD) and Single-wavelength anomalous diffraction (SAD). Molecular

Replacement requires an existing molecular model of the same protein or have a C_α rmsd of below around 2 Å – generally a structure with 35% or more identity to the target (78). Using this model, MR algorithms such as MOLREP or PHASER (79, 80) are used to test the search model in different orientations and positions to find the point at which the predicted electron diffraction pattern for the model is closest to the observed diffraction (81). From here, the phases from the search model are extended to the newly collected data to allow an initial electron density map, which can be refined to form a completed protein structure (81).

As previously stated, for MR to be successful, a suitable search model is needed – this is clearly not a viable phasing option when such a model doesn't exist, or if the target protein has no homologues of known structure. In instances such as this, MAD or SAD can be employed as a method to determine phases of experimental X-ray diffraction data.

The diffraction of X-rays by atoms in protein crystals is predominantly represented by the atomic scattering factor, f^0 which is dependent on the diffraction angle of X-rays. If atoms absorb X-rays, the scattering factor is expressed with the addition of wavelength-dependent factors f' and if''. Generally, proteins contain few anomalous X-ray scatterers and this anomalous signal (f' and $if'')$ is small (82). However, if atoms which anomalously scatter X-rays – such as selenium or iodine – are incorporated into the protein, much more anomalous signal will be detected if the wavelength of X-rays nears that atom's absorption edge (77). When anomalous scattering occurs, f' and if'' are not in phase with one another; reflections F(h, k, l) and F(−h, −k, −l) which, when obeying Friedel's law are normally identical now exhibit differences in phase and intensity (83) (82). The difference between the two normally matching

intensities is known as a Bijvoet Difference (84). Using the SAD method - with data collected at a single wavelength - these differences can be used to estimate the positions of anomalous-diffracting atoms in the structure by phasing algorithms such as ShelX (85), and with subsequent density modification (built in to ShelX), this determined substructure can be used to build an electron density for the crystallised protein (86). Using MAD phasing, data is instead collected from crystals at several wavelengths, to maximise absorption and dispersion of X-rays (*f'* and *f'')* (86). In comparison to SAD, this technique can result in crystal damage from repeated X-ray exposure. From the two anomalous differences, phase information can be calculated and substructure detected, again using software such as ShelX (85). Once the phase problem has been solved, the observed structure factors can be used to calculate electron density using a Fourier Transform (76).

Structural models can be built manually into the calculated electron density, but are often constructed with model-building software such as Buccaneer (87). This initial model (whether manually or automatically built) will need refinement to ensure the built protein matches the electron density, and that the amino acids and their bonds are in energetically feasible conformations (88). This refinement can be carried out using automated software such as REFMAC5, as well as manually in the COOT viewer (89, 90). To determine the quality of a refined protein structure, and how well it corroborates observed reflection data, the R-factor can used. This is calculated by measuring the percentage difference between the observed structure factor amplitudes from diffraction data, and that determined from the current model; as a general rule of thumb, R factors of below 25 % and considered indicators of well-refined structures (91). However, given that the R factor can manipulated, and structures with acceptable R values have been shown to contain serious errors (92),

R-free is often used in its place as a measure of model quality (93). In R-free determination, 5-10% of the data is set aside as a test set, and from this data an R-free value is determined for the current structure. This takes place using the same calculation as for the standard R factor. R-free is therefore a measure of how well the current protein model predicts the diffraction data in the R-free dataset, which was not used to build the model (94).

Given that R-free is not determined from the data used to build the model, it is preferred as a model quality measure. Values below 30% considered indicative of a 'good' model – this value indicates the percentage difference between the diffraction intensities observed from diffraction, and the diffraction intensities predicted from the refined model (94).

1.7 Specific Aims

The following work aims to elucidate the structure and function of members of the PqiABC and YebST pathways in the generation and maintenance of the inner leaflet of the outer membrane in Gram-negative bacteria.

This work intends to determine the structure for PqiC using X-ray crystallography, to help better understand its involvement – and that of the PqiABC pathway as a whole – in phospholipid trafficking in *E. coli*. In addition, it aims to purify, crystallise, and determine the structure of YebT and YebS, to investigate the role of the YebST pathway in the transit of phospholipids between inner and outer membranes.

CHAPTER 2 – Structural and Functional Characterisation of Putative Phospholipid Transport Pathway PqiABC

2. 1 – Materials and Methods

2.1.1 Amplification of *PqiC* Gene

For the overexpression of PqiC in *E. coli* cells, the *PqiC* gene was amplified from *E. coli K12* genomic DNA. All primers were analysed with basic bioinformatic tools to ensure each pair's melting temperatures were within 5°C and that the risk of primer dimerization was minimal. For the first cloning of *PqiC*, sites for *NcoI* and *HindIII* were incorporated into the forward and reverse primers respectively. With the initial expression plasmid of pET28b, the reverse primer also included codons to introduce a hexa-his tag to expressed protein; these tags were incorporated to allow nickel-affinity purification. Subsequent expressions used a separate primer pair with *NcoI and BamHI* restriction sites, without the need for introduced histidine codons. All primers were synthesised by EuroGentec; full sequences of all primers are available in **Table 2.1**.

In each PCR reaction carried out, the Q5 hot-start DNA polymerase (New England Biolabs) system was used, with individual component volumes listed in **Table 2.2**; the 10mM dNTP mixture was produced from individual dNTPs (dATP, dTTP, dGTP and dCTP) from Thermofisher Scientific. For the first round of PCR, 21 ng of *E. coli* gDNA was used as a template, with subsequent rounds using 30 ng of the constructed pET28b:*PqiC* plasmid. The PCR program used for each reaction consisted of an initial denaturation of 98°C for 30 seconds, followed by 34 cycles of: 98°C for 10 seconds; Tm°C (primer melting temperature) for 30 seconds and elongation phase of 72°C for 30 seconds. Each reaction concluded with a final elongation phase at 72°C for 120 seconds, before holding at 4°C. Melting temperature (Tm°C) varied between primer pairs, with 65°C used for full-length *PqiC* and 61°C for truncated *PqiC*. PCR

reactions were carried out in a Bio-Rad thermal cycler; for multi-sample PCR reactions a master mix of all shared components was made and distributed to each required tube.

Following PCR reactions, samples were analysed using agarose gel electrophoresis to determine the length of the amplified DNA products. To each 25 µl PCR reaction, 5 µl of DNA loading dye (Thermofisher) was added, before loading onto a 40 mL 1% Agarose-TAE (Tris-acetate EDTA) gel containing 2.5 µl Ethidium Bromide as a DNA dye. Each gel included 5 µl of GeneRuler 1kb plus ladder (Thermofisher) and was run in 1x TAE for 45 minutes at 110 volts. Gels were imaged by transillumination with UV light, using the self-optimisation feature on a Bio-Rad EZ GelDoc system. Once gels

Table 2.1 – Primers used in cloning components of the PqiABC pathway. The primers shown here were used to clone genes from the PqiABC operon and perform site-directed mutagenesis on PqiC.

Primer Name	Sequence (5'-3')
PqiC_FL_F	CTGGTACATATGAAAAAGTGGCTAGTGACGATTGCAGC
PqiC_FL_R	GAGGCAGGATCCTTAGTGATGGTGATGGTGATGAGGTAGACGCTTTATCTCTTGTGC
PqiC_Trunc_F	CTACTACCATGGCTTGCAGCTCCGGCGAAATTAATAAAAACTATTACC
PqiC_Trunc_R	GAGGCAGGATCCTTAAGGTAGACGCTTTATCTCTTGTGCAATAGAAGC
PqiC_Y64E_F	GTGGTTGAACAAACCAGTGATGTGAAGTATGTGATTGC
PqiC_Y64E_R	GGTTTGTTCAACCACACCATTCCCCGCCAG
PqiC_W79E_F	CAACTTGGAGGCCAGCCCGTTGGATCAACAGTTGCG
PqiC_W79E_R	CTGGCCTCCAAGTTGTTGTTGGCAATCACATACTTCACATCAC
PqiC_F123E_F	TAACGGAGGAGAACGGTCGCTATGATGGCAAGG
PqiC_F123E_R	CGTTCTCCTCCGTTACGGTAACATTGAGCGTGTC
PqiC_Y161E_F	TGGTGAAGATGAGATGGTTAAAGTGCTGGCCGGTGTCTG
PqiC_Y161E_R	CTTTAACCATCTCATCTTCACCA TCCTGAGTTTGCACTCCTTCCAG
PqiC_M164D_F	TGGTTACGATGAGGACGTTAAAGTGCTGGCCGGTGTCTG
PqiC_M164D_R	CTTTAACGTCCTCATCGTAACCATCCTGAGTTTGCACTCCTTCCAG
PqiA_F	ATTACATATGTGCGAACATCATCATGCCG
PqiA_R	AATTAAGCTTTCAGTGATGGTGATGGTGATGGGACTCCTCATGCTCTGATTC
PqiAB_F	ACTAGTCATATGATGTGCGAACATCATCATGCCGC
PqiAB_R	GAGACTAAGCTTTCAGTGATGGTGATGGTGATGTTGTTTCGCCCTCTTCGGCTCTG
PqiABC_F	ACTAGTCATATGATGTGCGAACATCATCATGCCGC
PqiABC_R	GAGGACAAGCTTTTAGTGATGGTGATGGTGATGAGGTAGACGCTTTATCTCTTGTGC

were imaged, DNA bands were excised using a scalpel under UV illumination in a darkroom and transferred to 1.5 mL microcentrifuge tubes. DNA was extracted from the gel slices using a GeneJet Gel Extraction Kit (Thermofisher), with a final elution volume of 30 µl. DNA concentration was assayed using a Nanodrop 2000c set to

Table 2.2 – Components of PCR mixture for gene amplification and mutagenesis work. Using the Q5 Hot Start system (NEB), PCR reactions were carried out using the below volumes of reaction components.

Component	Volume Added (µl)
5X Q5 Reaction Buffer	5.00
Q5 Hot Start DNA Polymerase	0.25
10 mM dNTPs	0.50
10 µM Forward Primer	1.25
10 µM Reverse Primer	1.25
Template DNA	1.00
Nuclease-Free Water	15.75

nucleic acid detection and stored at -20°C.

2.1.2 Generation of *PqiC* Expression plasmids

To allow its transformation into a cell line suitable for protein overexpression, the amplified *PqiC* gene was ligated into an expression plasmid. To enable this, *PqiC* was double-restricted using FastDigest restriction enzymes (ThermoFisher) appropriate for the selected primer pairs: *NcoI /HindIII* for full length and *NcoI /BamHI* for truncated. To 17 µl of DNA sample, 2 µl of FastDigest buffer and 0.5 µl of either restriction enzyme were added. For the corresponding digestion of plasmid DNA, 10 µl of DNA was added to 2 µl of FastDigest buffer, 0.5 µl of either restriction enzyme and 1 µl of alkaline phosphatase, to a total volume of 20 µl. For both digestions,

mixtures were incubated for 30 minutes at 37°C, followed by thermal inactivation for the restriction enzymes at 85°C for 5 minutes.

Following restriction digestion*, PqiC* and plasmid samples were purified prior to ligation using the GeneJet Gel Extraction and DNA Cleanup kit (ThermoFisher), with final elution volume of 30 µl. Ligation reactions used a 3:1 ratio of gene to plasmid, with sample concentrations determined with a Nanodrop 2000c. To a 9 µl aliquot of plasmid-gene mix, 1 µl Quick Ligase (NEB) and 10 µl Quick Ligase Buffer (NEB) were added, with reactions taking place for 30 minutes at room temperature.

After *PqiC* was ligated into its recipient plasmids, these plasmids were amplified in Top10 *E. coli* cells. To transform constructed plasmids into Top10, a heat shock protocol was used, adapted from Inoue *et al.* (95). Samples of 50 µl of chemically competent cells were added to 10 µl of the ligation mixture for transformation. The cells were then incubated on ice for 30 minutes, prior to a heat shock at 42°C for 70 seconds. After a 5-minute incubation on ice, 700 µl of LB media was added and samples were incubated at 37°C with shaking at 200rpm for 1 hour for recovery. After the recovery phase, the cells were spun at 3000rpm for 5 minutes, with the pellet resuspended in 100 µl of supernatant and plated on 1.5% agar - LB supplemented with 100 µg/ml ampicillin for pLou3 transformations, or 50 µg/ml kanamycin for pHistev and pET28b transformation. Plates were incubated overnight at 37°C, with single colonies used to inoculate 10 ml aliquots of LB medium with appropriate antibiotics the next day. Inoculations were incubated overnight at 37°C with 200rpm shaking. Plasmids were extracted from the overnight cultures using the GeneJet Plasmid miniprep kit (ThermoFisher), using a final elution volume of 50 µl.

2.1.3 Validation of *PqiC* Expression plasmids

To determine the presence of a successful *PqiC* insert in novel plasmids, 2 µl samples of prepared plasmid were double-restricted using 0.5 µl of each suitable Fast Digest enzyme and 2 µl Fast Digest Buffer, in a total volume of 20 µl. Samples were incubated for 30 minutes at 37°C , then 85°C for 5 minutes. Each restriction mixture was run on an agarose gel and imaged as previously described. Samples with detectable inserts of suitable length were submitted for sequencing, with 5 µl of 100 ng/ml plasmid sequenced by Source Biosciences.

After successfully cloned *PqiC* plasmids had been identified, they were transformed into *E. coli* C43 (DE3) or *E. coli* SoluBL21 (DE3) cells for overexpression. This was carried out using the same transformation procedure previously outlined. Single colonies from the resulting plates were used to inoculate 10 ml of LB medium and grown overnight at 37°C 200rpm. An aliquot of 250 µl overnight culture was mixed with 750 µl glycerol, flash frozen in liquid nitrogen and stored at -80°C for future inoculations.

2.1.4 Overexpression of PqiC

Full length PqiC was expressed in *E. coli* C43 (DE3), with truncated PqiC expressed *E. coli* SoluBL21 (DE3), using IPTG induction (96). Frozen glycerol stocks were used to inoculate 300 ml LB media supplemented with 100 µg / ml ampicillin or 50 µg / ml kanamycin, which was grown overnight at 37°C, with 200rpm shaking. Overnight culture was used to inoculate fresh antibiotic-containing LB media, using 25ml overnight culture per litre. Cells were grown at 37°C with 200rpm shaking. Culture was periodically assayed with a spectrophotometer at 600nm, and cells were grown

until OD600 was 0.6 – 0.8. Protein expression was induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.2mM – expression time and temperature was varied, with 16 hours at 20°C commonly used. Cells were harvested using centrifugation – culture was spun for 20 minutes at 4,000 g using a JLA 8.1000 rotor (Beckman Coulter). Wet cells were stored at -20°C until purification.

2.1.5 Purification of PqiC

To purify protein from frozen cell pellets, they were resuspended at room temperature in 1ml Lysis buffer per gram of cells. Protein purification methodology was developed using the work of Ekiert *et al.* as a starting point (60). The buffer contained: 20mM Tris-HCl pH8, 300mM NaCl, 10mM Imidazole, 1 cOmplete EDTA-free protease inhibitor tablet (Roche), 1 µg/ml DNase I (SigmaAldrich) 100 µg/ml lysozyme (SigmaAldrich). Cells were lysed twice by two passes at 30k psi through a Cell disrupter (Constant Systems) cooled to -5°C. Intact cells and cellular debris were pelleted via centrifugation for 30 minutes at 30,000 g with JA 25.50 rotor (Beckman Coulter).

Supernatant from this spin was loaded onto a 5 ml HP HisTrap column (GE HealthCare), pre-equilibrated to binding buffer containing 20mM Tris-HCl pH8, 300mM NaCl and 10mM Imidazole. Clarified supernatant was loaded onto the column using a peristaltic pump at 5 ml/min, with the flowthrough loaded again. The column was washed with 50 ml binding buffer with 50mM imidazole at 5 ml/min. Protein was eluted from the column in 1.5 mL fractions using 20 ml of binding buffer with 300mM imidazole and stored on ice. Protein concentration in fractions was determined using a Nanodrop 2000c, with the first fraction with over 10 mg/ml protein to the last fraction above 1 mg/ml pooled. Pooled PqiC was concentrated to

5 ml using a 5kDa MWCO Vivaspin 20 (Sartorius) spin concentrator at 3,000 g and 4°C. Concentrated protein was desalted to allow for optimum protease activity for affinity tag removal. A HiPrep 26/10 column (GE) was used, run on an ÄKTA Purifier (GE) at 2.5 ml/min into 300mM NaCl, 20mM Tris pH8. Desalted PqiC was treated overnight with rocking at 4°C, using 1 mg TEV proteinase per 50 mg protein to remove the hexa-histidine tag.

TEV-cleaved PqiC was separated from uncleaved protein using a pulldown through a 5 ml HisTrap HP column (GE); protein still attached to its affinity tag would still bind to the column matrix. The column was equilibrated to binding buffer, and the protein mixture was supplemented with 20mM imidazole before being run through the column 3 times at 5 ml/min using a peristaltic pump. The pulldown was concentrated to 5 ml using spin concentration as previously. The sample was further purified using SEC, with an ÄKTA Purifier and HiLoad 16/60 Superdex 200 prep grade column (GE) equilibrated to 300mM NaCl, 20mM Tris pH8 buffer. The sample was run at 1 ml/min in the same buffer used for equilibration for 2 hours and separated into 1.5 ml fractions.

Steps of the purification were analysed using Polyacrylamide Gel Electrophoresis (PAGE) to separate constituent proteins by molecular weight. A 10 µl aliquot of each sample was added to 10 µl and 4 µl of Bolt LDS buffer and reducing agent respectively (ThermoFisher), to a final volume of 40 µl. Samples were heated at 85°C for 2 minutes, and subsequently loaded onto a Bolt precast 4 to 12%, Bis-Tris gel (ThermoFisher), along with 5 µl of PageRuler Prestained protein ladder (ThermoFisher). Gels were run for 22 minutes at 200 V, stained for 30 minutes in

Quick Coomassie (Generon) and rinsed overnight with rocking in water. Imaging was performed using a Bio-Rad EZ GelDoc.

2.1.6 PqiC Crystallisation Trials

After successful PqiC purification, protein was concentrated using spin concentration from 5-15 mg/ml, as determined by a nanodrop 2000c. Initial crystal screening was performed using commercial crystallisation screen kits for soluble proteins from Molecular Dimensions and Hampton Research, with a full list of crystallisation screens in **Table 2.3**. Trials used sitting drop vapour diffusion method on 2-well MRC plates (SwissSci), with top well volume of 1 µl and bottom well volume 0.6 µl; both wells used 1:1 protein: screen solution. Crystal plates were prepared using a Crystal Gryphon robot (Art Robbins Instruments), sealed with ClearVue sheets (Molecular Dimensions) and incubated at 21°C.

Promising crystallisation conditions arising from initial screen plates were further optimised using optimisation screens. These screens were hand-made in deep-well blocks, using successful well conditions with pH and precipitant concentration varied across the wells. As for commercial screens, these optimisation screens were used to make sitting-well vapour diffusion plates with drop sizes of 1 µl and 0.6 µl.

2.1.7 Data Collection from PqiC crystals

Protein crystals selected for data collection were soaked in cryoprotectants to prevent them from being damaged by flash freezing. Cryoprotectants were made by matching the contents of the mother liquor supplemented to 20% glycerol and with or without the addition of 200mM Sodium Iodide for phasing. Crystals were soaked in cryoprotectant for between 5 and 15 minutes. Following buffer soaking, individual

crystals were collected using 0.2mm Litholoops (Molecular Dimensions) and flash frozen in liquid nitrogen. Loops were inserted into cooled Unipucks (Molecular Dimensions) and transported to Diamond Light Source in a Cryo-Express Dry Shipper (Molecular Dimensions) for data collection using beamlines I03 and I04. Native data were collected at a wavelength of 0.9795Å, with anomalous Iodide scattering data collected at 1.8223Å. In each case, 3600 images were collected at an exposure of 0.1 seconds; 25% and 100% beam transmission were used for anomalous and native data collections, respectively.

Table 2.3 – Crystallisation screens used in soluble protein crystal trials. Each screen kit was used to produce a 96-condition screen in a deep well block suitable for use with the Crystal Gryphon liquid handling robot.

Crystallisation Screen	Manufacturer
Morpheus®	Molecular Dimensions
PACT Premier™	Molecular Dimensions
The PGA Screen™	Molecular Dimensions
Structure Screen 1	Molecular Dimensions
Structure Screen 2	Molecular Dimensions
The BCS Screen	Molecular Dimensions
LMB Crystallization Screen	Molecular Dimensions
JCSG Plus™	Molecular Dimensions
PEG/Ion	Hampton Research
PEG/Ion 2	Hampton Research
PEGRx 1	Hampton Research
PEG Rx2	Hampton Research
Index	Hampton Research

2.1.8 PqiC Data Processing

PqiC X-ray data were processed using the CCP4 programme suite, through the CCP4i2 graphical user interface (97) (98). SAD datasets taken at the iodine absorption edge were processed and used for structure determination first. X-ray diffraction images were integrated with DIALS, using the Xia2 program – this processing step measures the intensity of the diffracted X-rays (99) (100). Data reduction and merging was carried out using aimless, to normalise data and combine symmetry-related measurements. Resolution was limited to 2.79 Å and data assigned to spacegroup $P6_1$ (101). The merged data in the form of a .mtz file output by aimless were used for initial model building with the Crank2 phasing and building pipeline (102).

2.1.9 PqiC Structure Solution and Validation

Substructure determination and improvement was carried out using 2000 trials of ShelX (85); this assigned positions to Iodine atoms, yielding phase information which could be extended to PqiC itself. Three iterations of Solomon were used for further phase improvement and hand determination (103). Density modification and model building were performed by Parrot and Buccaneer, with 25 cycles of density modification in each of 50 model building cycles (87). Refmac5 was utilised for refinement of the initial model using 100 cycles, to better fit the known amino acid sequence into the calculated electron density (89). Missing residues were manually added using COOT, with further structure refinements also made. Bond angles were adjusted in accordance with the Ramachandran Plot, again using COOT, to ensure that they were of an energetically permitted nature (90).

A higher resolution PqiC model was built using this refined structure as a molecular replacement model. MOLREP was used for this, with initial refinement carried out using Refmac. Following manual refinement in COOT, water molecules were added and refined using ARP/wARP and Refmac5 (104). Structures derived from native data were refined and adjusted in the same manner as the experimentally phased data. Small regions of protein absent from the iodine-phased model were built into the appropriate electron densities using COOT (90). Structure validation was performed by MolProbity, with scores improved by manual adjustments using COOT (88).

2.1.10 Primer Design for Site-Directed Mutagenesis of PqiC

To determine the involvement of specific residues in the function of PqiC, several mutant *PqiC* genes were generated with single residue substitutions. Using the determined PqiC structure, five positions were selected for mutagenesis. Each of these mutations took the form of individual amino acid substitutions which were chosen as potential functional sites for phospholipid binding. The five mutants were: Y64E, W79E, F123E, Y161E and M164D – in each case a larger, hydrophobic sidechain was replaced with a charged Glutamate or Aspartate residue. Primer pairs were designed in accordance with Liu's protocol (105), with an overlapping region containing the point mutation of melting temperature around 10°C less than that of the non-overlapping complementary regions. Primers were synthesised by EuroGenTec, with full sequences available in **Table 2.1.**

2.1.11 Site-Directed Mutagenesis of PqiC

For each given mutant, a full-length gene was mutated for *in vivo* functional studies. In addition, a gene lacking a signal peptide sequence was produced for each mutant,

to be used for protein purification and *in vitro* work. The previously constructed pHistev:*PqiC* plasmids (with and without the signal peptide region) were used as templates for the site-directed mutagenesis reactions. Mutations were introduced by amplifying the whole pHisTev plasmid using the Q5 Hot Start PCR system (NEB) (reaction mixture in **Table 2.2**), with overlapping primers containing the desired mutations. The standard PCR programme was modified for mutagenesis as per Liu's protocol (105), and consisted of: 98°C for 5 minutes, followed by 15 cycles of: 95°C for 1 minute; 5°C below melting temperature for non-overlapping region for 30 seconds; and an elongation phase of 72°C for 15 minutes. Each reaction concluded with a 1-minute incubation at 5°C below overlapping region melting temperature and a final phase of 72°C for 30 seconds before a 4°C hold. To degrade the *E. coli*-grown template DNA, PCR products were treated with the restriction enzyme *DpnI,* which recognises methyl groups – these are not present in the new mutant plasmids generated by PCR.

To confirm that plasmids had been amplified successfully, reaction mixtures were analysed with gel electrophoresis. To 17µl of each PCR reaction, 2µl of FastDigest Buffer and 1µl of FastDigest *DpnI* (both Thermofisher) were added, with the mixture incubated at 37°C for 1 hour, followed by 5 minutes at 85°C. For each mutant, 5 µl samples of *DpnI*-treated DNA were added to 5 µl of DNA loading dye and 20 µl water and run on a 1% agarose gel for 45 minutes at 110V. Samples with bands apparent at around 6kbp had successfully amplified. Successful amplifications were used to transform Top10 *E. coli* cells to generate enough plasmid for confirmatory sequencing. For each, 10 µl of the *DpnI*-treated sample was used to transform cells using the heat shock method as previously described. Colonies of transformed cells were used to inoculate 10 mL samples of LB, which were incubated overnight at 37°C,

200rpm. Plasmids were purified from Top10 cells using a GeneJet Plasmid miniprep Kit (Thermofisher) and submitted to Source Biosciences for sequencing to confirm successful *PqiC* mutagenesis.

2.1.12 Restriction and Ligation of mutant *PqiC* genes

For use in cell-based assays, the successfully mutated full-length *PqiC* genes were inserted into a different cloning vector. For this, confirmed mutant genes were removed from their existing pHisTev plasmid and inserted into the pJH113 plasmid. Plasmids containing full-length mutant *PqiC* were restricted with *NdeI & BamHI* – to 5 µl of plasmid 2 µl of FastDigest Buffer and 1 µl of either FastDigest restriction enzyme was added in a final volume of 20 µl. After a reaction of 50 minutes at 37°C, restricted plasmids were run on a 1.5% agarose gel containing ethidium bromide for 45 minutes at 110V to separate *PqiC* from the remaining DNA. Mutant *PqiC* bands were cut from the gel, and DNA purified using a GeneJet Gel Extraction Kit (Thermofisher). Genes were ligated to into the pJH113 plasmid (which was cleaved in the same manner) in a 9 µl aliquot with a ratio of 3:1 gene to plasmid, in a reaction mixture containing 10 µl Quick Ligase Buffer and 1 µl Quick Ligase (NEB). Ligated plasmids were transformed via heat shock into Top10 *E. coli* cells to replicate the plasmid. Newly produced plasmids were extracted from Top10 cells using the GeneJet Miniprep Kit. Success of transformation into pJH113 was confirmed with a double restriction and sequencing run, as previously described.

2.1.13 Mutagenesis functional Assay in PqiC Deletion Strain

The *E. coli* K12 strain JW5127-1 was obtained from the Keio Collection (106). This strain has a kanamycin resistance, and the *PqiC* gene has been deleted from its

genome: this PqiC knockout line was used for cell-based assays. The cells delivered from the Keio Collection were grown overnight in LB containing 50 µg/ml kanamycin at 37°C with 200rpm shaking. The grown cells were then tested for sensitivity to the drug chlorpromazine - 5 mL of LB with kanamycin was inoculated with 500 µl of overnight culture and grown at 37°C, 200rpm until OD_{600} reached 0.5. Cells were diluted in fresh LB with kanamycin in series from 10^{-1} to 10^{-7}. A 2 µl sample of each dilution was plated on 1.5% Agar-LB plates with kanamycin, with or without 120 µg / ml of Chlorpromazine and incubated overnight at 37°C. The remaining overnight culture was flash frozen in liquid Nitrogen with glycerol to a final concentration of 20%.

To determine whether growth of the PqiC cell line could be recovered by each of the mutant PqiC species generated, plasmids containing these mutant genes were transformed into separate aliquots of JW5127-1. To allow transformation using heat shock, aliquots of chemically competent Δ*PqiC* cells were prepared using a modified protocol derived from Chang *et al.* (107). Frozen JW5127-1 cells were used to inoculate a 10 mL sample of LB with 50 µg/ml kanamycin; this was grown overnight at 37°C, with 200rpm shaking. A 100 mL sample of LB with kanamycin was inoculated with 1 mL of overnight culture and grown for 2 hours at 37°C, 200rpm. Cells were chilled on ice for 30 minutes, then centrifuged for 5 minutes at 2,500 g, 4°C. The resulting cell pellet was resuspended in 25 mL of buffer containing 50 mM $CaCl_2$ and 20 mM $MgSO_4$ and chilled on ice for a further 30 minutes. The solution was spun again for 5 minutes at 2,500 g, 4°C with the resulting pellet resuspended in 5 mL of 50 mM $CaCl_2$ 20 mM $MgSO_4$ buffer. Sterile glycerol was added to a final concentration of 10% and cells were flash frozen in liquid Nitrogen in 50 µl aliquots and stored at -80°C for further use.

All 5 full-length *PqiC* mutant genes, as well as the wild type gene in the plasmid pJH113 were transformed into competent JW5127-1 cells using the heat shock method. With chlorpromazine appearing to inhibit growth of *PqiC* knockout cells, the newly transformed cell lines were used for a cell-based assay. The same method was used as for the initial chlorpromazine sensitivity test, with cells grown for 20 hours at 37°C prior to plate imaging.

2.1.14 Purification of Mutant PqiC protein

For use in *in-vitro* assays, mutant PqiC protein was overexpressed and purified. As with wild type protein, mutant PqiC lacking its signal peptide was expressed in *E. coli* SB21 (DE3) cells using the pHisTev plasmid. Protein was expressed in volumes of 6 litres per mutant for 16 hours at 20°C with IPTG induction of 0.2mM. Cells were harvested, lysed, and spun in the same manner as WT protein. Mutant protein was purified using IMAC, again following the same methods as the wild type. Following elution from the IMAC, protein was concentrated to a volume of 5mL using a 3 kDa MWCO spin concentrator. Concentrated protein was run though a 16/600 200pg SEC column at 1 ml/min in 20mM Tris-HCl pH8 300mM NaCl buffer. Fractions containing protein were collected and concentrated to around 1 mg/ml, before flash freezing in liquid Nitrogen in 50 µl aliquots and storage at -80°C. These purifications resulted in a mutant PqiC protein with a hexa-his tag still attached.

2.1.15 PqiC Phospholipid Binding Assay

To determine the ability of both wild-type and mutant PqiC to bind phospholipids, an enzyme-linked immunosorbent assay (ELISA) was carried out, with PLs immobilised on a plate investigated for their ability to interact with PqiC (108). Non-

treated 96-well cell culture plates were coated with 50 µg *E. coli* total phospholipids in chloroform (Avanti) per well overnight at 4°C. Plates were then washed 5 times with PBS supplemented with 0.05 % Tween-20 (PBST). PqiC was added to each well in PBS with 5 % milk, in a dilution series from 0.1-100 µg per well in columns 1-12. Protein was incubated in the PL-coated plates for 2 hours at 37°C, before its removal with 5 PBST washes. Mouse α-His antibodies were added to wells diluted 1:200 in 5 % milk; 100 µL per well was used. After a 16-hour incubation at 4°C, wells were washed 5 times in PBST, before the addition of 75 µL per well of goat α-mouse-HRP conjugated antibody to wells, diluted 1:3000 in 5% milk PBS. After a 2-hour incubation at room temperature, wells were washed 5 times with PBST. For detection of the secondary HRP conjugated antibody, 50 µL per well of 3,3',5,5'-Tetramethylbenzidine was added and incubated for 30 minutes at room temperature with gentle mixing. The reaction was quenched with 50 µL per well of 0.18 M H_2SO_4, with plates' absorbances at 450 nm determined with a plate reader. Higher absorbance values point to a stronger interaction between PqiC and the immobilised PLs.

2.1.16 PqiA and Pqi Complex expression plasmid generation

For investigations on the other components of the Pqi pathway, they were initially cloned from genomic *E. coli* DNA in the same manner as *PqiC*. PqiA, PqiAB and PqiABC genes were amplified and cloned into the pHisTev plasmid for expression, with the incorporation of C-terminal his tags on PqiA, PqiB and PqiC, respectively. Primer pairs were designed using a common forward primer with restriction site NdeI, and separate reverse primers with the HindIII site; full sequences available in **Table 2.1**.

PCR reactions to amplify the genes were carried out with the Q5 hot-start system (NEB) as previously outlined, with reaction components in **Table 2.2**. The PCR program used for each reaction was: 98°C for 30 seconds, followed by 34 cycles of: 98°C for 10 seconds; 69°C for 30 seconds and 72°C for 3 minutes. Reactions concluded with incubation at 72°C for 120 seconds, and were held at 4°C. Completed reactions were supplemented with 5 µl DNA loading dye, then run on a 1% agarose gel with 2.5 µL EtBr for 45 mins at 110V to determine PCR success. Bands of desired size were cut from the gel with a scalpel, with DNA extracted using a GeneJet Gel Extraction Kit.

Purified genes and pHisTev were restricted with NdeI and HindIII using the Fast Digest system and ligated using the Quick Ligase system in a ratio of 3:1 gene: plasmid. Ligated plasmid was transformed into Top10 *E. coli* cells with heat shock, and plasmid extracted using a GeneJet miniprep kit. Gene insert presence was confirmed with a further double restriction and gel electrophoresis test, followed by sequencing at Source Biosciences. Validated plasmids were transformed into expression cells using the heat shock method previously described.

2.1.17 Expression of Pqi Complex

PqiA, PqiAB and PqiABC are membrane-associated proteins, are were expressed in *E. coli* C43 (DE3) cells, which are better suited for membrane protein expression than the cell lines used for PqiC expression (109). LB media supplemented with 50 µg / ml kanamycin was inoculated with frozen cell culture and grown overnight at 37°C, with 200rpm shaking. Fresh antibiotic-supplemented culture was inoculated with 25 mL culture per litre, and grown at 37°C, 200rpm to an OD600 of 0.6. Protein expression was induced with 0.2 mM IPTG, and cells grown for a further 16 hours at 20°C before

harvest by centrifugation at 4,000 g for 20 minutes. Cells were stored at -20°C prior to purification.

2.1.18 Purification of Pqi Complex

As membrane – localised proteins, PqiA, PqiAB and PqiABC required a modified purification procedure and the inclusion of detergents in their buffers. The Pqi complex was purified using methods adapted from Dong *et al. (44).* Harvested cell pellets from overexpression were resuspended in lysis buffer (300 mM NaCl, 20mM Tris – HCl pH 8, 10 mM Imidazole) supplemented with 1 cOmplete EDTA-free protease inhibitor tablet, 1 µg/ml DNase and 100 µg/ml lysozyme. Cells were lysed twice at 30 kPsi, then spun for 30 minutes at 30,000 g. The supernatant was collected, and the fraction pelleted with centrifugation for 1 hour at 100,000 g with a Type 45 Ti rotor (Beckman Coulter). The membrane pellet was resuspended in lysis buffer containing 1% n-Dodecyl β-D-maltoside (DDM) with a plunger homogeniser. This resuspension was mixed gently for 16 hours at 4°C to solubilise membranes in detergent micelles. The solubilised solution was spun once more for an hour at 100,000 g, with the solubilised supernatant used for IMAC.

The sample was loaded onto a 5 mL HP HisTrap column equilibrated to lysis buffer with an additional 1% DDM. The column was washed with 50 mL lysis buffer supplemented with 50 mM Imidazole and 0.05% DDM. Protein was eluted in 25 mL lysis buffer with 300 mM Imidazole and 0.05% DDM. Samples from throughout the process were analysed with a PAGE gel run using the Bolt system.

For PqiAB and PqiABC, samples were concentrated to 5 mL by spin concentration at 3,000 g in a 10kDa MWCO Vivaspin 20 concentrator. The complexes were purified

using SEC with a 16/60 Superdex 200 pg column run at 1 mL/min in 20mM Tris pH 8, 150mM NaCl, 0.05% DDM buffer. Protein was collected in 1.5 mL fractions and analysed using PAGE.

2. 2 – Results

2.2.1 Cloning of the *PqiC* gene

The first step for structural characterisation of the PqiC protein was the cloning of its gene from *E. coli* genomic DNA. Suitable primers were designed **(Table 2.1)**, incorporating sites for restriction enzymes *NcoI* and *HindIII.* Following 34 cycles of the desired PCR program, the product was run on an agarose gel containing EtBr and imaged under Ultra-violet transillumination (**Fig2.1)**. Initial signs were promising, with strong bands apparent at around the 600bp expected for *PqiC*. Putative *PqiC* bands were excised from the gel to allow for extraction of the gene, before ligation into plasmid pET28b. After amplification of this novel expression plasmid, it was assessed with a double-restriction digest using *NcoI* and *HindIII* **(Fig2.2)**. When investigated using agarose electrophoresis, the double-restriction showed bands at both the expected size of pET28b and that of *PqiC.* This already suggested cloning success, which was further validated with results obtained through Sanger sequencing, **(Fig2.2)** which yielded a sequencing read with only single mismatched base to the known *PqiC* sequence. This mismatch changes a codon of AAG to that of AAA, and as such is of minimal significance to an expressed protein, as both would result in a lysine residue at the corresponding position.

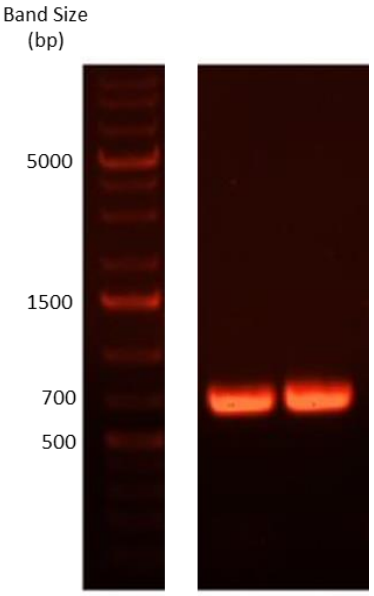
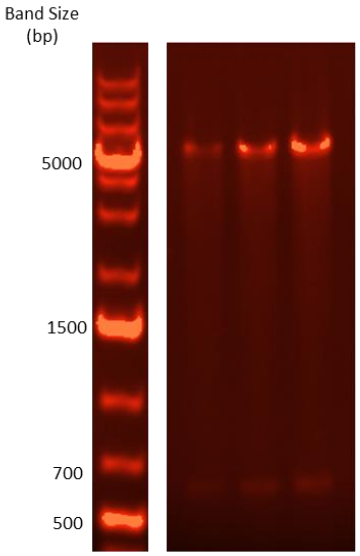


Figure 2.1 –Agarose gel of full length *PqiC* PCR products containing EtBr, imaged under UV. Both repeats showed bands at around the expected size of 600bp following 34 cycles of the PCR program.



Score	Expect	Identities	Gaps	Strand
1031 bits(558)	0.0	560/561(99%)	0/561(0%)	Plus/Plus

```
Query  59   ATGAAAAAGTGGCTAGTGACGATTGCAGCACTGTGGCTGGCCGGATGCAGCTCCGGCGAA  118
Sbjct  1    ATGAAAAAGTGGCTAGTGACGATTGCAGCACTGTGGCTGGCCGGATGCAGCTCCGGCGAA  60
Query  119  ATTAATAAAAACTATTACCAGTTACCTGTGGTGCAGAGCGGTACACAAAGTACCGCCAGC  178
Sbjct  61   ATTAATAAAAACTATTACCAGTTACCTGTGGTGCAGAGCGGTACACAAAGTACCGCCAGC  120
Query  179  CAGGGCAATCGTCTGTTATGGGTAGAGCAGGTCACTGTTCCTGACTATCTGGCGGGGAAT  238
Sbjct  121  CAGGGCAATCGTCTGTTATGGGTAGAGCAGGTCACTGTTCCTGACTATCTGGCGGGGAAT  180
Query  239  GGTGTGGTTTATCAAACCAGTGATGTGAAGTATGTGATTGCCAACAACAACTTGTGGGCC  298
Sbjct  181  GGTGTGGTTTATCAAACCAGTGATGTGAAGTATGTGATTGCCAACAACAACTTGTGGGCC  240
Query  299  AGCCCGTTGGATCAACAGTTGCGCAACACCCTGGTTGCCAACCTGAGTACGCAACTGCCC  358
Sbjct  241  AGCCCGTTGGATCAACAGTTGCGCAACACCCTGGTTGCCAACCTGAGTACGCAACTGCCC  300
Query  359  GGCTGGGTGGTTGCCTCCCAGCCTCTGGGAAGCGCCCAGGACACGCTCAATGTTACCGTA  418
Sbjct  301  GGCTGGGTGGTTGCCTCCCAGCCTCTGGGAAGCGCCCAGGACACGCTCAATGTTACCGTA  360
Query  419  ACGGAGTTTAACGGTCGCTATGATGGCAAGGTCATTGTCAGTGGTGAGTGGCTGTTGAAC  478
Sbjct  361  ACGGAGTTTAACGGTCGCTATGATGGCAAGGTCATTGTCAGTGGTGAGTGGCTGTTGAAC  420
Query  479  CACCAGGGACAACTGATCAAACGTCCGTTCCGTCTGGAAGGAGTGCAAACTCAGGATGGT  538
Sbjct  421  CACCAGGGACAACTGATCAAACGTCCGTTCCGTCTGGAAGGAGTGCAAACTCAGGATGGT  480
Query  539  TACGATGAGATGGTTAAAGTGCTGGCCGGTGTCTGGAGTCAGGAAGCCGCTTCTATTGCA  598
Sbjct  481  TACGATGAGATGGTTAAAGTGCTGGCCGGTGTCTGGAGTCAGGAAGCCGCTTCTATTGCA  540
Query  599  CAAGAGATAAAACGTCTACCT  619
Sbjct  541  CAAGAGATAAAGCGTCTACCT  561
```

Figure 2.2 – Validation of *PqiC* ligation into plasmid pET28b by double restriction (left) and plasmid sequencing (right). An agarose gel (eft) of pET28b:*PqiC* restricted with *Nco*I & *Hind*III, containing EtBr was run, and imaged under UV. In all three repeats, pET28b is visible at around 5kbp, with PqiC bands apparent around the expected 600bp. An alignment between known *PqiC* sequence and sequencing run of putative pET28b:*PqiC* (right) showed 99% identity, with only a single difference observed; this mismatch would result in the same lysine residue in the primary structure of expressed PqiC.

2.2.2 Expression of Full-length PqiC

With the full length *PqiC* gene cloned into plasmid pET28b, it next needed to be transformed into competent cells suitable for protein expression. The cells used were *E.coli* C43 (DE3), so chosen for their resistance to membrane protein overexpression (109). For an initial expression text, a total volume of 1 litre of cells was used. After an overnight IPTG-induced protein expression, cells were harvested by centrifugation. Due to the membrane localisation of PqiC, cellular membranes were collected and solubilised in detergent (1% DDM). Owing to the His-tag incorporated as a feature of the pET28b plasmid, PqiC was further purified using Nickel affinity chromatography, with its effectiveness analysed using polyacrylamide gel electrophoresis **(Fig 2.3)**. This initial expression test proved ineffective – none of the collected fractions from throughout the purification exhibited staining indicating protein of 20kDa, as would be expected for PqiC. The lack of discernible target protein indicated a potential problem in the cloning process, or protein expression occurring at low levels.

At this stage, it was decided to increase the volume of protein-producing cells in an effort to detect successful PqiC expression. The process was amplified eightfold, with 8 litres of cell culture used for protein expression – other than this the methodology remained consistent. In this instance, potential PqiC was observed on the resulting PAGE gel **(Fig 2.4)**, albeit at a low level and lacking in purity. The gel was also treated with a nickel-conjugate d fluorescent dye (InVision stain), to try and further authenticate the presence of PqiC through its binding to the His-tag attached to the protein **(Fig 2.4).** While some signal was apparent in the range of 20kDa, bands of over 100kDa also exhibited staining. This, coupled with the fact that stained bands were collected from late fractions of the IMAC elution, would point to the likelihood

that if the observed bands *were* from PqiC, then the expressed protein is prone to aggregation.

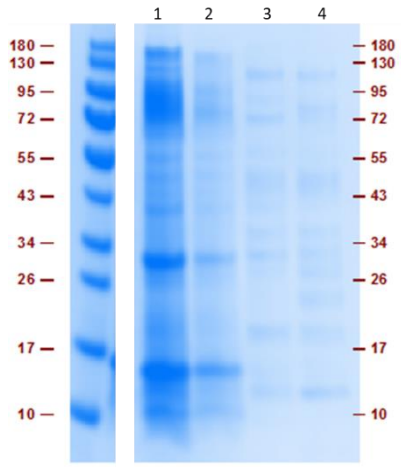


Figure 2.3 – SDS-PAGE gel of initial PqiC expression test, stained with Quick Coomassie. *E. coli* C43 (DE3) cells transformed with plasmid pET28b:PqiC were induced with 0.2mM IPTG to express protein overnight, with total volume 1 litre. Following a membrane collection and solubilisation in 1% DDM, the soluble fraction was used for Ni^{2+} IMAC. The column flowthrough (1), washes with binding buffer (2) and wash buffer (3) were collected, along with column elution (4). None of the fractions tested suggested the presence of PqiC at the expected 20kDa.

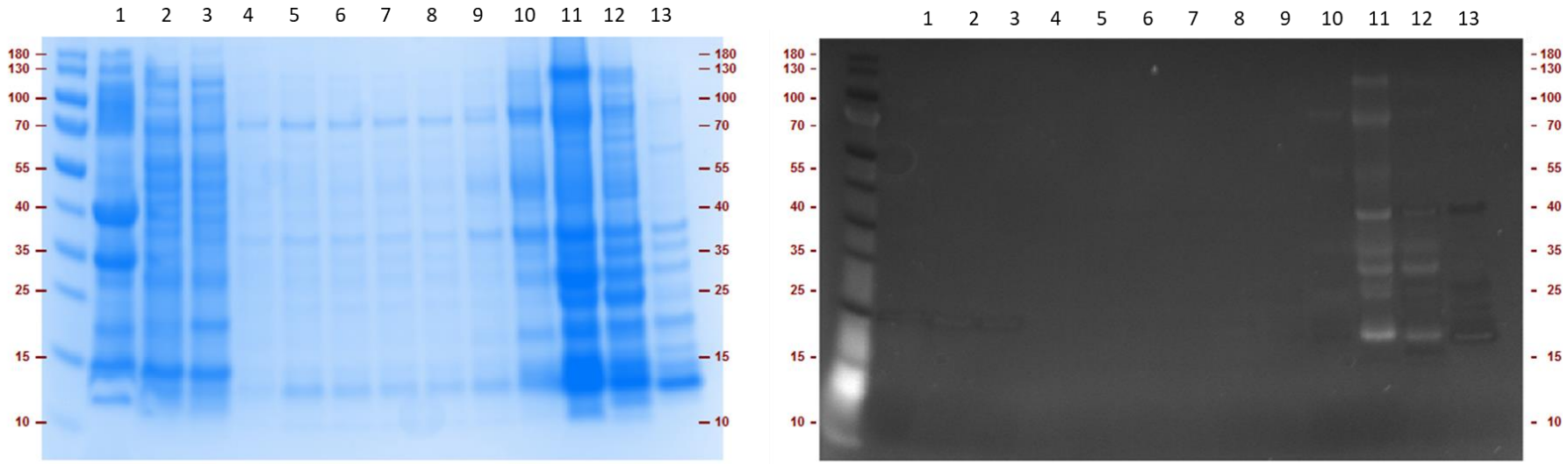


Figure 2.4 – SDS PAGE Gel of large scale PqiC expression test stained with Quick Coomassie (left) and InVision His-Tag stain (right). *E. coli* C43 (DE3) cells transformed with pET28b:PqiC were induced overnight with 0.2mM IPTG, with total volume 8 litres. Following cell lysis, membranes were collected and solubilised in 1% DDM – the resulting soluble fraction was used for Ni^{2+} IMAC. The non-soluble fraction (1), column flowthrough (2), binding buffer (3) and wash buffer washes (4) were collected and run on the gel, along with the column elution in separate 1ml fractions (5-13). A potential PqiC band is visible between 15 and 25kDa with Coomassie staining, with a small amount of signal visible after InVision staining – this staining does suggest protein aggregation, with several larger bands also exhibiting staining.

2.2.3 Generation of *PqiC* constructs with N-terminal Truncation

Given the apparent tendency of full length PqiC to aggregate under current expression conditions, either the expression construct or the conditions themselves needed to be altered. With such a large number of combinations of buffer, pH and detergent needing to be screened, it was instead decided that modifying the expression construct would be a more efficient course of action. Analysis of the *PqiC* primary sequence was performed, using LipoP 1.0 (110) and SignalP 5.0 (111) to determine the location of a potential signal peptide. These analyses **(Fig 2.5)**, coupled with prior investigations proposing that PqiC is anchored in the outer-membrane (68), strongly suggested that the first 15 residues of the nascent protein are a lipoprotein signal peptide. Indeed, the protein also contains the lipobox - appropriate motif of LAGC from residue 12-15 (71). This peptide is likely cleaved from the protein, with the mature protein prospectively anchored to the outer membrane (given its lack of a +2 aspartate residue (25)) through attached to the cysteine residue directly following the signal peptide (110). In an effort to express PqiC detached from the outer-membrane, new primers were designed to enable the cloning of a truncated *PqiC* gene, lacking the first 45 bases coding for the signal peptide. This was hoped to improve yield and avoid the previously seen aggregation issues.

New primers for truncated PqiC were designed, using restriction sites for *NcoI* and *BamHI*. These were used to allow compatibility with the plasmids pLou3 and pHisTev. Both plasmids were used to tag the protein with an N-terminal TEV-proteinase cleavable hexa-his tag, with pLou3 also encoding a cleavable MBP fusion partner. In this instance, the previously developed pET28b:*PqiC* plasmid was used as PCR template, with the resulting product was run on an EtBr-containing agarose gel. This gel was imaged under UV **(Fig 2.6)**, which indicated a successful PCR with a single

band between 500 and 700 bases visible. The new DNA was excised from this gel and ligated into pLou3 and pHisTev, with the effectiveness of this ligation assayed with a double-restriction digest using *BamHI* and *NcoI.* The agarose gel run of this digest **(Fig 2.7)** appeared to show positive results, with each plasmid visible on the gel as a distinct band with bands for *PqiC* evident at the same size throughout.

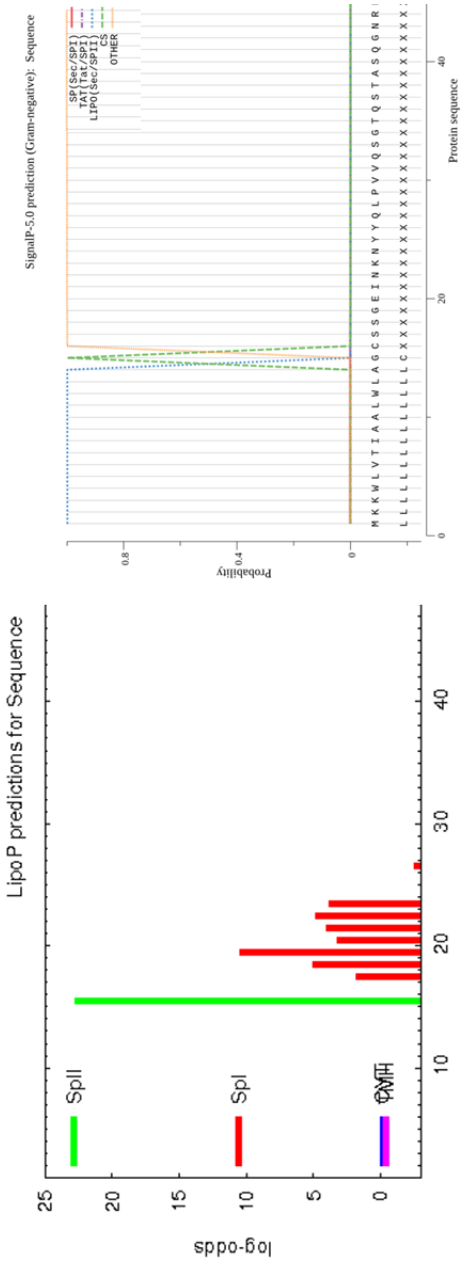


Figure 2.5– Signal peptide predictions for PqiC using LipoP 1.0 (Left) and SingalP 5.0 (Right). Both prediction algorithms suggest a high likelihood of a lipoprotein signal peptide (SpII) present at the N terminus of the protein, cleaved at around residue 15, with the cysteine at position 16 the first residue of the mature protein.

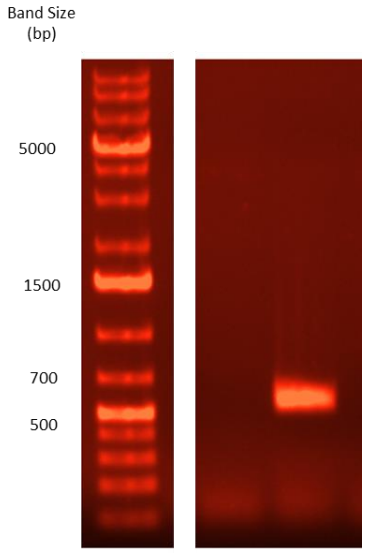


Figure 2.6 – Agarose gel of PCR reaction for truncated *PqiC* containing EtBr, imaged under UV. A band is clearly visible at the expected size of 550bp, following 34 cycles of the PCR program

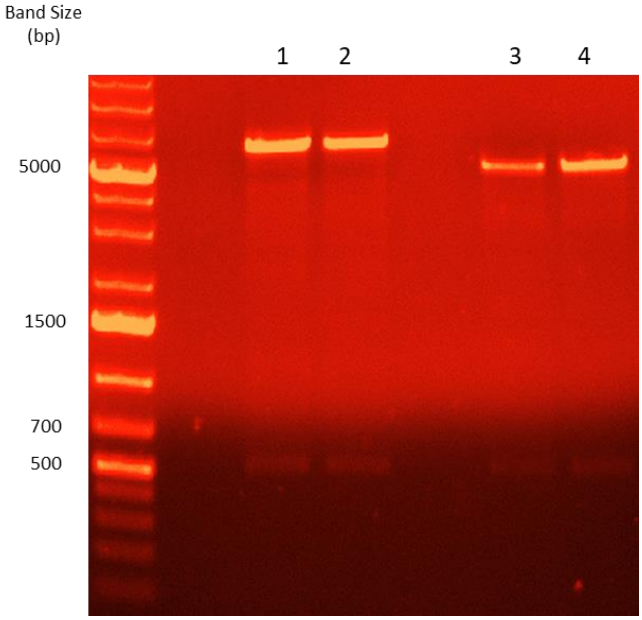


Figure 2.7 – Validation of truncated *PqiC* ligation into plasmids plou3 & pHisTev. Validation was achieved through double restriction digest, run on an Agarose gel. pLou3:*PqiC* (1, 2) & pET28b:*PqiC* (3,4) were restricted with *Nco*I & *Hind*III, run on an EtBr containing gel and imaged under UV. In each sample, a faint but discernible band was visible around 550bp as expected for truncated *PqiC,* with a band at around 5kb or 6kb corresponding to the linearised plasmid.

2.2.4 Expression tests of N-terminally Truncated PqiC

Having shown the cloning of truncated *PqiC* to be successful, the novel *PqiC*-containing plasmids needed transformation into a suitable expression strain. The strain SoluBL21 (DE3) (Genlantis) was selected for transformation, given its propensity to successfully express proteins that would often prove toxic or prone to aggregation in *in vitro* expression systems. Initially, small scale expression tests were carried out using SoluBL21 host cells and either the pHisTev or pLou3 truncated *PqiC* construct. First, a 1 litre purification utilising the pLou3 construct (including its MBP tag) was carried out; cells were again induced with IPTG for an overnight expression. Due to the truncation of signal peptide, the resulting cell lysate was in this instance not solubilised with detergent, and instead used directly for IMAC. A PAGE gel run of samples taken at various steps of the purification **(Fig 2.8)** indicated a high level of protein expression, with a thick band showing at around 65kDa – as expected for a PqiC-MBP fusion. A band was also visible at 45 kDa – as expected for MBP – which indicated that a small amount of the fusion protein was being cleaved into its constituents during the purification process. While there were other contaminant bands present at over 100 kDa and below 40 kDa, the densely stained 65 kDa band suggested successful expression of MBP-PqiC.

Following the success of a pLou3 test expression, the pHisTev construct was used for a separate 1 litre expression. This was carried out in identical fashion to the pLou3 expression system, with the same buffers and conditions throughout. As before, a PAGE gel was used to analyse fractions arising from the purification process **(Fig 2.9)**. In this case, a single significant band was seen, at around the expected 20kDa. The level of PqiC expressed appeared greater when using pLou3 system, although this is in part likely due to increased staining of a higher molecular weight protein. With

improved initial purity and a reduced number of further purification steps required to remove MBP (reducing subsequent protein loss), it was decided that the pHisTev expression system should be tested with a larger-scale expression.

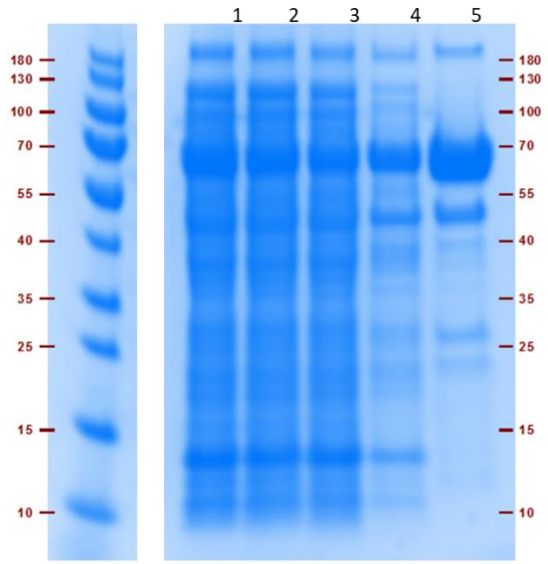


Figure 2.8 – SDS PAGE gel of expression test for truncated PqiC as a fusion protein with His-MBP, stained with Quick Coomassie. *E. coli SB21* (DE3) cells transformed with pLou3:PqiC were induced overnight with 0.2mM IPTG, total volume 1 litre. Clarified cell lysate (1) was used for Ni^{2+} IMAC. Column flowthrough (2), binding (3) and wash buffer (4) washes, along with elution (5) were collected and run on the gel. PqiC fused to MBP was expected to be 62kDa, with bands clearly visible at this size in each tested fraction. A large amount of presumed PqiC-MBP was observed in the elution, albeit alongside several other significant bands.

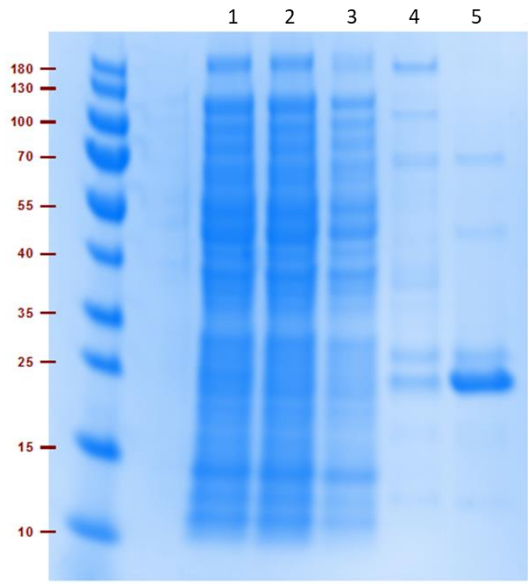


Figure 2.9 – SDS PAGE gel of expression test for truncated His Tagged PqiC, stained with Quick Coomassie. *E. coli SB21* (DE3) cells transformed with pHisTev:PqiC were induced overnight with 0.2mM IPTG, total volume 1 litre. Clarified cell lysate (1) was used for Ni^{2+} IMAC, with the flowthrough (2), binding buffer wash (3), wash buffer wash (4) and column elution (5) collected separately and run on the gel. While a band at the expected 20kDa wasn't apparent in lysate or wash fractions, a putative PqiC band was clearly visible in the column elution, with minimal impurities seen.

2.2.5 Large Scale Production of Truncated PqiC

A volume of 12 litres of cell culture was used chosen for a secondary test using the pHisTev construct. The same preliminary purification steps were used, followed by size-exclusion chromatography (SEC) to further purify the protein sample, and determine whether the protein appeared to be aggregating. The size exclusion chromatograph **(Fig 2.10)** had a small peak at around 50mL retention volume, but a single very large peak of over 1000 mAU at a retention volume of around 80mL. This peak indicated a good yield, and its relatively high retention volume suggested aggregation was unlikely – large aggregates would be expectedly seen around the column void volume of 40mL. The prominent peak was collected, with protein treated using an overnight digestion using TEV proteinase, aiming to remove the hexa-his tag incorporated in the protein. A pulldown through a nickel IMAC column was used to remove TEV proteinase and uncleaved protein from the solution. Fractions from before and after TEV cleavage were analysed with a PAGE gel **(Fig 2.10)** As previously seen, a strong PqiC band around 20KDa was apparent, although after proteolytic cleavage, very little of this was recovered from the IMAC pulldown – the majority of protein remained bound to the column and was detectable following a high-imidazole elution. With the post-cleavage band migrating marginally further down the gel, this pointed to non-optimal conditions during the pulldown causing this large loss of protein.

To prepare the sample for recovery of cleaved protein with a second pulldown, the high level of imidazole required for elution was removed from the protein sample using a desalting column. In its place, 10mM imidazole was added to the binding buffer, and to the protein sample before application to the nickel IMAC column. Following this, the IMAC flowthrough was applied to a SEC column, with fractions

from this and previous stages used to run a PAGE gel **(Fig2.11)**. Samples taken before and after the pulldown appeared very similar, with relatively small amounts of protein present in a wash and elution fraction. These observations pointed to a vast improvement in the recovery of PqiC. The protein samples obtained following SEC appeared to be of good purity, with strong bands assumed to be cleaved PqiC and a reduced proportion of uncleaved product visible, compared to the initial IMAC pulldown fraction **(Fig 2.10)**. PqiC was collected following SEC and concentrated to 10 mg/ml for use in initial crystallisation trials.

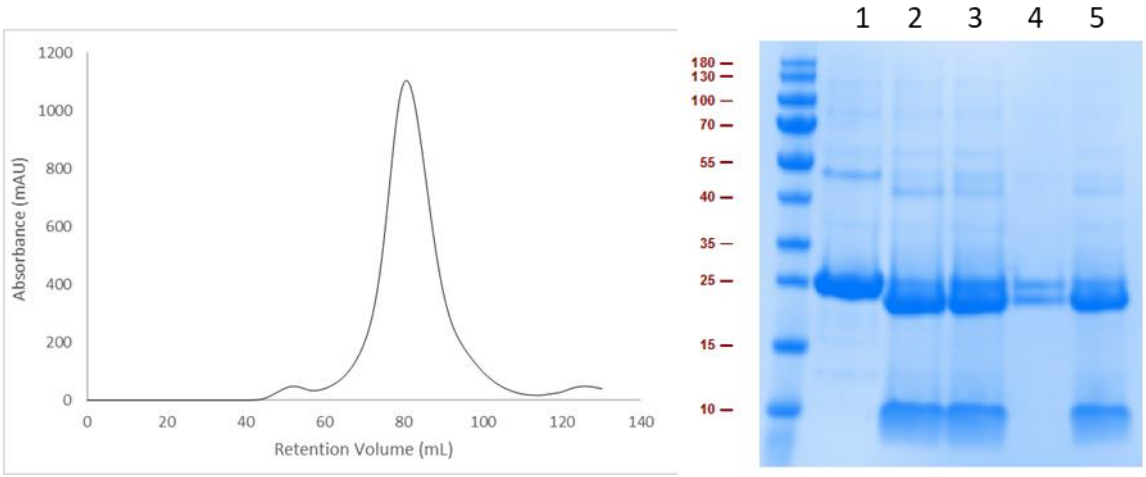


Figure 2.10 – SEC chromatograph and PAGE gel of large scale PqiC purification. *E. coli SB21* (DE3) cells transformed with pHisTev:*PqiC* were induced overnight with 0.2mM IPTG, total volume of 12 litres. Clarified cell lysate was used for IMAC, with column elution used for SEC (left); a peak of over 1000 mAU was observed at 80mL. This peak was collected and treated with 1:50 TEV proteinase overnight, followed by a pulldown through a Nickel IMAC column. Samples taken before (1) and after (2,3) cleavage, as well as flowthrough (4) and elution (5) from the pulldown were run on a PAGE gel (right). The majority of the putative PqiC at around 20kDa was only observed during column elution, so a further pulldown was needed to recover cleaved PqiC.

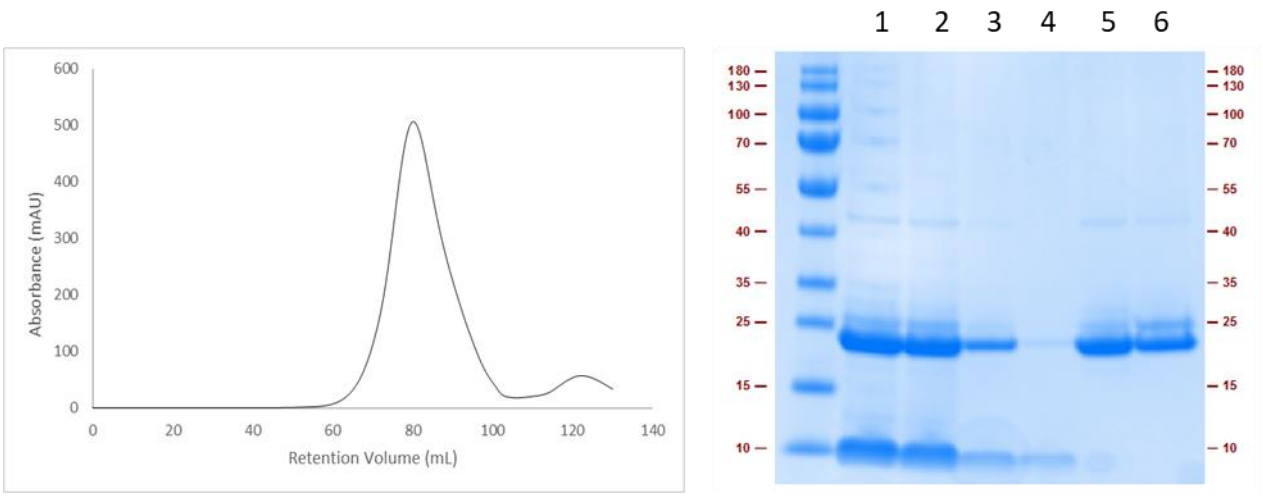


Figure 2.11 – SEC chromatograph and PAGE gel of large scale PqiC purification following overnight cleavage by TEV proteinase. Following initial purification **(Fig 2.10)**, 10mM imidazole was added to the TEV proteinase cleaved PqiC sample, which was then pulled down through a Nickel IMAC column. This flowthrough was used for a second run of SEC (left), which resulted in a peak at around the same retention volume of 80mL, but at a significantly reduced absorbance of ~500 mAU. An analytical PAGE gel (right) was run, using a sample taken before further purification (1), IMAC flowthrough (2), wash (3) and elution (4), as well as samples from the peak at retention volume 80 mL (5,6). The gel indicated some loosely bound protein was retained by the column (removed in the wash), with minimal protein present in the elution. These results suggested successful recovery of PqiC, with a relatively pure protein sample observed after SEC, albeit with a small amount of protein visible at around 25 kDa.

2.2.6 Initial PqiC Crystallisation Trials

Initial PqiC crystal trials were carried out using the sitting drop vapour diffusion method, with double-well plates. Drop sizes were 1 µl and 0.6 µl in top and bottom wells respectively, with equal proportion of protein solution and crystallisation screen solution. Initial crystal trials used several commercial soluble protein screens – listed in **Table 2.3** – and plates were incubated at 21°C.

After an incubation period of 96 hours, crystal growth had become apparent on the Peg Rx screen plate **(Fig 2.12)**. Specifically, wells B4 and C7 showed good initial hits, in relatively similar conditions of 0.1M Sodium Acetate Tetrahydrate pH 4, 10% PEG 4,000 and 0.1 M Citric Acid pH 3.5, 14 % PEG 1,000. To determine whether these were crystals of PqiC or if they were crystals of salt, they were sent for X-ray diffraction studies at Diamond Light Source. The crystals were soaked in cryoprotectants and picked in loops; these harvested crystals were flash frozen and dispatched in a dry shipper. Data was collected from the crystals using an X-ray beam with wavelength 0.97625 Å and an exposure time of 0.1 seconds. The diffraction spectra collected from the crystals **(Fig 2.13)** was consistent with that seen from protein crystals, rather than their salt counterparts. There were many distinct spots, both close to and further from the beam stop; crystals diffracted to around 4 Å. Had the crystal have been salt-based, there likely would have been fewer, larger spots, without diffraction detected close to the centre of the pattern. Data collected from these initial crystals confirmed that they were indeed formed of protein and represented good starting points for optimisation and eventual structure solution.

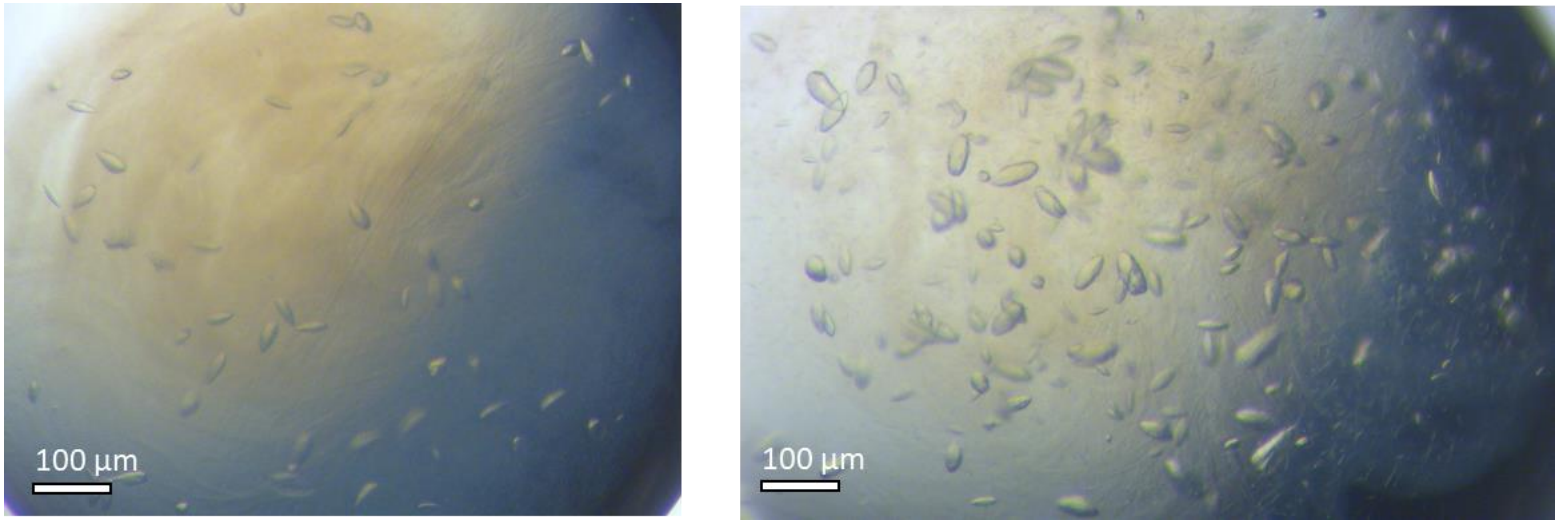


Figure 2.12 – Crystals obtained from initial PqiC screening. Plates were incubated for 4 days at 21°C, using a number of commercially available crystal screens. Both crystal-containing wells were from the PEG Rx 1 screen. The well C7 (left) had a well solution of 0.1 M Sodium Acetate Tetrahydrate pH 4 & 10% PEG 4,000. Well B4 (right) used a screen solution of 0.1 M Citric Acid pH 3.5 & 14 % PEG 1,000. Both of these initial hits were used to design optimised conditions for crystal growth. The protein used for trials was purified in figure 2.11.

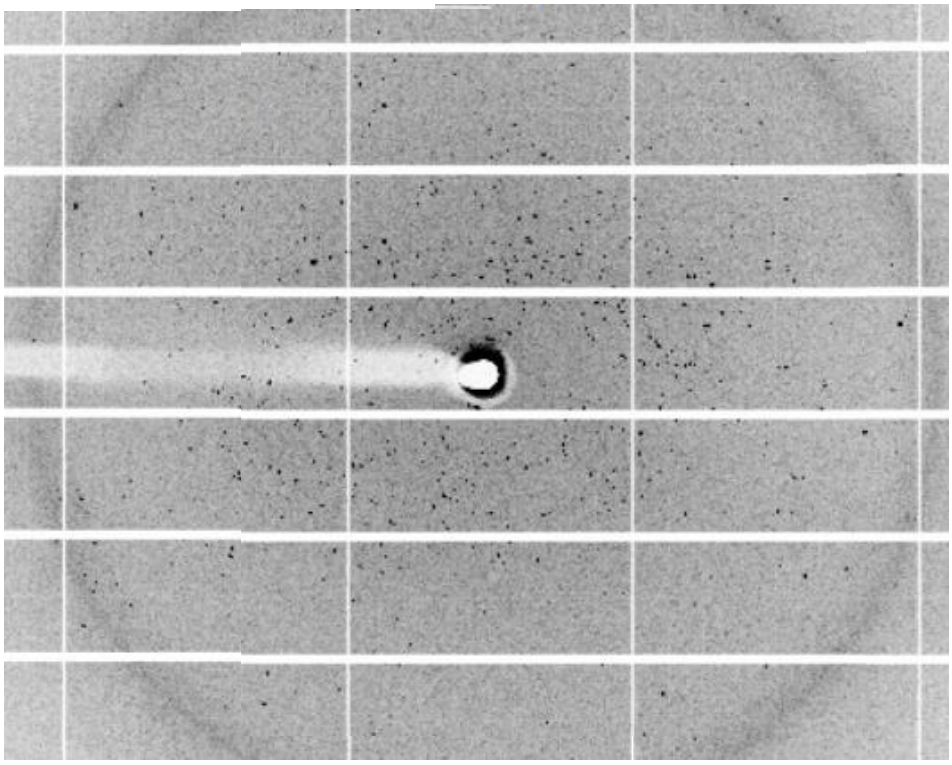


Figure 2.13 – X-ray diffraction pattern obtained from PqiC crystals generated from initial trials. Diffraction data was collected at a wavelength of 0.97625 Å, with an exposure time of 0.1 seconds at a beam transmission of 50%. The maximum resolution of observed spots was around 4 Å. The nature of the diffraction pattern, with a number of spots close to and far from the beam centre strongly suggests a protein crystal.

2.2.6 Further PqiC Purification and Crystallisation Trials

To generate more PqiC for further crystallisation trials to optimise crystal growth and improve resolution, a protein purification was carried out with a slightly modified method. This aimed to enhance yield and reduce the time between cell lysis and crystallisation trials. Protein was expressed from the same cell stock as previously used, given the success of the pHisTev vector. Steps for cell harvest, lysis and centrifugation remained the same. At the IMAC purification phase, the sample loading flowthrough was itself loaded onto the column once again, to ensure maximal amounts of PqiC could bind to the column. The column was washed and eluted as previously, with the protein exchanged to low imidazole buffer using a 26/10 desalting column as opposed to a 16/600 SEC column, reducing the time taken for this step by over an hour. Protein was subjected to TEV cleavage overnight at 4°C as before, and the sample was supplemented with 10 mM imidazole prior to a pulldown reaction, to prevent the previously seen loss of protein at this step. The flow-through from this step was purified with SEC **(Fig 2.14)**, with the purification analysed with a PAGE gel **(Fig 2.14)**. While PAGE indicated that a decent amount of protein was retained in the pulldown column, the overall yield was increased, with the main SEC peak increasing from around 500 mAU to over 700 mAU. Moreover, the PqiC band seen after SEC contained a lesser proportion of the larger 25 kDa band previously seen. Protein was collected from the fractions in this peak and concentrated to 10 mg/ml for use in further crystallisation trials and optimisations.

Using this newly purified PqiC protein, plates were made using optimisation screens put together by hand **(Table 2.4/5)**. These optimisation plates varied the PEGRx B4 and C7 conditions with changing pH and precipitant concentration. Plates were

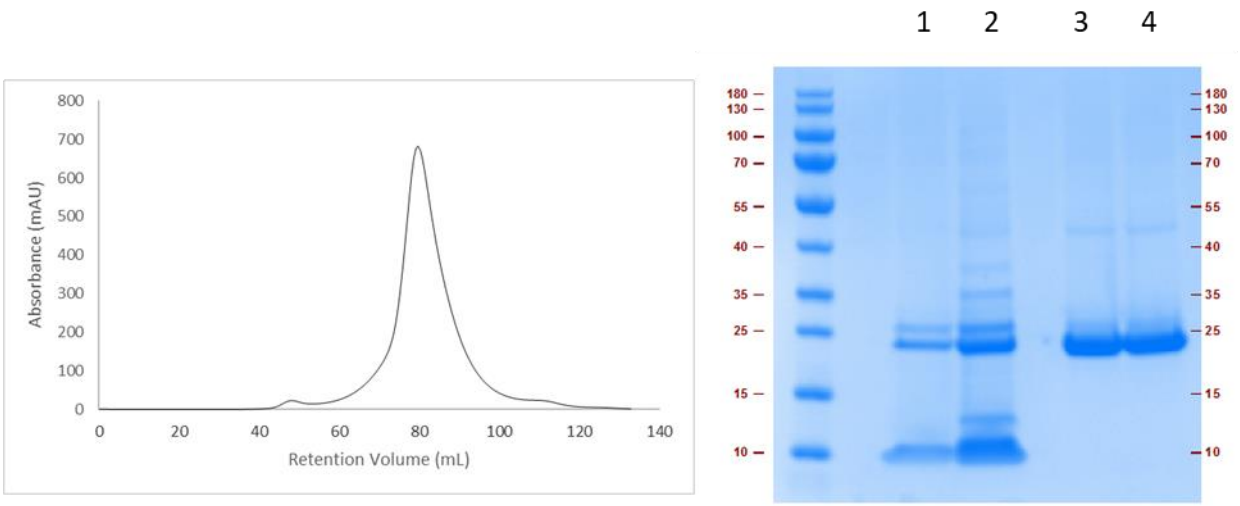


Figure 2.14 – SEC chromatograph and PAGE gel of optimised PqiC purification. As previously described, *E. coli SB21* (DE3) cells were used for protein expression in a volume of 12 litres. Clarified cell lysate was applied to a Nickel IMAC column with additional loading of flowthrough to ensure full protein binding. PqiC was eluted with high imidazole buffer, before desalting and overnight tag cleavage with TEV proteinase. Cleaved protein with added imidazole was pulled down through a Nickel column, with the pulldown applied to a SEC column (left), yielding a single significant peak of ~700 mAU at a retention volume of around 80mL, as observed in previous PqiC purifications. Fractions from the Wash (1) and elution (2) of the pulldown were run on a PAGE gel (right) stained with Quick Coomassie, alongside samples taken from the peak seen in SEC (3,4). While it appeared that an amount of protein was captured by the pulldown step (evident in the elution fraction, 2), a large amount of pure protein was derived from the purification, owing to the large absorbance value from SEC and the heavy staining at around 20kDA on the gel.

made at 5 mg/ml, 10 mg/ml, and 15 mg/ml – each plate used the same 1 µl and 0.6 µl drop sizes trialled in initial crystallisation experiments. In addition to the new optimisation screens, crystal plates were made at 10 mg/ml using the complete PEGRx and Crystal Screen 1/2 screens, in the chance that enhanced protein purity from the modified purification technique could improve crystal growth.

After eight days' incubation at 21°C, crystal growth was observed on the newly created optimisation plates made using the citric acid screen **(Fig 2.15)**. The crystals appeared generally larger in size than the initial screen hits, reflected more light, and had a more angular appearance. The 5 mg/ml and 10 mg/ml plates both contained hits, particularly in lower pH and higher PEG concentration wells. After 50 days incubation, these crystals remained unchanged. However, significantly larger crystals had grown wells from the PEG Rx (H9), and Crystal Screen 1/2 plates made with PqiC from the optimised purification process. These crystals **(Fig 2.16)** were much larger than those previously seen, and in the form of long needles. The newly observed needle crystals were considered suitable for data collection, alongside crystals obtained from the optimised citric acid conditions. Crystals selected for data collection were soaked for between 5-15 minutes in cryoprotectants; in many cases 200 mM Sodium Iodide was included in cryoprotectants to allow phasing through Single-wavelength Anomalous Diffraction (SAD) using the positions of iodine atoms. Cryoprotectant-soaked crystals were flash frozen in liquid nitrogen and shipped to Diamond Light Source for data collection.

Table 2.4 – PEG Rx B4 Optimisation conditions for PqiC crystallisation. Using the PEG Rx B4 condition as a starting point, optimisation screen blocks were made as shown; each well in the screen contained 0.1M Citric Acid.

		PEG 1000 (%)				
		10	12	14	16	18
pH	3.1					
	3.3					
	3.5					
	3.7					
	3.9					

Table 2.5 – PEG Rx C7 Optimisation conditions for PqiC crystallisation. A screen was produced based on the Peg Rx C7 condition. A screen block was produced shown; each well contained 0.1 M sodium acetate.

		PEG 4,000 (%)				
		6	8	10	12	14
pH	3.8					
	4					
	4.2					
	4.4					
	4.6					

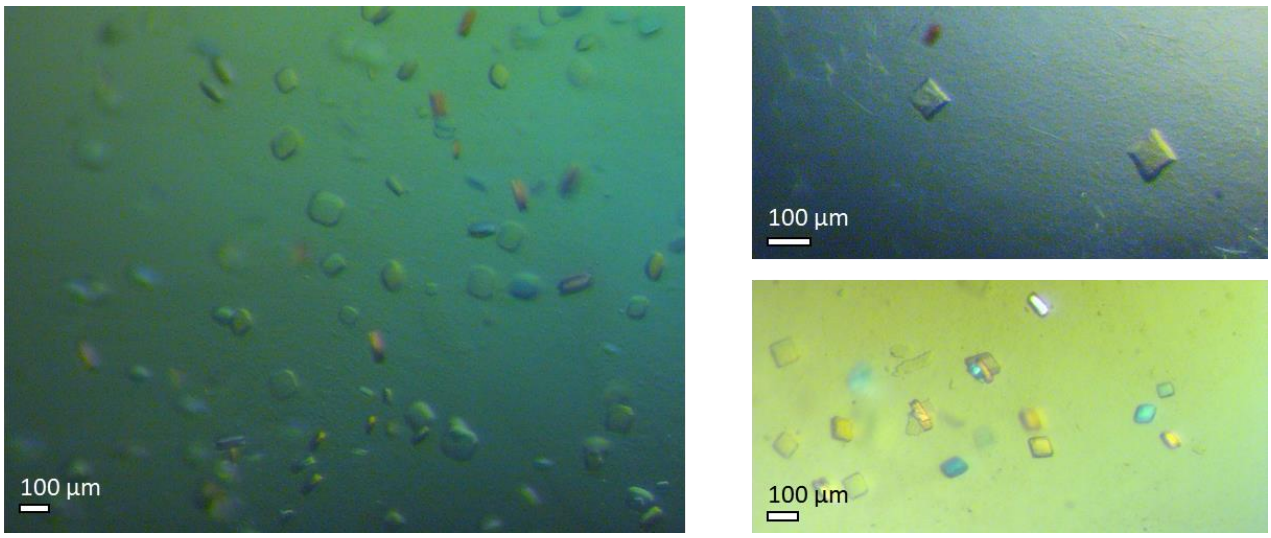


Figure 2.15 – Crystals obtained from PqiC optimisation screening. Crystals were imaged from the citric acid optimisation screen (Table 2.4). Plates were incubated at 21°C for 8 days prior to photographing. The plate made at 5 mg/ml yielded the greatest number of hits (Left, Upper Right), mainly in wells with lower PEG concentration and lower pH. The 10 mg/ml well (Lower Right) also contained several hits.

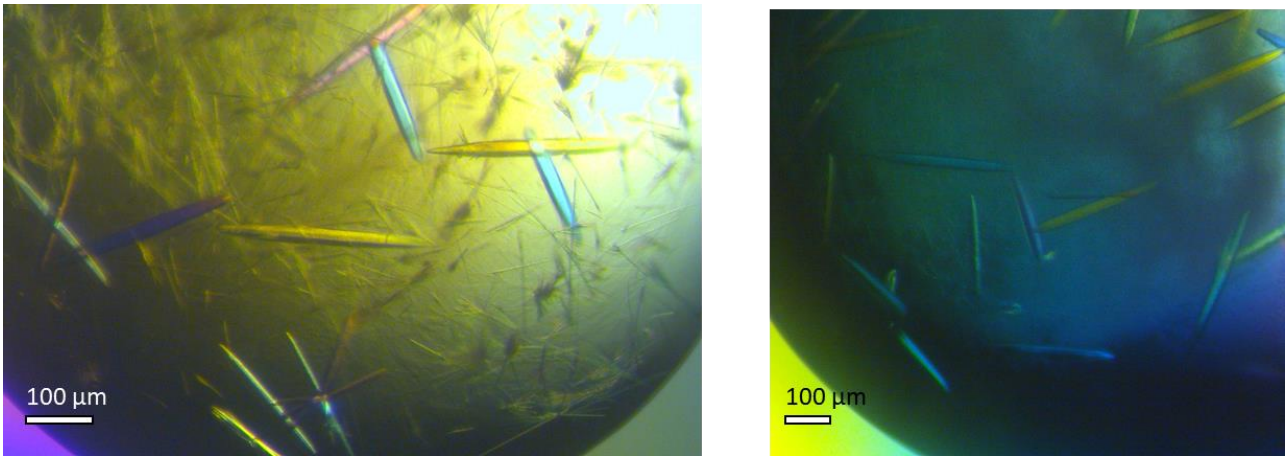


Figure 2.16 – Needle-like PqiC crystals observed in screening plates. These screen plates were made with protein derived from improved purification methods at 10 mg/ml. Wells from PegRx H9 (Left) and Crystal Screen 1/2 D6 (Right) both contained larger, needle shaped crystals with sharp, defined edges and highly reflective surfaces. The crystallisation solution in PegRx H9 was 5% v/v 2-Propanol, 0.1 M Citric acid pH 3.5, 6% w/v Polyethylene glycol 20,000. The well solution for Crystal Screen 1/2 D6 was: 0.05 M Potassium phosphate monobasic, 20% w/v Polyethylene glycol 8,000.

2.2.7 PqiC Structure Solution

Using the submitted crystals, data was collected on beamline i04 at 1.8223 Å, with 3600 images taken at 0.1 second exposures, using a beam transmission of 25 % and oscillation of 0.1°; all 3600 images were transferred from servers at DLS for data processing. Images were integrated using xia2, run through the DIALS interface, with space group P 6 1 assigned by pointless. Data was subsequently reduced and scaled using Aimless, with an initial R_{meas} value of 0.568 at a data resolution of 2.02 Å. After limiting resolution to 2.79 Å, data was processed with Aimless with an R_{meas} of 0.233 (full data parameters from phasing data in **(Table 2.6)**. Phasing and initial model building was carried out using Crank2; phasing of the data was performed using 2000 ShelX trials. ShelX determined 4 iodine sites, resulting in an initial PqiC model built by Buccaneer **(Fig 2.17)**. This initial model was refined with 100 Refmac cycles, before manual adjustments using COOT.

This refined initial model was used as a molecular replacement model with higher-resolution data collected at 0.9795 Å, with 1800 images taken at 0.04 seconds exposure and 100% beam transmission. As for iodine-phased data, the full dataset was integrated with xia2, before scaling and reduction with Aimless. An initial structure was built using MOLREP and refined using 50 Refmac cycles, with manual refinement using COOT; water molecules were added using ARP/wARP. Structure validation was carried out using MolProbity, with COOT used to manually rectify rotamer and Ramachandran outliers, yielding a refined PqiC structure **(Fig 2.18).** PqiC takes the form of a hair clip- like shape, with N and C-terminal subdomains. The N-terminal subdomain contains seven short β-sheets, and an α helix. The C-terminal domain comprises three β-sheets and a longer α helix, which forms a

Table 2.6 - Data collection and refinement statistics for PqiC data used to construct initial model. This was constructed from phasing information from Iodine SAD. Statistics from Resolution to multiplicity are given as Overall (outer shell).

Data Collection	
Beamline	i04
Wavelength (Å)	1.8223
Spacegroup	P 6 2 2
Unit Cell Dimensions	
a, b, c (Å)	63.721, 63.721, 86.283
α, β, γ (°)	90.000, 90.000, 120.000
Resolution (Å)	55.28 - 2.79 (2.94-2.79)
R_{merge}	0.230 (1.078)
R_{meas}	0.233 (1.128)
$CC_{1/2}$ (%)	0.983 (0.952)
Completeness (%)	99.2 (96.6)
Anomalous Completeness (%)	98.6 (96.2)
Mean I/σ (I)	12.7 (6.7)
Multiplicity	18.7 (19.0)
Anomalous Multiplicity	9.5 (9.5)
Refinement	
Molecules per AU	1
Total Atoms	1173
Number of Reflections	4945
R_{factor}	0.27
R_{free}	0.302

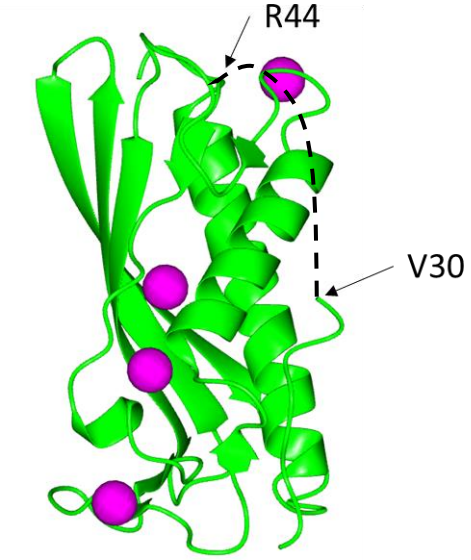


Figure 2.17 - Initial PqiC model built using Buccaneer. Iodine sites (magenta) were determined by ShelX. The model was unable to be built between valine 30 and arginine 44 and is instead indicated by a dashed line. Data was collected at 1.8223 Å, with substructure detected using 2000 trials of ShelX.

cavity with the shorter N-terminal helix within the protein. Statistical parameters for this data and model are available in **Table 2.7**. The N-terminal region appears to be more flexible, with residues from V30 to R44 in particular not able to be built into the generated electron density. The flexibility of this N-terminal region could help facilitate the involvement of some of its residues in embedding the protein into the outer membrane through its anchor. Incidentally, further data collected from square PqiC crystals **(Fig 2.15)** also allowed the protein's structure to be solved using molecular replacement, but in the space group P 43 21 2 with 8 PqiC monomers in the asymmetric unit. Alignment of this model to the first MR-solved structure showed no appreciable conformational changes between the separate crystal forms.

Table 2.7 - Data collection and refinement statistics for PqiC structural model built using molecular replacement. Data collected at the native wavelength was used with the experimentally phased structure as a molecular replacement model to build a structure of PqiC in spacegroup P1 21 1 at a higher resolution.

Data Collection	
Beamline	i04
Wavelength (Å)	0.9795
Spacegroup	P 1 21 1
Unit Cell Dimensions	
a, b, c (Å)	60.666, 82.245, 60.666
α, β, γ (°)	90.000, 117.744, 90.000
Resolution (Å)	82.24 - 2.40 (2.49-2.40)
R_{merge}	0.074 (0.234)
R_{meas}	0.086 (0.303)
$CC_{1/2}$ (%)	0.993 (0.904)
Completeness (%)	94.9 (70.0)
Mean I/σ (I)	9.2 (3.0)
Multiplicity	3.0 (2.0)
Refinement	
Molecules per AU	3
Total Atoms	3637
Number of Reflections	59133
Number of Unique Reflections	19648
R_{factor}	0.194
R_{free}	0.260
Mean atomic B-factor (Å^2)	25.810
Ramachandran Analysis	
Favoured (%)	97.33
Allowed (%)	2.67
Outliers (%)	0
RMSD	
Bonds (Å)	0.0055
Angles (°)	0.855

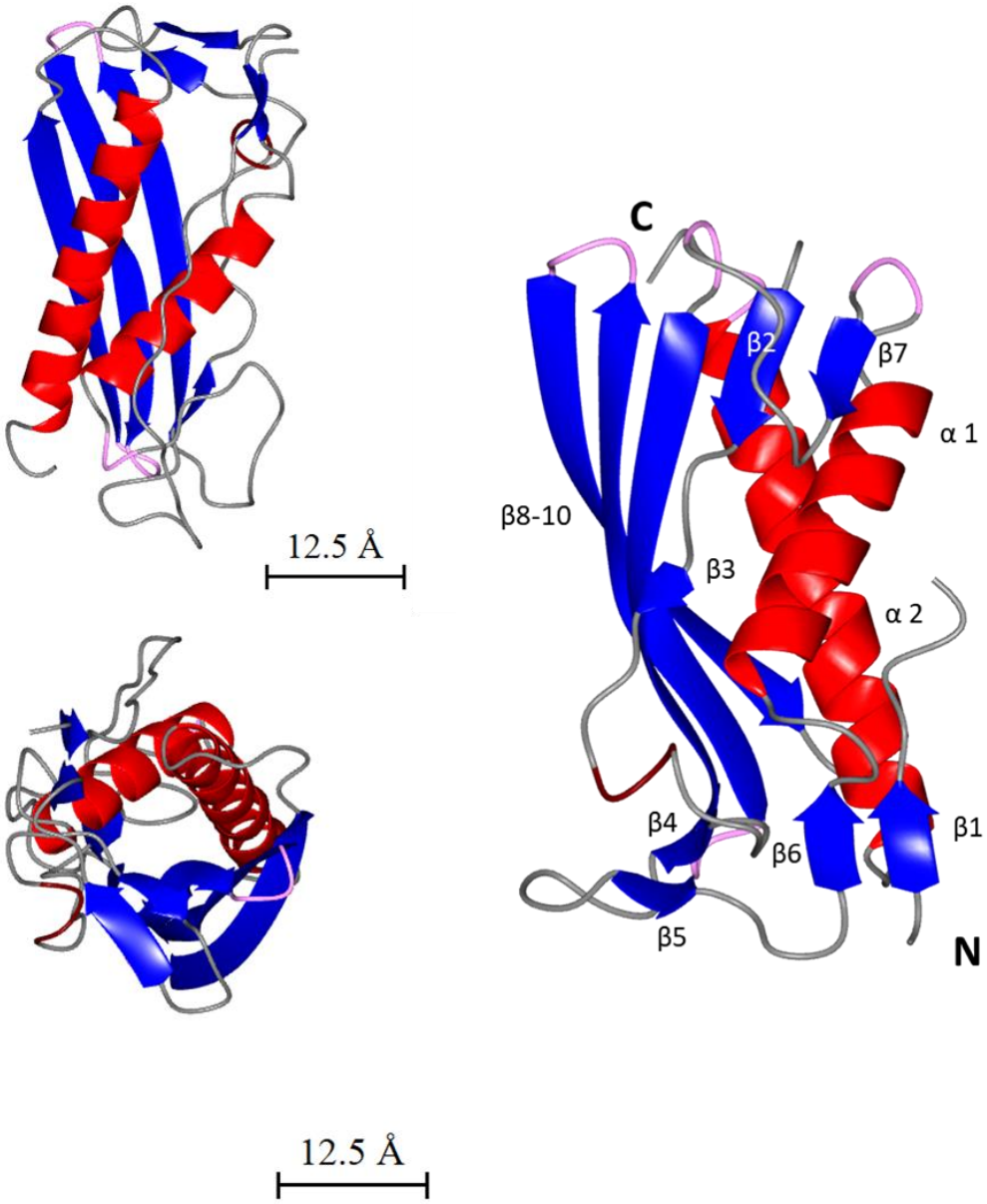


Figure 2.18 - PqiC structure as determined through X-ray crystallography studies to a resolution of 2.40 Å. An initial model phased using SAD with iodine atoms was used as a molecular replacement model to build the initial final model, before its refinement with Refmac5 and COOT.

PqiC has a hairclip shape, with N and C-terminal subdomains. The N-terminal subdomain contains seven short β-sheets, and an α helix around 22 Å long. The C-terminal domain consists of three long anti-parallel β-sheets and a long α helix 37 Å long, forming a deep cavity with the shorter N-terminal helix.

Not shown is a flexible region from V30 to R44 for which strong electron density was not visible.

2.2.8 PqiC Structural Analysis

Interestingly, when analysed using the DALI protein comparison server (112), the newly-solved structure bore a striking resemblance to that of LPS-transport protein LptE. When superimposed **(Fig 2.19)**, the two structures were visibly very similar, with the same structural motif of a pair of helices facing three anti-parallel sheets, albeit with longer β-sheets apparent in LptE. Previous structural studies on *E. coli* LptE have indicated that protein function is dependent on residues found in the loop regions between the first large β-sheet pair, and between the third large β-sheet and the first helix (113). Given the Lipid-A binding properties of LptE, the equivalent region in PqiC could potentially be equally important in its binding of Phospholipids.

The solved PqiC structure was used to search for potential Phospholipid (PL) binding sites using COOT, by examining unmodelled electron densities. A long, thin electron density was discovered at one end of the protein, between α-helices and β-sheets **(Fig 2.20)**. This density could belong to a bound PL molecule, with a fatty acid tail of Phosphatidylethanolamine appearing to be a good fit **(Fig 2.20)**. The PqiC residues Y64, W79, Y161 and M164 were in close proximity to this putative PL binding region and could indicate amino acids vital for the function of PqiC. Intriguingly, this was an equivalent region of the protein to the proposed Lipid A binding site in aforementioned LptE, further suggesting that this is a functionally relevant site for PqiC.

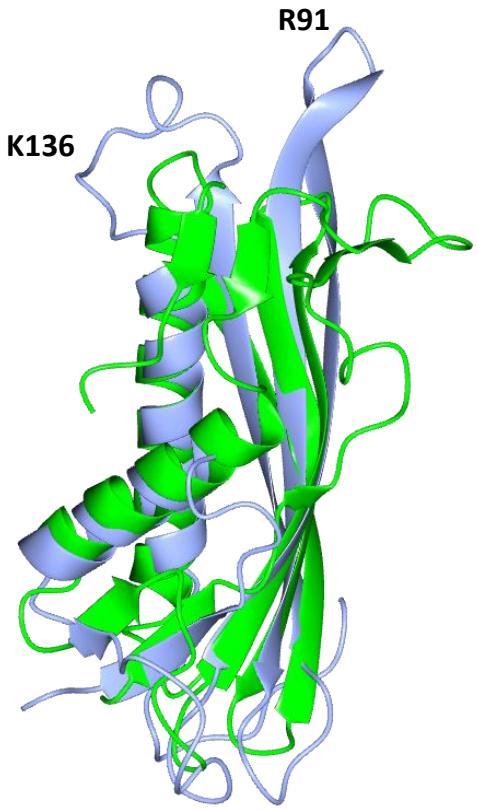


Figure 2.19 - Superimposition of LptE (blue) on solved PqiC structure (green). LPS-binding protein LptE superimposed with PqiC with an RMSD of 4.168 Å. The two proteins share the same general topology, with their hairclip shape and three β-sheets opposite a pair of α-helices of unequal length. Indicated on LptE are arginine 91 and lysine 136, both of which have been implicated in LPS interaction (113). Structure from PDB 4NHR (113).

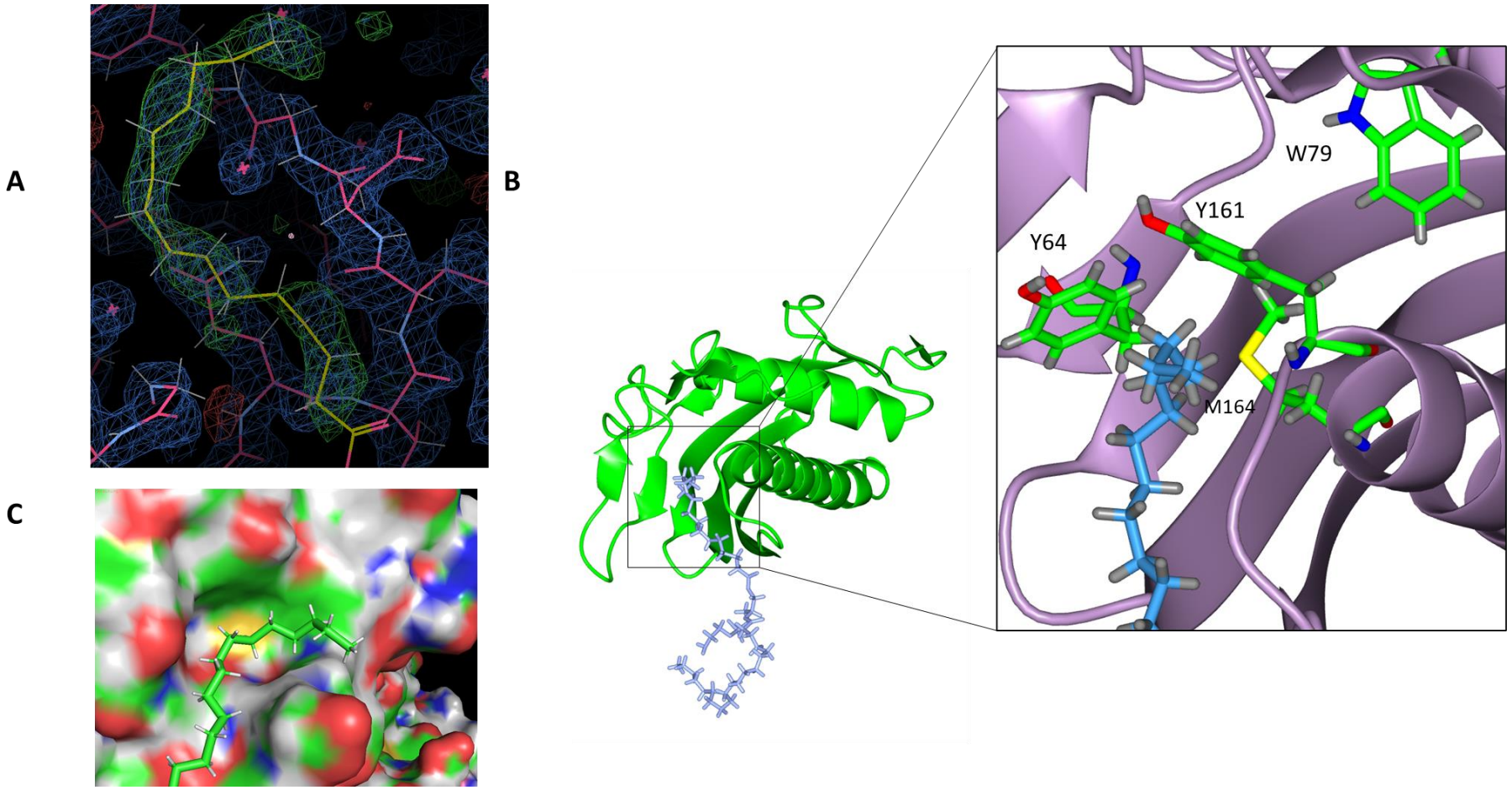


Figure 2.20 - A potential Phospholipid Binding site was determined from the crystal structure of PqiC. During manual refinement with COOT (A), an unfilled electron density (green) was noticed at the interface between the N and C-terminal subdomains of the protein, at the C-terminal of the long helix and the shorter, N-terminal β-sheets. This density was close to several aromatic residues and could suggest a binding cavity for Phospholipid (PL) molecules. A Fatty-acid tail of *E. coli* Phosphatidylethanolamine (yellow chain), fit well into this density, with Y64, W79, Y161 and M164 close by (B). A surface representation of PqiC (C) indicated that the prospective PL interacting site forms a possible binding cleft in the protein.

2.2.9 Site-Directed Mutagenesis of PqiC

With the structure of PqiC elucidated, attention turned to insights into the protein's function and mechanism of action. Using the protein structure which had been constructed, five sites with potential functional or structural significance were identified: Tyrosine 64, Tryptophan 79, Tyrosine 161, Methionine 164, and Phenylalanine 123. While Y64, W79, Y161 and M164 were close to PqiC' s putative phospholipid binding site, F123 was found deeper in the protein's cavity between the two α-helices, and was deemed another potential PL binding region, despite its lack of unmodelled electron density.

Primer pairs were designed for each site to mutate the target residue into a glutamate residue (or aspartate in the case of M164). Primers were designed to overlap and used with a for PCR with corresponding program, as outlined previously (114). To allow good protein expression for *in vitro* lipid binding tests, the first mutants to be cloned used the truncated *PqiC* gene found in previously constructed pHisTev expression vector as a template. The PCR reactions performed using this template were treated with *DpnI* to digest the *E. coli* – grown template DNA, then run on an agarose gel with Ethidium Bromide and imaged under UV **(Fig 2.21)**. Each lane contained a sharp band at around 6000bp, indicating that the template plasmid had been successfully replicated – any pre-existing DNA should have been digested. Following confirmation of successful mutagenesis with sequencing, the new plasmids were transformed into *E. coli* SB21 cells using heat shock for expression of mutant proteins.

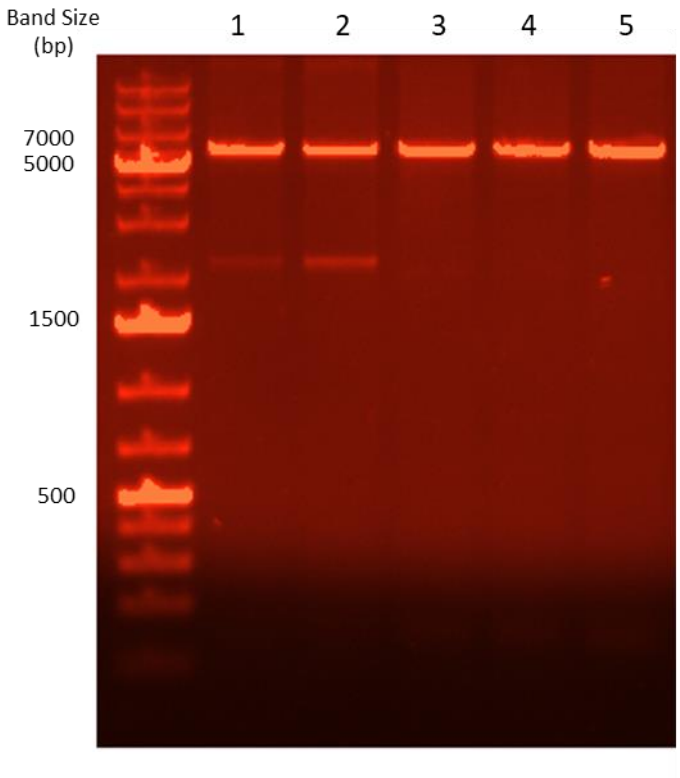


Figure 2.21 – Agarose gel of PCR for truncated *PqiC* mutants, containing EtBr and imaged under UV. Each reaction used the pHisTev:*PqiC*_truncated plasmid as a template, and a primer pair designed to introduce amino acid substitutions at mechanistically significant positions. The mutants produced here were: Y64E (1), W79E (2) F123E (3) Y161E (4) and M164D (5). Template DNA was digested with *DpnI* prior to running samples on the gel.

2.2.9 Purification of Mutant Truncated PqiC

The mutant proteins were individually purified following 6 litre expressions in SB21 cells, using the same expression and purification procedures which had proved effective for the Wild Type truncated protein.

Purification of the Y64E mutant gave rise to 3 distinct peaks following SEC **(Fig 2.22)**. The first of these, at 45 mL retention volume, was likely severely aggregated PqiC. The second peak at 60 mL was also potentially aggregated protein, with a much-reduced retention volume when compared to previous purification of WT protein; the 85 mL SEC peak was more likely to contain soluble, intact PqiC-Y64E. Sample from the SEC peaks were investigated with PAGE **(Fig 2.22)** with generally similar profiles across the three – each had predominant bands at just over 20 kDa, with some larger bands apparent, particularly in the 85 mL sample. The predominant peak of 85 mL was around 5 mL larger than that seen for the WT PqiC expression. Protein from the 85 mL peak was collected, concentrated to 1 mg/ml, and stored at -80°C for use in further investigations.

Protein purified with the W79E mutation was of a significantly reduced yield, with the highest peak on its SEC chromatograph **(Fig 2.23)** of around 45 mAU, compared to the 140 mAU observed for mutant Y64E. As previously seen, protein aggregation seemed to be an issue, with a sizeable peak again seen at 45 mL retention volume. PAGE **(Fig 2.23)** indicated that target PqiC protein was, as before, present in the 85 mL peak. However, there was a significantly increased presence of a marginally larger contaminant band at around 25 kDa. In line with the previous mutant, the 85 mL peak was collected, concentrated, and stored. While the identity of the protein present in the contaminant bands is unknown, its nature could be determined by

cutting and digesting the bands from the PAGE gel, followed by analysis using mass-spectrometry (115).

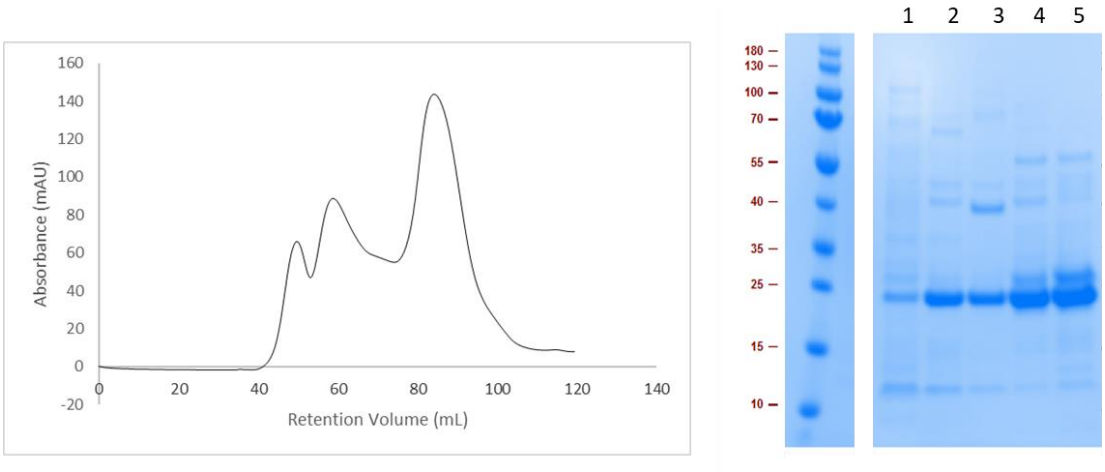


Figure 2.22 - SEC chromatograph and PAGE gel of truncated PqiC Y64E purification. Mutant protein was expressed and purified in the same manner as the truncated wild type protein, but without TEV cleavage. SEC of the sample purified by IMAC (left) gave rise to three distinct peaks, with the highest absorbance seen around 85mL – close to what was observed for WT PqiC. PAGE (right) was performed using samples from the first (1), second (2,3) and third (4,5) peaks. Results suggested that each of the fractions was likely to contain PqiC (all were around 20kDa), but given previous observations with the WT protein, the third peak was collected for concentration and further study.

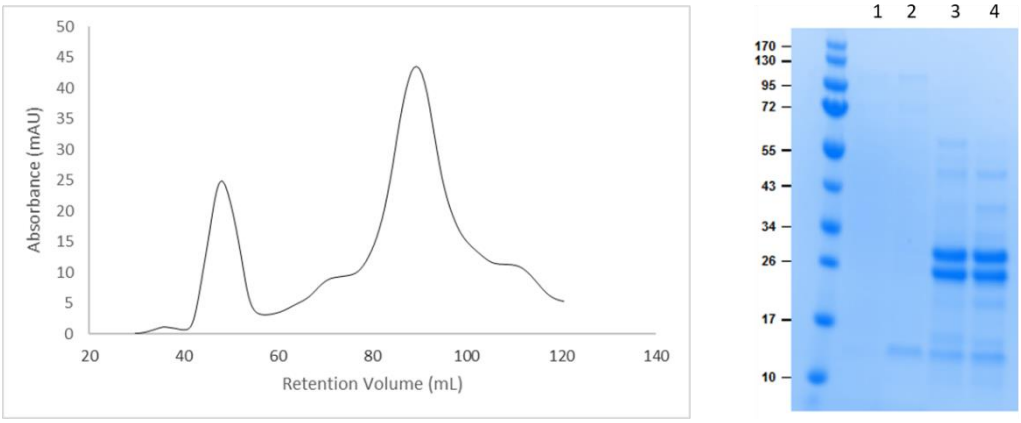


Figure 2.23 - SEC chromatograph and PAGE gel of truncated PqiC W79E purification. Mutant protein was expressed and purified in the same manner as the truncated wild type protein, without TEV cleavage. SEC carried out after IMAC (left) yielded two major peaks, with the highest absorbance seen around 85-90mL, slightly higher than seen with WT PqiC. PAGE (right) was run with two samples from both the first (1,2) and second (3,4) peaks. The first peak seemed to lack the target protein, but the second appeared to contain the desired mutant PqiC at around 20kDa, albeit in the presence of another slightly larger protein.

The purification for PqiC-F123E proceeded in a similar manner to that for W79E. Yield was similarly low, and SEC **(Fig 2.24)** indicated that aggregation was an issue, with a peak at 45 mL of almost the same absorbance as the now characteristic 85 mL peak. PAGE analysis **(Fig 2.24)** also proved familiar, with the 85 mL peak containing PqiC and the accompanying 25 kDa contaminant band. PqiC-F123E was collected and stored at 1 mg/ml as before.

Protein expression using the Y161E mutant resulted in a greatly increased yield. A SEC peak **(Fig 2.25)** of over 300 mAU was observed – a massive increase on W79E and F123E purifications, particularly. While aggregation was still an issue, the 45 mL peak was a reduced proportion of the higher retention volume signals. Interestingly, the common peak at 85 mL was split, with sub-peaks at 80 mL and 85 mL both apparent. This 80 mL peak was the first peak from mutant protein matching the 80 mL retention volume of WT PqiC. PAGE showed that the 80 mL peak contained a single 20 kDa band likely to be PqiC, with the 85 mL peak containing PqiC alongside the 25 kDa contaminant.

The fifth and final mutant protein to be purified was M164D. This mutant protein, when purified with **(Fig 2.26)**, appeared to be expressed at greater levels than the other proteins, with a peak of over 400 mAU detected. However, this peak was at 45 mL, implying that a large proportion of this expressed protein had become aggregated at some point during the expression or purification process. There were also peaks at just below and above 80 mL retention volume, which were likely to signal soluble PqiC presence. Results from PAGE **(Fig 2.26)** confirmed these observations, with PqiC bands visible from each SEC peak. The peak at just over 80 mL also contained the previously seen 25 kDa band. The peak at just below 80 mL

retention comprised almost exclusively 20 kDa PqiC, and the protein from this peak was collected and stored accordingly.

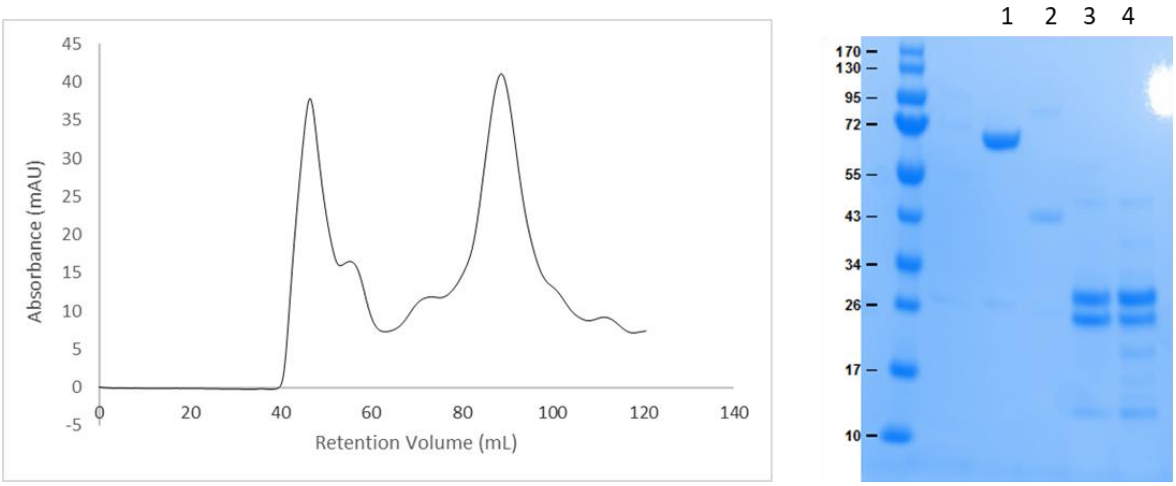


Figure 2.24 - SEC chromatograph and PAGE gel of truncated PqiC F123E purification. Protein expression was carried out using the same method as for WT protein without TEV cleavage. Following purification via IMAC, use of SEC (left) lead to two major peaks, one of around 50mL and one at around 85-90mL. PAGE (right) was used to determine the presence of PqiC, with samples run from the first major peak and its sub-peak (1 and 2 respectively), as well as the second significant peak (3,4). The first peak didn't appear to contain any PqiC, with neither the main peak nor the sub-peak containing any protein at 20kDa. The second peak, however, did contain protein likely to be PqiC. Observations were similar those made with W79E; retention volume was slightly increased from WT PqiC, and another larger band was visible on the PAGE gel.

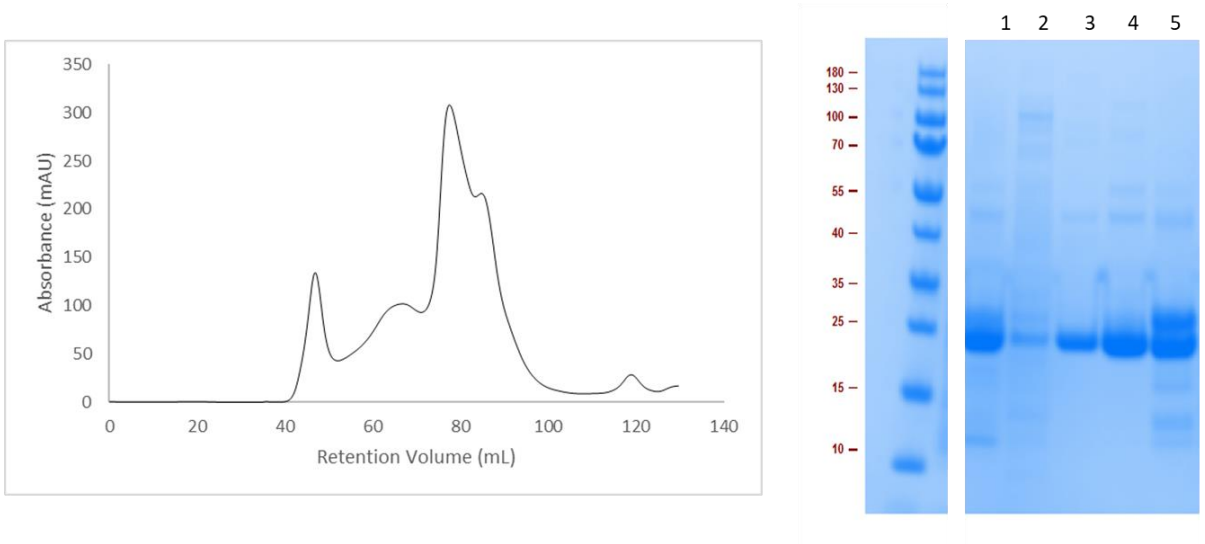


Figure 2.25 - SEC chromatograph and PAGE gel of truncated PqiC Y161E purification. Mutant PqiC was expressed in *E. coli* cells and purified using IMAC and SEC. A sample taken from after the IMAC process (1) showed a diffuse band at around 20 kDa, likely pointing to the successful expression of PqiC. The SEC chromatograph (left) contained 4 discernible peaks, each of which was further analysed with a PAGE gel (right). The peak at around 45 mL retention volume (2) gave rise to faint bands around 20 kDa and is likely caused by aggregated protein. The second and third peaks (3 and 4) both resulted in thick PAGE bands at the expected size for PqiC of 20 kDa, with the protein at around 80 mL retention volume consistent with observations in WT protein. The final peak (5) at 85 mL contained presumed PqiC, as well as the slightly larger contaminant band also found prominently in mutants W79E and F123E.

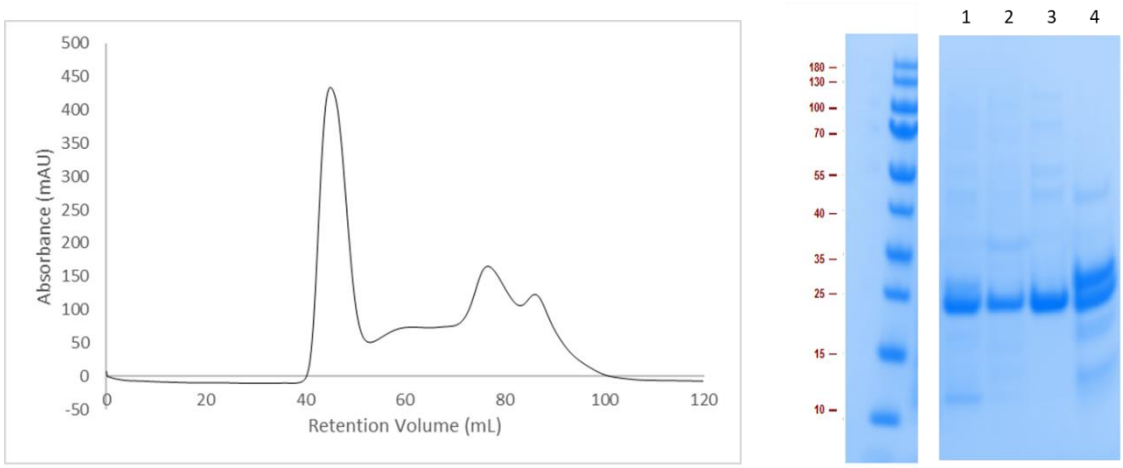


Figure 2.26 - SEC chromatograph and PAGE gel of truncated PqiC M164D purification. Protein purification was carried out in line with the other mutant PqiC species, without His tag cleavage. SEC (left) yielded a very large peak of over 400mAU at around 45mL. Interestingly, this peak is notably larger than those seen in the other four mutant proteins, yet at a reduced retention volume. A PAGE gel (right) was run to examine the purification results, with samples taken of the total SEC input (1), the prominent 45 mL peak (2), as well as the minor peaks just above and below 80 mL (3 and 4). The large 45 mL peak was almost exclusively PqiC, in a likely aggregated form. The higher retention volume peak also seemed to contain 20kDa PqiC at good purity, albeit with reduced yield.

To test whether Phospholipid binding was compromised by the introduced mutations, and indirect ELISA was used. A stock solution containing *E. coli* phospholipids was dried onto plates, allowing the PL molecules to adsorb to the surface. Then, samples of each mutant protein were added to wells, and allowed to bind overnight to the immobilised PLs. Bovine serum albumin (BSA) was included as a non-PL binding negative control. Taking advantage of the his-tags still attached to the PqiC mutants, protein was detected after washing steps using α-His antibodies, which were themselves the target of α-mouse antibodies. Using the conjugated HRP enzyme, a colorimetric reading was taken using the colour-changing substrate TMB. From the ELISA, the only protein to exhibit phospholipid binding was Wild-Type PqiC, with absorbance values of up to 0.658 seen, compared to near-zero values for mutant species - full data is available in the Appendix. While this assay indicated that PqiC does indeed bind PLs, optimisation of this preliminary test is required to draw further conclusions about the PqiC mutant species.

2.2.10 Preliminary PqiC Cell-based Assays

To test the effects of the introduced PqiC mutants in live cells, cell-based complementation assays were carried out. To determine its suitability as an anti-bacterial agent for cell-based assays against proteins from the Pqi pathway, the effects of Chlorpromazine on knockout cells lacking *PqiA, PqiB* or *PqiC* were first assayed **(Fig 2.27)**. Chlorpromazine was selected for testing given previous work showing its causing of growth defects in *E. coli* cells deficient in MlaA – a protein from the Mla pathway involved in phospholipid transport (59). While *PqiB* knockout cells appeared to be unaffected in these preliminary studies, the *PqiA* and *PqiC* cells seemed to be diminished by Chlorpromazine, and it appeared suitable to test the effects of single-residue PqiC mutations.

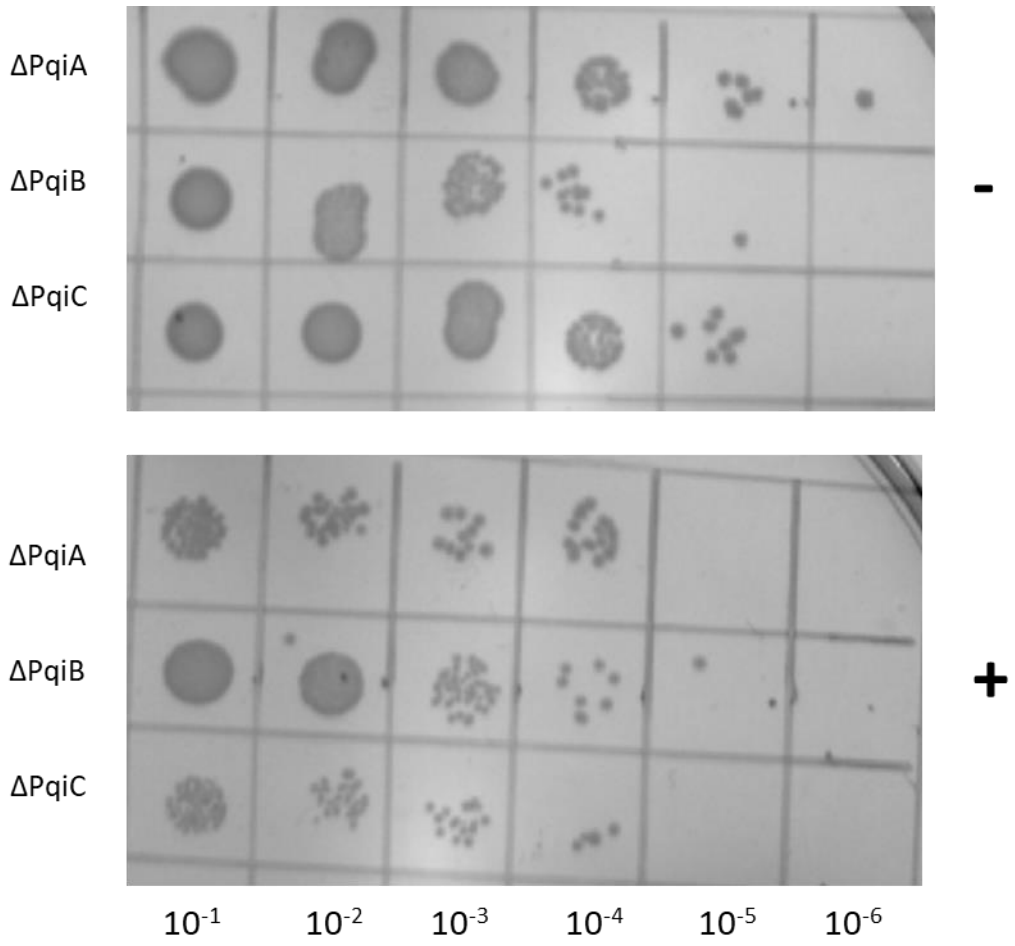


Figure 2.27 - Effects of the drug Chlorpromazine on *E. coli* cells deficient in components of the Pqi pathway. *E. coli* strains lacking the genes *PqiA*, *PqiB* or *PqiC* (all from the Keio Collection) were grown in the presence (+) and absence (-) of 120 µg/ml chlorpromazine. While the *PqiB* knockout appeared largely unaffected by the compound, both *PqiA* and *PqiC* exhibited reduced growth in its presence, with cells observed at lesser dilutions. At lesser dilutions, there was a marked reduction in the density of colonies.

To determine the effects of the introduced mutations *in vivo,* full-length mutant *PqiC* genes were produced, using the full-length pHisTev construct as an initial template. After the whole plasmid was replicated with PCR **(Fig 2.28)**, genes were transferred

to the pTrc99A-derived pHJ113 plasmid using a restriction digest and subsequent ligation, to allow for leaky expression of PqiC without significant overexpression (116). These pJH113 plasmids were transformed into chemically competent JW5127-1 cells from the Keio collection, deficient in the *PqiC* gene. Cells were grown with each of the five mutants, as well as wild-type PqiC and empty pJH113 plasmid **(Fig 2.29)**. Mutants Y64E and W79E appeared to be the most affected by chlorpromazine, with F123E similar manner to wild type protein; Y161E-containing cells were affected, but only moderately by the drug. These results indicate that Y64E and W79 are important to the function of PqiC, be that directly in Lipid binding, or in the structural stability of the protein's binding site.

However, these assays are inconclusive given the variable growth of knockout cells without chlorpromazine presence, and the lack of repeat datasets taken. This initial test represents only a preliminary investigation into the effects of chlorpromazine on *E. coli* cells expressing mutant PqiC species. Indeed, colonies of mutant M164D appeared to survive better than the equivalent wild-type cells in the presence of chlorpromazine; repeat tests are needed to confirm this observation. Further optimisation of these tests continues, and should allow for more definite conclusions to be drawn.

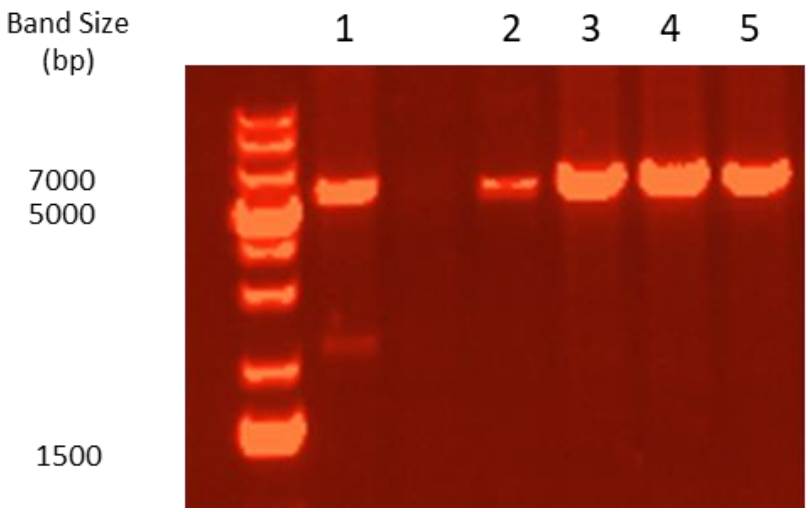


Figure 2.28 – Agarose gel of PCR full-length *PqiC* mutant plasmids, containing EtBr and imaged under UV. Each reaction used the full-length *PqiC* gene in the pHisTev plasmid as a template, and a primer pair intended to introduce single-codon differences in replicated DNA. The mutants Y64E (1), W79E (2) F123E (3) Y161E (4) and M164D (5) were all generated simultaneously. Template DNA was digested with *DpnI* prior to running samples on the gel.

2.2.11 Cloning of the Pqi Complex

To investigate the structure and function of the larger PqiABC complex, the genes encoding PqiA, PqiAB and PqiABC were all cloned into pHisTev to allow their expression. Each gene was amplified using PCR **(Fig 2.30)**, with expected sizes of around 1500 bp for PqiA, around 3000bp for PqiAB and 3500 bp for PqiABC all seen following gel electrophoresis. The amplified genes were restricted and ligated into the pHisTev vector, with cloning success confirmed with plasmid sequencing.

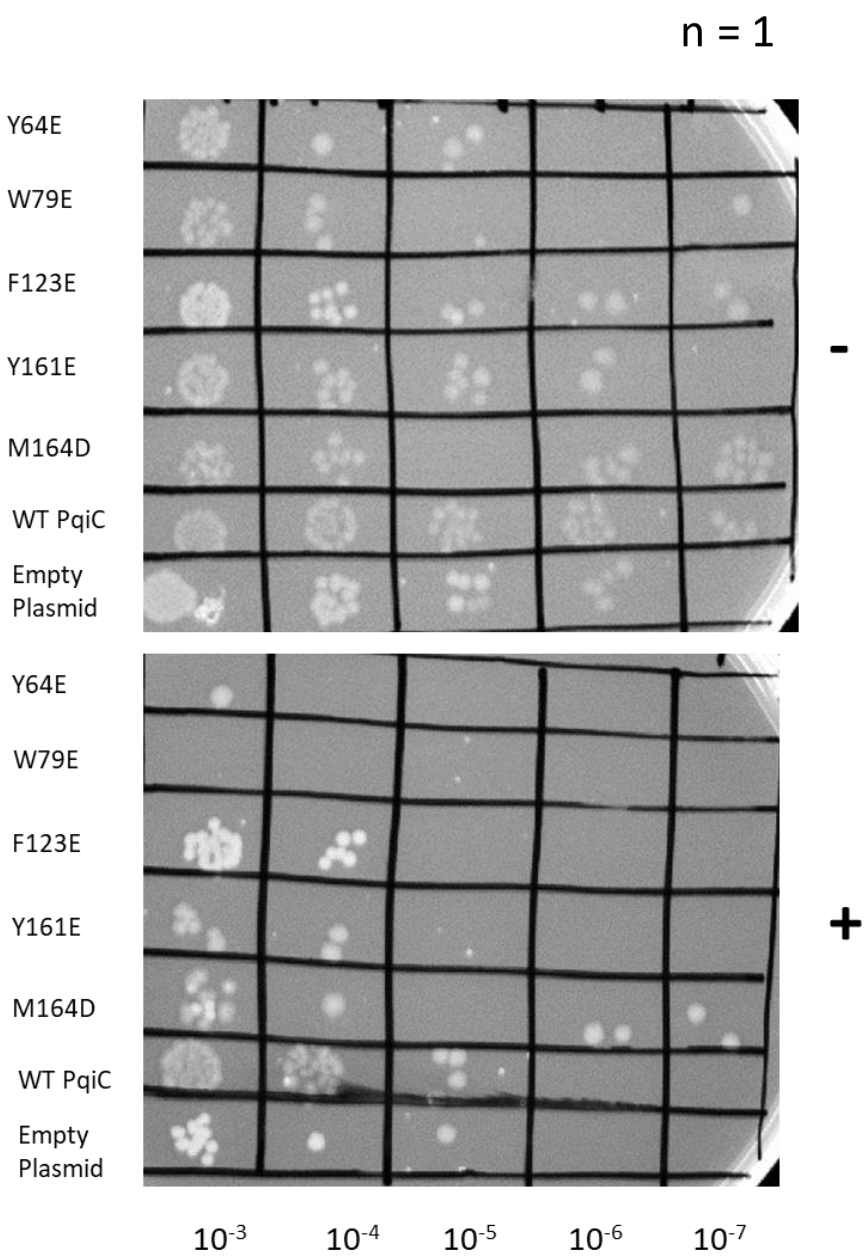


Figure 2.29 – Preliminary cell-based assays to determine Chlorpromazine sensitivity of *E. coli* cells with mutant PqiC. PqiC-knockout cells were transformed with pJH113 plasmid containing individual mutant PqiC species. As observed in initial testing, cells without functioning PqiC (and supplemented only with an empty plasmid) show growth only at lesser dilutions, and with reduced cell density. Cells transformed with the WT protein are still partially inhibited but continue to growth with similar cell densities at lesser dilutions. Mutants Y64E and W79E showed generally reduced growth, but both were heavily affected by chlorpromazine, with only a handful of colonies visible at even a 10^{-3} dilution. Cells carrying mutant F123E appeared to be affected similarly to the WT protein, with cell density comparable between the two plates. Y161E cells appeared to be partially affected by chlorpromazine, with a reduction in growth more apparent than the WT protein but not as severe as in Y64E or W79E. M164D mutant PqiC *appeared* to improve growth in the presence of chlorpromazine, but this is yet to be confirmed by repeat assays, given the preliminary nature of this test.

2.2.12 Expression of the Pqi Complex

For PqiA to be suitable target for structural analysis using X-ray crystallography, a pure concentrated protein sample was required. The assembled pHisTev_PqiA construct was transformed into C43 (DE3) *E. coli* cells with the heat shock method and used for protein expression. Cells were incubated overnight at 20°C with 0.2mM IPTG for expression, before cell lysis and debris removal. The sample was spun to pellet the membrane fraction, which was resuspended for solubilisation in 1% DDM, with any non-soluble components subsequently removed with a further spin. This clarified, soluble sample was used to purify PqiA from using Immobilised Metal Affinity Chromatography. Fractions from the purification were run on a PAGE gel **(2.31)** to assess whether PqiA had been successfully purified. After protein was purified with IMAC, the most prevalent band was indicated to be from a protein between 35 kDa and 40 kDa; this is contradictory of the known size of PqiA which is around 46 kDa. This difference could likely be an example of gel shifting, whereby the migration of membrane proteins on PAGE gels does not always directly correspond to molecular weight, potentially due to detergent binding behaviours (117). With this in mind, this initial test for PqiA tentatively suggests that the protein has been successfully expressed, and with further optimisation of this process could be a candidate for protein crystallisation trials.

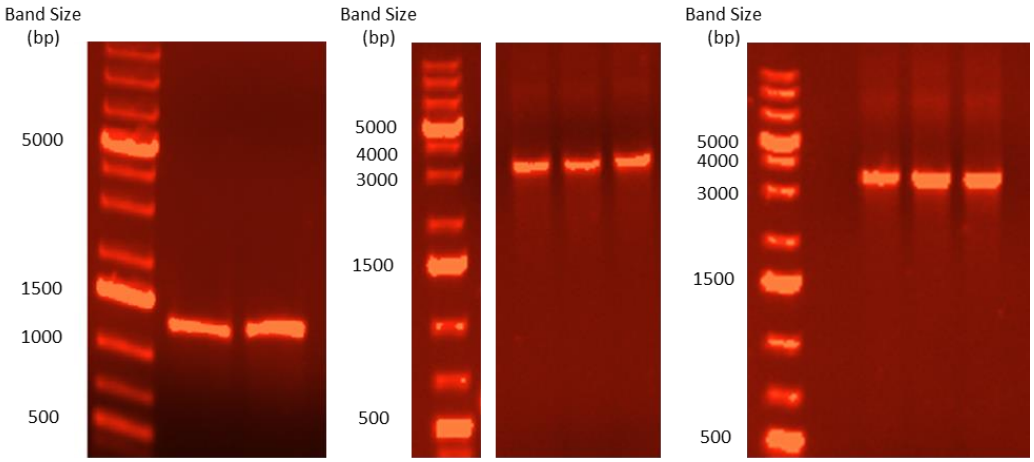


Figure 2.30 –Agarose gels of PCR products for cloning Pqi complex containing EtBr, imaged under UV. The *PqiA* gel (Left) had sharp bands indicated at around the expected 1400 bp region for the target gene, indicating successful amplification. A gel run following *PqiAB* PCR (Centre) contained bands showing at around 3000 bp, consistent with the length of the *PqiA* and *B* genes together. The gel of *PqiABC* PCR (right) showed slightly larger gene products, suggesting success in cloning the full *Pqi* operon of around 3500 bp.

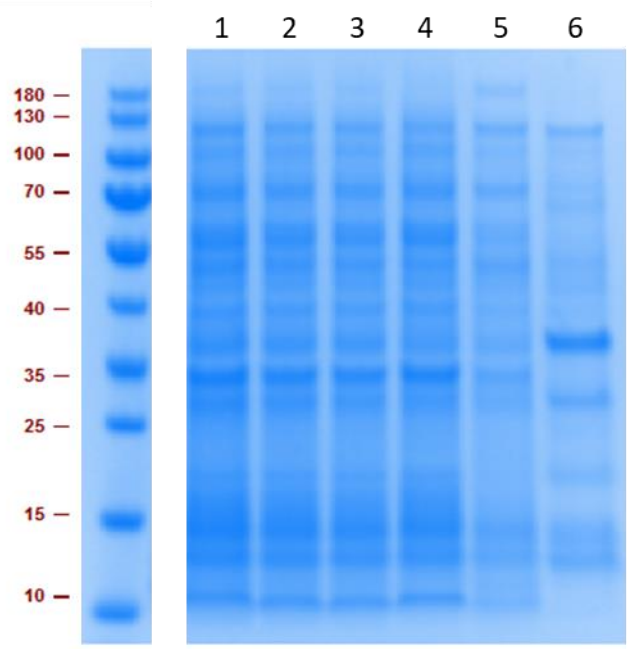


Figure 2.31 - PAGE analysis of PqiA expression test. PqiA was expressed in *E. coli* C43 (DE3) cells, in a total volume of 6 litres. The membrane pellet was solubilised in 1% DDM, with the successfully solubilised fraction used for IMAC purification. PqiA is difficult to pinpoint in samples taken before or after the DDM solubilisation step (1, 2), and the two samples appear similar. The solubilised protein sample following removal of non-soluble components (3) appears to be no different to previous samples. The IMAC loading flowthrough (4) and wash (5) steps looked alike, albeit with a reduction total protein amount. The IMAC column elution (6) contained a principal band at between 35 and 40 kDa, potentially pointing to successful PqiA expression, given the noted effects of gel shifting on membrane proteins.

With PqiA identified as a potential crystallisation target, expression of the protein was tested as a complex of PqiAB and PqiABC. These test expressions could help to confirm whether the three proteins form a stable complex, and suggest whether crystallisation of the entire complex would be feasible. For these protein preparations, pHisTev constructs incorporating C-terminal His-tags into either the PqiB or PqiC proteins were used, again with *E. coli* C43 (DE3) cells as the expression platform. As for PqiA, the membrane fraction was solubilised in 1% DDM and used for IMAC followed by SEC for purification. When run on SEC **(Fig 2.32)** the largest peak observed was found at 45 mL retention volume, around the value of the column's void volume. Another smaller, broad peak was seen at 80 mL. Running samples from either peak on a PAGE gel **(Fig 2.32)** indicated that the larger 45 mL peak contained two predominant proteins – one indicated at around 65 kDa, and one between 35 and 40 kDa. With its expected size at 60 kDa, it is quite likely that this larger molecular weight band indicates successful PqiB expression, particularly given the liability of membrane proteins in detergent to migrate non-uniformly during PAGE analysis. The 35-40 kDa band appeared to correspond to the PqiA from its own expression test and serves as an indicator that PqiA and PqiB formed a stable complex during purification, given that only PqiB contained the His-tag necessary to bind to the IMAC column. Interestingly, the small 80 mL peak contained a double band at around 25 kDa, to observations from PqiC expressions. With further investigation, this could indicate that the PqiC protein also forms a complex with PqiAB, albeit in a more transient nature, given its evident separation by SEC.

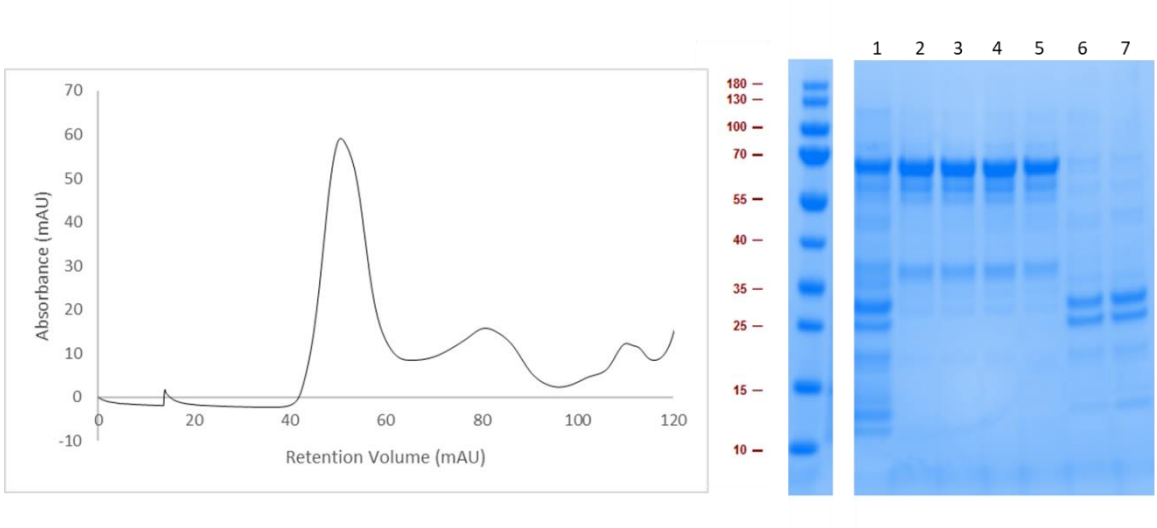


Figure 2.32 - Coomassie-stained PAGE gel of PqiAB purification. PqiAB was expressed in C43 (DE3) *E. coli* cells, with a C-terminal His-tag on PqiB. A PAGE gel (right) was run of samples from SEC (left) peaks, as well as the total SEC input (1). The bulk of the protein was detected in a peak with retention volume of around 50 mL, with a smaller peak at around 80 mL containing some protein. Samples taken across the first peak (2-5) contained two major bands – one at around 65 kDa, and one at between 35 and 40 kDa. From a previous expression test, the 35-40 kDa band is likely to be PqiA; PqiB has a molecular weight of 60 kDa, so the larger band is potentially indicative of this. The 80 mL peak (6 and 7) contains a double band centred around 25 kDa, which could suggest PqiC presence, although confirmation with MS is required for this.

After expression of PqiAB had suggested the formation of a stable complex, PqiABC was expressed, with only the PqiC protein itself labelled with a His-tag. Following purification in the same manner as for PqiAB, results from SEC and PAGE were analysed **(Fig 2.33).** The SEC chromatograph had two peaks; a smaller 30 mAU peak at 45 mL and a peak of almost 200 mAU centred at around 75 mL. PAGE run of the first peak generated bands at just below 70 kDa and 25 kDa, pointing to PqiB and PqiC presence, given previous observations of the proteins at these sizes. The larger SEC peak contained a very strong band at around 20 kDa, pointing to significant expression and purification of PqiC, not unexpectedly. However, also visible are bands at 55 kDa and a very faintly stained band at 35-40 kDa. This faint band matches that seen from PqiA purification and could suggest that PqiC was purified as a complex containing this protein. The 55 kDa band is a new observation, and may be a PqiB cleavage product, or perhaps an as yet undefined protein. While the possible presence of PqiA from the PqiC-only tagged construct strongly suggests a relatively stable PqiABC complex, identification of the 55 kDa band could provide important insight into the behaviour of PqiABC *in vivo*.

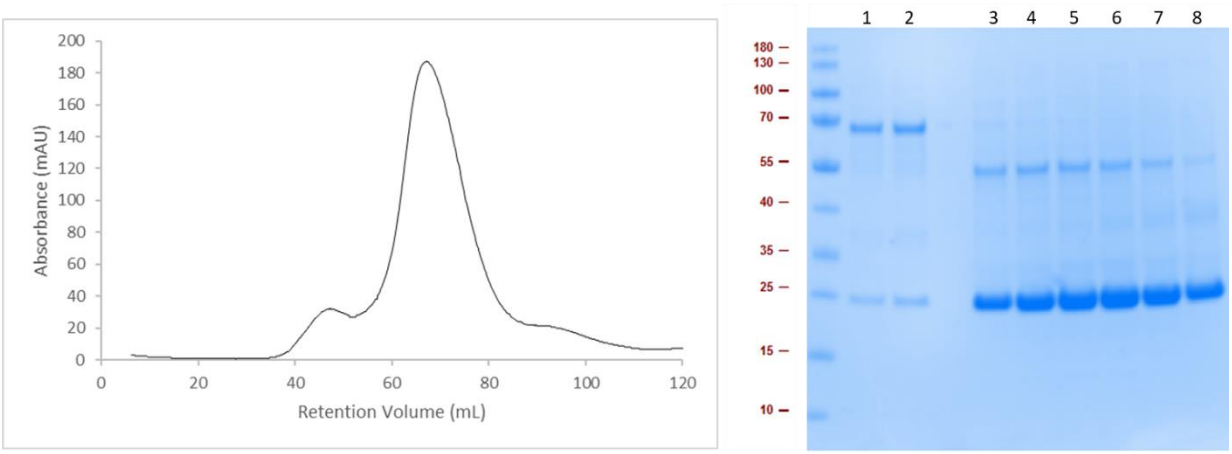


Figure 2.33 - PAGE gel stained with Coomassie of PqiABC purification. PqiAB was expressed in C43 (DE3) *E. coli* cells, with a C-terminal His-tag on PqiC. The protein eluted from the SEC column (left) in two peaks – a very small peak at 45 to 50 mL of about 20 mAU and a much larger peak at 70 mL of around 180 mAU. Samples from each peak were run on a PAGE gel (right). The first peak (1 and 2) contained two bands, one at just under 70 kDa and one below 25 kDa; previous experiments suggest these are potentially PqiB and PqiC, respectively. The Larger peak at higher retention volume (3-8) was primarily composed of a likely PqiC band below 25 kDa, given its tagging. Also in these fractions is a band at around 55 kDa of unknown origin, and a very faint band of between 35 and 40 kDa which points to possible PqiA presence.

2. 3 – Discussion

The Gram-negative outer membrane protects cells against toxic compounds and thus can convey antibiotic resistance to pathogenic bacteria (118). Furthering the understanding of the maintenance and assembly of this bacterial structure is important, as advances in this knowledge could provide targets for novel antimicrobial agents. Such new targets are increasingly important, given the dearth of new antibiotic agents under clinical trials (119).

The outer membrane forms a barrier around Gram-negative cells, with its layer of Lipopolysaccharides (LPS) at the outer leaflet, and an inner leaflet containing Phospholipids (4). The pathway by which LPS has been well studied, with structural and functional analysis carried out on its transport machinery, the Lpt complex (38). This is a large, transmembrane protein complex which passes the hydrophobic LPS across the periplasmic space prior to its insertion into the OM through the β-barrel LptD protein, with mediation by LptE (43). By contrast to the well-studied structures and mechanisms of Lpt action, the transport of Phospholipids (PLs) between the IM and OM remain mysterious. The best described transport machinery for PLs is the Mla. The Mla system is made up of proteins in the periplasmic space, as well as inner membrane proteins and an outer membrane portion (56, 57). Key to its function is the soluble periplasmic MlaC protein, which shuttles proteins between inner and outer membrane components, although the directionality of its transport remains contentious (63). While not initially determined as a part of the pathway, the β-barrel OmpC protein was later seen to interact with the OM localised MlaA protein, and play a key role in PL translocation (57).

Given the mammalian cell entry domain of its PqiB component (60), the *PqiABC* operon has been suggested to be important in PL trafficking in Gram-negative bacteria (68). The *PqiABC* genes are conserved in pathogenic Gram-negative organisms, including *Pseudomonas aeruginosa,* and has the potential to be a good target for novel therapeutic agents (120). While experimental evidence has implicated PqiB directly in PL transport and identified PqiC as an interactor, little else is known about their roles *in vivo* and how they interact with phospholipids.

The work outlined here focused on elucidating the structure of PqiC, to allow its significance in both the PqiABC pathway and in phospholipid trafficking in *E. coli* as a whole to be better understood. The structure of the protein was solved to 2.40 Å using X-ray crystallography **(Fig 2.18)**, which revealed a 'hairclip' like structure, with two α-helices of different lengths facing a region of 3 long β-sheet, with a cleft left between the helices. Examination of this new model lead to the discovery of unmodelled electron density, which could be attributed to phospholipid binding **(Fig 2.20)**. *E. coli* PL was shown to fit this density well, with phosphatidylethanolamine shown as an example. This crystal structure can be used to tentatively confirm the binding of Phospholipids by PqiC.

Comparison of the newly built PqiC structure with that of existing proteins noted its similarity to several putative lipoproteins, as well as the LPS transport-associated LptE protein, previously described as the 'plug' into the β-barrel of LptD (43). A superimposition of PqiC and LptE confirmed how similar the two proteins were, with an RMSD of 4.2 Å **(Fig 2.19)**. This similarity was a promising indicator that PqiC may be indeed be a PL binding protein, given that the structure of LptE allows it to bind similar hydrophobic regions in LPS (113).

While LptE appeared markedly similar to PqiC structurally, the possible Mla pathway orthologs of MlaA and MlaC bore far less resemblance to PqiC. While superimpositions between the structures were attempted, these proved of little significance **(Fig 2.34)**. MlaA – the OM anchored lipoprotein involved in PL transfer to MlaC (57)– is structurally very distinct from PqiC. MlaC – the periplasmic PL transporter (58)- is more similar to PqiC, and has been shown to bind potential phospholipid tails in its central binding cavity (60). The similarity to LptE was of particular interest, especially considering the proposed binding site of LptE (a loop between the first long β-sheets) was at a similar site to the presumed PL density in the novel PqiC structure (113).

Using the putative LptE LPS binding site, as well as the proposed PL density in PqiC, 4 residues were selected for mutagenesis, to determine their effects on PqiC. Additionally, a phenylalanine residue in the central groove of PqiC (analogous to the binding site of MlaC) was chosen for study, given the similar functions between the two proteins and the potential that a PL binding site in PqiC could extend further than the observed electron density would suggest. Genes for these single amino-acid substitution mutants (Y64E, W79E, F123E, Y161E and M164D) were used to produce soluble protein for PL-binding assays truncated of their signal peptide, as well as plasmids containing full length PqiC for preliminary cell-based assays. Incidentally, the M164D mutant **(Fig 2.26)** PqiC species was expressed in much greater volume than its counterparts under the same conditions, and would seem to suggest a structural significance to this residue. Mutants W79E and F123E **(Fig 2.23/4)** both purified with poor yield, and a larger proportion of contaminant protein of around 25 kDa. While this is potentially due to protein degradation and subsequent interaction with full-length PqiC, further investigation with mass-spectrometry is

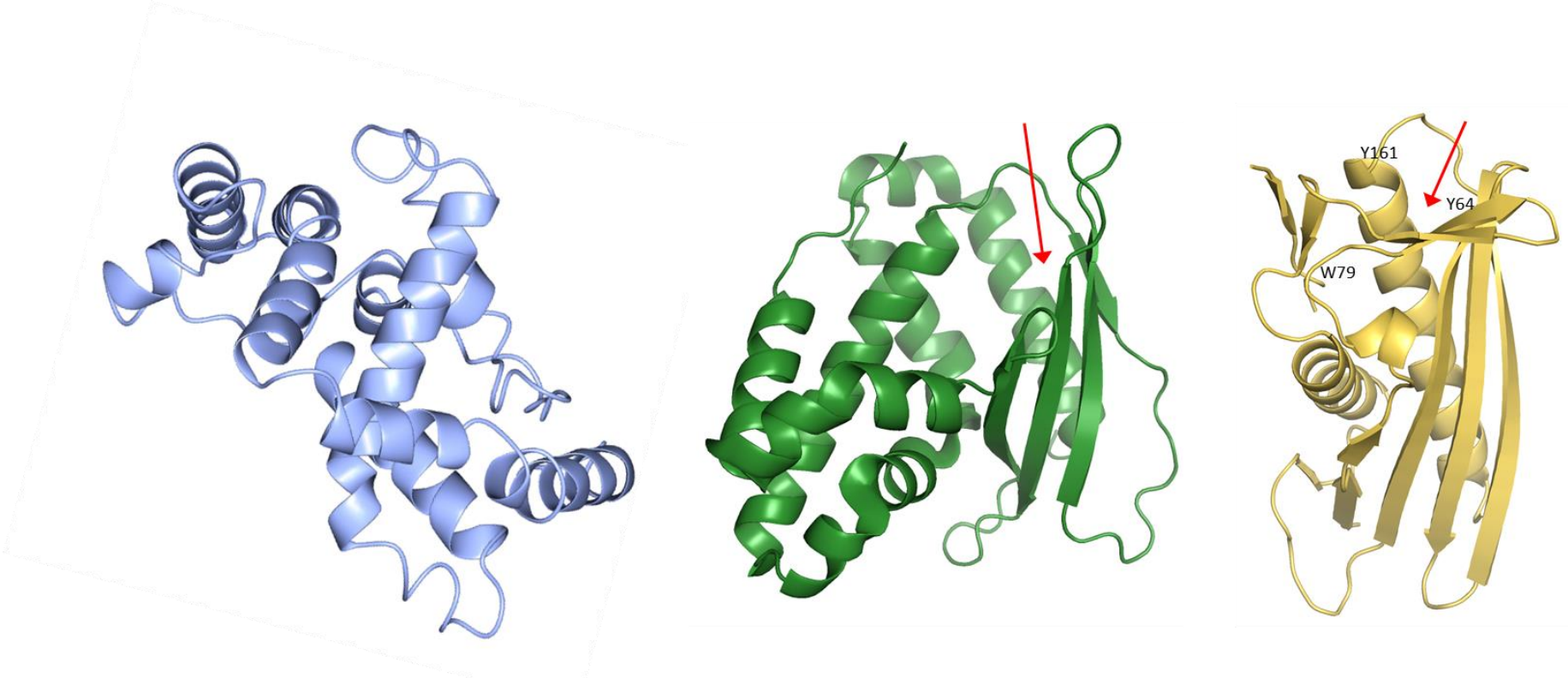


Figure 2.34 - Structures of MlaA (left), MlaC (centre) and PqiC (right). Given their potentially related cellular functions in phospholipid transport, the structures of OM lipoprotein MlaA and periplasmic PL transporter MlaC were compared with PqiC. Superimpositions were attempted which proved unfruitful. MlaA is structurally very different to PqiC, despite their shared presumed cargoes of phospholipid. MlaC *does* have a pair of helices facing a region of β-sheet, and has a binding site for PLs in the deep groove between the two (indicated with a red arrow), which may indicate a similar binding site in PqiC. The putative phospholipid binding site on PqiC is indicated with an arrow, with the location of three potentially involved residues shown. MlaA (5NUP) and MlaC (5UWA) structures from (57) and (58), respectively.

needed to confirm whether these unexpected protein bands represent instability of mutant PqiC or a separate protein. In any case, the poor yields of both W79E and F123E would suggest that the residues are of some structural importance to the protein.

Preliminary cell-based complementation assays **(Fig 2.29)** indicated that the growth of PqiC-deficient cells was reduced with the addition of the drug chlorpromazine. While WT PqiC recovered cell growth to a degree, this recovery was hampered in Y64E, W79E and Y161E mutants. This information - alongside the potential observed PL binding site in the PqiC crystal structure – tentatively suggest that Y64, W79 and Y161 are residues important in the binding of Phospholipids in PqiC. For verification of these findings, optimisation of these assays is continuing. To further investigate this, the purified soluble PqiC mutant proteins were used for lipid-binding assays. Initially, extraction of PLs and identification using thin-layer chromatography (TLC) was trialled as outlined for PqiB and YebT in (60), but results from this proved inconclusive. Instead, the purified PqiC was used for an indirect ELISA, against *E. coli* phospholipids immobilised on a 96-well plate. However, from this work, only wild-type PqiC was shown to bind to phospholipids, with repeat measurements needed for confirmation. While this was not suitable data to confirm the involvement of specific residues in PL binding, it can be used to further indicate the binding of phospholipids by PqiC. Optimised cell-based and *in vitro* assays would allow for more definite conclusions to be drawn regarding the involvement of specific residues in PL binding in PqiC.

Alongside direct work on PqiC, investigations were carried out on the PqiABC complex as a whole. Initially, expression of PqiA was tested **(Fig 2.31)**. This proved to

be a moderate success, identifying it as a target for future crystallisation studies to glean more information about its role in potential PL transport. In addition, PqiAB and PqiABC were purified using affinity tags on PqiB and C respectively, to determine whether the proteins form a complex *in vivo,* given the noted interaction between PqiB and PqiC in previous studies (68). PqiAB purified using tagged PqiB **(Fig 2.32)** gave rise to a protein sample containing two distinct species, including a protein of identical size to that yielded from PqiA overexpression. Furthermore, a protein of similar size to PqiC was detected from this purification. However, identification of these bands using mass-spectrometry is important in further work to determine if this is indeed PqiC, or just coincidental degradation products of PqiAB. From the expression, it can be inferred that PqiA and PqiB do indeed form a complex in *E. coli* cells, given that both were observed after their overexpression despite an affinity tag only present on one partner. Should the potential PqiC band be confirmed, it would indicate that the interaction between PqiC and PqiB previously demonstrated with a yeast two-hybrid assay also occurs in *E. coli* cells (10). Given that PqiA interacts with PqiB, an ineraction between PqiB and PqiC, would suggest that the three proteins may form a complex *in vivo,* albeit in a transient manner.

A subsequent overexpression test of PqiABC (with tagged PqiC) yielded comparable results. In one fraction from size-excusion chromatography during purification **(Fig 2.33)**, clear bands for PqiC and PqiB could be seen, giving further evidence that they associate with one another *in vivo* to form a transmembrane complex. However, a more porminent SEC peak contained a band signifying PqiC, a very small amount of possible PqiA and an appreciable amount of an unknown protein at 55 kDa. This is an exciting result, and can suggests one of two possibilities. First, that the 55 kDa band is a product of PqiB, and PqiABC does form a relatively stable transmembrane

complex. Second, it could mean that PqiC interacts with another protein altogether under physiolocigal conditions. As previously outlined, LPS insertion and PL transport using the Mla system are both dependent on integral outer membrane proteins, and this evidence could suggest that PqiC also interacts with an OMP. Further analysis of the detected protein is required, but this result could represent the first observations of an OMP bindning partner for PqiC, akin to OmpC and its interaction with MlaA (57). Indeed, it's not beyond the realms of possibility that PqiC could form the plug in an outer membrane barrel, given its striking similarity to LptE (43).

CHAPTER 3 – Structural Studies on the YebST proteins

3. 1 – Materials and Methods

3.1.1 Amplification of *YebST* genes

To allow their expression and purification from *E. coli* cells, Yeb pathway genes were transformed into cells using a plasmid vector – the first step of this process was to amplify the desired genes. The *YebT*, *YebS* and *YebST* genes were amplified from *E. coli K12* genomic DNA using PCR. Suitable primers for each reaction were designed with melting temperatures within 5°C of one another. Primers for *YebS* and *YebST* contained sites for the restriction enzymes *NdeI* and *HindIII,* with the reverse primer including codons to introduce a C-terminal hexa-histidine tag when expressed using the pHisTev plasmid. Both of these primer pairs cloned their respective genes in their full lengths.

The designed *YebT* primers contained sites for *BamHI* and *HindIII,* resulting in N-terminal tags from expression plasmids pLou3 or pHisTev. This primer pair cloned the *YebT* gene with a 120 base-pair N-terminal truncation, to allow its expression as a soluble protein. Full sequences for all three primer pairs are available in **Table 3.1**, all were synthesized by EuroGentec.

Table 3.1 – Primers used for cloning the components of the YebST pathway. Each primer pair was used with *E. coli* gDNA as a template for gene amplification.

Primer Name	Sequence (5'-3')
YebT_F	CTACTCGGATCCAGTTATCAGGACCGGGGTAATACCGTC
YebT_R	AGAGCAAGCTTTTATTTGGGAAGCGCAGTTCCCCATTC
YebT_Kpneu_F	CTACTACCATGGCTAGCTTTGACGATCGCGGATCGACGATTACG
YebT_Kpneu_R	GAGGCTGGATCCTTACTGCGGCAGAGCGGTGCCCCACTCCCGCCAC
YebST_F	CTACTGCATATGATGGCTCTTAACACACCACAAATTAC
YebST_R	CGAGCAGGATCCTTAATGATGATGGTGATGATGTTTGGGAAGCGCAGTTCCCC
YebS_F	ATTACATATGGCTCTTAACACACCACAAATTAC
YebS_R	AATTAAGCTTTCAGTGATGGTGATGGTGATGGTCGTCGAAGCGGGCGTTTC

For all 3 separate PCRs, the Q5 hot-start system (NEB) was utilised, with individual reagent volumes in **Table 2.2**; each reaction used 21 ng of *E. coli* gDNA as a PCR template. The primer pairs for *YebS, YebST* and *YebT* had melting temperatures of 69°C, 63°C and 71°C, respectively. The PCR program used was: initial denaturation at 98°C for 30 seconds, followed by 34 cycles of: 98°C for 10 seconds; Tm°C for 30 seconds and elongation phase at 72°C for 120 seconds. Reactions had a final elongation phase at 72°C for 120 seconds, and were subsequently held at 4°C.

The success of PCR reactions was determined using agarose gel electrophoresis, using the same methodology as for *PqiC* cloning validation, with *BamHI* and *HindIII* enzymes (Section 2.1.3).

3.1.2 Generation of YebST expression plasmids

Following a successful amplification of *Yeb* genes using PCR, genes were digested with restriction enzymes and inserted into plasmids. FastDigest restriction enzymes (Thermofisher) were used to cut synthesised *Yeb* genes, using the appropriate enzyme pair matching their primers. For 17 µl of each gene, 2 µl FastDigest buffer and 1 µl of each restriction enzyme was added. These reactions were incubated at 37°C for 30 minutes, followed by heat inactivation at 85°C for 5 minutes. Restriction reactions underwent cleanup using a GeneJet gel extraction kit (Thermofisher) and eluted in a volume of 20 µl. DNA was mixed in a ratio of 3:1 gene to plasmid, with 9 µl of the resulting mixture used for ligation reactions. Ligation took place at room temperature for 30 minutes, with 1 µl of Quick Ligase and 10 µl Quick Ligase Buffer (both NEB) added to the 9 µl DNA mixture.

Plasmids ligated with *Yeb* genes were transformed into chemically competent *E. coli* Top10 cells for amplification using heat shock. The same heat shock process used to transform *PqiC*-containing plasmids into competent cells in Section 2.1.2 was used for this.

A GeneJet Plasmid extraction Kit (Thermofisher) was used to extract plasmids from transformed cell cultures. The presence of *Yeb* genes in amplified plasmid was established via a double restriction reaction, using 10 µl of plasmid, mixed with 2 µl Fast Digest buffer and 1 µl of each required FastDigest restriction enzyme. Samples were incubated at 37°C for 30 minutes, then run on a 1% agarose gel for 45 minutes at 110V. Samples with visible gene inserts under UV illumination were submitted to Source Biosciences for sequencing as a final verification step. With gene inserts confirmed, plasmids were transformed into expression strains using the heat shock method.

3.1.3 Overexpression of YebST proteins

The different components of the Yeb pathway were expressed in different strains of *E. coli. YebT*, was expressed as a soluble protein without its transmembrane domain in *E. coli* SB21 (DE3) cells. *YebS* and *YebST* were expressed in *E. coli* C43 (DE3) cells, a suitable cell line for membrane protein expression. Initially, 300 mL of LB media was inoculated with cells from frozen glycerol stocks and incubated overnight at 37 °C, with 200rpm shaking and 50 µg / ml kanamycin. Fresh LB media supplemented with kanamycin was inoculated with 25 ml per litre of overnight culture and grown at 37 °C with 200rpm shaking until OD_{600} reached 0.6. Protein expression was induced with IPTG to a final concentration 200 µM. Protein expression was carried out overnight

at 20°C, with 200rpm shaking, before harvesting via centrifugation. Cells were spun at 3000 g for 20 minutes and stored at -20 °C prior to protein purification.

3.1.4 Selenomethionine Labelling of YebST

In addition to its native expression in LB, truncated YebT was also labelled with Selenomethionine (SeMet) during some expression procedures for use as a phasing agent from any resulting protein crystals. The basic SeMet labelling process was adapted from (121). As before, YebT was expressed in *E. coli* SB21 (DE3) cells; these cells were grown overnight at 37 °C in 300 mL LB with the addition of 50 µg / ml kanamycin. For use in growth medium, a sterile solution of 20 x M9 salts was made with: 60 g Na_2HPO_4, 30 g KH_2PO_4, 10 g NH_4Cl and 5g NaCl in a total volume of 500 mL. In addition, an amino acid solution containing Lysine, Phenylalanine and Threonine (all 1g) and Isoleucine, Leucine and Valine (all 0.5g) in a volume of 250 mL was made and filter sterilised.

A volume of 8 litres of SeMet labelling media was produced, with 900 mL sterile H_2O, 50 mL of 20x M9 Salts, 50 mL of 1x Selenomethionine Nutrition Mix (Molecular Dimensions) and 50mg of Kanamycin per litre. Cells from the overnight culture were spun for 10 minutes at 2,000 g and the pellet resuspended in 250 mL 1x M9 salts. This washing step was repeated, to ensure that LB was not carried over into the SeMet labelling media. Each litre of labelling media was inoculated with 30 mL of washed overnight culture and grown at 37 °C with 200rpm shaking until OD_{600} reached 0.7. At this point, each flask had 25 mL of the amino acid solution added, and was incubated for a further 15 minutes, now at 20°C. Next, Selenomethionine was added to a final concentration of 60 mg /L and incubated for 15 minutes. Finally, IPTG was added to a concentration of 0.2mM and the cells were incubated overnight

to express SeMet-labelled YebT. Cells were harvested and stored in the same manner as with native protein expressions.

3.1.5 Purification of YebT

YebT was purified at a later date from the cell pellets frozen after protein overexpression. Frozen cell pellets from YebT expressions (both native and SeMet labelled) were resuspended in lysis buffer, lysed using a cell disrupter and clarified with centrifugation, using the same buffers and methods as PqiC (Section 2.1.5). Following these stages, YebT was purified using nickel affinity chromatography and size exclusion chromatography, again using the same protocols and buffer compositions as for purifications of PqiC (Section 2.1.5).

Polyacrylamide gel electrophoresis (PAGE) was used to investigate individual purification steps by separating protein by molecular weight. The Bolt system tank and reagents were used, with a 4-12% Bis-Tris precast gel (Thermofisher). The gels were run for 22 minutes at 220 Volts, with 5 µl of PageRuler Prestained Ladder (Thermofisher) and 26 µl of each prepared sample. Gels were stained for 30 minutes using Quick Coomassie (Generon) and rinsed overnight in water.

3.1.6 Purification of YebS

As with the Pqi complex, as a membrane protein, YebS required a purification method using detergents and membrane fractionation steps. The purification process used was the same as for the Pqi Complex (Section 2.1.18), except buffers were supplemented with 1% n-Dodecyl β-D-maltoside (DDM) for membrane solubilisation and column binding, reducing to 0.05% DDM during IMAC washing and elution steps, with 0.025% DDM used during SEC

An additional YebS purification was carried out using these methods, with the detergent n-Dodecyl-N, N-Dimethylamine-N-Oxide (LDAO) in place of DDM for IMAC wash and elution, as well as in the SEC running buffer, at 0.05% in each case.

3.1.7 Crystallisation Trials for YebT and YebS

Once protein had been shown to be pure by PAGE analysis, it was used for crystallisation trials. Protein was concentrated for crystallisation trials using Vivaspin concentrators at 3,000 g; concentrations from 5 – 15 mg/ml were used. Crystal trials for both YebT and YebS used sitting drop vapour diffusion method with 2-well plates. The top well used a volume of 1 µL and bottom well 0.6 µl; both used 1:1 protein: screen solution. Plates were made using a Crystal Gryphon robot, sealed with ClearVue sheets and incubated at 21°C. YebT trials utilised a combination of commercially available soluble protein-optimised screens **(Table 2.3**), with initial YebS trials carried out primarily using screens developed for crystallisation of membrane proteins **(Table 3.2)**. To improve crystal size and quality, initial crystal hits were further optimised with specific hand-made screens, by varying salt or precipitant concentrations, as well as pH across the condition from which the initial hit was seen.

Table 3.2 – Crystallisation screens used in YebS crystal trials. The below screening solutions were transferred to a deep well block for use with the Crystal Gryphon liquid handling robot.

Crystallisation Screen	Manufacturer
Mem Start	Molecular Dimensions
Mem Sys	Molecular Dimensions
Mem Gold	Molecular Dimensions
Mem Gold2	Molecular Dimensions
Mem Channel	Molecular Dimensions
Mem Trans	Molecular Dimensions
JCSG Plus	Molecular Dimensions
PACT Premier	Molecular Dimensions

3.1.7 Data Collection from YebT Crystals

Crystallisation plates were periodically checked for crystal growth manually using a stereomicroscope. Once selected for diffraction studies, selected crystals of YebT were soaked for 5-15 minutes in cryoprotectant buffers. These consisted of mother liquor with glycerol to 20% and included 200 mM Sodium Iodide in cases where a phasing agent was absent – cryoprotectant was added in a 1:1 ratio with drop size. Crystals were removed from the plate wells in litholoops from 0.1 - 0.3 mm in diameter, before flash freezing in liquid nitrogen and insertion into unipucks. These were loaded into a Cryo-Express Dry Shipper and sent to Diamond Light Source for remote use of their X-ray beamlines. Data for Iodide phasing was collected at 1.8501 Å, with data collections for Selenomethionine carried out at 0.9795 Å. Other parameters were varied, with 3600 images at 0.1 second exposure, 40% beam transmission commonly used as a starting point.

3. 2 – Results

3.2.1 Cloning of YebST pathway genes

The initial step for investigating the YebST pathway was the cloning of the genes for both constituent proteins, in addition to the complex as a whole. For this, 3 specific primer pairs were used, with each containing the recognition sequence for a specific restriction enzyme. Using these primers, the full-length genes of *YebS* and *YebST*, along with the periplasmic region of *YebT* were amplified from *E. coli* genomic DNA using a PCR reaction, which was analysed using agarose gel electrophoresis **(Fig 3.1)**. When imaged under UV, bands were seen for *YebS* were seen at 1200bp, *YebST* at around 4000bp and for truncated *YebT* at 2500bp. Each of these bands corresponded to the known sizes of the genes, and thus the PCR was deemed successful. Following successful PCR, the genes were restricted with

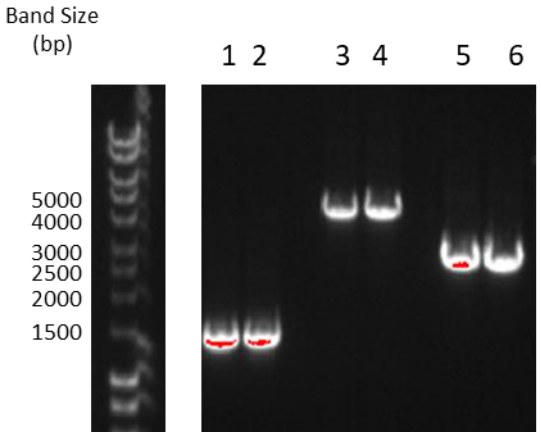


Figure 3.1 –Agarose gel of *YebST, YebS and YebT* PCR products containing EtBr, imaged under UV. PCR products for *YebS* (1,2) were visible around the predicted 1200bp. *YebST* was seen as expected at around 4000bp (3,4), and bands corresponding to *YebT* with its N-terminal truncation were apparent around the anticipated size of 2500bp (5,6).

the appropriate restriction enzymes given the primers used and ligated into the pHisTev or pLou3 plasmids. After a restriction digest, ligated plasmid samples were run on an agarose gel to confirm the presence of a gene insert **(Fig 3.2/3)**. For each construct, bands were observed of consistent size to those seen from electrophoresis following PCR.

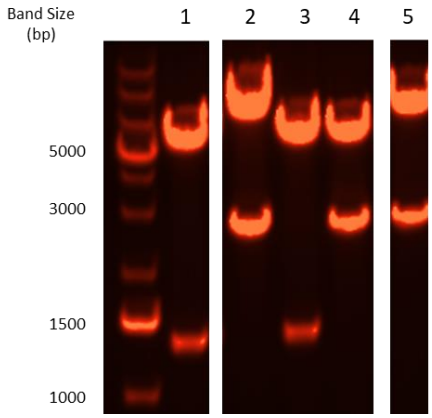


Figure 3.2 – Validation of *YebS and* truncated *YebT* ligations into expression plasmids through double restriction digest, run on an Agarose gel. pHisTev:*YebS* (1, 3) was restricted with *NdeI & HindIII*, while pLou3:*YebT_truncated* (2,5) and pHisTev:*YebT_truncated* (4) were restricted with *BamHI & HindIII.* The plasmids were all run on an EtBr containing gel and imaged under UV. In each case, plasmids were visible at over 5000bp, with pLou3 noticeably larger at around 7000bp. In all lanes, there are sharp bands in the expected range for the genes, with *YebS* around 1200bp and *YebT* at 2500bp in its truncated form.

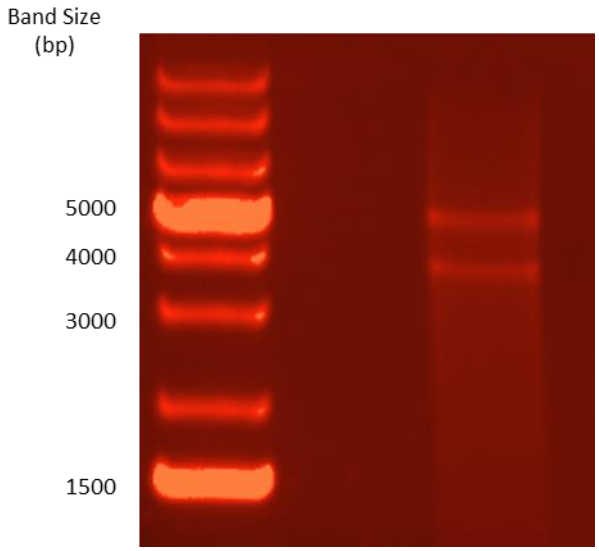


Figure 3.3 – Validation of *YebST* ligation into plasmid pHisTev through double restriction digest, run on an Agarose gel. pHisTev:*YebST* was restricted with *NdeI & HindIII*, and run on an EtBr containing gel and imaged under UV. A plasmid band is visible at around 5000bp, with a band corresponding to Y*ebST* seen at just under 4000bp.

3.2.2 Expression of the YebT periplasmic region

With both individual *Yeb* genes and the two-protein complex cloned into expression plasmids, it was decided that the truncated YebT would be the first protein to be expressed. With its membrane anchor lacking, it was postulated to be easier to express in greater amounts as a soluble protein. The truncated *YebT*:pHisTev plasmid was transformed into SoluBL21 (DE3) *E. coli* cells for protein expression with a reduced tendency for aggregation. With an IPTG-induced overnight expression, 4 litres of cells were used for an initial expression test. YebT was initially purified with IMAC, with a PAGE gel used to analyse the initial expression level **(Fig 3.4)**. From this preliminary purification, it was apparent that YebT was expressing in good amounts, with a very intense band visible at around 95kDa in the IMAC elution fraction, with minimal protein of the same size visible in column wash or flowthrough fractions. As a result, protein resulting from this initial test was purified further, using TEV proteinase tag-cleavage, a pulldown through IMAC column and Size exclusion chromatography. When these further steps were analysed **(Fig 3.5)** it was clear that YebT had been purified successfully, with a single intense band of the desired size present on an SDS-PAGE gel. While the SEC chromatograph suggested that a proportion of purified YebT was aggregating, a larger amount seemed to be soluble, with a retention volume in the SEC column of around 70 mL.

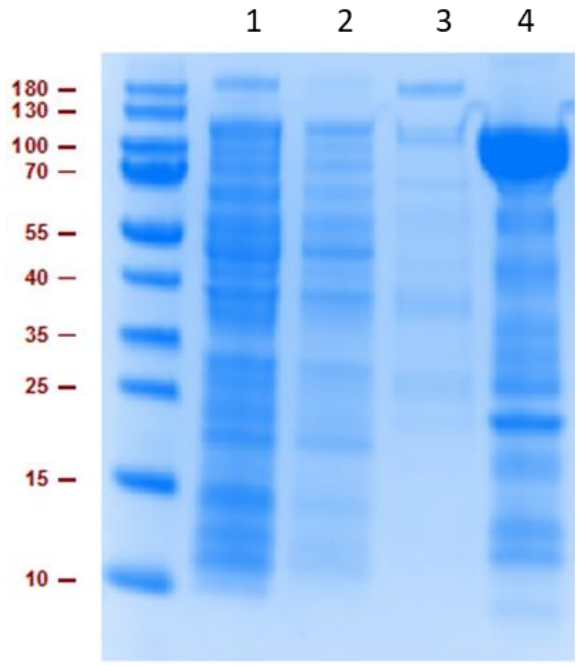


Figure 3.4 – SDS-PAGE gel of truncated YebT expression test, stained with Quick Coomassie. *E. coli SB21* (DE3) cells transformed with plasmid pHisTev:YebT were induced with 0.2mM IPTG to express protein overnight, with total volume 4 litres. Following cell lysis, the soluble fraction was used for Ni^{2+} IMAC. The column flowthrough (1), washes with binding buffer (2) and wash buffer (3) were collected, along with column elution (4). From this initial expression, it appeared that the vast majority of expressed YebT (around 95kDa) was successfully biding to the IMAC column. There was minimal YebT present in either the column loading flowthrough, either wash step, with a very intense band present in only the column elution fraction.

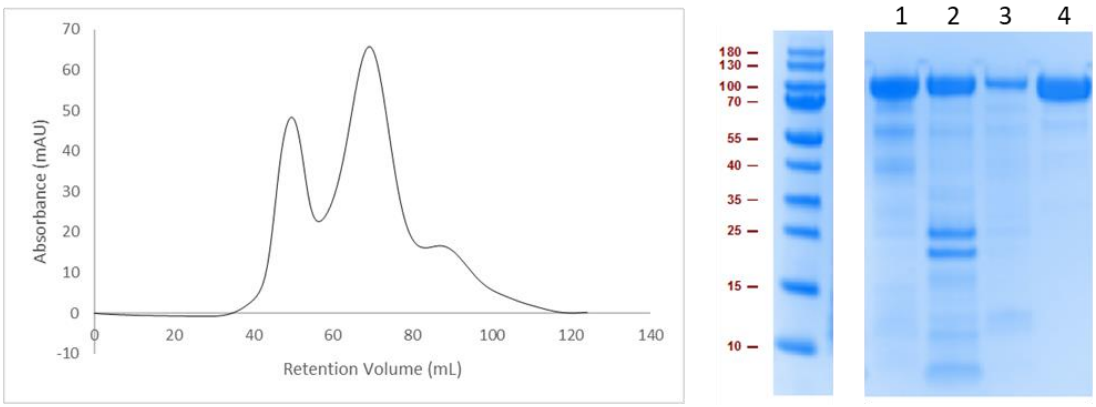


Figure 3.5 – SDS-PAGE gel and SEC chromatograph of further YebT purification following expression test. With a substantial amount of protein detected in the expression test, YebT was further purified for crystallisation. After tag cleavage with TEV proteinase, the protein was purified with an IMAC pulldown. Cleaved protein was purified further still, using size-exclusion chromatography (left). This yielded two distinct peaks, at volume, with a small sub-peak at 90 mL. The subsequent PAGE gel (right) revealed that the pulldown fraction (1) showed good purity, but the pulldown elution fraction (2) still contained good amounts of YebT at around 100kDa – while it had successfully removed any remaining TEV (bands around 25kDa), it seemed some YebT was still his-tagged. The 50 mL and 70 mL SEC peaks (3 and 4 respectively) both showed good purity of YebT, with the peak at lower retention volume representing aggregated YebT.

3.2.3 Initial YebT Crystallisation Trials

After its initial purification, sufficient truncated YebT was obtained for initial crystal trials, using the sitting drop vapour diffusion method. Adequate protein was available to 6 initial crystallisation plates at 5 mg/ml YebT, using commercial screens such as JCSG-Plus and PACT-Premier. Plates were made with drop sizes of 1 µl and 0.6 µl, in a 1:1 proportion of well solution to protein. After a week-long incubation, two conditions had promising initial crystal hits: PegRx H11 and PACT Premier D11 **(Fig 3.6)**. The crystals from PegRx took the form of flat, plate-like square crystals, with the PACT crystal appearing much more 3D, with flat, shiny faces. Crystals from either condition were soaked in cryoprotectant containing 20% glycerol, and 200 mM Sodium Iodide, for potential phasing through SAD if crystal quality was sufficient. These were then harvested in crystal-holding loops, flash frozen in liquid nitrogen and cryo-shipped to beamline i03 at Diamond Light Source.

Data was collected at 1.85 Å, a wavelength suitable for using anomalous scattering data from Iodine (122). For both crystal forms, 3600 images were collected with 0.1 second exposure times at 40% beam transmission. The resulting X-ray diffraction patterns **(Fig 3.7)** indicated that crystals from the PACT D11 condition were most likely salt, and not a worthwhile optimisation target. The PegRx crystal looked more promising, with many more spots visible. However, while some spots were seen at higher resolutions, the overall data quality and resolution was poor from initial crystals. Data integration of complete datasets from these crystals using Xia2, as well as automated processing steps within the data collection software all failed. To try and improve collected data, crystal quality needed improvement – further crystal trials and optimisation of the PegRx H11 condition was the next step.

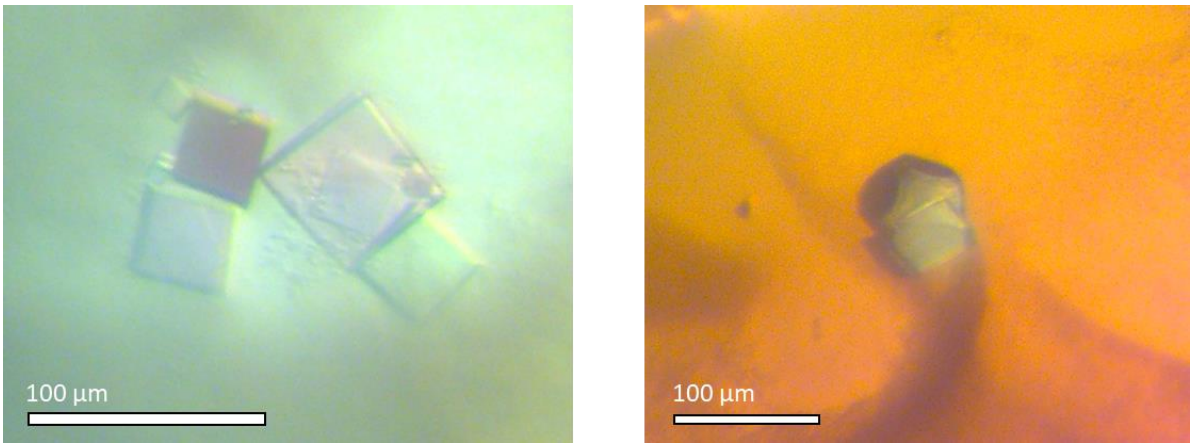


Figure 3. 6 - Crystals were formed from initial YebT crystallisation trial plates. PEGRx well H11 (Left) gave rise to flat, square crystals; conditions were: 0.2M Magnesium Chloride, 0.1 M Sodium Citrate Tribasic pH 5 10% PEG 20,000. PACT Premier D11 (Right) contained a pair of seemingly fused crystals, with a more three-dimensional structure; conditions were: 0.2M Calcium Chloride, 0.1 M Tris-HCl pH 8, 20% PEG 6,000.

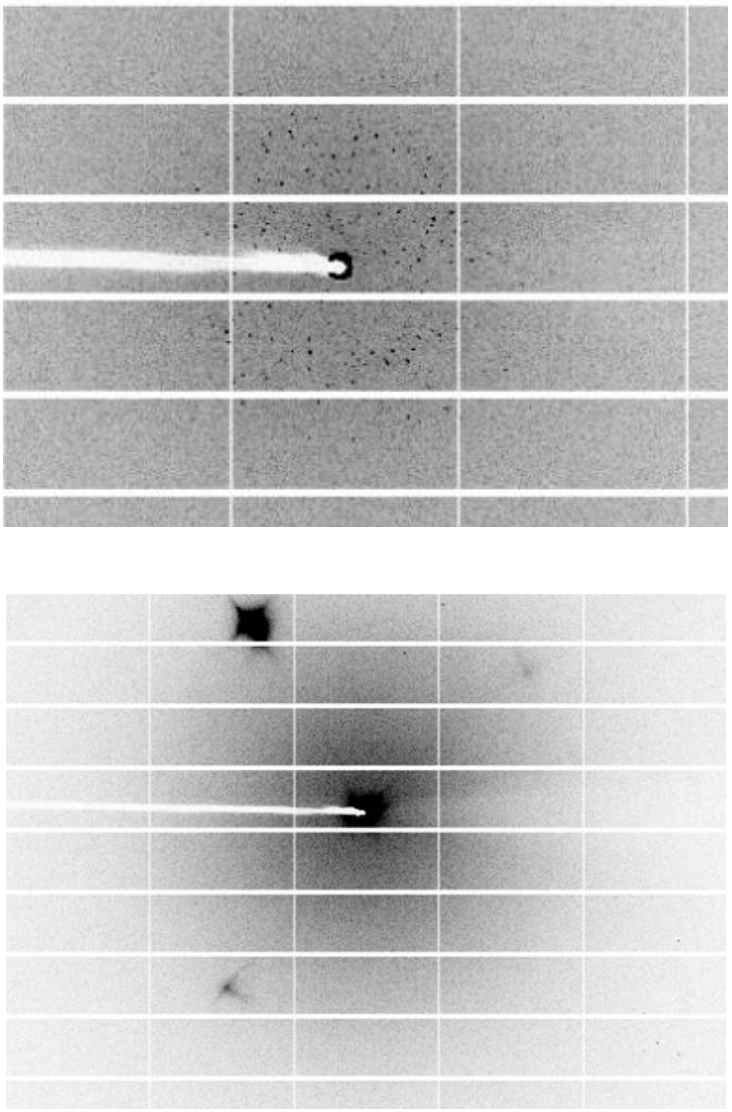


Figure 3.7 - X-ray Diffraction patterns from putative YebT crystals harvested from trial plates. The flat, thin crystals from PEGRx well H11 (Upper) resulted in many spots, with reducing intensities at higher resolutions; this crystal's diffraction pattern is consistent with a protein crystal. Data collected from the more three-dimensional PACT Premier D11 crystal (Lower) was less promising, with a very large, dark spot at around 3 Å resolution and few others; this pattern is more likely to arise from a salt crystal.

3.2.4 Optimisation of YebT Crystallisation

For use in optimising these initial hits, and for other further crystal trials, several larger-scale purifications of truncated YebT were carried out. Given the successes seen with the initial purification, the same basic methods and buffers were used, but with more stringent sample cooling and repeated sample loading steps with the IMAC column. In addition, each expression was from 12 litres of cells as opposed to the previous 4 litre volumes. Following the size-exclusion chromatography step of the first such purification, the pooled YebT sample was analysed with an SDS-PAGE gel **(Fig 3.8)**. This gel indicated that YebT was obtained as a very pure sample, with minimal contaminant proteins detected, even with a thick, concentrated target protein band. Data from the SEC step **(Fig 3.8)** showed a good increase in soluble protein yield, with the peak 70 mL retention increasing to 250 mAU from 70 mAU. In addition, the proportion of presumed protein aggregate at 50 mL was greatly reduced from the initial purification - the difference in peak size increasing to over 200 mAU from only 20 mAU. With far more soluble YebT generated, more sitting drop vapour diffusion crystallisation plates could be produced, with screening plates using protein concentrations of 10 mg/ml and 14 mg/ml both made. To optimise the previously tested protein crystals found in PegRx H11, a 5x5 optimisation screen was made by hand and also used to make crystallisation plates **(Table 3.3)**

After 2 months incubation at 21°C, few more hits were discovered on crystal plates. While crystals were present on the new optimisation plate around the Peg Rx condition, they were visually no different to those previously harvested. The PGA screen (not previously used due to a lack of available protein) was one screen plate to contain some new hits, with particular interest in well A4 **(Fig 3.9)**. This well

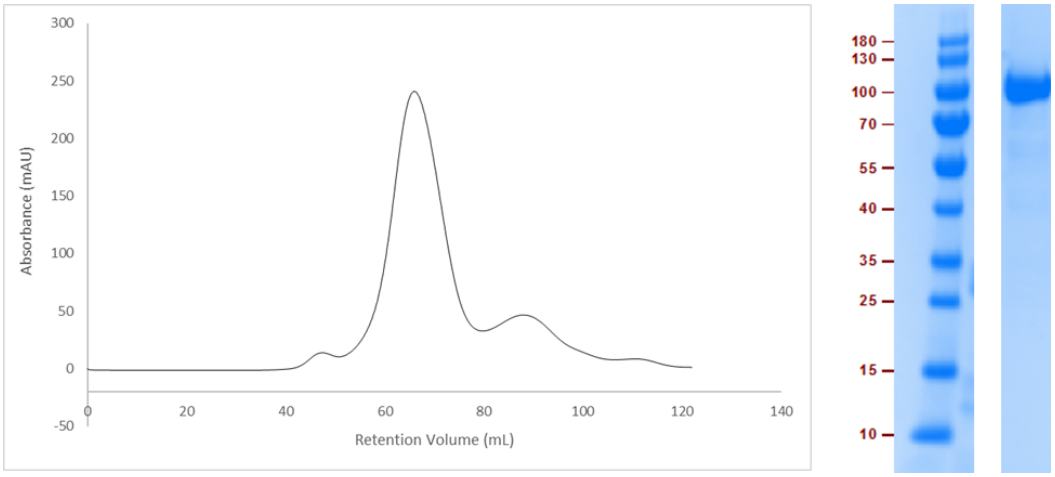


Figure 3.8 –SEC chromatograph and PAGE gel of large-scale YebT purification. YebT was expressed in *E. coli SB21* (DE3) cells using the plasmid pHisTev:YebT. A volume of 12 litres of cells was induced with 0.2mM IPTG to express protein overnight. Following cell lysis, protein was purified using IMAC with a nickel column, before tag cleavage using TEV proteinase. After an IMAC pulldown, protein was finally purified using SEC. The SEC chromatograph (left) contained a large peak at around 70 mL – this is the same as seen in the expression test, but in this instance, there appears to be a reduced aggregation peak, with only a small absorbance at 50 mL. A PAGE gel run prior to crystallisation trials (right), indicated that YebT had been purified with good purity, and there were no other identifiable proteins present.

Table 3.3 – YebT Optimisation screen based on PegRx H11. After yielding likely protein crystals, the PegRx H11 condition was used as an optimisation condition, with PEG concentration and pH varied as shown. Each well contained 0.2M Magnesium Chloride and 0.1M Sodium Citrate Tribasic.

		PEG 20,000 (%)				
		6	8	10	12	14
pH	4.6					
	4.8					
	5					
	5.2					
	5.4					

contained flat, plate-like crystals which were larger than those derived from PegRx. These, along with the some of the new RxH11 crystals were soaked in 200 mM Sodium Iodide cryoprotectants and harvested for data collection at DLS. As before, data was collected at 1.85 Å, with 3600 images, 0.1 second exposure times and 40% transmission. Data collected from Rx H11 optimisation crystals was largely the same as previously seen, with only a moderate number of spots detected, albeit with some at resolutions of around 3 Å. Data from PGA A4 crystals though **(Fig 3.9)**, appeared more promising, with a much greater number of spots seen closer to the beam stop. Auto processing at DLS using xia2 was successful, and suggested the dataset had a resolution of around 8 Å. While this data was insufficient to solve the YebT structure, it did point to improvements in crystal quality.

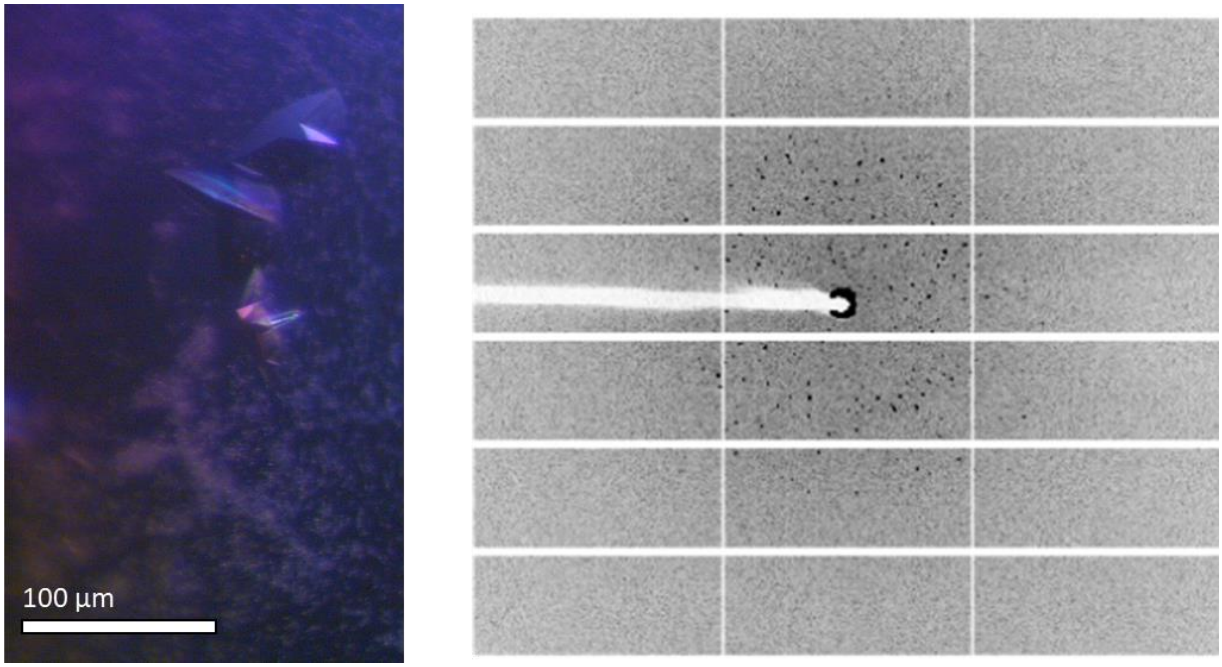


Figure 3.9 - A YebT crystal and its X-ray Diffraction pattern from further YebT crystals trials. Well A4 from the PGA screen made at 10 mg/ml YebT contained thin, plate like crystals (left). This condition consisted of: 0.6 M Sodium formate, 0.1 M Sodium acetate pH and 5.0 8 % w/v γ-PGA. Once submitted for data collection at Diamond Light Source, its X-ray diffraction pattern confirmed its protein nature. Auto processing suggested the resolution of the collected dataset was around 8 Å.

3.2.5 Purification and Crystallisation of YebT at reduced salt concentration

In a further attempt to enhance protein crystal growth, YebT was purified with an exchange into lower salt buffer during the SEC step. This consisted of a reduction in NaCl confrontation of purification buffer from 300 mM to 150 mM. Other than this methodological change, the same lysis and purification procedures were used as with higher salt YebT preparations. The SEC chromatograph for the NaCl reduction **(Fig 3.10)** appeared similar to that observed from previous purifications, with peaks at around 50 mL, 70 mL, and 90 mL retention volume. Notably, the 50 mL and 90 mL peaks, likely indicating aggregation or proteolytic cleavage respectively, appeared at an increased proportion of the larger 70 mL peak than previously demonstrated. A PAGE gel run of protein pooled from the 70 mL peak **(Fig 3.10)** indicated that YebT was successfully purified under these conditions, albeit with some observed impurities. This concentrated protein was used for crystallisation trials, in the hope of obtaining improved crystals with a reduction in NaCl concentration in the mother liquor. Alongside commercial screens, an optimisation plate using the previously more successful PGA A4 condition as a starting point was produced **(Table 3.4)**. Several other novel 5x5 screens were hand-made from conditions with minor crystal 'hits', such as Crystal Screen 1/2 B1, JCSG Plus D2 and PGA E10, with varying pH and precipitate concentrations.

After an incubation period of 8 weeks, new crystal growth was seen on newly made plates at reduced salt concentration **(Fig 3.11)**. Many wells contained numerous flat, plate crystals, with an optimisation screen around PGA E10 **(Table 3.5)** yielding a particularly large flat crystal. These crystals were harvested after soaking in

cryoprotectants with 200 mM NaI and sent for data collection. Data was collected at 1.85 Å, with 3600 images, 0.1 second exposure times and 40% transmission.

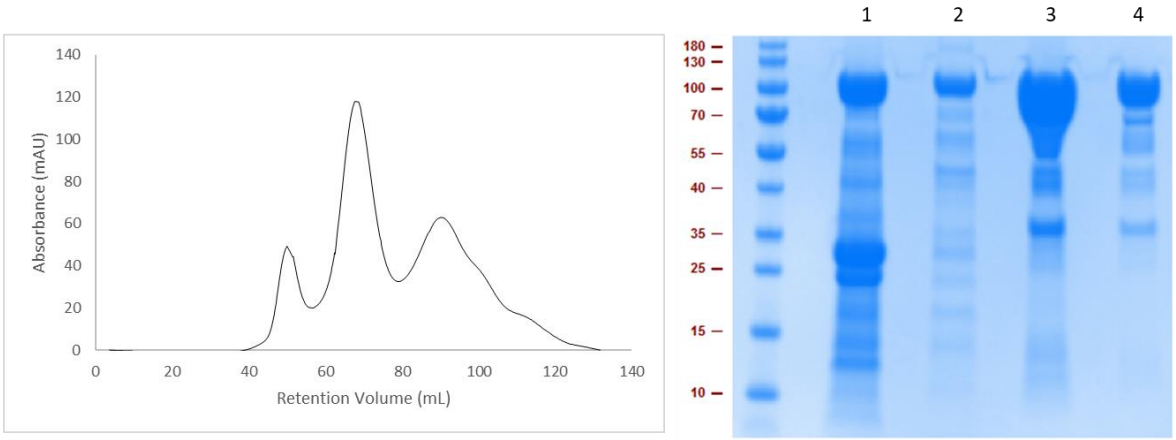


Figure 3.10 – SEC chromatograph and SDS-PAGE gel of YebT purification and exchange into lower-salt buffer. YebT was expressed in *E. coli SB21* (DE3) cells using a total volume of 12 litres. Protein expression was induced with 0.2mM IPTG overnight. Protein was purified using Nickel IMAC, and the hexa-HisTag was removed with TEV proteinase. Following a pulldown through the column, cleaved protein was purified for crystallisation using SEC. Compared to previous purifications, NaCl concentration was reduced from 300mM to 150mM during SEC. The Chromatograph from SEC (left) has the same three previously observed peaks at around 50 mL, 70 mL, and 90 mL. Noticeably, the 50 and 90 mL peaks were larger than previously seen, suggesting higher levels of aggregation (50 mL peak) and degradation (90mL peak) occurring. A PAGE gel (right) of samples taken from the pulldown column elution (1), pulldown wash (2) and from 70 mL retention volume peak (3, diluted in 4) showed that a significant amount of YebT was lost in the pulldown step, but a large amount of YebT was still obtained. While purity appears poorer than previously seen, this perceived drop in purity may be a result of too high a concentration of YebT sample being run on the gel.

Table 3.4 – YebT Optimisation screen based on PGA A4. After a promising data collection, the PGA screen A4 well solution was used as the basis for an optimisation plate, with PGA concentration and pH varied as shown. Each well contained 0.2M Sodium Formate and 0.1M Sodium Acetate.

		y-PGA (%)				
		4	6	8	10	12
pH	4.6					
	4.8					
	5.0					
	5.2					
	5.4					

Table 3.5 – Optimisation screen for YebT from the on PGA E10 Condition. To optimise crystal growth seen following lower-salt YebT screening, an optimisation plate was made which varied the two precipitants in the PGA E10 well solution. All wells contained 0.1 M Ammonium sulfate, 0.3 M Sodium formate and 0.1 M Sodium cacodylate; PEG500 MME was at 20% in the first 6 columns and y-PGA was at 3% in the second 6 columns.

		y-PGA (%)						PEG500 MME (%)					
		0	1	2	3	4	5	0	5	10	15	20	25
pH	6.1												
	6.2												
	6.3												
	6.4												

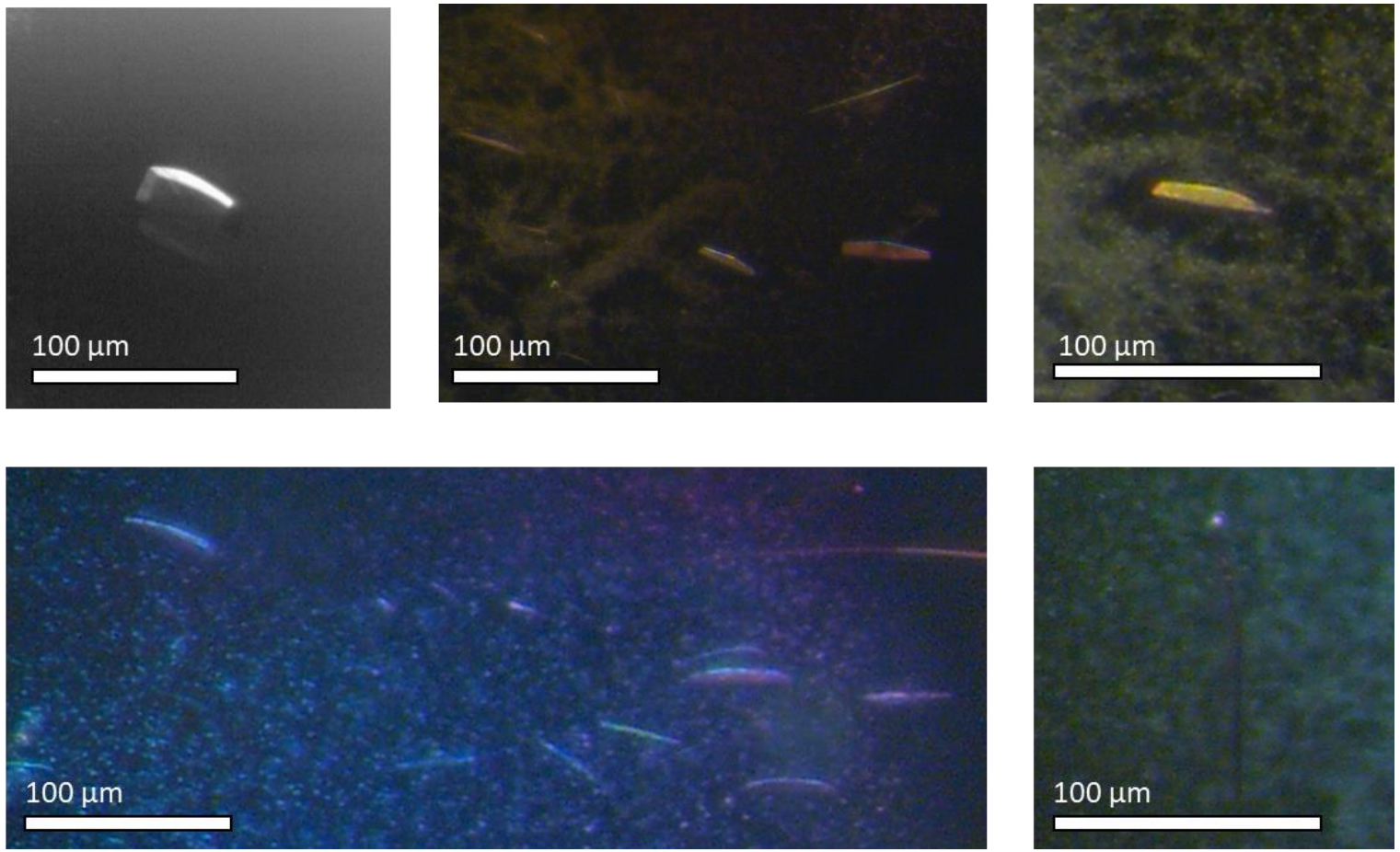


Figure 3.11 - Crystals of YebT formed following screening after a reduction in salt concentration. Several wells showed hits, including a larger, thicker crystal formed in well H9 of the LMB screen (Top left). Most crystals seen were in the form on flat, plate-like crystals, including those seen on optimisation plates from the PGA E10 condition (Upper Centre, Upper Right). Similar crystals were seen in the JCSG Plus D2 optimisation plate (lower left), with a large, single plate observed in the optimisation plate for PGA condition E10 (Lower right).

Unfortunately, data taken at 1.85 Å was poor, with few spots recorded. A test data collection at 0.9795 Å using crystals from the optimised PGA E10 plate yielded data capable of successful integration with xia2 to a resolution of 3.2 Å. Similar data was obtained from the larger, thicker crystal from the LMB H9 condition. While this data could not be used to solve the structure of YebT without any phasing information, it indicated that a solution could be possible with current YebT crystals or their data, with an appropriate solution to the phase problem.

3.2.6 YebT Crystallisation Trials with Sodium Iodide supplementation

While some YebT crystals capable of successfully diffracting X-rays had been produced, gathering phase information from these crystals when soaked in Iodine-containing cryoprotectants proved ineffective. An alternative method would be to grow protein crystals in the presence of Iodine atoms, in the hope that they would become incorporated into the crystals during their growth. For use with this method, YebT was purified using the same methods which had so far proved effective. During the size-exclusion chromatography step, the protein was exchanged from buffer containing 300mM NaCl, to a buffer containing 150mM NaI instead. The chromatograph from this purification step **(Fig 3.12)** suggested that protein aggregation may be increased from this change in buffer conditions, with the putative soluble YebT peak at around 70 mL around half the size of a peak at 50 mL retention volume. Despite this, PAGE analysis of the SEC peaks **(Fig 3.12)** showed a single band present at around 100kDa in either case, pointing to a successful purification of YebT without significant degradation.

From here, the protein was concentrated as before, and used in crystallisation trials with the sitting-drop vapour diffusion method. Plates were made using the

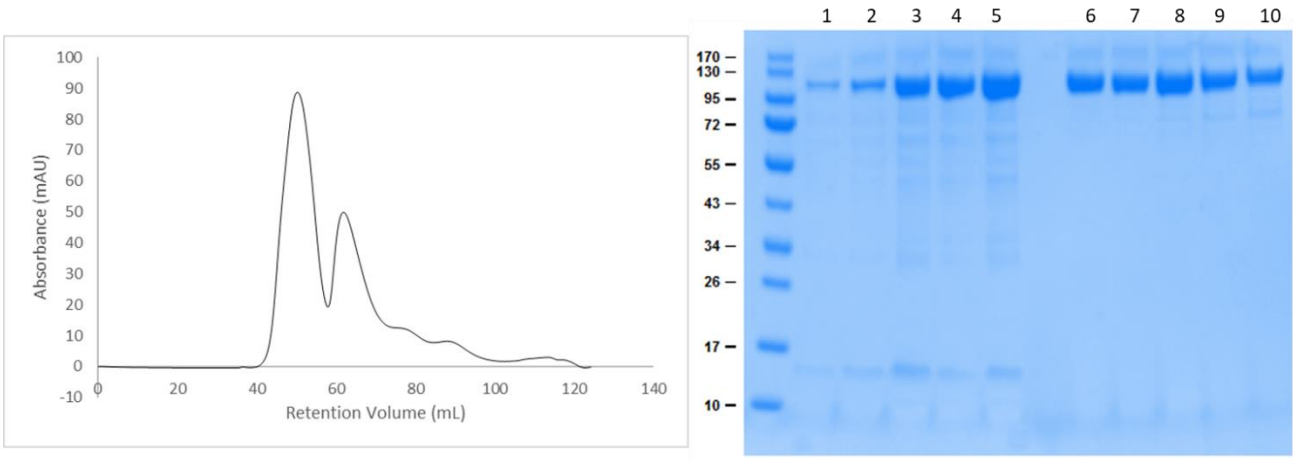


Figure 3.12 – SEC chromatograph and SDS-PAGE gel of YebT purification and its exchange into iodide – containing buffer. Truncated YebT was expressed in SB21 (DE3) *E. coli* cells and initially purified in the same manner as before using IMAC and a TEV cleavage. At the final purification step, instead of 300 mM NaCl-containing buffer, the YebT sample was exchanged to buffer containing 150mM NaI. The SEC chromatograph (left) suggests that aggregation might have been an issue, with a larger peak visible at around 50 mL retention volume as opposed to the smaller peak at around 70 mL that has become characteristic in YebT purification. A PAGE gel run from samples using the peaks (right) indicated that both peaks were indeed YebT (at around 100kDa). The first peak (1-5) did show a minor band at around 15kDa, but the second SEC peak (6-10) appeared to have negligible co-purifying proteins. Protein collected from the second peak was concentrated and used for crystallisation trials.

optimisation conditions based on PGA E10 and A4, as these conditions had led to crystals with promising diffraction capability. After 3 months of growth, some large, flat crystals similar in morphology to those previously observed were found in the PGA E10 optimisation screen **(Fig 3.13)**. These were harvested and submitted for data collection; data was again collected at 1.85 Å, 3600 images and 0.1s transmission. The data collected from these crystals was very poor, with fewer than 20 spots detected in the majority of images taken.

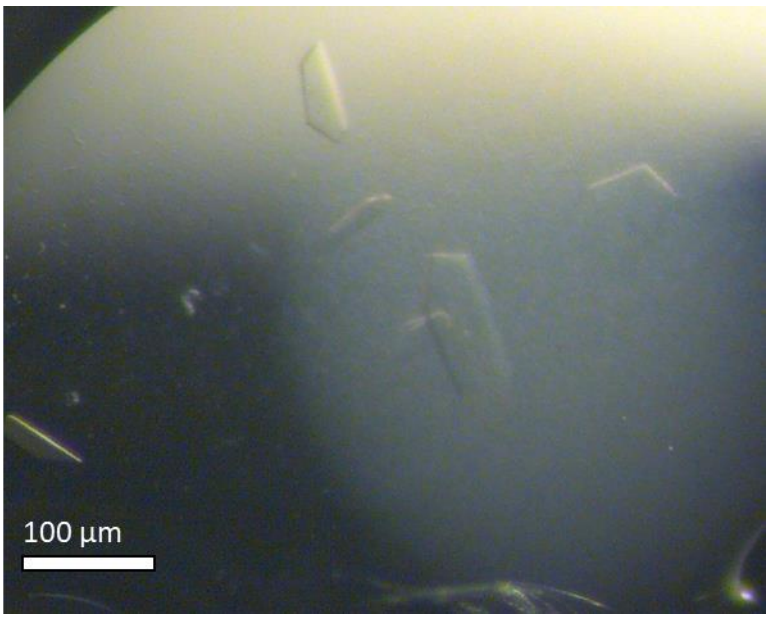


Figure 3.13 - Crystals of YebT formed in Sodium Iodide-containing purification buffers. These flat, plate like crystals were very similar in appearance to previously seen YebT crystals, and were grown on a plate optimising the PGA E10 condition.

3.2.7 Crystallisation of Selenomethionine-Labelled YebT

With iodide soaking and co-crystallisation both proving unfruitful, the next phasing method attempted was the incorporation of Selenomethionine (SeMet) within YebT during its expression from *E. coli* cells. The truncated form of YebT contains 12 methionine residues and is thus a good candidate for this technique with 12 potential selenium sites. Protein was expressed in the same SB21 *E. coli* cells as used for non-selenium protein expression. Cells to a volume of 8 litres were grown in M9 minimal media with SeMet nutrition mix, to allow selenium incorporation.

After an overnight expression, cells were harvested via centrifugation, before a cells lysis and protein purification process using IMAC and SEC, with the same conditions as for the native protein. The results of the size exclusion chromatography run **(Fig 3.14)** indicated potential issues with protein aggregation within the SeMet prep, with a single clearly defined peak at around 50 mL retention. A PAGE gel run to analyse the protein contents of this peak **(Fig 3.14)** indicated that YebT had successfully been purified, with single, thick bands visible at around 100kDa in samples taken from the 50 mL SEC peak. Despite potential aggregation owing to the low retention volume observed, protein was concentrated to 10 mg/ml and used to make sitting drop vapour diffusion crystallisation plates from the two broad matrix screens, as well as optimisation condition to match previous YebT crystal formation in PGA E10 and A4 conditions. After 3 months growth, some small crystals were formed from Selenomethionine-incorporated protein samples **(Fig 3.15)**. These were visually distinct from previously seen crystals, and appeared to be smaller, thicker formations compared to the thin sheets of crystal previously seen. There were 32 crystals harvested for X-ray diffraction; data was collected at 0.9750 Å, with 3600

images at 0.05s exposure. Unfortunately, none of the harvested crystals gave rise to data suitable for the phasing of YebT data.

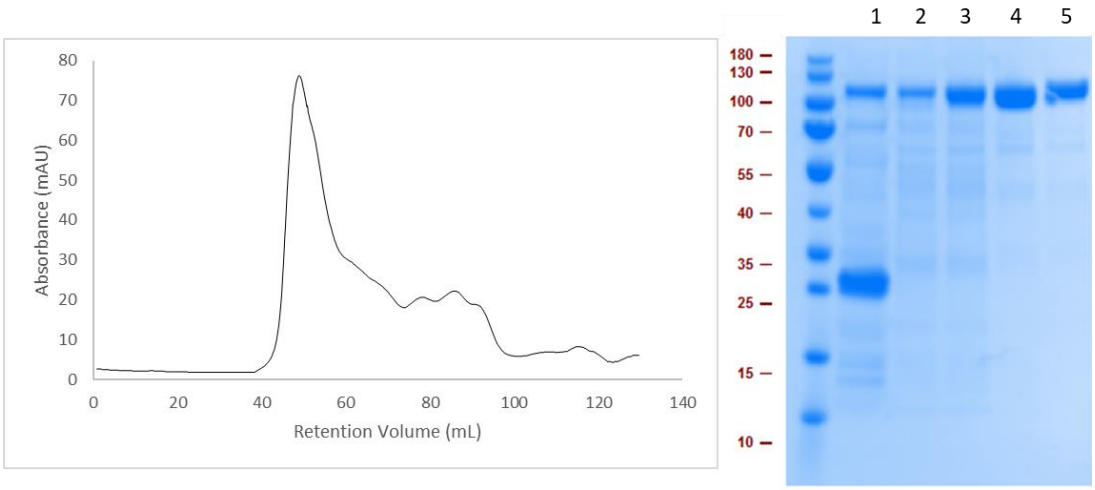


Figure 3.14 – Purification of Selenomethionine-incorporated YebT. YebT was expressed in SB21 (DE3) *E. coli* cells grown in M9 minimal media supplemented with SeMet nutrition mix and 60 mg/L Selenomethionine in a total volume of 8 litres. YebT was purified in the same manner as non-selenium tagged protein. The chromatograph from size exclusion chromatography (left) had its largest peak at around 50 mL retention volume, with a series of small, poorly defined peaks between 65 mL and 100 mL. In previous purifications, 50 mL appeared to be the retention volume corresponding to aggregated protein, although the incorporation of selenium may have effected retention volume. A PAGE gel was run from the samples taken from the purification points (right). The elution from the IMAC pulldown (1) indicated that some YebT was retained in the column, but a large amount of probable TEV proteinase was captured at 27 kDa. With only a single clearly defined peak, four samples (2-5) were taken from increasing retention volumes of the peak at 50 mL. Samples from the 50 mL peak showed prominent bands at around 100 kDa with minimal impurities. Protein from this peak was pooled and concentrated for Se-Met crystal trials.

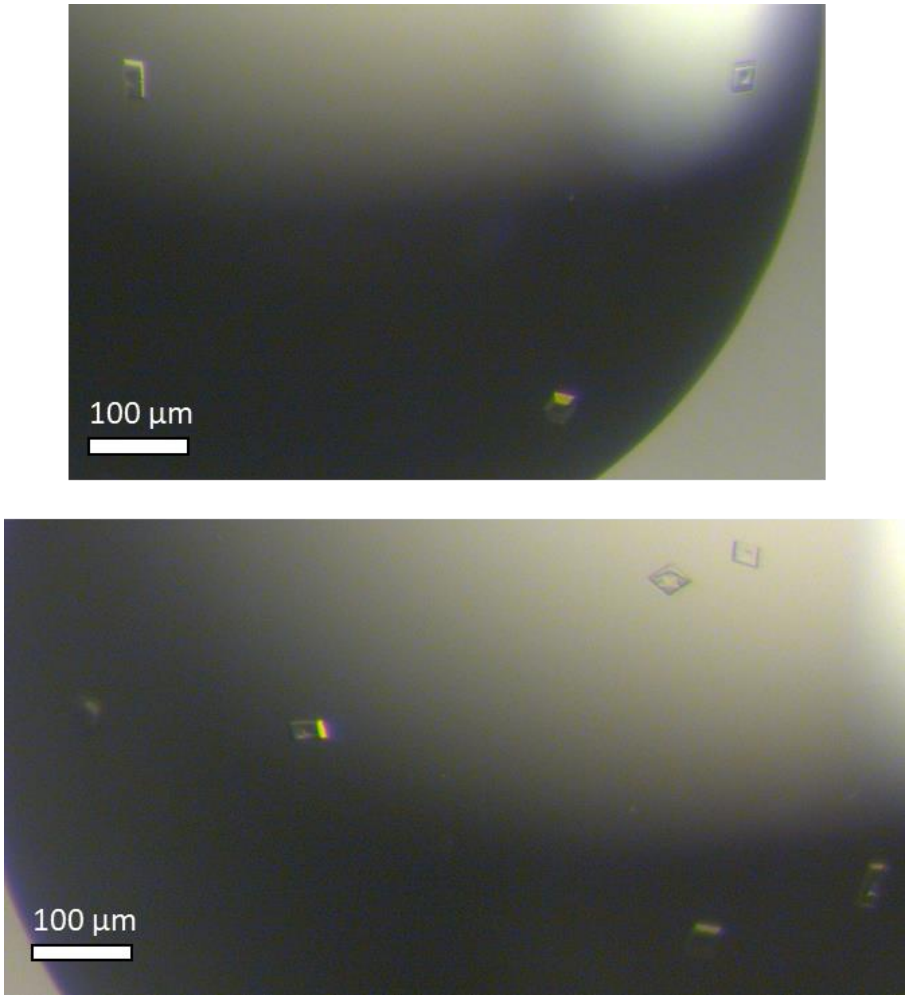


Figure 3.15 - Crystals of Selenomethionine-incorporated YebT. Crystals were formed in optimisation conditions for PGA E10, containing 0.1M Sodium Cacodylate pH 6.5-6.8, 3% PGA, 20% PEG 500MME, 0.3M Sodium Formate and 125-150mM Ammonium Sulfate. These crystals were morphologically different to other YebT crystals previously generated and were much smaller.

3.2.8 Native Data Collection and Attempted Molecular Replacement of YebT

During data collection for Se-met crystals, some remaining flat plate crystals originally used for attempted iodine-soaked phasing **(Fig 3.11)** were used for X-ray diffraction. As opposed to collecting data at the absorption edge for iodide, data was instead collected at 0.9750 Å, with 3600 images at 0.01s exposure, 100% beam transmission. Of these crystals, one seemed to diffract well, with data processed by xia2 at DLS to an estimated resolution of 2.02 Å **(Table 3.6)**.This integrated data was reduced with Aimless and used for a molecular replacement using MOLREP as using the previously outlined MR methods for PqiC. The search model chosen was a MCE domain from PqiB (PDB 5UVN), which has been shown to from hexameric stacked rings in the same manner as YebT (60). Unfortunately, this molecular replacement was unsuccessful, and a model of YebT from this X-ray diffraction data was not successfully built.

Subsequently, the low-resolution negative stain EM map of YebT (EMD-8611) at a resolution of around 25 Å (60) was used as a search model for MR. This map served as initial model for molecular replacement through PHASER-MR (80) with the Phenix GUI (123), using methods specifically outlined in (124). This method again proved ineffective, with no suitable model built.

Table 3.6 - Data collection statistics for native YebT crystal for data processed at DLS using xia2. Data was collected from remaining crystals prepared for iodide soaking at 0.9750 Å for use in potential molecular replacement structure solution.

Data Collection	
Beamline	i03
Wavelength (Å)	0.9750
Spacegroup	P 1 21 1
Unit Cell Dimensions	
a, b, c (Å)	45.16, 214.67, 55.86
α, β, γ (°)	90.00, 95.83, 90.00
Resolution (Å)	107.34 - 2.02 (2.05-2.02)
R_{meas}	0.117 (1.499)
$CC_{1/2}$ (%)	0.983 (0.952)
Completeness (%)	99.2 (96.6)
Mean I/σ (I)	7.9 (1.1)
Multiplicity	6.6 (4.7)

3.2.9 *In situ* limited proteolysis of YebT

With crystallisation and subsequent structure solution of intact YebT proving difficult, protease tests were conducted to determine candidate enzymes for use with *in situ* limited proteolysis crystallisation. Proteases were used at a ratio of 1:100 with YebT and incubated at room temperature for 90 minutes. After digestion, YebT samples were run on an SDS-PAGE gel to assess which proteases would be suitable for further use **(Fig 3.16)**. Several of the proteases caused significant protein degradation, without near-to full length protein bands remaining – these proteases were deemed unsuitable for crystallisation. Trypsin and V8 protease appeared promising, with bands present in both instances at around 70kDa indicating that

small portions of the protein were cleaved, and these two enzymes would be appropriate for crystal trials. YebT at 5 mg/ml was mixed 1:100 with Trypsin or V8 proteinase, and used to make plates of PACT Premier, JCSG Plus and the PGA screen, given previous crystal hits obtained with the screen. Plates were made using 1 µl and 0.6 µl drop sizes, and incubated at 21°C. Sadly, no crystal formation was seen on any of the YebT plates made using the *in situ* limited proteolysis technique.

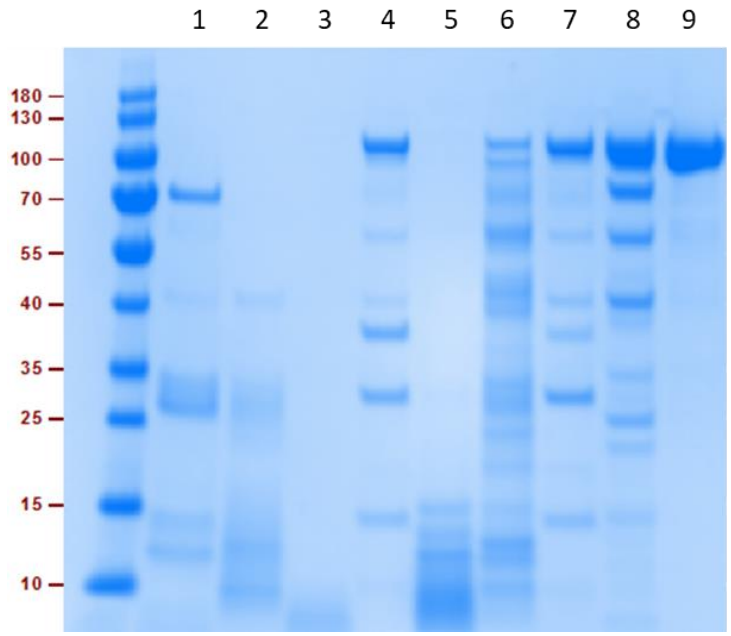


Figure 3.16 – PAGE gel of Limited proteolysis of truncated YebT. Purified YebT was treated with 8 separate proteases, each at a ratio of 1:100 protease: protein for 90 minutes. The enzymes used were: Trypsin (1), Chymotrypsin (2), Subtilisin (3), Thermolysin (4), Papain (5), Elastase (6), Dispase (7) and V8 protease (8); an aliquot of purified YebT was also run as a control (9). Chymotrypsin, Subtilisin, Elastase and Papain all appeared to completely degrade the protein and were not considered viable for crystallisation trials. Trypsin and V8 protease were selected for *in situ* limited proteolysis crystal trials, owing to the large size of the observed protein fragments. Each protease resulted in a cleavage product of around 70kDa, suggesting a large proportion of YebT was still intact and available for crystallisation.

3.2.10 Cloning of *YebT* from *Klebsiella pneumoniae*

In an effort to produce crystals capable of better diffracting X-rays and with better uptake of phasing agents such as Iodine atoms, YebT from *Klebsiella pneumoniae* was cloned for expression in *E. coli* cells. The YebT ortholog from *K. pneumoniae* was used for purification and crystal trials given previous work showing the increased likelihood of growing protein crystals by screening homologous proteins from different sources (125, 126).The first step in the cloning process was to run a PCR reaction to amplify the *YebT* gene from *K. pneumoniae* genomic DNA. This PCR reaction was run on an agarose gel **(Fig3.17)** which showed bands at the expected size of around 2500bp for truncated *YebT.* The gene and expression plasmid pHisTev were both restricted using the enzymes *NcoI* and *BamHI* and the two genetic components joined together with DNA ligase. Following amplification of the newly transformed plasmid in a high-copy number *E. coli* strain, said plasmid was double restricted with *NcoI* and *BamHI* to determine the presence of the gene insert. This reaction was again run on an agarose gel for analysis **(Fig3.18)** which showed a successful integration of truncated *YebT* from *K. pneumoniae* into the expression plasmid.

YebT from *K. pneumoniae* is very similar to that found in *E. coli*, with 81% of the proteins' amino acids being identical – this rises an identity 92% if taking similar amino acids into account , when aligned using BLAST (127). The similarity between the YebT proteins from the two species suggested that they would likely be stable under the same buffer conditions. Moreover, should the structure of *K. pneumoniae* YebT be elucidated, it would serve as a suitable molecular replacement map to allow for the phasing of existing *E. coli* diffraction data at the native wavelength.

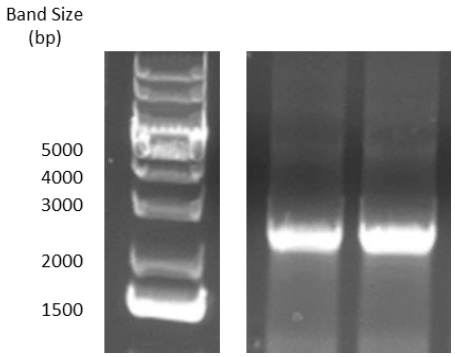


Figure 3.17 –Agarose gel of *Klebsiella pneumoniae YebT* PCR product. The gel contained EtBr and was imaged under UV illumination. In both repeats, a single band corresponding to *K. pneumoniae YebT* with its N-terminal truncation was seen at around 2500bp, as expected.

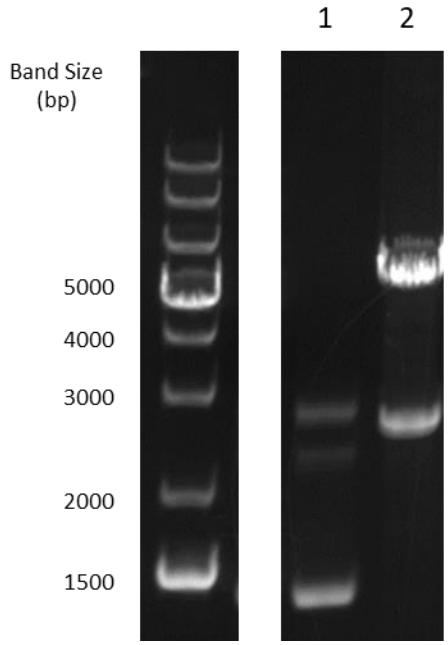


Figure 3.18 – Confirmation of *Klebsiella pneumoniae YebT* ligation into an expression plasmid using a restriction digestion, run on an Agarose gel. The *YebT* gene was ligated into the pHisTev plasmid. The putative expression construct was restricted with *NcoI & BamHI* to test for gene insert. Of the two samples tested, lane 2 showed success, with a strong band at over 5000bp indicating the presence of the pHisTev plasmid, and the band at around 2500bp indicative of *YebT* insertion.

3.2.11 Expression of *Klebsiella pneumoniae* YebT

Given the noted similarities, the same experimental parameters were used for the *K. pneumoniae* YebT expression and purification as were used for its *E. coli* counterpart. The newly constructed plasmid was transformed into SB21 (DE3) *E.* coli cells. Using 12 litres of these transformed, YebT production was induced using IPTG, and followed by an overnight protein expression. Following cell lysis, protein was purified using IMAC, followed by size-exclusion chromatography. PAGE was used to analyse the outcome of the purification **(Fig 3.19)**. From this analysis, the purification appeared to have been unsuccessful – no bands were visible around the expected 95kDa range, with only a diffuse band seen following SEC at 75kDa. The lack of full-length protein - coupled with a SEC peak at 50 mL suggesting aggregation – pointed to different expression or purification conditions being a necessity to obtain a pure sample of *K. pneumoniae* YebT for crystallisation trials.

To increase the probability of successful purification and subsequent crystal growth of YebT, further screening could be undertaken including a much larger number of orthologs from different species.

3.2.12 Expression and Crystallisation Trials of YebS

To test its suitability for crystallisation and subsequent structure solution, an initial expression test for YebS was carried out, using the pHisTev_*YebS* construct previously produced. The plasmid was transformed into *E. coli* C43 (DE3) cells using heat shock, with protein expressed overnight in 6 litres of cell culture. Cells were lysed and debris removed in the same manner as for YebT purification, with a subsequent spin at 100,000 g to collect cell membranes. YebT-containing

membranes were solubilised in 1% DDM for further purification with IMAC and SEC, again carried out in a manner in line with YebT purifications, except for the inclusion of DDM throughout the process. The purification's success was analysed from its SEC chromatograph, as well as a PAGE gel with samples taken from its peaks **(Fig 3.20)**. The chromatograph contained two predominant peaks at 45 mL and 60 mL, of 80 and 60 mAU in size, respectively. On the PAGE gel, samples taken from across either of these peaks contained a main band between 35 and 40 kDa. While YebS is expected to be around 48 kDa, this band probably corresponds to the protein, given the effects of 'gel shift' on membrane proteins during electrophoretic analyses. With the 45 mL peak likely formed of protein aggregate, fractions from the second peak were collected and used for initial crystal screening at 10 mg/ml. A number of commercial membrane protein screens were used; these included Mem Sys, Mem Start and Mem Gold 1/2. As for YebT, drop sizes of 1 µL and 0.6 µL were used, with incubation at 21°C.

After 2 weeks' incubation, no prospective crystal hits were seen across the screen plates made. To try and improve chances of YebS crystal formation, another purification was carried out, using a different detergent – n-Dodecyl-N, N-Dimethylamine-N-Oxide (LDAO). LDAO was selected, as its smaller micelles may allow for better packing of proteins in solution and therefore greater potential for crystallisation. Protein purification proceeded in the same way, except for the change from DDM to LDAO in buffers used for IMAC and SEC. However, the SEC chromatograph **(Fig 3.21)** indicated that this change had caused a much greater amount of protein aggregation the previously seen, with only a peak at 45 mL present of around 140 mAU – equal to a combination of the 45 and 60 mL peaks

yielded from DDM purification. Given the greater aggregation issues compared to the previous purification, this protein was not used for subsequent crystal trials.

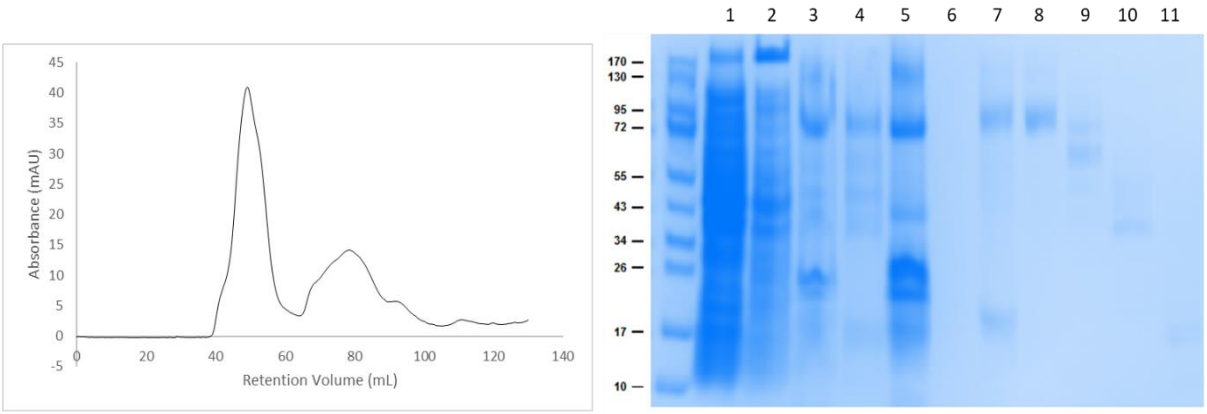


Figure 3.19 – SEC chromatograph and SDS-PAGE gel of Klebsiella *pneumoniae* YebT purification. YebT was expressed in SB21 (DE3) *E. coli* cells using the pHisTev expression plasmid under the same conditions used for *E. coli* – derived protein, in a total volume of 12 litres. YebT was purified in using the same methods and buffers as in previous experiments. The SEC chromatograph from the final purification step (left) showed a sharp peak at 50 mL retention volume, and a broad combination of peaks between 60 and 100 mL retention. Notably, all peaks observed were below 40 mAU, indicating a very poor yield when compared to a peak of almost 250 mAU seen in 12 litre expressions of *E. coli* protein. After the purification was complete, a PAGE gel was run (right) to determine the presence of YebT in each step. In the Lysate flowthrough (1) and IMAC column wash (2), there was no identifiable band likely to correspond to YebT. In the IMAC elution (3), pulled-down sample (4) and pulldown column elution (5), a smeared band at around 75 kDa is apparent, potentially indicating a degraded form of YebT is present; it is notable that at least some of this was retained in the pulldown column, though. Samples from the first peak (6,7,8) also showed this 75 kDa band, again in a faint and diffuse manner. The wider peak (9,10,11) contained minimal detectable protein.

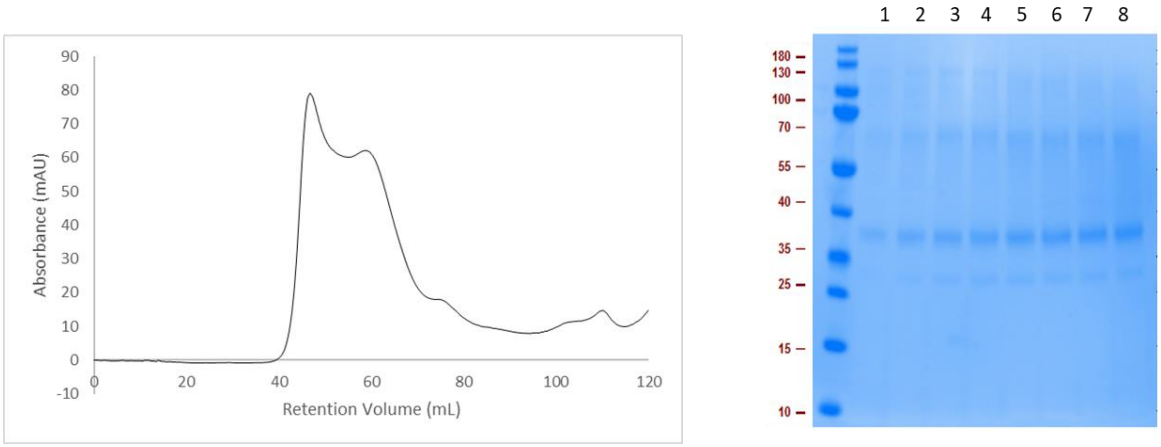


Figure 3.20 – SEC chromatograph and SDS-PAGE gel of initial *E. coli* YebS purification. Protein was expressed in C43 (DE3) *E. coli* cells using the pHisTev expression vector in a total volume of 12 litres. YebS was purified using IMAC and SEC, with protein solubilised in 1% DDM. The SEC chromatograph (left) showed two distinct peaks at 45 and 60 mL respectively. A PAGE gel (right) of the first (1-4) and second (5-8) peaks showed a similar profile across all samples. The predominant band was between 35 and 40 kDa; while this isn't at the 48 kDa of YebS, the phenomenon of gel shift in membrane proteins raises the likelihood that this indeed is the target protein.

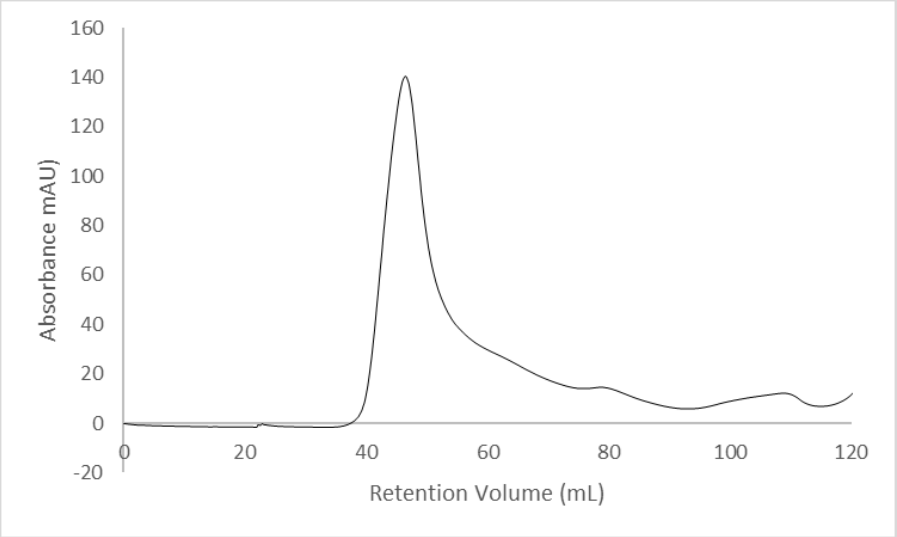


Figure 3.21 – SEC chromatograph of YebS from purification with 0.05% LDAO. To try to increase the proportion of YebS solubilised during the purification process, the detergent LDAO was used in place of DDM for IMAC and SEC steps. However, results from the SEC purification step indicated that this change had in fact increased aggregation levels, with no peak present at 60 mL and the majority of protein eluting from the column around the void volume of 40 mL. The 140 mAU peak corresponds to both main peaks combined from the DDM purification **(Fig 3.20)**, suggesting that overall YebS yield was unchanged.

3. 3 – Discussion

The outer membrane of Gram-negative bacteria is important for the protection of bacterial cells against harmful compounds, and is a significant barrier to antimicrobial agents (7). The outer portion of the outer membrane (OM) is comprised predominantly of Lipopolysaccharides (LPS) (26). The synthesis, role and transport of LPS is well described, with the Lpt complex of proteins responsible for its translocation across the membrane (38). This machinery has been shown to span the transmembrane space – the periplasm- to shield hydrophobic LPS molecules from their surroundings and deliver them to the OM (128).

The inner portion of the OM more closely resembles the inner membrane, and is composed of the Phospholipids Phosphatidylethanolamine (PE), Phosphatidylglycerol (PG) and Cardiolipin (CA) in *E. coli* (129).These phospholipids are of clear structural importance to the membrane, but also play important roles in the function of the proteins found there (50). It is interesting then, that given the significance PLs play to the OM, more is not known about their travel between the IM and the OM. The only pathway by which this occurs which has been characterised is the Maintenance of Lipid Asymmetry (Mla) pathway (56). This pathway is primarily thought to move PLs in the retrograde direction, but some evidence has pointed to an accompanying anterograde action (63). Aside from this pathway, the Tol-Pal, PqiABC and YebST pathways have been put forward as potential phospholipid transporters in Gram-negative bacteria (42). PqiABC and YebST were both initially suggested as PL transporters due to the possession of mammalian cell entry (MCE) domains by PqiB and YebT (68). These known-lipid-binding domains have been shown to form a hexameric ring-like structure in Mla pathway member MlaD (60, 62).

Cryo-EM studies previous to the work here have elucidated intriguing general structures for PqiB and YebT **(Fig 1.9)**. PqiB was shown to form a needle and syringe-like structure, with YebT appearing to form a long tube; both of these structures appeared to be long enough to span the whole intermembrane space (60). These large hexameric structures (like MlaD) were formed of rings from MCE domains. These general structures were solved with Cryo-EM at ~25 Å resolution, with the same data able to resolve the ring formation in PqiB to ~4 Å (60).

To better understand the structure of YebT and how this could relate to potential PL binding and transport, the work presented here aimed to use the technique of X-ray crystallography to solve this structure at higher resolution. This would allow the identification of potentially functionally significant residues and help begin to determine the involvement of YebT in PL transport. The preparation of YebT at suitable purity and concentration for crystallisation studies proved a success, and protein crystals *were* generated as part of the investigations.

Whilst crystals were produced which were apparently protein and likely to be YebT, the phasing of X-ray diffraction data collected from these crystals proved very challenging. Despite attempts using iodide soaking, iodide co-crystallisation and Selenomethionine incorporation into YebT, a higher-resolution structure for YebT remained elusive. Similarly, molecular replacement using a dataset collected at 0.9750 Å was not successful in solving the structure of YebT, using either the low-resolution YebT EM map or the higher resolution MCE domains of PqiB as search models.

While the work displayed here showed successful crystallisation of the periplasmic YebT architecture, the structure of the entire tube-like transmembrane domain was

recently solved at 3.46 Å resolution using Cryo-EM **(Fig 3.22)** (130). Incidentally, another group also solved the structure of YebT using Cryo-EM at 3.10 Å resolution, in separate structures encompassing MCE 1-4 and 5-7 (131). This two part-structure of YebT included a potential lipid binding site between MCE domains 5 and 6, although the authors admit it is a potential artefact of Cryo-EM averaging, and may be more due to this narrow portion of the tube taking longer for putatively transported PLs to traverse (131). Both novel structures suggested an element of flexibility within the YebT 'tube'; the two-part structure proposed two distinct structural conformations: one straight and one 'bent' form, with the MCE 5-7 region being displaced by around 9° from the vertical axis (131). The one-part structure suggested that flexibility was confined to the inner face of the YebT tube. The inner faces of MCEs 1,5,6 and 7 (named as pore-lining loops) appeared to exist in open or closed states **(Fig 3.22)**, with the potential to permit transit of cargoes of differing sizes (130).

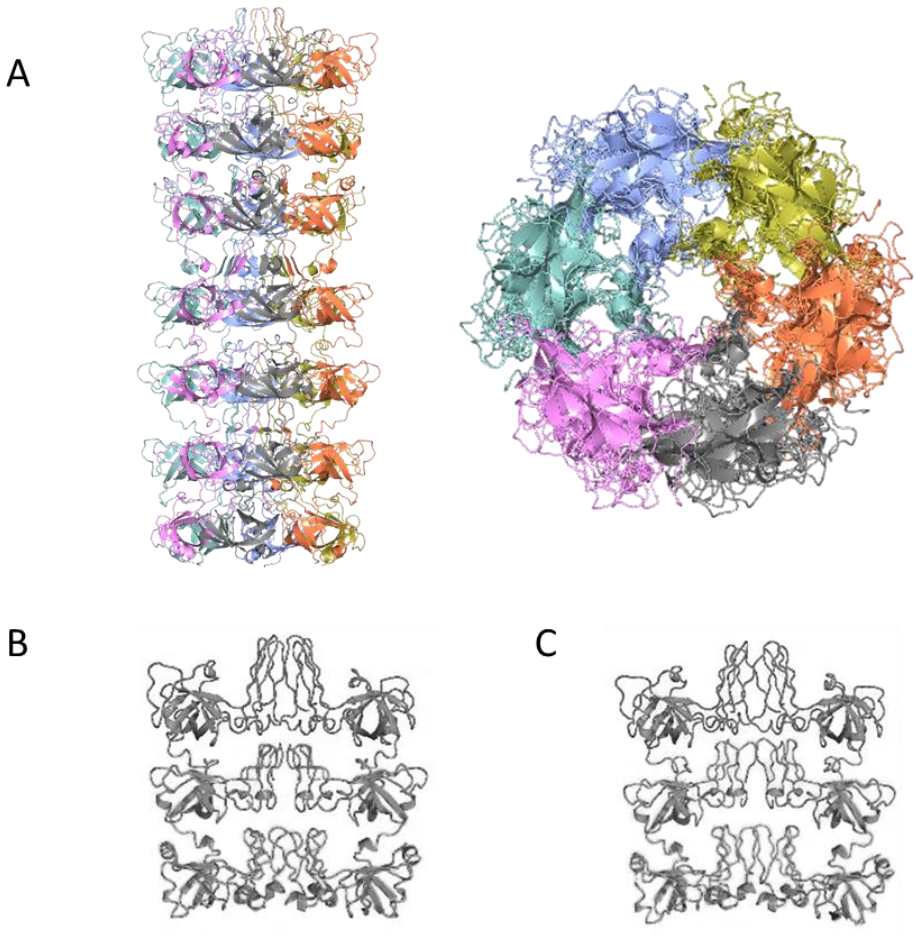


Figure 3.22 - YebT periplasmic structure solved using Cryo-EM to a resolution of 3.46 Å. While the phasing of X-ray diffraction data from YebT crystals was unsuccessful, the structure of YebT was solved without the need for phasing agents with the use of Cryo-EM by Isom *et al.* (130). The protein is formed of 6 YebT monomers (one subunit per colour), arranged into rings through their interacting MCE domains (A). The domains within the tube (termed pore-lining loops) were shown to exist in both closed (B) and open conformations (C). PDB 6V0C shown from (130).

Cryo-EM has the advantage of requiring significantly less protein at lower concentrations than X-ray crystallography (132). Moreover, the need to crystallise the protein is removed entirely, which – as outlined by the work here – can be a very lengthy and labour-intensive process. Without the need for crystals, the solution of the phase problem which escaped our grasp is equally redundant. Crystal formation can often prove particularly difficult for proteins with significant flexible regions (132), and the observed flexibility of YebT in (131) could go some way as to explaining why structure determination using crystallography proved problematic. While Cryo-EM was once used for large complexes at low resolution, the technique now growing rapidly, with structure solutions at 3 Å resolution of proteins in the range of 50 kDa possible (74, 133). While X-ray crystallography remains the premier technique for protein structure solution for smaller proteins and at resolutions of 2 Å and better, Cryo-EM is fast becoming a very powerful tool in structural biology (73).

Still, functional information about YebT is yet to be fully uncovered; the inner membrane anchoring region at the protein's N-terminus has yet to be categorised, and with it its potential interaction with YebS. Additionally, while the model presented in (131) sought to suggest a binding site for PLs in the cavity of YebT' s tube, this is still in need of confirmation and identification of interacting residues. Using the plasmid constructs and protein purification methods outlined here, potential functionally relevant residues could be identified and tested with future work. Through the optimisation of protein overexpression using of the produced YebST construct, there is potential that insights into the interaction between YebS/T could be gleaned with a crystallographic solution. To this end, the structure of YebS, other than its general topology, is still yet to be determined. Work shown here indicates that a successful expression and purification strategy for it has been

implemented and could be the first step in solving its structure. Indeed, crystallisation trials for YebS remain ongoing, and whilst further detergent screening may be necessary, this is still a promising avenue for pursuit of structural and functional information about the YebST system. The lipidic cubic phase method, wherein protein is crystallised in a membrane-like lipid structure, could also be used for YebS crystallisation (134).

In addition to the work described in this chapter, individual members of the Mla pathway (including MlaFEDB as a complex) have been successfully cloned into vectors suitable for their overexpression or use in functional assays. This will allow further studies in tandem with work on PqiABC and YebS and could help in functional characterisation of phospholipid transport in Gram-negative bacteria.

CHAPTER 4 – Investigation into the binding capabilities of *H. sapiens* SHP2 Protein

4. 1 – Introduction

The proven aptitude for the technique of X-ray crystallography to determine protein structures at high resolutions allows its application in the study of proteins and their interactions across a broad range of targets (75). While the bulk of research presented here is related to the biogenesis and maintenance of the outer membrane in Gram-negative bacteria, work was also carried out in parallel on a *Homo sapiens* derived protein. The binding properties of Tyrosine-protein phosphatase non-receptor type 11 (PTPN11), known henceforth as Src homology region 2 domain-containing phosphatase-2 (SHP2) were investigated at length. This work was carried out in a similar manner as for the previously discussed prokaryotic protein studies, with protein from *H. sapiens* produced using expression plasmids containing appropriate cDNA for SHP2.

4.1.1 Protein Phosphorylation

Phosphorylation is a widespread *in vivo* protein modification and is carried out by protein kinases; these phosphorylation events can have a range of effects. Phosphorylation causes conformational changes, and can result in the activation or inactivation of a protein, as well as permitting interaction with a separate phospho-binding protein (135). Protein phosphorylation is of great biological significance; for example, phosphorylation and dephosphorylation has been shown to be of vital importance to the Eukaryotic cell cycle, which is in part controlled by Cyclins and Cyclin-dependent Kinases (CDKS) (136). Once activated by interactions between Cyclin and CDK helical domains, the protein CDK2 can control G1/S cell-cycle transition through its phosphorylation of Rb, which in turn influences transcription factor E2F (137-140).

Another example as to the significance of protein phosphorylation is the Mitogen-activated protein kinase (MAPK) signalling pathways. The MAPK cellular signalling processes are involved in the transduction of extracellular signals such as that from

epidermal growth factor (EGF) from cell-surface receptors all the way to the activation or suppression of transcription factors (141). This pathway is dependent on a cascade of protein phosphorylation by MAP kinases, MAPK kinases and MAPKK kinases; each kinase is activated by phosphorylation by the previous kinase in the chain (142). This signalling is significant, as the controlled transcription factors - such as Myc - are involved in cell growth and can be a cause of oncogenesis if improperly regulated (143). In Eukaryotes, protein kinases are generally defined by their target residue – these broad classes are tyrosine kinases and serine/threonine kinases (144).

Given that the phosphorylation of proteins is of biological significance, it is logical that their dephosphorylation can be important to cellular function, too (145). For example, the phosphatase CDC25 is involved in the cell cycle, and is an activator of aforementioned CDKs through the removal of inhibitory phosphates (146). Phosphatase activity is also of importance the MAPK signalling cascades –MAPK phosphatases (MKPs) such as MKP3 remove phosphate groups from signalling kinases to reduce transcription factor activation (147). Aberrant action of phosphate enzymes has been shown to cause cancer and neurodegenerative diseases, so a greater understanding of their behaviour is valuable (145).

4.1.2 Src homology region 2 domain-containing phosphatase-2 (SHP2)

SHP2 falls into the category of protein tyrosine phosphatases. The function of these Protein tyrosine phosphatases (PTPs) is the removal of the phosphate from tyrosine residues in signalling molecules, and are opposite in function to protein tyrosine kinases (148, 149). SHP2 is expressed across mammalian tissues, and is essential for growth factor signalling transduction, as well as embryonic development (150). The observed functions of SHP2 are wide-ranging, and it has been shown to regulate growth factor-activate signalling cascades, as well as activation of MAPK pathways (151, 152).

SHP2 contains a pair of SH2 domains at its N-terminus, followed by a PTP domain, with a flexible c-terminal tail region **(Fig 4.1)**. (153) SH2 (Src homology region 2) domains are conserved regions of around 100 amino acids, and take the form of a pair of α-helices around a 4-stranded central β-sheet (154, 155). The function of these domains is to mediate interaction with proteins through 'docking' with phosphorylated tyrosine residues (153). In its resting state, SHP2 is autoinhibited, with the N-terminal SH2 domain blocking the active site of PTP, the activity of which is only displayed when SH2 domains interact with signalling proteins such as growth factor receptors (156). SHP2's protein tyrosine phosphatase domain is structurally very similar to catalytic regions of other protein tyrosine phosphatases, with the most notable feature a long 10-strand β-sheet which wraps around one of the 9 α-helices (157).

Inappropriate activation of SHP2 has been noted in many diseases, including a large number of cancers (151). While the manner in which SHP2 is involved with the causes of many of these cancers is not completely clear, it appears to be vital for the survival and growth of tumour cells (158). In recent years, the SHP2 allosteric inhibitor SHP099 has been developed, which inhibits the protein by stabilising the interaction between SH2 and PTP domain, effectively enhancing the existing autoinhibition (159). This inhibitor has been shown to reduce cancer cell growth, and is both an interesting potential drug and further evidence that SHP2 activity is important in tumour development (160). Incidentally, inhibition of SHP2 has proven to be more effective in reducing tumour growth when programmed cell death protein 1 (PD-1) is concurrently targeted with an antibody (160).

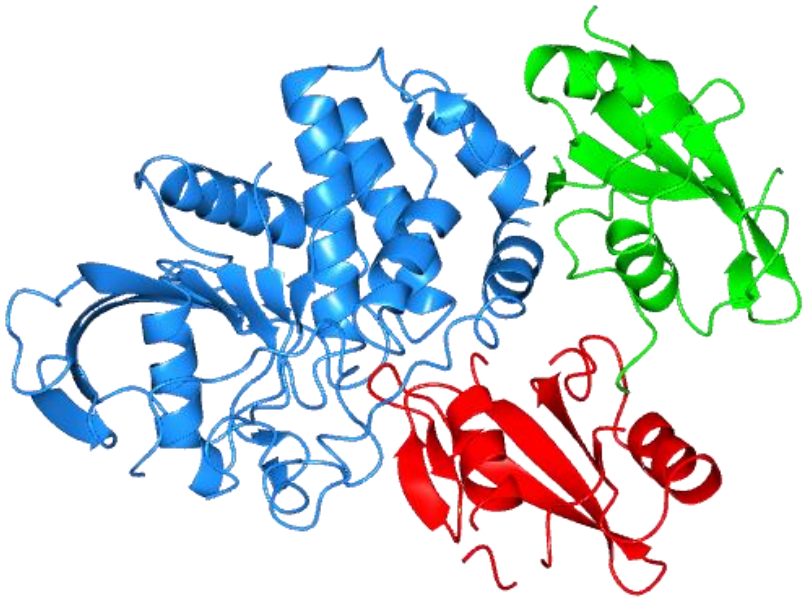


Figure 4.1 - Protein structure of SHP2 at 2.00 Å, without its C-terminal tail. SHP2 is composed of a pair of SH2 domains (red & green) with a β-sheet sandwiched between a pair of α-helices, followed by a protein tyrosine phosphatase domain (Blue), with its long β-sheet twisted around a single helix next to 8 packed helices. PDB code used: 2SHP (155).

4.1.3 Programmed cell death protein 1 (PD-1)

Programmed cell death protein 1 (PD-1) is a transmembrane receptor protein, and is a key regulator of the immune response (161). PD-1 prevents the activation of T cells through binding with their cell-surface ligands PD-L1 and PD-L2 (161, 162). Through this interaction, PD-1 triggers a signalling pathway which prevents autoimmunity by downregulating self-reactive T cells, whilst simultaneously upregulating regulatory T cells (163, 164).

Previous studies have shown that SHP2 interacts with PD-1 through its immunoreceptor tyrosine-based inhibitory motif (ITSM), a short structural feature which acts as a docking region for SHP2 when phosphorylated (165, 166). This recruitment of SHP2 triggers its phosphatase activity due to the displacement of the autoinhibitory SH2 domain (156). Now, SHP2 can act to downregulate signalling pathways activated by T cell binding through dephosphorylation, and ultimately reduce the activity of transcription factors which would otherwise contribute to T cell activation and survival (163). It is through this evasion of T cell activity which reduces autoimmunity that action that PD-1 is thought to contribute to cancer cell survival; cancer cells are often high in PD-1 expression. Monoclonal antibodies preventing PD-1 and PD-L1 interaction have been relatively effective in preventing wide scale evasion of T cell activity, and have been rapidly approved for use in cancer patients (167).

Initial studies from the work of our collaborators had suggested that interactions between the SHP2 protein and PD-1 had been compromised by the compound Methylene Blue, and their data had suggested that MB was an SHP2-binding molecule. This has the potential to be an interesting drug target for cancer treatment, given its potentially improved specificity, with a direct targeting of SHP2-PD-1 interaction. Indeed, severe and potentially deadly side effects have been observed in patients with the use of PD-1 inhibitors, and particularly so the use of antibodies targeting both PD-1 and PD-L1 (168). Methylene Blue, therefore, is an

exciting potential SHP2-PD1 inhibitor; the work described here outlines characterisation of the proposed MB/SHP2 interaction, primarily using X-ray crystallography. Additionally, the docking of SHP2 on phosphorylated ITSM domain of PD-1 has been described but the precise nature of the binding between it and the SH2 domains of SHP2 is still unclear. The work here also seeks to elucidate the specific details of this interaction.

4.1.4 Specific Aim

The work here aims to determine the site within SHP2 at which the compound Methylene Blue binds using X-ray crystallography. Furthermore, the investigation aims to characterise the interaction between PD-1 and SHP2 through the former's ITSM regions.

4. 2 – Materials and Methods

4.2.1 Generation of SHP2 expression plasmids

SHP2 was expressed from *E. coli* cells through their transformation with plasmids containing the *SHP2* gene. Initial expression plasmids pHisTev and pLou3 containing *SHP2* cDNA with N-terminal His tags were generated from SHP2 mRNA using reverse transcription at the UEA by Changjiang Dong (unpublished). To enable the expression of SHP2 with a C-terminal his tag, primers pairs were designed – a pair for the full-length gene, and a pair designed to enable the expression of the first 530 residues of SHP2. The primer pair for truncated *SHP2* contained the restriction sites *NdeI* and *HindIII,* with recognition sites for *NdeI and BamHI* incorporated into primers for the full gene; full primer sequences available in **Table 4.1**.

For both reactions, mixtures were assembled with the Q5 Hot Start system (NEB), with volumes in **Table 2.2**. A volume of 1 µl of the initial N-terminally tagged pHisTev:*SHP2* construct was used as a template in both cases. The PCR program used was: 98°C for 30 seconds, followed by 34 cycles of: 98°C for 10 seconds; Tm°C for 30 seconds and 72°C for 180 seconds. Reactions concluded with incubation at 72°C for 120 seconds, and were held at 4°C. The Tm values used were 63°C and 64°C for truncated and full length *SHP2* respectively.

Table 4.1 – Sequences of Primers used for C-terminally tagged *SHP2* amplification. From an initial N-terminally tagged plasmid constructed by other group members, these primers were used to clone the gene with a C-terminal tag.

Primer Name	Sequence (5'-3')
SHP2_FL_Cter_F	GACAGGCATATGATGACATCGCGGAGATGG
SHP2_FL_Cter_R	GATACGGGATCCTCAGTGATGGTGATGGTGATGTCTGAAACTTTTCTGCTGTTGC
SHP2_530_Cter_F	GACAGGCATATGATGACATCGCGGAGATGG
SHP2_530_Cter_R	GATACGAAGCTTTCAGTGATGGTGATGGTGATGCTGTAGTGTTTCAATATAATGCTGG

PCR success was determined by running samples on a 1% agarose-TAE gel with 6.25% Ethidium bromide for 45 minutes at 110V, and imaging under UV. Successful PCR products, along with recipient pHisTev plasmid, were double restricted with *NdeI* and either *HindIII* or *BamHI* using Thermofisher Fastdigest enzymes. Restricted *SHP2* genes were ligated into pHisTev using Quick Ligase (NEB), in a ratio of 3:1 gene: plasmid.

To amplify plasmids for analysis and transformation into expression cells, they were transformed into Top10 *E. coli* cells using the heat shock method previously outlined. From the resulting cultures, amplified plasmids were purified with a GeneJet plasmid miniprep kit (Thermofisher). Gene insert presence was confirmed by a double restriction of newly cloned plasmids, followed by agarose electrophoresis to confirm gene insert presence. Gene sequence was validated with sequencing at Source Biosciences. Using the heat-shock method, the new plasmids were transformed into expression cells.

4.2.2 Overexpression of SHP2

SHP2 was expressed in *E. coli* Rosetta (DE3) cells, which are designed for Eukaryotic protein expression through their optimisation for codons rarely found in prokaryotes. Transformed cell stocks were initially incubated overnight in LB supplemented with Chloramphenicol and either ampicillin (for pLou3) or Kanamycin (pHisTev) at 37°C with shaking at 200rpm. Fresh LB with antibiotics was inoculated with 20 mL of overnight culture per litre and grown at 37°C 200rpm until OD_{600} reached 0.6. Protein expression was induced with 0.2mM IPTG, with expression taking place at 20°C for either 6 or 16 hours. Cells were harvested with centrifugation and stored at -20°C.

4.2.3 Purification of SHP2

Frozen pellets of Rosetta cells containing overexpressed SHP2 were the starting point for protein purification. The purification methodology was adapted from recent work buy LaRochelle *et al.* (169). Cell pellets were resuspended at room temperature in lysis buffer: 50mM Tris-HCl pH8.5, 500mM NaCl, 2.5mM $MgCl_2$ and 10mM Imidazole, 1 cOmplete EDTA-free protease inhibitor tablet (Roche), 1 µg/ml DNase I (SigmaAldrich) 100 µg/ml lysozyme (SigmaAldrich); lysis buffer was supplemented with either 10% glycerol, 2mM DTT or 1mM TCEP in some purifications.

The resuspended cell mixtures were lysed using a cell disrupter, clarified with centrifugation, and purified by nickel affinity chromatography using methods consistent with those outlined for PqiC (Section 2.1.5). The binding buffer used for SHP2 purification was: 50mM Tris-HCl pH8.5, 500mM NaCl, 2.5mM and 10mM Imidazole, supplemented in some instances with glycerol, DTT or TCEP. Imidazole concentration was varied in line with PqiC methods to elute protein from the HisTrap column. SHP2 was cleaved from its His-tag using TEV proteinase and further purified using a pull down by the same methods described in 2.1.5.

In some SHP2 purifications, protein was purified with an addition step utilising anion exchange chromatography, whereby protein is separated based upon its charge. For this, protein samples were desalted into 50mM Tris-HCl pH8.5, 50mM NaCl, 1mM TCEP buffer on a HiPrep 26/10 desalting column at 2.5 ml /min, run on an ÄKTA Purifier. These samples were then purified using a 5 mL HiTrap Q HP column (GE) run on an ÄKTA Purifier at 2 ml / min, with NaCl concentration increasing to 500 mM

over a volume of 50 mL to elute bound SHP2 from the column. Protein – containing fractions were collected and concentrated as before to 5 mL.

For all SHP2 samples, the final purification step was size-exclusion chromatography (SEC) using a HiLoad 16/60 Superdex 200 pg column (GE) at 1 ml/min. The column was run with buffer containing 20mM Tris-HCl pH 8.5 and 150 mM NaCl, with glycerol or TCEP additives in some instances. Following SEC, protein was concentrated SHP2-530 was concentrated using a 30 kDa MWCO Vivaspin concentrator spun at 3,000 g

4.2.4 Crystallisation of SHP2

Once purity had been established using a PAGE gel, samples of SHP2 were used for crystallisation trials. Protein was used at a concentration of 15 mg/ml, as determined using a 1.5 µl aliquot of protein sample on a Nanodrop 2000c, zeroed to the applicable SEC buffer.

Initial crystallisation trials were carried out using the sitting-well vapour diffusion method, using the same methods and initial screens as used for PqiC trials in section 2.1.6. After initial screening, a 12x6 optimisation screen was made by hand, optimising the condition C4 from PEGION1/2 screen with PEG 3,350 concentrations from 10- 32% and Calcium Acetate concentrations from 160-240 mM. To investigate protein – ligand binding, crystallisation plates were made using this screen with 15 mg/ml SHP2-530 with 20mM Methylene blue or 0.24 mM PD-1 ITSM. All pates were sealed with Clearvue sealing sheets (Molecular Dimensions), incubated at 21 °C and checked for crystal growth periodically under a stereomicroscope.

4.2.5 Data Collection from SHP2 Crystals

Crystals selected for analysis were sent to the synchrotron at Diamond Light Source (DLS) for X-ray diffraction studies. Prior to their transport, crystals were soaked in cryoprotectant matching their mother liquor for 15 minutes to 1 hour, with added glycerol to a final concentration of 20%. Crystals grown in the absence of ligand were soaked in buffers with an additional 50 mM Methylene Blue. Crystals were picked individually in 0.3 mm Litholoops (Molecular Dimensions), flash frozen and added to cooled Unipucks (Molecular Dimensions) to be shipped in a Cryo-Express Dry Shipper (Molecular Dimensions).

Diffraction data was collected from beamlines i03 and i04 at a wavelength of 0.9795 Å. Crystals were exposed for 0.02 seconds in each of 1000 images at 100% beam transmission; these values were varied during the data collection process.

4.2.6 Processing of SHP2 Diffraction Data

Following the collection of X-ray diffraction data sets, they were processed through automated systems. This took place automatically at DLS, with the automated xia2 pipeline (99) generating an MTZ file. This data was reduced and scaled with Aimless (101), which allows the crystal to be assigned a spacegroup, and determines the resolution of the dataset, as well as calculating quality metrics. SHP2-530 crystal structure was determined using MOLREP (79), with existing SHP2 structure 5IBM as its search model (156). The MOLREP software uses the search model to extend phase information to the collected dataset and build a structure of SHP2. The resulting model was refined initially with Refmac (89), to better fit the built structure into the electron density calculated form the diffraction data. COOT was

used to manually refine the built model (90). The structure of apo-SHP2-530 was aligned with the initial search model using PyMol (170), to confirm the protein's identity. Through COOT, structures from soaking and co-crystallisation studies were examined to search for electron density belonging to Methylene Blue or PD-1 ITSM.

Initially, the target ligand was imported and the find ligand feature was used to attempt to assign it an empty region of density. Following this, unmodelled blobs from the electron density map were identified and examined for similarities to ligand structure. Finally, the structures were searched through manually residue by residue to ensure no electron densities which could contain ITSM or MB had been missed.

4.2.7 Isothermal Titration Calorimetry

Full-length SHP2 was titrated against Methylene Blue (MB) using Isothermal Titration Calorimetry (ITC) to further investigate whether MB binds to SHP2 by measuring heat changes during a potential interaction. Methods from (171) were modified and used for this purpose.

Frozen SHP2 was thawed on ice, and dialysed to 150 mM NaCl, 20 mM Tris-HCl pH8.5 buffer devoid of additives. Dialysis was carried out using a 10 kDa mwco Slide-A-Lyzer™ Dialysis Cassette (ThermoFisher) overnight at 4°C with gentle stirring. Dialysed protein was concentrated to 1.4 mg/ml (20 µM) using spin concentration, and MB was diluted to 200 µM in dialysis buffer.

ITC runs were carried out using a MicroCal PEAQ-ITC (Malvern Panalytical). Initial test runs were carried out with water injections into water, and buffer into buffer to ensure that any energy changes were the result of protein – ligand interaction rather

than varying buffer conditions. Experiments used 13 separate injections of protein into the cell (buffered MB) over 45 minutes. Tests were run at 20 µM protein with 200 µM, 450 µM and 2 mM Methylene Blue. Subsequent experiments used 55 µM SHP2 with 450 µM and 2 mM MB.

4. 3 – Results

4.3.1 Expression and purification of SHP2 as an N-terminal MBP fusion

To determine the existence and nature of the interaction between Methylene Blue and SHP2 - and to which part of SHP2 the ITSM of PD-1 associates – a pure, soluble sample SHP2 itself needed to be obtained. Plasmids containing SHP2 cDNA were transformed into *E.coli* Rosetta (DE3) cells, a BL21-derived cell strain optimised for Eukaryotic protein expression by the addition of tRNAs rare in *E. coli* (172).

The initial expression of SHP2 was performed using the pLou3 plasmid vector. This produces SHP2 as a fusion protein with His-tagged Maltose-Binding Protein (MBP) at its N-terminal; this tag was cleavable using TEV proteinase. Protein was expressed overnight after induction with IPTG, with cells harvested using centrifugation. MBP-SHP2 was purified using Immobilised Metal Affinity Chromatography (IMAC), with protein eluted by increasing imidazole concentration. After protein was desalted into imidazole-free buffer, the MBP tag was cleaved overnight. MBP was removed from the protein mixture using a pulldown through an IMAC column, with SHP2 finally purified using Size-Exclusion Chromatography (SEC), with various steps of the purification process analysed using Poly-acrylamide Gel Electrophoresis (PAGE) **(Fig 4.2)**. Samples examined from lysed cells before IMAC showed thick bands at 110 kDa indicating a successful expression of a fusion between SHP2 (~65 kDa) and MBP (45~kDa). However, while some 110 kDa protein *was* observed in the elution from

the IMAC column, large smears of protein at around 65 kDa were seen in column washing fractions, suggesting that some SHP2 was prematurely separated from its fusion partner and lost within the purification procedure. Following TEV cleavage, the amount of MBP visible on the gel was greatly increased, indicating effective tag removal; however, the protein sample tested after the IMAC pulldown was largely devoid of 65 kDa protein, with only a faint, thin band present. When this post-pulldown sample was used for SEC, the predominant peak was at 45 mL retention did appear to contain a small amount of 65 kDa protein when tested with PAGE. The 90 mL SEC peak contained smaller proteins, potentially cleavage products of full-length SHP2. The presence of SHP2 at a retention volume of 45 mL (similar to the column's void volume) suggested that SHP2 had been purified in an aggregated form. In any case, the expression system seemed ineffective at producing stable SHP2 from *E. coli* cells.

4.3.2 Expression and purification of SHP2 with an N-terminal His tag

With expression of SHP2 from pLou3 proving relatively unsuccessful, the protein was instead produced using the pHisTev construct. This aimed to produce SHP2 with an N-terminal His-tag, again removable with TEV proteinase. Expression and purification were carried out in the same manner as the initial pLou3 test, with identical buffers and conditions utilised. As before, PAGE was used to analyse the outcome of the purification **(Fig 4.3)**. The IMAC wash fraction collected contained likely-SHP2 protein (around 70 kDa) which had not bound strongly to the Nickel column. The elution and post-TEV fractions had sharp bands around 70kDa, with a slight reduction in molecular weight after TEV treatment, indicating successful proteolysis; both samples contained a mass of protein at 35 kDa and below, either in the form of contaminants or SHP2 cleavage products. While SHP2 appeared to be

well-expressed, little 70 kDa protein was present in the pulldown fraction, with a thick band visible when this column was eluted suggesting a large amount of protein wasn't successfully cleaved by TEV proteinase. Despite this, a sample taken

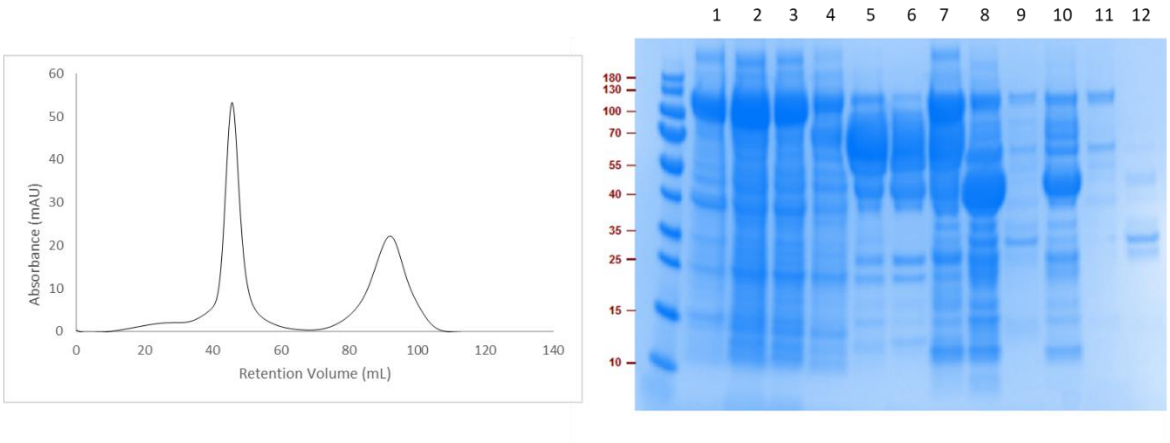


Figure 4.2 - SEC Chromatograph and PAGE gel of initial SHP2 purification. SHP2 was expressed in 12 litres of Rosetta (DE3) *E. coli* cells, as a TEV proteinase cleavable MBP fusion. After cell lysis, protein was purified using IMAC and SEC (Left). Samples from SEC, and various steps from the IMAC process were used to run a PAGE gel (Right). Fractions taken from the post-lysis pellet, its supernatant and initial IMAC flowthrough (1,2 and 3 respectively) all contain thick bands at around 110 kDa, corresponding to the SHP2 (65kDa) and MBP (45kDa) fusion. In the three IMAC washing steps (4,5 and 6), some 110 kDa fusion was apparent, with diffuse bands around 65 kDa. Between the elution (7) and TEV-cleaved fractions (8), the amount of protein seen at 45 kDa increases markedly, with a smaller band at 65 kDa still visible. The pulldown fraction (9) contained little protein, although the column elution (10) contained a large band at 45 kDa, likely the his-tagged MBP after its cleavage from SHP2. Samples obtained from SEC peaks at 45 mL and 90 mL (11 and 12) showed only small amounts of protein visible, with a minor amount of protein visible at 65kDa corresponding to full length SHP2.

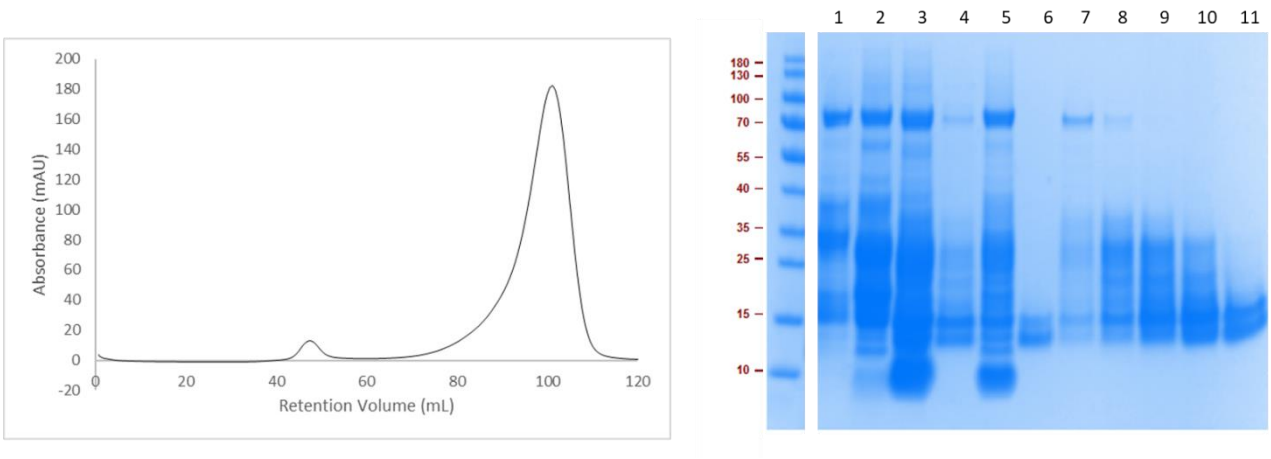


Figure 4.3 - SEC Chromatograph and PAGE of SHP2 purification from pHisTev expression vector. SHP2 was again expressed in 12 litres of Rosetta (DE3) *E. coli* cells, with a TEV cleavable his tag. After cell lysis, protein was purified using IMAC and SEC (Left). Samples from SEC the IMAC processes were used to run a PAGE gel (Right). The IMAC Wash (1), IMAC elution (2) and TEV-cleaved sample (3) all contained putative SHP2 bands at around 70 kDa. The TEV pulldown (4) and pulldown column elution (5) both also contained this band, although most putative SHP2 appeared to be lost to the pulldown phase. Fractions from throughout the broad SEC peak (6-11) contained trace amounts of full – length SHP2, along with many smaller proteins.

from the lower retention volume region of the largest SEC peak at around 85 mL did contain a band at 70 kDa when run on a PAGE gel, suggesting that at least some target protein had remained intact and soluble, in contrast to the pLou3 system. To further investigate the merits of the pHisTev construct, the elution from the IMAC pulldown column was concentrated and purified using SEC, and again analysed with a PAGE gel **(Fig 4.4).** From the additional SEC step, a distinct peak at around 85 mL retention was seen, corresponding to the region in which SHP2 was previously detected in its soluble form. Examination of the PAGE gel showed thick, dark protein bands at around 70 kDa present in samples from this new peak, with diminishing protein of this size present as samples were taken from the larger, previously observed peak at around 100 mL. In test samples from both peaks, there was significant amounts of protein at around 25 kDa, suggesting that while much more SHP2 had been successfully purified using this construct, some was still being lost to degradation.

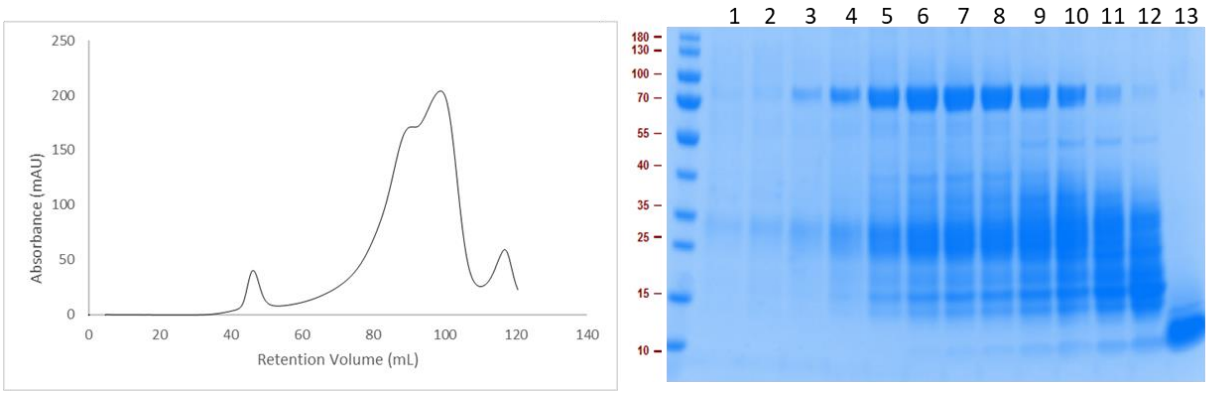


Figure 4.4 - Chromatograph and PAGE of continued SHP2 purification from pHisTev expression pulldown fraction. Following a purification from frozen cells of SHP2 **(Fig 4.3),** a sample eluted from the column used in the pulldown purification step was re-cleaved with TEV proteinase and used for SEC (Left). This chromatograph contained 2 distinct small peaks at 45mL and 115 mL, as well as a much larger peak split into two sections at around 90 mL and 100 mL. Samples were taken from each of these peaks and analysed with a PAGE gel (right). The small 45 mL peak (1,2) contained relatively little protein, with faint bands at 70 kDa. The first of the split peaks (3-9) contained large amounts of protein in the 70 kDa range, but with significant contamination below 40 kDa. The second of these split peaks (10-12) appeared similar, albeit with a diminishing 70 kDa band at higher retention volumes. The final peak at 115 mL (13), contained a group of smaller bands at around 10 kDa.

4.3.3 Expression and purification of SHP2 with glycerol supplementation

In an attempt to combat the apparent degradation of SHP2, it was expressed once more using the pHisTev expression system in Rosetta cells, but purified in buffers supplemented with 10% glycerol in an effort to enhance its stability. In addition, the protein's His-tag remained uncleaved, both to prevent the previously seen loss of protein and to allow the purification to be completed in a shorter timeframe to try to reduce the rate at which SHP2 was degrading. The SEC chromatograph **(Fig 4.5)** indicated a large increase in protein yield, with a broad peak encompassing the previously noted 85 and 100 mL peaks of over 700 mAU, compared to the maximum of 200 mAU recorded in previous SEC runs. PAGE analysis of this broad peak **(Fig 4.5)** yielded thick, dark protein bands at around 70 kDa in samples taken from 70 to 85 mL retention volume, with reduced amounts of putative SHP2 seen in the samples taken from the largest mAU region of the peak around 100 mL. Samples at higher retention volume exhibited significantly more low molecular weight contaminants (likely degraded SHP2), and while these contaminating proteins were reduced in amount in the lower retention region around 85 mL, the improvement from the inclusion of glycerol was seemingly minor.

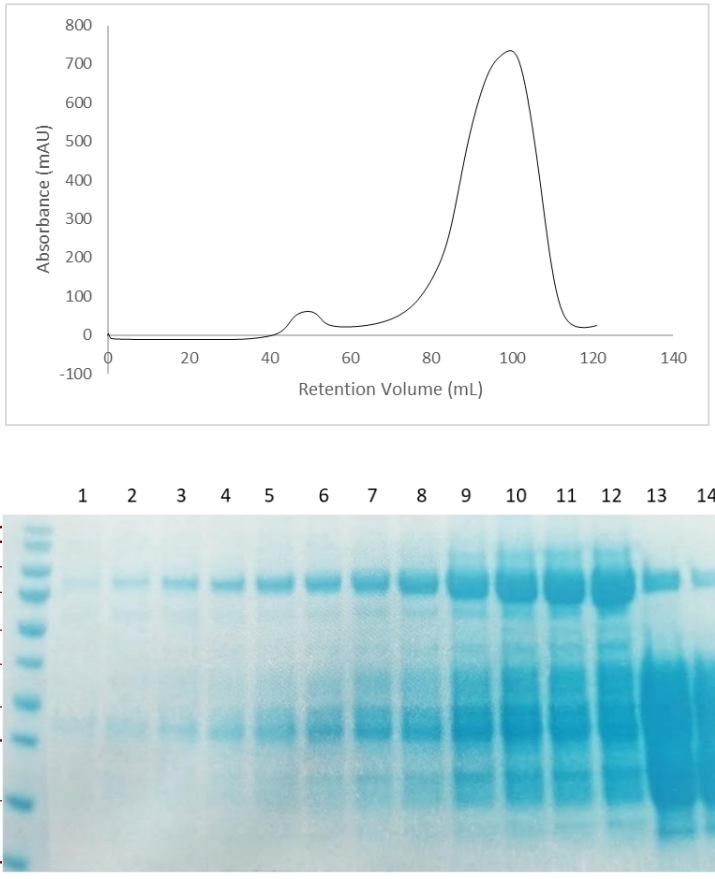


Figure 4.5 - SEC Chromatograph and PAGE of SHP2 purification with the addition of glycerol. His-tagged SHP2 was expressed in 12 litres of Rosetta cells. SHP2 was purified from harvested cells using IMAC and SEC (upper), in the presence of the reducing agent TCEP and 10% glycerol. As in previous expressions, a small peak at around 45 mL retention volume was seen, as well as a large, broader peak at between and 80 and 115mL. Samples from both of these peaks were taken and analysed on a PAGE gel (lower). The small 45 mL peak (1) seemed to contain a small amount of protein around the size of 70 kDa characteristic of SHP2. Samples obtained of the peak from 70 to 85 mL (2-12) indicates that the region between the two main peaks contains protein of around 70 kDa, with increasing amounts of likely SHP2 in addition to increased impurity levels. Samples from the highest absorbance region of the peak (13 and 14) still contain a band at around 70 kDa, but severe amounts of contaminant protein in the 15 – 40 kDa region.

4.3.4 Purification of SHP2 following a 6-Hour Expression

After glycerol as an additive proved ineffective at preventing SHP2 degradation, the protein was instead expressed over a period of 6 hours, as opposed to the 16 hours used up until now. The resulting chromatograph **(Fig 4.6)** was largely similar to previous purifications. There was no discernible aggregated protein peak at 45 mL retention volume, although degradation, as opposed to aggregation had been an issue with the previous experiments. While there was still a large, broad peak at 100 mL, the previously observed 85 mL peak was more noticeable from SEC alone. The PAGE experiment carried out using the SEC samples **(Fig 4.6)** indicated that, as previously seen, the 85 mL peak contained full length SHP2, with the amount of 70 kDa SHP2 reducing as samples neared the 100 mL peak. While a second SEC peak was more pronounced with a reduced expression period, there was still large amounts of smaller protein present in samples run from SEC fractions, and evidently SHP2 degradation was still an issue.

4.3.5 Generation of C-terminally tagged SHP2 Expression Plasmids

With persistent degradation plaguing purification attempts, expression constructs were altered in an effort to maintain SHP2's stability during the purification process. Other studies had shown SHP2 to be successfully purified and subsequently crystallised as a 530-residue truncation, lacking the protein's C-terminal tail (159, 173). For the generation of this new, shortened gene, primers were designed which introduced a stop codon (and hexa-his tag) directly following the codon for residue 530. In addition, a primer pair was designed to allow the conversion of the moderately successful N-terminally tagged pHisTev construct to a C-terminally tagged construct, to see if this would have an appreciable effect on

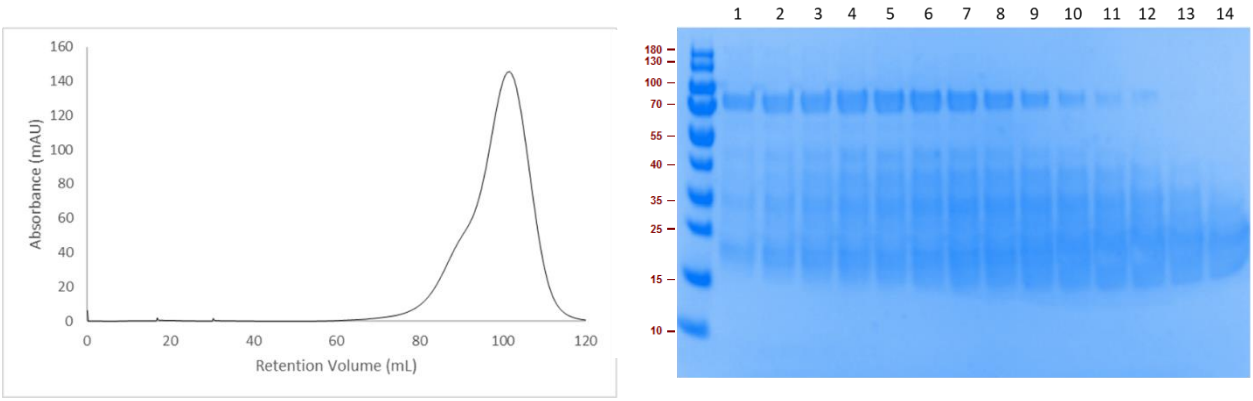


Figure 4.6 - Purification of SHP2 from Rosetta (DE3) cells after a reduced period of protein expression. An expression was carried out for 6 hours at 20°C, to determine the effects on protein degradation. The SEC chromatograph (left) from the purification had no previously seen aggregation peak at around 45 mL retention, and while the characteristic wide peak centred around 100 mL was still observed, it appeared to have a slightly more prominent side peak at around 85 mL. Samples from across the wide peak were analysed with PAGE (right); the side peak centred around 85 mL (1-10) contained good amounts of likely SHP2 at 70 kDa, with the main peak at 100 mL (11-14) exhibiting a reducing proportion of this protein. However, samples taken from each peak still contained similar levels of contaminants, likely degraded protein.

protein purification. Using PCR **(Fig 4.7)** the new primers were used to amplify *SHP2* DNA from the existing N-terminally tagged pHisTev plasmid, with these "new" genes inserted into separate pHisTev vectors using restriction and ligation reactions. These C-terminally tagged SHP2 expression plasmids were transformed with heat shock into Rosetta (DE3) *E. coli* cells for protein production.

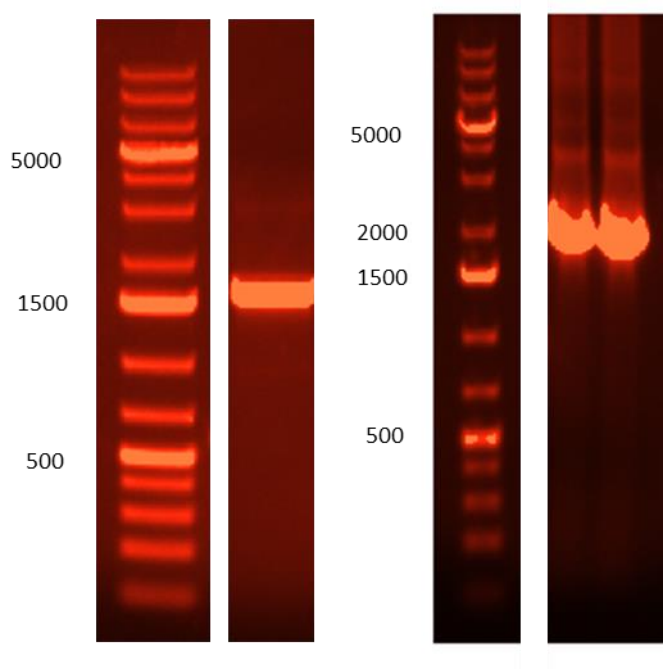


Figure 4.7 - Agarose gels of PCR reactions for the amplification of the *SHP2* gene, imaged under UV light. Using the N-terminally tagged pHisTev:*SHP2* construct as a template, PCR reactions were used to generate two C-terminally tagged *SHP2* genes: a 530-amino acid variant (left) and a full-length product (right).

4.3.6 Purification of SHP2-530 truncation protein

The SHP2-530 construct was the first of the two to be tested in Rosetta cells, with protein produced for 16 hours at 20°C. The cells were harvested by centrifugation and lysed, as before, in a cell disruptor. Protein was purified in buffer supplemented with 1mM TCEP throughout, to prevent potential aggregation. SHP2 was captured using IMAC, and subsequently desalted into a low NaCl buffer (50mM from the 500mM in IMAC elution buffer) to allow for further purification using anion exchange chromatography. SHP2 was loaded onto a HiTrap Q HP column using an ÄKTA Purifier and eluted with an increasing NaCl gradient from 50mM to 500mM, with protein eluted at a range of 150 to 350mM salt concentration **(Fig 4.8)**. This protein was collected and concentrated to allow for the final purification step of SEC and subsequent PAGE analysis, as carried out in previous purification. Results from Size-Exclusion Chromatography **(Fig 4.8)** looked very positive, with a large, sharp peak at around 85 mL corresponding to the retention volume of full-length, soluble SHP2 seen in N-terminal constructs. Furthermore, the absorbance of this peak was over 1400 mAU, double the size of the next-highest peak observed from previous attempts. The fractions in this peak were collected and concentrated to 15 mg/ml for crystallisation trials. To confirm the nature of the protein present following SEC, a sample was taken from the sample part-way through concentration and run on a PAGE gel **(Fig 4.8).** The undiluted sample contained a very high level of protein, with a thick band covering a range of molecular weights, and some protein of lower sizes. A more dilute sample indicated the protein collected was between 55 and 70 kDa – as expected for SHP2 – with minimal amounts of contaminant protein present at this concentration.

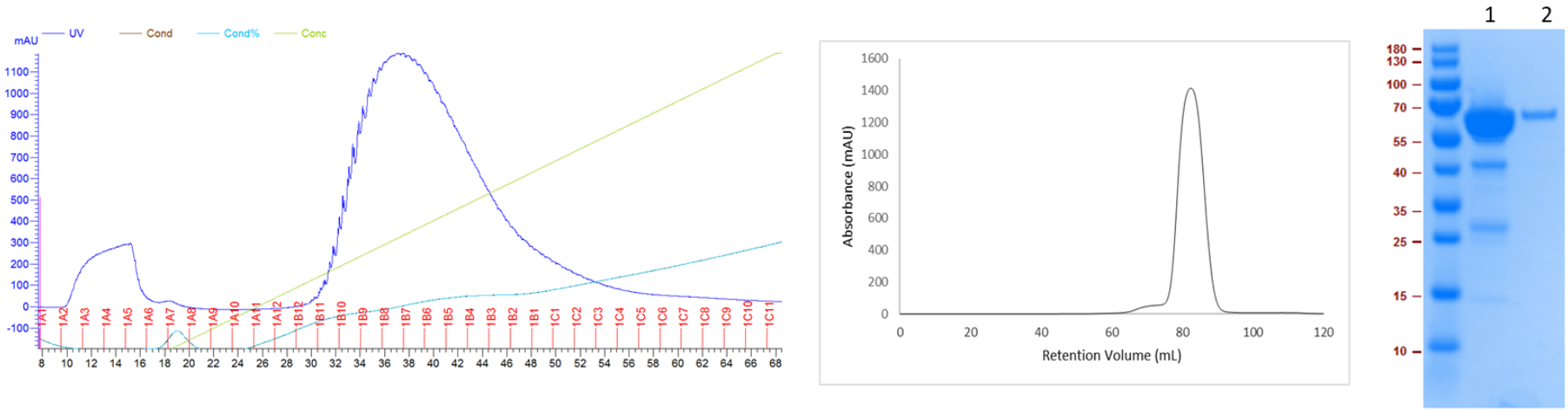


Figure 4.8 - SHP2-530 purification using Immobilised Metal Affinity Chromatography (IMAC), Anion Exchange Chromatography (AEC) and Size-exclusion Chromatography (SEC). SHP2-530 was expressed with a C-terminal His-tag in Rosetta (DE3) *E. coli* cells, in a volume of 12 litres for 6 hours; protein was purified in buffer containing reducing the agent TCEP. During the AEC step (Left), a large peak of over 1000 mAU was observed at a retention volume between 30 mL and 50 mL, corresponding to a NaCl concentration of 150mM to 350mM. Protein from this large peak was used for SEC (centre), resulting in a chromatograph with a small speak at 70 mL retention, and a large, relatively sharp peak of over 1400 mAU at around 80 mL. Protein samples corresponding to this large peak were concentrated to 15 mg / ml and analysed with a PAGE gel (right). The concentrated sample (1) contained small amounts of contaminants alongside a very large band at around 60 kDa, well exceeding the gel' s protein capacity. A 20x dilution of this sample (2) yielded a much sharper band, with no contaminant bands detectable.

4.3.7 Initial Crystallisation of SHP2-530

The purified SHP2-530 sample was deemed suitable for crystal trials, with 10 double-well sitting drop vapour diffusion crystallisation plates constructed using commercially available screens including PACT Premier, JCSG Plus and Peg ION 1/2; plates were incubated at 21°C. After an incubation period of 48 hours, a large number of putative SHP2 crystals were found, with many wells on the PEG ION 1/2 screening plate containing large crystals **(Fig 4.9)**. These were tentatively determined to be protein in nature, due to their fragility and propensity to shatter under minimum pressure with a needle. To confirm whether crystals were indeed SHP2-530, a small number were soaked in cryoprotectants containing 20% glycerol and harvested in crystal loops. These crystals were frozen in liquid nitrogen and shipped to Diamond Light Source (DLS) for data collection. Data was collected from the native crystals at 0.9795 Å, with 1000 images taken with 0.2° oscillation at 0.02 seconds exposure with 100 % beam transmission **(Fig 4.10)**. The structure of SHP2-530 was solved using Molecular replacement, with the existing structure of a 526-atom SHP2 truncation with PDB code 5IBM used as a search model. Data was initially processed at DLS using xia2, with the resulting MTZ file reduced and scaled with Aimless. The data was assigned to the spacegroup P 1 21 1, with an indicated resolution of the dataset of 2.77 Å with an R_{meas} of 0.120. Using the aforementioned search model, the structure of SHP2-530 was solved using MOLREP and refined using 20 Refmac cycles, resulting in an R value of 0.2 and R_{Free} of 0.29. Further refinement including modifications according to the Ramachandran plot were not carried out, given the purpose of the data collection was purely to confirm the identity of the protein. Data parameters from this

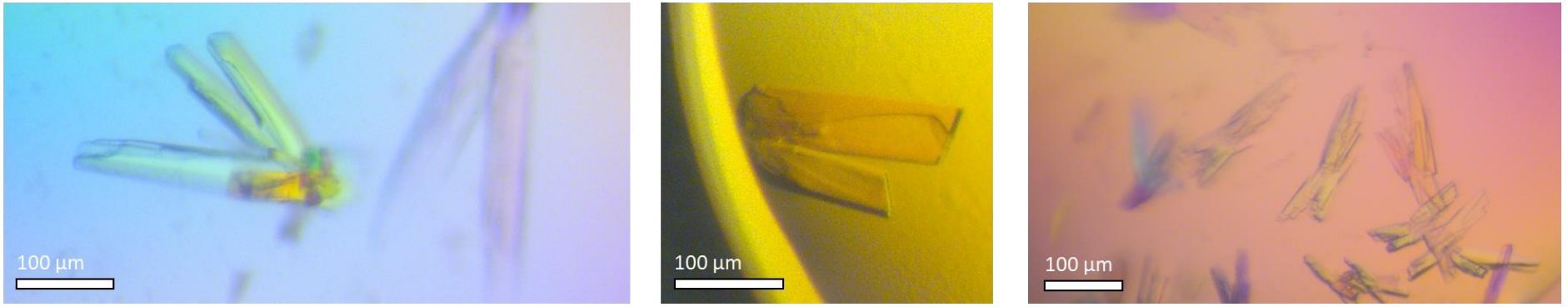


Figure 4.9 - Images of SHP2-530 Crystals obtained after a 24-hour incubation period. After protein samples were incubated for 48 hours using the sitting drop vapour diffusion method with various commercial screens, large numbers of putative protein crystals were obtained. The PEGION 1/2 screen yielded crystals in 27 out of the 96 screened well' s crystallisation solutions, with crystals in wells C3 (Left), C4 (Centre) and C5 (Right) shown. Each of the shown crystals was formed in a well solution containing 20% PEG 3350 and 0.2M of differing acetate salts. Drop size was 0.5 µl well solution: 0.5 µl protein solution.

dataset for the SHP2-530 native structure are in **Table 4.2**. The solved SHP2-530 structure was superimposed on the initial 5ibm search model **(Fig 4.11)**, which indicated that the crystallised protein was almost certainly SHP2-530. Even without stringent refinement, the models aligned with an RMSD of 0.488 Å over 6468 atoms.

Table 4.2 - Data collection and refinement statistics for Native SHP2-530. Statistics from Resolution to multiplicity are given as Overall (outer shell).

Data Collection	
Beamline	i04
Wavelength (Å)	0.9795
Spacegroup	P 1 21 1
Unit Cell Dimensions	
a, b, c (Å)	45.067, 213.120, 55.460
α, β, γ (°)	90.0, 95.743, 90.0
Resolution (Å)	29.89-2.77 (2.92-2.77)
R_{merge}	0.085 (0.528)
R_{meas}	0.12 (0.747)
$CC_{1/2}$ (%)	0.993 (0.661)
Completeness %	98 (92.4)
Mean I/σ (I)	5.8 (1.2)
Multiplicity	1.9 (1.9)
Refinement	
Molecules per AU	2
Total Atoms	15358
Number of Reflections	25879
R_{factor}	0.201
R_{free}	0.288

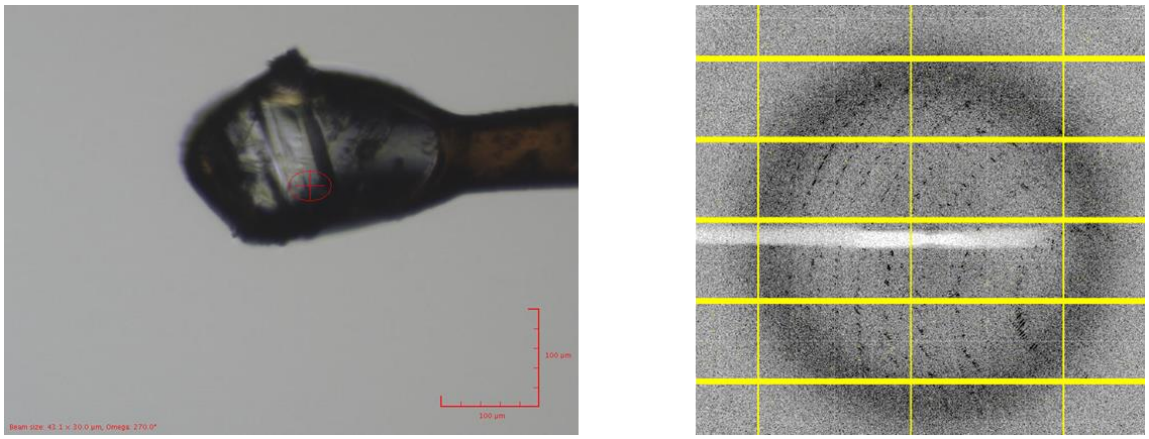


Figure 4.10 - Data collection from Native SHP2-530 crystal at Diamond Light Source, beamline i04. A crystal from the PEG ION C4 condition (0.2 M Calcium acetate, 20% w/v Polyethylene glycol 3,350) was used to confirm the grown crystals were indeed composed of SHP2. Data was collected from the crystal (left) held in a 0.1 mm loop and resulted in a diffraction pattern seen right.

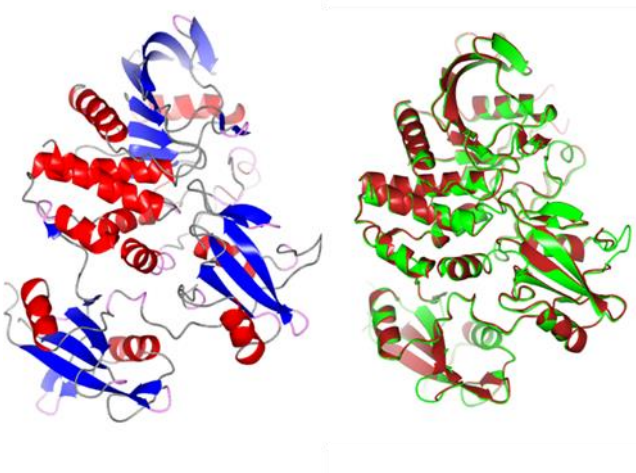


Figure 4.11 - Structure of SHP2-530 solved from native crystal data using Molecular Replacement. This solved structure (left) – including the assignment of side chains – was built almost entirely from molecular replacement, indicating that the tested crystal was formed from SHP2. This was confirmed with a superimposition on existing SHP2 model 5ibm (right), with solved structure shown in green and 5ibm shown in red. Structures were aligned with an RMSD of 0.488 Å over 6468 atoms.

4.3.8 Co-Crystallisation of SHP2-530 with Methylene Blue

With the crystals confirmed as SHP2-530 through the solution of their structure, a further 12 x 6 optimisation screen was made by hand (**Table 4.3)** for use in co-crystallisation trials with Methylene Blue. Co-crystallisation plates were made using this optimisation screen, with a solution of SHP2 at 15 mg/ml, supplemented with 20mM Methylene blue. After 14 days of growth, the formation of crystals similar in shape to those from initial screening had formed **(Fig 4.12)**. Crystals were harvested from wells with 0.5 µl of cryoprotectant added, consisting of the well solution with the addition of 20% glycerol. As before, crystals were submitted to DLS for the collection of X-ray diffraction data.

Data was collected from co-crystallisation samples at 0.9795 Å, at 0.04 seconds exposure with 1000 images taken with 0.2° at 100% beam transmission. As for the native SHP2-530, MTZ files from data integration using xia2 were compiled from DLS following data collection. The crystal forms were incredibly similar to that previously seen, with all datasets assigned to spacegroup P 1 21 1, with very similar unit cell dimensions and bond angles to the native crystal. SHP2 structures from co-crystallisation were solved in exactly the same manner as the apo-protein, using MOLREP followed by Refmac. In total, 15 separate structures of SHP2-530 from co-crystallisation plates with Methylene Blue were built, with a best resolution of 2.41 Å. Structures were examined in COOT to search for potential electron density belonging to MB. This was done manually, as well as by searching for unmodelled blobs and using the 'find ligands' feature. None of the 15 structures contained electron densities likely to correspond to Methylene Blue, with most electron density presence on difference maps too small to be MB **(Fig 4.13)**, even using models without the most rigorous refinement.

Table 4.3 – Optimisation conditions for SHP2-530 crystallisation. Designed around the original PEG ION 1/2 C4 condition (0.2M Calcium Acetate Hydrate, 20% PEG 3,350 pH 7.5), the salt and precipitate concentrations were varied as shown.

		PEG 3,350 (% w/v)											
		10	12	14	16	18	20	22	24	26	28	30	32
Calcium Acetate (mM)	160												
	180												
	200												
	220												
	240												

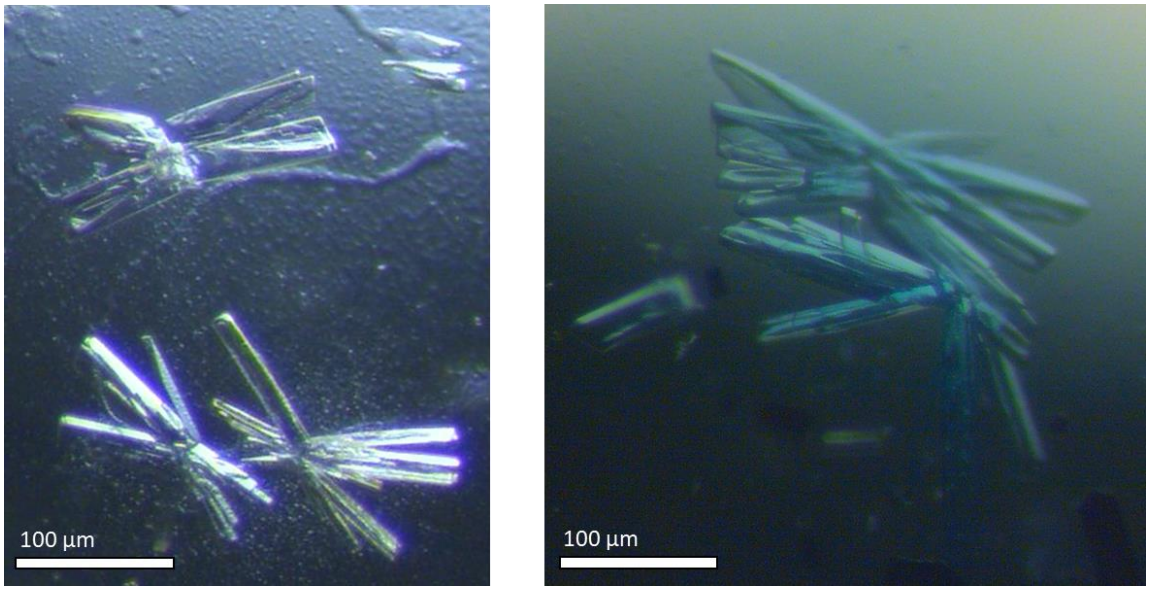


Figure 4.12 - Images of SHP2-530 Co-Crystallised with Methylene Blue (MB) after an incubation of 14 days. Protein samples were supplemented with 20mM MB before crystallisation plates were made. These long, rectangular crystals were formed in wells from the calcium acetate optimisation screen from **Table 4.2**.

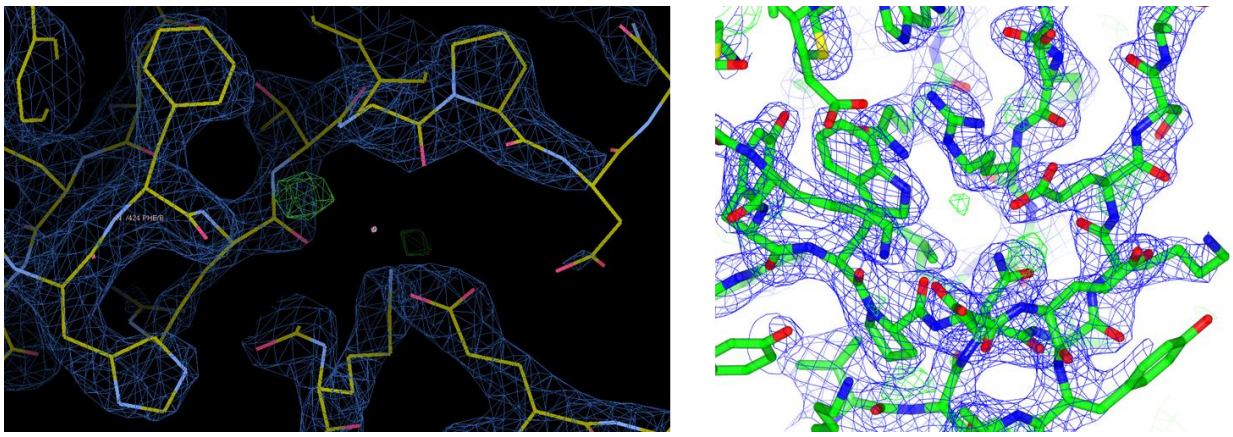


Figure 4.13 - Representative views of SHP2-530 Methylene Blue co-crystallisation electron densities. Across the 15 distinct structures build from co-crystallisation data, MB could not be positively identified in any. While some electron densities were unaccounted for (green regions), none of these were of sufficient size or appropriate shape to suggest Methylene Blue presence in the crystal structures.

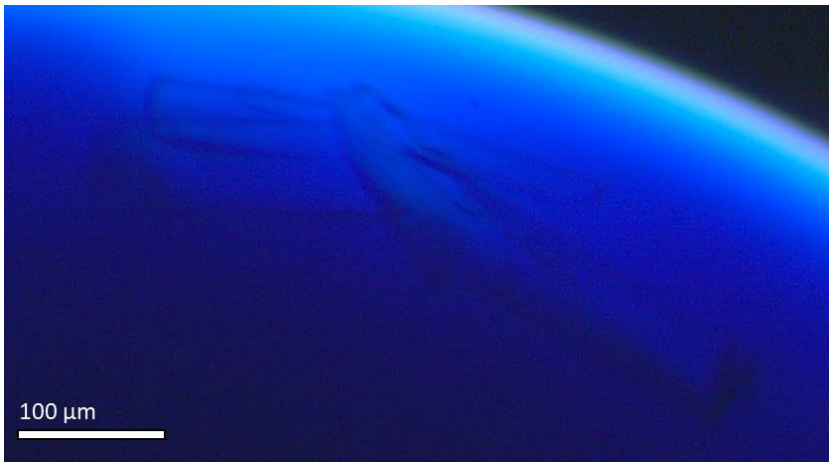


Figure 4.14 - Crystals of SHP2-530 were soaked in cryoprotectants containing 50 mM Methylene Blue (MB). Protein crystals formed from optimisation **(Table 4.2)** and original screen plates were selected for soaking. Owing to the high concentration of MB, the imaging of crystals proved challenging.

4.3.9 Soaking of SHP2-530 with Methylene Blue

With Methylene Blue's presence unable to be determined in solved SHP2-530 co-crystallisation structures, a modified approach was used. Instead of crystallising the protein with 20mM MB, native crystals arising from both the hand-built optimisation screen and the initial PEG ION1/2 screen plate were instead soaked for 15 minutes to 1 hour in suitable cryoprotectants supplemented with 50 mM Methylene Blue **(Fig 4.14)**.These crystals were then harvested and flash frozen as before and submitted to Diamond Light Source for data collection.

Data was collected at 0.9795 Å, with 3600 images taken with 0.1° oscillation at 0.015 seconds exposure with 35 % beam transmission. MTZ files from xia2 processing were retrieved from DLS, and structures solved for 10 different datasets using molecular replacement in the same manner as for apo and co-crystallised data. Again, each crystal was assigned spacegroup P 1 21 1, and had unit cell parameters akin to the apo & co-crystallised forms. For these higher-concentration soaking studies, 10 different structures were produced, to a maximum resolution of 1.98 Å. COOT was again used to examine models and their electron densities to search for molecules of Methylene Blue. From 9 of these models, no unexplained densities were seen. However, the model of 1.98 Å resolution contained an unexpected density in a small cavity near residues K276, Q331, L334 and N336, in the PTP domain of SHP2 **(Fig. 4.15)**. While this seemed to be a possible binding site for Methylene Blue, the ligand appeared too large to fit this density when overlaid on the structure. To determine whether or whether not this was a 'true' binding site, further analyses were required.

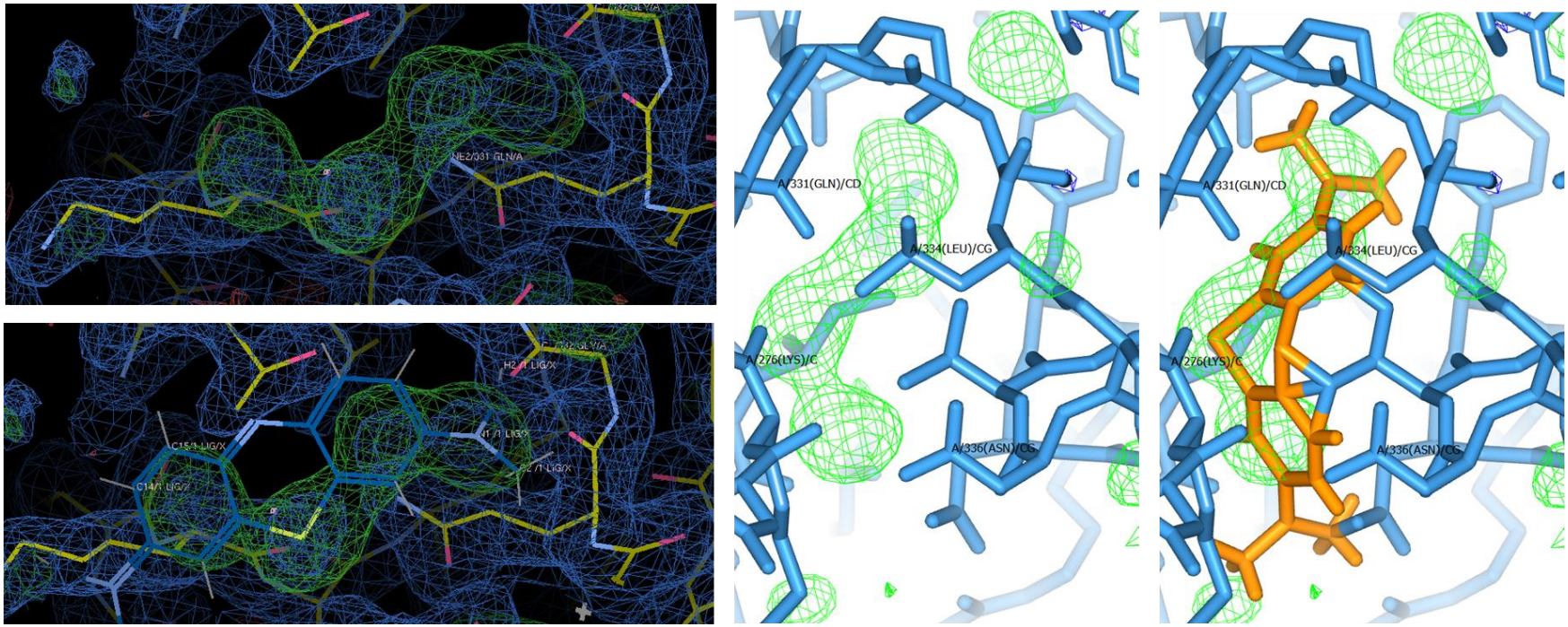


Figure 4.15 - An empty electron density was found near residues K276, Q331, L334 and N336 in the PTP domain of SHP2-530.

This density was not seen in apo or MB co-crystallised SHP2-530 structures. While initially promising, Methylene Blue appeared to be too large and not flexible enough to fit this observed density entirely. MB is shown as blue (left) and orange (right).

4.3.10 Isothermal Titration Calorimetry of SHP2-530 against Methylene Blue

With Crystallisation studies suggesting a potential MB interaction with SHP2, the use of Isothermal Titration Calorimetry (ITC) was considered to investigate the binding of SHP2 and Methylene Blue. As MB binding had been inconclusive following co-crystallisation and crystal soaking experiments, using full-length SHP2 was likely to increase any chances of observing binding to the ligand. To this end, the C-terminally tagged SHP2 pHisTev construct was used to express protein in Rosetta cells for purification. Given the success of SHP2-530 purification, the same 3-step purification method with IMAC, anion exchange chromatography and SEC was used in sample generation. The purification was successful, with a sharp peak of around 600 mAU present on the SEC chromatograph **(Fig 4.17)**. This peak was centred at around 75 mL, with the reduction in retention volume compared to SHP2-530 SEC experiments a good indication of reduced protein size. While protein yield appeared below half of that seen for SHP2-530, it was still a large improvement on the levels of around 200 mAU achieved using N-terminally tagged full-length protein. Analysis of the concentrated C-terminally tagged SHP2 with PAGE **(Fig 4.17)** yielded a tick band of protein between 55 and 70 kDa with minimal contaminants. Diluted samples confirmed this, with sharp bands at around 70 kDa and negligible contaminants visible. Protein was concentrated to 5 mg/ml and stored at – 20°C for use in ITC.

Prior to ITC, SHP2-FL was dialysed into 150 mM NaCl, 20 mM Tris-HCl pH 8.5 using a Dialysis Cassette. Protein was re-concentrated to 1.4 mg/ml (20 µM), and a separate solution of 200 µM Methylene Blue was made in dialysis buffer. ITC was carried out with 13 injections over a run time of 45 minutes, with MB solution as the cell contents and SHP2 in the syringe. Following three separate runs, no discernible energy change was detected by the instrument from any of the injections. Repeat

runs using MB at 450 µM and 2 mM, and protein at 55 µM yielded the same result, with the differential power between reference and sample cells remaining relatively constant **(Fig 4.16).** From these experiments, no detectable binding between SHP2-FL and MB was observed. This information suggested that the mystery density found in SHP2-530 was likely not that belonging to MB. The detected density could have related to other molecules found in the crystallisation or cryoprotectant solution. The density appeared to fit a pair of glycerol molecules well **(Fig 4.18),** which could be a likely explanation as to the nature of the density's cause.

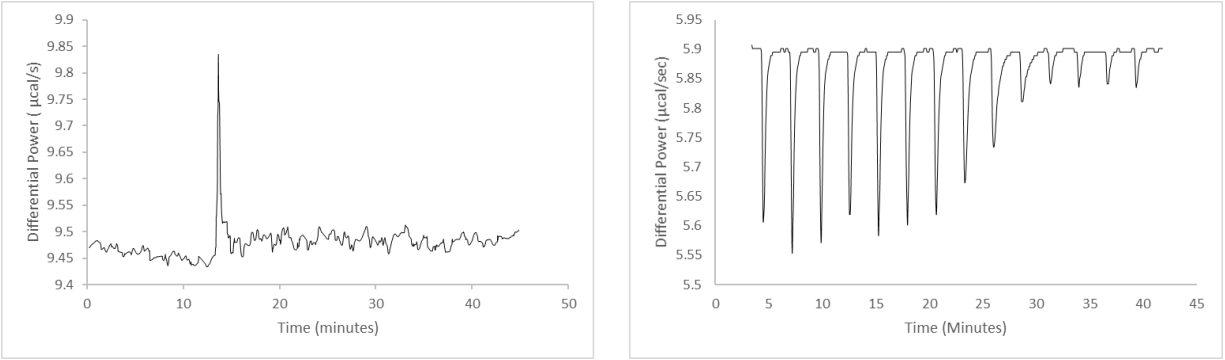


Figure 4.16 ITC failed to show successful binding between SHP2 and Methylene Blue. Traces taken during the mixing of SHP2 and MB (left), failed to indicate binding. Similar results were obtained across three repeats. Differential Power (DP) between the reference and sample cells largely remained at the baseline, except for an aberrant increase in DP at around 15 minutes, potentially due to heterogeneity of the MB or protein mixture. In an example of a successful ITC run from the same instrument (right), DP is reduced during sample injection, before a return to the baseline – each injection has a noticeable effect on the system, in stark contrast to experimentally derived SHP2 data. This positive example was from and interaction between E3 ubiquitin ligase WWP2 and a peptide from Smad7.

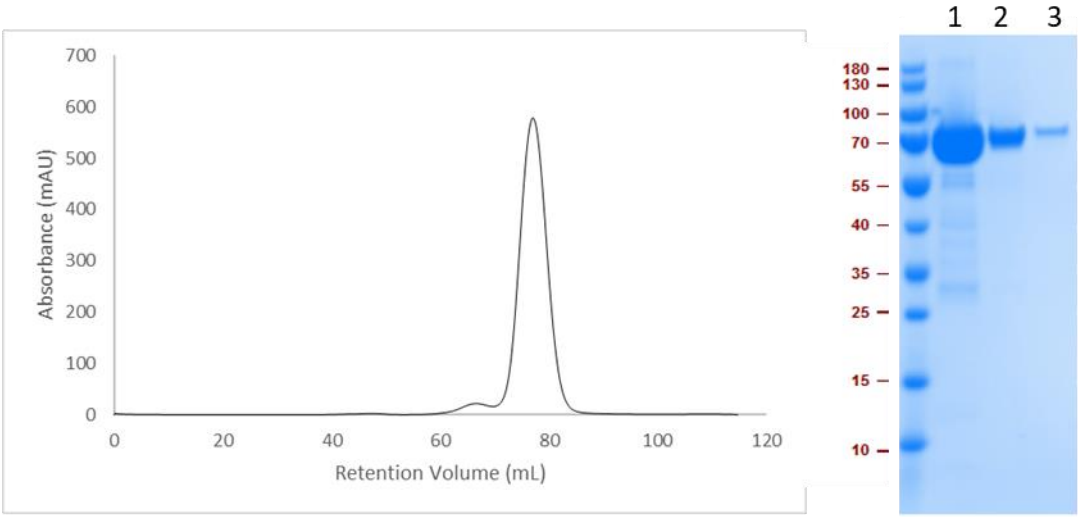


Figure 4.17 - Purification of Full-length SHP2 with a C-terminal His-Tag. Following the success of SHP2-530 purification, the C-terminally tagged full length protein was expressed and purified in the same manner. The SEC chromatograph (left) had a large, sharp peak at around 75 mL, marginally lower in retention that that observed in the SHP2-530 construct. Protein from this peak was collected and concentrated, with the resulting concentrated sample and two serial dilutions run on a PAGE gel (Right, 1-3). The samples all contained a thick band at around 70 kDa, indicating successful purification; while the most concentrated sample (1) had some visible contaminants at lower molecular weights, none of these were visible in the more dilute samples.

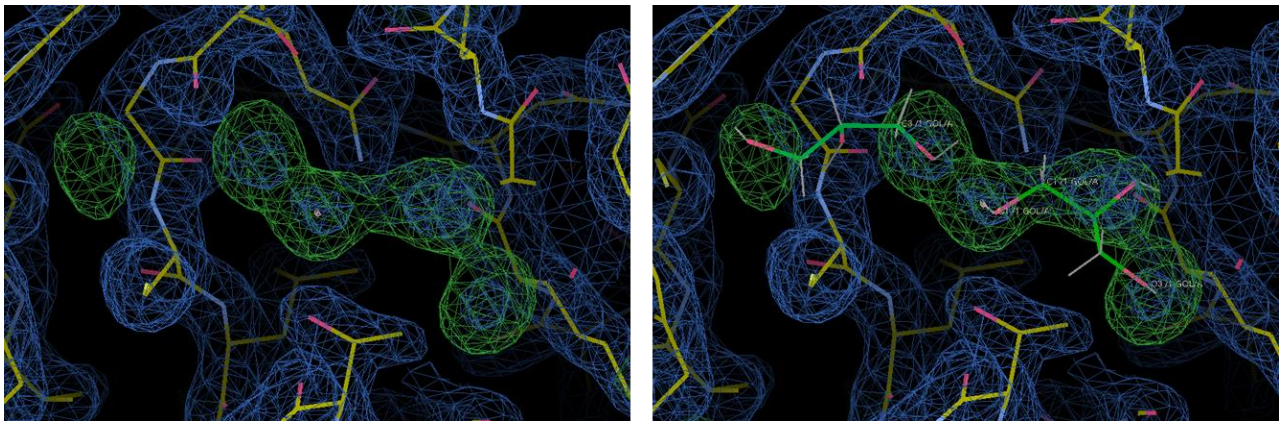


Figure 4.18 - A pair of Glycerol molecules fit the empty electron density in SHP2-530. With ITC indicating a lack of binding between SHP2 and MB, components from the crystallisation solution or cryoprotectant were examined for their fit to this density. In this 1.99 Å structure, glycerol appeared to fit better than MB into this unmodelled electron density.

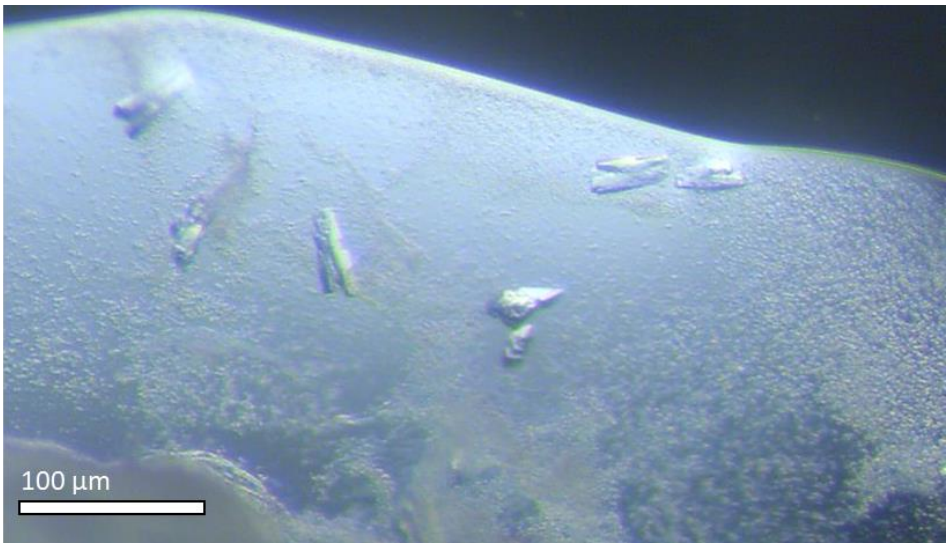


Figure 4.19 - SHP2-530 was co-crystallised with the PD-1 ITSM peptide in a 1:1 molar ratio. Crystals were discovered in wells from the PEGION1/2 C4 optimisation screen **(Table 4.2)** made at a protein concentration of 15 mg/ml.

4.3.11 Co-Crystallisation of SHP2 and the ITSM motif of PD-1

While it appeared that Methylene Blue was not binding to SHP2 – either in its full length or truncated forms – the ITSM of PD-1 was still a potential SHP2 binding partner. To determine if this was the case, a 30-residue peptide containing the motif and SHP2-530 were co-crystallised. Protein was purified as previously outlined, with screening plates made at a SHP2-530 concentration of 15 mg/ml, supplemented with 0.24 mM PD-1 ITSM (a 1:1 molar ratio with SHP2). Crystals were observed after 10 days incubation at 21°C **(Fig 4.19)** from crystallisation plates using the optimisation screen around PEGION1/2 condition C4 **(Table 4.2)**; these were notably shorter than native crystals from the same condition. Crystals were soaked in cryoprotectants containing glycerol to 20% and submitted for data collection at DLS.

Data was collected at 0.9795 Å, with 3600 images taken with 0.1° oscillation at 0.02 seconds exposure with 70 % beam transmission **(Fig 4.20)**. Protein structures were solved using molecular replacement as previously outlined, using integrated MTZ files processed at Diamond with xia2. Spacegroup P 1 21 1 was again selected for the data, with unit cell dimensions matching those for the apo protein; 3 full models were built with a best resolution of 1.96 Å. None of these models contained electron densities suggesting the presence of the 30-residue ITSM peptide, which would have a large, obvious density were it present. Of the models searched, no empty density was discovered which would point to ITSM presence **(Fig 4.21)**.

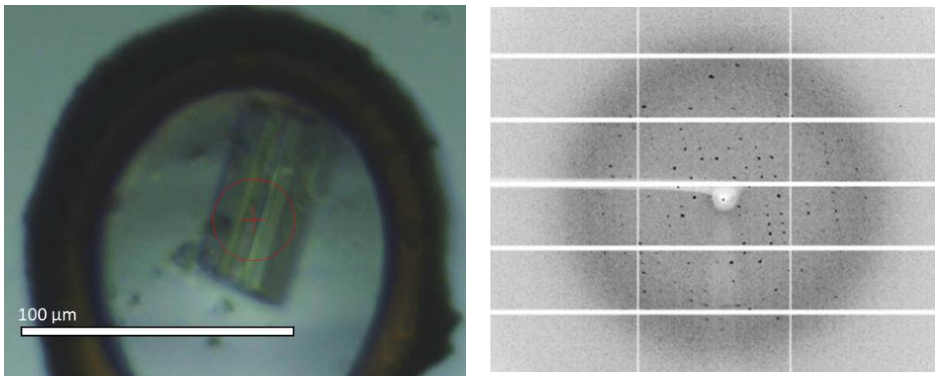


Figure 4.20 - Data collection from SHP2-530 co-crystallised with PD-1 ITSM at Diamond Light Source, beamline i24. The crystal (left) was grown in 1.8 M Calcium Acetate Hydrate, 32% PEG 3,350 pH 7.5 and harvested in a 0.1mm loop. Data colletion resulted in the diffraction pattern seen right.

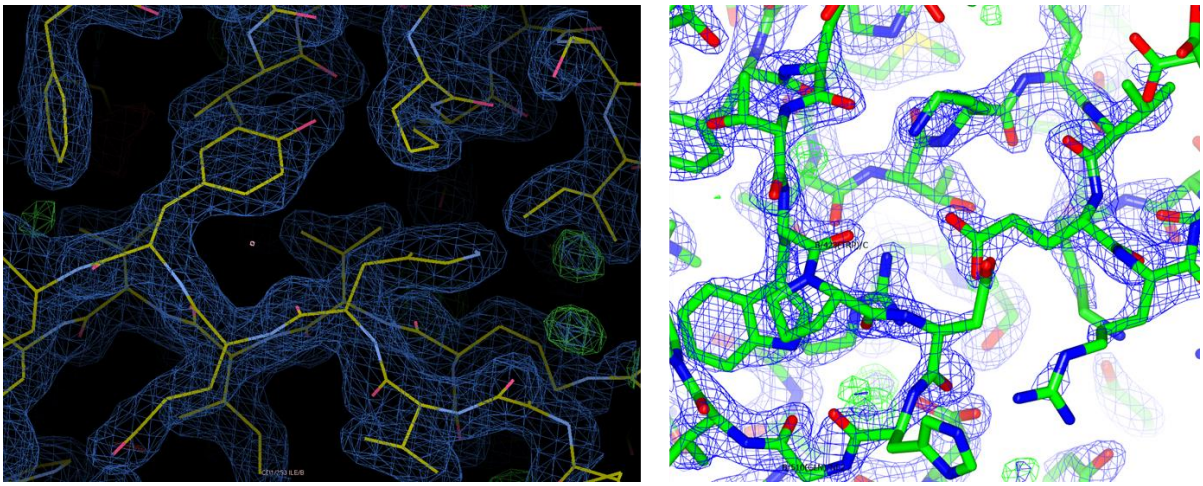


Figure 4.21 - Representative views of SHP2-530 PD-1 ITSM co-crystallisation electron densities. On none of the 3 structures solved from co-crystallisation between SHP2-530 and ITSM could electron density for the 30-residue peptide ITSM be discovered. While small amounts of empty density were present, none of these were large enough to represent a single missing amino acid, let alone a molecule composed of 30.

4. 4 – Discussion

The work outlined here sought to elucidate the binding site and mechanism of SHP2 and the compound Methylene Blue, as well as the known interacting protein PD-1 and its ITSM motif.

Obtaining SHP2 in a pure enough form for crystallisation trials was the initial task at hand, and this process did take some time with iterative improvements. After a change was enacted from N-terminal to C-terminally tagged constructs, purifications improved appreciably, and allowed for crystallisation trials to take place. Unlike the struggles encountered with previous work on the YebT protein, crystal formation from the truncated SHP2-530 protein was relatively straightforward, with protein crystals capable of diffracting X-rays to resolution greater than 3 Å generated relatively soon after a pure protein sample was available. However, while the solution of the protein structure contained within these crystals was relatively straightforward given the availability of a molecular replacement model with 100 % identity in the structure of SHP2, binding of Methylene Blue was not detected. Of the number of crystals leading to solved structures, only one even contained any appreciable electron density which could be attributed to MB presence.

For confirmation as to whether MB was indeed binding to SHP2, the technique of isothermal titration calorimetry (ITC) was used, which measured the energy differences observed when SHP2 was added to a solution containing Methylene Blue. Unfortunately, no such differences were detected, despite repeat measurements with varied concentrations of both protein and putative ligand, requiring a large amount of purified protein. With no detectable binding from ITC, all likelihood suggested that the observed electron density present in the SHP2-530 crystal structure after MB soaking was caused by trapped molecules from the crystallisation solution. Indeed, a pair of glycerol molecules which were introduced as a cryoprotectant seemed to fit better into the density than MB had initially. With

hindsight, it seems apparent that this ITC experiment should have been carried out immediately once a pure SHP2 sample had been obtained, to determine whether crystallographic methods were warranted, and this, must be factored into subsequent work where ITC is a suitable technique to investigate protein-ligand interaction.

However, while this work showed that Methylene Blue does not bind to SHP2, it does help provide evidence that MB interacts instead with PD-1, in the light of work carried out by collaborators. Their work (174) showed that MB specifically prevents SHP2 and PD-1 binding; assays indicated that MB did not prevent PD-1 associating with its known ligand PD-L1, and instead an ELISA using the interaction between SHP2 and PD-1 showed that MB blocked the SHP2-PD1 interaction in a dose dependent manner. With the work here precluding an SHP2-MB interaction, it appears very likely that MB instead blocks the SHP2-PD1 interaction through binding to PD1.This work on MB (174) indicated that Methylene Blue was an effective drug for the suppression of tumour growth in mouse models and appeared to restore the effectiveness of T cells which would otherwise have been downregulated by PD-1 mediated signalling. However, the direct interaction between MB and PD-1 is still in need of further investigation. Future work should focus on this and could proceed in the same manner as that for SHP2. Using the same methodologies outlined here, PD-1 expression could be optimised, with the pure sample initially used for ITC to confirm its interaction with MB. From here, crystal trials of the protein could take place alongside PD-1. Given the transmembrane nature of PD-1, it would be prudent to clone different domains separately for use crystallography experiments to enhance the likelihood of crystal formation.

Following work on the binding of MB and SHP2, the optimised protein production method was used to allow co-crystallisation trials of SHP2-530 and the ITSM motif derived from PD-1. As with MB experiments, though, the ITSM peptide was unable to be located in solved structures. Shortly after this work was carried out, data was

published showing the interaction of ITSM with SHP2, obtained using X-ray crystallography and NMR (175). In this work, the individual SH2 domains of SHP2 were crystallised alongside synthetic 11 residue ITSM-containing peptides. By contrast, the work here used a 530-residue stretch of SHP2 containing both SH2 domains and PTP domain, in co-crystallisation studies with a 30-residue PD-1 constituent peptide containing the ITSM motif. It is possible that co-crystallisation was successful in the work of Marasco *et al.* given the higher concentrations of ITSM utilised in their studies – co-crystallisation experiments here only used a maximum concentration of 0.24 mM ITSM, and it is possible that at this concentration binding between ITSM and SHP2 would not be widespread enough to be determined in a crystal structure. Additionally, the crystallisation of only a single SHP domain alongside the ITSM motif may have led to success due to the reduction of the steric hindrance resulting from the interaction of SHP and PTP domains of SHP2. It is also is possible that co-crystallisation was not observed simply because the 530 residue SHP2 protein was more favourable to be crystallised than that of the complex of SHP2 and the PD-1 ITSM peptide.

Concluding Remarks

The work presented in this thesis sought to utilise the technique X-ray crystallography to determine protein structures and gain functional information on protein - ligand interactions, in combination with other molecular biological techniques.

Undoubtedly, the greatest success in regard to these initial aims was the solution of the novel structure of the *E. coli* PqiC protein to 2.40 Å. Using this discovered structure, functional studies were undertaken which allowed the tentative assignment of specific residues important for phospholipid binding, pending further confirmatory work. In addition, an initial ELISA displaying the ability of wild-type PqiC to bind to PLs could represent an exciting opportunity to further explore PqiC and the Pqi pathway as a whole in regard to its involvement in outer membrane biogenesis.

While the determination of a crystal structure from PqiC can be viewed as an achievement, the attempted structure solution of YebT proved more challenging and was ultimately unsuccessful. While the protein did appear to crystallise successfully, the quality of the data obtained was not of sufficient quality to build a model of the protein, in large part due to the difficulty faced in overcoming the phase problem. While a novel structure of YebT has since been uncovered by cryo-electron microscopy, a structure determined from a crystallographic source could still be of value, to potentially increase resolution, and to corroborate findings from cryo-EM using a distinct technique. Furthermore, a crystal structure of YebT could show the protein in a different state to the existing cryo-EM derived models, and could help better understand the phospholipid transport mechanism of YebT.

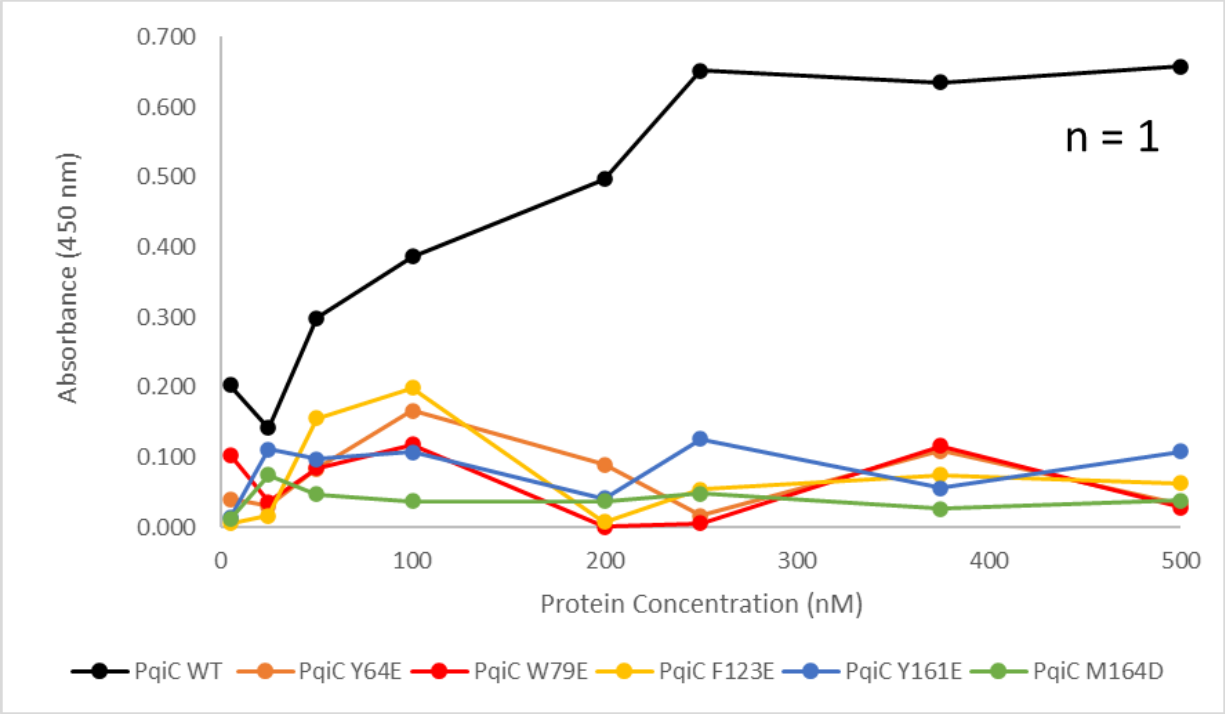
Studies on the SHP2 protein were an interesting avenue to explore, and widened the scope of work undertaken to encompass protein from a Eukaryotic source. While in this case the protein was successfully crystallised and its structure determined, the search for potential ligand binding sites was the aim of this investigation. In combination with co-crystallisation and crystal soaking studies, Isothermal titration calorimetry showed that the target ligand (Methylene Blue) did not interact SHP2. On reflection, once a stable SHP2 sample had been purified, binding studies like ITC should have been carried out prior to co-crystallisation, to hypothetically allow focus to shift onto characterising the interaction between PD-1 and MB, instead.

Overall, the experimentally derived data presented here shows the ability of novel structural information to inform studies on protein function. Further work into the Pqi and Yeb pathways, or the interaction between PD-1, SHP2 and Methylene Blue can be directed by the data here, and serves to show the value of structural biology techniques in the wider field.

Appendix

Appendix Table 1 – Absorbance values taken at 450 nm from preliminary Phospholipid binding ELISA. The test was used investigate phospholipid binding of PqiC mutants; each well was coated with 5 µg of *E. coli* total lipids, before incubation with a varied amount of each purified PqiC species. Bovine serum albumin (BSA) was included as a negative control. Of each of the proteins assayed for PL binding, only wild-type PqiC appeared to exhibit PL binding under these conditions. While this confirms that PqiC binds PLs, this assay will require further optimisation and repeat readings to allow conclusions to be drawn about each purified mutant.

		Protein Concentration (nM)							
		5	25	50	100	200	250	375	500
Protein Species	PqiC WT	0.204	0.143	0.298	0.386	0.497	0.652	0.635	0.658
	PqiC Y64E	0.041	0.030	0.084	0.167	0.090	0.017	0.110	0.033
	PqiC W79E	0.104	0.036	0.085	0.119	0.001	0.006	0.116	0.029
	PqiC F123E	0.006	0.017	0.156	0.199	0.008	0.054	0.075	0.063
	PqiC Y161E	0.014	0.111	0.097	0.107	0.041	0.127	0.057	0.109
	PqiC M164D	0.012	0.075	0.047	0.038	0.038	0.048	0.027	0.038
	BSA	0.116	0.063	0.026	0.126	0.101	0.061	0.050	0.025



Appendix figure 1 - Preliminary ELISA to determine phospholipid binding capability of PqiC mutants. *E. coli* Phospholipid was immobilised on a plate, before PqiC was allowed to bind. After washing steps, protein was detected using a primary α-His antibody and a secondary α-mouse antibody conjugated to HRP. The colour change of substrate TMB was measured at 450 nm to determine the amount of protein which had remained bound to the phospholipid coating. From each PqiC species assayed, only wild-type PqiC appeared to exhibit PL binding under the tested conditions. Further work is needed to optimise and repeat this assay to confirm the PL binding capabilities of the PqiC mutants.

References

1. Schleifer KH, Kandler O. Peptidoglycan types of bacterial cell walls and their taxonomic implications. Bacteriological reviews. 1972;36(4):407.

2. Brown L, Wolf JM, Prados-Rosales R, Casadevall A. Through the wall: extracellular vesicles in Gram-positive bacteria, mycobacteria and fungi. Nature Reviews Microbiology. 2015;13(10):620-30.

3. Zgurskaya HI, Löpez CA, Gnanakaran S. Permeability Barrier of Gram-Negative Cell Envelopes and Approaches To Bypass It. ACS Infect Dis. 2015;1(11):512-22.

4. Silhavy TJ, Kahne D, Walker S. The bacterial cell envelope. Cold Spring Harbor perspectives in biology. 2010;2(5):a000414.

5. Miller SI, Salama NR. The gram-negative bacterial periplasm: Size matters. PLoS Biol. 2018;16(1):e2004935.

6. De Duve C, Wattiaux R. Functions of lysosomes. Annual review of physiology. 1966;28:435-92.

7. Nikaido H. Molecular basis of bacterial outer membrane permeability revisited. Microbiology and molecular biology reviews. 2003;67(4):593-656.

8. Galdiero S, Falanga A, Cantisani M, Tarallo R, Della Pepa ME, D'Oriano V, et al. Microbe-host interactions: structure and role of Gram-negative bacterial porins. Curr Protein Pept Sci. 2012;13(8):843-54.

9. Tamm LK, Hong H, Liang B. Folding and assembly of β-barrel membrane proteins. Biochimica et Biophysica Acta (BBA) - Biomembranes. 2004;1666(1–2):250-63.

10. Lauber F, Deme JC, Lea SM, Berks BC. Type 9 secretion system structures reveal a new protein transport mechanism. Nature. 2018;564(7734):77-82.

11. Ricci DP, Silhavy TJ. The Bam machine: a molecular cooper. Biochim Biophys Acta. 2012;1818(4):1067-84.

12. Driessen AJ, Nouwen N. Protein translocation across the bacterial cytoplasmic membrane. Annu Rev Biochem. 2008;77:643-67.

13. Bechtluft P, Nouwen N, Tans SJ, Driessen AJM. SecB—A chaperone dedicated to protein translocation. Molecular BioSystems. 2010;6(4):620-7.

14. Soltes GR, Schwalm J, Ricci DP, Silhavy TJ. The Activity of Escherichia coli Chaperone SurA Is Regulated by Conformational Changes Involving a Parvulin Domain. Journal of Bacteriology. 2016;198(6):921-9.

15. Rigel NW, Silhavy TJ. Making a beta-barrel: assembly of outer membrane proteins in Gram-negative bacteria. Curr Opin Microbiol. 2012;15(2):189-93.

16. Iadanza MG, Higgins AJ, Schiffrin B, Calabrese AN, Brockwell DJ, Ashcroft AE, et al. Lateral opening in the intact β-barrel assembly machinery captured by cryo-EM. Nature Communications. 2016;7(1):12865.

17. Wilson MM, Bernstein HD. Surface-Exposed Lipoproteins: An Emerging Secretion Phenomenon in Gram-Negative Bacteria. Trends in Microbiology. 2016;24(3):198-208.

18. Miyadai H, Tanaka-Masuda K, Matsuyama S, Tokuda H. Effects of lipoprotein overproduction on the induction of DegP (HtrA) involved in quality control in the Escherichia coli periplasm. J Biol Chem. 2004;279(38):39807-13.

19. Vollmer W, Bertsche U. Murein (peptidoglycan) structure, architecture and biosynthesis in Escherichia coli. Biochimica et Biophysica Acta (BBA) - Biomembranes. 2008;1778(9):1714-34.

20. Okuda S, Tokuda H. Lipoprotein sorting in bacteria. Annu Rev Microbiol. 2011;65:239-59.

21. Kaplan E, Greene NP, Crow A, Koronakis V. Insights into bacterial lipoprotein trafficking from a structure of LolA bound to the LolC periplasmic domain. Proc Natl Acad Sci U S A. 2018;115(31):E7389-E97.

22. Dalbey RE, Chen M. Sec-translocase mediated membrane protein biogenesis. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2004;1694(1):37-53.

23. Dalbey RE, Wang P, Kuhn A. Assembly of Bacterial Inner Membrane Proteins. Annual Review of Biochemistry. 2011;80(1):161-87.

24. Lorenz C, Dougherty TJ, Lory S. Correct Sorting of Lipoproteins into the Inner and Outer Membranes of by the *Escherichia coli* LolCDE Transport System. mBio. 2019;10(2):e00194-19.

25. Grabowicz M. Lipoprotein Transport: Greasing the Machines of Outer Membrane Biogenesis. BioEssays. 2018;40(4):1700187.

26. Rietschel ET, Kirikae T, Schade FU, Mamat U, Schmidt G, Loppnow H, et al. Bacterial endotoxin: molecular relationships of structure to activity and function. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 1994;8(2):217-25.

27. Matsuura M. Structural Modifications of Bacterial Lipopolysaccharide that Facilitate Gram-Negative Bacteria Evasion of Host Innate Immunity. Frontiers in Immunology. 2013;4:109.

28. Erridge C, Bennett-Guerrero E, Poxton IR. Structure and function of lipopolysaccharides. Microbes and infection. 2002;4(8):837-51.

29. Ramachandran G. Gram-positive and gram-negative bacterial toxins in sepsis: a brief review. Virulence. 2014;5(1):213-8.

30. Raetz CR, Reynolds CM, Trent MS, Bishop RE. Lipid A modification systems in gram-negative bacteria. Annual review of biochemistry. 2007;76:295.

31. Raetz CR, Ulevitch RJ, Wright S, Sibley C, Ding A, Nathan C. Gram-negative endotoxin: an extraordinary lipid with profound effects on eukaryotic signal transduction. The FASEB Journal. 1991;5(12):2652-60.

32. Lerouge I, Vanderleyden J. O-antigen structural variation: mechanisms and possible roles in animal/plant–microbe interactions. FEMS Microbiology Reviews. 2002;26(1):17-47.

33. Raetz CR, Whitfield C. Lipopolysaccharide endotoxins. Annu Rev Biochem. 2002;71:635-700.

34. Putker F, Bos MP, Tommassen J. Transport of lipopolysaccharide to the Gram-negative bacterial cell surface. FEMS Microbiology Reviews. 2015;39(6):985-1002.

35. Heinrichs DE, Yethon JA, Whitfield C. Molecular basis for structural diversity in the core regions of the lipopolysaccharides of Escherichia coli and Salmonella enterica. Molecular Microbiology. 1998;30(2):221-32.

36. Amor K, Heinrichs DE, Frirdich E, Ziebell K, Johnson RP, Whitfield C. Distribution of Core Oligosaccharide Types in Lipopolysaccharides from Escherichia coli. Infection and Immunity. 2000;68(3):1116-24.

37. Wang X, Quinn PJ. Lipopolysaccharide: Biosynthetic pathway and structure modification. Progress in lipid research. 2010;49(2):97-107.

38. Sperandeo P, Martorana AM, Polissi A. The lipopolysaccharide transport (Lpt) machinery: A nonconventional transporter for lipopolysaccharide assembly at the outer membrane of Gram-negative bacteria. The Journal of biological chemistry. 2017;292(44):17981-90.

39. Okuda S, Freinkman E, Kahne D. Cytoplasmic ATP hydrolysis powers transport of lipopolysaccharide across the periplasm in E. coli. Science. 2012;338(6111):1214-7.

40. Sperandeo P, Villa R, Martorana AM, Samalikova M, Grandori R, Dehò G, et al. New insights into the Lpt machinery for lipopolysaccharide transport to the cell surface: LptA-LptC interaction and LptA stability as sensors of a properly assembled transenvelope complex. Journal of bacteriology. 2011;193(5):1042-53.

41. May JM, Sherman DJ, Simpson BW, Ruiz N, Kahne D. Lipopolysaccharide transport to the cell surface: periplasmic transport and assembly into the outer

membrane. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2015;370(1679):20150027.

42. Lundstedt E, Kahne D, Ruiz N. Assembly and Maintenance of Lipids at the Bacterial Outer Membrane. Chemical Reviews. 2020.

43. Freinkman E, Chng S-S, Kahne D. The complex that inserts lipopolysaccharide into the bacterial outer membrane forms a two-protein plug-and-barrel. Proceedings of the National Academy of Sciences. 2011;108(6):2486-91.

44. Dong H, Xiang Q, Gu Y, Wang Z, Paterson NG, Stansfeld PJ, et al. Structural basis for outer membrane lipopolysaccharide insertion. Nature. 2014;511(7507):52-6.

45. Botos I, Majdalani N, Mayclin SJ, McCarthy JG, Lundquist K, Wojtowicz D, et al. Structural and Functional Characterization of the LPS Transporter LptDE from Gram-Negative Pathogens. Structure. 2016;24(6):965-76.

46. Raetz CR, Dowhan W. Biosynthesis and function of phospholipids in Escherichia coli. J Biol Chem. 1990;265(3):1235-8.

47. Sohlenkamp C, Geiger O. Bacterial membrane lipids: diversity in structures and pathways. FEMS Microbiol Rev. 2016;40(1):133-59.

48. Zhang Y-M, Rock CO. Membrane lipid homeostasis in bacteria. Nature Reviews Microbiology. 2008;6(3):222-33.

49. Rowlett VW, Mallampalli VKPS, Karlstaedt A, Dowhan W, Taegtmeyer H, Margolin W, et al. Impact of Membrane Phospholipid Alterations in Escherichia coli on Cellular Function and Bacterial Stress Adaptation. Journal of bacteriology. 2017;199(13):e00849-16.

50. Dowhan W. Understanding phospholipid function: Why are there so many lipids? The Journal of biological chemistry. 2017;292(26):10755-66.

51. Dowhan W. Molecular genetic approaches to defining lipid function. J Lipid Res. 2009;50 Suppl(Suppl):S305-10.

52. Lu YH, Guan Z, Zhao J, Raetz CR. Three phosphatidylglycerol-phosphate phosphatases in the inner membrane of Escherichia coli. J Biol Chem. 2011;286(7):5506-18.

53. DeChavigny A, Heacock PN, Dowhan W. Sequence and inactivation of the pss gene of Escherichia coli. Phosphatidylethanolamine may not be essential for cell viability. J Biol Chem. 1991;266(8):5323-32.

54. Nishijima S, Asami Y, Uetake N, Yamagoe S, Ohta A, Shibuya I. Disruption of the Escherichia coli cls gene responsible for cardiolipin synthesis. Journal of bacteriology. 1988;170(2):775-80.

55. Jones NC, Osborn MJ. Translocation of phospholipids between the outer and inner membranes of Salmonella typhimurium. J Biol Chem. 1977;252(20):7405-12.

56. Malinverni JC, Silhavy TJ. An ABC transport system that maintains lipid asymmetry in the gram-negative outer membrane. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(19):8009-14.

57. Chong ZS, Woo WF, Chng SS. Osmoporin OmpC forms a complex with MlaA to maintain outer membrane lipid asymmetry in Escherichia coli. Mol Microbiol. 2015;98(6):1133-46.

58. Chi X, Fan Q, Zhang Y, Liang K, Wan L, Zhou Q, et al. Structural mechanism of phospholipids translocation by MlaFEDB complex. Cell Res. 2020.

59. Abellón-Ruiz J, Kaptan SS, Baslé A, Claudi B, Bumann D, Kleinekathöfer U, et al. Structural basis for maintenance of bacterial outer membrane lipid asymmetry. Nat Microbiol. 2017;2(12):1616-23.

60. Ekiert DC, Bhabha G, Isom GL, Greenan G, Ovchinnikov S, Henderson IR, et al. Architectures of Lipid Transport Systems for the Bacterial Outer Membrane. Cell. 2017;169(2):273-85.e17.

61. Lou H, Chen M, Black SS, Bushell SR, Ceccarelli M, Mach T, et al. Altered antibiotic transport in OmpC mutants isolated from a series of clinical strains of multi-drug resistant E. coli. PLoS One. 2011;6(10):e25825.

62. Isom GL, Davies NJ, Chong Z-S, Bryant JA, Jamshad M, Sharif M, et al. MCE domain proteins: conserved inner membrane lipid-binding proteins required for outer membrane homeostasis. Sci Rep. 2017;7(1):8608-.

63. Ercan B, Low W-Y, Liu X, Chng S-S. Characterization of Interactions and Phospholipid Transfer between Substrate Binding Proteins of the OmpC-Mla System. Biochemistry. 2019;58(2):114-9.

64. Powers MJ, Trent MS. Intermembrane transport: Glycerophospholipid homeostasis of the Gram-negative cell envelope. Proceedings of the National Academy of Sciences of the United States of America. 2019;116(35):17147-55.

65. Shrivastava R, Jiang XE, Chng S-S. Outer membrane lipid homeostasis via retrograde phospholipid transport in Escherichia coli. Molecular Microbiology. 2017;106(3):395-408.

66. Yakhnina AA, Bernhardt TG. The Tol-Pal system is required for peptidoglycan-cleaving enzymes to complete bacterial cell division. Proceedings of the National Academy of Sciences. 2020;117(12):6777-83.

67. Krachler AM, Ham H, Orth K. Outer membrane adhesion factor multivalent adhesion molecule 7 initiates host cell binding during infection by gram-negative

pathogens. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(28):11614-9.

68. Nakayama T, Zhang-Akiyama QM. pqiABC and yebST, Putative mce Operons of Escherichia coli, Encode Transport Pathways and Contribute to Membrane Integrity. J Bacteriol. 2017;199(1).

69. Matias VRF, Al-Amoudi A, Dubochet J, Beveridge TJ. Cryo-transmission electron microscopy of frozen-hydrated sections of Escherichia coli and Pseudomonas aeruginosa. Journal of bacteriology. 2003;185(20):6112-8.

70. UniProt: the universal protein knowledgebase. Nucleic Acids Research. 2017;45(D1):D158-D69.

71. Kovacs-Simon A, Titball RW, Michell SL. Lipoproteins of bacterial pathogens. Infection and immunity. 2011;79(2):548-61.

72. Lorenz C, Dougherty TJ, Lory S. Correct Sorting of Lipoproteins into the Inner and Outer Membranes of Pseudomonas aeruginosa by the Escherichia coli LolCDE Transport System. mBio. 2019;10(2):e00194-19.

73. Kühlbrandt W. The Resolution Revolution. Science. 2014;343(6178):1443-4.

74. Callaway E. Revolutionary cryo-EM is taking over structural biology. Nature. 578. England2020. p. 201.

75. Smyth MS, Martin JH. x ray crystallography. Mol Pathol. 2000;53(1):8-14.

76. Wlodawer A, Minor W, Dauter Z, Jaskolski M. Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. FEBS J. 2013;280(22):5705-36.

77. Taylor G. The phase problem. Acta Crystallographica Section D. 2003;59(11):1881-90.

78. Abergel C. Molecular replacement: tricks and treats. Acta crystallographica Section D, Biological crystallography. 2013;69(Pt 11):2167-73.

79. Vagin A, Teplyakov A. Molecular replacement with MOLREP. Acta Crystallographica Section D. 2010;66(1):22-5.

80. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. Journal of Applied Crystallography. 2007;40(4):658-74.

81. Evans P, McCoy A. An introduction to molecular replacement. Acta crystallographica Section D, Biological crystallography. 2008;64(Pt 1):1-10.

82. Dodson E. Is it jolly SAD? Acta Crystallographica Section D. 2003;59(11):1958-65.

83. Rose JP, Wang B-C. SAD phasing: History, current impact and future opportunities. Archives of Biochemistry and Biophysics. 2016;602:80-94.

84. Blow DM. How Bijvoet made the difference: the growing power of anomalous scattering. Methods Enzymol. 2003;374:3-22.

85. Sheldrick G. A short history of SHELX. Acta Crystallographica Section A. 2008;64(1):112-22.

86. Taylor GL. Introduction to phasing. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 4):325-38.

87. Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallographica Section D. 2006;62(9):1002-11.

88. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 1):12-21.

89. Kovalevskiy O, Nicholls RA, Long F, Carlon A, Murshudov GN. Overview of refinement procedures within REFMAC5: utilizing data from different sources. Acta Crystallogr D Struct Biol. 2018;74(Pt 3):215-27.

90. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallographica Section D. 2004;60(12 Part 1):2126-32.

91. Rondeau J-M, Schreuder H. Chapter 22 - Protein Crystallography and Drug Discovery. In: Wermuth CG, Aldous D, Raboisson P, Rognan D, editors. The Practice of Medicinal Chemistry (Fourth Edition). San Diego: Academic Press; 2015. p. 511-37.

92. Bränd´en C-I, Alwyn Jones T. Between objectivity and subjectivity. Nature. 1990;343(6260):687-9.

93. Kleywegt GJ, Brünger AT. Checking your imagination: applications of the free R value. Structure. 1996;4(8):897-904.

94. Brünger AT. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature. 1992;355(6359):472-5.

95. Inoue H, Nojima H, Okayama H. High efficiency transformation of Escherichia coli with plasmids. Gene. 1990;96(1):23-8.

96. Dubendorff JW, Studier FW. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. J Mol Biol. 1991;219(1):45-59.

97. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. Acta Crystallographica Section D. 2011;67(4):235-42.

98. Potterton L, Agirre J, Ballard C, Cowtan K, Dodson E, Evans PR, et al. CCP4i2: the new graphical user interface to the CCP4 program suite. Acta Crystallographica Section D. 2018;74(2):68-84.

99. Winter G. xia2: an expert system for macromolecular crystallography data reduction. Journal of Applied Crystallography. 2010;43(1):186-90.

100. Waterman DG, Winter G, Gildea RJ, Parkhurst JM, Brewster AS, Sauter NK, et al. Diffraction-geometry refinement in the DIALS framework. Acta Crystallographica Section D. 2016;72(4):558-75.

101. Evans PR, Murshudov GN. How good are my data and what is the resolution? Acta Crystallographica Section D. 2013;69(7):1204-14.

102. Pannu NS, Waterreus W-J, Skubak P, Sikharulidze I, Abrahams JP, de Graaff RAG. Recent advances in the CRANK software suite for experimental phasing. Acta Crystallographica Section D. 2011;67(4):331-7.

103. Abrahams JP, Leslie AGW. Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Crystallographica Section D. 1996;52(1):30-42.

104. Perrakis A, Harkiolaki M, Wilson KS, Lamzin VS. ARP/wARP and molecular replacement. Acta Crystallographica Section D. 2001;57(10):1445-50.

105. Liu H, Naismith JH. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnology. 2008;8(1):91.

106. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2006;2:2006.0008-2006.0008.

107. Chang AY, Chau V, Landas JA, Pang Y. Preparation of calcium competent Escherichia coli and heat-shock transformation. JEMI methods. 2017;1:22-5.

108. Gharavi AE, Harris EN, Asherson RA, Hughes GR. Anticardiolipin antibodies: isotype distribution and phospholipid specificity. Annals of the rheumatic diseases. 1987;46(1):1-6.

109. Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, et al. Tuning Escherichia coli for membrane protein overexpression. Proc Natl Acad Sci U S A. 2008;105(38):14371-6.

110. Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci. 2003;12(8):1652-62.

111. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol. 2019;37(4):420-3.

112. Holm L. DALI and the persistence of protein shape. Protein Science. 2020;29(1):128-40.

113. Malojčić G, Andres D, Grabowicz M, George AH, Ruiz N, Silhavy TJ, et al. LptE binds to and alters the physical state of LPS to catalyze its assembly at the cell surface. Proceedings of the National Academy of Sciences of the United States of America. 2014;111(26):9467-72.

114. Liu H, Naismith JH. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnol. 2008;8:91.

115. Shevchenko A, Tomas H, Havli J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nature Protocols. 2006;1(6):2856-60.

116. Warren JW, Walker JR, Roth JR, Altman E. Construction and characterization of a highly regulable expression vector, pLAC11, and its multipurpose derivatives, pLAC22 and pLAC33. Plasmid. 2000;44(2):138-51.

117. Rath A, Glibowicka M, Nadeau VG, Chen G, Deber CM. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. Proceedings of the National Academy of Sciences. 2009;106(6):1760-5.

118. Dong H, Tang X, Zhang Z, Dong C. Structural insight into lipopolysaccharide transport from the Gram-negative bacterial inner membrane to the outer membrane. Biochim Biophys Acta Mol Cell Biol Lipids. 2017;1862(11):1461-7.

119. Exner M, Bhattacharya S, Christiansen B, Gebel J, Goroncy-Bermes P, Hartemann P, et al. Antibiotic resistance: What is so special about multidrug-resistant Gram-negative bacteria? GMS Hyg Infect Control. 2017;12:Doc05-Doc.

120. Schmidt H, Hensel M. Pathogenicity islands in bacterial pathogenesis. Clin Microbiol Rev. 2004;17(1):14-56.

121. Walden H. Selenium incorporation using recombinant techniques. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 4):352-7.

122. Melnikov I, Polovinkin V, Kovalev K, Gushchin I, Shevtsov M, Shevchenko V, et al. Fast iodide-SAD phasing for high-throughput membrane protein structure determination. Science Advances. 2017;3(5):e1602952.

123. Liebschner D, Afonine PV, Baker ML, Bunkoczi G, Chen VB, Croll TI, et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallographica Section D. 2019;75(10):861-77.

124. Jackson RN, McCoy AJ, Terwilliger TC, Read RJ, Wiedenheft B. X-ray structure determination using low-resolution electron microscopy maps for molecular replacement. Nat Protoc. 2015;10(9):1275-84.

125. Grimm C, Klebe G, Ficner R, Reuter K. Screening orthologs as an important variable in crystallization: preliminary X-ray diffraction studies of the tRNA-modifying enzyme S-adenosylmethionine:tRNA ribosyl transferase/isomerase. Acta Crystallographica Section D. 2000;56(4):484-8.

126. Hui R, Edwards A. High-throughput protein crystallization. Journal of Structural Biology. 2003;142(1):154-61.

127. Altschul SF. Amino acid substitution matrices from an information theoretic perspective. J Mol Biol. 1991;219(3):555-65.

128. Villa R, Martorana AM, Okuda S, Gourlay LJ, Nardini M, Sperandeo P, et al. The *Escherichia coli* Lpt Transenvelope Protein Complex for Lipopolysaccharide Export Is Assembled via Conserved Structurally Homologous Domains. Journal of Bacteriology. 2013;195(5):1100-8.

129. Pluschke G, Overath P. Function of phospholipids in Escherichia coli. Influence of changes in polar head group composition on the lipid phase transition and characterization of a mutant containing only saturated phospholipid acyl chains. Journal of Biological Chemistry. 1981;256(7):3207-12.

130. Isom GL, Coudray N, MacRae MR, McManus CT, Ekiert DC, Bhabha G. LetB Structure Reveals a Tunnel for Lipid Transport across the Bacterial Envelope. Cell. 2020;181(3):653-64.e19.

131. Liu C, Ma J, Wang J, Wang H, Zhang L. Cryo-EM Structure of a Bacterial Lipid Transporter YebT. Journal of Molecular Biology. 2020;432(4):1008-19.

132. Förster F, Villa E. Chapter Three - Integration of Cryo-EM with Atomic and Protein–Protein Interaction Data. In: Jensen GJ, editor. Methods in Enzymology. 483: Academic Press; 2010. p. 47-72.

133. Fan X, Wang J, Zhang X, Yang Z, Zhang J-C, Zhao L, et al. Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution. Nature Communications. 2019;10(1):2386.

134. Caffrey M. A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes. Acta Crystallogr F Struct Biol Commun. 2015;71(Pt 1):3-18.

135. Ardito F, Giuliani M, Perrone D, Troiano G, Lo Muzio L. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). Int J Mol Med. 2017;40(2):271-80.

136. Swaffer MP, Jones AW, Flynn HR, Snijders AP, Nurse P. CDK Substrate Phosphorylation and Ordering the Cell Cycle. Cell. 2016;167(7):1750-61.e16.

137. Malumbres M. Cyclin-dependent kinases. Genome Biology. 2014;15(6):122.

138. Fisher RP. Getting to S: CDK functions and targets on the path to cell-cycle commitment. F1000Res. 2016;5:2374-.

139. Lim S, Kaldis P. Cdks, cyclins and CKIs: roles beyond cell cycle regulation. Development. 2013;140(15):3079-93.

140. Harbour JW, Dean DC. The Rb/E2F pathway: expanding roles and emerging paradigms. Genes Dev. 2000;14(19):2393-409.

141. Pawson T, Scott JD. Protein phosphorylation in signaling – 50 years and counting. Trends in Biochemical Sciences. 2005;30(6):286-90.

142. Zhang W, Liu HT. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. Cell Research. 2002;12(1):9-18.

143. Miller DM, Thomas SD, Islam A, Muench D, Sedoris K. c-Myc and cancer metabolism. Clin Cancer Res. 2012;18(20):5546-53.

144. Paul MK, Mukhopadhyay AK. Tyrosine kinase - Role and significance in Cancer. Int J Med Sci. 2004;1(2):101-15.

145. Cheng H-C, Qi RZ, Paudel H, Zhu H-J. Regulation and Function of Protein Kinases and Phosphatases. Enzyme Research. 2011;2011:794089.

146. Boutros R, Lobjois V, Ducommun B. CDC25 phosphatases in cancer cells: key players? Good targets? Nature Reviews Cancer. 2007;7(7):495-507.

147. Courtney TM, Deiters A. Optical control of protein phosphatase function. Nature Communications. 2019;10(1):4384.

148. Tonks NK. Protein tyrosine phosphatases: from genes, to function, to disease. Nature Reviews Molecular Cell Biology. 2006;7(11):833-46.

149. Qu CK. The SHP-2 tyrosine phosphatase: Signaling mechanisms and biological functions. Cell Research. 2000;10(4):279-88.

150. Salmond RJ, Alexander DR. SHP2 forecast for the immune system: fog gradually clearing. Trends in Immunology. 2006;27(3):154-60.

151. Zhang J, Zhang F, Niu R. Functions of Shp2 in cancer. J Cell Mol Med. 2015;19(9):2075-83.

152. Zehender A, Huang J, Györfi AH, Matei AE, Trinh-Minh T, Xu X, et al. The tyrosine phosphatase SHP2 controls TGFβ-induced STAT3 signaling to regulate fibroblast activation and fibrosis. Nat Commun. 2018;9(1):3259.

153. Kaneko T, Joshi R, Feller SM, Li SSC. Phosphotyrosine recognition domains: the typical, the atypical and the versatile. Cell Communication and Signaling. 2012;10(1):32.

154. Schlessinger J, Lemmon MA. SH2 and PTB domains in tyrosine kinase signaling. Sci STKE. 2003;2003(191):Re12.

155. Filippakopoulos P, Müller S, Knapp S. SH2 domains: modulators of nonreceptor tyrosine kinase activity. Curr Opin Struct Biol. 2009;19(6):643-9.

156. LaRochelle JR, Fodor M, Xu X, Durzynska I, Fan L, Stams T, et al. Structural and Functional Consequences of Three Cancer-Associated Mutations of the Oncogenic Phosphatase SHP2. Biochemistry. 2016;55(15):2269-77.

157. Hof P, Pluskey S, Dhe-Paganon S, Eck MJ, Shoelson SE. Crystal structure of the tyrosine phosphatase SHP-2. Cell. 1998;92(4):441-50.

158. Song Z-D, Wang M-J, Ge Y, Chen X-P, Xu Z-Y, Sun Y, et al. Tyrosine phosphatase SHP2 inhibitors in tumor-targeted therapies. Acta Pharmaceutica Sinica B. 2020.

159. Pádua RAP, Sun Y, Marko I, Pitsawong W, Stiller JB, Otten R, et al. Mechanism of activating mutations and allosteric drug inhibition of the phosphatase SHP2. Nat Commun. 2018;9(1):4507.

160. Zhao M, Guo W, Wu Y, Yang C, Zhong L, Deng G, et al. SHP2 inhibition triggers anti-tumor immunity and synergizes with PD-1 blockade. Acta Pharmaceutica Sinica B. 2019;9(2):304-15.

161. Riley JL. PD-1 signaling in primary T cells. Immunol Rev. 2009;229(1):114-25.

162. Marchetti A, Di Lorito A, Buttitta F. Why anti-PD1/PDL1 therapy is so effective? Another piece in the puzzle. J Thorac Dis. 2017;9(12):4863-6.

163. Sharpe AH, Pauken KE. The diverse functions of the PD1 inhibitory pathway. Nature Reviews Immunology. 2018;18(3):153-67.

164. Francisco LM, Sage PT, Sharpe AH. The PD-1 pathway in tolerance and autoimmunity. Immunol Rev. 2010;236:219-42.

165. Boussiotis VA. Molecular and Biochemical Aspects of the PD-1 Checkpoint Pathway. N Engl J Med. 2016;375(18):1767-78.

166. Sheppard K-A, Fitz LJ, Lee JM, Benander C, George JA, Wooters J, et al. PD-1 inhibits T-cell receptor induced phosphorylation of the ZAP70/CD3ζ signalosome and downstream signaling to PKCθ. FEBS Letters. 2004;574(1-3):37-41.

167. Akinleye A, Rasool Z. Immune checkpoint inhibitors of PD-L1 as cancer therapeutics. Journal of Hematology & Oncology. 2019;12(1):92.

168. Moslehi JJ, Salem JE, Sosman JA, Lebrun-Vignes B, Johnson DB. Increased reporting of fatal immune checkpoint inhibitor-associated myocarditis. Lancet. 2018;391(10124):933.

169. LaRochelle JR, Fodor M, Vemulapalli V, Mohseni M, Wang P, Stams T, et al. Structural reorganization of SHP2 by oncogenic mutations and implications for oncoprotein resistance to allosteric inhibition. Nature Communications. 2018;9(1):4508.

170. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 2.0. 2020.

171. Duff MR, Jr., Grubbs J, Howell EE. Isothermal titration calorimetry for measuring macromolecule-ligand affinity. Journal of visualized experiments : JoVE. 2011(55):2796.

172. Tegel H, Tourle S, Ottosson J, Persson A. Increased levels of recombinant human proteins with the Escherichia coli strain Rosetta(DE3). Protein Expr Purif. 2010;69(2):159-67.

173. Garcia Fortanet J, Chen CH, Chen YN, Chen Z, Deng Z, Firestone B, et al. Allosteric Inhibition of SHP2: Identification of a Potent, Selective, and Orally Efficacious Phosphatase Inhibitor. J Med Chem. 2016;59(17):7773-82.

174. Fan Z, Tian Y, Chen Z, Liu L, Zhou Q, He J, et al. Blocking interaction between SHP2 and PD-1 denotes a novel opportunity for developing PD-1 inhibitors. EMBO Molecular Medicine. 2020;12(6):e11571.

175. Marasco M, Berteotti A, Weyershaeuser J, Thorausch N, Sikorska J, Krausze J, et al. Molecular mechanism of SHP2 activation by PD-1 stimulation. Science Advances. 2020;6(5):eaay4458.