CHARACTERISATION OF THE ASMA-LIKE PROTEIN FAMILY IN *ESCHERICHIA COLI*

by

RACHEL HOLYFIELD

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

School of Biosciences

College of Life and Environmental Sciences

University of Birmingham

September 2024

ABSTRACT

Understanding the mechanisms of cell envelope biogenesis in Gram-negative bacteria could reveal novel therapeutic targets to combat rising antimicrobial resistance. Despite decades of research, the fundamental process of phospholipid transport from the inner membrane to the outer membrane of the Gram-negative cell envelope has long remained an elusive process. Recently, the AsmA-like superfamily have emerged as likely contenders of this intermembrane phospholipid transport. With six AsmA-like members in *E. coli* (YhdP, TamB, YdbH, YhjG, AsmA and YicH) and mostly unknown structures and functions for each, many questions have now emerged about their exact role in the cell and their potential mechanisms of action.

Here, we present the phenotypic profiling of the *asmA*-like superfamily, supporting recent evidence of an essential and redundant function in outer membrane biogenesis between *yhdP* and *tamB*, identifying a novel gain-of-fitness phenotype and providing evidence of a genetic interaction between *yhdP* and *asmA*. We also describe trials to express and purify these proteins, discovering that they are often prone to breakdown and aggregation. Nevertheless, we then present low-resolution cryo-EM structures of YhjG and YicH, the first experimental structures of any of the full-length AsmA-like proteins. We show the first evidence of phospholipid transport via YdbH, YhjG and YicH using an *in vitro* FRET-based phospholipid transport assay and reveal a spontaneous mechanism of phospholipid entry through coarse-grained molecular dynamics simulations. This investigation sheds light on the role of AsmA-like proteins in phospholipid transport and lays the groundwork for future structural and mechanistic studies of both this protein family and other phospholipid transporter proteins.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr Tim Knowles for his guidance and support throughout the project. My scientific career started as a Masters student in the Knowles lab, and it is through encouragement and scientific freedom from Tim that I have been able to develop my own research ideas and scientific mindset. I would also like to thank my co-supervisor, Professor Andy Lovering, for providing both scientific advice and an infectious enthusiasm for all things structural biology. I will miss hearing about the latest cool structure you've found.

Thank you to the whole Knowles group, past and present. Dave, Pete, Pooja, Hannah, Giedrė, Ben, Gareth, Adam and Charlotte – thank you all for being such a fun and supportive lab, for providing excellent advice and mentorship, and helping with any problem, big or small. To the whole 7th floor – thank you for creating such a welcoming and fun work environment. I've had so many fun memories throughout my time here because of you all: lab socials, conferences, lunchtime crosswords, and many coffee breaks. Particularly big thanks must go to my ever-supportive friends and fellow PhD-ers Beth and Lauren – we made it! I'm very lucky to have met you both during the PhD and am very grateful for our friendship.

Whilst on the theme of friends, huge thanks to my friends, new and old, who remind me that there is so much more to life than the PhD. I am particularly indebted to my best friends from home, Beth, Ursi and Lauren – they have been by my side through every moment.

I must give the biggest thanks to my family: Gransha – for teaching me how to be patient and kind no matter what, Ben – for reminding me to have fun and enjoy life, and Mum and Dad – for always picking up the phone on my walks into campus, travelling to Birmingham just to take me out for food, and for always believing in me.

The work in this thesis was performed with the help of many incredible scientists. In no particular order, thank you to Dr Simon Caulton, who provided invaluable scientific and project advice. Thank you to Dr Manuel Banzhaf, Dr Jack Bryant, Michael Aleo, Dr Gabrielle Boelter and Dr Hannah Doherty for so much help on the genetics side of the project - your eagerness to help and willingness to teach me new skills were instrumental to the findings presented in this thesis. A big thank you to Dr Andrew Quigley, Dr Peter Harrison, Claudia Burn-Leefe, and all the other lab members, for your expertise, help and friendliness during my time at the Membrane Protein Laboratory. Thank you to my project student Olamide Babatunde who performed some of the final complementation and FRET transport assays. Thank you to Dr Natalie Allcock, Dr Christos Savva, Dr Claudia Lancey, Dr TJ Ragan and Dr Emma Hesketh at the Leicester Institute of Structural and Chemical Biology for their teachings and expertise in all things EM. Thank you also to Professor Natasha Ruiz, who kindly provided the knockout strains studied in this thesis.

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) and University of Birmingham funded Midlands Integrative Biosciences Training Partnership [grant number BB/T00746X/1]. Part of this work was also supported by the proposal MX35602: “Determining optimal purification conditions for some of the *E. coli* AsmA-like proteins”, funded by the Membrane Protein Laboratory and Diamond Light Source.

I dedicate this work to Gransha, who I miss every day.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A22	*S*-(3,4-dichlorobenzyl) isothiourea
AA	All-atom
ABC	ATP binding cassette
AI	Auto-induction
AMR	Antimicrobial resistance
ANOVA	Analysis of variance
ATG	Autophagy-related
ATP	Adenosine triphosphate
AUC	Analytical ultracentrifugation
Bam	β-barrel assembly machinery
BPA	p-benzoyl-L-phenylalanine
BLTP	Bridge-like lipid transport protein
CDP-DAG	Cytidine-diphosphate diacylglycerol
Cds	CDP-DAG synthase
CG	Coarse-grained
CHS	Cholesteryl hemisuccinate
CL	Cardiolipin
Cls	Cardiolipin synthase
CMC	Critical micelle concentration
CPS	Capsular polysaccharride
Cryo-EM	Cryogenic electron microscopy
CTAB	Cetryltrimethylammonium bromide
CTF	Contrast transfer function
CV	Column volume
DDM	*n*-dodecyl-β-D-maltoside
DNA	Deoxyribonucleic acid
DSBU	Disuccinimidyldibutyric urea
DSF	Differential scanning fluorimetry
DUF	Domain of unknown function
ECA	Enterobacterial common antigen
ECL	Enhanced chemiluminescence

EDTA	Ethylenediaminetetraacetate
ERMES	ER-mitochondria encounter structure
EM	Electron microscopy
EPL	*E. coli* polar lipids
FC-12	Fos-choline 12
FEG	Field emission gun
FL	Full-length
FRET	Fourier resonance energy transfer
FSC	Fourier shell correlation
F-SEC	Fluorescence size exclusion chromatography
FT	Flow-through
GDN	Glyco-diosgenin
GPL	Glycerophospholipid
HPLC	High performance liquid chromatography
HRP	Horseradish peroxidase
HSD	Honestly significant difference
IM	Inner membrane
IMP	Inner membrane protein
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
KDO	3-Deoxy-D-manno-oct-2-ulosonic acid
KLD	PNK, T4L and DpnI
LB	Lysogenic broth
LDAO	Lauryldimethylamine oxide
Let	Lipophilic envelope spanning tunnel
LMNG	Lauryl maltose neopentyl glycol
Lol	Localisation of lipoprotein
LPD	Lipid depleted
Lpp	Braun's lipoprotein
LPS	Lipopolysaccharide
Lpt	Lipopolysaccharide transport
LSB	Lauryl-sulfobetaine
LTP	Lipid transfer protein

MCE	Mammalian cell entry
MD	Molecular dynamics
MES	2-(N-morpholinoethanesulphonic acid)
ML	Machine learning
Mla	Maintenance of lipid asymmetry
MW	Molecular weight
MWCO	Molecular weight cut-off
NAG	N-acetylglucosamine
NAM	N-acetylmuramic acid
NAPol	Non-ionic amphipol
NBD	Nitrobenzoxadiazole
NPT	Isothermal-isobaric ensemble
NSPr	Nanodisc scaffold peptide
NTA	Nitrilotriacetic acid
NVT	Canonical ensemble
OD	Optical density
OM	Outer membrane
OMP	Outer membrane protein
PA	Phosphatidic acid
PAE	Predicted aligned error
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PC	Phosphatidylcholine
PCR	Polymerase chain reaction
PDB	Protein data bank
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
PGN	Peptidoglycan
PGP	Phosphatidylglycerol phosphate
PITP	Phosphatidylinositol transfer protein
Pgp	Phosphatidylglycerol phosphate phosphatase
Pgs	Phosphatidylglycerol phosphate synthase

PL	Phospholipid
pLDDT	Predicted local distance difference test
PMA	Phosphomolybdic acid
PME	Particle-Mesh Ewald
Pmf	Proton motive force
PMM	Plate middle mean
PNK	T4 polynucleotide kinase
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPE	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
POPG	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol
POTRA	Polypeptide-transport-associated
Pqi	Paraquat inducible
PS	Phosphatidylserine
Psd	Phosphatidylserine decarboxylase
Pss	Phosphatidylserine synthase
RBG	Repeating β-groove
Rhod	Rhodamine
RT	Room temperature
SAXS	Small angle X-ray scattering
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Sec	Secretion
SEC	Size exclusion chromatography
SEM	Standard error of the mean
SMA	Styrene maleic acid co-polymer
SMP	Synaptotagmin-like mitochondrial-lipid-binding domain
SOC	Super optimal broth with catabolite repression
SPA	Single particle analysis
SRP	Signal recognition particle
STARD	Steroidogenic acute regulatory protein-related lipid transfer domain
STAS	Sulphate transporter and anti-sigma factor antagonist

SUV	Small unilamellar vesicle
T4L	T4 ligase
TAE	Tris-acetate-EDTA
Tam	Translocation and assembly module
Tat	Twin arginine translocation
TB	Terrific broth
TBST	Tris-buffered saline with Tween 20
TCA	Tricarboxylic acid
TCEP	Tris-(2-carboxyethyl)phosphine
TEM	Transmission electron microscopy
TEMED	N,N,N',N'-tetramethylethylenediamine
TEV	Tobacco etch virus
TLC	Thin layer chromatography
TM	Transmembrane
T_m	Melting temperature
TULIP	Tubular lipid binding protein
UDP	Uridine diphosphate
UV	Ultraviolet
WT	Wild-type
β-OG	n-Octyl β-D-glucopyranoside

CHAPTER 1. INTRODUCTION

1.1 Gram-negative bacteria and the importance of the cell envelope

1.1.1 Antimicrobial resistance

The overconsumption and misuse of antibiotics worldwide has led to the enormous global public health challenge of antimicrobial resistance (AMR), where bacteria are developing resistance to multiple drugs, resulting in routine infections becoming deadly. Government reports estimate that if immediate action is not taken to combat AMR, it could result in the deaths of 10 million people annually by 2050 (O'Neill, 2016). The proposed strategies to combat this threat include reducing the use of antimicrobials in agriculture and medicine, increasing public awareness, and developing new antimicrobials. Furthermore, the discovery of novel antibiotics has drastically slowed since the 1980s, with research and development of these drugs not seen as a commercially viable pipeline by the pharmaceutical industry (Theuretzbacher et al., 2020).

Gram-negative (diderm) bacteria are more resistant to commercially available antibiotics compared to Gram-positive bacteria, with 9 out of 15 bacteria on the World Health Organisation priority list of the most important-to-tackle multidrug-resistant bacteria being Gram-negative species (WHO, 2024). Therefore, understanding the fundamental biology of these organisms is essential for the identification of novel therapeutic targets and enabling the next generation of antimicrobials.

1.1.2 The Gram-negative cell envelope

One of the reasons Gram-negative bacteria pose such a problem is due to the complex architecture of their cell envelope (Silhavy et al., 2010). This barrier consists of an inner

membrane (IM), a periplasm containing the thin peptidoglycan cell wall, and an additional outer membrane (OM) not present in Gram-positive bacteria (Figure 1.1).

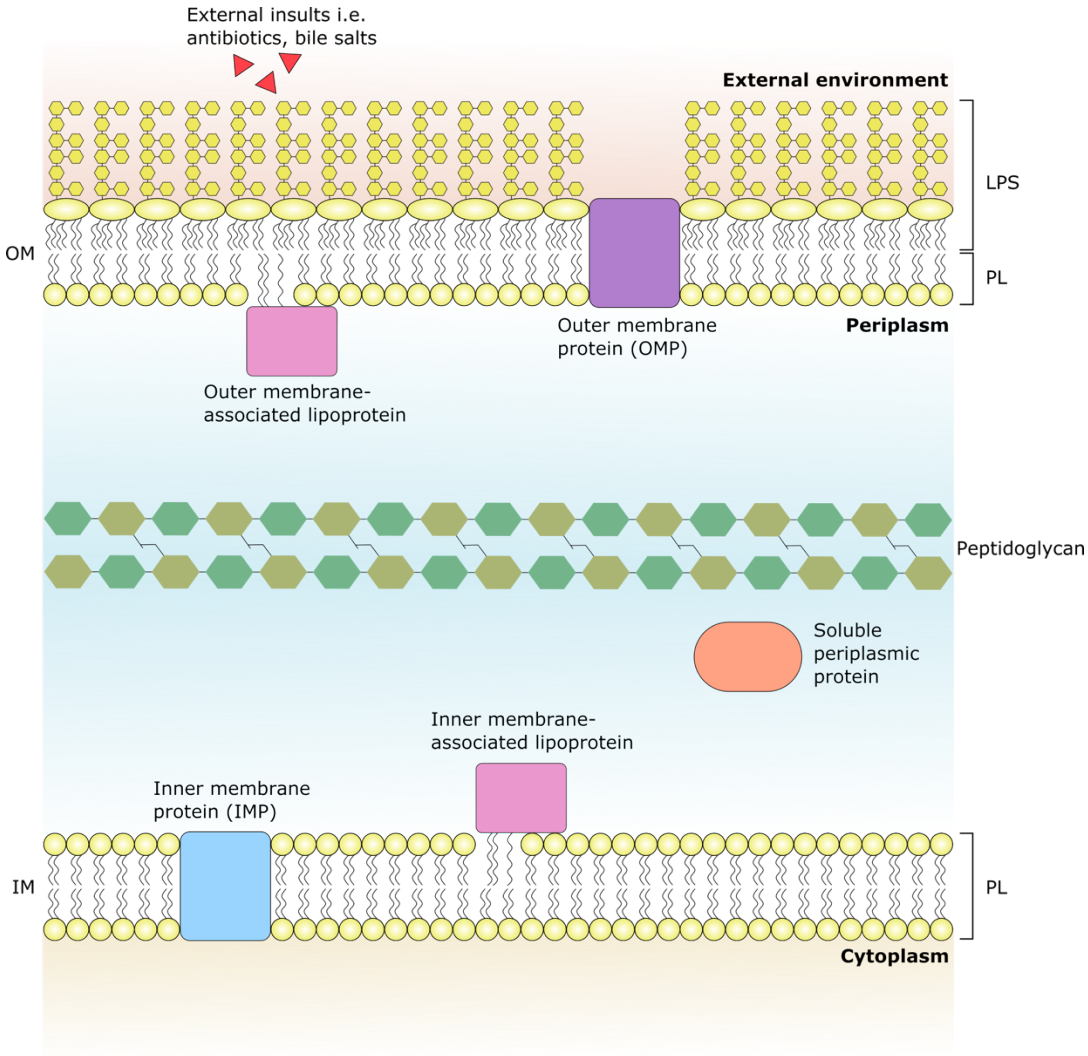


Figure 1.1. The Gram-negative cell envelope. The double membrane architecture of Gram-negative bacterial cell envelopes serves as a formidable barrier against external insults such as antibiotics and bile salts. The inner membrane (IM) is a symmetrical phospholipid (PL) bilayer containing inner membrane proteins (IMP) and lipoproteins. The asymmetric outer membrane (OM) consists of an inner leaflet of PL and an outer leaflet of lipopolysaccharide (LPS). The aqueous periplasm separates the inner and outer membranes and contains the peptidoglycan cell wall and soluble periplasmic proteins.

The IM consists of a symmetric glycerophospholipid (GPL) bilayer, whereas the outer membrane has an asymmetric distribution of lipids, with an inner leaflet of GPLs and an outer

leaflet of lipopolysaccharides (LPS). The LPS molecules are tightly packed and bridged by divalent cations, giving further structural integrity to the LPS layer (Silhavy et al., 2010). The OM is a formidable barrier that is impermeable to large and charged molecules due to the hydrophobic lipid bilayer, but also to hydrophobic molecules due to the dense LPS layer (Nikaido, 2003). The ~200 Å wide aqueous periplasm houses the peptidoglycan layer which gives the cell structural integrity, allowing it to withstand significant turgor pressure (Vollmer et al., 2008). The cell envelope also harbours proteins involved in cell envelope biogenesis and maintenance, nutrient and waste influx and efflux, and cell defence, including inner membrane proteins (IMPs), outer membrane proteins (OMPs), periplasmic proteins and lipoproteins tethered to either membrane (Silhavy et al., 2010).

The synthesis and upkeep of this complex envelope are essential for the cell to survive and adapt to various environmental conditions. The varied processes involved in its biogenesis and maintenance are therefore a promising target for novel drug discovery (Riu et al., 2022). Novel cell envelope targeting compounds could render the unique barrier ineffective, killing the bacteria or making the bacteria more susceptible to existing antibiotics. Despite decades of research, the complex and multifaceted processes behind cell envelope biogenesis are still relatively poorly understood (Silhavy et al., 2010). Elucidating these processes is therefore necessary to identify novel targets for antimicrobial drug development and to improve our understanding of fundamental process within Gram-negative bacteria.

1.2 Biogenesis and transport of cell envelope components

Biogenesis of cell envelope components occurs at the cytoplasm and IM where cellular forms of energy and precursors can be utilised (Silhavy et al., 2010). Once synthesised, these amphipathic biomolecules must be transported across the hydrophobic lipid bilayers and the

aqueous periplasm. Adenosine triphosphate (ATP) and the proton motive force (pmf) are limited to processes at the IM, leading to questions about how components are transported across and into the OM (Duong et al., 1997). The biogenesis and transport of components must also be tightly regulated to maintain cell envelope homeostasis (Saha et al., 2021; Hummels et al., 2023). The next sections will introduce the current understanding of the synthesis and transport of the most abundant cell envelope components, including proteins, the peptidoglycan layer and lipids.

1.2.1 Insertion and transport of cell envelope proteins

1.2.1.1 Inner membrane proteins and periplasmic proteins

Inner membrane proteins (IMPs) play important roles in nutrient transport, protein translocation, lipid biosynthesis, signal transduction and energy production (Duong et al., 1997). IMPs are typically α-helical in structure and can have cytoplasmic and periplasmic domains. After synthesis, the nascent polypeptide chain is inserted into the IM via the Sec translocon or through the IMP YidC (Figure 1.2). Targeting to Sec occurs either co-translationally via the signal recognition particle (SRP), or post-translationally via SecB (Figure 1.2). Normally, SRP-dependent transport is used for IMPs and involves elongation arrest at the ribosome before targeting to the SecYEG channel (Steinberg et al., 2020). SecB-dependent translocation is normally used for secretory proteins and involves chaperoning the unfolded polypeptide to SecYEG. Both SecA ATPase activity and the pmf drive the translocation of proteins via SecYEG (Allen et al., 2020). The transmembrane (TM) region of the unfolded polypeptide is threaded into the SecYEG channel and moved laterally into the phospholipid (PL) bilayer, stitching into the membrane (Silhavy et al., 2010). The signal peptide of periplasmic proteins is then cleaved for release into the periplasm, where molecular chaperones aid protein folding (Kaushik et al., 2022).

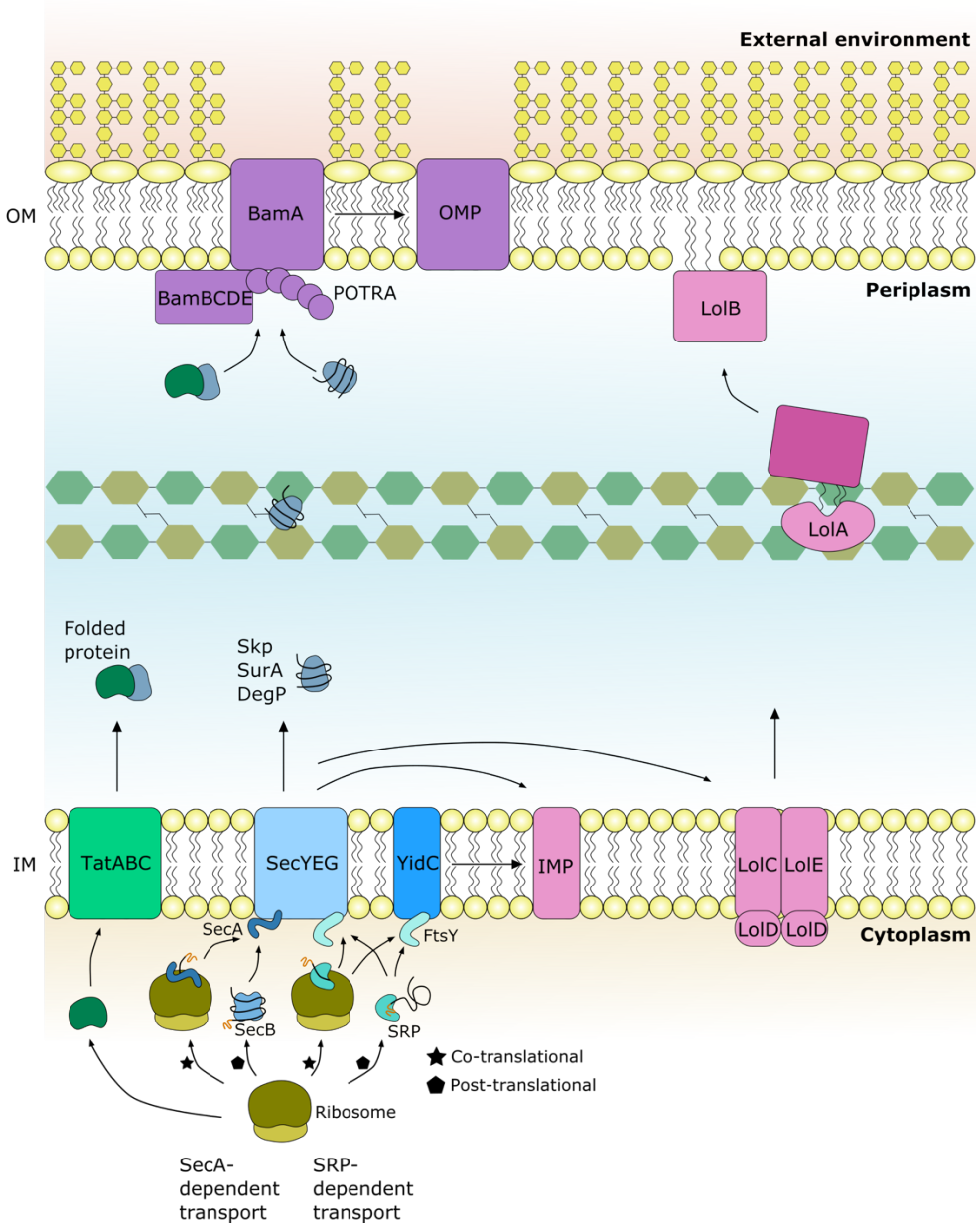


Figure 1.2. IMP, OMP and lipoprotein transport across the cell envelope. Proteins are inserted into or translocated across the inner membrane (IM) by the Sec machinery or the Tat pathway. Unfolded IM proteins go through SRP-dependent transport for insertion into the IM through the SecYEG translocon and/or YidC. Proteins for export into the periplasm go through the Sec translocon via Sec-dependent transport for unfolded proteins or through the Tat pathway for folded proteins. Skp, SurA or DegP are chaperones that deliver unfolded outer membrane (OM) proteins across the periplasm to the Bam complex for insertion into the OM (Otswald, 2021). Lipoproteins are transported across the IM by the ABC transporter LolCDE, passed to the molecular chaperone LolA and inserted into the OM by the lipoprotein LolB (Okuda and Tokuda, 2011).

1.2.1.2 Outer membrane proteins

The OM harbours many important proteins needed for nutrient uptake, efflux and cell defence (Duong et al., 1997). OMPs are primarily β-barrel proteins, embedded across both leaflets of the OM, and can be adorned with periplasmic or extracellular domains. These often large, amphipathic proteins need to be transported across the IM, across the periplasm and finally inserted into the OM. To do this, Sec-dependent transport is primarily used to pass unfolded OMPs across the IM via recognition of an amino leader sequence (Figure 1.2). Once in the periplasm, molecular chaperones such as Skp, SurA and DegP carry the unfolded protein to the OM. Additionally, the twin-arginine translocation (Tat) pathway passes folded proteins from the cytoplasm across the IM via TatABC (Alcock et al., 2016). Proteins are targeted to the Tat pathway by two arginine residues in the N-terminal signal peptide. The signal peptide is cleaved after translocation and the folded protein is delivered to the OM. It is suggested that the Sec translocon and Tat work co-operatively to transport OMPs across the IM (Oswald et al., 2021).

At the OM, the β-barrel assembly machinery (BAM) receives the protein, carried by periplasmic chaperones such as SurA, for subsequent folding and insertion via recognition of a C-terminal β-signal consensus sequence (Figure 1.2) (Kaur et al., 2021; Fenn et al., 2024). The Bam complex is made up of the core BamA and its 5 polypeptide-transport associated (POTRA) domains, and 4 other lipoproteins BamBCDE (Malinverni et al., 2006). An elegant study elucidating intermediate states of OMP folding using non-functional folding mutants revealed the formation of a hybrid-barrel of BamA and the OMP, before barrel formation of the OMP and insertion into the OM, potentially driven by membrane tension (Doyle et al., 2022).

In addition to the Bam complex, the translocation and assembly module (TAM) complex has been shown to secrete OMPs, such as the adhesin Ag43 in *E. coli*, and can fold OMPs *in vitro*

(Selkrig et al., 2012; Wang et al., 2024a). The Tam complex is made up of the IM protein TamB that is predicted to span across the periplasm to the OMP, TamA. TamA shares structural homology with BamA as they are both Omp85 proteins with multiple POTRA domains, and a TamB-BamA interaction has been seen in some organisms (Heinz et al., 2015). Currently, the function of TamB remains unclear with evidence suggesting possible roles in transporting and folding OMPs, transporting PLs or perhaps playing a regulatory role for the proper function of TamA (Heinz et al., 2015; Josts et al., 2017; Ruiz et al., 2021). TamB is a member of the *E. coli* AsmA-like family, and its putative role in PL transport will be further discussed.

1.2.1.3 Lipoproteins

As well as IMPs and OMPs, the cell envelope harbours lipoproteins that can sit in the periplasmic-facing leaflet of the IM and both leaflets of the OM. Lipoproteins serve various functions in *E. coli*, with the most abundant lipoprotein, Lpp (Braun's lipoprotein), necessary for maintaining structural integrity by linking the OM and the peptidoglycan cell wall (Cohen et al., 2017). Lipoproteins are synthesised in the cytoplasm and translocated across the IM via the Sec machinery under instruction of a lipobox sequence: L-[A/S/T]-[G/A]-C (Zückert, 2014). After cleavage of the signal peptide, the lipobox cysteine is acylated to allow attachment to the membrane. Lipoproteins destined for the OM transit the periplasm via the localisation of lipoproteins (Lol) pathway, whereas lipoproteins with a Lol avoidance signal (an aspartate residue at the +2 position post-processing) remain in the IM (Masuda et al., 2002).

The Lol pathway involves an ABC transporter, LolCDE, at the IM which utilises ATP hydrolysis to shuttle the lipoprotein onto a periplasmic chaperone, LolA (Figure 1.2). The LolA-lipoprotein complex then travels across the periplasm to the OM lipoprotein LolB for insertion

into the inner leaflet of the OM (Okuda and Tokuda, 2011). How the lipoprotein enters the lipid bilayer from LolB without an energy source is not yet understood.

1.2.2 Biogenesis of the peptidoglycan cell wall

The peptidoglycan cell wall is responsible for determining cell shape and providing rigidity against osmotic pressure (Vollmer et al., 2008). Its mesh-like structure excludes large particles from passing through the cell envelope. The cell wall consists of alternating N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM) polymers that are cross-linked by short peptides. Initial synthesis of peptidoglycan precursors occurs in the cytoplasm, converting fructose-6-phosphate to UDP-NAM via multiple intermediary steps (Garde et al., 2021). The pentapeptide [L-Ala]-[D-Glu]-[mDAP]-[D-Ala]-[D-Ala] is then added before attachment to the IM carrier undecaprenyl pyrophosphate, resulting in Lipid I (Figure 1.3). Lipid II is then formed through the addition of UDP-NAG before being flipped across the IM by MurJ (Kuk et al., 2022). Polymerisation of these building blocks then occurs in the periplasm through transglycosylation between the sugars and transpeptidation between the short peptides of two NAM units (Vollmer et al., 2008).

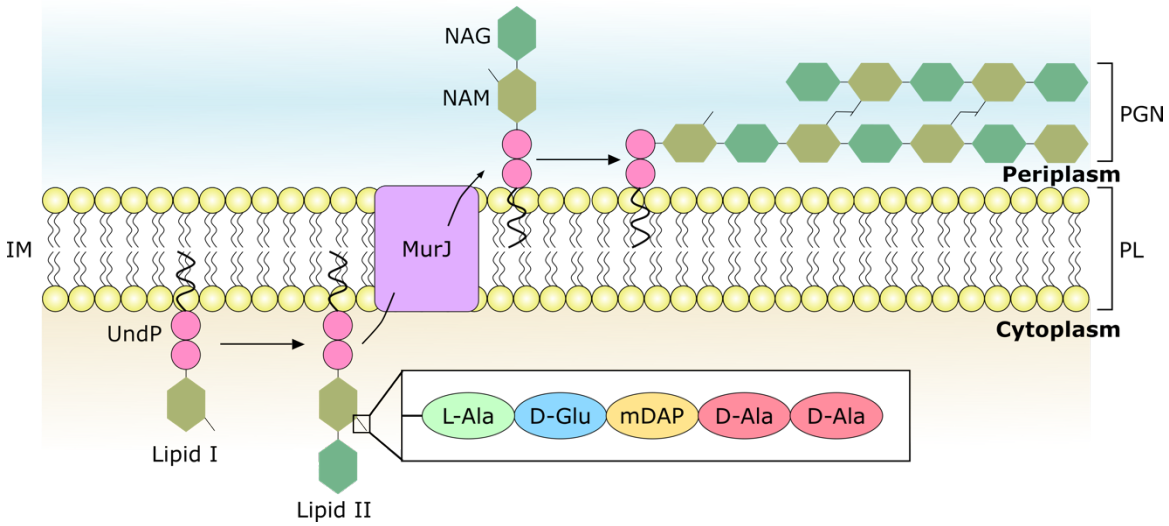


Figure 1.3. Biogenesis of the peptidoglycan cell wall. The peptidoglycan (PGN) cell wall is made from polymers of alternating N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM), cross-linked by short peptides. A pentapeptide-linked NAM attached to the IM by undecaprenyl phosphate (UndP) called Lipid I is converted to Lipid II by the addition of NAG. MurJ then flips Lipid II to the periplasmic facing leaflet where polymerisation occurs through transglycosylation of the sugars and transpeptidation of the short peptides between two NAM sugars (Garde et al., 2021).

1.2.3 Biogenesis and transport of lipopolysaccharide

The lipopolysaccharide (LPS) outer leaflet in the Gram-negative OM is an important barrier against hydrophobic compounds that would otherwise pass through a phospholipid bilayer. There are three main components of LPS: lipid A, the core sugars and O-antigen (Figure 1.4, inset). Lipid A sits in the membrane and is followed by the core sugars, with a conserved inner core of 3-Deoxy-D-manno-oct-2-ulosonic acid (KDO) and heptose and an outer core of more varied sugar groups. The LPS structure is complete with O-antigen, which has varying lengths of a repeating oligosaccharide unit (Liu et al., 2019).

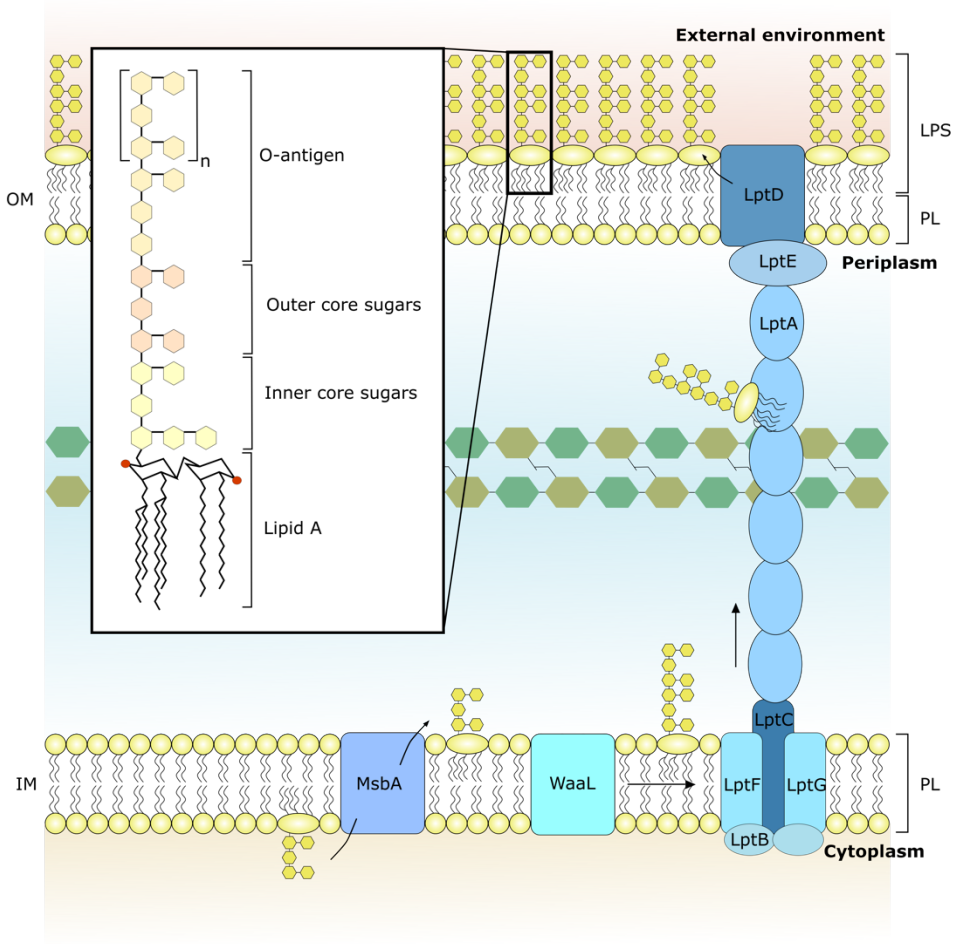


Figure 1.4. Synthesis and transport of lipopolysaccharide (LPS). The structure of LPS is shown in the inset, with lipid A connected to the inner and outer core sugars, followed by a varied number of repeating oligosaccharide units in the O-antigen. Lipid A is synthesised in the cytoplasm and flopped by MsbA to the outer leaflet of the IM, where WaaL ligates the O-antigen and lipid A components. The Lpt system then transports the full LPS structure through the inner membrane $LptB_2CFG$ component, across repeating units of LptA, to the outer membrane components LptDE where it is inserted into the outer leaflet of the OM.

The synthesis of LPS starts in the cytoplasm with the formation of lipid A, the biosynthesis of which has been well-characterised and is the molecule typically associated with Gram-negative septic shock (Trent, 2004). Lipid A is then converted to 'rough LPS' through the addition of the core sugars. MsbA then flops (inner leaflet to outer leaflet) the rough LPS molecule to the periplasmic-facing leaflet of the IM (Figure 1.4). O-antigen repeat units are carried to the periplasmic facing leaflet by undecaprenyl diphosphate, before WaaL-mediated ligation onto

the rough LPS (Heinrichs et al., 1998). The now 'smooth LPS' molecule is transported to the OM outer leaflet via the transenvelope lipopolysaccharide transport (Lpt) machinery (Figure 1.4). $LptB_2CFG$ is an IM ATP-binding cassette (ABC) transporter that provides the energy for LPS movement from the IM across the periplasm, against a concentration gradient, via a bridge of LptA monomers (Sherman et al., 2018). Lpt functions in a 'PEZ dispenser'-like manner, pushing one LPS molecule after another through the bridge, before insertion into the outer leaflet of the OM via LptD and LptE (Okuda et al., 2016).

Many of the Lpt proteins share a structure resembling a slightly twisted β-jellyroll fold (Suits et al., 2008). This fold creates interior pockets capable of protecting the acyl chains of lipid A from the aqueous periplasm during transport. The Lpt machinery is essential, with the deletion of any component causing an accumulation of LPS in the IM periplasmic leaflet (Sperandeo et al., 2008).

1.2.4 Biogenesis and transport of phospholipids

1.2.4.1 Biosynthesis of phospholipids

In *E. coli*, phospholipid membranes are typically composed of 75% phosphatidylethanolamine (PE), 20% phosphatidylglycerol (PG) and 5% cardiolipin (CL) (Figure 1.5A) (Raetz and Dowhan, 1990). Phospholipid biogenesis in *E. coli* is well-characterised, with biosynthetic pathways being determined as early as the 1960s (Kanfer and Kennedy, 1964). Phosphatidic acid (PA) precursors are synthesised through two acylation steps of glycerol-3-phosphate, before being modified with cytidine-triphosphate to form cytidine-diphosphate diacylglycerol (CDP-DAG) by the enzyme CDP-DAG synthase (CdsA) (Figure 1.5B) (Parsons and Rock, 2013; Rowlett et al., 2017). At this point, headgroup variation is introduced. Conversion of CDP-DAG by phosphatidylserine synthase (PssA) forms the short-lived phospholipid

phosphatidylserine (PS), which is decarboxylated to PE by PS decarboxylase (Psd). Alternatively, CDP-DAG is modified with glycerol-3-phosphate to form phosphatidylglycerol phosphate (PGP) via PGP synthase (PgsA). The dephosphorylation of PGP by PGP phosphatases (PgpABC) results in PG. CL synthases then form CL through the condensation of two PG molecules (ClsAB) or the condensation of PG and PE (ClsC) (Li et al., 2016; Rowlett et al., 2017).

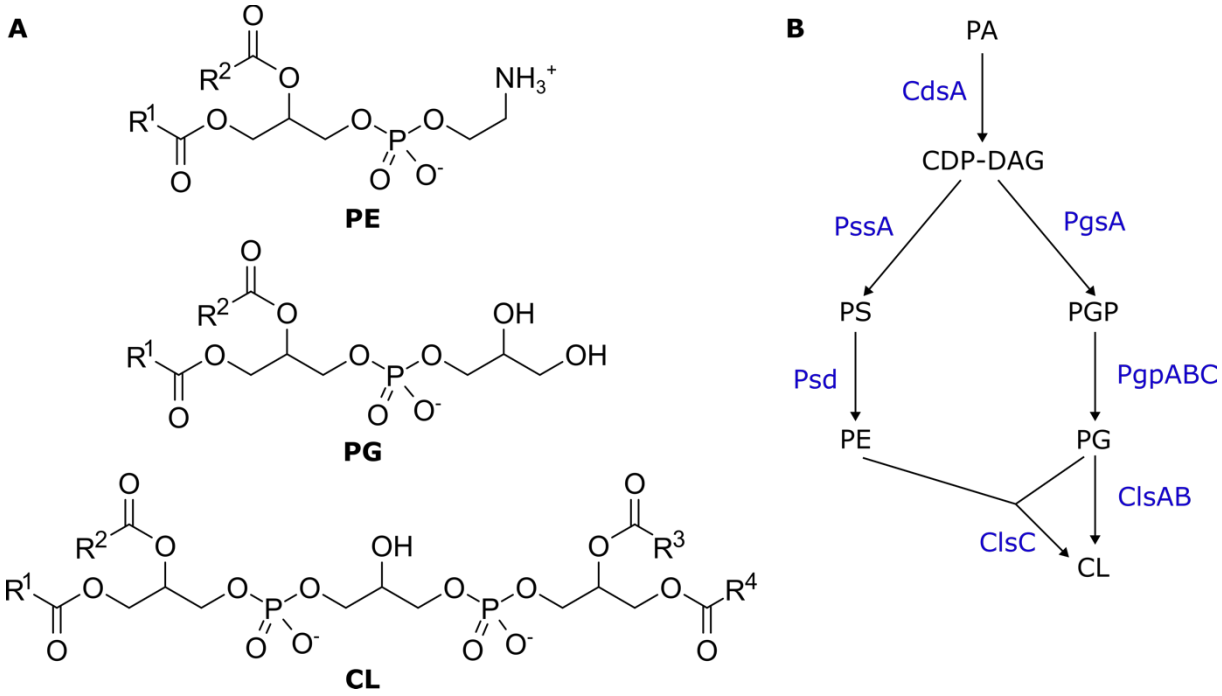


Figure 1.5. *E. coli* phospholipid composition and biosynthesis. A – Chemical structures of the main glycerophospholipid species found in *E. coli*: phosphatidylethanolamine (PE), phosphatidylglycerol (PG) and cardiolipin (CL). B – Phospholipid synthesis from the precursor phosphatidic acid (PA). Cardiolipin synthase A (CdsA) converts PA to cytidine-diphosphate diacylglycerol (CDP-DAG). Phosphatidylserine synthase A (PssA) converts CDP-DAG into phosphatidylserine (PS). PS decarboxylase (Psd) converts PS to PE. CDP-DAG can also be converted to PG phosphate (PGP) via PGP synthase. PGP phosphatases (PgpABC) convert PGP to PG. CL is formed from either two PG molecules via CL synthases (ClsAB) or one PG and one PE molecule via ClsC. R groups represent acyl chains of varying length. Figures adapted from (Rowlett et al., 2017; Cooper, 2022).

1.2.4.2 Phospholipid translocation

Phospholipids are synthesised at the inner leaflet of the IM and must be translocated to the periplasmic-facing leaflet of the IM to ensure equilibration of the symmetric bilayer. Phospholipids in this leaflet are also used to modify lipid A with phosphoethanolamine moieties, to acylate lipoproteins, and as a reservoir for OM biogenesis (Roney and Rudner, 2023). Phospholipid translocation involves the movement of the hydrophilic headgroup across the hydrophobic membrane core, an energetically unfavourable process (Guo et al., 2021). Phospholipid flipping is rapid in *in vivo* studies, but these high rates have not been replicated in *in vitro* experiments investigating spontaneous phospholipid flip-flop, suggesting translocation is facilitated by membrane proteins (Bai and Pagano, 1997).

Despite decades of research into phospholipid movement, only two bacterial phospholipid flippases have been identified, the *S. aureus* MprF that flips the minor Gram-positive phospholipid species lysyl-phosphatidylglycerol and the *B. subtilis* PetA that flips PE (Ernst et al., 2015; Roney and Rudner, 2023). PetA is a member of the DedA superfamily, ubiquitous across many Gram-positive and Gram-negative bacteria. Members of this family are often found in operons next to phospholipid synthases, and bioinformatic analyses of DedA domains suggest they may be capable of transporting diverse phospholipids (Roney and Rudner, 2023). Alternatively, the lipid A floppase MsbA has been implicated in also translocating phospholipids across the cell membrane, with a temperature-sensitive mutant of MsbA resulting in a 90% reduction of the export of all major OM lipids, and *in vitro* assays showing that MsbA flops PE in an ATP- and ΔpH gradient-dependent manner (Doerrler et al., 2001; Guo et al., 2021). Additionally, some studies suggest the presence of integral membrane proteins alone may promote phospholipid flopping (Kol et al., 2003).

1.2.4.3 Phospholipid transport

Defined transport pathways of phospholipids between the inner and outer membrane of Gram-negative bacteria are yet to be determined (Shrivastava and Chng, 2019). Early studies showed that phospholipid transport between membranes of the cell envelope was bidirectional and likely diffusive (Jones and Osborn, 1977). Later studies on radiolabelled phospholipids provided further evidence of bidirectional phospholipid movement, with both anterograde (IM to OM) and retrograde (OM to IM) transport seen (Donohue-Rolfe and Schaechter, 1980; Langley et al., 1982).

Phospholipid transport across the aqueous periplasm is proposed to be facilitated either by proteins that shield the acyl chains during transport or by the direct exchange of phospholipids at sites of contact between the two membranes (Figure 1.6). Initially, regions of membrane adhesion called Bayer junctions were identified by electron microscopy (EM) and proposed to facilitate the transfer of phospholipids between membranes (Bayer, 1968). However, this has since been disputed as an artefact of EM sample preparation (Kellenberger, 1990). Vesicle budding is another possible method of phospholipid transport; although no evidence of this mechanism has been seen, and the size limits of the peptidoglycan layer and the periplasm make this mechanism unlikely (Shrivastava and Chng, 2019). It is more likely that protein-mediated transport is used. Here, soluble periplasmic chaperones or trans-envelope proteinaceous bridges protect the acyl chains of phospholipids as they traverse the periplasm, akin to LolA in lipoprotein transport and the Lpt machinery in LPS transport (Shrivastava and Chng, 2019).

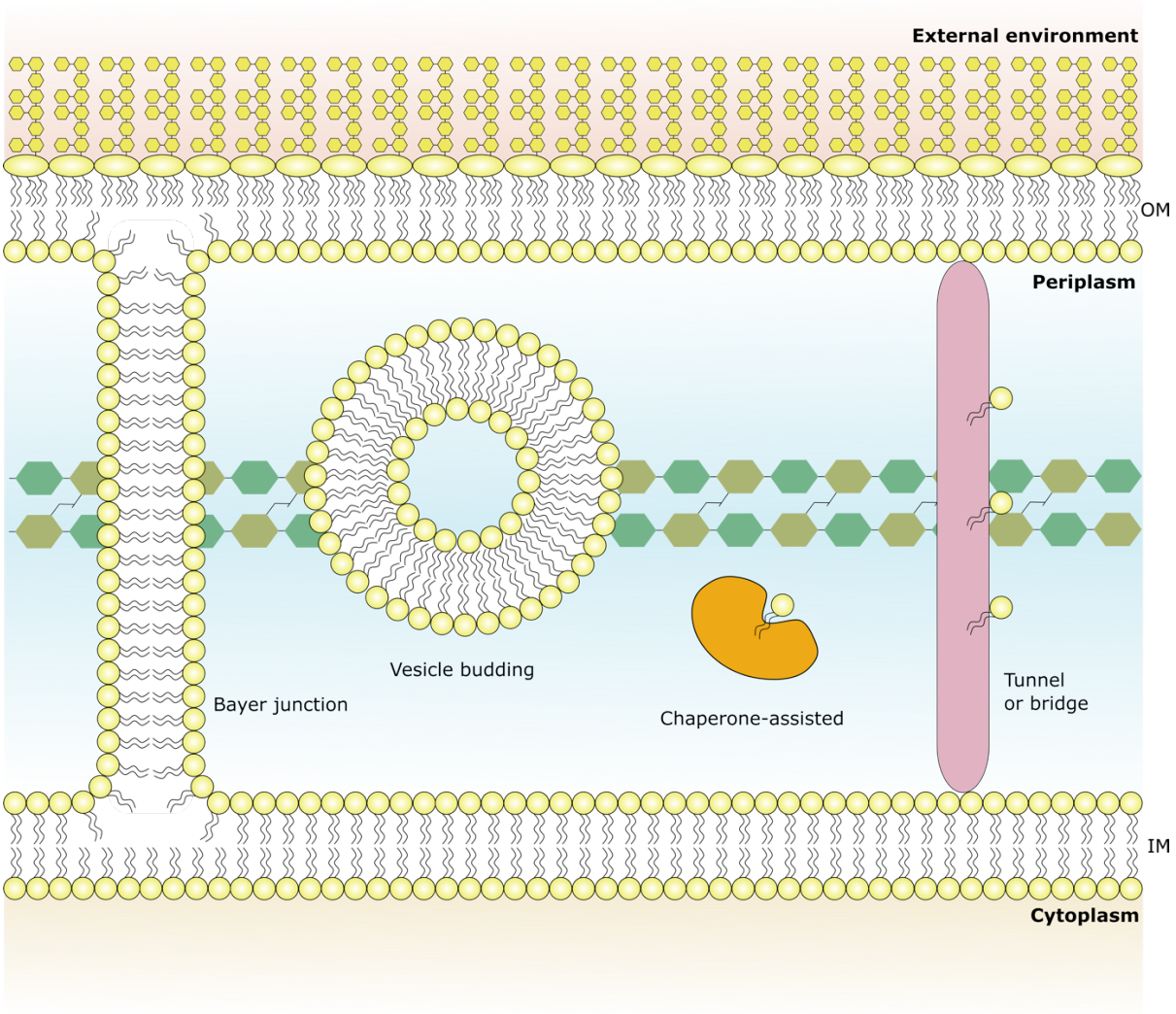


Figure 1.6. Proposed mechanisms of phospholipid transport in the cell envelope. Transport of phospholipids in the cell envelope of Gram-negative bacteria is proposed to occur via sites of membrane adhesions, called Bayer junctions, vesicle budding between membranes, or protein-mediated transport such as soluble chaperones and protein tunnel/bridges. Figure adapted from (Shrivastava and Chng, 2019).

The discovery of the maintenance of lipid asymmetry (Mla) pathway in *E. coli* provided the first evidence of a proteinaceous system involved in phospholipid transport (Malinverni and Silhavy, 2009). Since then, other pathways and proteins have been proposed, including the Pqi pathway, the LetAB pathway, the Tol-Pal complex and the AsmA-like proteins (Figure 1.7) (Yeow and Chng, 2022). The following sections will introduce these pathways and the current evidence towards their role as phospholipid transport systems in *E. coli*.

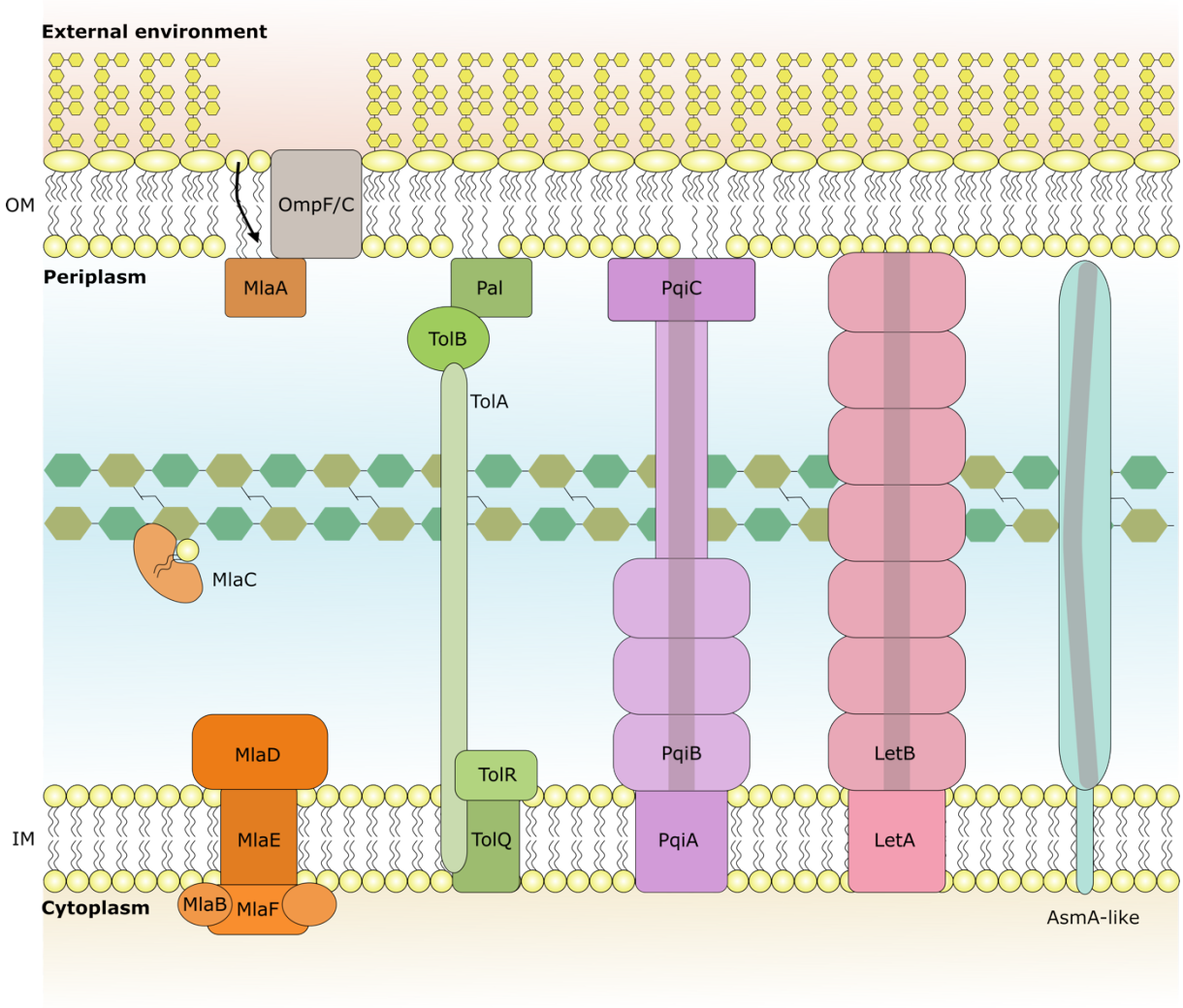


Figure 1.7. Putative phospholipid transport systems in *E. coli*. The Mla pathway returns mislocalised phospholipids in the outer leaflet of the OM by the MlaA-OmpF/C complex to the IM MlaFEDB complex via a soluble phospholipid carrier MlaC. The Tol-pal complex is proposed to also be involved in retrograde transport. The PqiABC, LetAB and AsmA-like proteins are trans-envelope complexes or protein bridges proposed to be involved in anterograde phospholipid transport by acting as a conduit for phospholipids across the periplasm.

1.2.4.4 The Mla pathway

A loss of OM lipid asymmetry can occur under certain stress conditions and during growth and division of the cell, with mislocalised phospholipids in the outer leaflet of the OM compromising the barrier function of the OM. In response, these phospholipids are either

degraded by the OM proteins PldA and PagP, or returned to the IM through the Mla pathway (Malinverni and Silhavy, 2009). PldA is a phospholipase that deacylates phospholipids, and PagP is a palmitoyl transferase that transfers a palmitate moiety from phospholipids onto lipid A molecules to form hepta-acylated LPS, which is more resistant to host defences (Guo et al., 1998; Bishop et al., 2000; Dekker, 2000). Alternatively, the Mla pathway transports these mislocalised phospholipids across the periplasm, back to the IM (Malinverni and Silhavy, 2009).

The Mla system is composed of an IM ABC transporter MlaFEDB, a periplasmic chaperone MlaC, and an OM lipoprotein MlaA, associated with the major OM porins OmpF/C (Figure 1.8). Original evidence of its involvement in OM maintenance came from the sensitivity of *mla* knockout mutants to the OM stressors SDS and EDTA (Malinverni and Silhavy, 2009). Suppressor mutations repairing this OM permeability defect were found in PldA, with increased expression levels found to compensate for the higher amounts of outer leaflet phospholipids. Additionally, PagP activity increased in a Δ*mla*CΔ*pldA* background, furthering the evidence towards a role in recovering the lipid asymmetry of the OM.

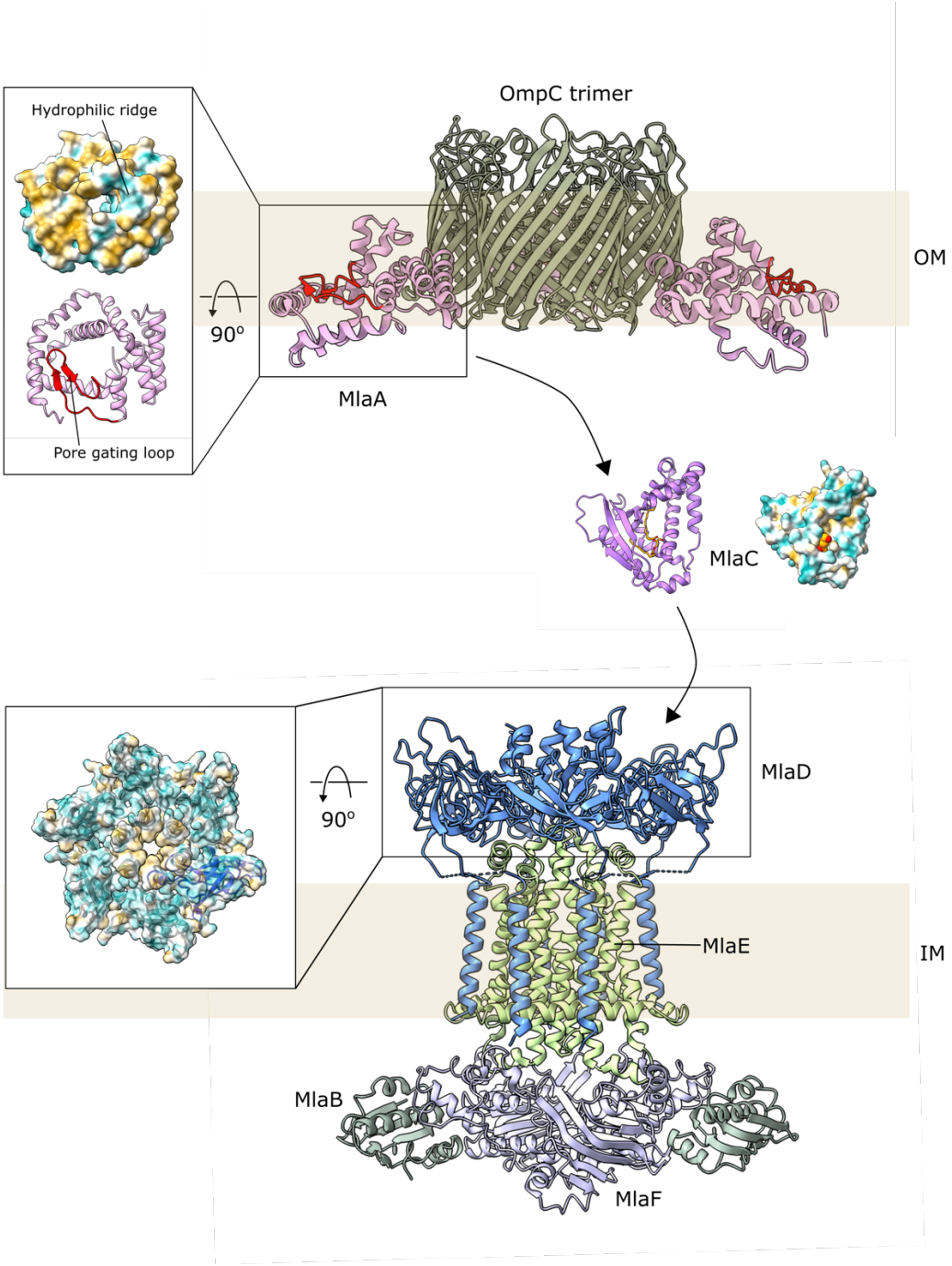


Figure 1.8. Structures of the Mla pathway components. The OM lipoprotein MlaA (pink) associates with trimeric OmpC (brown, PDB: 5NUP). The MlaA hydrophilic channel ('ridge') is highlighted (hydrophobic surface and cartoon), with the adjacent β-hairpin pore-gating loop coloured red. Phospholipid bound MlaC is shown (cartoon MlaC in purple, phospholipid in orange and hydrophobic surface, PDB: 5UWA), responsible for shuttling phospholipids from MlaA to the IM ABC transporter MlaFEDB (MlaD - blue, MlaE - lime green, MlaF– lilac, MlaB – dark green, PDB: 7CGE). The hexameric MCE ring of MlaD is highlighted (hydrophobic surface), showing the hydrophobic pore, with a single MlaD protomer in blue.

The OM lipoprotein MlaA and its association with OmpC trimers are responsible for removing the mislocalised phospholipid from the outer leaflet before transfer to the periplasmic MlaC (Chong et al., 2015). Structural characterisation of the complex revealed that MlaA embeds deep within the bilayer and comprises a 'donut'-shaped architecture with a vertical 'ridge' that forms a central hydrophilic channel (Figure 1.8) (Yeow et al., 2018). This channel likely provides an exit path for amphipathic phospholipids from the outer leaflet into the periplasmic MlaC. An adjacent β-hairpin loop motif is thought to be responsible for the gating of this channel. An interesting variant called MlaA*, containing deletions of two residues near this hairpin loop, is thought to allow the uncontrolled movement of inner leaflet phospholipids into the outer leaflet and exhibits increased drug sensitivity and early cell death during stationary phase (Sutterlin et al., 2016).

The phospholipid transfer between MlaA and MlaC is energetically favourable and has been replicated in *in vitro* studies (Yeow et al., 2023). MlaC is recruited to the OM through electrostatic interactions with the C-terminal helix of MlaA, before membrane deformation adjacent to the complex potentially facilitates phospholipid transfer. Phospholipids then bind with high affinity to MlaC, with the acyl chains residing in its deep hydrophobic pocket (Figure 1.8) (Ekiert et al., 2017). A pivoting β-sheet mechanism is thought to open and close the pocket upon interaction with MlaA or MlaFEDB, allowing either binding or release of the phospholipid (Huang et al., 2016; Hughes et al., 2019).

Once at the IM, lipid transfer from MlaC to the binding cavity of MlaFEDB occurs. MlaFEDB forms a stable IM complex in a 2:2:6:2 stoichiometry (Figure 1.8) (Thong et al., 2016; Ekiert et al., 2017). MlaF is an ATPase that is stabilised by the sulphate transporter and anti-sigma factor antagonist (STAS)-domain containing MlaB, both facing the cytoplasm (Kolich et al., 2020). MlaE is an IM permease that forms a hydrophobic pocket at the dimer interface capable

of binding phospholipids (Coudray et al., 2020). MlaD is a periplasmic-facing mammalian cell entry (MCE)-domain containing protein that interacts directly with MlaC for phospholipid retrieval (Isom et al., 2017; Wotherspoon et al., 2024). A pore within the centre of the hexameric assembly of MlaD MCE domains is formed from hydrophobic loops of each subunit, allowing passage of phospholipids from MlaC into the MlaE pocket via MlaD.

Structures of the full MlaFEDB complex reveal potential mechanisms of action, with an outward-facing "V"-shaped pocket at the MlaE dimer interface in the absence of nucleotide and a closed pocket at the dimer interface in the ATP-bound state (Chi et al., 2020). Structural plasticity in the positioning of the MlaD ring relative to the rest of the complex may also be important for phospholipid movement throughout the complex (Mann et al., 2021; Wotherspoon et al., 2024). Recently, investigations into the MlaC-MlaD complex suggest that phospholipid transfer occurs with MlaC binding the outward-state MlaFEDB complex, before ring deformation of MlaD causes conformational changes to MlaE, which in turn stimulates ATP-binding. Simultaneously, partial collapse of the MlaC pocket is thought to force phospholipids to enter the MlaD pore via tail-first entry of a single acyl chain, before further movement into the MlaE cavity upon ATP hydrolysis. The resulting closed MlaFEDB state may force phospholipids within its cavity out into the surrounding bilayer (Wotherspoon et al., 2024).

Some evidence for anterograde transport (IM to OM) has also been seen through the spontaneous phospholipid transfer from both MlaD and a reconstituted MlaFEDB complex to MlaC, though further research has revealed that ATP binding and hydrolysis abolishes transport in this direction (Hughes et al., 2019; Low et al., 2021; Tang et al., 2021). Exactly how phospholipids enter the IM and into which leaflet is not yet determined.

Interestingly, homologues of the Mla pathway in other bacterial species consist of bridge-like complexes, with MlaD-like MCE domain-containing proteins spanning the periplasm, suggesting retrograde transport of phospholipids could also occur through bridge-like complexes (Grasekamp et al., 2023; Chen et al., 2023). This could potentially allow high-flux lipid transport, as bridge-like structures would likely be more capable of transporting higher amounts of lipids compared to a protein chaperone. However, whether phospholipids are indeed the transported cargo of these homologues is unknown, as is the direction of transport.

1.2.4.5. The Tol-Pal complex

Evidence also exists for a role of the *E. coli* Tol-Pal complex in retrograde phospholipid transport to maintain OM lipid asymmetry (Shrivastava et al., 2017). The Tol-Pal pathway consists of a trans-envelope complex, with the IM TolQRA complex able to utilise the pmf to drive interactions with the periplasmic protein TolB and the OM lipoprotein Pal (Figure 1.7). Transduction of the potential energy of the pmf dissociates the TolB-Pal complex, allowing Pal to bind peptidoglycan, providing a direct link between the OM and the cell wall (Webby et al., 2022).

Some studies suggest that this complex is responsible for bulk retrograde phospholipid transport from the inner leaflet of the OM to the IM to control overall levels of phospholipid in the OM, lowering the likelihood of phospholipids flipping to the outer leaflet. *Tol-pal* mutants accumulate excess phospholipids in the OM, are sensitive to OM stressors and exhibit OM blebbing (Bernadac et al., 1998; Shrivastava et al., 2017; Tan and Chng, 2024). Additionally, overexpression of the Mla pathway components in this background can partially recover these effects, furthering the evidence towards its role in retrograde phospholipid transport.

However, no evidence has yet been found of phospholipid binding within any of the Tol-pal components and a potential mechanism of action is unknown. Additionally, the Tol-Pal complex has an important and defined role in cell division, with septal localisation of all its components and a requirement of the Tol-Pal complex for proper OM invagination (Gerding et al., 2007). Therefore, further investigation of this complex and its putative phospholipid transport role is required.

1.2.4.6 The Pqi pathway

The paraquat-inducible (Pqi) complex is another putative phospholipid transport pathway in *E. coli*. The Pqi complex spans the cell envelope, with an IM PqiA and an IM-tethered PqiB that spans the periplasm to the OM lipoprotein PqiC (Nakayama and Zhang-Akiyama, 2016). Like MlaD, PqiB contains an MCE domain. However, in PqiB, three copies of this domain are found, along with a long C-terminal helix. Structural investigations revealed a homohexameric assembly of PqiB, forming three stacked MCE rings and a hollow coiled coil needle-like structure (Figure 1.9) (Ekiert et al., 2017). The structure revealed a continuous hydrophobic tunnel throughout (~230 Å) that may provide a conduit for phospholipid transport across the periplasm (Ekiert et al., 2017). Recently, it was determined that PqiC exists as an octomeric toroid that stabilises the PqiB needle, with an interaction that is necessary for Pqi function (Cooper et al., 2024). There has been no experimental structure determined for the IM PqiA, but structural similarity of the AlphaFold predicted PqiA structure to that of the *Pyrococcus furiosus* multidrug and toxic compound extrusion (PfMATE) transporter may indicate a pmf-driven mechanism of action (Cooper, 2022).

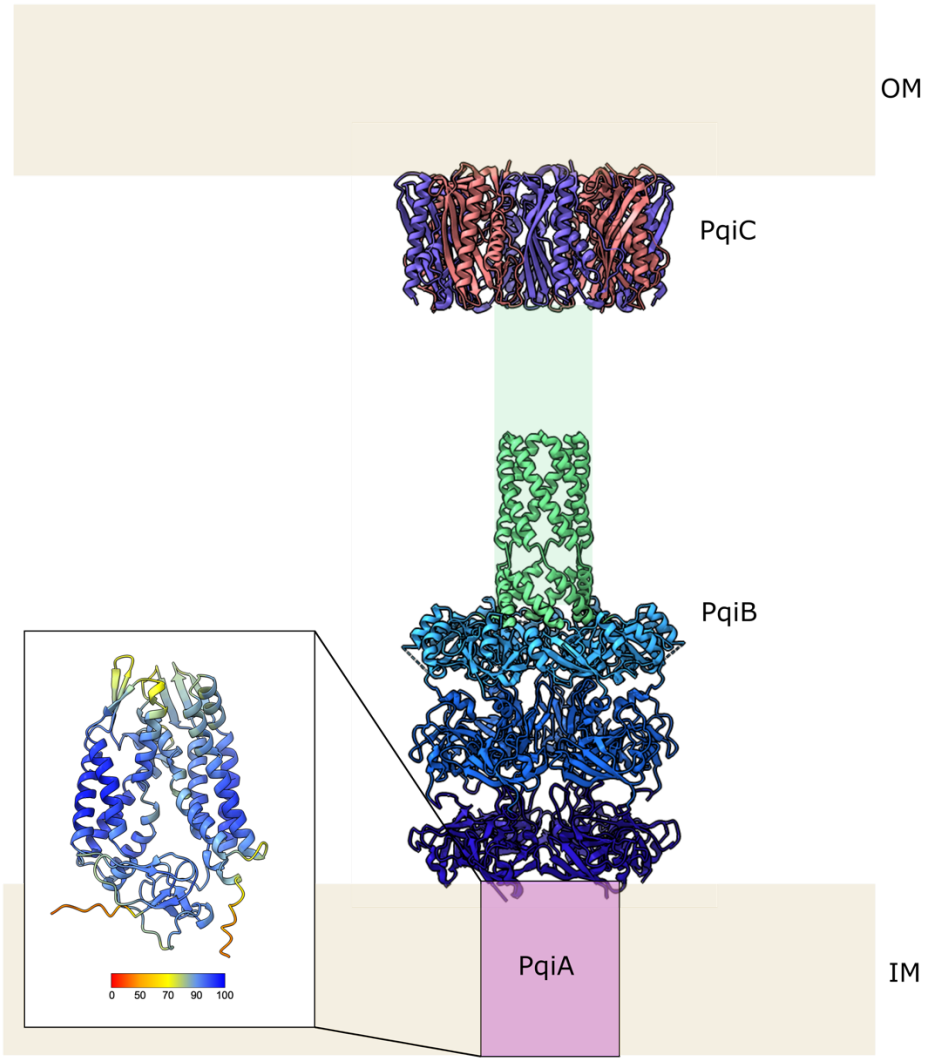


Figure 1.9. Structure of the PqiABC complex. The PqiABC complex putatively transports phospholipids across the cell envelope. The AlphaFold prediction of IM PqiA is shown coloured by pLDDT score (key shown, red to blue as confidence increases), along with the experimentally determined 3 MCE domain-containing PqiB (blue and green, PDB: 5UVN), that spans across the periplasm to the octomeric assembly of the OM lipoprotein PqiC (pink and purple, PDB: 8Q2C) (Jumper et al., 2021).

Further evidence for a role in phospholipid transport comes from the co-purification of Pqi components with phospholipids and by the increased sensitivity of *mla* mutants to SDS and EDTA with deletion of *pqiABC* and another MCE domain-containing complex *yebST* (Nakayama and Zhang-Akiyama, 2016; Cooper, 2022). However, deletion of *pqiABC* alone, or in the *mla* mutant background, was not enough to confer this OM permeability phenotype.

Sensitivity to a different compound, lauryl-sulfobetaine (LSB) was instead found, allowing complementation studies to investigate residues of important function (Isom et al., 2017). Using this sensitivity, a similar role was found for the PqiB MCE ring pore lining loops as that of MlaD, indicating the pore is important for substrate binding and overall complex function (Cooper, 2022). Nevertheless, details of the Pqi complex and its putative phospholipid transport role require further investigation.

1.2.4.7 The Let pathway

E. coli contains a third MCE domain-containing protein, LetB. Structural investigations of LetB revealed a homohexameric arrangement, with 7 stacked MCE rings (~230 Å) capable of spanning the periplasm, again providing a hydrophobic tunnel capable of shielding phospholipids (Figure 1.10) (Ekiert et al., 2017; Isom et al., 2020). LetA is the IM component of the LetAB (formally YebST) complex, and like PqiA, no experimental structures of this protein have been reported.

As previously stated, the deletion of *yebST* and *pqiABC* in the *mla* mutant background exacerbates the OM permeability phenotype, providing evidence for a similar role (Nakayama and Zhang-Akiyama, 2016). The number of MCE rings was found to be important for the function of LetAB, with *letB* mutants that had fewer than 5 MCE rings unable to complement a sensitivity phenotype of Δ*letAB* on LSB, likely due to being unable to span the periplasm. Interestingly, *letB* mutants lacking the third or sixth MCE ring had increased resistance to LSB than the wild-type (WT) protein, indicating that they have an important regulatory function in substrate transport (Vieni et al., 2022). Also, the MCE ring pore lining loops were found to be important for LetAB function, again suggesting a similar role to that of MlaD and PqiB (Ekiert

et al., 2017). Whilst current evidence suggests a role in phospholipid transport across the cell envelope, the exact substrate and transport directionality of LetAB are yet to be confirmed.

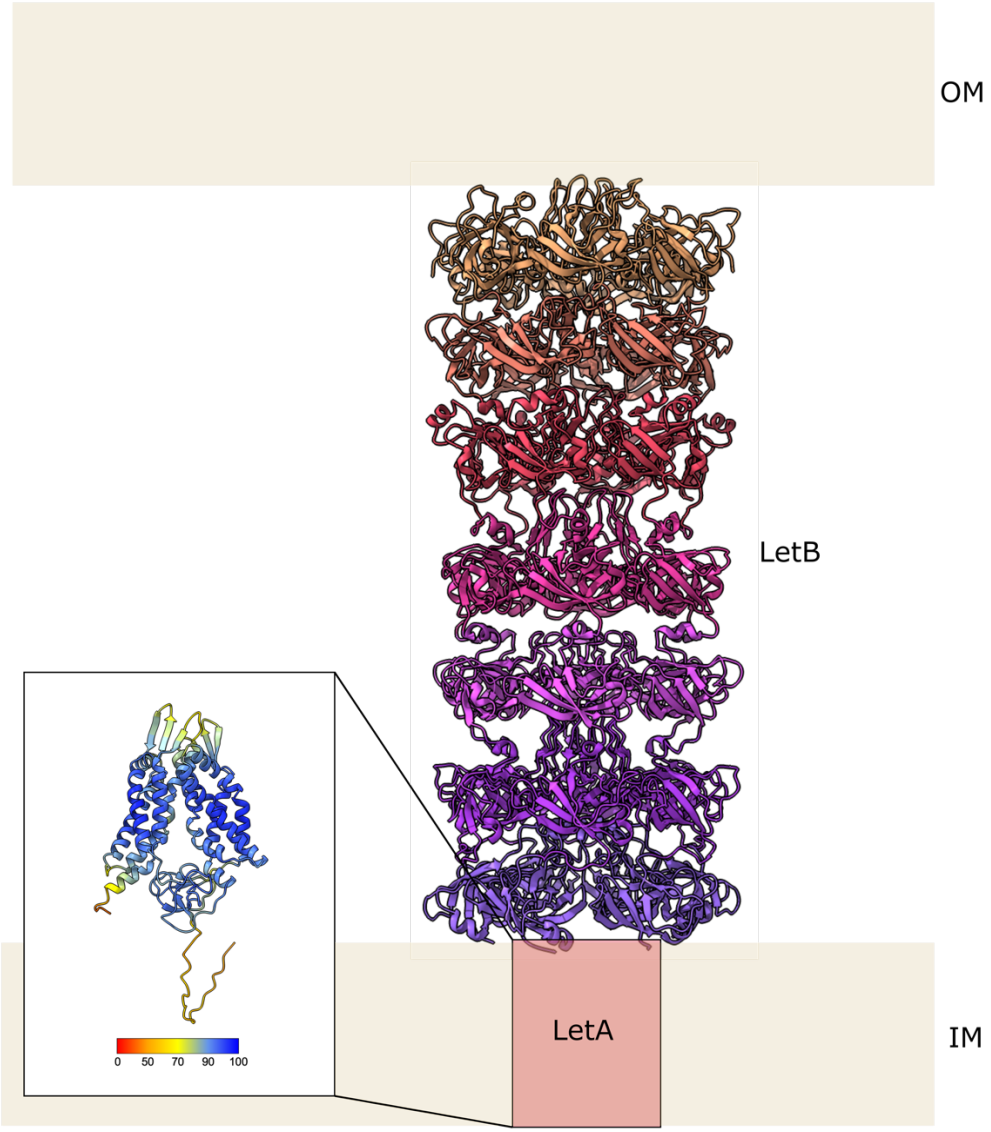


Figure 1.10. Structure of the LetAB complex. The LetAB complex putatively transports phospholipids across the cell envelope. The MCE domain-containing LetB cryo-EM structure is shown, coloured by MCE ring (PDB: 6V0C). The AlphaFold LetA prediction is shown, coloured by pLDDT score with colour key below (red to blue as confidence increases) (Jumper et al., 2021).

1.3 The *E. coli* AsmA-like proteins

1.3.1 The link to phospholipid transport in the cell envelope

Until recently, evidence of an anterograde phospholipid transporter that is needed to build the OM remained elusive. The first protein to be identified with a role in this transport was YhdP, due to the finding that the deletion of *yhdP* delayed the early cell death phenotype of the *mlaA** mutant (Grimm et al., 2020). In this mutant, the uncontrolled mislocalisation of phospholipids to the outer leaflet triggers an increase in LPS biosynthesis, which in turn, increases phospholipid transport to the OM (Sutterlin et al., 2016). Upon entry into stationary phase, the rate of this transport becomes higher than the rate of phospholipid synthesis at the IM due to nutrient scarcity, causing lethal IM shrinking. In Δ*yhdP mlaA** mutants, anterograde phospholipid transport reduced by 50%, slowing IM shrinkage and instead causing cell death through OM rupture. The authors also found that Δ*yhdP* cells exhibited increased sensitivity to SDS+EDTA and vancomycin, a hallmark phenotype of an impaired OM (Grimm et al., 2020). The findings suggested a role of YhdP in modulating high-flux and diffusive phospholipid transport from the IM to the OM.

YhdP is a member of the AsmA-like superfamily, of which *E. coli* encodes six genes: *yhdP*, *tamB*, *ydbH*, *yhjG*, *asmA* and *yicH* (Figure 1.11A) (Ruiz et al., 2021). With evidence for *yhdP* being the long sought-after anterograde phospholipid transporter, investigations into the whole family began.

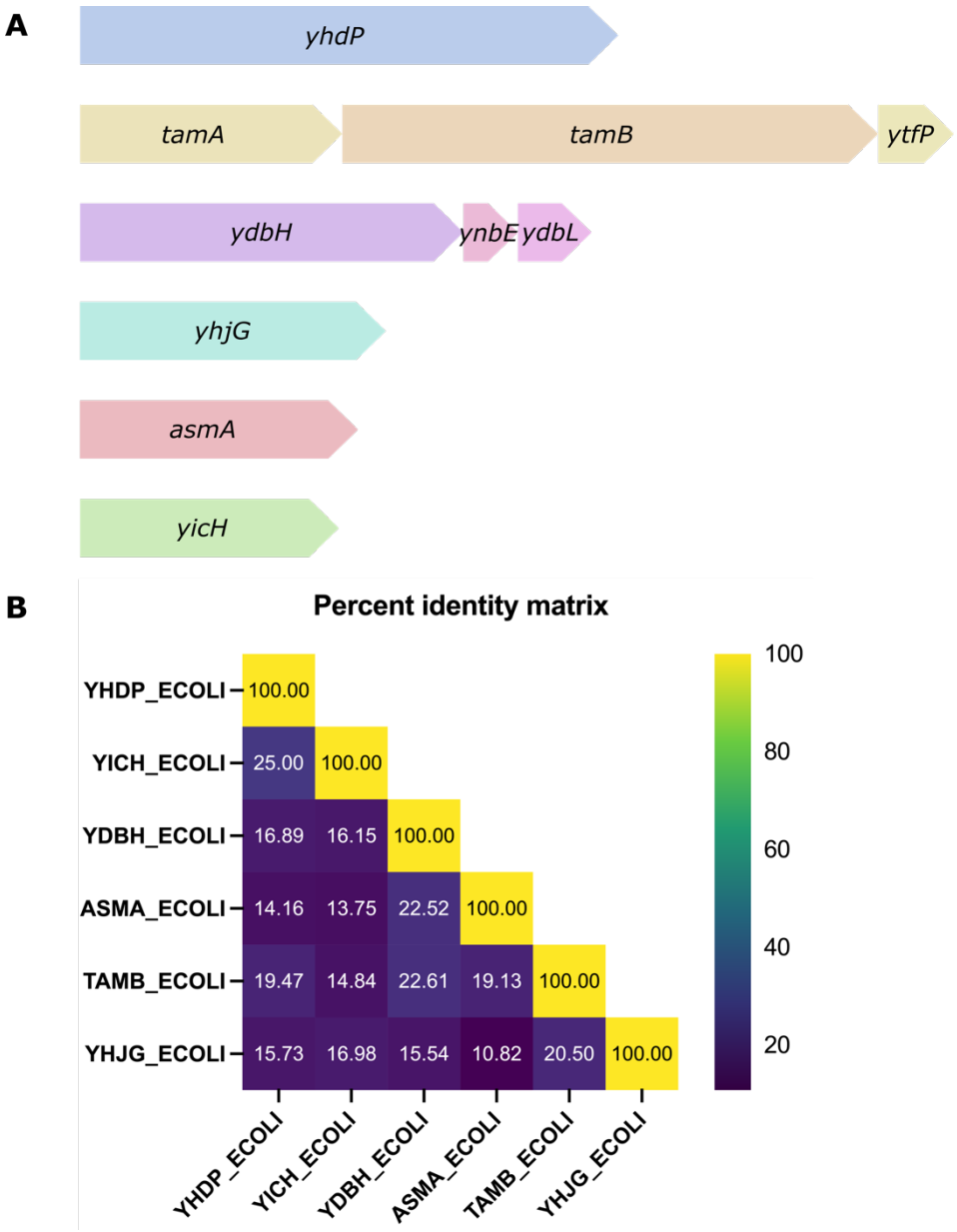


Figure 1.11. The *E. coli* AsmA-like superfamily. A. Operonic units of the six *asmA*-like superfamily genes in *E. coli*. B. Percent identity matrix of the six *E. coli* AsmA-like proteins made using Clustal Omega (pairwise % sequence identity).

The AsmA-like protein family are highly conserved across diderm bacteria, particularly across the *Enterobacteriaceae*, with the sequence conservation between the six *E. coli* proteins and their respective homologues across bacterial species shown in Appendix 1. TamB is the most well conserved protein out of the family, with YicH, YdbH and YhjG the least conserved.

The family were first identified through comprehensive hidden Markov model (HMMER) searches of TamB across Gram-negative bacteria using a lenient cut-off followed by Markov clustering (Heinz et al., 2015). This approach grouped proteins containing distinct domains which were found to share underlying sequence similarity. The AsmA-like proteins were identified through this search, with six *E. coli* proteins in the dataset, each in an individual cluster. The AsmA-like family are also sometimes referred to in the literature as being part of the Pfam clan CL0401, which contained each of the conserved domains found in the family, but this entry has since been removed from the InterPro database (Ruiz et al., 2021; Cooper et al., 2023). Also of note is that YicH and YhjG are paralogs of AsmA, with both YhjG and AsmA being members of the "AsmA-related" InterPro protein family (IPR052894) (Heinz et al., 2015).

Although the pairwise sequence identity between the six proteins is low (all ≤25%, Figure 1.11B), explaining the need for a lenient cut-off in the initial HMMER search, the underlying sequence similarity of their conserved domains, their grouping in the earlier Pfam clan CL0401, along with later predicted structural similarity through secondary structure predictions and AlphaFold, has solidly established their classification as a protein family within the literature.

Very little was previously known about this family. The identification of *asmA* came from the discovery that *asmA* null mutants suppress the assembly defects of mutant OmpF (Misra and Miao, 1995). Further investigation then revealed that LPS levels decreased in an *asmA* null mutant (Deng and Misra, 1996). It was also found that *asmA* nonsense mutations led to increased resistance to the LPS-acting antibiotic colistin (Marciano et al., 2022). Evidence for a role of TamB in the transport of OMPs and autotransporters had been seen as part of the TAM complex with the Omp85 protein TamA, as discussed previously (Selkrig et al., 2012; Shen et al., 2014; Wang et al., 2024a). Finally, *yhdP* was known to be involved in the maintenance of

the OM through its functional interaction with cyclic Enterobacterial Common Antigen (cyclic ECA) (Mitchell et al., 2018). Deletion of *yhdP* caused a decrease in the levels of cyclic ECA and inversely, a deletion of *wzzE*, encoding a protein needed in the synthesis of cyclic ECA, overcame the increased OM permeability phenotype of Δ*yhdP* cells. Additionally, *yhdP* was proposed to have a role in the stationary phase stress response as it was necessary for RpoS-dependent SDS resistance in carbon-limited stationary phase (Mitchell et al., 2017). There were no previously characterised roles of the other three AsmA-like family: *ydbH*, *yhjG* and *yicH*.

Many genetic investigations have since been performed on the whole family. Most notably, a functional redundancy between the three largest proteins in the family, YhdP, TamB and YdbH, was found in two independent studies (Ruiz et al., 2021; Douglass et al., 2022). A study in *P. aeurignosa* found a similar functional redundancy of AsmA-like homologues (Sposato et al., 2024). The combined loss of these three proteins was synthetically lethal, indicating that their function is essential. A double knockout mutant of Δ*yhdP*Δ*tamB* also resulted in severe permeability defects and smaller, rounder cells, with further sensitivity to OM stressors seen with the additional deletion of Δ*asmA* (Ruiz et al., 2021). Interestingly the increased lysis of the Δ*yhdP*Δ*tamB* mutant was suppressed with *mlaA* and *pldA* deletions, which would limit the removal of outer leaflet phospholipids. Additionally, the overexpression of *yhdP* rescued the viability of the triple knockout Δ*yhdP*Δ*tamB*Δ*ydbH* and resulted in phospholipids accumulating at the IM, suggesting *tamB* and/or *ydbH* are also needed for proper GPL trafficking (Douglass et al., 2022).

Different phenotypes were also found between Δ*yhdP*, Δ*tamB* and Δ*ydbH*, indicating they likely have distinct functions (Kumar and Ruiz, 2023). Recently, genetic evidence of differing behaviour from the combined deletion of these genes with Δ*fadR*, a regulator of fatty acid

degradation and unsaturated fatty acid production, suggests differential phospholipid substrate preferences for TamB and YhdP of unsaturated and saturated phospholipids, respectively (Rai et al., 2024). Further experimentation is required to confirm this preference. Despite these recent studies, no phenotypic effects have been observed for *yhjG* and *yicH*.

Recent biochemical and *in silico* investigations on YhdP provide further evidence of its phospholipid transport role (Cooper et al., 2023). Here, *in vivo* cross-linking found phosphate-containing molecules, most likely phospholipids, throughout the length of YhdP, and MD simulations showed spontaneous phospholipid entry into the predicted cavity of YhdP.

1.3.2 Known and predicted structures of the AsmA-like proteins

Bioinformatic and AlphaFold predictions reveal that all six proteins share a similar β-sheet rich fold and are IM-tethered by a single TM helix (Figure 1.12) (Jumper et al., 2021). They vary in size, with some able to potentially span the entire periplasm and insert into the OM, particularly YhdP (~270 Å) and TamB with its OM partner TamA (together: ~200 Å). YdbH is the third largest protein but at ~180 Å long, it may not be able to span the periplasm on its own. The AlphaFold-multimer prediction of YdbH with two proteins encoded for within its operon, the periplasmic YdbL and the OM lipoprotein YnbE (Figure 1.11A), resulted in a likely complex that would link the protein to the OM (Figure 1.12). This has very recently been experimentally confirmed through *in vivo* site-directed cross-linking, revealing interactions between the final β-strand of YdbH and YnbE (Kumar et al., 2024). AsmA, YhjG and YicH are predicted to be shorter in length (~150-170 Å), and as they aren't present in operons, no binding partners are yet to be predicted or identified.

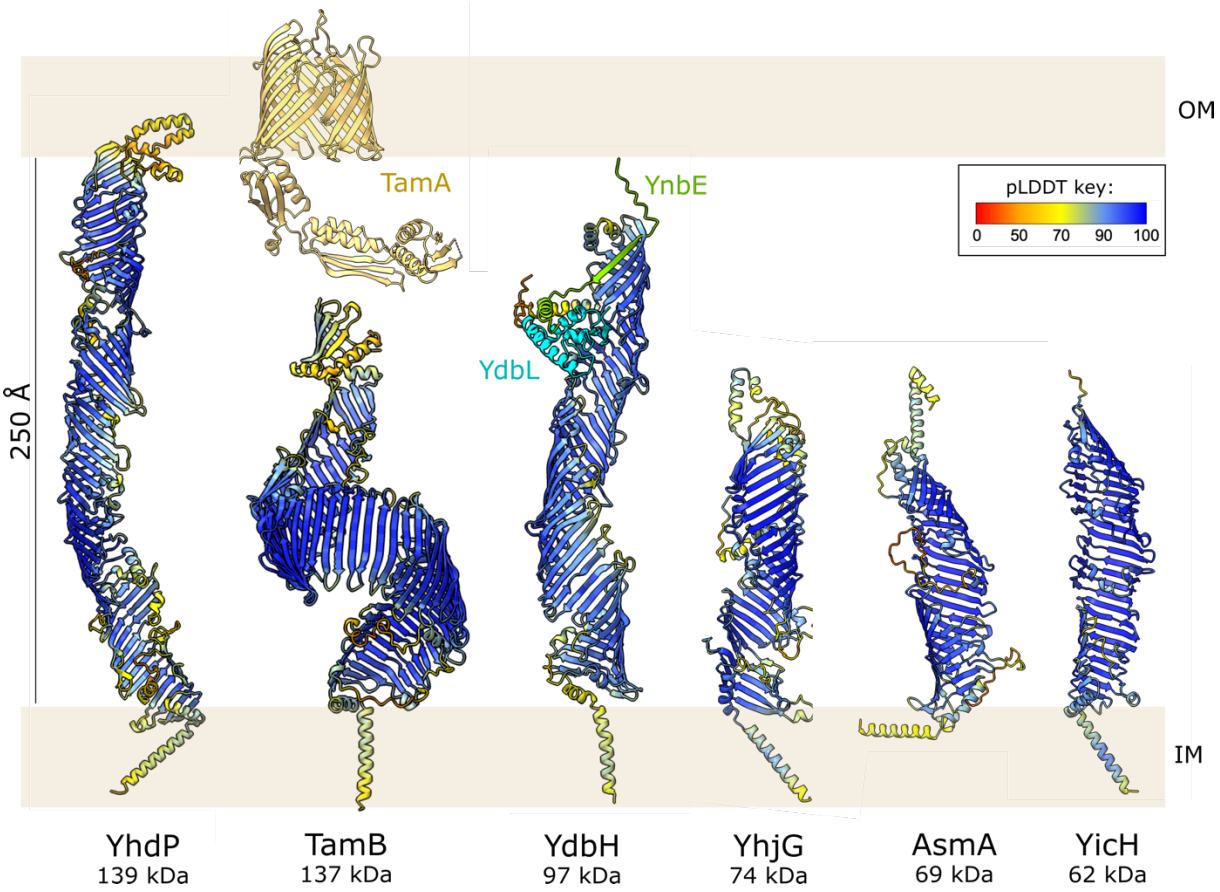


Figure 1.12. AlphaFold structure predictions of the AsmA-like protein family in *E. coli*. The AsmA-like superfamily of proteins are IM-tethered proteins that span into or across the periplasm and share an AsmA-like periplasmic domain. TamB interacts with the OMP TamA (beige, PDB: 4C00) to form the TAM complex. All other proteins are AlphaFold predictions, with the pLDDT key shown (red to blue, increasing confidence) (Jumper et al., 2021). YnbE (lime green) and YdbL (cyan) are predicted to form a complex with YdbH.

Part of the C-terminal DUF490 domain of TamB has been crystallised to reveal a novel β-taco fold, and atomic force microscopy of TamB and negative stain EM 2D classes of YhdP reveal low-resolution long, elongated structures, similar in length to the predicted structures (Josts et al., 2017; Shen et al., 2014; Cooper et al., 2023). No other experimentally solved structures of the full-length AsmA-like proteins have yet been reported.

The β-taco fold comprises a concave β-sheet structure that forms a hydrophobic cavity (Figure 1.13A), capable of protecting phospholipid acyl chains or other hydrophobic substrates, and is

reminiscent of the Lpt β-jellyroll fold (Figure 1.13B) (Josts et al., 2017). Unmodelled density within the groove, likely the detergent that was needed for protein stability, suggests that the groove can indeed harbour amphipathic substrates. The AlphaFold structural predictions of the AsmA-like proteins reveal that repeats of this β-taco fold make up the large periplasmic domains of all six proteins, forming a continuous hydrophobic groove throughout each structure (Figure 1.12). Many of the proteins also share a hydrophobic C-terminal region, which is modelled less confidently, typically consisting of a helix-turn-helix which may be able to tether or insert into the OM. Evidence for this tethering at both ends has recently been seen in MD simulations of the full-length YhdP model in a simulated double bilayer (Cooper et al., 2023).

The N-terminal regions of the AsmA-like proteins (typically, the first 100-150 residues of the periplasmic domain) share remote homology with a conserved fold found in eukaryotic bridge-like lipid transport proteins (BLTPs) called the Chorein-N domain (Figure 1.13C). This fold is again reminiscent of the β-taco and β-jellyroll folds, forming a concave hydrophobic cavity. Eukaryotic BLTPs are known to be responsible for bulk lipid transport between membranes in eukaryotic cells (Wong et al., 2019). Structural elucidation of these proteins has revealed the presence of a characteristic repeating β-groove (RBG) domain throughout the structure, akin to the AsmA-like proteins (Figure 1.13C). BLTPs of known structure and function include Atg2, which transports lipids from the endoplasmic reticulum (ER) to the autophagosome, Vps13, which transport lipids from the ER to various other membranous organelles, and LPD-3, which transports lipids from the ER to the plasma membrane (Kumar et al., 2018; Wang et al., 2024b; Kang et al., 2024). It has been suggested that eukaryotic BLTPs may have indeed evolved from bacterial RBG proteins due to the predicted structural homology, furthering the evidence for a role of the AsmA-like protein family in bulk phospholipid transport (Neuman et al., 2022).

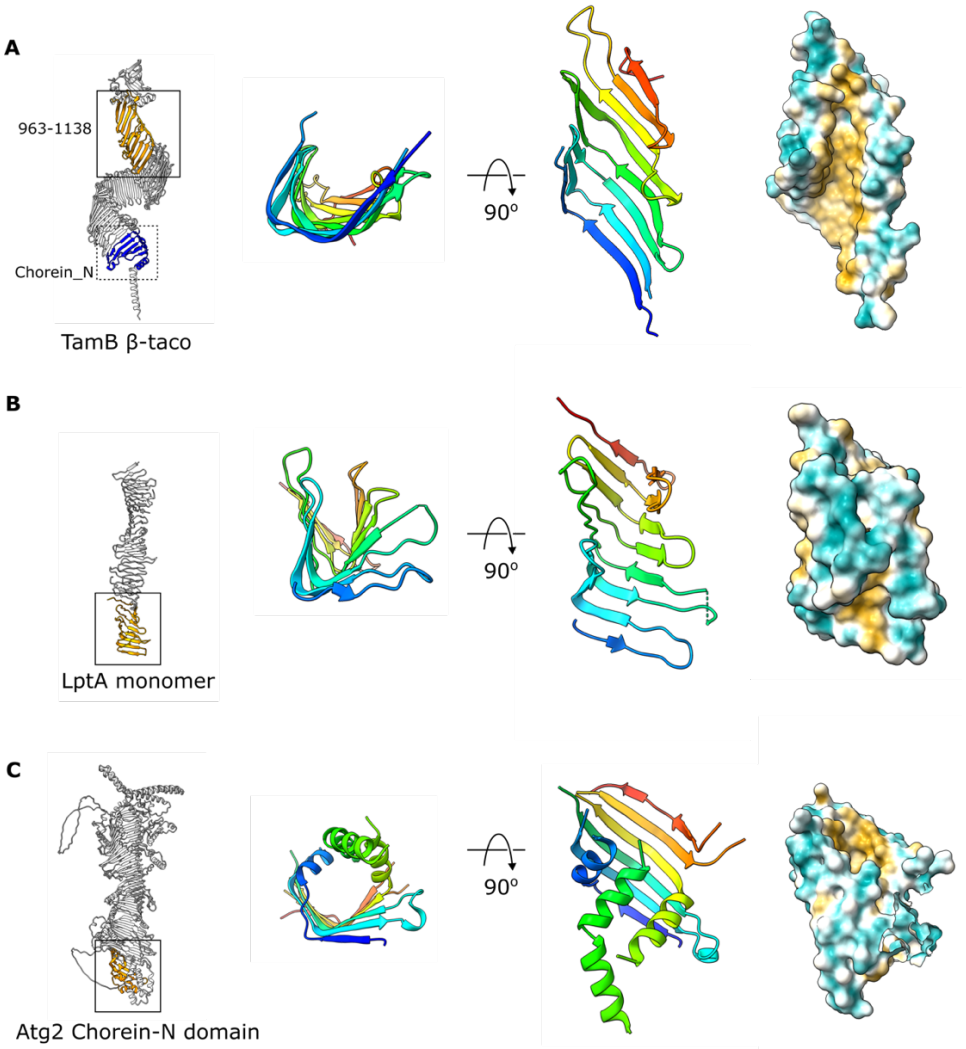


Figure 1.13. Similarity between β-taco fold, β-jellyroll fold and Chorein-N. The crystallised portion of TamB 963-1138 (orange) revealed a novel β-taco fold with hydrophobic interior, this fold is shared by all members of the AsmA-like family (PDB: 5VTG). B – The crystallised LptA monomer (orange) of the LPS transport machinery, showing a similar β-sheet concave structure capable of binding lipid tails (PDB: 2R1A). C – The Chorein-N domain (orange) of *Schizosaccharomyces pombe* Atg2, a lipid-binding domain, which shares remote homology to the N-termini of the AsmA-like proteins (blue on TamB structure) (PDBs: 689E and 8KBY). The AlphaFold structure of full-length Atg2 is shown (Jumper et al., 2021). Figure adapted from (Kumar and Ruiz, 2023).

Typically, function can be inferred by conserved domains, however the annotated domains within the *E. coli* AsmA-like family, shown in Figure 1.14, are typically of unknown structure and function. The only domain that is suggestive of a function is the DctA-YdbH domain, a dicarboxylate transport domain (Janausch et al., 2002). However, no other evidence of this

function for YdbH has yet been seen. These domains do not appear as structurally distinct regions, with the repeating β-groove (RBG) structure continuous throughout the periplasmic region of each protein (Cooper et al., 2023). This structural similarity suggests that the homology between AsmA-like proteins may be stronger than sequence data alone implies. The lack of distinct domains of known function makes predicting the functions of these proteins by sequence homology difficult, furthering the need for structural and functional investigation.

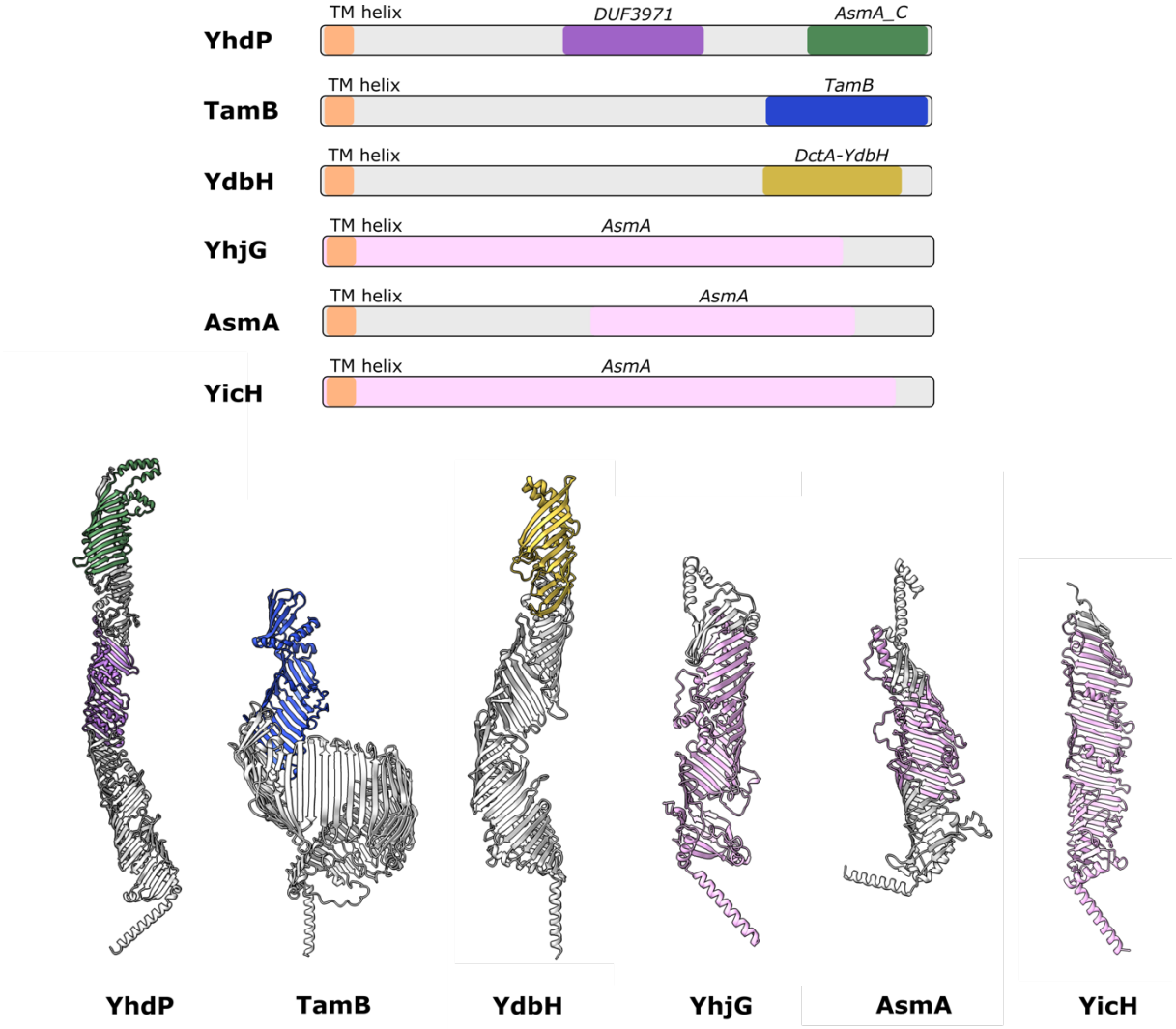


Figure 1.14. AsmA-like domains predicted by sequence homology. InterPro/Pfam domains for each of the AsmA-like proteins in *E. coli*. Annotated on the sequence and on the cartoon of each AlphaFold model. Interpro domains: DUF3971 – IPR025263, AsmA_C - IPR032712, TamB – IPR007452 (also the position of Pfam DUF490), DctA-YdbH - IPR021730, AsmA – IPR007844. Figure adapted from (Cooper et al., 2023).

1.4 Biological significance and aims of the study

Growing evidence for a role of the AsmA-like proteins in bulk, diffusive anterograde phospholipid transport from the IM to the OM in Gram-negative cell envelopes sheds light on an area of OM biogenesis that has long remained elusive. Further investigation into these pathways not only improves our understanding of this fundamental process but could also provide novel targets for antibiotic drug discovery in the combat against rising antimicrobial resistance.

Many questions are still left unanswered. Biochemical evidence of phospholipid binding and transport through these proteins is still lacking, as is the mechanism of action. There is also very little structural data for any of the AsmA-like proteins. Furthermore, there is a significant lack of information regarding the smaller AsmA-like proteins, YhjG and YicH.

Many of the papers discussed have been published since the beginning of the project, showing just how fast-moving and exciting this research area is. The project focus has therefore shifted throughout with influence from experimental results, project feasibility and findings from other research groups. Therefore, the chapters presented in this thesis approached the following aims:

1. To characterise the *asmA*-like genes using phenotypic profiling screens and phenotypic complementation assays.
2. To express, purify and biochemically characterise the *E. coli* AsmA-like proteins.
3. To elucidate the structures of two of the AsmA-like proteins: YhjG and YicH.
4. To develop a phospholipid transport assay and determine mechanisms of transport to investigate the functions of YdbH, YhjG and YicH.

CHAPTER 2. MATERIALS AND METHODS

All materials were purchased from Sigma-Aldrich unless otherwise stated.

2.1 Construct design and mutagenesis

2.1.1 Construct design

The initial constructs that were designed and synthesised for each *E. coli* AsmA-like protein and the putative complex of YdbH, YnbE and YdbL are shown in Table 2.1.

Table 2.1. Synthesised DNA construct designs.

Protein	Plasmid	His-tag location	TEV site	Source
YhdP	pET-26b(+)	C-terminal	N	GenScript
YdbH	pET-26b(+)	C-terminal	N	GenScript
YicH	pET-26b(+)	C-terminal	N	GenScript
YhjG	pET-26b(+)	C-terminal	N	GenScript
AsmA	pET-26b(+)	C-terminal	Y	GenScript
TamB	pET-22b(+)	C-terminal	Y	GenScript
YdbH-YnbE-YdbL	pET-29b(+)	C-terminal of YdbL	N	Twist

2.1.2 Q5 site-directed mutagenesis

Site-directed mutagenesis was performed to alter existing constructs via deletion, addition or substitution of base pairs when needed. Q5 site-directed mutagenesis (New England Biolabs) was performed for all mutagenesis following the manufacturer's protocol. The reaction mix is listed in Table 2.2 and the PCR conditions are listed in Table 2.3. Primers were designed and corresponding annealing temperatures (T_a) were calculated using NEBaseChanger. Agarose gel electrophoresis was used to check for successful amplification (2.1.3).

The resulting PCR product was treated with the enzymes T4 polynucleotide kinase (PNK), T4 ligase (T4L) and DpnI; either as separate components or as a KLD enzyme mix (New England

Biolabs). For individual enzymes, the reaction was left for 1 hour at RT. The reaction with the premade KLD enzyme mix was left for 20 minutes at RT. Reaction constituents are listed in Table 2.4. 5 µL of the KLD reaction was then transformed into *E. coli* DH5α competent cells (New England Biolabs) before plating on antibiotic selection LB agar plates. Mutagenesis was confirmed by plasmid sequencing (2.1.4).

Table 2.2. Q5-site directed mutagenesis reaction mix.

Reagent	Amount (µL)
10x Hot Start High-Fidelity 2x Master Mix	12.5
10 µM Forward primer	1.25
10 µM Reverse primer	1.25
Template DNA (1-25 ng/µL)	1
Nuclease-free water	9

Table 2.3. PCR conditions for site-directed mutagenesis.

Step	Temperature (°C)	Time
Initial denaturation	98	30 sec
30 cycles	98	10 sec
	T_a	30 sec
	72	30 sec/kb
Final extension	72	2 min
Hold	4	

Table 2.4. KLD reaction mix.

Reagent	Amount (µL)
KLD enzyme mix	
PCR product	1
2x KLD reaction buffer	5
10x KLD enzyme mix	1
Nuclease-free water	3
Separate enzymes	
PCR product	1
T4 PNK	1
T4 ligase	1
DpnI	1
T4 ligase buffer	1
Nuclease-free water	5

2.1.3 Agarose gel electrophoresis

1% (w/v) agarose was made by dissolving agarose powder (Sigma Aldrich) in 1x TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8.5) and heating the solution. SYBR Safe DNA stain (Thermo Fisher Scientific) was then added in a 1:10000 (v/v) ratio before casting the gel in a Bio-Rad casting tray. Gel samples were prepared by mixing PCR product with 6x gel loading dye purple (New England Biolabs). A 1 kb DNA ladder (New England Biolabs) was used as a marker. Gels were run in 1x TAE buffer for 40 minutes at 100 V, visualised under UV light and imaged using a GelDoc (Bio-Rad).

2.1.4 Plasmid sequencing

Plasmid DNA was purified using the Qiagen QIAprep Spin Miniprep kit following the manufacturer's protocol. DNA concentrations were measured using a DeNovix DS-11 FX spectrophotometer. Sequencing to check for mutations was performed using Source Bioscience's Sanger Sequencing service with their in-house T7 Forward or T7 Reverse primers.

Full plasmid sequencing was performed using PlasmidNG's Oxford Nanopore plasmid sequencing service.

2.2 Cell culture

2.2.1 Strains used in this study

The *E. coli* strains used for protein expression, plasmid amplification, and genetics experiments are listed in Table 2.5.

Table 2.5. *E. coli* strains used in this study.

Strain	Purpose	Source
BL21(DE3)	Protein expression	New England Biolabs
C43(DE3)		(Miroux and Walker, 1996)
Rosetta^{TM}(DE3)		Novagen
Lemo21(DE3)		New England Biolabs
BL21(DE3)pLysS		Novagen
NEB5α	Plasmid amplification	New England Biolabs
DH5α		New England Biolabs
MG1655 WT	*asmA*-like knockout strains used in genetics studies (Chapter 3)	(Ruiz et al., 2021)
NR3182 Δ*yhdP*::frt		
NR3183 Δ*tamB*::frt		
NR3184 Δ*asmA*::frt		
NR5158 Δ*yicH*::frt		
NR5159 Δ*yhjG*::frt		
NR5160 Δ*ydbH*::frt		
NR3191 Δ*yhdP*::frt Δ*asmA*::kan		
NR3193 Δ*tamB*::frt Δ*asmA*::kan		
NR5161 Δ*yhdP*::frt Δ*tamB*::frt		
NR5166 Δ*yicH*::frt Δ*yhdP*::kan		
NR5167 Δ*yicH*::frt Δ*tamB*::kan		
NR5168 Δ*yhjG*::frt Δ*asmA*::kan		
NR5169 Δ*yhjG*::frt Δ*yhdP*::kan		
NR5171 Δ*yhjG*::frt Δ*tamB*::kan		
NR5172 Δ*ydbH*::frt Δ*yhdP*::kan		
NR5173 Δ*ydbH*::frt Δ*yicH*::kan		
NR5174 Δ*ydbH*::frt Δ*tamB*::kan		
NR5218 Δ*yicH*::frt Δ*asmA*::kan		
NR5219 Δ*ydbH*::frt Δ*asmA*::kan		
NR5221 Δ*ydbH*::frt Δ*yhjG*::kan		
NR5236 Δ*yhjG*::frt Δ*yicH*::frt		

2.2.2 Preparation of culture media

Liquid culture media used in this study are shown in Table 2.6. Media were prepared according to the manufacturer's instructions and supplemented with antibiotics if necessary for selection. LB agar was prepared by adding 15 g/L of nutrient agar to LB media and following the manufacturer's instructions.

Table 2.6. Liquid culture media used in this study.

Media	Constituents
Miller's lysogeny broth (LB)	10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract
Terrific broth (TB)	12 g/L peptone, 24 g/L yeast extract, 9.5 g/L K_2PO_4, 2.2 g/L KH_2PO_4
Auto-induction (AI)	10 g/L tryptone, 5 g/L yeast extract, 3.55 g/L Na_2HPO_4, 3.4 g/L KH_2PO_4, 2.675 g/L NH_4Cl, 2.5% trace metals, 1 mM $MgSO_4$, 20 mL/L 50x sugar mix (0.14 M glucose, 0.3 M lactose, 2.7 M glycerol)

2.2.3 Preparation of competent cells

E. coli competent cells were prepared by inoculating 100 mL of LB with 1 mL of an overnight culture and incubating at 37°C with shaking until an OD_{600} of 0.5 was reached. Cells were then placed on ice for 10 minutes before pelleting at 4000 x *g* for 10 minutes at 4°C. The pellet was resuspended in 30 mL TFB1 (100 mM RbCl, 50 mM $MnCl_24H_2O$, 30 mM potassium acetate, 10 mM $CaCl_2$ and 1.5% (v/v) glycerol) and kept on ice for 90 minutes. The cells were then centrifuged at 4000 x *g* for 10 minutes at 4°C before resuspension in ice-cold 0.1 M $CaCl_2$ with 15% (v/v) glycerol. Cells were then aliquoted into 50 µL aliquots and snap-frozen in liquid nitrogen, before long-term storage at -80°C.

2.2.4 Bacterial transformation

A 20-50 µL aliquot of competent *E. coli* cells were transformed with 50-100 ng plasmid DNA via incubation on ice for 10 minutes, followed by heat shock at 42°C for 30 sec, then recovery on ice for 2 minutes. 200-950 µL SOC (super optimal with catabolite repression) broth was then added before incubation at 37°C for 1 hour with shaking. The cells were then plated on LB agar supplemented with the appropriate selection antibiotic and grown at 37°C overnight.

2.2.5 Expression trials

E. coli cell lines (BL21(DE3), C43(DE3), Rosetta^{TM}(DE3), BL21(DE3)pLysS and Lemo21(DE3), New England Biolabs) were transformed with an expression construct and grown in LB media overnight at 37^{o}C with shaking, supplemented with the appropriate antibiotic. 50 mL day cultures of either LB media, TB media or AI media supplemented with the appropriate antibiotic were inoculated with 0.5 mL of the overnight culture and grown to a range of OD_{600} values between 0.3-0.8. Pre-induction gel samples were taken. A range of IPTG concentrations were then used to induce expression (0.1 mM – 2 mM). For *E. coli* Lemo21(DE3) cells, 0.4 mM IPTG was used in all conditions, with varying L-rhamnose concentrations (0 mM – 2 mM). Expression conditions were trialled at temperatures between 18^{o}C - 37^{o}C for between 3 hours to overnight. Post-induction samples were taken. The samples were mixed with 2x Laemmli sample buffer (Merck), sonicated or heated at 95^{o}C for 10 minutes to break open the cells, centrifuged at 10,000 x *g* to pellet debris, and 10 µL of the supernatant was subsequently analysed by SDS-PAGE (2.4.4).

2.2.6 Full-scale growth conditions

E. coli C43(DE3) cells were used for membrane construct expression. *E. coli* BL21(DE3) cells were used for soluble construct expression. 20 mL LB media supplemented with the appropriate antibiotic was inoculated with a single colony and grown overnight at 37^{o}C with shaking (180 rpm). This was then added to 2 L of LB media supplemented with appropriate antibiotic and grown until OD_{600} = 0.5, at which point 0.5 mM IPTG was added to induce protein expression, which continued overnight at 18^{o}C. Cells were then pelleted by centrifugation at 4,000 x *g* for 15 minutes at 4^{o}C, then stored at -80^{o}C until required. Expression of YhdP and TamB required 6 L of media, whereas YdbH, YhjG and YicH required 2 L of media.

2.3 Genetic methods

2.3.1 Phenotypic profiling of the deletion mutant strains

The single and double knockout strains of the AsmA-like proteins in *E. coli* are listed in Table 2.5. AsmA-like knockout strains were generated by and kindly provided by Natividad Ruiz (Ruiz et al., 2021). In brief, the gene deletion was taken from the Keio library and introduced into an *E. coli* MG1655 background strain using P1 transduction. The kanamycin cassette was then removed by overexpressing FLP recombinase on a PCP20 plasmid, leaving a short scar sequence. Confirmation of the gene deletion strains was performed by colony PCR using primers upstream and downstream of the gene of interest, followed by agarose gel electrophoresis.

The *E. coli* MG1655 wild-type (WT) and knockout strains were grown overnight in LB media at 37^{o}C and then aliquoted into 96-well culture trays. A source plate was made of all the strains by mixing the culture equally with 60% (v/v) glycerol. The strains were then pinned from the source plate onto 2% (w/v) LB agar plates (or 2% (w/v) MacConkey agar) supplemented with different chemical perturbations using a pinning robot (Biomatrix 6). The chemicals and concentrations used are detailed in Chapter 3, and were found through the manual curation of online high-throughput phenotype screen databases (Nichols et al., 2011; Shiver et al., 2016; Price et al., 2018). Agar plates were incubated at 37^{o}C for 12 hours and imaged using an 18-megapixel Canon Rebel T3i (Canon). Colony size, opacity and circularity were quantified using the Iris imaging analysis software (Kritikos et al., 2017). The chemical screen was performed with the help of Michael Alao from Manuel Banzhaf's lab.

2.3.2 Analysis of the fitness of the mutants

Data analysis and data visualisation were kindly performed by Michael Aleo and Hannah Doherty from Manuel Banzhaf's lab (Doherty et al., 2023). Colony size data for each plate was checked for plate effects via a Wilcoxon rank sum test between the distributions of colony sizes for the outer two edges of the plate and the centre colonies. The outer edge colonies were scaled to the plate middle mean (PMM) if there was a significantly different distribution. The PMM of each plate was normalised to the median colony size of all colonies on the plate. For the heatmap, a fitness ratio was calculated by dividing the average colony size for each set of mutants within each plate by the average colony size of the WT of the same plate. The average fitness ratios for each mutant in each condition were then calculated between replicate plates. For the swarm plots, each individual mutant colony within each plate was divided by the mean colony size of the WT from the same plate. A one-way ANOVA was performed between each mutant's distribution of fitness ratios and the wild-type distribution of fitness ratios followed by a Tukey HSD analysis to determine if distributions were significantly different for each condition. Significance values between the mutant and the wild-type were denoted as such: *: 1.00e-02 < p <= 5.00e-02; **: 1.00e-03 < p <= 1.00e-02; ***: 1.00e-04 < p <= 1.00e-03; ****: p <= 1.00e-04.

2.3.3 Growth curve analysis

E. coli MG1655 wild-type (WT), *ΔyhdP*, *ΔtamB* and *ΔtamBΔyhdP* mutants were grown in LB overnight at 37ºC, then normalised to an OD_{600} of 0.04. A 96-well plate was prepared containing LB media and the antibiotic A22 (*S*-(3,4-dichlorobenzyl) isothiourea) serially diluted from 20 µg/mL in row A to 0 µg/mL in row G, with a negative control of no culture inoculation in row H. Each normalised culture was added to the plate in triplicate. A clear, permeable membrane was used to cover the wells and the plate was read using a Clariostar plate reader (BMG

Labtech) following the OD_{600} at 37ºC for 72 cycles with 10 minutes per cycle and shaking between each step.

2.3.4 Phase contrast microscopy

Phase contrast microscopy was performed with the help of Gabrielle Boelter from Manuel Banzhaf's lab, following the protocol in Boelter *et al.*, (2022). *E. coli* MG1655 WT, Δ*yhdP*, Δ*tamB* and Δ*tamB*Δ*yhdP* mutants were grown in LB media overnight at 37ºC and normalised to OD_{600} = 0.01. These were then grown in LB media or LB media supplemented with 40 µg/mL A22 at 37ºC until OD_{600} = 0.4. The cultures were then adjusted to 33 mM sodium phosphate, pH 7.4 and fixed with formaldehyde and glutaraldehyde to 2.4% (v/v) and 0.04% (v/v), respectively, for 30 minutes at RT. The cells were washed with sterile 1x PBS and resuspended in sterile 1x PBS to a final OD_{600} of 2.5 before applying to agarose pads.

Microscope slides were prepared by adding 1% (w/v) agarose in 1x PBS to a Gene Frame (Thermo Scientific) on a glass microscope slide, before a second glass slide was placed on top and the agarose left to set. The top slide was then removed and a channel was cut out of the agarose to allow two samples per slide. 2 µl of the fixed cells was applied to the pad and a coverslip was added once dry. Microscope slides were visualised using a Zeiss AxioObserver microscope using a Plan-Apochromat 100 x/Oil Ph3 objective lens and illumination from HXP 120 V for phase contrast images. Images were analysed using the MicrobeJ plugin for Fiji (Ducret et al., 2016). Cell size, length and area were analysed using a two-way ANOVA followed by a Tukey HSD analysis comparing against the WT cells using GraphPad Prism v10.0.

2.3.5 Phenotypic complementation assay

Phenotypic complementation assays were performed in the loss-of-fitness condition for Δ*yhdP*Δ*tamB* knockouts of LB agar supplemented with 0.1% (w/v) SDS and 0.1 mM EDTA, found from the phenotype screen and optimisation of stressor concentrations. *E. coli* MG1655 WT and Δ*yhdP*Δ*tamB* knockout mutants were made competent and transformed with either an empty plasmid or plasmid containing the WT or mutant gene (pET-26b(+) for *yhdP*, pET-22b(+) for *tamB*) (2.2.3, 2.2.4). Mutations were made via Q5 site-directed mutagenesis (2.1.2). Strains were grown in LB media overnight at 37ºC with shaking, supplemented with 30 µg/mL kanamycin for pET-26b(+) transformants and 100 µg/mL ampicillin for pET-22b(+) transformants. The cultures were normalised to an OD_{600} of 1 and serial diluted in LB from 10^{0}-10^{-7} in a 96-well plate. 2.5 µL of each culture was then spotted on LB agar plates and LB agar plates with 0.1% (w/v) SDS and 0.1 mM EDTA. Plates were incubated at 37ºC for 16 hours before imaging with a U:Genius system (Syngene). Leaky expression from the T7 promoter was sufficient for complementation, confirmed by Western blotting of whole cells with an anti-His antibody (2.4.5).

2.4 Protein purification

2.4.1 Purification of soluble proteins

Cell pellets were resuspended in 50 mM Tris, 500 mM NaCl, 10 mM imidazole, pH 8 (or pH 8.5 for YhjG soluble constructs) supplemented with cOmplete™ EDTA-free protease inhibitor tablets (Roche). They were then lysed via 3-6 passes through a C3 Emulsiflex cell disruptor (Avestin) at 17,000 psi at 4ºC, or via sonication on ice 20 seconds on, 20 seconds off for a total of 2 minutes sonication time. Lysed cells were clarified by centrifugation at 50,000 x *g* for 1 hour at 4ºC. Supernatants were filtered with a 0.45 µm syringe filter (Merck) before mixing with pre-equilibrated Ni-NTA beads (Qiagen) for 1 hour or they were passed through a HisTrap

Ni-NTA 5 mL column (Cytiva). The beads/column were then washed with 5-10 column volumes (CV) of 50 mM Tris, 500 mM NaCl, 50 mM imidazole, pH 8/8.5 before the remaining protein was eluted from the resin using 50 mM Tris, 500 mM NaCl, 500 mM imidazole, pH 8/8.5. Samples of each fraction were taken and mixed with 2x Laemmli sample buffer (Merck) to be run on SDS-PAGE gels for analysis (2.4.4). Samples containing protein of interest were pooled, concentrated using centrifugal spin concentrators with an appropriate molecular weight cut-off (MWCO) (Merck Millipore) and gel filtered through a Superdex® 200 size exclusion column pre-equilibrated in 50 mM Tris, 150 mM NaCl, pH 8/8.5 using an ÄKTA purifier (GE Healthcare). Samples from peak fractions were mixed with 2x Laemmli sample buffer (Merck) and analysed by SDS-PAGE (2.4.4). Protein concentrations were determined using a DeNovix DS-11 FX spectrophotometer.

2.4.2 Purification of membrane proteins

Cell pellets were resuspended in 50 mM Tris, 500 mM NaCl, 10 mM imidazole, pH 8 (or pH 8.5 for YhjG constructs) supplemented with cOmplete EDTA-free protease inhibitor tablets (Roche). They were then lysed via 3-6 passes through a C3 Emulsiflex cell disruptor (Avestin) at 17,000 psi at 4°C. The lysed cells were centrifuged at 10,000 x *g* for 30 minutes at 4°C to pellet cell debris, and the resulting supernatant was centrifuged at 100,000 x *g* for 1 hour at 4°C to pellet the cell membranes. The pellet was removed and weighed, before resuspension in solubilisation buffer (50 mM Tris, 500 mM NaCl, 10 mM imidazole, and 1% (w/v) n-dodecyl-β-D-maltoside (DDM) or 2.5% (w/v) styrene maleic acid co-polymer (SMA), pH 8 (or pH 8.5 for YhjG constructs)) using a dounce homogensier at a ratio of 40 mg membrane to 1 mL buffer which was left to mix on a rotor overnight at 4°C. Additionally, for YhdP and TamB, 5% (v/v) glycerol and 0.5 mM tris-(2-carboxyethyl)phosphine (TCEP) were added to each buffer for improved protein stability.

After solubilisation, any insoluble material was pelleted out via centrifugation at 100,000 x *g* for 30 minutes at 4ºC. The membranes were filtered through a 0.45 µm filter and mixed with pre-equilibrated Ni-NTA beads (Qiagen) for between 2 hours to overnight at 4ºC on a rotor. The beads were then applied to a gravity flow column and the flow-through was collected, before the beads were washed with 5-10 CV of 50 mM Tris, 500 mM NaCl, 30-50 mM imidazole and 0.03% (w/v) DDM (DDM was omitted if SMA-solubilised), pH 8/8.5. Protein was then eluted in 1 mL fractions with 50 mM Tris, 150 mM NaCl, 500 mM imidazole and 0.03% (w/v) DDM (unless SMA-solubilised), pH 8/8.5. Samples of each fraction were taken and mixed with 2x Laemmli sample buffer (Merck) to be analysed by SDS-PAGE (2.4.4). Samples containing the protein of interest were pooled and concentrated prior to size exclusion chromatography (SEC) using centrifugal spin concentrators of an appropriate MWCO (Merck Millipore). YdbH was not concentrated due to aggregation.

SEC was then performed on the pooled fractions on a pre-equilibrated column (a Superose® 6 10/300 for YhdP and TamB and a Superdex® 200 10/300 or 16/600 for YdbH, AsmA, YicH and YhjG) using an ÄKTA purifier system (GE Healthcare). Equilibration buffers were the same as the elution buffers without imidazole. To exchange into the amphipol A8-35, a 4x molar excess of A8-35 was incubated with the purified DDM-solubilised protein overnight at 4ºC on a rotor. Pre-equilibrated and degassed Bio-Beads SM-2 resin (Biorad) were then added to the sample according to the manufacturer's instructions and mixed for 90 minutes on a rotor at 4ºC. The biobeads were then filtered out of the sample before SEC purification (as above). Columns were pre-equilibrated with 50 mM Tris, 150 mM NaCl, pH 8/8.5. Detergent exchange was performed by loading purified, DDM-solubilised protein onto a SEC column pre-equilibrated in the original elution buffer but with 0.7% (w/v) n-octyl-β-D-glucopyranoside (β-OG) or 0.6% (w/v) lauryldimethylamine N-oxide (LDAO) as the detergent. Samples from peak

fractions were mixed with 2x Laemmli sample buffer (Merck) and run on SDS-PAGE gels for analysis (2.4.4). Protein concentrations were determined using a DeNovix DS-11 FX spectrophotometer.

2.4.3 Solubilisation screen

Solubilisation screens were performed at the Membrane Protein Laboratory (MPL) at the Research Complex at Harwell (RCaH) using a standard protocol with the aid of Claudia Burn-Leefe (Gamage et al., 2023). Membrane pellets were prepared as previously described (2.4.1) using 50 mL of culture and resuspended using a dounce homogeniser in 12.5 mL buffer (50 mM Tris, 500 mM NaCl, pH 8 with a cOmplete EDTA-free protease inhibitor tablet). An extraction 96-well block was used to trial solubilisation of YhdP, TamB, YdbH and AsmA in the following conditions: 1% (w/v) DDM, 1% (w/v) DDM and 0.2% (w/v) cholesteryl hemisuccinate (CHS), 1% (w/v) lauryl maltose neopentyl glycol (LMNG), 1% (w/v) digitonin, 1% (w/v) glycol-diosgenin (GDN) and 1% (w/v) fos-choline 12 (FC-12, used as a positive control) and 2.5% (w/v) SMA. For exchange into amphipols, nanodiscs and peptidiscs, 1% (w/v) DDM was used to first solubilise the membrane. The resuspended membrane was added to each condition. The 96-well extraction block was then covered with a foil seal and mixed using a rotor for 1 hr at 4ºC, before centrifugation at 3500 x *g* for 30 mins at 4ºC.

The supernatants were then transferred to a 96-deep well block. TALON resin (Takara Bio) was washed and pre-equilibrated in 50 mM Tris, 150 mM NaCl, pH 8 before adding to each well of the block, except for the SMA well which used Ni-NTA resin (Thermo). The block was covered with a foil seal and mixed for 1 hr on a rotor at 4ºC. The block was then centrifuged for 30 min at 3500 x *g* at 4ºC. The supernatants were transferred to a 96-well filter plate on top of a waste deep-well block for small-scale purification. Flow-through was collected by centrifugation of the block for 1 min at 500 x *g* at 4ºC. Wash buffer (50 mM Tris, 500 mM

NaCl, 20 mM imidazole, 3x CMC of detergent, pH 8) was then added before the block was centrifuged for 1 min at 500 x *g* at 4ºC and repeated 3 times.

For improved protein purity and stability, exchange from 1% (w/v) DDM solubilised protein into the following membrane mimetics were trialled: amphipol 18, A8-35, non-ionic amphipol (NAPol), a poly(maleic anhydride-alt-1-decene), 3-(dimethylamino)-1-propylamine derivative (PMAL-C8) and nanodisc scaffold peptide (NSPr). Polymers were added to the wells to a final concentration of 0.1 mg/mL and incubated for 15 min at 4ºC on a rotor. The block was centrifuged at 500 x *g* for 1 min at 4ºC before wash buffer (50 mM Tris, 500 mM NaCl, 20 mM imidazole, pH 8) was added to these wells, repeating two more times. The block was then centrifuged for 500 x *g* for 1 min at 4ºC to completely remove the residual wash buffer.

Elution buffer (50 mM Tris, 500 mM NaCl, 300 mM imidazole, pH 8, 3x CMC detergent if necessary) was then added to the filter plate which was placed on top of a 96-well microtitre plate. Centrifugation at 500 x *g* for 1 min at 4ºC was performed to elute the protein. Protein yield, monodispersity and stability were then analysed by SDS-PAGE (2.4.4), F-SEC, nanoDSF and negative stain EM (2.5.1).

F-SEC was performed by Claudia Burn-Leefe at the Membrane Protein Laboratory (MPL). Briefly, 20 µL of the eluate was transferred into a fresh microtiter plate and sealed with a pierceable seal. 10 µL of the eluate was automatically injected by an autosampler onto an SRT-C-300 HPLC column connected to an SRT-C 300 guard column. Samples were run at a flow rate of 1.0 mL/min at 4ºC. UV absorbance at 280 nm and tryptophan fluorescence (ex. 280 nm, em. 350 nm) were monitored.

Nano-differential scanning fluorimetry (nanoDSF) was performed if protein concentrations of the eluate were above 0.1 mg/mL. Concentrations were adjusted so that they were similar for

each condition using the relevant elution buffer. 10 µL of each sample were loaded into a glass capillary (Prometheus) and placed in a Prometheus NanoDSF machine (Prometheus). Intensity of tryptophan fluorescence (ex. 280 nm, em. 330-350 nm) was adjusted so that the signal was within the range of the machine. A thermal melt of 1°C per second was then performed from 25°C to 95°C, measuring the tryptophan fluorescence throughout. Data analysis was then performed on the Prometheus Panta to report T_m values for each condition.

2.4.4 SDS-PAGE

Samples mixed with 2x Laemmli sample buffer (Merck) were loaded into pre-cast SurePAGE Bis-Tris 4-12% polyacrylamide gels (GenScript) or laboratory-cast 10% polyacrylamide gels. 10% polyacrylamide gels were made using a resolving gel of 0.39 M Tris-HCl pH 8.8, 0.1% (w/v) SDS, 10% (v/v) acrylamide, 0.1% (w/v) ammonium persulphate and 0.01% (v/v) tetramethylethylenediamine (TEMED), and a stacking gel of 0.13 M Tris-HCl pH 6.8, 0.1% (w/v) SDS, 5% (v/v) acrylamide, 0.1% (w/v) ammonium persulphate and 0.01% (v/v) TEMED. A Precision Plus Protein all blue prestained protein standard (BioRad) was used as the molecular weight ladder. Gels were run in MES-SDS running buffer (GenScript) at a constant voltage of 165 V for 35 minutes for pre-cast gels and at 135 V for 50 minutes for laboratory-cast gels. Quick Coomassie Stain (Neo Biotech) was used to stain the gels for Coomassie staining. Gels were imaged using a GS-800 densitometer (BioRad). Gels made at the MPL used pre-cast Criterion XT gels (BioRad) with SeeBlue Pre-Stained Standard 1x molecular weight ladder (Invitrogen). Gels were visualised with a ChemiDoc imaging system (BioRad).

2.4.5 Western blot

Proteins were separated by SDS-PAGE and the peqGOLD protein ladder VI (VWR) was used as the molecular weight standard. SDS-PAGE gels were transferred to nitrocellulose membranes using a TransBlot Turbo transfer system (Biorad) using the Turbo Mini-Gel setting.

Membranes were blocked with either 5% (w/v) skimmed milk powder in TBST (20 mM Tris, 150 mM NaCl, 0.1% (w/v) Tween® 20) or 5% (w/v) bovine serum albumin (BSA) in TBST with rocking for 1 hour at RT or 4ºC overnight. After blocking, anti-His monoclonal antibody (Takara Bio) was diluted at 1:2500 in blocking buffer and added to the membrane for 1 hour at RT or 4ºC overnight. Membranes were then washed 3 times for 10 mins each in TBST. Anti-mouse IgG HRP conjugated secondary antibody (Abcam) was diluted 1:10000 in TBST and incubated with the membrane for 1 hour at RT. Membranes were washed a further 3 times before activation using ECL (Cytiva) and visualisation using an Amersham 600 series imager (GE Healthcare).

2.4.6 Mass spectrometry

Samples for mass spectrometry analysis were excised from Coomassie-stained SDS-PAGE gels using a clean scalpel. Trypsin digest mass spectrometry and analysis were performed by the University of Birmingham Advanced Mass Spectrometry Facility. Peptides were sequenced and compared to *E. coli* databases for peptide identification.

2.4.7 Analytical ultracentrifugation

Sedimentation velocity experiments were performed at the University of Birmingham Analytical Ultracentrifugation (AUC) Facility by Pooja Sridhar. Proteins were diluted to a concentration of 0.1-0.5 mg/mL in 50 mM Tris, 150 mM NaCl, 0.03% (w/v) DDM, pH 8/8.5 (YicH/YhjG). Experiments were performed in a Beckman Coulter ProteomeLab XL-I AUC with a Ti50 rotor at 20ºC at 30,000 rpm, with absorbance measured at 280 nm. Data was analysed in Sedfit applying the continuous c(s) distribution model (Schuck, 2000). The buffer density and viscosity were calculated with SEDNTERP (Laue et al., 1992).

2.4.8 Thin layer chromatography

Lipids were extracted from protein samples by chloroform-methanol extraction. 2 mL of 0.5 mg/mL protein was mixed with 1 mL of chloroform and 2 mL of methanol. The samples were vortexed for 5 minutes, incubated at 50°C for 30 minutes and vortexed for a further 5 minutes. Centrifugation at 2250 x *g* for 10 minutes at RT separated the organic and inorganic phases, allowing the lower chloroform phase to be extracted using a glass Pasteur pipette. The chloroform was evaporated from this extracted phase and the dried lipids were resuspended in 100 µL of fresh chloroform. 5 µL of sample and 1 µL of *E. coli* polar lipid standards were spotted onto the origin of a silica TLC plate (Sigma-Aldrich) before being run in a 6.5:2.5:1 (chloroform:methanol:acetic acid) solvent until the solvent front was 1 cm from the top of the TLC plate. The plate dried for 30 minutes before staining in 10% (w/v) phosphomolybdic acid (PMA) in 100% ethanol for 3 seconds. A heat gun was used to develop the stain.

2.5 Structural determination

2.5.1 Negative stain EM grid preparation

200 mesh carbon-coated copper grids (Agar Scientific) were glow discharged for 60 sec at 35 mA using a GloQube (Quorum) glow discharger. 10 µL of 4-100 µg/mL purified protein was applied to the grid and incubated for 1 minute before blotting with filter paper. 10 µL of 2.5% (w/v) uranyl acetate solution was then applied to the grid, incubated for 1 minute and blotted with filter paper. YdbH negative stain grids were made and viewed at the MPL by Dr Peter Harrison.

2.5.2 Negative stain EM data collection and processing

Images were recorded on a JEOL1400 microscope (JEOL) at 60,000x magnification. For YicH, 13 micrographs of a 26 µg/mL sample were taken. For YhjG, 23 micrographs of a 4 µg/mL

sample were taken. Single particle analysis was performed in Relion v3.0 (Scheres, 2012). Motion correction was performed using MotionCorr2 and CTF was estimated using CTFFIND4 (Rohou and Grigorieff, 2015; Zheng et al., 2017). Manual picking resulted in 1335 particles for YicH and 4057 particles for YhjG, which was followed by 2D classification, and the generation of three *de novo* 3D models from particles in the most populous 2D class. Visualisation and model fitting of the EM maps were then performed in UCSF ChimeraX (Pettersen et al., 2021). For YdbH, negative-stain EM was performed as part of the detergent and polymer screen, with 1-5 micrographs taken for each sample and no further processing taking place.

2.5.3 Cryo-EM grid preparation

Graphene oxide coated grids were prepared by glow discharging QUANTIFOIL® 300 mesh R1.2/1.3 AU grids (Quantifoil) for 180 sec at 40 mA using a GloQube glow discharger (Quorum). 3 µL of 0.2 mg/mL graphene oxide was applied to the grid and incubated for 1 minute before blotting off any excess with filter paper. The grid was then washed with 20 µL dH_2O and blotted with filter paper three times before being left to air-dry.

Carbon grids were prepared by glow discharging QUANTIFOIL® 300 mesh R1.2/1.3 AU grids (Quantifoil) for 60 sec at 35 mA using a GloQube® glow discharger (Quorum) for all grids except for YicH and YhjG in A8-35. YicH and YhjG in A8-35 grids were prepared at the MPL by glow discharging QUANTIFOIL® 200 mesh R1.2/1.3 copper grids for 60 sec + 60 sec at 30 mA using a GloQube® glow discharger (Quorum) with the help of Dr Peter Harrison.

Grids were frozen by applying 3 µL of sample to the grid before plunging into liquid ethane using a Vitrobot mark IV (ThermoFisher Scientific) at 4ºC and 95% humidity. Various blot forces and times were used. The cryo-EM data collections presented in this study were on grids

which used a blot force of 3, a blot time of 2 and a wait time of 10 sec. The grids were then clipped and transferred to grid box storage in liquid nitrogen until ready for screening.

2.5.4 Cryo-EM screening, data collection and processing

Grid screening was performed either at the Leicester Institute for Chemical and Structural Biology (LISCB) or at the electron Bio-Imaging Centre (eBIC). Grids screened at LISCB used a Titan Krios TEM (ThermoFisher Scientific) operating at 300 kV. Grids screened at eBIC used either a Glacios II TEM (ThermoFisher Scientific) or a Talos Arctica operating at 200 kV. Grid screening and data collection at eBIC was aided by Dr Peter Harrison from the MPL. Data collection parameters are listed in Chapter 5 for each grid data was collected on.

Data collection and processing pipelines are described in detail in Chapter 5. In summary, micrographs were motion-corrected and the CTF was estimated using cryoSPARC implementations (Punjani et al., 2017). Particles were picked with a variety of strategies and subject to several rounds of 2D classification. *Ab initio* model generation was then performed on the good classes followed by further refinements. Higher quality maps were subject to local resolution estimation before visualisation and model fitting in UCSF ChimeraX (Pettersen et al., 2021). Dr Peter Harrison performed the processing of eBIC datasets (YhjG in A8-35).

2.6 Functional characterisation

2.6.1 Preparation of liposomes

Liposomes were composed of phospholipids purchased from Avanti Polar Lipids. Donor liposomes were made up of 97.5% POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine), 1.5% head-labelled NBD-PE (1,2-dioleoyl-sn-glycero-3-phosphoethanolamine-N-(7-nitro-2-1,3-benzoxadiazol-4-yl)), and 1.5% head-labelled Rhodamine-PE (L-α-phosphatidylethanolamine-N-(lissamine rhodamine B sulfonyl)). Acceptor liposomes were

made up of 100% POPC. Lipids were dried under nitrogen gas to deposit multilamellar stacks and resuspended in 5 mL of 50 mM Tris, 150 mM NaCl, pH 8/8.5 (YicH and YdbH/YhjG) to a final suspension concentration of 50 µM. Liposomes were then made into single unilamellar vesicles (SUVs) using a sonicating water bath at 30ºC for 30 minutes until clear.

2.6.2 Lipid transport assay

The reactions were prepared in triplicate in a 96-well microplate. Each 150 µL reaction mixture consisted of 20 µL of 1-5 µM purified protein (final concentrations 0.13-0.67 µM), 10 µL donor liposomes (final concentration in reaction mix 3.3 µM), 10 µL acceptor liposomes (final concentration in reaction mix 3.3 µM) and 110 µL 50 mM Tris, 150 mM NaCl, pH 8/8.5. Control reactions used the protein SEC buffer instead of the protein, i.e. 50 mM Tris, 150 mM NaCl, 0.03% (w/v) DDM, pH 8/8.5. Controls were also performed using only donor liposomes, only fluorescent liposomes, and no liposomes, and all showed no fluorescent changes. Maximum NBD fluorescence was determined by popping liposomes with the addition of 0.6% (w/v) DDM (final concentration).

NBD fluorescence was read using a fluorescence intensity reader (Tecan Infinite® 200 Pro). The excitation wavelength was set at 460 nm and fluorescence emission was recorded at 538 nm. Readings were taken every 10 seconds using 10 flashes, with the gain and z position calculated from the well with popped liposomes. Baseline NBD fluorescence was measured for 10 minutes before the addition of the protein, where the fluorescence was immediately followed for >1 hour. Graphs report NBD fluorescence as a percentage of the maximum NBD fluorescence, shown as the mean value and standard error of the mean. Phospholipid transport initial rates were calculated by taking the gradient of a simple linear regression fit of the linear rate between 7 - 14 mins (after the burst kinetics phase). Maximum NBD fluorescence was

reported as the highest mean percentage of the total NBD fluorescence achieved, with standard error of the mean also shown.

2.6.3 Molecular dynamics simulations

All molecular dynamics (MD) simulations were performed using GROMACS 2021.4 (Abraham et al., 2015). Proteins were oriented in a bilayer using predicted orientations from the Orientation of Proteins in Membranes (OPM) server or by manual placement of the N-terminal helix in the bilayer (Lomize et al., 2022). Proteins were converted to coarse-grained (CG) representations using a Martini3 forcefield (Souza et al., 2021). Proteins were embedded into a preformed membrane with a symmetrical 1:4 POPG:POPE ratio using INSANE 3.0 (Wassenaar et al., 2015). The protein was in an elastic network with an upper and lower elastic bond cut-off of 1.0 nm and 0.5 nm, respectively. All systems were solvated with water and neutralised with 0.15 M NaCl. Electrostatics were modelled using reaction field and van der Waals interactions both with a cut-off of 1.1 nm.

The steepest descent algorithm was used to energy minimise the system before a 20 ns equilibration was performed over 0.01 ps time steps. Equilibration used V-rescale group temperature coupling and semi-isotropic Berendsen pressure coupling with constants of 1 ps and 12 ps, respectively, with a reference temperature of 310 K and 1 bar pressure. The 1 μs simulation was run over 0.02 ps time steps with V-rescale group temperature coupling (with a constant of 1 ps) and Parinello-Rahman semi-isotropic pressure coupling using 1 bar pressure and a constant of 12 ps. CG simulations were converted to atomistic representations using CG2AT (Vickery and Stansfeld, 2021). Simulations were visualised in PyMOL v2.4.1 or UCSF ChimeraX v1.7.1 (Schrodinger, 2015; Pettersen et al., 2021).

CHAPTER 3. GENETIC CHARACTERISATION OF THE ASMA-LIKE GENES IN *E. COLI*

3.1 Introduction

3.1.1 Phenotypic screening

Phenotypic screening involves growing strains where a gene of interest has been deleted or knocked out on various stressors to infer potential functions of the gene (Doherty et al., 2023). For example, if a wild-type (WT) cell can grow on an antibiotic, but a deletion strain has impaired growth, then the gene that was knocked out is likely to have a role in the mechanism of action of that antibiotic. As discussed in Chapter 1.3, there are now many studies looking at phenotypes of the *asmA*-like genes in various conditions and alongside other genetic deletions. Most notably, multiple papers have identified that *yhdP*, *tamB* and *ydbH* are functionally redundant, with an essential function associated with bulk, diffusive anterograde phospholipid (PL) transport (Douglass et al., 2022; Ruiz et al., 2021). There are still no reported phenotypes for the single deletions of the two smaller AsmA-like paralogues: *yhjG* and *yicH*.

In addition to publications specifically looking at the *asmA*-like genes in the presence of membrane stressors, there are also multiple high-throughput phenotype screens performed on entire libraries, such as the Keio collection of every viable *E. coli* single gene deletion. In these screens, a multitude of conditions are tested which can be used as a starting point for further genetic investigation (Nichols et al., 2011; Shiver et al., 2016; Price et al., 2018). These screens inspired the chemical genomics screen developed by Manuel Banzhaf's laboratory at the University of Birmingham (Figure 3.1A) (Doherty et al., 2023). This method utilises automated robots capable of pinning the strains of interest onto large quantities of agar plates with various conditions, reducing human error and allowing more repeats to be made. Strain fitness is then

assessed by colony morphology characteristics such as size, opacity and circularity. The mutant strains can then be compared to the behaviour of the WT strain for each condition plate and the resulting fitness ratio assessed, giving positive scores for increased fitness effects (pink) and negative scores for decreased fitness effects (blue).

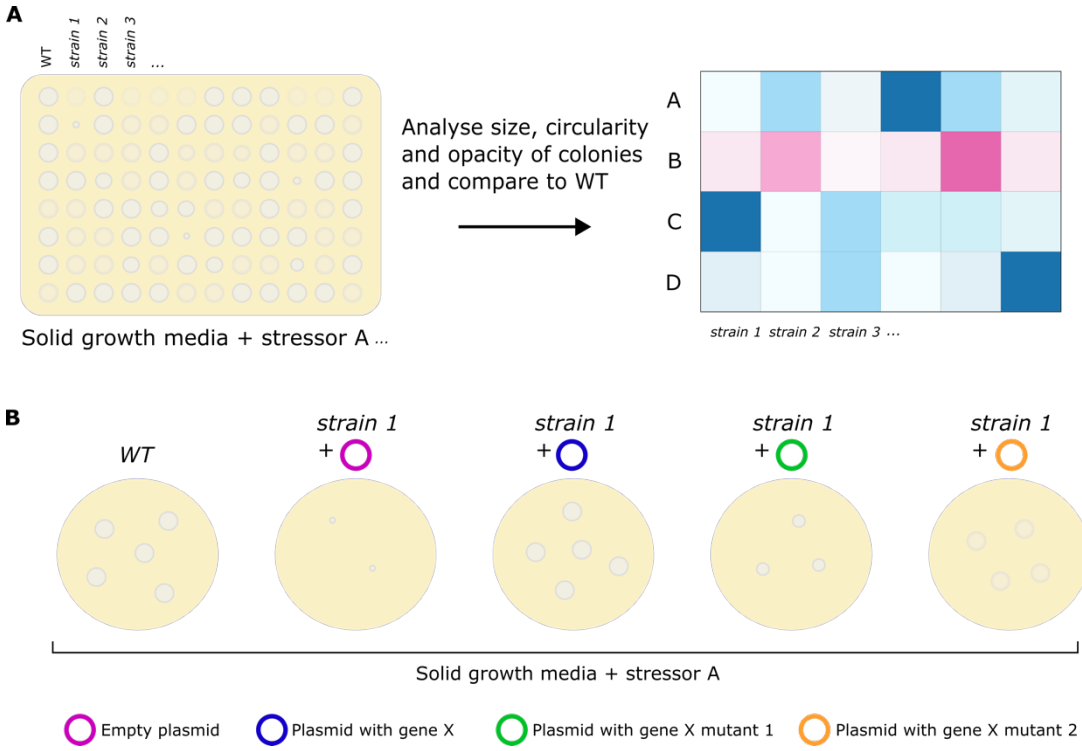


Figure 3.1. Phenotype screening and complementation assays. A – High-throughput phenotypic screening of strains on various stressors. Wild-type (WT) and mutant strains are pinned onto solid agar plates containing a stressor and the resulting colony size, circularity and opacity of each strain on each plate is assessed and compared to the WT, calculating a fitness ratio. An increase in fitness levels compared to the WT leads to a positive fitness ratio (pink), whereas a decrease in fitness levels leads to a negative fitness ratio (blue). B – Complementation assay. A growth condition that leads to a phenotypically different knockout strain compared to the WT is used to assess complementation. Reintroduction of the gene on a plasmid recovers the WT phenotype, whereas reintroduction of mutated genes on a plasmid may hinder the complementation. Level of complementation can inform residue importance.

3.1.2 Complementation assays

Aside from any inferences you can make from phenotypes just from the association of the deleted gene with increased or reduced fitness effects in a chemical or environmental stress,

you can also take advantage of a measurable phenotype to characterise the protein. Taking a gene deletion strain that is phenotypically different to the WT when exposed to a certain condition and then re-introducing the gene on a plasmid should recover the WT phenotype (i.e. complement it), thus confirming the gene-phenotype relationship (Figure 3.1B). The re-introduced gene can contain mutations that allow you to assess the contribution of individual residues or regions within a protein to the function of the protein in the cell. An important residue for protein function would likely weaken the complementation, whereas a residue of no functional importance would likely make no difference to the recovery of the WT phenotype. This is therefore a powerful method to determine regions of importance within a protein without having to overexpress or purify the protein *in vitro*.

3.1.3 Aims of genetic studies

This chapter focussed on two clear aims:

1. To evaluate the growth of each of the *asmA*-like gene single and double knockouts on a wider variety of conditions than have been previously reported to identify novel phenotypes and establish phenotypic profiles.
2. To characterise residues of functional importance using complementation assays, giving insight into protein function.

3.2 Results

3.2.1 Analysis of existing high-throughput screens

The existing high-throughput phenotype screens for *E. coli* single gene deletion mutants were used to influence the design of the in-house phenotype screen in 3.2.2. Many conditions that give a phenotypically different outcome compared to WT strains were found during database curation. A list of all conditions tested in high-throughput screens that gave an arbitrary score greater than a cut-off of ±2.5 for the six *asmA*-like genes are listed in Appendix 2. The list of conditions was then curated, looking at cost, availability and the mechanism of action of the chemical. For example, with the hypothesis that the *asmA*-like genes are involved in cell envelope biogenesis, it seemed unlikely that a ribosome-inhibiting chemical would be directly linked to the pathway of these genes. For that reason, only three protein synthesis inhibitors out of the many hits in the databases were used in the final screen. The final conditions taken forward for the in-house screen along with the genes that were affected in the high-throughput screens are shown in Table 3.1.

Table 3.1. Final conditions for in-house phenotype screen. Each condition is shown along with the type of stress induced and the associated hits from the high-throughput screens. Blue genes are those that had a fitness score lower than the cut-off of -2.5, pink genes are those that had a fitness score higher than the cut-off of 2.5. *ytfN* is the original gene name of TamB. Data curated from online databases (Nichols et al., 2011; Shiver et al., 2016; Price et al., 2018).

Stress	Type	Knockouts
SDS	Envelope stress	*yhdP*
SDS+EDTA	Envelope stress	*yhdP*
EDTA	Envelope stress	*yhdP*
MacConkey agar (bile salts)	Envelope stress	*asmA*, *yhdP*, *asmA*
Triton X-100	Envelope stress	Added due to envelope stress effect
Benzalkonium	Envelope stress	
A22	Cytoskeleton	*yhdP*, *asmA*, *ydbH*, *yhdP*
Vancomycin	Cell wall synthesis	*asmA*, *yhdP*
Ampicillin	Cell wall synthesis	*yhdP*
Bacitracin	Cell wall synthesis	*asmA*, *yhdP*
Temperature	Stress response	*asmA*, *yhdP*, *yicH*
NaCl	Osmotic stress	*asmA*, *yhdP*
Chloramphenicol	Protein synthesis	*asmA*, *ytfN*
Erythromycin	Protein synthesis	*asmA*, *yhdP*
Chloramphenicol+erythromycin	Protein synthesis	*asmA*

3.2.2 Phenotype screen

The designed in-house screen was then used to assess the phenotypes of *asmA*-like single and double knockout mutants (Ruiz et al., 2021). The resulting agar plates are shown in Figure 3.2 and Figure 3.3, with a key showing the order in which the 22 strains were pinned onto each agar plate. The heatmap in Figure 3.4 shows the colony sizes relative to the wild-type colony on that condition plate, going from smaller (blue) to larger (red) colonies.

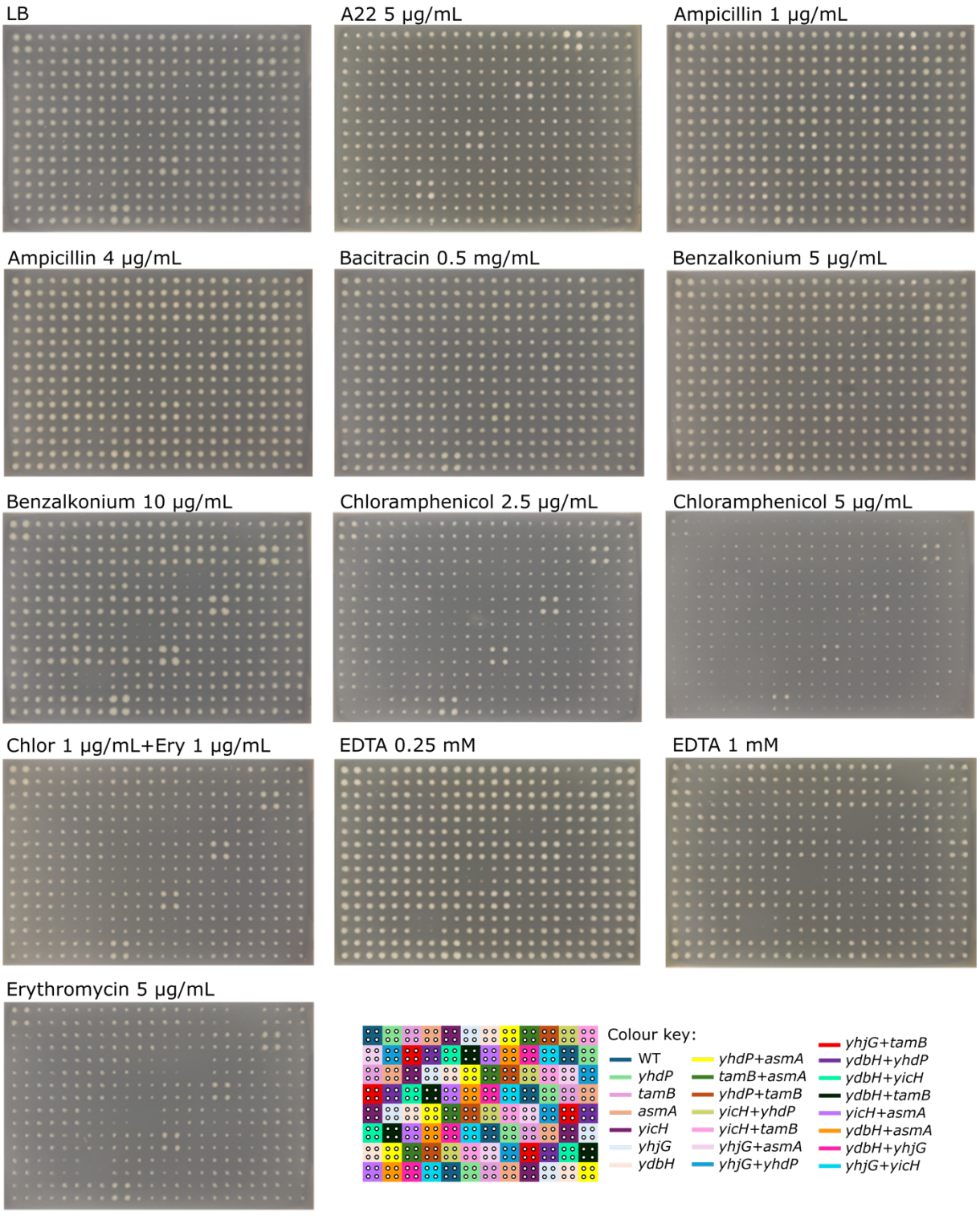


Figure 3.2. Phenotypic profiling plates. The *asmA*-like single and double knockout mutants pinned onto agar plates with various conditions for a phenotype screen. Each of the 22 strains were pinned as a group of four and repeated across the plate as the key shows. For clarity, the plate images were split across two figures.

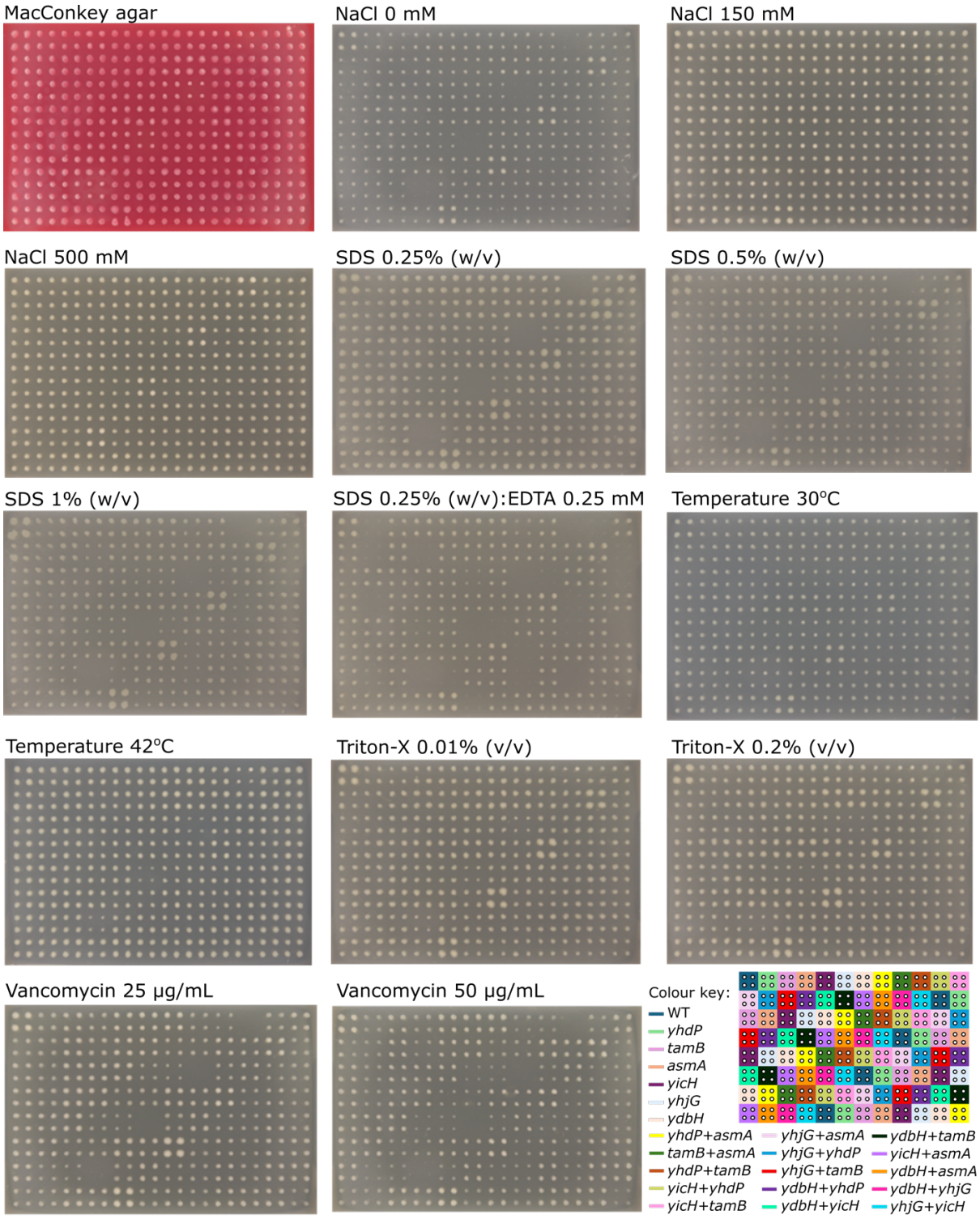


Figure 3.3. Phenotypic profiling plates (continued). The *asmA*-like single and double knockout mutants pinned onto agar plates with various conditions for a phenotype screen. Each of the 22 strains were pinned as a group of four and repeated across the plate as the key shows. For clarity, the plate images were split across two figures.

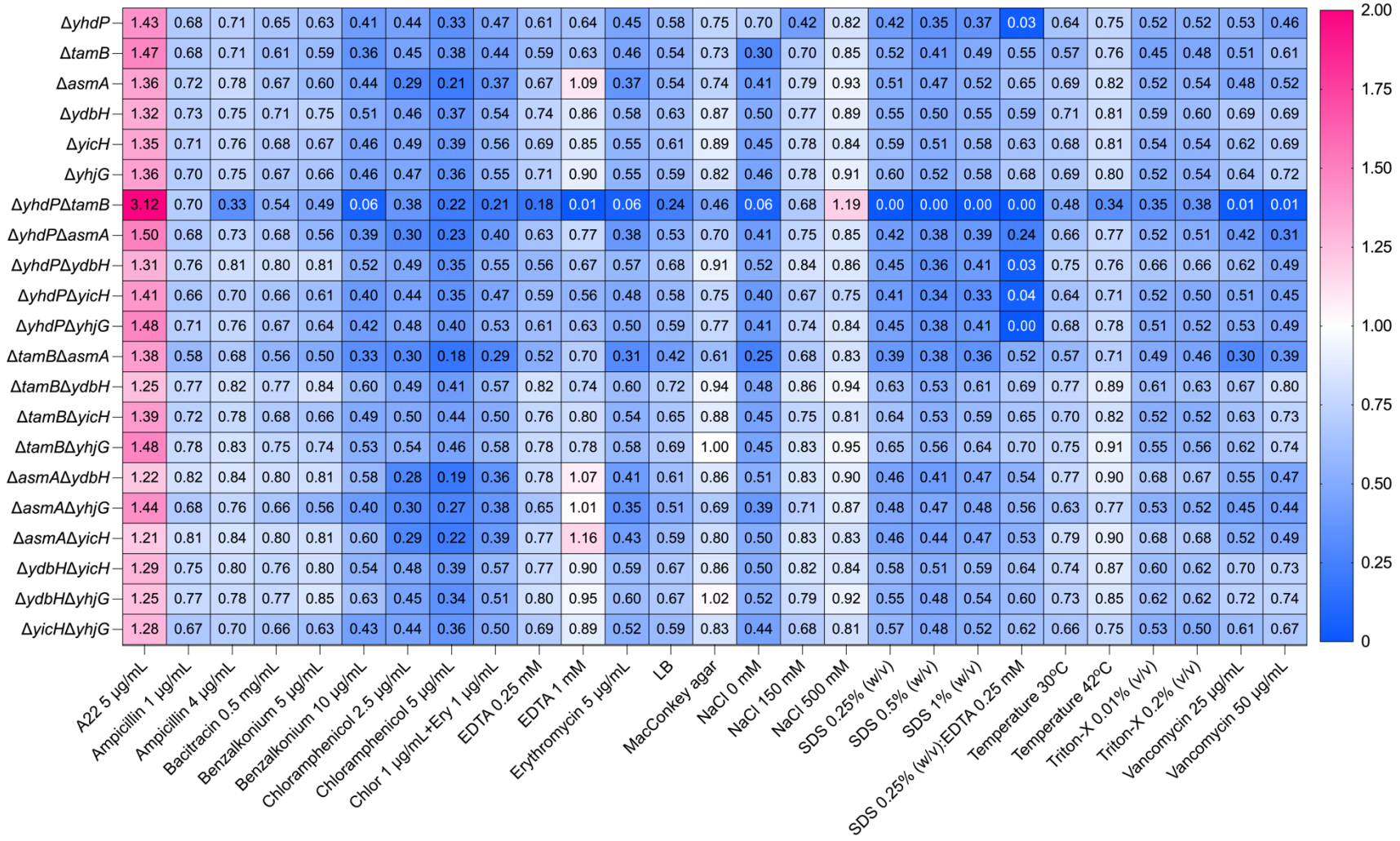


Figure 3.4. Heatmap of colony size distribution of mutants against wild-type cells across different conditions. Scale is relative to wild-type (WT) of the same condition, with colonies smaller than WT in blue and colonies larger than WT in pink.

There was a clear trend of severe fitness defects for the double knockout of Δ*yhdP*Δ*tamB* in a variety of stressors that indicate a perturbed cell envelope, namely SDS, EDTA, vancomycin, benzalkonium and ampicillin, and in the presence of the macrolide erythromycin and under osmotic stress due to no salt (Figure 3.4). The data support the finding that *yhdP* and *tamB* are functionally redundant, as the fitness defect is more severe than in the single knockouts of each gene. This function is likely important to cell envelope integrity due to the many membrane stressors in the screen perturbing cell growth. This aligns with the published studies that find *yhdP* and *tamB* to be important in outer membrane biogenesis (Ruiz et al., 2021; Douglass et al., 2022).

High salt conditions seemed to slightly recover the general fitness defects seen in the Δ*yhdP*Δ*tamB* double knockout, with growth on this condition resulting in larger colonies than on LB agar alone (Figure 3.4). High salt also resulted in less perturbed fitness of the other single and double knockout mutants than most other conditions.

Furthermore, in all mutants containing Δ*asmA*, there was increased sensitivity to chloramphenicol and erythromycin compared to the other knockouts (except the Δ*yhdP*Δ*tamB* knockout) (Figure 3.4). These chemicals are protein synthesis inhibitors, and it is likely that the reduction of LPS seen in Δ*asmA* mutants allows entry of these antibiotics, an effect that has been previously reported (Deng and Misra, 1996; Liu et al., 2010). There is further evidence of a link between *asmA* and LPS levels, as depletion of ions from exposure to EDTA, which would normally destabilise the LPS layer, resulted in slightly improved fitness compared to the wild-type and other knockouts (Figure 3.4).

Interestingly, there was a clear fitness gain when these strains were exposed to the antibiotic A22 (*S*-(3,4-dichlorobenzyl)isothiourea); a previously unreported phenotype (Figure 3.5A).

A22 binds MreB, the bacterial homologue of actin, which plays an important role in maintaining cell structure (Gitai et al., 2005). The binding of A22 prevents ATP-dependent polymerisation of MreB and causes the normally rod-shaped *E. coli* cell to become round (Awuni and Mu, 2019). A drastic increase in colony size of the Δ*yhdP*Δ*tamB* mutant, which were on average 3x as large as the wild-type strain, can be seen in Figure 3.5A. There could therefore be a link between MreB, the cytoskeleton and these two cell envelope proteins.

Another interesting outcome of this screen was the importance of *yhdP* when grown in 0.25% (w/v) SDS + 0.25 mM EDTA (Figure 3.5B). There was a general trend of fitness defects in this condition across all knockouts. In mutants without *yhdP*, there is a clear severe loss of fitness, except interestingly, in combination with Δ*asmA*. Here, whilst the colony size was still significantly smaller than WT, the loss of *asmA* seems to recover some fitness of the *yhdP* gene deletion. Thus, there is most likely some genetic interaction between *yhdP* and *asmA*.

Overall, different phenotypic profiles were seen for each of the mutants, indicating multiple roles for these genes, or at least, varying degrees of importance of a similar function for these genes. The most severe phenotypes were present when *yhdP* and *tamB* were both knocked out, which aligns with published studies, and allowed further study of these two genes via complementation assays.

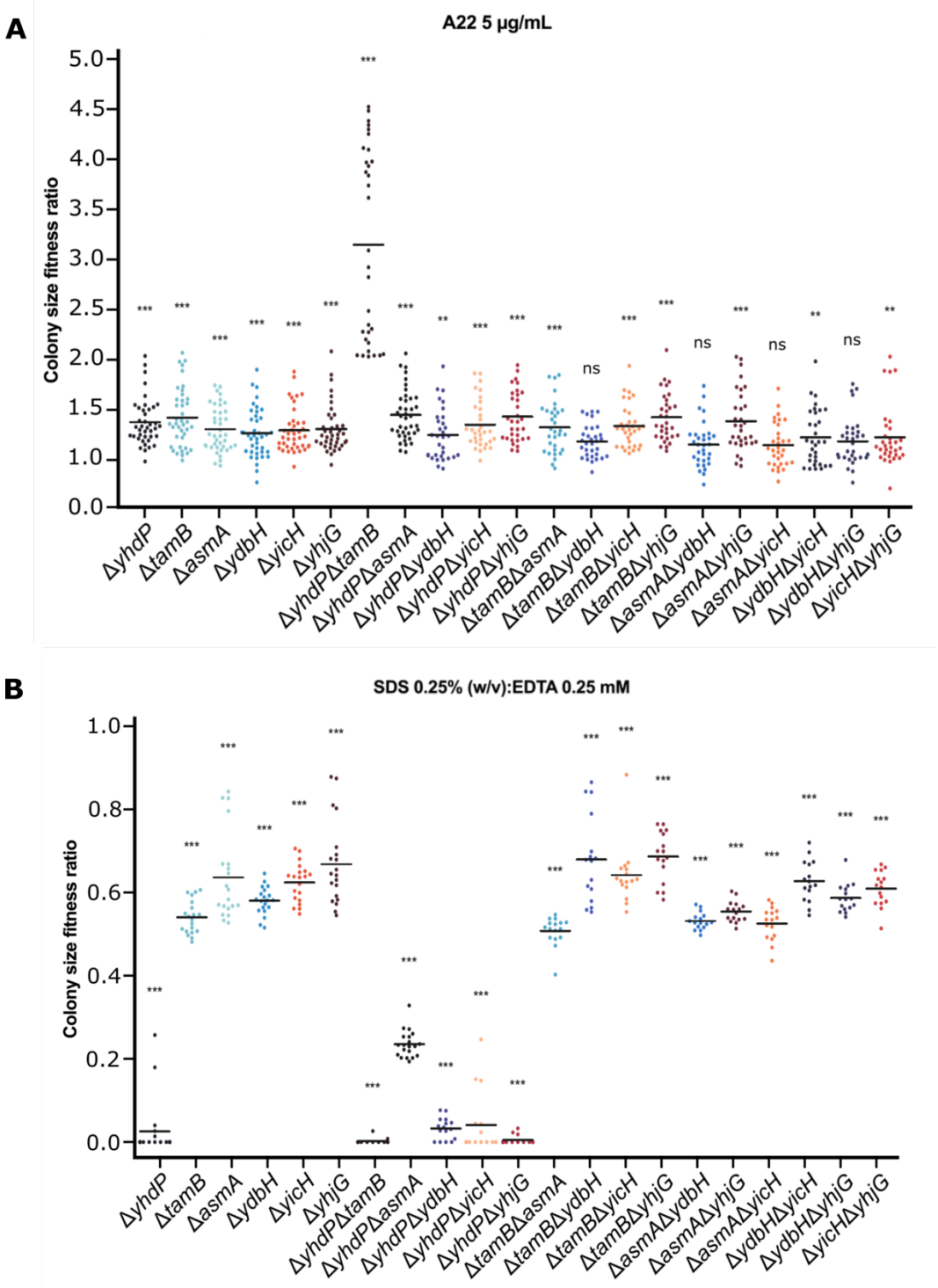


Figure 3.5. Colony size of the single and double knockouts of the *asmA*-like family compared to the wild-type. A - 5 µg/mL A22. B - 0.25% (w/v) SDS and 0.25 mM EDTA. Plots show replicates for each condition, with significance values representing significant difference to WT on same condition denoted such as ns: p > 0.05; *: 1.00e-02 < p <= 5.00e-02; **: 1.00e-03 < p <= 1.00e-02; ***: 1.00e-04 < p <= 1.00e-03; ****: p <= 1.00e-04.

3.2.3 Growth curve analyses of Δ*yhdP*Δ*tamB* with A22

The gain-of-fitness phenotype in the double knockout Δ*yhdP*Δ*tamB* when grown in the presence of the MreB-inhibiting antibiotic A22 was of particular interest. This mutant normally exhibits growth defects, even when grown on just LB. To further probe this interesting effect, OD_{600} growth curves, performed in triplicate, were assessed in the presence and absence of A22 (Figure 3.6). For the WT *E. coli* cells, the overall trend showed that increasing concentrations of A22 either killed or delayed the growth of cells (Figure 3.6A). The WT cells exhibited a significant lag period when exposed to the highest concentrations of A22 before growth started again after ~18 hours. The reason for this is unclear, but it could be the result of secondary mutations arising that can evade the cell-death phenotype, and that this cell line then outcompetes the susceptible cells. To confirm this, full genome sequencing would be required. Alternatively, A22 could be having a bacteriostatic effect rather than a bactericidal effect at these concentrations, as both effects have been previously reported (Kotzialampou et al., 2021).

The growth curves of Δ*yhdP*Δ*tamB* were then assessed in the same concentrations of A22 (Figure 3.6B). Here, the presence of A22 allowed a greater maximal OD_{600} to be reached compared to LB alone. Cell death then seemed to occur after this maximal OD_{600} was reached, with a faster decline of cell density as concentrations of A22 increased. This effect was not observed when assessing colony sizes and might only be present, or observable, in liquid culture media.

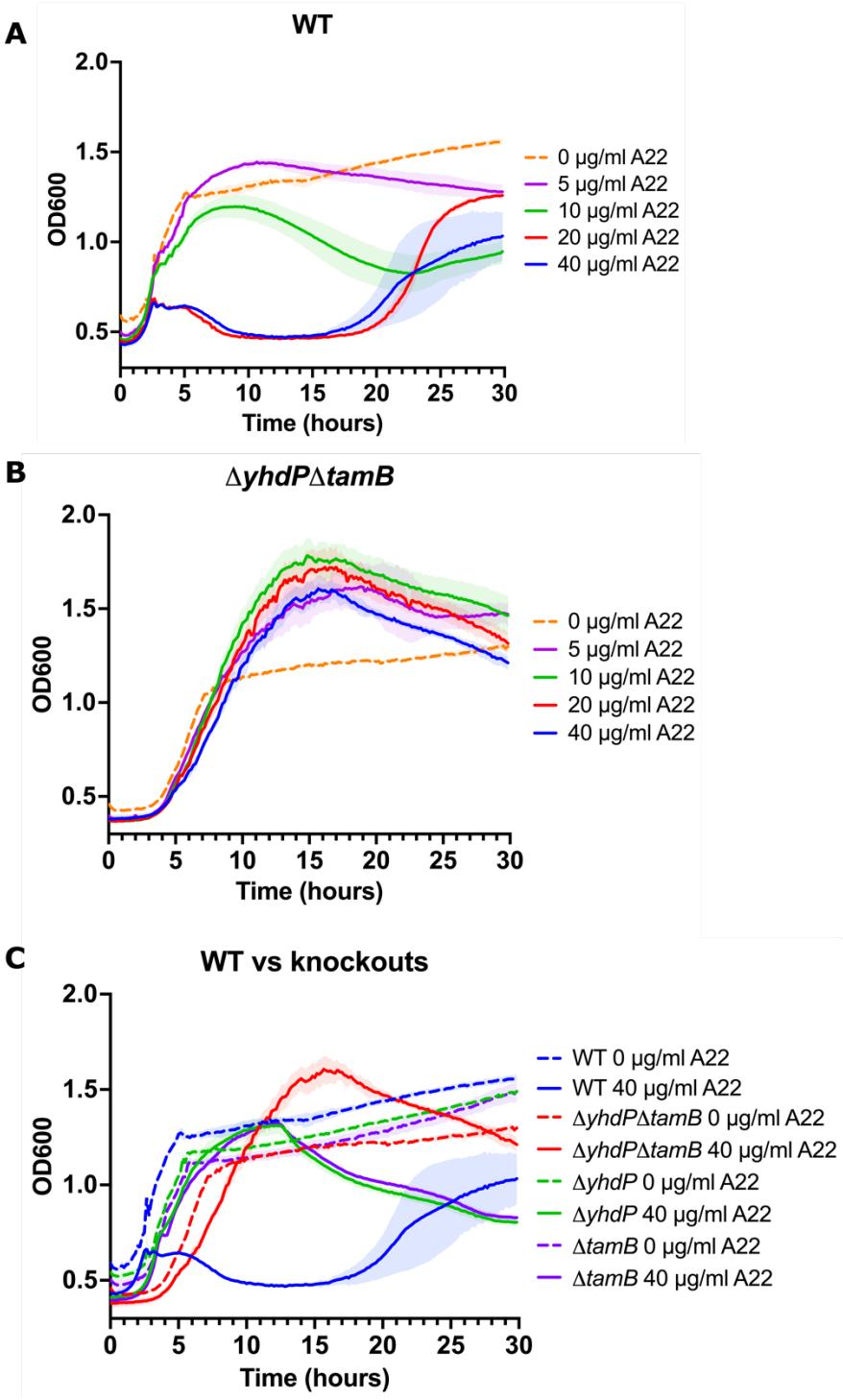


Figure 3.6. Growth curves of the *yhdP* and *tamB* mutants in the antibiotic A22. A – Wild-type (WT) *E. coli* MG1655 cells in concentrations of A22 from 0 µg/mL to 40 µg/mL. B - Δ*yhdP*Δ*tamB* cells in concentrations of A22 from 0 µg/mL to 40 µg/mL. C – WT, Δ*yhdP*, Δ*tamB* and Δ*yhdP*Δ*tamB* cells in the presence (40 µg/mL) and absence (0 µg/mL) of A22.

The growth curves for WT, Δ*yhdP*, Δ*tamB* and Δ*yhdP*Δ*tamB* at the highest concentration of A22 (40 μg/mL) and in the absence of A22 were then compared (Figure 3.6C). In LB media only (dashed lines), the mutants, and particularly the double knockout mutant, showed slightly delayed growth compared to the WT, and reached a lower maximum OD_{600}, which aligns with previously reported findings (Ruiz et al., 2021). In the presence of A22 (solid lines), the single knockout mutants grew similarly to LB only, whereas the double knockout mutant exhibited a slight lag compared to its growth in LB. As previously stated, the WT did not exhibit normal growth in this condition. The double knockout mutant reached a higher OD_{600} maximum in LB and A22 than both the single knockout mutants and WT cells. All mutants then started to die in LB and A22 after reaching the maximal OD_{600}, at ~12 hours for the single knockout mutants and ~17 hours for the double knockout mutant.

Why the mutant cells could grow in the presence of A22 and to a higher maximal OD_{600} than WT cells in LB is unclear, but the results do confirm a gain-of-fitness compared to its growth in LB only and some level of resistance to the effect of A22 when *yhdP* and *tamB* are knocked out, as the phenotype screen had shown.

3.2.4 Phase-contrast microscopy

A22 inhibits MreB polymerisation, which prevents the formation of the peptidoglycan cell wall during elongation, causing a loss of structural integrity (Gitai et al., 2005). This causes the cells to become swollen and round instead of rod-shaped, which is what leads to cell lysis of WT cells (Iwai et al., 2002). Due to this, it was of interest to see if the strong gain-of-fitness phenotype seen by *yhdP* and *tamB* mutants prevented the rounding effect of A22. This could provide information about whether the gain-of-fitness effect was due to preventing A22 interacting with MreB or if they were otherwise recovering the loss of MreB function. It may also explain the higher maximal OD_{600} that was achieved in the growth curves of the knockouts,

as round cells would likely scatter more light and therefore give a larger signal compared to the same number of rod-shaped cells. Therefore, to assess cell shape in the presence of A22, phase-contrast microscopy of the WT, Δ*yhdP*, Δ*tamB* and Δ*yhdP*Δ*tamB* cells was performed.

The phase contrast microscopy images are shown in Figure 3.7 and violin plots of the calculated cell lengths, areas and circularities for each strain are shown in Figure 3.8, with the mean values listed in Table 3.2. Both the WT and single mutant strains showed normal rod-shaped cells in LB, with shorter and rounder cells (circularity closer to 1.0) seen for the double knockout (Figure 3.7 and Figure 3.8). This phenotype has been previously reported and characterised (Ruiz et al., 2021). In the presence of A22, cells became round in both the WT and the knockout mutants. This implies that A22 was still able to target MreB in the mutants, but that the cell death induced by the loss of structural integrity was being prevented or delayed somehow. Interestingly, in A22, the WT cells were significantly larger than the single and double knockout cells, with an average length of 3.5 ± 0.9 μm and an average area of 7.1 ± 3.0 μm^2 compared to the average length of 2.4 ± 0.6 μm and area of 3.4 ± 1.0 μm^2 for the double knockout (Figure 3.8, Table 3.2). The double knockout cells were also longer when grown in A22 compared to LB only, with an average length of 2.4 ± 0.6 μm with A22 and 1.9 ± 0.3 μm without (Figure 3.8, Table 3.2).

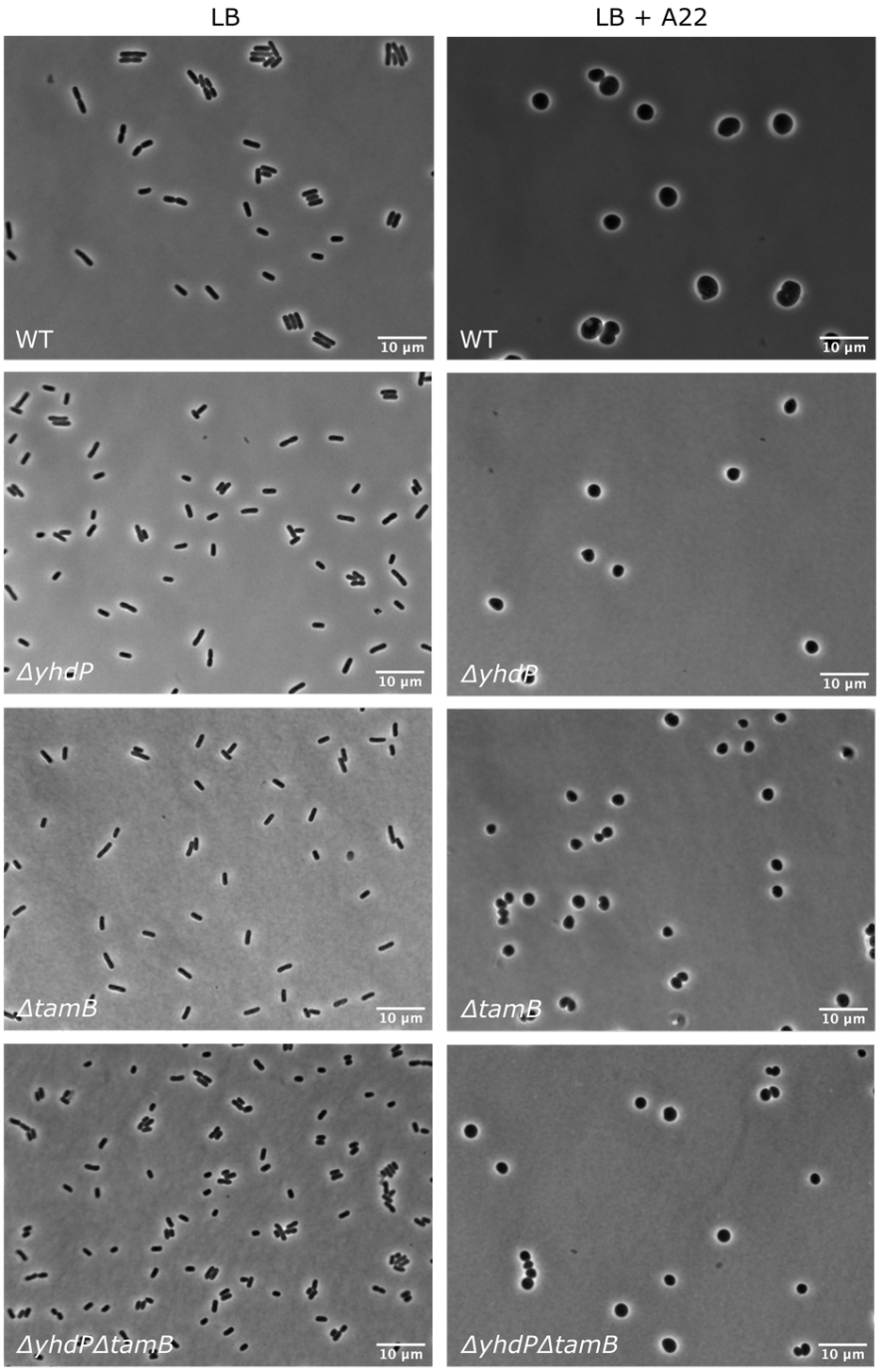


Figure 3.7. Phase contrast microscopy. Micrographs of the WT, single and double knockouts of *yhdP* and *tamB* grown in LB media or LB media with 40 μg/mL A22. The cell rounding effect of A22 is present in the WT and mutants. Scale bar = 10 μm.

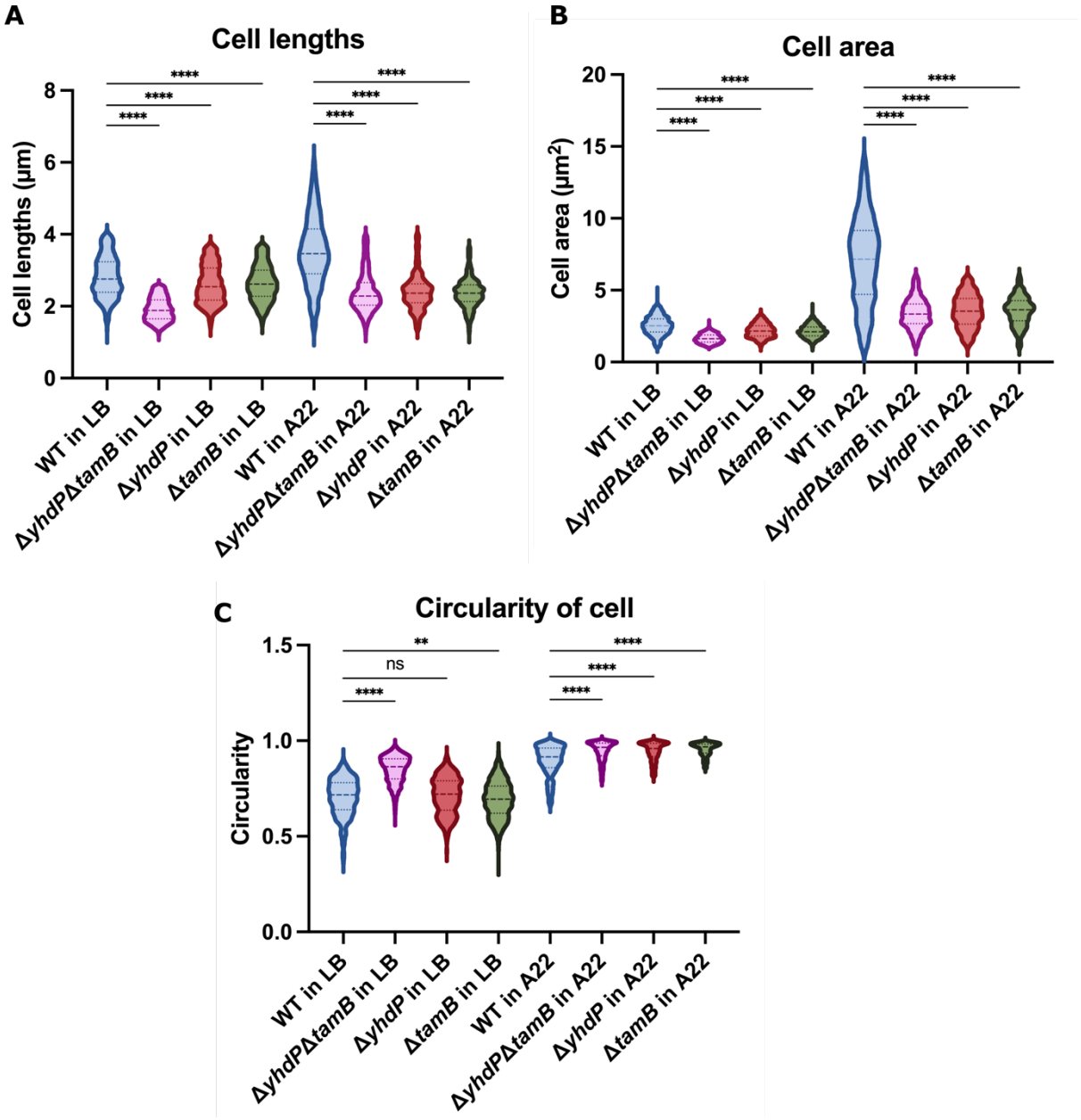


Figure 3.8. Cell shape analysis. Cell shape parameters were analysed from the micrographs of the WT, single and double knockouts of *yhdP* and *tamB* in LB or LB with 40 µg/mL A22. A – Cell lengths. B – Cell area. C – Cell circularity. MicrobeJ plugin on ImageJ used for cell shape analysis. Significance values denoted such as *: 1.00e-02 < p <= 5.00e-02; **: 1.00e-03 < p <= 1.00e-02; ***: 1.00e-04 < p <= 1.00e-03; ****: p <= 1.00e-04; ns: no significance.

Table 3.2. Cell shape analysis of WT and knockout mutants with and without A22. Cell length, area and circularity were measured from the phase contrast microscopy of wild-type (WT), Δ*yhdP*Δ*tamB*, Δ*yhdP* and Δ*tamB* cells in LB or LB with 40 µg/mL A22.

Genotype	A22	No. of cells	Average length (µm)	Average area (µm^2)	Average circularity
WT	-	1017	2.8 ± 0.6	2.5 ± 0.7	0.7 ± 0.1
	+	565	3.5 ± 0.9	7.1 ± 3.0	0.9 ± 0.1
ΔyhdPΔtamB	-	3207	1.9 ± 0.3	1.7 ± 0.3	0.9 ± 0.1
	+	992	2.4 ± 0.6	3.4 ± 1.0	1.0 ± 0.1
ΔyhdP	-	845	2.6 ± 0.5	2.2 ± 0.5	0.7 ± 0.1
	+	472	2.4 ± 0.5	3.6 ± 1.2	0.9 ± 0.1
ΔtamB	-	1213	2.7 ± 0.5	2.2 ± 0.4	0.7 ± 0.1
	+	616	2.4 ± 0.4	3.6 ± 1.0	1.0 ± 0.0

3.2.5 TamB complementation assay

Another outcome from the phenotype screen was the severe fitness defect of Δ*yhdP*Δ*tamB* cells in LB with the presence of the membrane stressors SDS and EDTA, a phenotype since reported in other published studies (Douglass et al., 2022). Phenotypic complementation assays were performed utilising this measurable phenotype whereby residues of importance for TamB function could be determined through the recovery of the phenotype.

Current evidence suggests that TamB either has a role in transporting phospholipids or OMPs from the IM to the OM of the cell envelope (Shen et al., 2014; Ruiz et al., 2021). The hydrophobic cavity throughout the AlphaFold model of TamB was proposed to provide a conduit across the periplasm for these amphipathic substrates. To test this, mutants were designed based on the AlphaFold structure. Whilst this is only a model and not experimentally determined, the high pLDDT scores seen throughout indicate a high degree of confidence in the model, particularly in the periplasmic β-taco region (Figure 3.9). If the substrates (i.e.

phospholipids or OMPs) are protected within this cavity, then introducing charged residues here may inhibit transport. Therefore, mutants were designed to change hydrophobic side chains (leucines) within the N-terminal cavity entrance to bulkier, negatively charged side chains (glutamates), which should block the movement of the substrate into the cavity. Site-directed mutagenesis of these mutants was only partially successful, with the two successful mutants, L77E and L96E, shown in Figure 3.9.

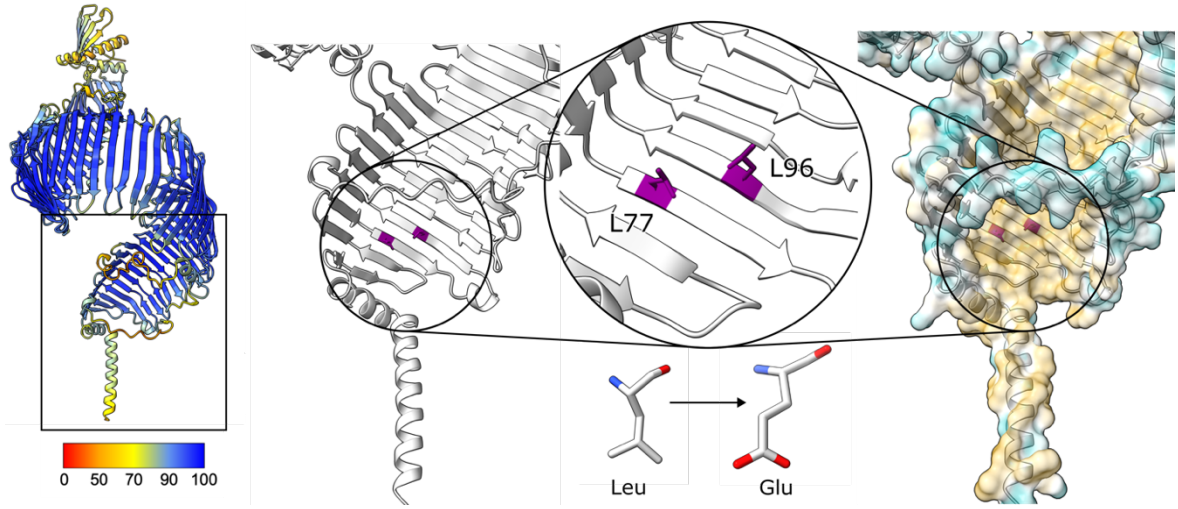


Figure 3.9. TamB functional mutations. The AlphaFold model of TamB is shown along with the pLDDT key (red to blue, increasing model confidence). The mutants L77E and L96E were designed to place a charged, bulky residue within the predicted hydrophobic cavity entrance where amphiphilic substrates are proposed to enter for transport through the cavity. Hydrophobic surfaces in yellow, hydrophilic surfaces in blue. Leucine and glutamate residues shown to indicate change of side chain size and charge after mutation.

Growth on LB agar was similar between all the strains, however in the presence of SDS and EDTA, severely perturbed growth was seen for the double knockout and for the mutants TamB L77E and L96E. Leaky expression of TamB WT recovered the growth phenotype, confirming the gene-phenotype relationship. This supports the finding that YhdP and TamB exhibit functional redundancy, as re-introduction of only one of the genes recovered the severe fitness defect (Ruiz et al., 2021; Douglass et al., 2022). The SDS+EDTA phenotype of TamB L77E was also more severe than that of TamB L96E, suggesting that the L96E mutation was better

tolerated and less detrimental to TamB function. Western blotting revealed comparable expression levels of the TamB WT and variants, indicating the growth defects were indeed due to impaired protein function and not protein instability (Figure 3.10).

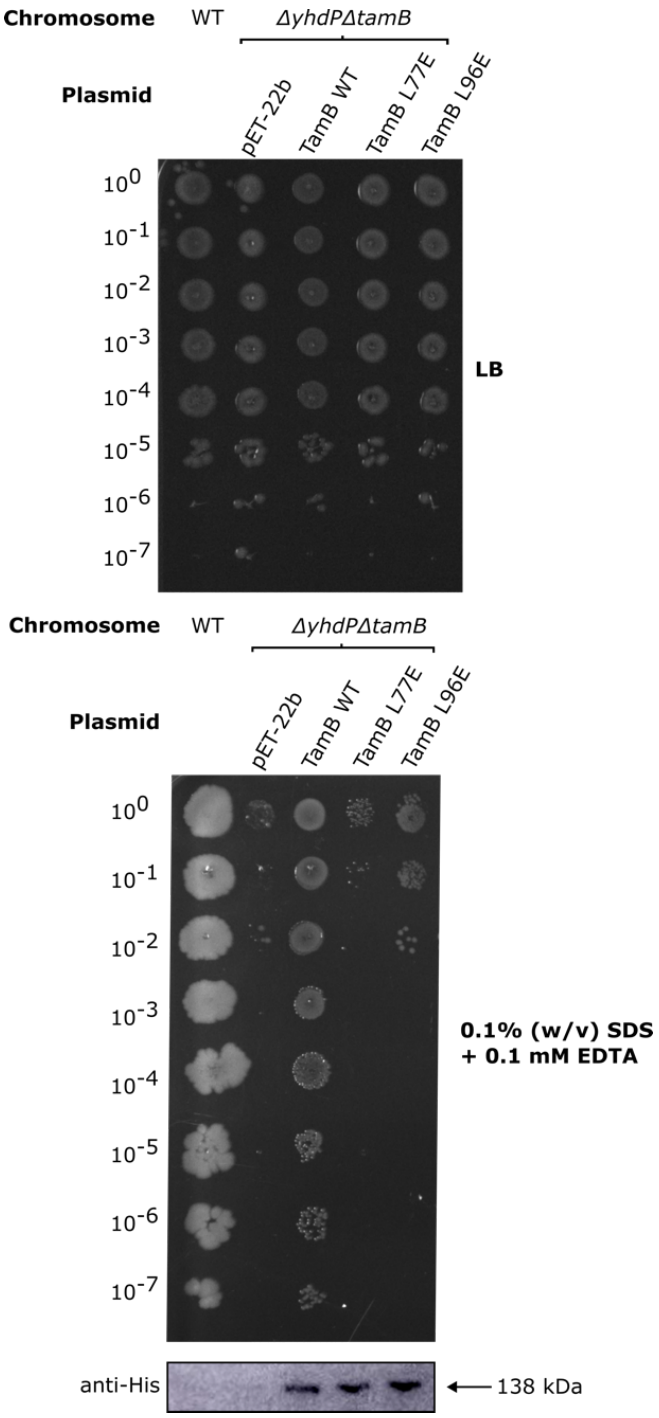


Figure 3.10. Complementation of TamB into Δ*yhdP*Δ*tamB* knockout cells. *E. coli* MG1655 WT cells and Δ*yhdP*Δ*tamB* knockout cells transformed with empty pET-22b, pET-22b with TamB WT, pET-22b with TamB L77E or pET-22b with TamB L96E were plated on LB agar (top) or LB agar supplemented with 0.1% (w/v) SDS and 0.1 mM EDTA (bottom). Anti-His western blot showing comparable expression levels of TamB variants.

3.2.6 YhdP complementation assay

The same approach as for the TamB complementation assay was taken for investigating YhdP function. YhdP has a proposed role in phospholipid transport from the IM to the OM with the hydrophobic cavity predicted by the AlphaFold model suggested to protect the acyl chains of phospholipids as they cross the periplasm (Cooper et al., 2023). The high pLDDT scores for the AlphaFold model of YhdP suggest a high degree of confidence in the model, particularly within the periplasmic region, which allowed the design of potential non-functional mutants (Figure 3.11). Akin to the rationale behind the design of TamB mutants, hydrophobic residues within the cavity entrance were mutated to larger, charged residues to potentially block the entrance of hydrophobic phospholipid tails. Site-directed mutagenesis was successful for the mutation of two leucines at the entrance of the cavity to negatively charged glutamates (L68E and L90E). A mutant was also made with the N-terminal helix removed to assess the requirement of IM association at the N-terminus of YhdP for its function (YhdP 30-1266). Attempts at creating a mutant of a potential redox active site within the structure, with the classic CXXC motif (C999-T1000-E1001-C1002), were made but were not successful (Quan et al., 2007).

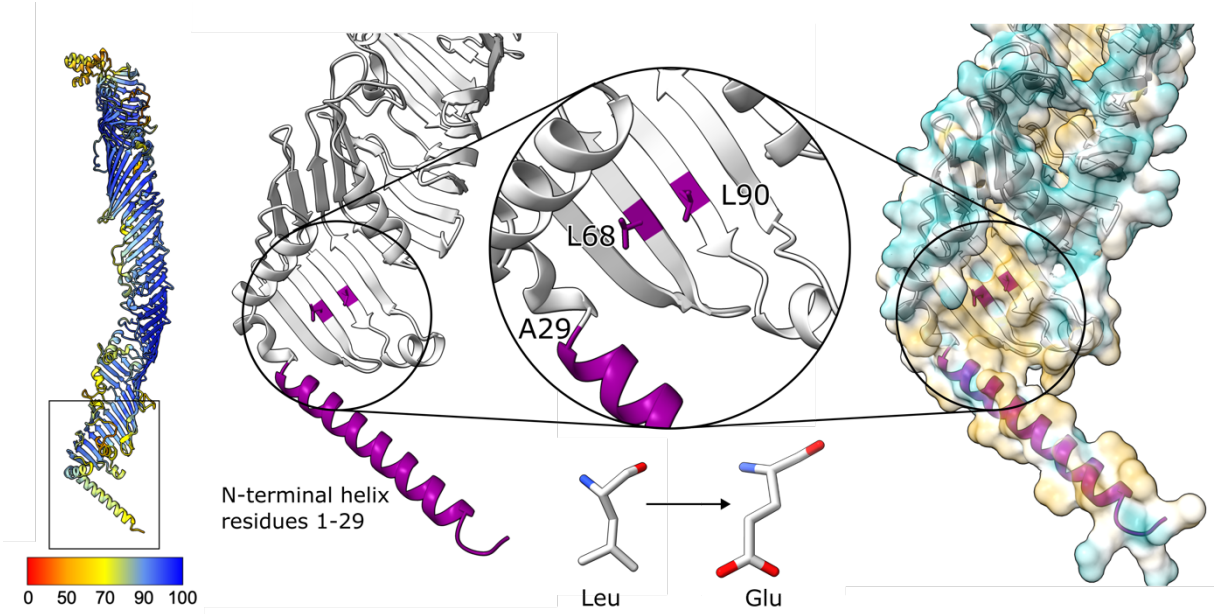


Figure 3.11. YhdP functional mutations. The AlphaFold model of YhdP is shown along with the pLDDT key (red to blue, increasing model confidence). The mutants L68E and L90E were designed to place a charged, bulky residue within the predicted hydrophobic cavity entrance where phospholipids are proposed to enter for transport through the cavity. The N-terminal helix (residues 1-29) was also removed to assess its requirement for YhdP's IM association and function. Hydrophobic surfaces in yellow, hydrophilic surfaces in blue. Leucine and glutamate residues shown to indicate change of side chain size and charge after mutation.

The strains all grew similarly on LB agar, and in the presence of SDS and EDTA, reintroduction of YhdP WT fully recovered the growth phenotype, again confirming the gene-phenotype relationship and supporting the functional redundancy between YhdP and TamB (Figure 3.12). Interestingly, complementation with YhdP 30-1266, YhdP L68E and YhdP L90E resulted in only slightly compromised growth when compared to YhdP WT. This indicated that YhdP can mostly function without its N-terminal anchor, potentially associated to the IM by other means or not needing the IM. Similarly, the mutations L68E and L90E were well-tolerated, suggesting that these residues are not important for YhdP function. Western blotting revealed comparable expression levels between the variants, confirming that observed differences in viability are due to protein function rather than instability.

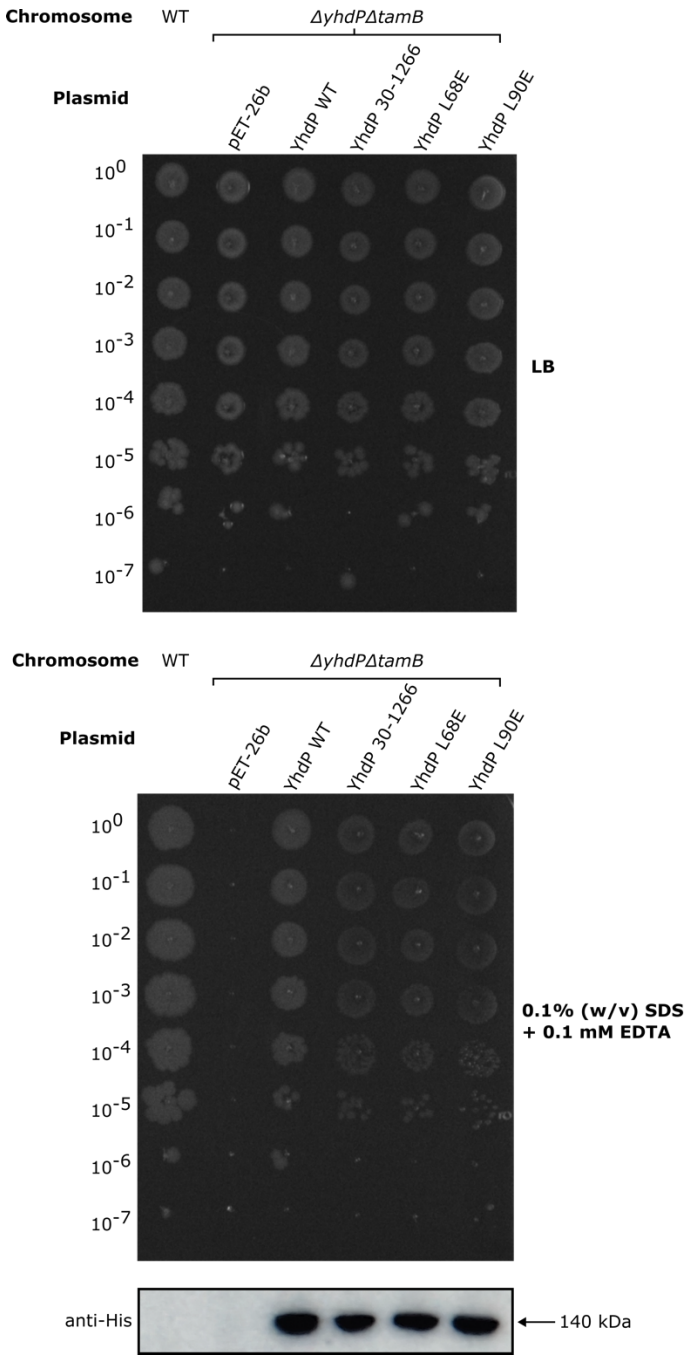


Figure 3.12. Complementation of YhdP into Δ*yhdP*Δ*tamB* knockout cells. *E. coli* MG1655 WT cells and Δ*yhdP*Δ*tamB* knockout cells transformed with empty pET-26b, pET-26b with YhdP WT, pET-26b with YhdP 30-1266, pET-26b with YhdP L68E or pET-26b with YhdP L90E were plated on LB agar (top) or LB agar supplemented with 0.1% (w/v) SDS + 0.1 mM EDTA (bottom). Anti-His western blot showing comparable expression levels of YhdP variants.

3.3 Discussion

In this chapter, the main aim was to find novel phenotypes of the single and double knockout mutants of the *E. coli* AsmA-like family. Some of the findings of this screen have since been supported by evidence from recent publications, with reported susceptibility to membrane stressors like SDS, EDTA and vancomycin for Δ*yhdP* mutants (Ruiz et al., 2021; Douglass et al., 2022; Cooper et al., 2023). This susceptibility suggests that YhdP is important in the biogenesis and/or maintenance of the cell envelope, which aligns with recent evidence proposing that YhdP has a role in phospholipid transport to the OM. Single mutants of the other *asmA*-like genes did not show as severe fitness defects in the conditions tested, indicating that these genes either play a less crucial role in the cell, have an alternate role that is not yet understood or selected for via the phenotypic screening performed here, or perhaps that compensation mechanisms are stronger for these genes.

3.3.1 Identification of a link between AsmA and YhdP using phenotype screens

The combination of any of the single gene knockouts with Δ*yhdP* revealed the same severe susceptibility to SDS and EDTA, except for Δ*yhdP*Δ*asmA*, where the fitness was slightly recovered and small colonies formed. This finding suggests that there is some genetic interaction between these genes and that the functions of AsmA and YhdP could be opposing. A small fitness recovery for Δ*yhdP* mutants when *asmA* was knocked out in the presence of polymyxin has also since been reported, but no explanation for the recovery was given (Douglass et al., 2022).

yhdP was first implicated in anterograde phospholipid transport due to the recovery of fitness when deleted in the *mlaA** mutant, where aberrant phospholipid flipping to the outer leaflet of the OM occurs (Grimm et al., 2020). The loss of *yhdP* delayed cell death in this mutant,

reminiscent to the loss of *asmA* recovering fitness in the Δ*yhdP* mutant. Perhaps then, AsmA is involved in retrograde phospholipid transport from the OM. However, this seems unlikely as the predicted AlphaFold structure of AsmA seems too short (~140 Å) to interact with both membranes. However, there could be times where the periplasm is smaller, allowing engagement with both membranes, for example during cell division. A similar suppression of Δ*yhdP* OM permeability was also seen with the deletion of *wzzE*, a protein involved in the synthesis of cyclic ECA (Mitchell et al., 2018). Perhaps AsmA is therefore involved in the biogenesis or transport of cyclic ECA, and its subsequent loss recovers *yhdP* sensitivity.

Other known phenotypes of Δ*asmA* are the suppression of OmpF/C mutant assembly and reduced levels of LPS in the OM (Misra and Miao, 1995; Xiong et al., 1996; Deng and Misra, 1996). If YhdP is an anterograde phospholipid transporter, then a reduction in LPS may enable cell viability in Δ*yhdP* mutants due to reduced levels of both the inner and outer leaflet components, maintaining a pseudo-homeostasis of the OM. Whereas with *asmA* present and normal LPS levels within the OM, the cell may not be able to cope with the demand of trafficking phospholipids to the OM to keep up with LPS insertion. Further evidence for this link has since been reported with mutations in lipid A synthesis genes recovering the resistance to vancomycin and polymyxin antibiotic in Δ*yhdP* mutants (Douglass et al., 2022). Whilst lower LPS levels are a likely mechanism, further research into this genetic interaction is required to determine how exactly loss of *asmA* contributes to the fitness recovery of Δ*yhdP* in membrane stressing conditions.

3.3.2 The novel A22 phenotype in Δ*yhdP*Δ*tamB*

The antibiotic A22 caused an interesting gain-of-fitness phenotype in Δ*yhdP*Δ*tamB*. Growth curve analyses showed that the single and double knockout mutants could grow in A22, whereas

the WT cells had severely delayed growth. The microscopy then revealed that A22 still caused cells to become round through the inhibition of MreB, part of the elongasome MreBCD complex and responsible for directing peptidoglycan synthesis machinery to a growing cell wall, and that these cells were much smaller than the WT in A22 (Kruse et al., 2005). Therefore, it is likely that the absence of YhdP and TamB, by an unknown mechanism, enables the *E. coli* cell to survive better when new peptidoglycan synthesis is inhibited despite the increased intracellular turgor pressure of a round cell.

Interestingly, the STRING database suggests there may be a functional link between *yhdP* and another component of the elongasome complex, *mreD*, due to their close intergenic distance (2127 bp) and from homologous genes being neighbours in other genomes (Szklarczyk et al., 2021). Similarly, another gene within the *mre* operon, *yhdE*, is also suggested to have a functional link with *yhdP* (intergenic distance = 1525 bp). This gene encodes a nucleotide pyrophosphatase that has a role in maintaining cell shape, with *yhdE* null mutants also exhibiting a more spherical cell shape and showing increased growth under very high salt concentrations, which is reminiscent of the slightly improved growth seen across all *asmA*-like mutants in 500 mM NaCl reported in this study (Jin et al., 2015). Whether there is indeed a link between YhdE, YhdP and the other AsmA-like proteins should be further investigated.

Overcoming the A22 phenotype has also been seen in slow growing cells, cells overexpressing the cell division proteins *ftsZAQ* and in TCA cycle mutants (Kruse et al., 2005; Bendezú and de Boer, 2008; Barton et al., 2021). Perhaps YhdP and TamB overcome A22 in a similar fashion, likely due to the Δ*yhdP*Δ*tamB* mutants being slow growing (which can be seen in the slight lag before exponential growth in the growth curves). With proposed roles in anterograde phospholipid transport, this slow growth is likely caused by slow membrane biogenesis from their absence. The slower OM formation may also feedback into the cell that less new

peptidoglycan needs to be synthesised, which would likely lessen the effect of MreB inhibition. Alternatively, perhaps there is a link to central metabolism in this mutant that indirectly overcomes the disruption of MreB, akin to the accumulation of cell wall precursors in TCA cycle mutants (Barton et al., 2021).

The result of A22 causing an increased fitness phenotype in Δ*yhdP*Δ*tamB* is incredibly interesting but requires further investigation to determine how exactly this effect is happening. One way to do this would be to use transposon directed insertion sequencing (TraDIS), a screening technique that looks at the positions of transposon insertion events in the genome of a strain under a particular growth condition. The lack of transposons inserted into some genes would indicate the essentiality of those genes for that strain in that growth condition. Therefore, the double knockout could be grown with and without A22 and the transposon insertions could be compared to the WT strain, providing details of those genes showing essentiality for the knockout's fitness gain. It would also be prudent to investigate the proteome and metabolite composition of the strain after growth in different conditions to identify differentially expressed genes and changes in membrane composition that may give an insight into the gain-of-fitness phenotype. Similar experiments could also be utilised to investigate the effect of *asmA* deletion on Δ*yhdP* mutants under SDS+EDTA conditions.

3.3.3 Determining residues of importance of YhdP and TamB

The successful complementation of *yhdP* or *tamB* into the Δ*yhdP*Δ*tamB* background confirmed the gene-phenotype relationships seen and allowed testing of the importance of specific residues for protein function. In this preliminary study, the mutants tested assessed the importance of the hydrophobic cavity at the potential site of entry of phospholipids or other amphipathic substrates from the IM by mutating hydrophobic residues into charged residues. In TamB, these mutations almost completely abolished growth, indicating that the presence of a single charged

residue within the hydrophobic cavity entrance impaired normal function, with an increased severity seen for L77E compared to L96E. This could be due to L77E likely being nearer to the IM, potentially preventing substrate entry more effectively than L96E.

The equivalent mutations in YhdP were more tolerated, with growth in these mutants almost the same as the reintroduction of YhdP WT. Perhaps this tolerance is due to a lower energy barrier for substrates to enter the cavity from the IM in YhdP compared to TamB, outcompeting any detrimental effect of the single point mutation, or perhaps this is due to ligand preference. Recent studies suggest a preference of TamB for unsaturated lipids, and it could be that these mutations are more preventative of unsaturated lipids moving into the cavity than saturated (Rai et al., 2024).

YhdP also functioned without its N-terminal helix which was thought to anchor the protein to the IM. This finding has since been reported in similar complementation assays using Δ*yhdP* knockouts, where interestingly, they then found a more detrimental effect with the removal of a small helix parallel to the membrane immediately following the N-terminal helix (named 'P-helix') (Cooper et al., 2023). The authors then report that in molecular dynamics simulations of YhdP without its N-terminal helix, this P-helix allowed the association of the N-terminus of the protein with a simulated bilayer. This is likely the reason that removal of the N-terminal helix did not abolish protein function here. This finding also implies that the N-terminal helix can be removed for *in vitro* structure/function analysis of the protein without being detrimental to protein function, which would likely improve its solubility and amenability to a variety of biochemical and biophysical techniques. It should also be noted that similar removal of the N-terminal helix was attempted for TamB, but that the site-directed mutagenesis was not successful. Future experiments should identify other residues of interest within TamB and YhdP utilising the complementation assay presented here. Particularly interesting could be the C-

terminal regions which are proposed or known to interact with the OM or an OM partner, to see if function is impaired (Cooper et al., 2023; Shen et al., 2014).

3.3.4 Phenotypes of the other *asmA*-like genes

Whilst phenotypes of the other *asmA*-like gene knockouts were not as obvious, the screening data does suggest that these mutants may be sensitive to benzalkonium, chloramphenicol and no salt conditions and exhibit slightly improved growth in A22. Whilst this was not further investigated in this study due to the more obvious phenotypes seen in *yhdP* and *tamB* mutants, future work should endeavour to investigate these other *asmA*-like phenotypes. Increasing the concentration of stressor may reveal a more severe phenotype, which would allow validation of the gene-phenotype relationship with complementation assays.

3.3.5 Conclusion

In this chapter, the phenotypic profiles of the *asmA*-like family in *E. coli* were determined across a variety of conditions and can now be used to influence the design of future investigations. Particularly interesting findings from the screen included the ability of Δ*asmA* to suppress the Δ*yhdP* SDS+EDTA sensitive phenotype and the increased fitness seen in the normally defective Δ*yhdP*Δ*tamB* mutant in the MreB-acting antibiotic A22 compared to the wild-type. Phenotypic complementation assays determined that residues at the hydrophobic cavity entrance were necessary for TamB function, but that equivalent residues and the N-terminal helix in YhdP were not required for its function. This preliminary work paves the way to understanding the function of all six AsmA-like proteins. Future experimentation should focus on understanding the potential genetic interactions and novel phenotypes determined here, and on identifying more severe phenotypes for the other *asmA*-like genes. These can then

be utilised in complementation assays to test residue importance and ultimately gain insight into the function of each family member.

CHAPTER 4. PURIFICATIONS OF THE *E. COLI* ASMA-LIKE PROTEINS

4.1 Introduction

With the initial broad aim of elucidating the structures and functions of the AsmA-like proteins, a large proportion of the project was spent attempting to purify the AsmA-like proteins to a high enough yield and purity for structural techniques like X-ray crystallography and cryo-EM. With all six family members being membrane-bound and having a large periplasmic domain that is predicted to contain a highly hydrophobic cavity throughout its length, the purification of these proteins was likely going to be challenging. And, as this chapter will describe, working with these proteins was indeed complex for a multitude of reasons.

4.1.1 Optimisation of protein expression and purification

Recombinant protein overexpression in *E. coli* can be optimized through the trialling of both expression constructs and conditions (Figure 4.1) (Joseph et al., 2015). Construct design involves choosing an appropriate vector backbone, determining optimal locations of purification tags, and choosing whether to include signal peptides, which are needed for some membrane and secretory proteins. Additionally, truncation constructs can be designed by removing non-essential or problematic regions of the protein and/or solubility tags can be added to improve protein stability and solubility. Optimizing expression conditions typically involves trialling different expression cell lines and culture media, concentration of the inducer (e.g. IPTG), point of induction and duration of expression.

Once optimal expression conditions are established, the protein is extracted and purified (Figure 4.1). For soluble proteins, this process involves lysing the cells, capturing the protein, and proceeding with purification. For membrane proteins, the membrane fraction is isolated, before

solubilisation using detergents or encapsulation agents, which enables their subsequent purification (Ratkeviciute et al., 2021). This typically involves initial protein capture by affinity chromatography followed by size exclusion chromatography (SEC). Optimisation of buffer conditions and chromatography protocols can significantly enhance the final protein yield, purity and stability.

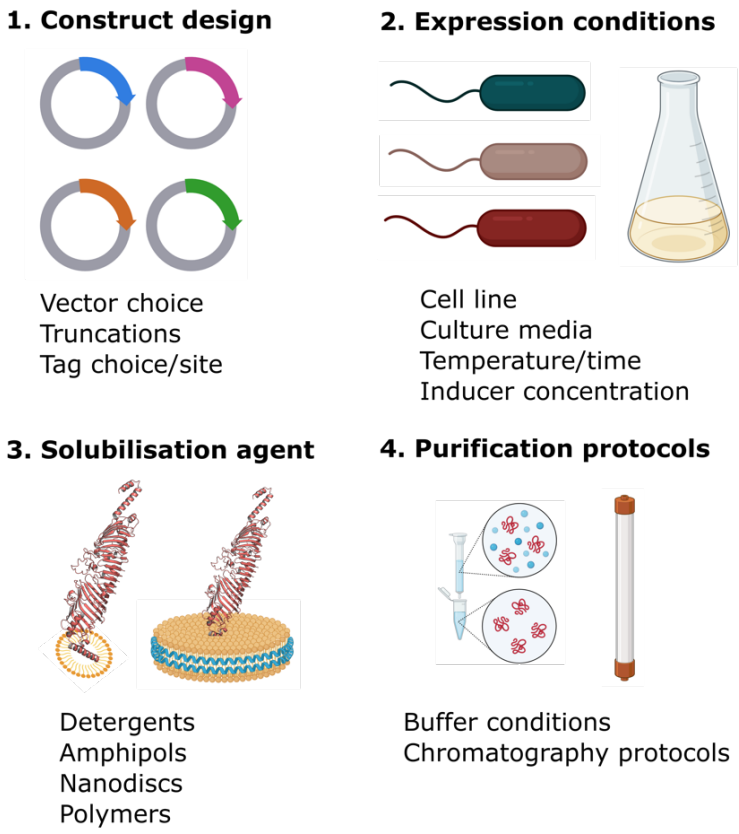


Figure 4.1. Expression and purification optimisation. 1. Construct design considers the choice of vector, truncations of the construct and the choice and location of tags. 2. Expression conditions can be optimised, including choice of cell line, culture media, expression temperature and time, and inducer concentration. 3. For membrane proteins, the choice of solubilisation agent can include detergents, amphipols, nanodiscs and polymers. 4. Purification is optimised by finding ideal buffer constituents and chromatography protocols.

4.1.2 Solubilising membrane proteins

Membrane proteins are notoriously difficult to study due to their amphipathic nature, with a hydrophobic region embedding within the phospholipid tails of a membrane and hydrophilic

surfaces where the protein is solvent-exposed. They require extraction from the membrane utilising various solubilisation agents which allow the protein to remain stable in aqueous environments. The detergent n-dodecyl-β-D-maltoside (DDM) is the most common detergent used for membrane protein solubilisation and stabilisation, but many other detergents and membrane mimetics now exist that can improve extraction from the membrane and/or enhance protein stability in a protein-dependent manner (Carpenter et al., 2008; Ratkeviciute et al., 2021). A solubilisation screen can test various membrane encapsulation agents for their ability to solubilise membrane proteins and improve purity and stability downstream.

4.1.3 Chapter aims

The aims of this chapter were therefore:

1. To find the optimal expression and purification conditions for the AsmA-like proteins, including the use of a solubilisation screen testing extraction in different detergents and assessing protein solubility and stability in different membrane mimetics.
2. To perform biophysical and biochemical studies on proteins that could be well expressed and purified.

4.2 Results

4.2.1 Construct design

Initial constructs were designed with the simple intention of expressing the full-length native protein with a C-terminal 6xHis-tag for each of YhdP, TamB, YdbH, YhjG, AsmA and YicH (Figure 4.2). Throughout the project, many constructs were designed from these initial constructs. These included the design of "soluble" constructs, truncated versions that may improve stability, and isolated domains. Site-directed mutagenesis of the initial constructs was performed to make these different constructs with mixed success. Figure 4.2 shows the constructs that were successfully made and used for subsequent study. The residue at which to delete the TM helix to make a soluble construct was determined by using the TMHMM server and visual inspection of the predicted structure (Krogh et al., 2001). The construct for the putative YdbH complex was designed based on the AlphaFold-multimer prediction of YdbH and its two neighbouring operonic genes: YnbE and YdbL, which folded into a likely complex.

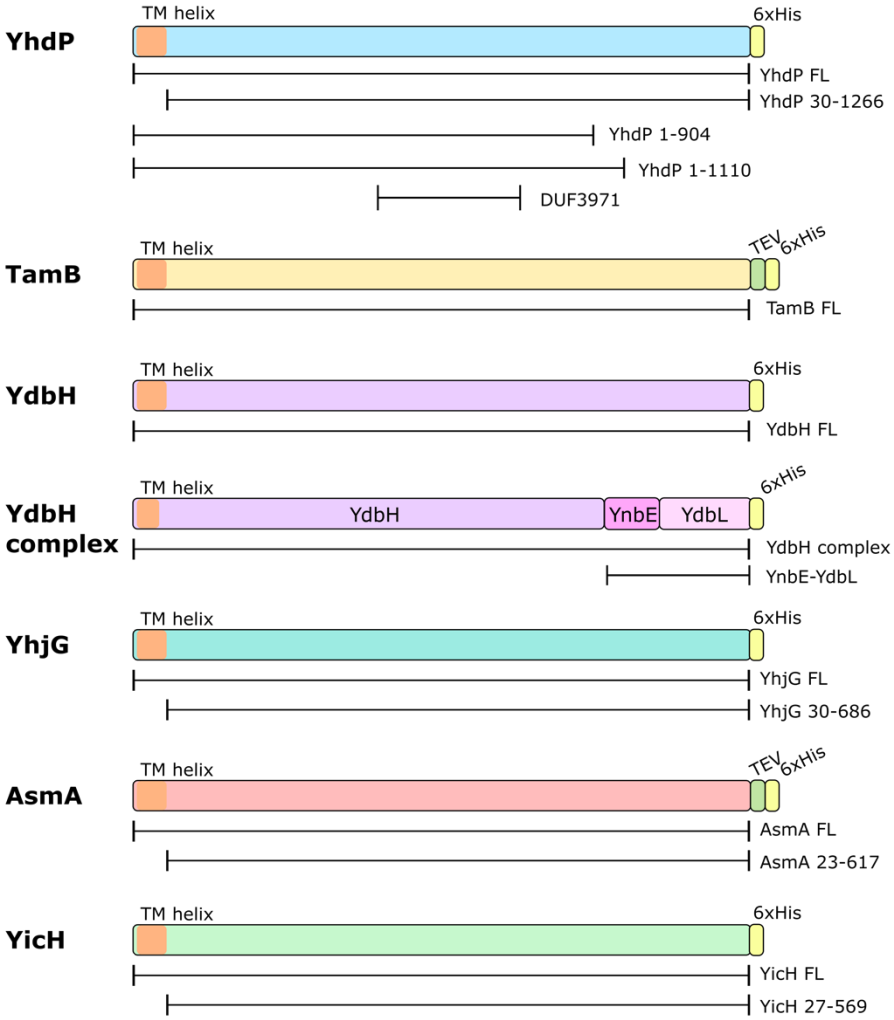


Figure 4.2. Construct design for overexpression of the AsmA-like proteins. Construct designs for each AsmA-like protein (YhdP, TamB, YdbH, YhjG, AsmA and YicH) and the putative YdbH-YnbE-YdbL complex. Various truncations that were successfully made are shown below each gene. These constructs were also used as the template DNA for any mutagenesis work and were used for leaky expression in genetic complementation studies.

4.2.2 Purification proved challenging for YhdP, TamB and AsmA

4.2.2.1 Purification of YhdP full-length

Expression trials determined that full-length YhdP (YhdP FL) was overexpressed in *E. coli* C43(DE3) cells grown in LB media, with a band at the expected molecular weight (MW) of ~140 kDa visible in the post-induction sample (Figure 4.3A). The low expression prompted

further testing of cell lines, including two cell lines that can tightly regulate expression and have seen improved handling of toxic proteins: *E. coli* BL21(DE3)pLysS and *E. coli* Lemo21(DE3) (Studier, 1991; Wagner et al., 2008). The overexpression of YhdP FL in these cell lines was comparable with the other cell lines and media conditions tested, with no clear improvement in yield (Figure 4.3B). To confirm that the band was YhdP, the post-induction sample of *E. coli* C43(DE3) cells in LB media was analysed on an anti-His Western blot, revealing a band at the expected MW (Figure 4.3C).

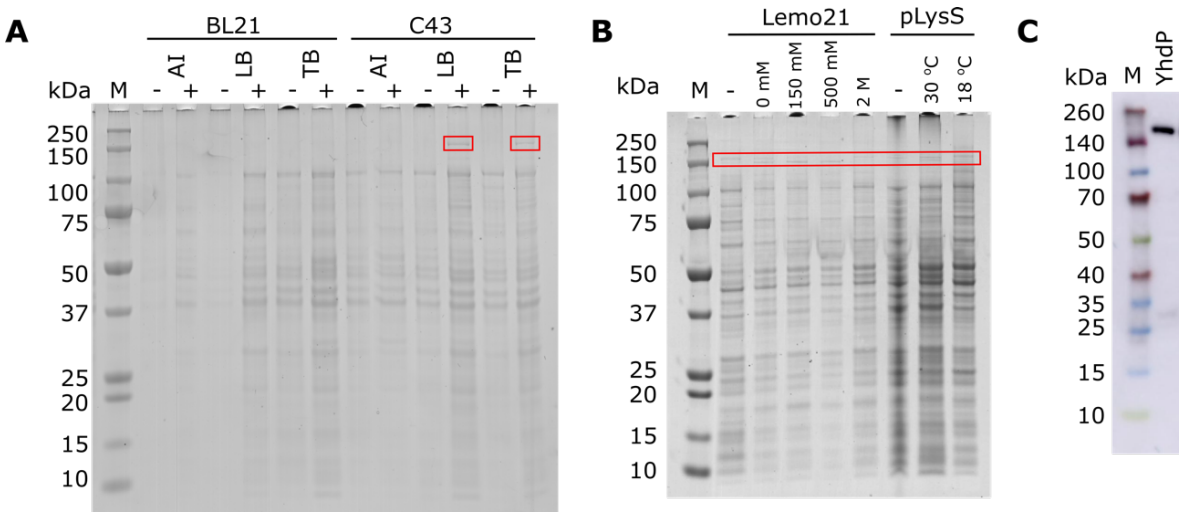


Figure 4.3. Expression trials of YhdP FL. A – SDS-PAGE gel of an expression trial of YhdP FL in *E. coli* C43(DE3) and *E. coli* BL21(DE3) cells using AI, LB or TB media. AI - auto-induction; LB – Miller's lysogeny broth; TB – terrific broth. '-' and '+' indicate pre- and post-induction samples respectively. B – SDS-PAGE gel of further expression trial of YhdP FL using *E. coli* Lemo21(DE3) cells and *E. coli* BL21(DE3)pLysS cells in LB media. Concentrations refer to the concentration of L-rhamnose in the media to control T7 expression. Temperatures correspond to the expression temperature. C – Anti-His western blot of over-expressed YhdP in *E. coli* C43(DE3) with LB media.

Due to low expression levels, 6 L of media were used for the full-scale grows. Subsequent Ni-NTA purifications and gel filtrations could not achieve high purity of the protein. The SDS-PAGE in Figure 4.4A shows the results of a Ni-NTA purification of DDM-solubilised YhdP FL. A band at the expected MW of YhdP of ~140 kDa was seen in the elution fractions (E1-E8). This band was also seen in the flow-through (FT), suggesting poor binding of the His-tag

to the Ni-NTA beads. Longer binding times were trialled but did not improve the purification. Along with the 140 kDa band, the elution fractions contained many smaller molecular weight bands and a higher band at ~200 kDa.

The Ni-NTA elution fractions were then subject to size exclusion chromatography (SEC) using a Superose 6 10/300 column to remove any contaminating proteins or YhdP break-down products that could be contributing to the lower molecular weight bands on the gels. The chromatogram of DDM-solubilised YhdP revealed multiple, unresolved peaks across the elution profile which started eluting near the void at ~8 mL (Figure 4.4B). SDS-PAGE analysis of the peak fractions revealed multiple bands were present across the most intense peak (13-15 mL) that had been carried through from the Ni-NTA purification, with two more prominent bands at ~140 kDa (the expected MW for YhdP) and at ~100 kDa (Figure 4.4C). The later peaks (16-18 mL) revealed a band at ~25 kDa, indicating the successful separation of this smaller protein. The SDS-PAGE and SEC traces presented here represent the most improved purification of DDM-solubilised YhdP FL, but with multiple peaks and multiple bands present after SEC, protein of sufficient purity for downstream use could not be obtained.

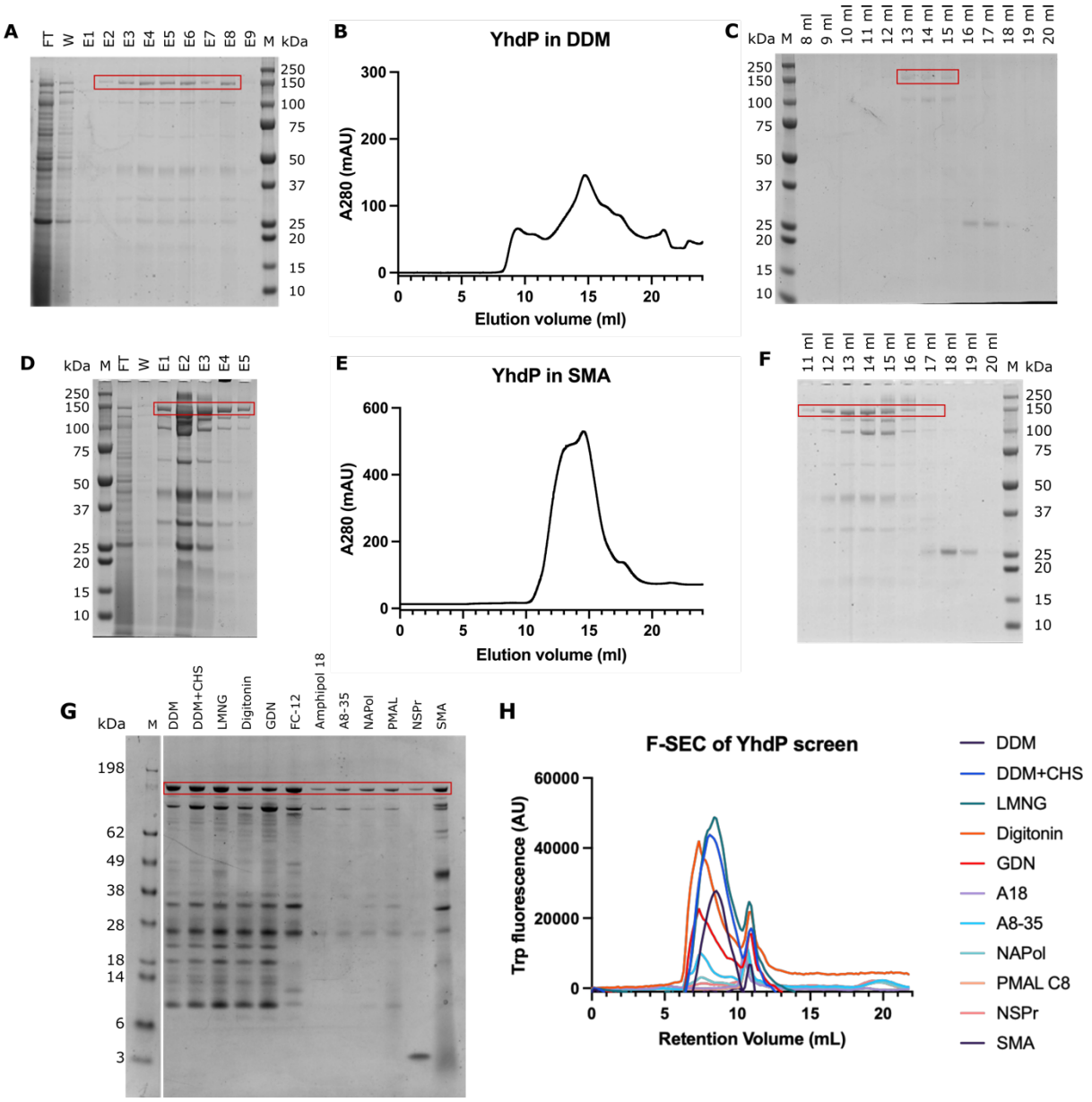


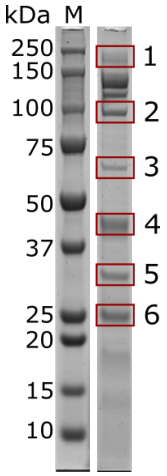
Figure 4.4. Ni-NTA affinity purification and SEC of full length YhdP. A – SDS-PAGE gel of the Ni-NTA affinity purification of YhdP in DDM. B – SEC trace of YhdP in DDM on a Superose 6 10/300 column. C – SDS-PAGE gel of the elution fractions of YhdP in DDM post SEC. D - SDS-PAGE gel of the Ni-NTA affinity purification of YhdP in SMA. E – SEC trace of YhdP in SMA on a Superose 6 10/300 column. F – SDS-PAGE gel of the elution fractions of YhdP in SMA post SEC. G – Solubilisation screen of full-length YhdP. Membranes were solubilised in various membrane mimetics and subject to a small-scale Ni-NTA affinity purification. Samples of the elution fractions were run on the SDS-PAGE gel shown. E – F-SEC traces of the Ni-NTA elution fractions from the solubilisation screen. M – Marker; FT – flow through; W – wash; E1 – 1st mL of elution fraction etc.

Styrene-maleic acid co-polymer (SMA) was then trialled to test for an improvement in the solubilisation of YhdP FL from the membrane. An improved protein yield was seen compared to the DDM-solubilised YhdP, with higher intensity bands at ~140 kDa in the elution fractions on the SDS-PAGE of the Ni-NTA purification presented in Figure 4.4D. The additional bands that were present in the DDM-solubilised YhdP Ni-NTA purification were also present here.

SEC of the pooled elution fractions from the Ni-NTA purification was performed, revealing a broad peak eluting at ~11 mL, later than DDM-solubilised YhdP (Figure 4.4E). SDS-PAGE gels of the eluted fractions post SEC showed multiple bands present across the peak fractions that had been carried through from the Ni-NTA purification (Figure 4.4F). These bands were more prominent in the SMA-solubilised YhdP gel filtration fractions compared to the DDM-solubilised YhdP peak fractions due to the higher amounts of protein present. The tailing shoulder of the broad peak (17-19 mL) contained a small band at ~25 kDa, like the DDM-solubilised YhdP gel filtration.

To identify the bands that were persisting throughout the purifications, six of the contaminating bands were sent for trypsin-digest mass spectrometry at the University of Birmingham Advanced Mass Spectrometry Facility. This was to assess whether the bands were potentially interesting co-purifying interactors, contaminants or YhdP breakdown products. Table 4.1 shows the proteins that were identified for each band sent for analysis with a Sequest-HT score above 100. This score was calculated from the sum of individual peptide scores for each protein, which is based on the number of times the peptide is sequenced (peptide spectrum matches, or PSMs) and the confidence of that result.

Table 4.1. YhdP mass spectrometry results. Trypsin digest mass spectrometry analysis of six contaminating bands in purifications of YhdP. Proteins listed are those with a Sequest score above 100. Accession ID, protein name, theoretical molecular weight and Sequest score are shown for each band.

Lane	Band	Accession	Protein	MW (kDa)	Sequest score
	1	P46474	YhdP	139	313.28
		P31224	AcrB	113.5	118.58
		P0A6X3	Hfq	11.2	162.47
	2	P31224	AcrB	113.5	1553.24
		P46474	YhdP	139	407.5
	3	P31433	YicH	62.2	702.18
		P46474	YhdP	139	382.86
	4	P0ABI8	CyoB	74.3	309.22
		P46474	YhdP	139	263.09
		P0AE06	AcrA	42.2	163.11
		P0CE47	TufA	43.3	132.65
	5	P0ABJ1	CyoA	34.9	730.93
		P46474	YhdP	139	485.01
	6	P0A9K9	SlyD	20.8	608.95
		P46474	YhdP	139	424.18

YhdP peptides were present in all bands, with other hits being common *E. coli* protein contaminants and YicH, another AsmA-like protein. AcrB, SlyD and CyoA have all been previously reported as Ni-NTA purification contaminants (Veesler et al., 2008; Robichon et al., 2011; Abeyrathne and Grigorieff, 2017). The other bands present are the DNA and RNA-binding proteins TufA and Hfq, and so are unlikely to be relevant interacting proteins, and CyoB and AcrA, which likely co-purified with CyoA and AcrB respectively (Vijgenboom et al., 1988; Cotter et al., 1990; Nikaido and Zgurskaya, 2001; Vogel and Luisi, 2011). Therefore, the additional bands present on the SDS-PAGE gels are most likely a result of YhdP breakdown and contamination. The presence of YicH was surprising, and not easily explained. It could simply be from contamination of the gel as work on purifying YicH had started at this point.

Further attempts at improving the purification of YhdP FL were then performed when work at the Membrane Protein Laboratory (MPL) commenced. Solubilisation screening was performed on membrane preparations of YhdP FL. A small-scale purification followed, and the resulting elution fractions were analysed by SDS-PAGE (Figure 4.4G). The gel showed that detergents and SMA had a higher yield of YhdP compared to the other membrane mimetics. The banding pattern was mostly consistent across the samples, with prominent bands at ~100 kDa, ~35 kDa, ~25 kDa and ~10 kDa. These bands were not as visible when the peptidisc NSPr was used, but the yield was also low for this protein.

Fluorescence SEC (F-SEC) of the elution fractions from the screen was then performed to look at protein dispersity using intrinsic tryptophan fluorescence. The chromatograms in Figure 4.2H revealed that most of the samples eluted close to the void volume at ~7 mL with often a second minor peak around ~11 mL. The polymers, nanodisc and peptidisc samples had too low a protein yield to achieve a good signal from F-SEC. Interestingly, the solubilisation of YhdP was different in this small-scale solubilisation screen than it was in the large-scale purifications in DDM and SMA. This was mostly attributed to the use of a different resin for the Ni-NTA purification and the use of different reagents. Unfortunately, there was no clear improvement in both the yield and purity of YhdP from the solubilisation screen.

4.2.2.2 Purification of truncated YhdP constructs

Ultimately, none of the purification trials for full-length YhdP produced sufficiently pure protein for structural and functional studies. With the predicted structure showing that YhdP is a long, likely flexible, protein that may insert into both the IM and the OM, various constructs were designed that aimed to shorten the length of the protein and to remove membrane-bound regions to increase stability and extraction. Unfortunately, attempts at making a construct of

YhdP without the C-terminal helical-turn-helix that potentially embeds into the outer membrane (deletion of residues 1201-1266) were unsuccessful.

A construct of YhdP without the N-terminal helix (deletion of residue 1-29) was made to remove the need for detergent throughout the purification (Figure 4.5A). Expression trials revealed over-expression of a band at ~140 kDa in *E. coli* BL21(DE3) cells in LB and further improved yield in TB (Figure 4.5B). Some leaky expression in the pre-induction samples was seen for these two conditions. *E. coli* BL21(DE3) cells in AI grew very poorly and samples were not able to be normalised to the other conditions, resulting in very faint bands. Unfortunately, initial Ni-NTA purifications revealed that the protein was mostly insoluble, as a prominent band at ~140 kDa was visible in the insoluble fraction, with very little protein in the elution fractions (Figure 4.5C). Protein degradation was likely an issue, with the same banding pattern that had been seen previously, indicating that the stability had not improved. The insolubility of the protein is not overly surprising, considering there may still be a membrane-bound region at the C-terminal end of the protein and due to the potential exposure of hydrophobic residues within the cavity at the N-terminus. It is likely that this construct required detergent or other amphipathic molecules to be present to stabilise these areas of hydrophobicity in solution. Due to this, further attempts at purifying this construct were not made as the main aim here was to remove the need for detergent.

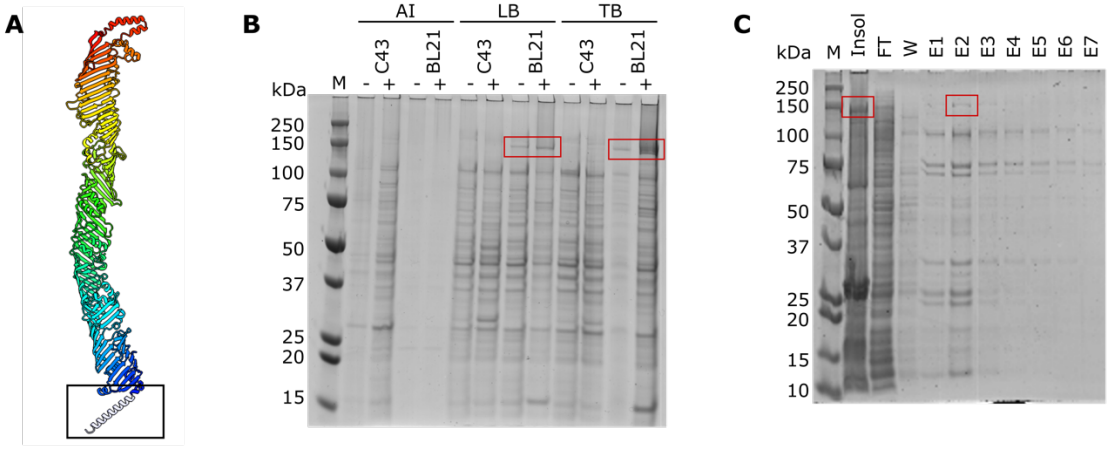


Figure 4.5. Design, Ni-NTA affinity purification and SEC of YhdP 30-1266. A – Model of deletion of YhdP residues 1-29 to create YhdP 30-1266. B – SDS-PAGE gel of an expression trial of YhdP 30-1266. *E. coli* C43(DE3) and *E. coli* BL21(DE3) cells were trialled in three media types: auto-induction (AI), LB and TB. '-' and '+' indicate pre and post-induction samples respectively. C – SDS-PAGE gel of the Ni-NTA affinity purification of YhdP 30-1266. FT – flow through; W – wash; E1 – 1st mL of elution fraction etc.; Insol – insoluble material after cell lysis.

Two shortened constructs were successfully designed and overexpressed: YhdP residues 1-904 and YhdP residues 1-1110. YhdP 1-904 was designed to end after a repeating β-groove (RBG) fold at what seems to be a point of constriction within the structure, with a predicted shortening by 84 Å from the C-terminal region (Figure 4.6A). The point of deletion for YhdP 1-1110 was selected to remove any C-terminal region that may be interacting with the OM, shortening it by potentially 32 Å (Figure 4.6B). Expression appeared to improve with these shortened constructs compared to YhdP FL, with a higher protein yield in trial purifications from the same expression conditions as YhdP FL.

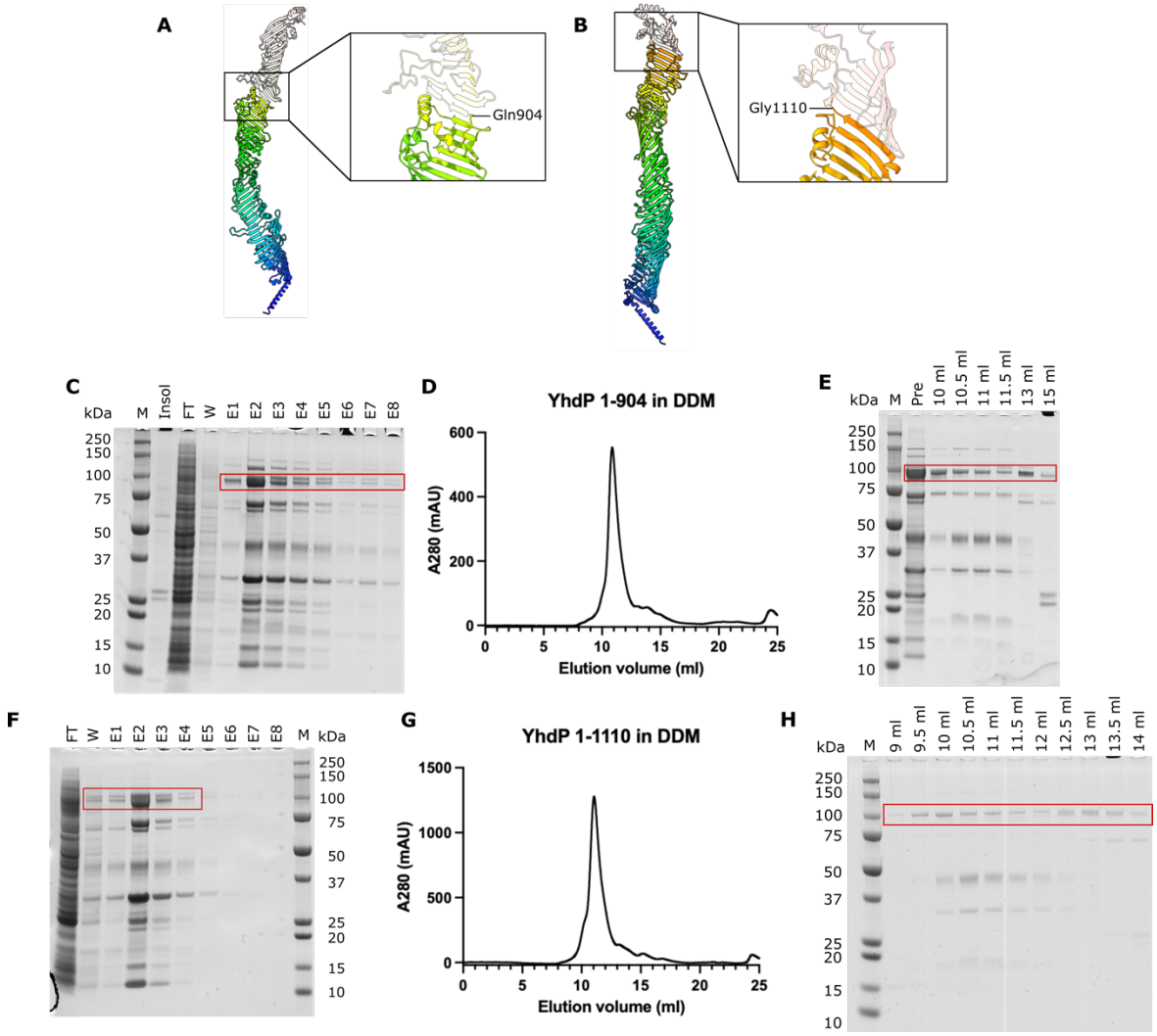


Figure 4.6. Design, Ni-NTA affinity purification and SEC of YhdP 1-904 and YhdP 1-1110. A – Model of deletion of YhdP residues 905-1266 to create YhdP 1-904. B – Model of deletion of YhdP residues 1111-1266 to create YhdP 1-1110. C - SDS-PAGE gel of the Ni-NTA affinity purification of YhdP 1-904. D – SEC trace of YhdP 1-904 on a Superose 6 10/300 column. E – SDS-PAGE gel of the elution fractions of YhdP 1-904 post SEC. F - SDS-PAGE gel of the Ni-NTA affinity purification of YhdP 1-1110. G – SEC trace of YhdP 1-1110 on a Superose 6 10/300 column. H – SDS-PAGE gel of the elution fractions of YhdP 1-1110 post SEC. FT – flow through; W – wash; E1 – 1st mL of elution fraction etc.; Insol – insoluble pellet after solubilisation; Pre – Concentrated sample prior to running on SEC.

The two proteins were purified via Ni-NTA purification and SEC following the same optimised conditions as full-length YhdP. The SDS-PAGE gels post Ni-NTA purification of the two shortened constructs looked very similar, with the same banding pattern present as in the full-length protein (Figure 4.6C and F). The most prominent band was around ~100 kDa in both

purifications, which aligned with the expected weights of 100 kDa for YhdP 1-904 and 123 kDa for YhdP 1-1110. There was little difference in the migration of the band between the two proteins, despite an expected molecular weight difference of 23 kDa. This was likely due to the poor resolving power at the top of the gel. The deletion of the C-terminal region increased the protein yield compared to the full-length construct, potentially due to an improved ability of DDM to solubilise YhdP 1-904 and 1-1110 from the membrane and/or due to a more accessible C-terminal His-tag able to reduce the amount getting lost in the FT. The SEC chromatograms of both proteins revealed a prominent single peak at ~11 mL with a leading and tailing shoulder (Figure 4.6D and G). SDS-PAGE analysis of this peak from both purifications showed that multiple bands were present (Figure 4.6E and H). The likelihood that these are the same proteins as previously identified with mass spectrometry for YhdP full-length is high, implying that shortening the length of the protein did not improve protein stability as breakdown products were still present.

Further investigation focussed on the purification of a smaller, soluble part of YhdP containing just the DUF3971 domain. It was hypothesised that this would have less protein breakdown and if successful, could be subject to crystallisation trials, akin to the determination of the crystal structure of a partial DUF490 domain in TamB (Josts et al., 2017). Figure 4.7A shows the DUF3971 region of YhdP, with Figure 4.7B showing an SDS-PAGE of a Ni-NTA purification of the protein. The expected weight of the DUF3971 construct is 32 kDa, and a band at this weight can be seen in the elution fractions of the purification (Figure 4.7B). The pooled fractions were then run on an S75 16/600 column, resulting in a peak at the void, likely aggregate, overlapping with a more intense peak at ~52 mL (Figure 4.7C). SDS-PAGE analysis of peak fractions revealed a band at the expected MW for DUF3971, but whilst the lower MW contaminant at ~12 kDa was successfully separated, a prominent band at 26 kDa and some

higher MW bands remained (Figure 4.7D). Unfortunately, the samples were prone to aggregation, and likely needed detergent for stabilisation. Indeed, the portion of the TamB DUF490 domain that was crystallised also required the presence of the detergent LDAO for protein stability (Josts et al., 2017). Some crystallisation trays were set up of DUF3971, but with contamination still present and apparent poor stability, it is unsurprising that no crystals formed.

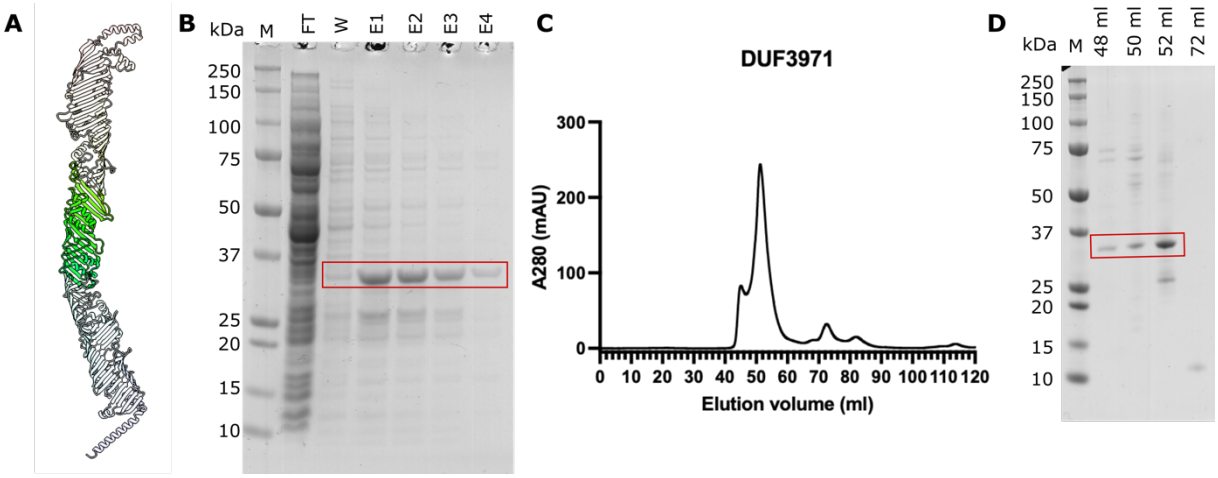


Figure 4.7. Design, Ni-NTA affinity purification and SEC of YhdP DUF3971. A – Model of deletion of YhdP residues 1-503 and 790-1266 to create YhdP DUF3971. B – SDS-PAGE gel of the Ni-NTA affinity purification of DUF3971. C – SEC trace of DUF3971 on a Superdex 75 16/600 column. D – SDS-PAGE gel of the SEC peak fractions. FT – flow through; W – wash; E1 – 1st mL of elution fraction etc.

4.2.2.3 Purifications of TamB and AsmA

At the beginning of the project, TamB and AsmA were the most well-studied of the AsmA-like protein family in terms of putative functions. AsmA mutants were known to affect LPS levels and TamB was known to have a role in the assembly of OMPs (Deng and Misra, 1996; Shen et al., 2014). Structural studies of TamB had also been reported, with a crystal structure of a portion of TamB, from which the β-taco fold was initially characterised, and atomic force microscopy of a soluble TamB construct that revealed a ~180 Å long, elongated structure (Shen

et al., 2014; Josts et al., 2017). With the difficulty in purifying YhdP, the project focus expanded to the characterisation of the other AsmA-like proteins. To assess the amenability of the other AsmA-like proteins to further structural and functional studies, attempts at expressing and purifying them were made. This section will focus on the attempts made to purify TamB and AsmA.

Expression trials revealed that both TamB full-length (TamB FL) and AsmA full-length (AsmA FL) were able to be overexpressed in *E. coli* C43(DE3) cells in LB media, with bands at the expected MW of 138 kDa for TamB and 70 kDa for AsmA in the post-induction samples (Figure 4.8A and D, red box). The overexpression of TamB was low, and so further expression trials were performed using Lemo21(DE3) and BL21(DE3)pLysS cells. Whilst a band was present at the expected MW for TamB again, expression levels were comparable to that of the *E. coli* C43(DE3) cells, and so were not used for full-scale expression (Figure 4.8B). An anti-His western blot was performed on the post-induction samples of *E. coli* C43(DE3) in LB media cells for both TamB and AsmA, revealing bands at the expected weight for both proteins (Figure 4.8C and E, respectively). Two lower bands at ~42 kDa and ~28 kDa were seen in both Western blots, either indicating similar breakdown patterns or indicating other *E. coli* proteins were present that can bind an anti-His antibody.

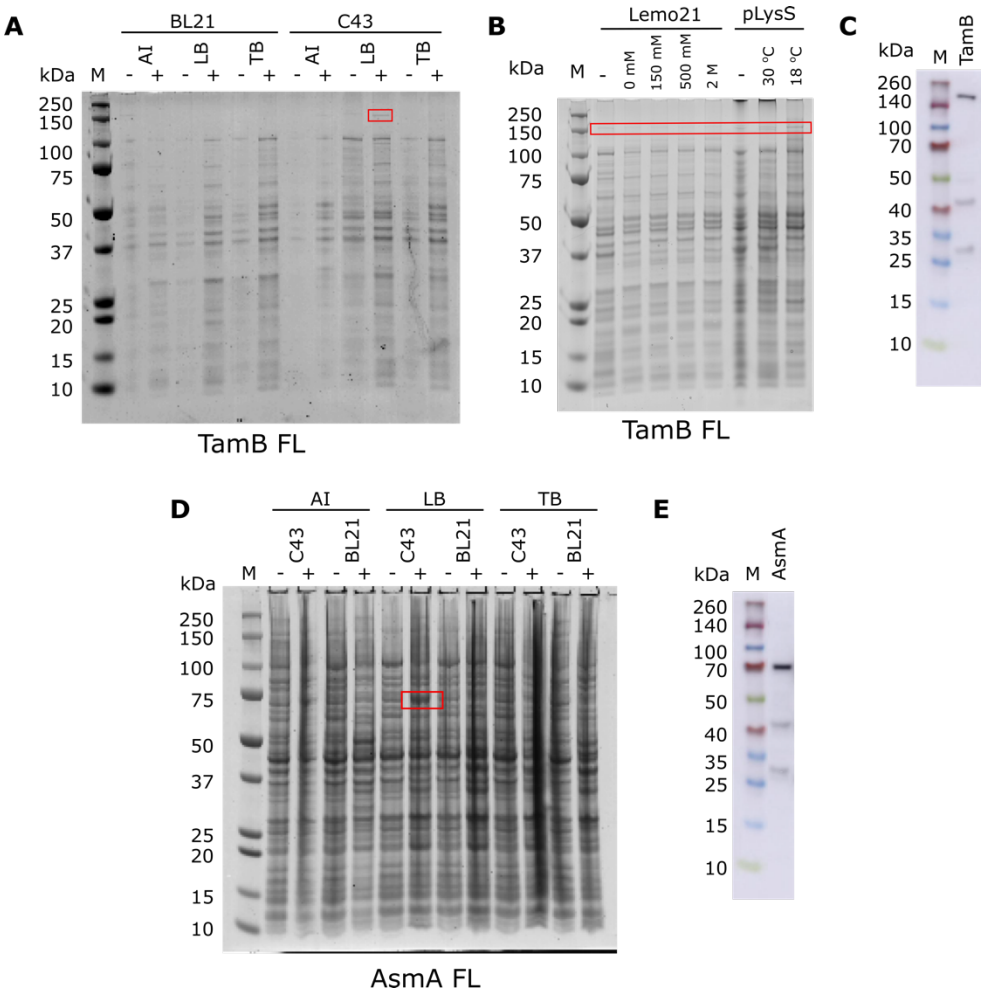


Figure 4.8. Expression trials of TamB FL and AsmA FL. A – SDS-PAGE gel of an expression trial of TamB FL in *E. coli* C43(DE3) and *E. coli* BL21(DE3) cells using AI, LB or TB media. B – SDS-PAGE gel of further expression trial of TamB FL using *E. coli* Lemo21(DE3) cells and *E. coli* BL21(DE3)pLysS cells in LB media. Concentrations refer to the amount of L-rhamnose added to the media to control T7 expression. Temperatures correspond to the growth and expression temperature. C – Western blot of over-expressed TamB in *E. coli* C43(DE3) with LB media. D - SDS-PAGE gel of an expression trial of AsmA FL in *E. coli* C43(DE3) and *E. coli* BL21(DE3) in AI, LB or TB media. AI - auto-induction; LB – Miller's lysogeny broth; TB – terrific broth. '-' and '+' indicate pre- and post-induction samples respectively.

TamB was DDM-solubilised and purified following the optimised protocol of YhdP FL. SDS-PAGE analysis of the Ni-NTA purification revealed a prominent band at the expected weight of ~138 kDa (Figure 4.9A). It was interesting to note a similar banding pattern to the purifications of YhdP with the presence of many lower molecular weight bands, implying

TamB may also be prone to breakdown. There was also no band representing TamA (65 kDa), which is known to form a complex with TamB (Shen et al., 2014).

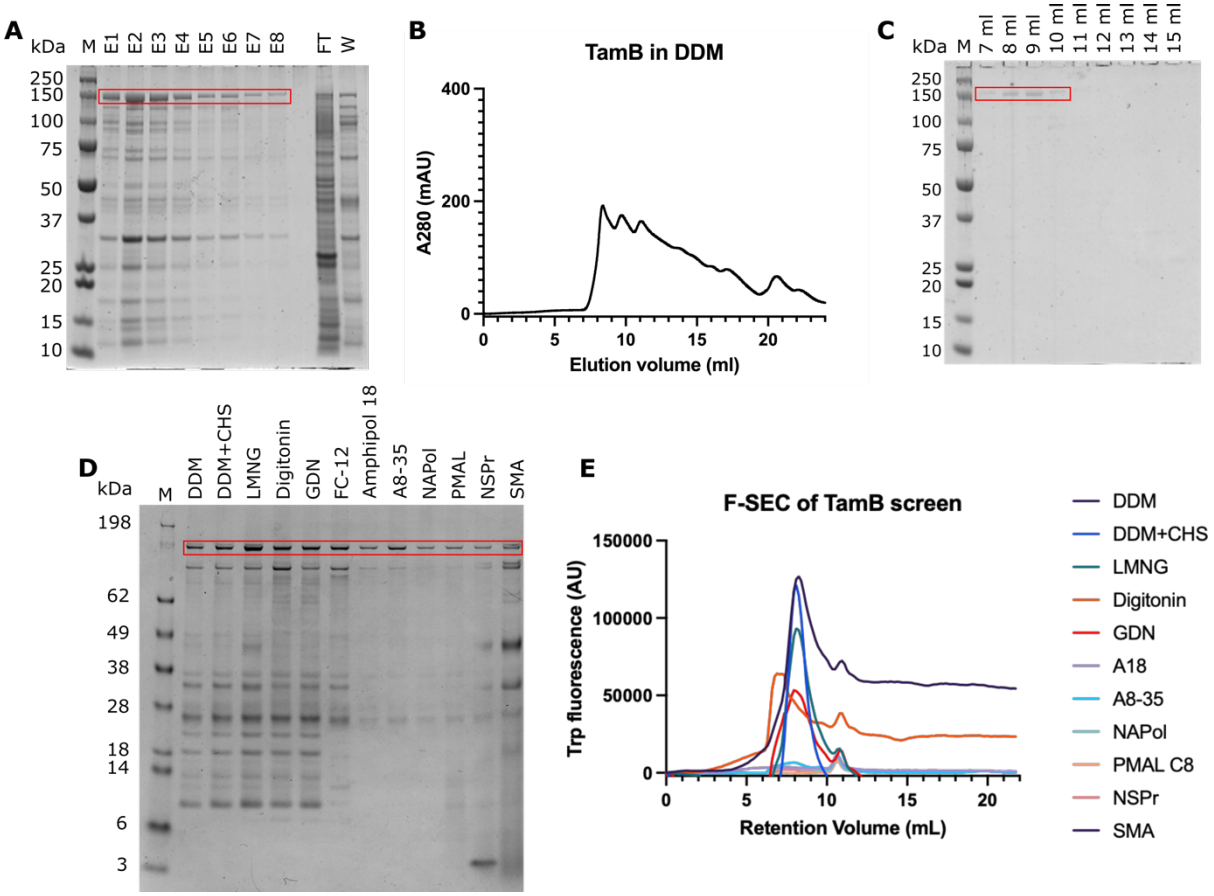


Figure 4.9. Purifications and solubilisation screen of full-length TamB. A – SDS-PAGE gel of the Ni-NTA affinity purification of TamB. B – SEC trace of TamB on a Superdex 200 10/300 column. C – SDS-PAGE gel of the elution fractions of TamB post-SEC. D – Solubilisation screen of full-length TamB. Membranes were solubilised in various membrane mimetics and subject to a small-scale Ni-NTA affinity purification. Samples of the elution fractions were run on the SDS-PAGE gel shown. E – F-SEC traces of the Ni-NTA elution fractions from the solubilisation screen. M – Marker; FT – flow through; W – wash; E1 – 1st mL of elution fraction etc.

Subsequent gel filtration of the pooled Ni-NTA elution fractions revealed a mixed population, with very broad, overlapping peaks (Figure 4.9B). SDS-PAGE analysis of the peak fractions revealed a mostly pure band at the expected molecular weight but with a low yield (Figure 4.9C). Faint bands for lower molecular weight species were also present. With the SEC profile

indicating a polydisperse sample, the stability and purity of the protein were questionable even with a seemingly pure sample by SDS-PAGE.

The same solubilisation screen as used for YhdP FL was used to test TamB solubilisation and protein stability in different membrane mimetics. The results show a similar pattern to YhdP, with a higher yield of protein when detergents were used compared to the polymers, nanodisc and peptidisc (Figure 4.9D). The use of LMNG seems to improve the yield of TamB compared to the other detergents, although a band at ~45 kDa was more prominent in this sample. The purity improved in the polymer/nanodisc/peptidisc samples although faint bands at lower molecular weights do suggest that the contaminants or breakdown products were still present, just in lower amounts. The F-SEC traces suggest a polydisperse sample, with an initial peak in the void volume at ~7 mL followed by a minor peak at ~11 mL (Figure 4.9E). The polymer/nanodisc/peptidisc samples were again too low of a yield to give a good signal. Some issues were also seen with the normalisation of the curves due to baseline drift, which can occur when the signal-to-noise ratio is too low. Like YhdP, there was no solubilisation agent that resulted in a high yield of monodisperse, pure protein.

Issues were also encountered in the purification of AsmA FL. The SDS-PAGE analysis of the initial Ni-NTA purification of DDM-solubilised AsmA FL revealed a faint band at the expected weight of 70 kDa, with many contaminating proteins or breakdown products visible (Figure 4.10A). As this may have been due to poor solubilisation with DDM, the solubilisation screen was performed on membranes of over-expressed AsmA. SDS-PAGE showed poor solubilising efficiency in all the detergents tested, including the harsh FC-12 detergent, and in SMA (Figure 4.10B). It was unsurprising that the polymers/nanodisc/peptidisc retained little protein, due to the exchange from DDM that is required. The F-SEC also revealed poor monodispersity and low signal due to low protein yield in all samples, except DDM and DDM+CHS (Figure 4.10C).

However, this is likely due to the more prominent band at ~100 kDa that is present on the gel (Figure 4.10B). AsmA FL is likely expressed into the insoluble fraction, as expression trials indicate high levels of overexpression but no extraction from the membrane fraction can be seen in the positive control of FC-12. It is also worth noting that the spin to remove cell debris after cell lysis resulted in a white pellet, which can indicate the presence of inclusion bodies.

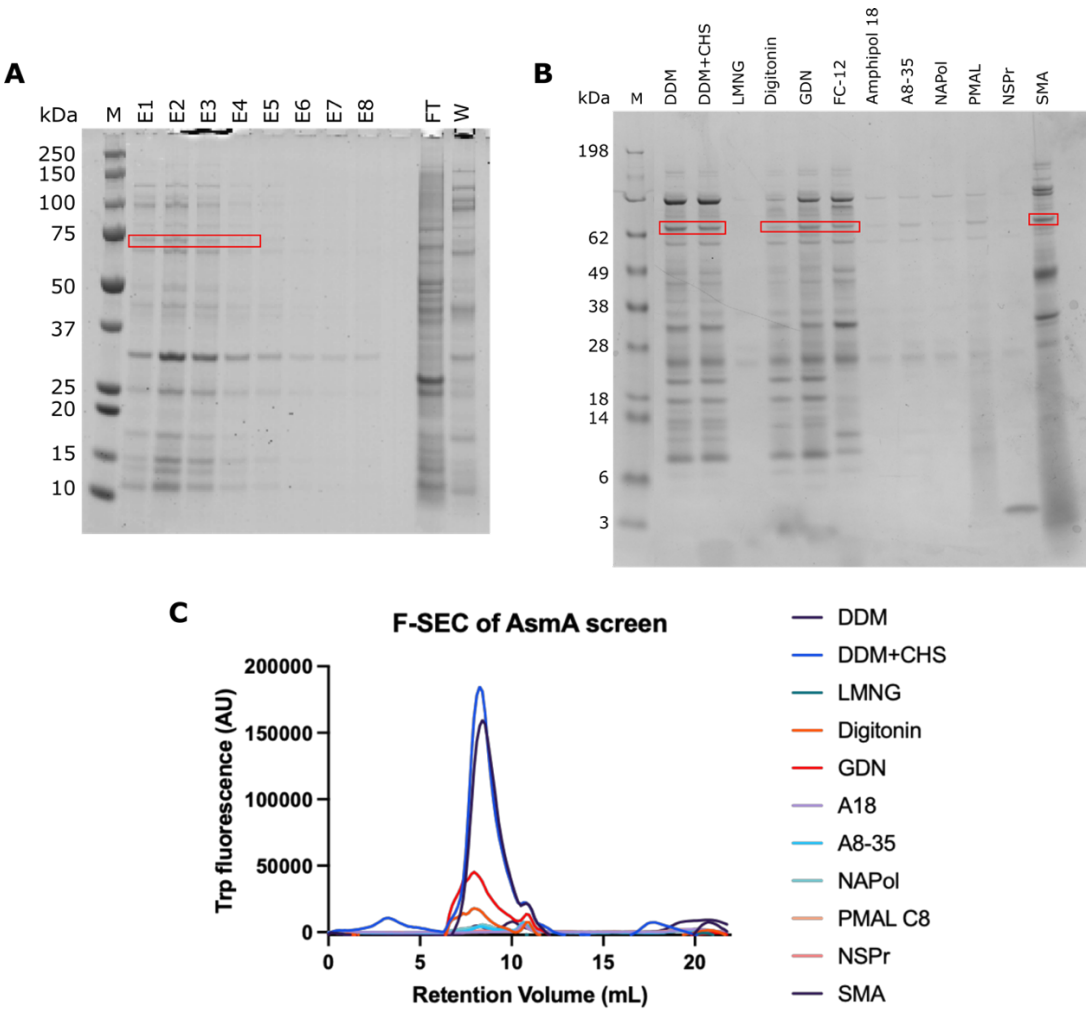


Figure 4.10. Ni-NTA affinity purification and solubilisation screen of full-length AsmA. A – SDS-PAGE gel of the Ni-NTA affinity purification of AsmA FL in DDM. B – Solubilisation screen of AsmA FL. Membranes were solubilised in various membrane mimetics and subject to a small-scale Ni-NTA affinity purification. Samples of the elution fractions were analysed by SDS-PAGE. C – F-SEC traces of the Ni-NTA elution fractions from the solubilisation screen. M – Marker; FT – flow through; W – wash; E1 – 1st mL of elution fraction etc.

Like YhdP FL, soluble constructs of TamB and AsmA were attempted to be made by removing the N-terminal helix of the protein. Unfortunately, this was unsuccessful in TamB. Removal of the residues 1-22 in AsmA was successful (Figure 4.11A), and so, a trial purification was performed. The optimal expression conditions determined for YhdP 30-1266 of *E. coli* BL21(DE3) in TB media were used for this preliminary test due to a prediction that they would behave similarly. The resulting Ni-NTA purification is shown in Figure 4.11B, revealing a band of high intensity at the expected MW for AsmA 23-617 of 67 kDa. There was another prominent unidentified band at ~25 kDa, and many other bands likely representing contaminating or breakdown products.

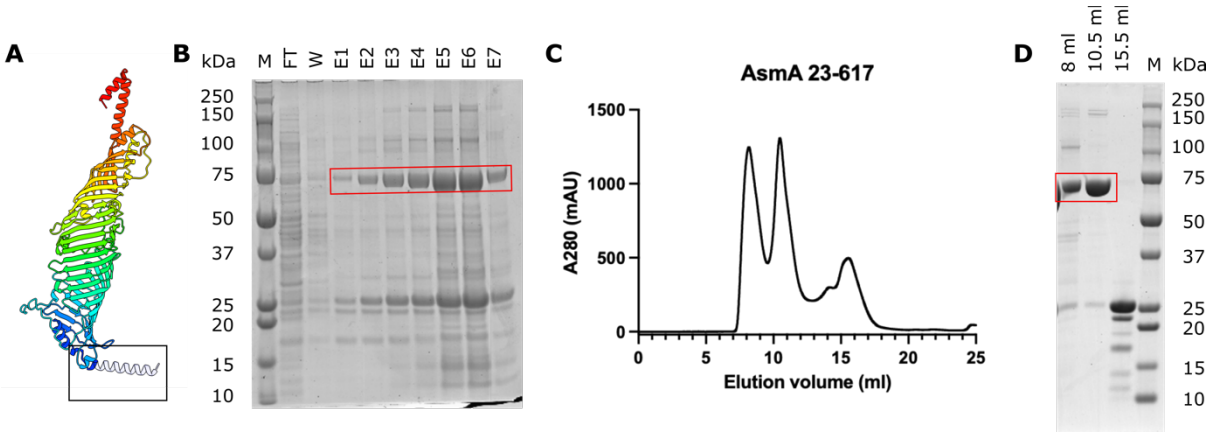


Figure 4.11. Design, Ni-NTA affinity purification and SEC of AsmA 23-617. A – Model of deletion of AsmA residues 1-22 to create AsmA 23-617. B – SDS-PAGE gel of the Ni-NTA affinity purification of AsmA 23-617. C – SEC trace of AsmA 23-617 on a Superdex 200 10/300 column. D – SDS-PAGE gel of the SEC peak fractions. FT – flow through; W – wash; E1 – 1st mL of elution fraction etc.

Subsequent SEC of the pooled elution fractions revealed multiple peaks that elute earlier than expected for a soluble protein of 67 kDa. The first peak is likely aggregate, due to sharply eluting in the void at ~7 mL, whereas the second peak (10 mL) is likely a more stable protein. The SDS-PAGE gel of the peak fractions revealed that the band for AsmA 23-617 was present in the first two peaks, with the prominent band at ~25 kDa mostly separated and coming off at

~15 mL (Figure 4.11C). The peak at ~10 mL was then concentrated for crystallisation trials, but protein instability seemed to be an issue, with visible aggregation occurring. Further purification may also be needed to remove the higher ~200 kDa band and lower ~25 kDa band for successful crystallisation.

4.2.3 YdbH, YhjG and YicH were more amenable to expression and purification

4.2.3.1 Purification of YdbH

YdbH was well overexpressed in the initial growth conditions using *E. coli* C43(DE3) cells. Ni-NTA purification of YdbH revealed a very strong band just under 100 kDa, which aligns with the expected molecular weight of 98 kDa for YdbH, with some bands at lower molecular weights which are likely contaminants (Figure 4.12A). Unfortunately, the protein was unstable at these high concentrations, with visible aggregation in the most concentrated elution fractions. The aggregation was more noticeable upon spin concentration when pooling together fractions prior to SEC; therefore, this step was omitted. Instead, individual elution fractions were taken immediately from Ni-NTA elution to gel filtration to minimise the amount of time in high imidazole concentrations and to not further increase the protein concentration. The resulting gel filtration using a Superdex 200 16/600 column revealed a single peak with a tailing shoulder which started eluting at ~73 mL (Figure 4.12B). SDS-PAGE analysis of the peak fractions revealed a prominent band at 100 kDa and a smaller band at 75 kDa that was present across all the fractions and is likely protein breakdown (Figure 4.12C). Samples of YdbH for further analysis were taken from the main peak as it was suspected that the shoulder was likely protein breakdown.

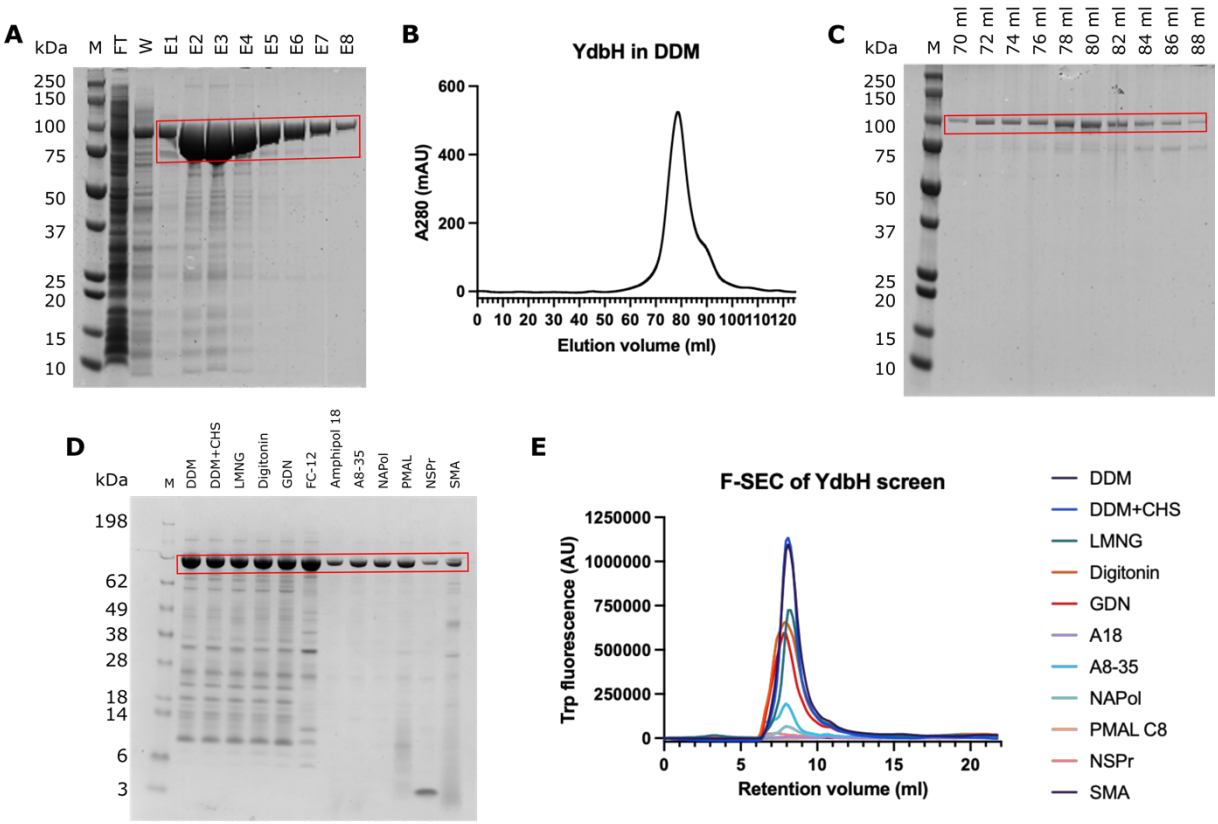


Figure 4.12. Purifications and solubilisation screen of YdbH. A – SDS-PAGE gel of the Ni-NTA affinity purification of YdbH. B – SEC trace of YdbH on a Superdex 200 16/600 column. C – SDS-PAGE gel of the elution fractions from the SEC of YdbH. D – Solubilisation screen of full-length YdbH. Membranes were solubilised in various membrane mimetics and subject to a small-scale Ni-NTA affinity purification. Samples of the elution fractions were run on the SDS-PAGE gel shown. E – F-SEC traces of the Ni-NTA elution fractions from the solubilisation screen. M – Marker; FT – flow through; W – wash; E1 – 1st mL of elution fraction etc.

To improve protein stability, the solubilisation screen was performed on membranes of overexpressed YdbH. Figure 4.12D shows the SDS-PAGE gel of this screen, with an obvious band at ~100 kDa representing YdbH. The detergents have a higher yield of YdbH than the other membrane mimetics, but with more contaminating bands. The F-SEC results for YdbH eluted as a single peak at ~8 mL in most conditions, which indicates a monodisperse sample (Figure 4.12E). The amphipol, nanodisc and peptidisc samples again gave too low a signal to be accurately measured.

Unlike the YhdP and TamB solubilisation screens, there was enough protein (at least ~0.1 mg/mL) from the purification trial to test protein stability in the different conditions using nano-differential scanning fluorimetry (nanoDSF). This technique uses intrinsic tryptophan fluorescence to assess changes in protein structure. Typically an increase in tryptophan fluorescence results from the exposure of hydrophobic regions, an indication of protein unfolding (Kim et al., 2021). This typically causes a bathochromic shift in tryptophan fluorescence from 330 nm (F330) to 350 nm (F350). The sample is exposed to rising temperatures and the tryptophan fluorescence is followed. An inflection point in the F350/F330 ratio indicates the melting temperature (T_m) of the sample, the temperature at which 50% of the protein is unfolded. The aim of the experiment is therefore to find the conditions in which the protein is the most stable, i.e. has the highest T_m.

The nanoDSF experiments on YdbH in detergents revealed a substantial increase in protein stability when glycol-diosgenin (GDN, T_m = 65ºC) and digitonin (T_m = 70ºC) were used compared to DDM (T_m = 58ºC), DDM+CHS (T_m = 59ºC) and LMNG (T_m = 58ºC) (Figure 4.13A). FC-12 behaved poorly, but this was to be expected due to its harsh nature. Figure 4.13B shows the nanoDSF results for YdbH in the amphipols, nanodisc, peptidisc and polymers. A substantial improvement in protein stability was seen with amphipol A18 compared to the other membrane mimetics, with a melting temperature of 72ºC.

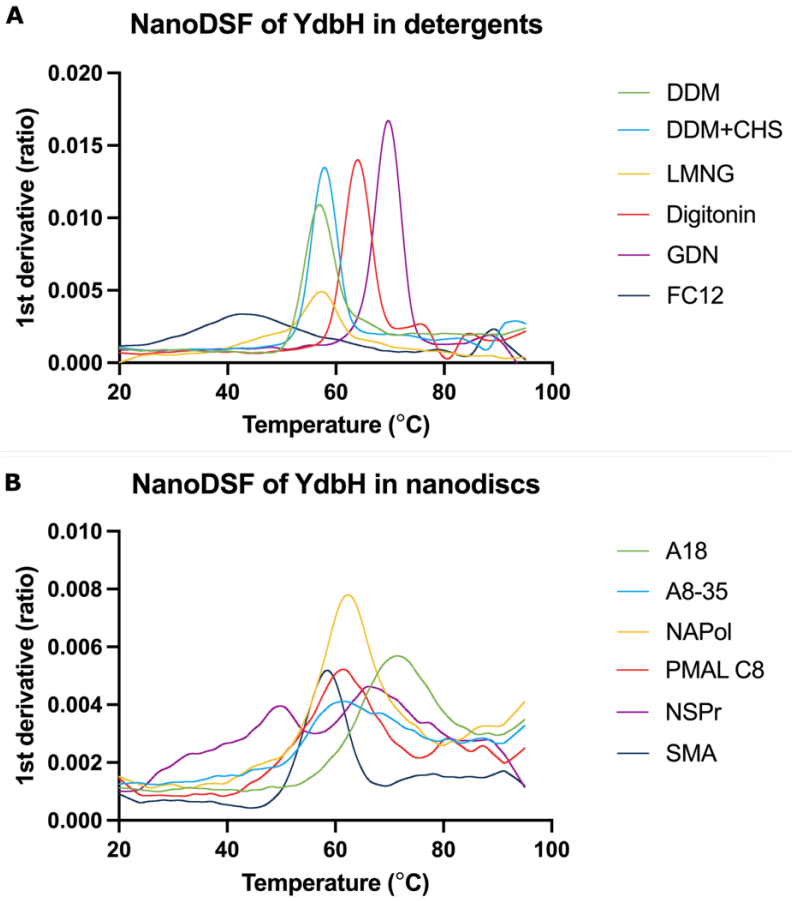


Figure 4.13. NanoDSF of the YdbH solubilisation screen. A – NanoDSF measurements of YdbH in the detergents from the screen. B – NanoDSF measurements of YdbH in the polymers, nanodiscs, peptidiscs and amphipols from the screen. Data is presented as the 1st derivative of the thermal denaturation curves measuring F350/F330, with the maximum representing the inflection point or T_m.

With the purity and quantity of protein resulting from the solubilisation screen, YdbH was able to be put on grids for negative stain electron microscopy (EM). The aim of this was to look at the behaviour of the purified protein on grids, assessing particle heterogeneity, purity and ultimately the membrane mimetic preference for further EM studies. A representative micrograph of each grid is shown in Figure 4.14. General stick-like particles could be seen in the micrographs, with dimensions in the range of those predicted for YdbH in the AlphaFold model of ~216 Å long and ~30 Å wide. Most of the micrographs revealed protein aggregation, particularly in the GDN sample, and severe particle overcrowding, particularly in the NAPol

sample. Particles were most easily identifiable when the amphipol A8-35 was used, with less contamination and aggregation visible compared to the other conditions. Considering these samples had not been purified after the small-scale Ni-NTA purification, this condition was promising for further EM analysis. Whilst cryo-EM grids were set up of YdbH in A8-35, stick-like particles were not visible in the grid holes, and due to time limitations of the project, further work optimising these grids was not able to be performed.

4.2.3.2 Purification of the putative YdbH complex

At the time of experimentation, the complex between YdbH and YnbE had not been experimentally proven. The complex formation was only predicted based on the AlphaFold-multimer model (Figure 4.15A). The β-strand augmentation of YnbE into the RBG folds of YdbH led to the hypothesis that exposure of this β-strand was leading to instability. YdbL is predicted to bind near an opening of the YdbH hydrophobic cavity, so it was hypothesised that in the absence of YdbL, this exposed region was also causing instability. Therefore, over-expression of the putative complex was attempted by expressing the whole operon, with a C-terminal His-tag on YdbL.

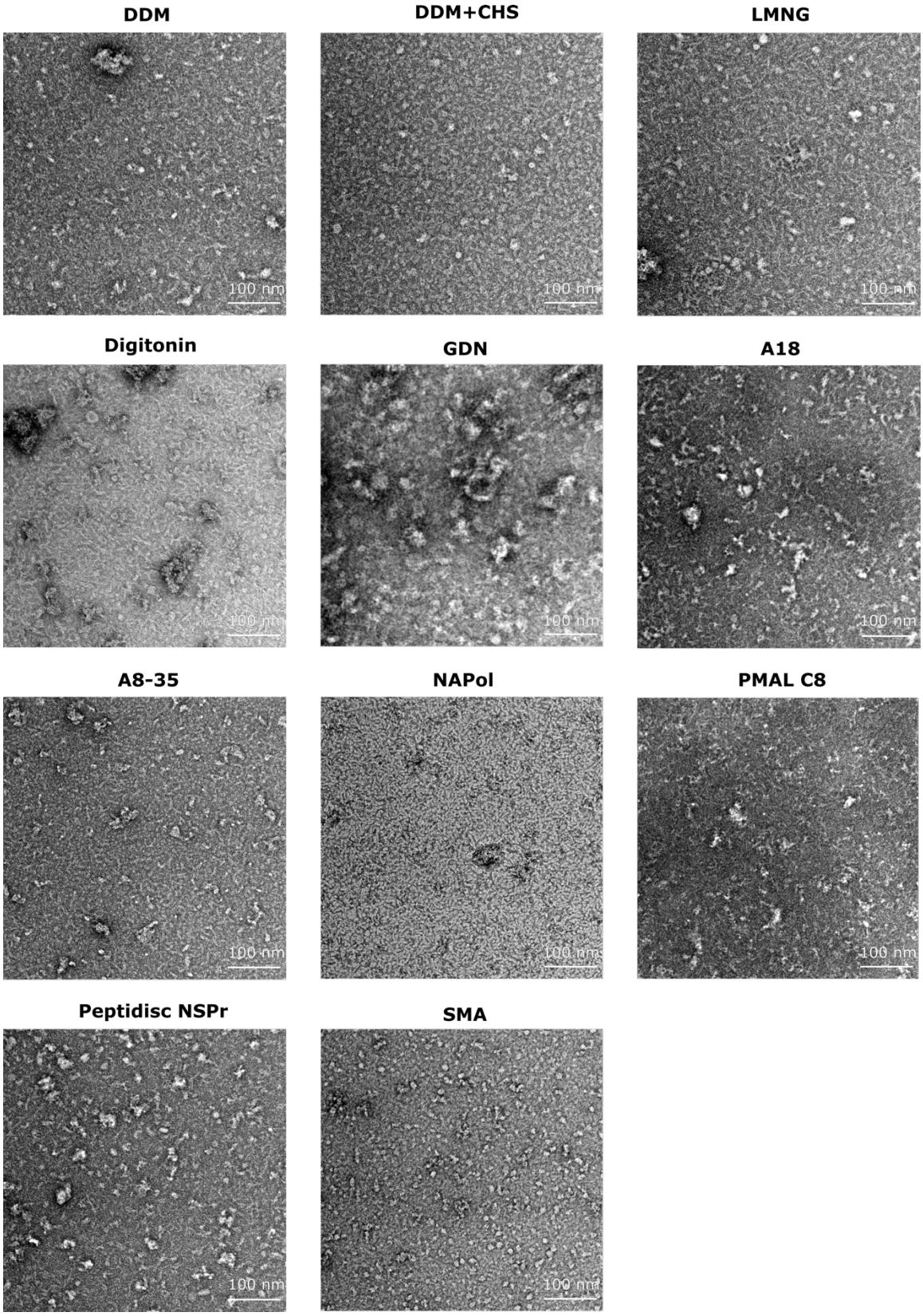


Figure 4.14. Negative stain EM of the solubilisation screen of YdbH. A representative micrograph of YdbH in each of the membrane mimetics tested is shown. Scale bar = 100 nm.

The expression conditions of YdbH FL of *E. coli* C43(DE3) cells in LB media were utilised. Subsequent Ni-NTA purification and SDS-PAGE analysis did reveal two bands for the expected molecular weights of 97 kDa for YdbH and 12 kDa for YdbL but did not reveal a band for YnbE at 7 kDa (Figure 4.15B). It was important to note a significant reduction in yield of YdbH here compared to the purifications of YdbH alone, which questions the identity of the 97 kDa band. An anti-His western blot confirmed that the His-tagged YdbL protein was present (Figure 4.15C). However, a stronger band at ~3x the MW of YdbL was seen, as well as some higher bands at ~70 kDa. This indicates that YdbL could form a trimer and potentially higher order oligomers. The molecular weights of these higher bands do not align with any complex formation between YdbH, YnbE and YdbL.

Pooled fractions from the Ni-NTA purification were subject to SEC, revealing a single peak at ~10 mL with a leading and tailing shoulder, and some smaller peaks eluting later (4.15D). Peak fractions were analysed by SDS-PAGE, revealing that the unexplained bands from the Ni-NTA purification and western blot were again present (Figure 4.15E). The peak at 17 mL revealed a band at ~12 kDa, which was likely monomeric YdbL eluting later than the higher order complex that had eluted at ~11 mL. Due to the lack of evidence of proper complex formation, further efforts at purifying this complex were not made.

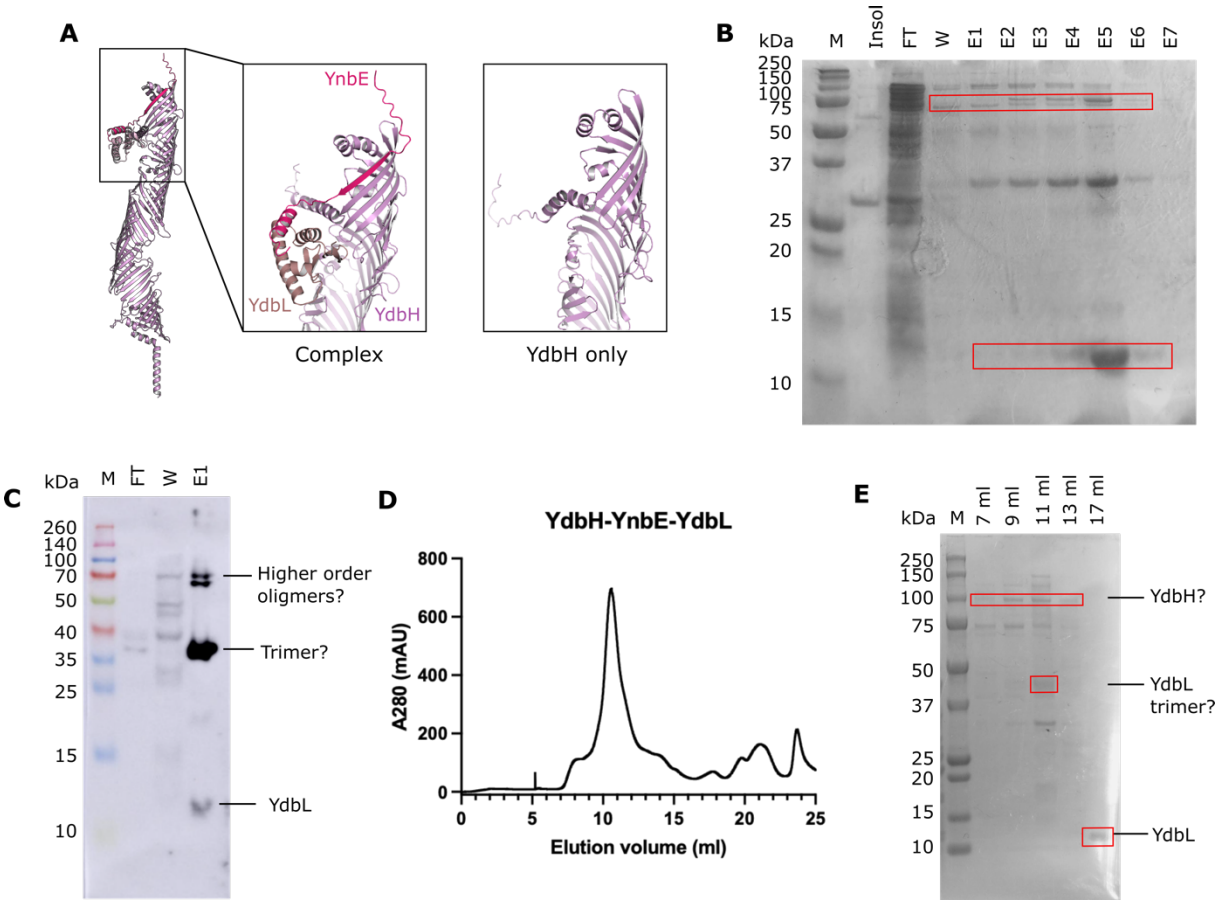


Figure 4.15. Purification of the putative YdbH complex. A – AlphaFold multimer prediction of the YdbH-YnbE-YdbL complex, showing β-augmentation of the last β-strand in YdbH (purple) by YnbE (pink). YdbL (beige) binds next to the exposed hydrophobic cavity. B – SDS-PAGE gel of the Ni-NTA affinity purification of YdbH-YnbE-YdbL. C – Anti-His western blot of the Ni-NTA affinity samples of the YdbH-YnbE-YdbL complex. D – SEC trace of the pooled Ni-NTA elution fractions on a Superdex 200 10/300 column. E – SDS-PAGE gel of the peak fractions from the SEC of YdbH-YnbE-YdbL. M – Marker; Insol – insoluble material after the first spin; FT – flow through; W – wash; E1 – 1st mL of elution fraction etc.

4.2.3.3 YhjG and YicH overexpressed and purified relatively well

YhjG was highly overexpressed in the initial test conditions of *E. coli* C43(DE3) cells in LB media, with subsequent Ni-NTA purifications purifying a protein at the expected molecular weight of ~75 kDa, as evidenced by the intense band (red box) on the SDS-PAGE gel in Figure 4.16A. Pooled fractions were further purified by SEC using a Superdex 200 10/300 column, with the main peak eluting at ~10 mL (Figure 4.16B). This eluted earlier than expected for a

protein of this size, but this was explained by the presence of a detergent micelle and its predicted stick-like structure resulting in a larger hydrodynamic radius than globular proteins of the same MW. The leading shoulder that eluted in the void volume was assumed to be aggregated YhjG. SDS-PAGE of the peak fractions revealed a mostly pure band at ~75 kDa, with some minor contaminants at ~100 kDa and ~60 kDa (Figure 4.16C).

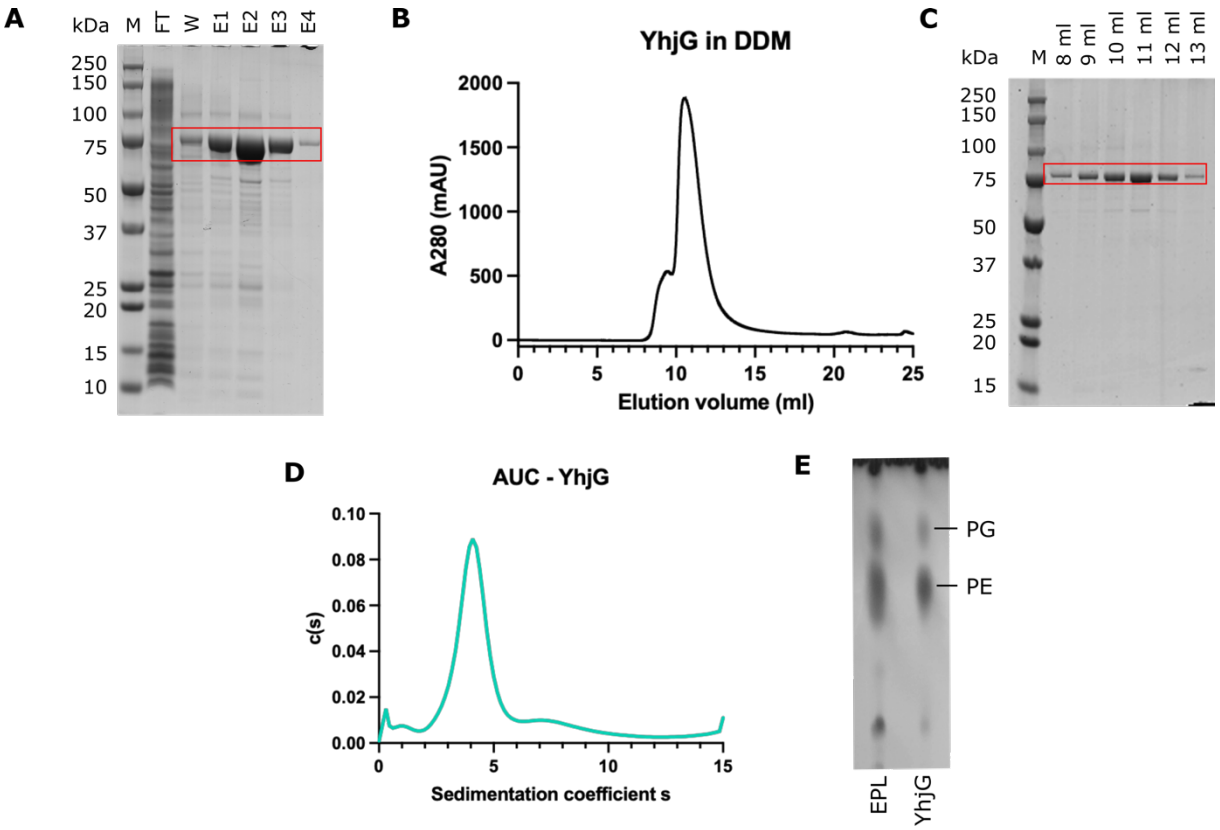


Figure 4.16. Purification of YhjG. A – SDS-PAGE gel of the Ni-NTA affinity purification of YhjG. B – SEC trace of YhjG on a Superdex 200 10/300 column. C – SDS-PAGE gel of the elution fractions of YhjG post-SEC. D – Sedimentation velocity data from the analytical ultracentrifugation of YhjG. E – Thin layer chromatography of lipids extracted from YhjG. FT – flow-through; W – wash; E1 – 1st mL of elution etc.; c(s) – sedimentation distribution; EPL – *E. coli* polar lipids; PG – phosphatidylglycerol; PE – phosphatidylethanolamine.

The purified YhjG protein was relatively stable compared to the other AsmA-like proteins and could withstand concentrations up to ~3 mg/mL. Due to this and its relative purity, further biophysical and biochemical characterisation could be performed. Analytical

ultracentrifugation (AUC) of the purified protein revealed that YhjG was most likely monomeric and homogenous in solution, with a single peak in the sedimentation velocity analysis of the data at s = 4 and a mass estimate of 64 kDa (Figure 4.16D). A minor peak at s = 7 was also present, which could indicate the presence of a higher-order species. Thin-layer chromatography of chloroform-methanol extracted lipids from the purified sample was performed, revealing two spots identified as the phospholipids PG and PE (Figure 4.16E). This indicated that YhjG co-purified with these phospholipids, even with detergent present throughout.

The final AsmA-like protein, YicH, overexpressed well in the initial expression conditions of *E. coli* C43(DE3) cells in LB media. SDS-PAGE analysis of the Ni-NTA purification of YicH revealed an intense band in the elution fractions at the expected molecular weight of 64 kDa (Figure 4.17A). Elution fractions were pooled prior to SEC on a Superdex 200 16/600 column, which revealed a single peak that eluted at ~72 mL (Figure 4.17B). Peak fractions were then analysed by SDS-PAGE, revealing a single band at 64 kDa. This protein was the most stable of the AsmA-like proteins and could be concentrated up to ~5 mg/mL. Due to this, crystallisation trays were set up, but no crystal hits were found.

With a high yield and purity of YicH, further characterisation could be performed. AUC revealed that the majority of the protein was likely monomeric and homogeneous in solution, with a single major peak in the sedimentation velocity analysis and an estimated molecular weight of 63 kDa from mass distribution analysis (Figure 4.17D). There was also a minor peak at s = 7, which could indicate a higher-order oligomer. Thin-layer chromatography was performed on chloroform-methanol extractions of the purified YicH sample, which revealed the presence of phospholipids (Figure 4.17E). Comparison to an *E. coli* phospholipid standard revealed that both PG and PE were co-purified.

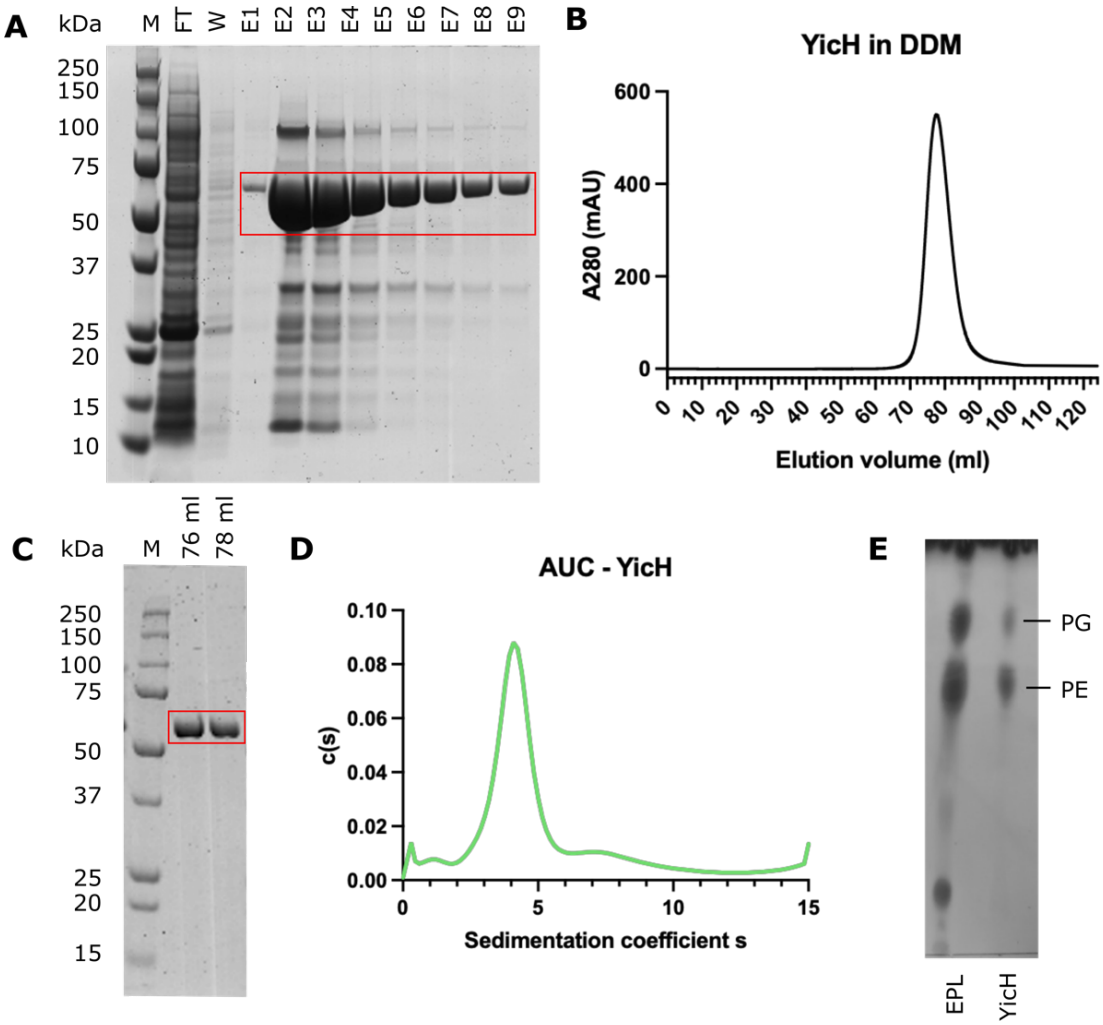


Figure 4.17. Purification of YicH. A – SDS-PAGE gel of the Ni-NTA affinity purification of YicH. B – SEC trace of YicH on a Superdex 200 16/600 column. C – SDS-PAGE gel of the elution fractions of YicH post-SEC. D – Sedimentation velocity data from the analytical ultracentrifugation of YicH. E – Thin layer chromatography of lipids extracted from YicH. FT – flow-through; W – wash; E1 – 1st mL of elution etc.; c(s) – sedimentation distribution; EPL – *E. coli* polar lipids; PG – phosphatidylglycerol; PE – phosphatidylethanolamine.

4.2.3.4 YhjG 30-686 and YicH 23-569 expressed well but were insoluble

The formation of soluble constructs of YhjG and YicH would improve the amenability to structural and functional techniques. The presence of detergent or other membrane encapsulation agents often introduces heterogeneity and can increase the background noise in a variety of techniques. The soluble construct of YhjG was designed to remove the N-terminal helix, residues 1-29 (Figure 4.18A). An expression trial was performed that identified strong

overexpression in both LB media and TB media using *E. coli* BL21(DE3) cells, with a band at the expected weight of ~72 kDa (Figure 4.18B). Initial purifications from cells grown in TB and LB media revealed no band at the expected MW. The insoluble and soluble material from the first centrifugation of cell debris post-cell lysis were analysed by SDS-PAGE, revealing a strong band for YhjG 30-686 within the insoluble material of both grows. This indicated that the protein expressed here was insoluble or otherwise attached to the membrane fraction. A construct of YhjG without the hydrophobic C-terminal helix-turn-helix (residues 642-666) that may be capable of binding to the OM was also attempted to be made, but site-directed mutagenesis was unsuccessful.

Similarly, a soluble construct of YicH was designed by removing the residues 1-26 (Figure 4.18D). Expression trials revealed prominent overexpression in *E. coli* BL21(DE3) cells in TB media with a band at ~60 kDa present in the post-induction sample (Figure 4.18E). However, initial purification revealed no protein. SDS-PAGE analysis of the soluble and insoluble fraction post-cell lysis revealed a more prominent band for YicH 27-569 (~60 kDa) in the insoluble fraction (Figure 4.18F). Again, this could be due to the protein being insoluble or somehow attached to the membrane.

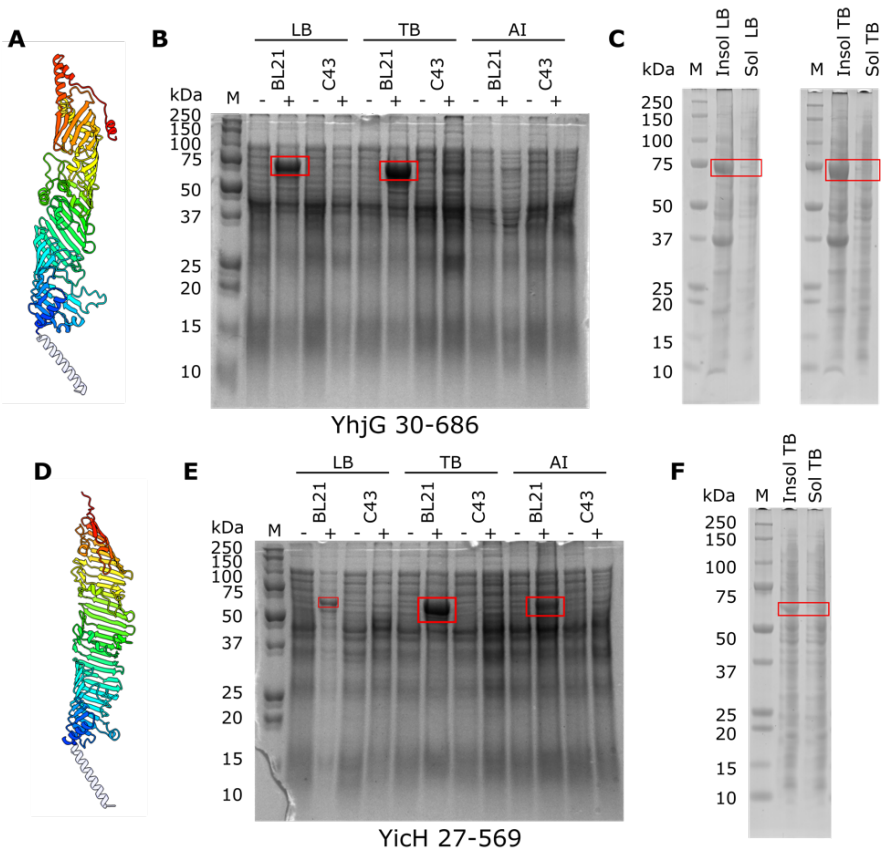


Figure 4.18. Expression trials of YhjG 30-686 and YicH 27-569. A – Construct design of YhjG 30-686. B – SDS-PAGE gel of an expression trial of YhjG 30-686. C – SDS-PAGE gel of the soluble and insoluble fractions of *E. coli* BL21(DE3) cells grown in TB and LB media after the first spin post-cell lysis. D – Construct design of YicH 27-569. E – SDS-PAGE gel of an expression trial of YicH 23-569. F – SDS-PAGE gel of the soluble and insoluble fractions of *E. coli* BL21(DE3) cells grown in TB media after the first spin post-cell lysis. Expression trials were performed in *E. coli* C43(DE3) and *E. coli* BL21(DE3) in LB, TB or AI media. AI - auto-induction; LB – Miller's lysogeny broth; TB – terrific broth. '-' and '+' indicate pre- and post-induction samples respectively.

4.3 Discussion

4.3.1 Varying success of expression and purification for the AsmA-like proteins

Recombinant protein expression was used to produce the AsmA-like proteins for further study with varying levels of success. YhdP and TamB were particularly challenging to express with low yields across many different expression conditions and a scale-up of expression required to perform trial purifications. The other AsmA-like proteins were more easily overexpressed. This could potentially indicate a more important role for YhdP and TamB in the cell compared to the other AsmA-like proteins. With the evidence of YhdP and TamB as phospholipid transporters, their overexpression likely causes significant disruption to membrane biogenesis and integrity (Ruiz et al., 2021; Douglass et al., 2022; Cooper et al., 2023). This could be toxic to the cell and may have meant that only low expressing cells grew. Additionally, other pathways within the cell may regulate the amount of YhdP and TamB, with proteases (e.g. DegP) degrading the unfolded or potentially misfolded protein (Pan et al., 2003). This could also explain the persistent breakdown products seen throughout protein purification trials.

Both proteins are predicted to span across the periplasm and into the OM. Perhaps they need an OM component (e.g. the OMP TamA for TamB) to encourage correct folding and protein stability (Shen et al., 2014). Truncated constructs of YhdP were designed, which would eliminate the need for an OM binding partner. Unfortunately, whilst these constructs saw improved expression, solubilisation and behaviour on gel filtration, there was still a consistent issue with protein breakdown and a pure sample was not obtained. The successful expression and purification of the full TamAB complex has since been reported for an investigation into the OMP folding capabilities of the TAM (Wang et al., 2024a). The authors note that there was difficulty in expressing and purifying the TamAB complex and note that they could not purify TamB alone. Even as a complex, they report high levels of TamB proteolysis, low expression

levels and the co-purification of many non-specific contaminants, which have all been issues encountered in this study.

Trials of the other AsmA-like proteins revealed much higher levels of overexpression. The shorter AsmA-like proteins may not be as important for cell function as YhdP and TamB, and they may not span the entire periplasm, which could mean an OM component is not needed for correct folding. With the genetic evidence for YdbH being a functionally redundant phospholipid transporter along with YhdP and TamB, the difference in behaviour is surprising (Ruiz et al., 2021; Douglass et al., 2022). Unlike YhdP and TamB, YdbH expressed without an OM binding partner at high levels. Though high expression was seen for YdbH, severe protein breakdown was a consistent issue.

Expression of the YdbH complex to potentially combat this instability was not successful, with bands at unexplained molecular weights present even after SEC. This may now be explained by the recent biochemical evidence of complex formation between YdbH and YnbE only (Kumar et al., 2024). The authors suggest YdbL plays a chaperone-like role, preventing extensive multimerization of the small YnbE lipoprotein. Perhaps because of this, the His-tagged YdbL was unable to co-purify with YdbH and YnbE. Mass spectrometry of the ~100 kDa band would be needed to confirm whether it was indeed from the co-purification of YdbH, or if it was a contaminating protein. Mass spectrometry of the other bands present may also be informative, perhaps revealing identities of proteins that co-purify with, and are potentially also chaperoned by, YdbL. Future experiments should focus on expressing and purifying a complex of only YdbH and YnbE to improve the stability of YdbH, as this would likely still protect the exposed C-terminal β-strand.

The soluble constructs trialled in this study all showed high levels of overexpression, including the YhdP 30-1266 construct. This difference in expression could indicate that it is the localisation of YhdP to the membrane that is limiting the overexpression of the IM-tethered YhdP constructs. Unfortunately, most of the constructs with a deletion of the N-terminal helix were found in the insoluble fraction after cell lysis. This was likely due to the exposure of hydrophobic residues within the predicted cavity upon deletion of the TM helix and the absence of a membrane mimetic to protect these regions. Interestingly, this was not an issue for AsmA, which as a full-length protein was insoluble, but as a truncated protein (AsmA 23-617), could be purified from the soluble fraction of the cell.

The solubilisation screens performed here revealed novel information to inform future experiments, particularly for YdbH. The screen was less informative for YhdP, TamB and AsmA, with improved expression likely needed for better characterisation of the detergents and other membrane mimetics tested. The improved protein stability and particle behaviour on grids from the YdbH nanoDSF and negative stain data resulted in the further use of amphipols as membrane mimetics during the remainder of the study. Further optimisation of cryo-EM grids of YdbH in A8-35 should be a priority for further work based on the promising negative stain micrograph presented here.

The two most successfully expressed and purified proteins were YhjG and YicH. Information about these two proteins is very limited, with references to them as being members of the AsmA-like family in the literature but no assigned structure or function. The improved protein stability and yield allowed further characterisation of the proteins as likely monomeric proteins that could co-purify with *E. coli* phospholipids. This exciting result provided evidence towards a putative involvement in phospholipid transport within the cell envelope.

With consistent issues in the protein expression and purification of YhdP, TamB and AsmA, the project focus for further structural and functional elucidation of the AsmA-like proteins turned to the other three proteins: YdbH, YhjG and YicH. It should be noted for the context of future chapters that whilst the purification of these proteins was significantly improved compared to the other AsmA-like proteins, there were still protein stability issues. Snap-freezing and thawing of the protein aliquots resulted in aggregation and the proteins were only stable at 4°C for a few days. This resulted in increased experimental complexity, particularly in terms of having a short time limit to perform experiments before sample aggregation. Nevertheless, these proteins were able to be further characterised, as Chapters 5 and 6 will discuss.

4.3.2 Recommendations for future work

There are seemingly endless ways that can be trialled to improve recombinant protein expression in *E. coli*. With the poor expression yields of YhdP and TamB found here, it is unlikely that further changes in expression conditions would result in significantly improved expression. Cell-free expression systems could instead be trialled, which have seen recent progress in mimicking a lipid bilayer and in fine-tuning expression (Steinkühler et al., 2024; Jiang et al., 2024). However, with the scale-up to 6 L of culture media producing enough protein for purification trials, further optimisation of purification steps can be performed. A priority for future work should be the use of alternative purification tags that are less likely to result in contaminants. Primers were designed to change the His-tag to a Strep-tag during this study, but time limitations and changes in project direction meant that this project never came to fruition. With even a small improvement in protein yield and purity, biophysical assays like nanoDSF could be performed to assess protein stability in a variety of conditions (Birch et al., 2020). These recommendations can also be applied to the other AsmA-like proteins.

Future studies on expressing YhdP, TamB, YdbH and YdbH in a complex could also utilise the recently published optimal conditions found for TamAB expression: a lower concentration of 50 μM IPTG to reduce the effect of toxic overexpression, incubation of the cells on ice at time of induction (OD_{600} = 0.4 – 0.5), cell lysis only by cell disrupters (not sonication or protein extraction reagents) and 12 L of 2xYT media (Wang et al., 2024a). Additionally, the purification of TamAB required PBS buffers and DDM for solubilisation, and they found a significant disruption of TAM activity when the following were present in purification buffers: protease inhibitor cocktails, β-mercaptoethanol, glycerol and the detergent LDAO. With such intricate measures needed to improve the expression and stability of the TamAB complex, it explains the difficulty in the expression and purification of some of the AsmA-like proteins trialled in this study.

Alternatively, computational protein design methods could be employed to improve protein stability. RFDiffusion is a machine learning (ML)-based *de novo* protein design tool that can be used to create stabilising peptides against regions of instability (Watson et al., 2023). The design of a peptide that complements the final β-strand of YdbH could result in improved stability. Similarly, a "plug" could be designed to protect the exposed hydrophobic residues upon removal of the N-terminal helix in the "soluble" constructs. There have also been advances in the rational design of making membrane proteins water-soluble through a process called "QTY design" (Li et al., 2024). This converts hydrophobic residues to one of glutamine (Q), threonine (T) and tyrosine (Y), effectively reversing the hydrophobicity. This could be applied to regions of hydrophobic residue exposure within the AsmA-like proteins (e.g. N-terminal TM helix, opening to the hydrophobic cavity, C-terminal regions). A ML-based approach for this has also been reported, using a combination of AlphaFold structure prediction and ProteinMPNN sequence design to convert membrane-bound regions to water-soluble

regions whilst retaining the fold of an input structure (Goverde et al., 2024). However, whilst these approaches may improve solubility and allow structural investigation, they would also likely impact their putative phospholipid transport function.

4.3.3 Conclusion

The AsmA-like protein family were overall challenging to produce, with consistent issues with low expression and protein instability. The initial project focus was on the characterisation of YhdP, and so many expression and purification conditions and different construct designs were trialled with little success. Initial expression and purification trials of the other AsmA-like proteins revealed that YdbH, YhjG and YicH were highly overexpressing and more amenable to purification. For this reason, the project focus moved towards these three proteins. Further characterisation of YhjG and YicH was performed, revealing that they co-purified with *E. coli* phospholipids, providing evidence for the putative phospholipid transport function. Further work utilising different construct designs, conditions and the use of engineered proteins could improve the purification and stability of the other AsmA-like proteins in future investigations.

CHAPTER 5. STRUCTURAL CHARACTERISATION OF YICH AND YHJG

5.1 Introduction

The structural information for the AsmA-like protein family is limited to AlphaFold structural predictions, low-resolution negative stain 2D classes of YhdP and the β-taco crystal structure of a portion of TamB (Cooper et al., 2023; Josts et al., 2017). The lack of structural knowledge is most likely attributed to the complexity of both the expression and purification of this family of proteins and their intrinsically flexible nature. Nevertheless, further structural information is required to elucidate their proposed role in inter-membrane lipid transport.

Distantly related eukaryotic bridge-like lipid transport proteins (BLTP) have been structurally characterised, with the high-resolution structure of an N-terminal fragment of a BLTP called LPD-3 revealing lipid densities throughout the hydrophobic cavity, further solidifying their role as lipid transporters and implying continuous, diffusive transport (Kang et al., 2024). The determination of a protein structure could therefore reveal similar information for the AsmA-like proteins, potentially identifying a bound ligand and revealing a possible transport mechanism (Biou, 2023). With evidence for both OMP transport and PL transport via the TamAB complex, identification of a ligand would address this controversy (Josts et al., 2017; Ruiz et al., 2021; Wang et al., 2024a). It could also explain the potential cargo preference between the AsmA-like proteins seen in some *in vivo* studies (Rai et al., 2024).

Structural determination typically requires the protein of interest to be of sufficient purity and quality. Therefore, due to the various issues in protein expression, purification or stability that were seen for YhdP, TamB, YdbH and AsmA, these proteins were not able to be structurally characterised. The other two AsmA-like proteins, YhjG and YicH, were more amenable to

protein expression and purification. Additionally, very little is known about their functional role in the cell from genetic studies. Structural characterisation of these two proteins is therefore desirable to hypothesise proposed functions and mechanisms of action. The approaches taken to determine the structures of YhjG and YicH could also be applied to the determination of the larger AsmA-like proteins due to the predicted structural similarity.

5.1.1 Introduction to negative stain-EM and cryo-EM

Transmission electron microscopy (TEM) utilises the very short wavelength of electrons to image electron dense particles. The technique has been used for decades, with negative stain EM giving low resolution but high contrast information, and cryogenic EM (cryo-EM) giving higher resolution but low contrast information (Bhella, 2019; Gallagher et al., 2019). Only in the last decade or so has the cryo-EM "resolution revolution", a by-product of drastic improvements to instrumentation and processing, meant that high-resolution protein structures could be determined (Neumann et al., 2018). This has resulted in an exponential growth of the use of cryo-EM and deposition of cryo-EM structures in the PDB, now rivalling X-ray crystallography as the technique of choice for structural biologists.

5.1.1.1 Sample preparation

To assess initial sample quality, negative stain EM can be performed. This is a lower resolution technique but requires less specialist equipment than cryo-EM. Negative stain grids are prepared by glow discharge of an EM grid, typically a copper mesh grid with continuous carbon film, which both removes organic contaminants and creates a hydrophilic surface for sample adsorption, followed by the addition of a small amount of very low concentration protein sample (Gallagher et al., 2019). Excess sample is blotted away from the grid before coating with a heavy metal stain such as uranyl acetate (Figure 5.1A). The heavy metal salts create a

mask of the particles on the grid and absorb electrons more readily, leading to a high-contrast image with low-resolution features such as particle shape and size.

Analysis of protein structure by cryo-EM requires the preparation of an EM grid with monodispersed particles at multiple orientations. Interactions with the support film and with the air-water interface are common, as are protein aggregation, preferential orientations of particles and uneven ice thickness across the grid (Drulyte et al., 2018). These can all make EM data collection and processing difficult, making sample preparation optimisation a major bottleneck within the whole process. This can involve extended periods spent optimising the protein sample to be homogenous prior to grid preparation, and then the screening of multiple protein concentrations, buffer components, grid types and blotting parameters to assess particle behaviour on grids.

Cryo-EM grid preparation requires cryogenic temperatures throughout. This is to ensure the rapid freezing of aqueous solutions such that amorphous ice is formed, vitrifying the sample in an almost native state (Dubochet et al., 1988). The vitrification also minimises radiation damage from electrons in the microscope (Bammes et al., 2010). To do this, glow-discharged EM grids are loaded with the protein sample, blotted to remove any excess, and plunge frozen into a cryogen such as liquid ethane, which has a higher heat capacity and higher thermal conductivity than liquid nitrogen, pre-cooled to cryogenic temperatures by the colder liquid nitrogen (Figure 5.1B). The EM grids are often made of a copper or gold mesh and are covered with perforated foil. Typically, a carbon support film is used at first, with further optimisation potentially using lacey carbon, graphene or graphene oxide supports, which can reduce background noise, or the use of gold supports, which often minimise beam-induced motion and help with particle orientation (Bammes et al., 2010).

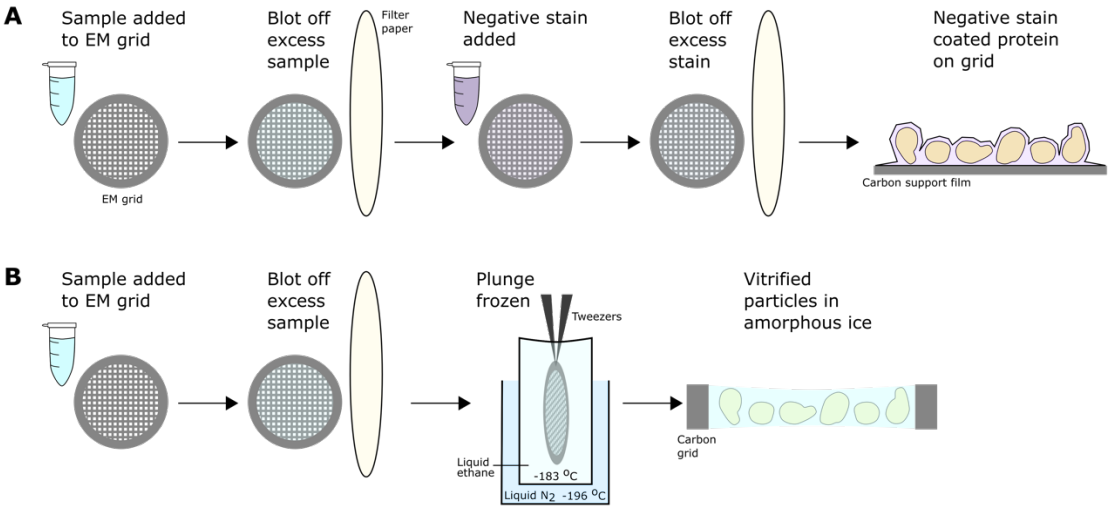


Figure 5.1. Grid preparation for electron microscopy (EM). A – Negative-stain EM grid preparation involves adding the sample to an EM grid, blotting off the excess sample, adding a negative stain salt such as uranyl acetate, blotting off the excess stain and leaving to air dry to produce a negative stain layer over the protein on the grid. B – Cryo-EM sample grid preparation involved adding the sample to an EM grid in high humidity, blotting off the excess sample and plunge freezing into a cryogen such as liquid ethane pre-cooled with liquid nitrogen to create a vitreous ice layer, keeping the protein in a hydrated state.

5.1.1.2 Electron microscopes

To image the sample, a transmission electron microscope is used. A typical electron microscope consists of an electron source, lenses that focus the electrons into a beam, apertures and a detector (Figure 5.2) (Thompson et al., 2016). The electron source is most commonly a field emission gun (FEG) in newer microscopes, where electrons are generated from a cathode tip with a supplied voltage and released into the column. The lenses then direct the beam towards the sample by creating electromagnetic fields that focus the electrons. Apertures within the microscope prevent highly scattered electrons that deviate from the optical axis from reaching the detector to refine the signal. These lenses and apertures mean the beam can be manually adjusted to change the size and shape of the electron beam, allowing for accurate control over where and with how much dose the electron beam hits the grid. The condenser system focuses

the beam onto the sample where the objective lens system then forms the image, and finally, the projection lens amplifies the signal prior to reaching the detector. The advent of direct electron detectors, which convert the electrons detected into an output of intensity and record movies instead of a single image, was fundamental to advancing the resolution revolution (Neumann et al., 2018). The multi-frame movies can then be processed to correct for beam-induced motion. The whole system is kept under vacuum to minimise noise and to help maintain a coherent beam.

As the wavelength of electrons is 0.0197 Å (at 300 kV), much shorter than photons, theoretically, a very high resolution could be achieved using this technique. However, various practical considerations limit this potential resolution, most from the electron microscope itself (Zhang and Hong Zhou, 2011). Spherical aberrations which arise from electrons being incorrectly focussed by the objective lens and chromatic aberrations which are caused by electrons not being focussed onto the same plane of focus both prevent the theoretical resolution from being reached. Another issue arises when waves of electrons that are perpendicular to each other are not focussed equally, leading to axial astigmatism. Aberration correctors, magnetic field correctors and stigmator coils have been created to account for these issues, which has led to resolutions nearing 1 Å now able to be achieved (Yip et al., 2020; Ratkevičiūtė, 2023).

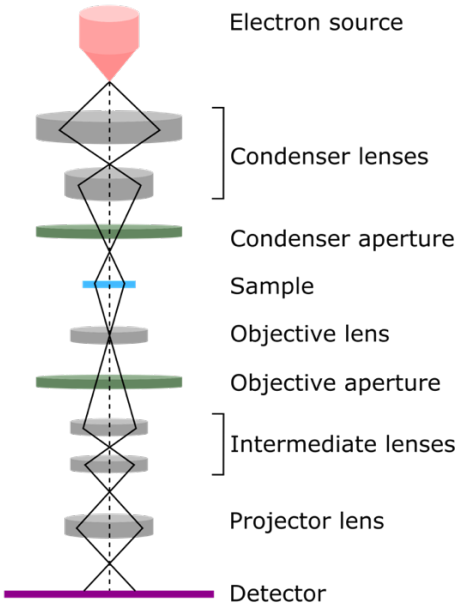


Figure 5.2. Diagram of a transmission electron microscope. The electron beam in black is produced by the electron source and focussed through a variety of lenses and apertures, passing through the sample and reaching the detector screen. Figure adapted from (Wotherspoon, 2022).

5.1.1.3 Image formation

When incident electrons interact with the sample, they can be unscattered, continuing through to the detector, scattered elastically or scattered inelastically. Most electrons pass through the sample unscattered. However, elastic scattering is the event that is most useful for image formation and involves the electron changing direction but maintaining its incident energy (Franken et al., 2020). Inelastic scattering involves a transfer of energy to the sample, which generates free radicals and subsequently causes radiation damage. In negative stain EM, the heavy metal salts have a large atomic number, whereas in cryo-EM, often only C, N or O atoms are present, which all have low atomic numbers. The higher the atomic number, the more electrons around the nucleus, resulting in higher scattering angles.

In negative stain EM, the image is mostly formed from amplitude contrast, as the strong electron scattering from the heavy metal salts causes high contrast levels between areas of high and low

electron density (Thompson et al., 2016). Proteins and other biological material show up light in colour, with dark regions contributed by the heavy metal stain.

In cryo-EM, the image is mostly formed from phase contrast, which is generated from the interference effects between the unscattered and elastically scattered electrons (Thompson et al., 2016). The contribution of each scattering angle to the final image can be described by the Contrast Transfer Function (CTF) (Figure 5.3, blue). This represents the positive and negative amplitudes that are contributed at different angles or spatial frequencies.

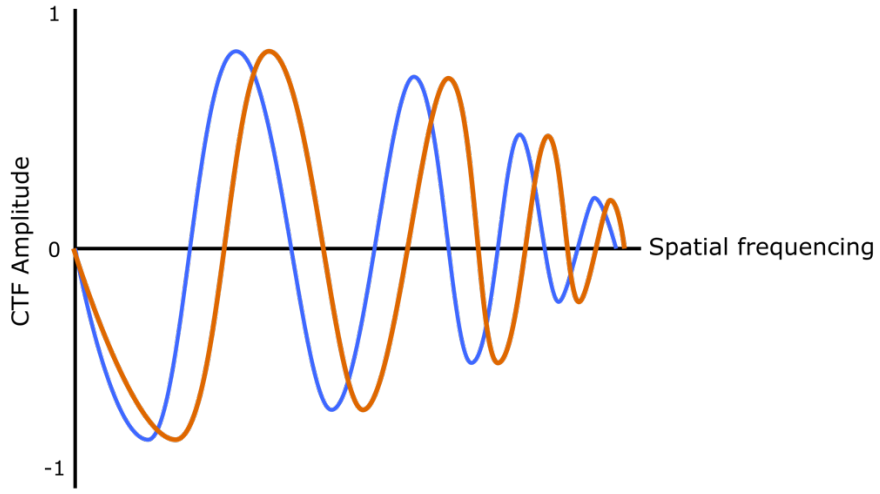


Figure 5.3. CTF diagram. The curves show the positive and negative amplitudes that are contributed at different angles or spatial frequencies for image formation in TEM. Blue and orange curves represent the CTF curves from two different defocus values, resulting in non-zero contrast values at all spatial frequencies.

Due to the low scattering potential of proteins, cryo-EM images at absolute focus contain little focus (Sigworth, 2016). To counteract this, images are recorded at various defocus levels, which will change the phase contrast observed due to changing the path lengths of scattered electrons. Information from across the different defocus levels is then used to generate contrast information at all spatial frequencies (Figure 5.3, blue and orange). Generally, at high levels of defocus, low spatial frequencies contribute more contrast and vice versa. Obtaining contrast

information for both low and high spatial frequencies is important to maintain a high enough level of contrast to identify particles but to also retain high-resolution information.

5.1.1.4 Data processing

Single particle analysis (SPA) is used to reconstruct TEM data and produce a model of a single particle (Figure 5.4) (Sigworth, 2016). The first most important step is to perform motion correction of each frame within a single movie collected, resulting in a single micrograph that is an average of the individual movie frames. This accounts for beam-induced motion and small amounts of stage drift. It is then vital to perform CTF estimation by comparing the experimental data of each micrograph with theoretical CTF curves at different defocus levels. Each micrograph produces a 2D Fourier power spectrum with spherical Thon rings caused by phase contrast, that are used to visualise the CTF curve produced from that micrograph. Once defocus levels are estimated, CTF correction can then be performed, wherein negative amplitudes are flipped to become positive and values from each defocus curve are merged.

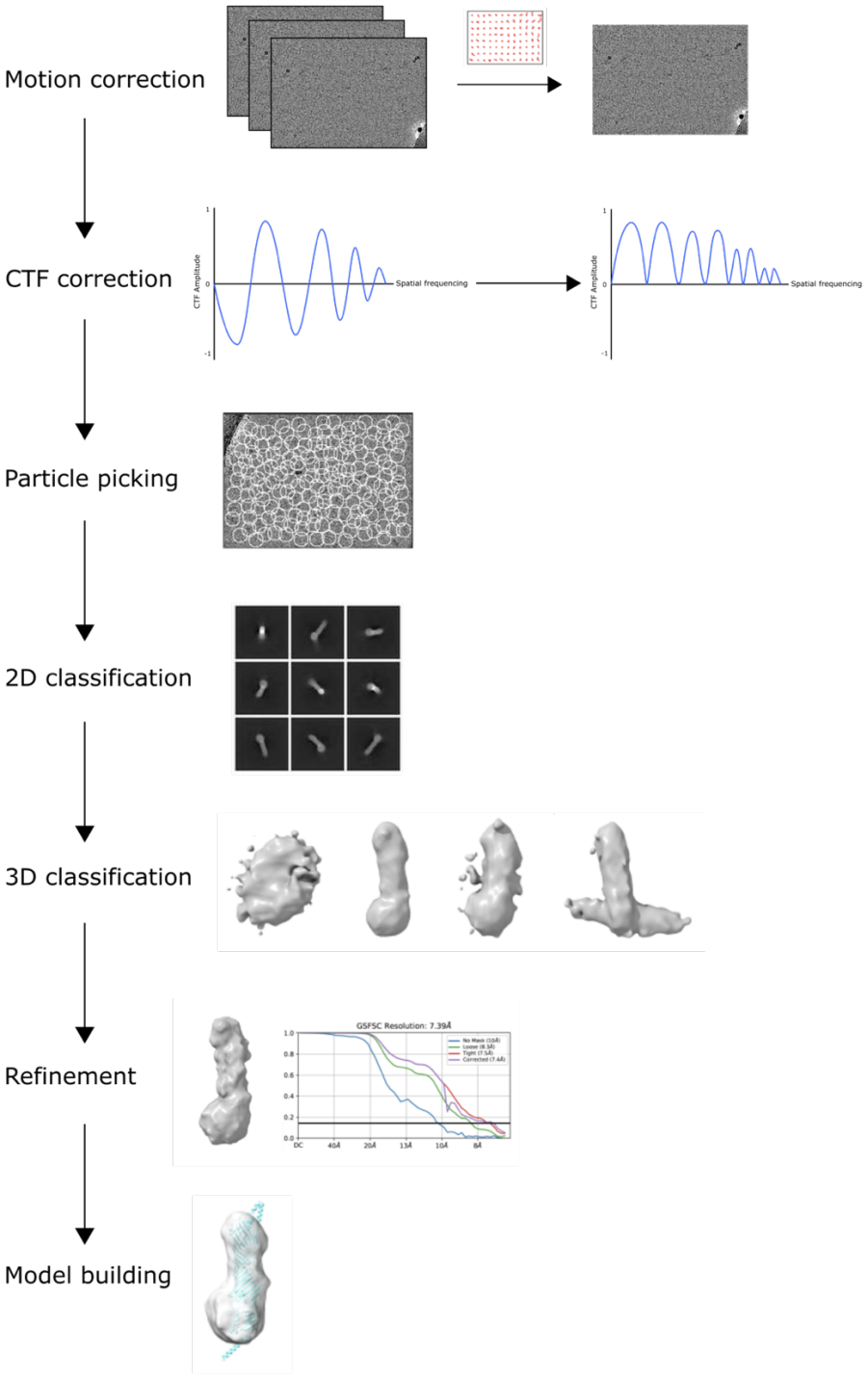


Figure 5.4. Single particle analysis (SPA) processing pipeline. Movie frames are motion corrected into an averaged micrograph. The Contrast Transfer Function (CTF) is then estimated and corrected to achieve positive amplitudes at all spatial frequencies. Particles are then picked before 2D classification, which shows particles from various views. Good 2D classes are then used to generate 3D models, which can be further refined. A map at sufficiently high resolution can then be used to model the protein.

The next step is to perform particle picking. This can be performed manually or by using various automatic particle-picking programs. Manual picking is often only used to assess small numbers of micrographs for initial processing, as it can introduce significant particle conformation and orientation biases. Automatic particle picking programs include template-based pickers, where the computer searches the micrographs for particles representing a supplied reference image, blob pickers, which identify any potential particle within a set size restraint, and machine learning algorithms, which can be used with pre-trained models or with a model trained on a small subset of particles (Sigworth, 2016; Bepler et al., 2019; Wagner et al., 2019).

2D classification is then performed to group particles into classes representing different views of a particle (Sigworth, 2016). This allows assessment of particle quality and particle orientation, with an ideal scenario of classes showing multiple different orientations of a single particle. This procedure also groups together junk particles, which can then be easily removed prior to downstream processing.

3D classification is then performed using the 2D classes to generate *ab initio* models. The Fourier slice theorem states that the Fourier transform of a 2D projection is equivalent to the Fourier transform of a slice of the 3D model generated (Sigworth, 2016). Therefore, if the projection angles are known for each 2D class, these can be projection-matched to iteratively improve a starting model. Multiple 3D volumes will be generated that match the 2D classes, and if there is any conformational heterogeneity within the particle set, this can result in the separation of these conformations into different 3D classes.

Refinement of the generated 3D volume is then performed to improve the map quality and to estimate map resolution. For this, particles are split into two half sets which are then independently refined and compared to each other using the Fourier shell correlation (FSC)

criterion. The FSC curve describes the correlation across spatial frequencies, with values decreasing from 1 to 0 as the resolution increases. The gold standard FSC cut-off of 0.143 is used to describe global map resolution (Rosenthal and Henderson, 2003). Further refinement can be performed, including the use of advanced processing algorithms that address non-uniformity, flexibility, and disordered regions. With a refined model at a sufficiently high resolution, *de novo* atomic model generation or fitting of a prior atomic structure can then be performed.

5.1.2 Chapter aims

This chapter discusses progress made in the structural determination of YicH and YhjG, utilising negative stain EM and cryo-EM. The main aims of the chapter were to:

1. Determine the structures of YicH and YhjG utilising electron microscopy techniques.
2. Determine optimal grid preparation conditions and data processing pipelines for recommendations of future structural investigations of the other AsmA-like proteins and other proteins exhibiting similar flexibility characteristics.

5.2 Results

5.2.1 Structural characterisation of YicH

5.2.1.1 Negative stain EM of YicH

The smallest of the AsmA-like proteins, YicH, has almost no structural and functional information in the literature and as discussed in Chapter 3, genetic investigations were not able to elucidate its role in the cell. This lack of knowledge, plus the ability to overexpress and purify YicH, meant structural characterisation of the protein was desirable and potentially feasible. Both protein crystallisation for X-ray crystallography determination and SEC-SAXS were employed but were unsuccessful, with no crystals forming and poor data quality from SEC-SAXS. Another approach was to investigate the protein with cryo-EM. YicH is a small protein to be investigated via EM at only 62 kDa. As a quick and easy way to check the viability of the approach, negative stain EM was first performed to see if YicH was identifiable on grids. It was also used to gain low-resolution structural information and to check sample quality prior to higher-resolution cryo-EM investigations.

Multiple concentrations of purified YicH in DDM were applied to carbon grids and stained with uranyl acetate. An optimal concentration of 26 μg/mL gave the most even distribution of particles, with a representative micrograph shown in Figure 5.5A. Stick-like particles of dimensions of about 30 Å wide and 150 Å long were visible, in agreement with the AlphaFold model of YicH. Also of note were rounder particles, which could be DDM micelles or dust, and particles that looked like protein density that were longer and had varying degrees of curvature when compared to the predicted model. This flexibility and range of particle lengths increased the heterogeneity, making both particle picking and class averaging difficult.

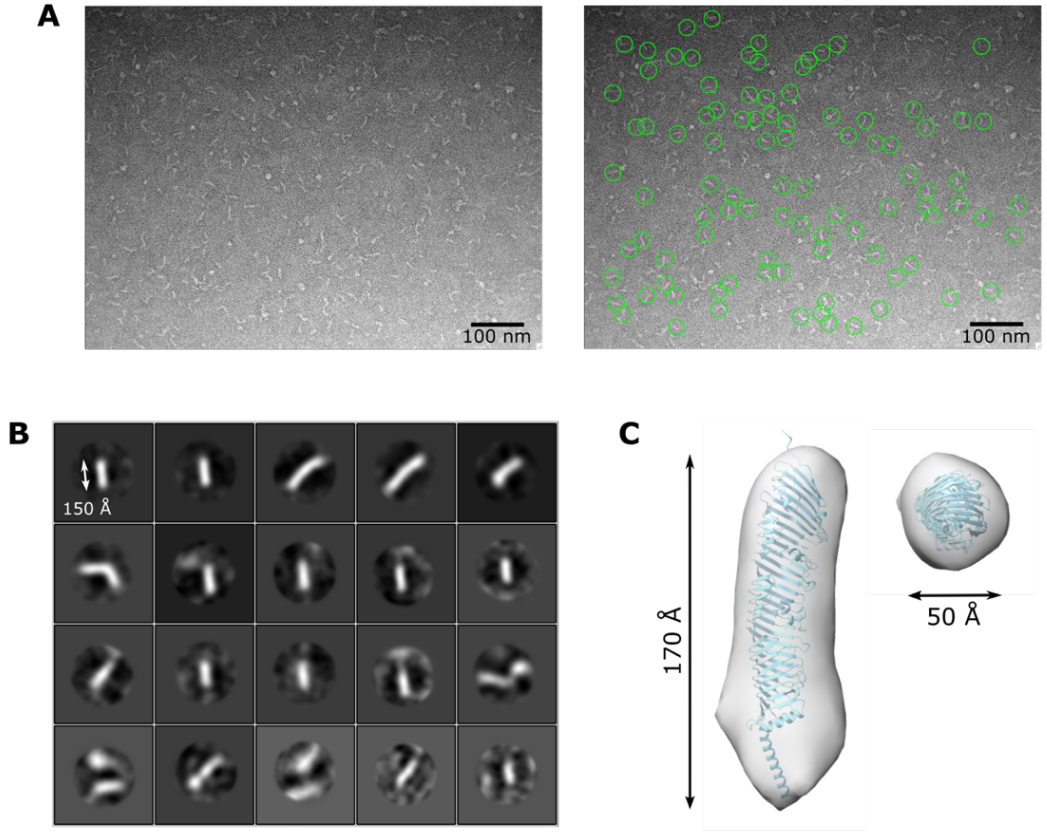


Figure 5.5. Negative stain EM of YicH. A – Representative micrograph of YicH at 26 µg/mL stained with 2% (w/v) uranyl acetate and the manual picks of particles. B – 2D class averages of the manual picks of 1335 particles from 12 micrographs into 20 classes. C – *De novo* 3D initial model generation into three classes of a sub-selection of particles was performed using particles from the top 2D classes. One class resembled expected particle dimensions at a high contour level, with the AlphaFold model of YicH (turquoise cartoon) fitting well to the volume (grey). Processing performed in Relion v3.0 and visualisation performed in UCSF ChimeraX (Scheres, 2012; Pettersen et al., 2021).

Regardless, processing of the dataset showed that 2D classes were able to be obtained from 12 micrographs, with 1335 particles manually picked (Figure 5.5A, green circles), extracted, aligned, and averaged (Figure 5.5B). The most populous classes were of the expected dimensions for YicH, but some classes show particles that were longer than expected and some classes show that the particles are too close to each other, resulting in multiple particles extracted in the same box. The longer particles seem to be two YicH particles sticking end-to-

end, potentially embedding into the same DDM micelle. This could be acting like a hinge, allowing the two particles to protrude at different angles, and giving rise to the flexibility seen in the 2D classes. There were also no classes representing an end-on view of the particle, and this was probably due to the manual picking introducing bias into the data as stick-like particles were more easily identifiable. An end-on view would show density as a very small circle of diameter ~40 Å and would be harder to spot by eye. Another reason could be a preferential side-on orientation of particles, and this could be due to an effect of the negative stain salt or due to an attraction of a particular side of the protein to the grid.

The most populous 2D class was the density that was of an expected dimension, and so these particles were used to generate 3 classes of *de novo* 3D initial models. Two of the models were either junk or noise, with one model showing very low-resolution density for a stick-like structure (Figure 5.5C). Overlaying the YicH AlphaFold model showed good agreement, although with no obvious features, including the DDM micelle, the fitting of the model was ambiguous. However, as an initial check of sample quality for further structural investigation, the negative stain data showed that YicH was amenable to further cryo-EM studies, with the knowledge that some issues may need to be addressed including particles sticking together end-on-end and potential preferred orientation.

5.2.1.2 Cryo-EM of YicH

After promising negative stain EM results, YicH was then frozen on grids for cryo-EM screening. Initially, many grids were made where no protein could be seen in the holes. Attempts to increase protein on the grid were made but this wasn't sufficient to observe enough particles in the holes, with the protein seemingly sticking to the hole edge and the support (Figure 5.6A). Graphene oxide-coated grids were then trialled as lower concentrations could be

used and proteins may adhere to the graphene oxide layer, improving particle distribution, as has been reported for other proteins that adsorb to the carbon support (Cooper, 2022). On graphene oxide grids, the protein could be seen as stick-like particles of varying lengths and curvatures, showing similar behaviour to the YicH particles on negative stain EM grids (Figure 5.6B and C). However, at concentrations of 0.1 mg/mL and 0.05 mg/mL, there were too many particles and visible aggregation in the holes for further data collection to be performed.

A lower concentration of 0.01 mg/mL YicH in DDM revealed a better distribution of particles, and so a small dataset of 207 micrographs was collected (Figure 5.7). This was the best grid screened but there were still some issues with particle overcrowding and an indication of preferred orientation along the long axis, as was seen during negative stain EM. Data was collected using the FEI EPU automated data collection software at the Midlands Regional Cryo-EM facility, with collection parameters listed in Table 5.1.

Initial processing was performed using cryoSPARC v4.4.1, with particles picked using the blob picker (Punjani et al., 2017). Unfortunately, the overcrowding of particles resulted in poor particle picking with this method. Two rounds of 2D classification of these particles led to ambiguous and fuzzy 2D classes that could not be taken forward for further processing (Figure 5.7). The 2D classes contained multiple particles in various orientations with respect to each other. In some classes, the particles lined up along their long axis, in some, they crossed over each other, and in others, they associated with each other end-to-end – as was seen in the negative stain EM 2D classes.

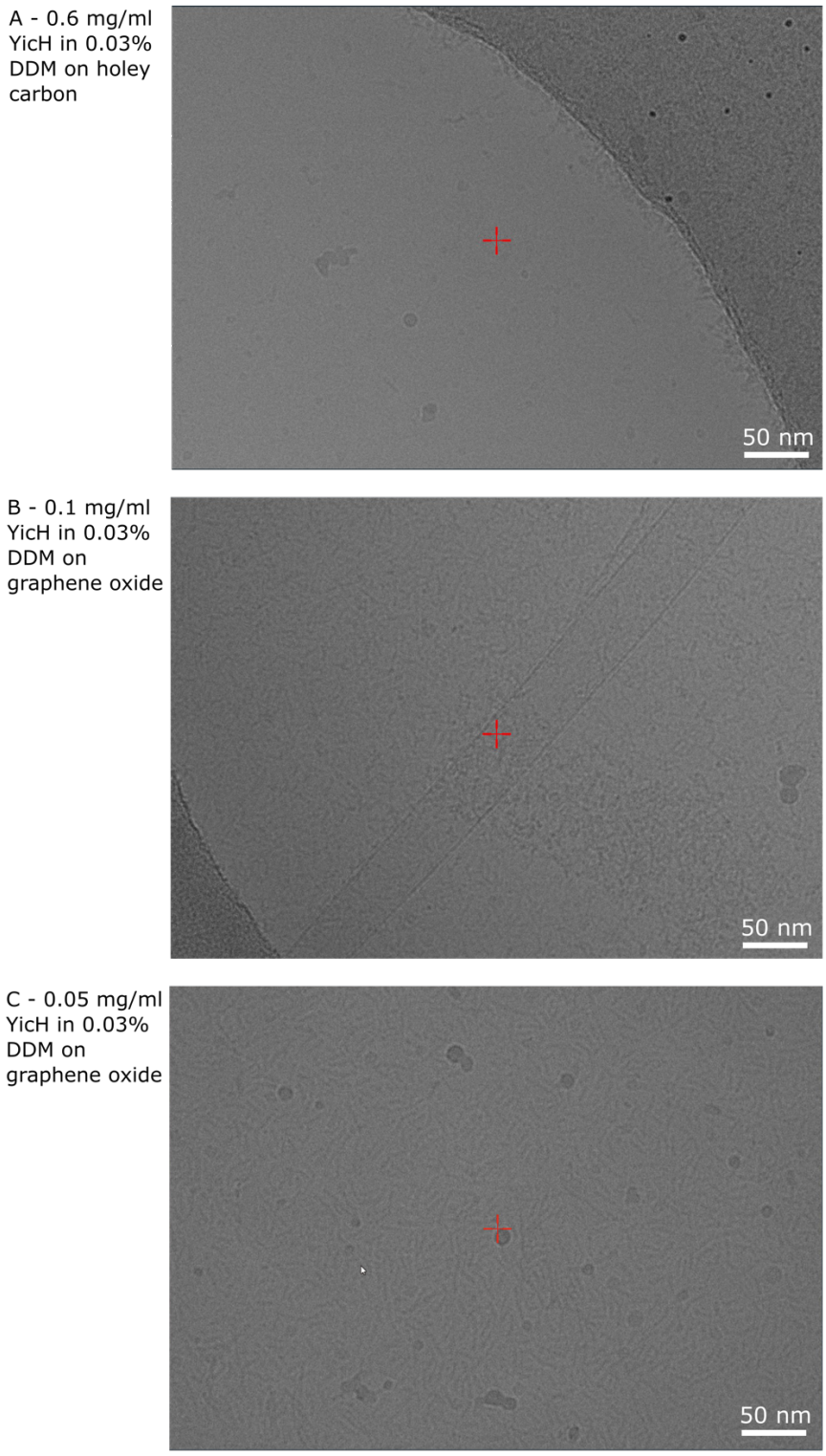


Figure 5.6. Cryo-EM grid screening of YicH. A – 0.6 mg/mL YicH in 0.03% (w/v) DDM on Quantifoil® grids. B – 0.1 mg/mL YicH in 0.03% (w/v) DDM on graphene oxide grids. C – 0.05 mg/mL YicH in 0.03% (w/v) DDM on graphene oxide grids.

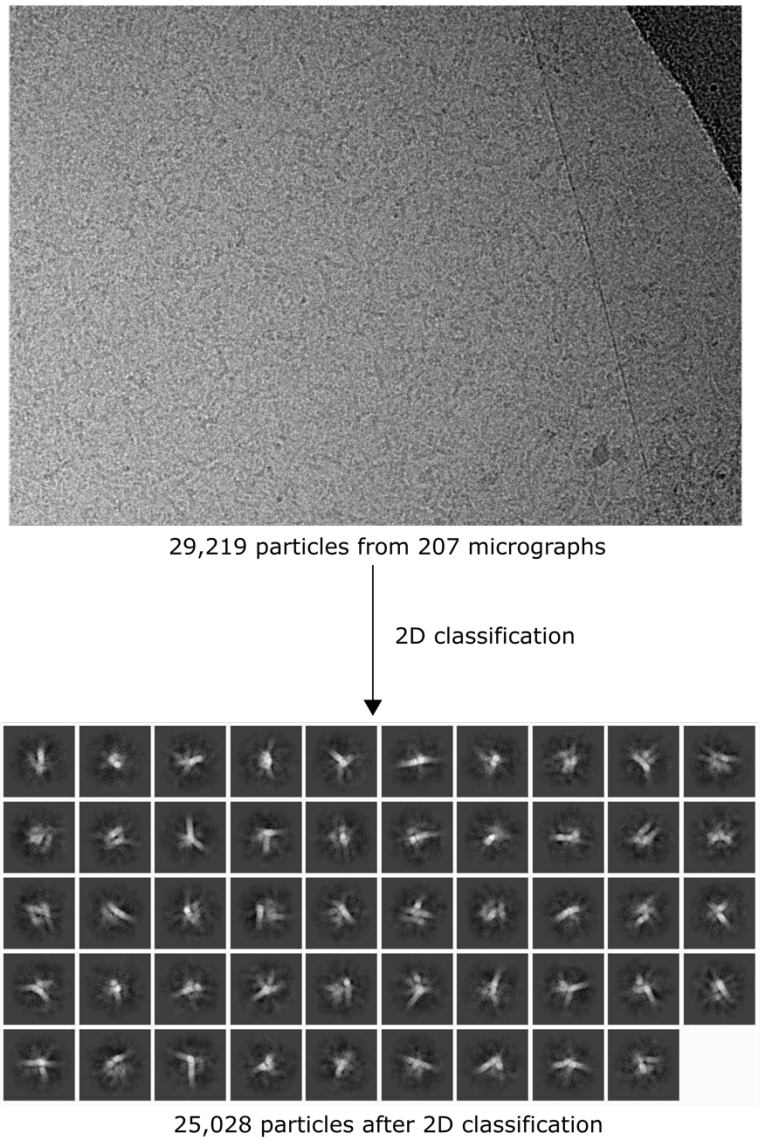


Figure 5.7. Cryo-EM workflow of YicH using blob picker. 0.01 mg/mL YicH in DDM was frozen on graphene oxide-coated grids and a small dataset of 207 micrographs were collected. Particles were picked using cryoSPARC's blob picker tool and 2D classification was performed. The resulting 2D classes were ambiguous, with multiple particles in various orientations per class. Processing was performed in cryoSPARC v4.4.1 (Punjani et al., 2017).

Table 5.1. YicH in DDM on graphene oxide cryo-EM dataset collection parameters.

Parameter	Value
Detector	K3
Detector mode	Super resolution counting bin 2
Voltage (kV)	300
Pixel size (Å)	0.835
Defocus (μm)	-2.7 to -1.2 in 0.3 μm intervals
Dose rate on specimen (e^-/px/s)	18
No. of frames	50
Exposure time (s)	2
No. of micrographs	207

An alternative picking procedure was implemented that used Topaz, a particle-picking program that trains a convolutional neural network using positive-unlabelled learning, with the processing shown in Figure 5.8 (Bepler et al., 2019). When the default ResNet16 Topaz model was used on this dataset, 130,419 particles were picked. After multiple rounds of 2D classification to remove junk particles and those showing more than one particle in the class, *ab initio* model generation was used to create 4 classes. The most populous of this class (33% of particles) resembled a tube-like structure and was used for further refinement. CryoSPARC's non-uniform refinement was performed and revealed a 9.31 Å map.

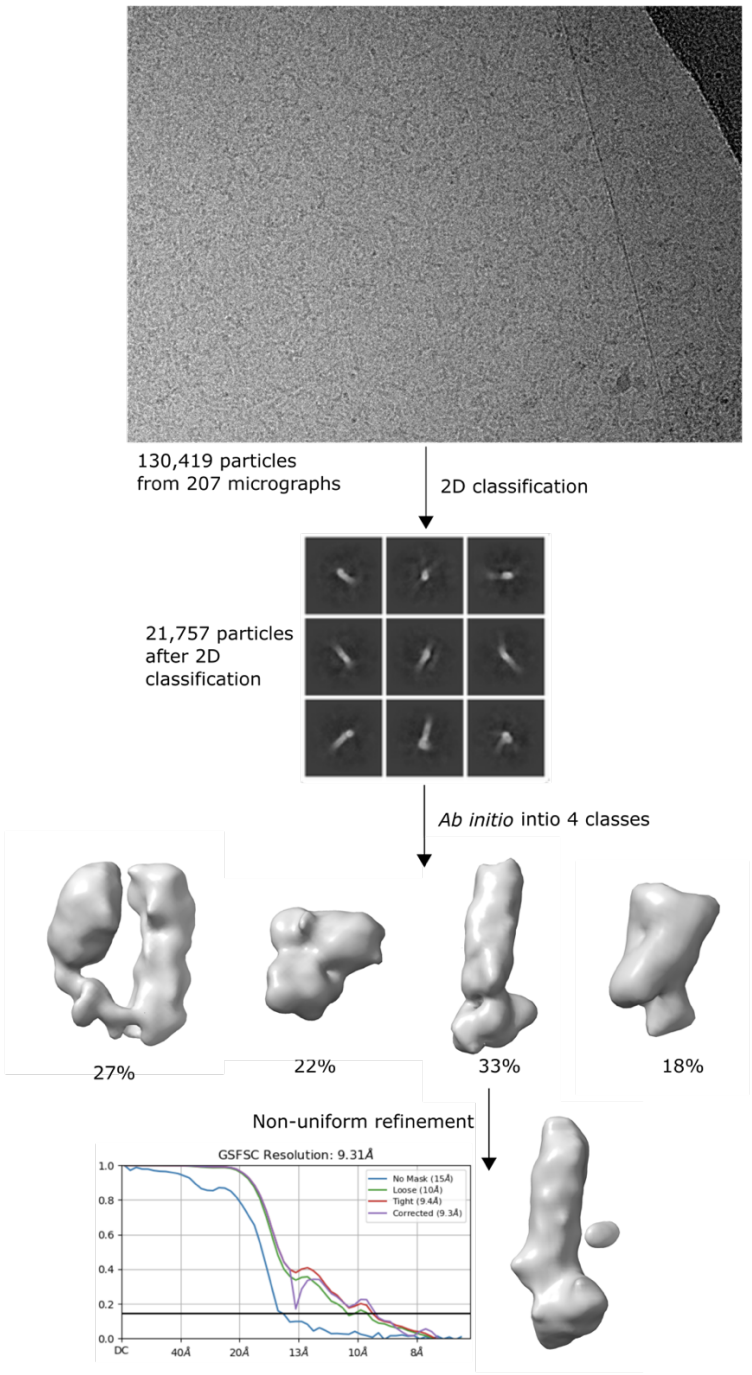


Figure 5.8. Cryo-EM data processing pipeline for the default Topaz model picks of YicH in DDM. The Topaz ResNet16 model was used to pick 130,419 particles from 207 micrographs. 2D classification to remove junk particles and poor 2D classes yielded 21,757 particles for *ab initio* model generation. The most populous 3D class was then subject to non-uniform refinement to yield a 9.31 Å map, with signs of overfitting and unexplained density. The refined map is shown along with the gold-standard Fourier shell correlation (GCFSC) curve, with resolution estimated at the gold standard cut-off of 0.143 (black line). Processing was performed in cryoSPARC v4.4.1 (Punjani et al., 2017).

Unfortunately, inspection of the map revealed some density separate from the main tube-like structure. This was likely due to overfitting of noise or the occasional presence of another particle within the extracted box. There was also a defined region towards the bottom of the tube-like region. This could be the detergent micelle or partial density for a second particle that is within the extracted box. The gold-standard Fourier shell correlation (GSFSC) curve is shown for this refinement, with the gold-standard cut-off of 0.143 used to determine global map resolution. The separation of the curves for the tight mask values (red) and the corrected values (purple) indicated overfitting of the data. The extra "bump" in the curves (here, at ~12 Å) has been previously observed for flexible, disordered or partially occupied membrane proteins, and those with particle box overlaps (Wang et al., 2024b).

Nevertheless, local resolution estimates were generated for the final ~9 Å map and are shown in Figure 5.9A. The area of highest resolution was the noise artefact, and the rest of the map showed poor resolution of above 13 Å. Interestingly, the predicted hydrophobic cavity in the AlphaFold model was visible at this resolution. When the YicH AlphaFold model was then overlayed with the map, a general agreement could be seen (Figure 5.9B). The N-terminal helix did not fit into the density with this fitting, which questions the conclusion that the lower region of density arises from a detergent micelle or a second particle in the extraction box. It could be that at this low a resolution, the fitting of the model is poor and that in reality the N-terminal helix is within this density. However, in this scenario, the C-terminal region would then extend out from the top of the density. It could be that the map is indicative of a very flexible N-terminal helix and lower third of the structure, and that is why the density for this region is poor. It is also possible that the map has been overfitted and is missing density due to this. It is likely a combination of these factors.

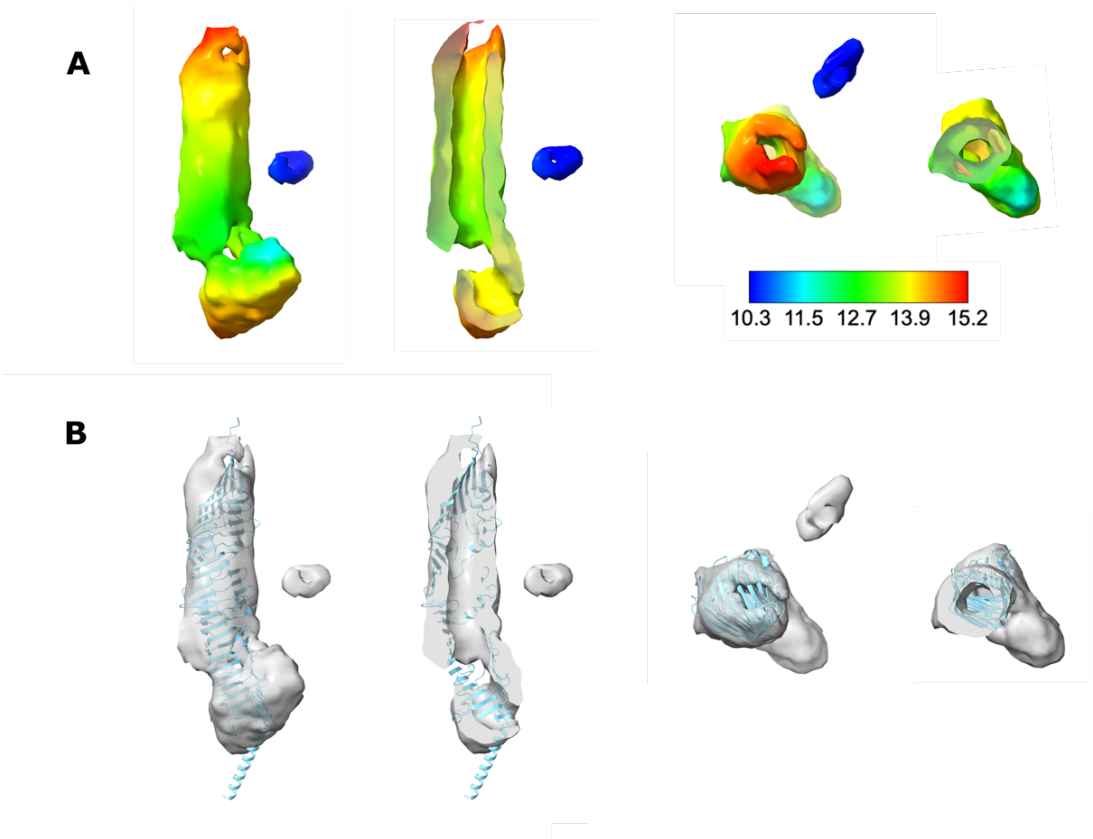


Figure 5.9. Final cryo-EM map of YicH in DDM. A – Local resolution estimates for the map at contour level 0.896, with resolution key shown. B – Transparent map overlayed with the AlphaFold YicH structure. Each structure is depicted as a side view, a side view sliced down the middle of the map, a top view, and a top view sliced across the middle of the map.

With the poor-resolution map and questions about how the density could be explained by the predicted structure, a higher-resolution structure was required. Whilst a larger dataset would likely yield improved results, it was decided that even then, a high-resolution structure would be unlikely as the conditions for grid preparation still needed optimising. Therefore, various other grids were set up to address the issues of overcrowded particles and the potential preferential orientation, including spiking with the detergent CTAB to alter the preferred orientation of YicH and to potentially interfere with the particles sticking together, higher concentrations of YicH in DDM on Quantifoil® grids, and the preparation of YicH in the amphipol A8-35 to change particle behaviour on the grid (Figure 5.10). CTAB is a detergent

used in commercial grid optimisation screening kits (ThermoFisher VitroEase) and the use of amphipol A8-35 saw improved particle distribution for YdbH, another AsmA-like protein, on negative stain EM grids during the detergent and polymer screen in Chapter 4.

Unfortunately, many of these grids either showed no protein in the holes again or had too much crystalline ice to be collected on. The preparation of YicH in the amphipol A8-35 was more promising, with visible stick-like particles and a better distribution (Figure 5.10C). The grid presented here was too icy to be collected on, and ideally, a lower concentration would be trialled to improve the particle distribution further. Interestingly, a concentration of 1.2 mg/mL YicH in A8-35 showed slight particle overcrowding in the holes, whereas a similar concentration of 1.5 mg/mL YicH in DDM had very few particles (Figure 5.10B and C). This reveals the significant effect on particle distribution and behaviour on grids that can be caused by changing the membrane protein encapsulation agent used.

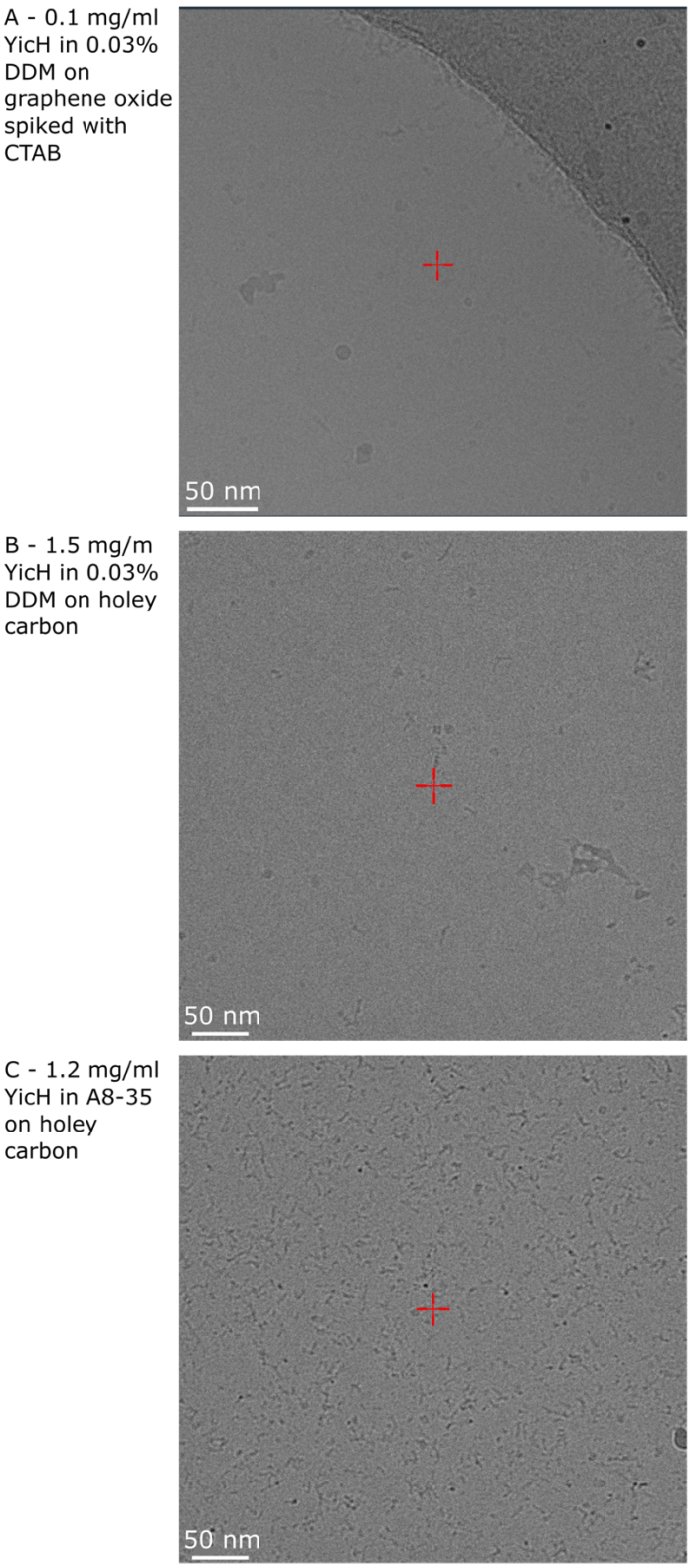


Figure 5.10. Further cryo-EM grid screening of YicH. A – 0.1 mg/mL YicH in 0.03% (w/v) DDM on graphene oxide spiked with CTAB. B – 1.5 mg/mL YicH in 0.03% (w/v) DDM on Quantifoil®. C – 1.2 mg/mL YicH in A8-35 on Quantifoil®.

5.2.2 Structural characterisation of YhjG

5.2.2.1 Negative stain of YhjG

The same approach taken for YicH was applied to YhjG, a slightly larger AsmA-like protein (YicH MW = 62 kDa, YhjG MW = 74 kDa) that could be expressed and purified reproducibly and to a relatively high purity. Like YicH, genetic investigations of *yhjG* did not reveal a clear function within the cell, and so obtaining a structure for functional insight was desirable.

Various concentrations of YhjG in DDM were applied to carbon grids and coated with uranyl acetate. Upon screening, it was found that overcrowding was an issue, with many stick-like particles overlapping. The optimal concentration was found to be 4 μg/mL, and Figure 5.11A shows a representative micrograph of this grid that shows how densely packed the particles still were at this low concentration. The particles appeared very similar to YicH, which was expected as there is only ~20 Å difference in length between the two structurally homologous AlphaFold models. Similar to the behaviour of YicH on grids, variations in the length of particles were apparent, with some particles showing the same flexibility with what looked like two particles end-to-end. There also appeared to be a preferred orientation of the long axis, although again, this could have been due to the top-end view being too small and ambiguous to identify by eye.

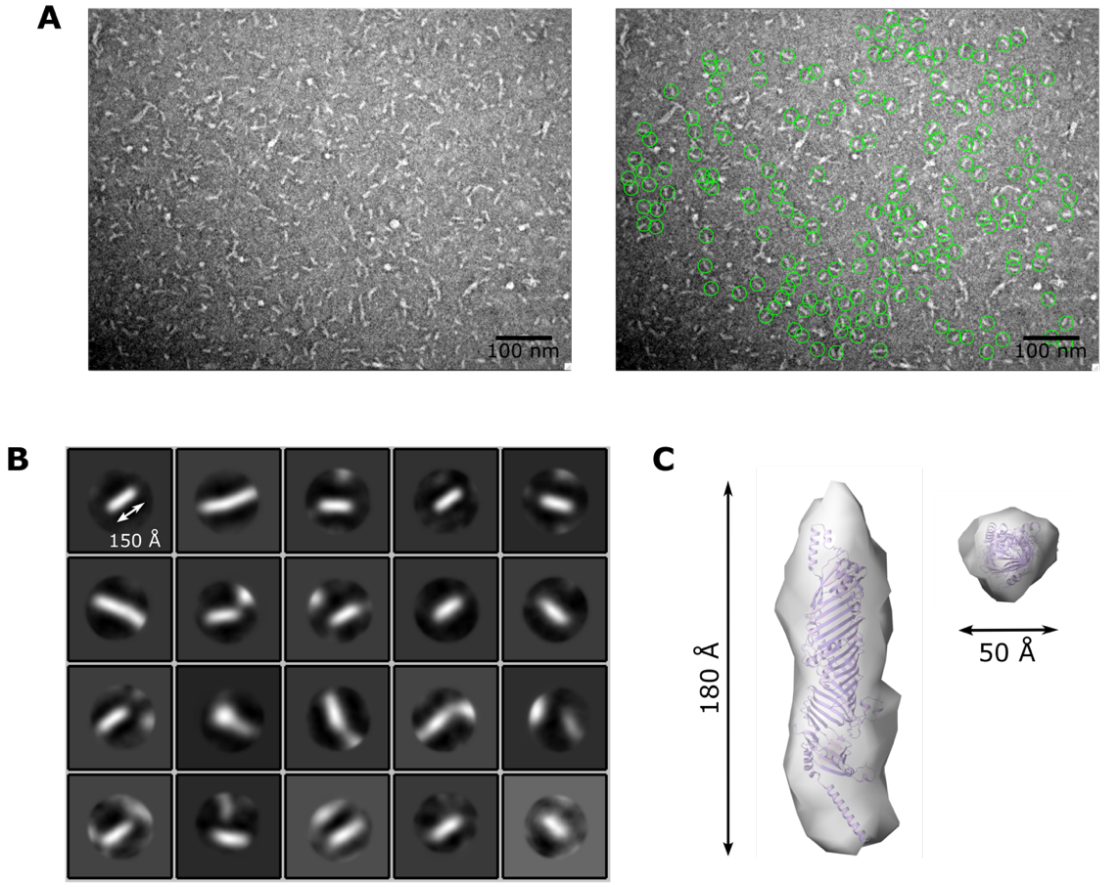


Figure 5.11. Negative stain EM of YhjG. A – Micrograph of 4 µg/mL YhjG in DDM stained with 2% (w/v) uranyl acetate and the manual picks of particles. B – 2D class averages after classification of the manual picks. C – One of the *de novo* 3D volumes generated from top three 2D classes, with the AlphaFold model of YhjG (purple cartoon) overlayed onto the volume at a high contour level (grey). Processing performed in Relion v3.0 and visualisation performed in ChimeraX-1.7.1 (Scheres, 2012; Pettersen et al., 2021).

Manual picking was performed on the 23 micrographs, with 4057 particles then extracted to create the 2D classes shown in Figure 5.11B. The more overcrowded grid resulted in multiple particles in some picks, which was evident in some of the 2D classes. However, the most populous class indicated the presence of a protein that was of the expected dimensions when compared to the YhjG AlphaFold model of 170 Å long and 40 Å wide. Particles from three of the classes were then used to generate three *de novo* 3D models. One of which had density for

a particle with dimensions representative of the YhjG AlphaFold model at very low resolution (Figure 5.11C).

5.2.2.2 Cryo-EM of YhjG

After promising negative stain EM results of YhjG, with particles that represented the predicted dimensions of the AlphaFold model, cryo-EM grids were set up. Finding the optimal preparation was time-consuming, with many grids screened at different concentrations, in different detergents and on different supports. Additionally, some grids had too much crystalline ice to be screened. Various issues during screening that reoccurred included too little protein in the holes due to particles sticking to the hole edge and carbon support shown in Figure 5.12A, visible protein aggregation and a high detergent background shown in Figure 5.12B, and particle overcrowding in Figure 5.12C. The following sections will cover the two purification conditions that led to the most homogenous, evenly distributed grids out of those screened: YhjG purified in the detergent LDAO and YhjG purified in the amphipol A8-35.

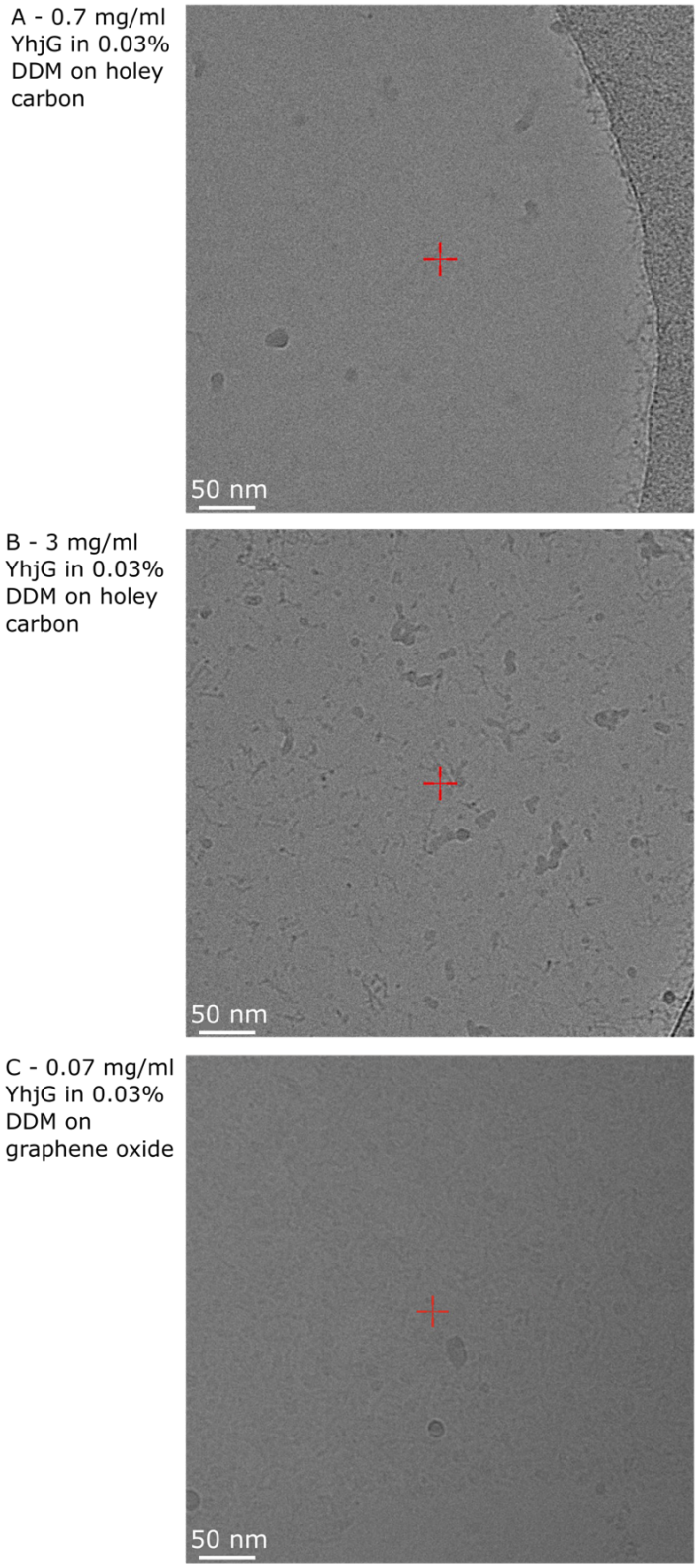


Figure 5.12. Cryo-EM grid screening of YhjG. A – 0.7 mg/mL YhjG in 0.03% (w/v) DDM on Quantifoil®. B – 3 mg/mL YhjG in 0.03% (w/v) DDM on Quantifoil®. C – 0.07 mg/mL YhjG in 0.03% (w/v) DDM on graphene oxide.

YhjG in LDAO

With unsuccessful grids of YhjG in DDM, different detergents were tested to see if particle monodispersity, distribution and orientations improved. It was reported that the truncated construct of TamB, the crystal structure of which first revealed the β-taco fold, required the detergent LDAO in its buffers for solubility and crystal formation (Josts et al., 2017). Due to similar predicted structures of TamB and YhjG, with the conserved β-taco fold present in the periplasmic domains, LDAO was used here to try and replicate this enforced stability. DDM-solubilised YhjG was detergent exchanged into LDAO via gel filtration, resulting in a single peak eluting at 12 mL with a slight shoulder (Figure 5.13A). Like the gel filtration of YhjG in DDM, the peak elutes earlier than expected for a monomeric size (Chapter 4). Again, this was assumed to likely be a combination of the detergent micelle adding molecular weight and the rod-shaped structure of the protein resulting in a large hydrodynamic radius with a shorter retention time. When peak fractions were analysed by SDS-PAGE, a single band at 75 kDa, the expected weight of YhjG, was apparent (Figure 5.13B).

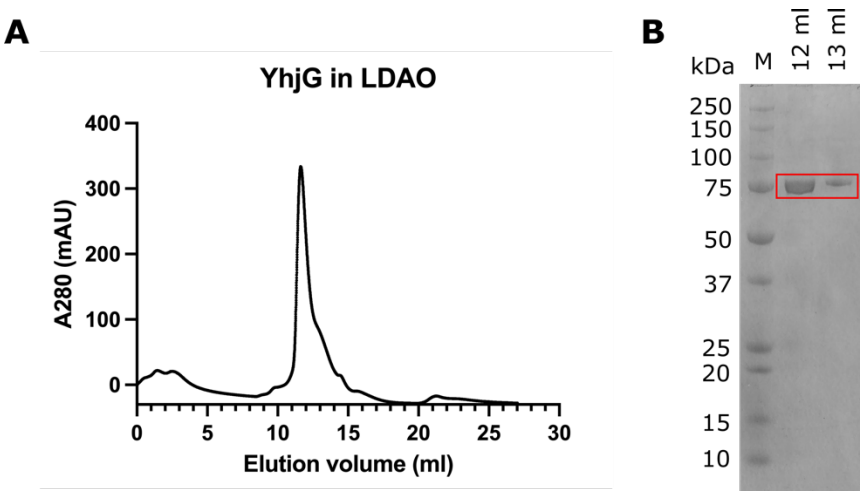


Figure 5.13. Purification of YhjG in the detergent LDAO. A – Detergent exchange gel filtration of YhjG from DDM to LDAO on a S200 10 300 increase column. B – SDS-PAGE analysis of peak fractions after SEC. M – molecular weight marker.

The best grid screened was of YhjG in LDAO frozen at 0.27 mg/mL on Quantifoil Au R1.2/1.3 300 mesh grids. A dataset of 2771 micrographs was collected using the FEI EPU automated data collection software using the parameters shown in Table 5.2 at the Midlands Regional Cryo-EM facility. Data processing was performed in cryoSPARC v4.4.1 (Punjani et al., 2017). Patch-based motion correction followed by patch-based CTF estimation was performed on all micrographs using the cryoSPARC implementation. Micrographs were then curated to bin any ice-contaminated micrographs and those with poor CTF estimation, leaving 2636 accepted micrographs for further processing.

Table 5.2. YhjG in LDAO cryo-EM dataset collection parameters.

Parameter	Value
Detector	K3
Detector mode	Super resolution counting bin 2
Voltage (kV)	300
Pixel size (Å)	0.835
Defocus (μm)	-2.7 to -1.2 in 0.3 μm intervals
Dose rate on specimen (e^-/px/s)	15.786
No. of frames	50
Exposure time (s)	2.2
No. of micrographs	2771

Particle picking was a time-consuming process as it required testing of various picking strategies. A high-quality particle set, with many different orientations of the protein, is necessary to yield a high-resolution map, meaning the accurate picking of a high number of particles is crucial. An overview of all the particle picking strategies employed for this dataset is shown in Figure 5.14.

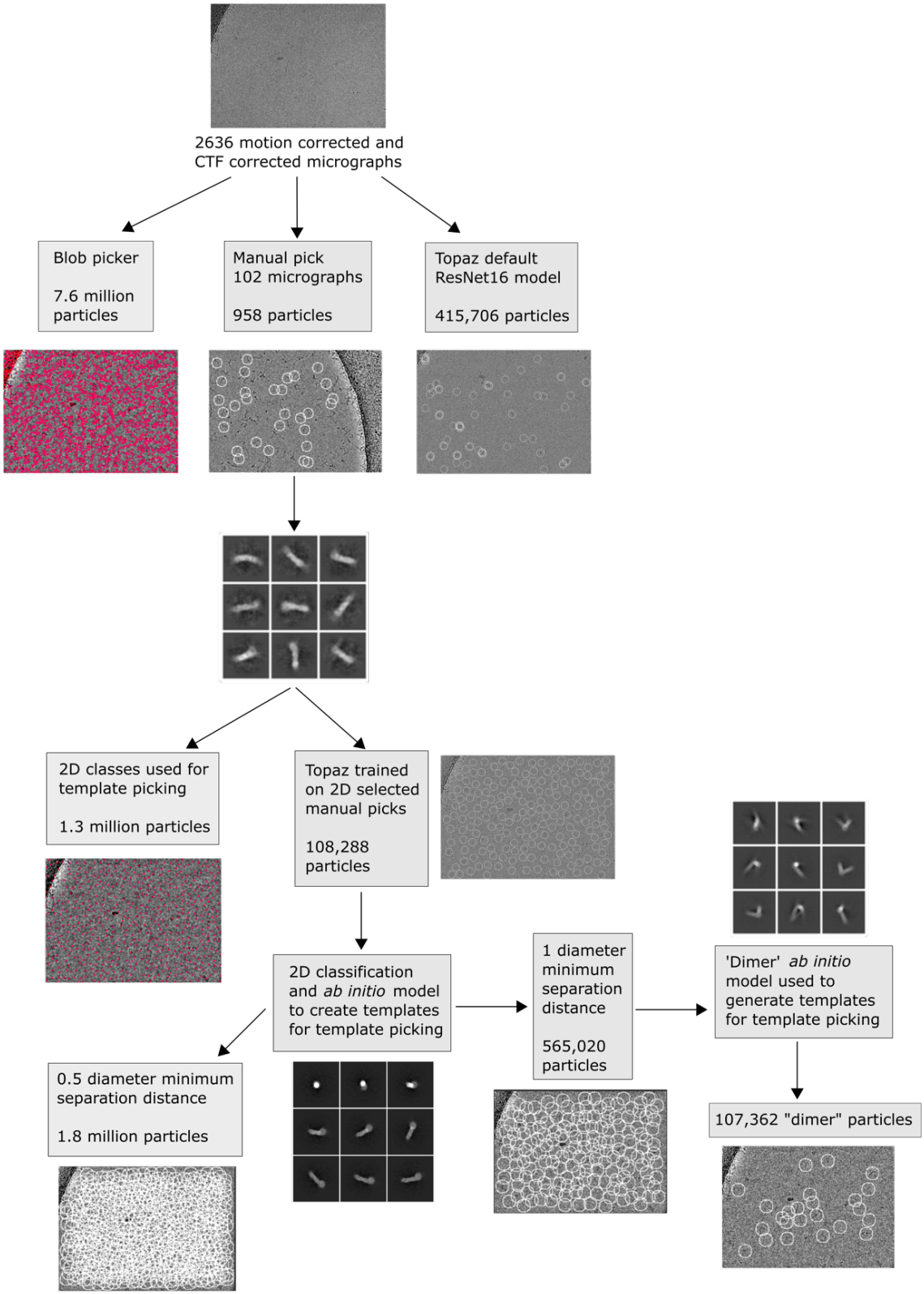


Figure 5.14. Overview of the various particle picking pipelines used for the YhjG in LDAO cryo-EM dataset. Micrographs and 2D templates used for picking are shown, along with the number of particles picked using each strategy.

Initially, particle picking was performed using cryoSPARC's blob picker tool, however, it was immediately apparent that a more stringent picking tool was needed, as ~7.6 million particles were picked using this method (Punjani et al., 2017). Inspection of particle picks did reduce this number significantly due to many false-positive picks, but the resulting particles did not yield good enough 2D classes for further processing.

The next approach to particle picking was to assess a small representative portion of the dataset and manually pick particles for template picking of the whole dataset. 102 micrographs were selected due to having good CTF estimations and a good distribution of particles across the ice, resulting in 958 manually picked particles. 2D classification of this small particle set yielded promising results, with classes representative of the dimensions of the YhjG AlphaFold model (Figure 5.14). The highest quality templates were then used for template picking of the full 2636 micrograph dataset; however, inspection of the picking results revealed a high number of false positive picks and some particles being missed.

Instead, a Topaz model was trained on the manually picked particles and used to pick from the whole dataset, as Topaz reportedly leads to fewer false-positive picks and the improved identification of smaller and irregular-shaped particles than other automated particle picking programs (Bepler et al., 2019). This resulted in improved particle picking, with 2D classification of the particle set yielding various orientations of a single YhjG particle (Figure 5.15). There were also some 2D classes showing density for more than one monomer in a single class, which will be discussed later in this section. 2D classes representing various orientations of a YhjG monomer were used to generate four *ab initio* models. Most particles (38%) belonged to a 3D class that agreed with the dimensions of the YhjG AlphaFold model of 170 Å long, with large "micelle-like" density at one end at 50 Å wide, and narrow "tube-like" density at 30 Å wide.

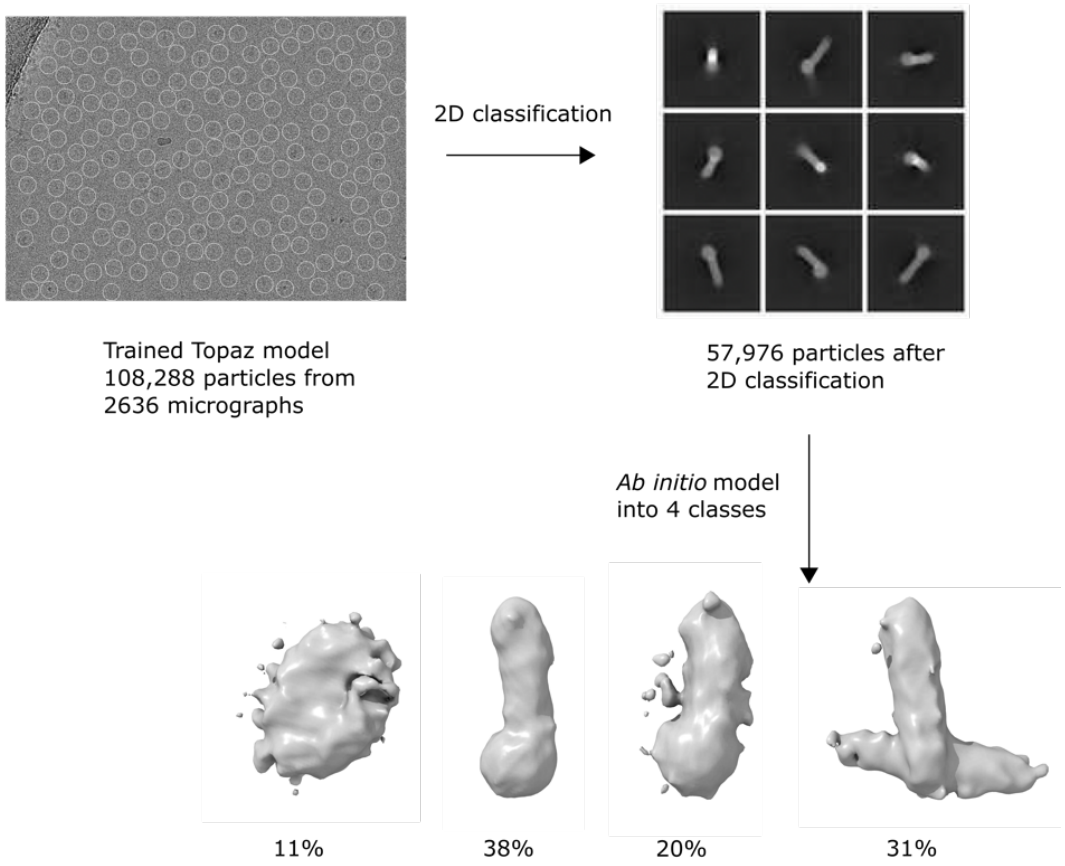


Figure 5.15. Cryo-EM pipeline of training a Topaz model on manual picks of YhjG in LDAO. Manual particle picking of 102 micrographs was used to train a Topaz model for particle picking of the full 2636 micrograph dataset. The resulting 108,288 particles were subject to multiple rounds of 2D classification to remove junk particles. The best 2D classes were used to generate 4 *ab initio* models, with a model resembling the anticipated dimensions of YhjG used to generate templates for further template picking of the dataset. Processing was performed in cryoSPARC v4.4.1 (Punjani et al., 2017).

With only ~22,000 particles in this class, it was anticipated that more particles were needed to generate a higher-resolution structure. Therefore, template picking was performed utilising templates generated from the *ab initio* model. This is a common strategy that sees an improvement in the particle set due to the elimination of noise and the generation of evenly sampled projection directions (Baldwin and Penczek, 2007; Punjani et al., 2017). When the default minimum separation distance between picks of 0.5 * picking diameter (= 110 Å) was imposed, 1.8 million particles were picked, whereas there were 565,020 particles picked when a 1 * picking diameter (= 220 Å) separation distance was imposed. Initial processing of particles

from the 0.5 * picking diameter selection are shown in Figure 5.16. Poor 2D classification results and poor *ab initio* class generation were generated, and so further processing was not performed.

Initial processing of particles from the 1 * picking diameter selection was more encouraging, and so these particles were taken forward for further processing (Figure 5.17). Inspection of the particles led to 389,301 particles, which were then subject to heterogeneous refinement using the 4 previously generated *ab initio* maps low pass filtered to 20 Å as input models. Non-uniform refinement of the best class resulted in an 8.95 Å map. Inspection of the map revealed tube-like density but also exhibited density artefacts along the sides that were likely due to overfitting. The GSFSC curve also indicated overfitting, with poor alignment between the corrected and tight mask values from ~9 Å. The curve indicated that the resolution may be hitting the Nyquist limit, but with a low-quality map containing density artefacts and a poor global resolution, improved particle picking and processing procedures were still required.

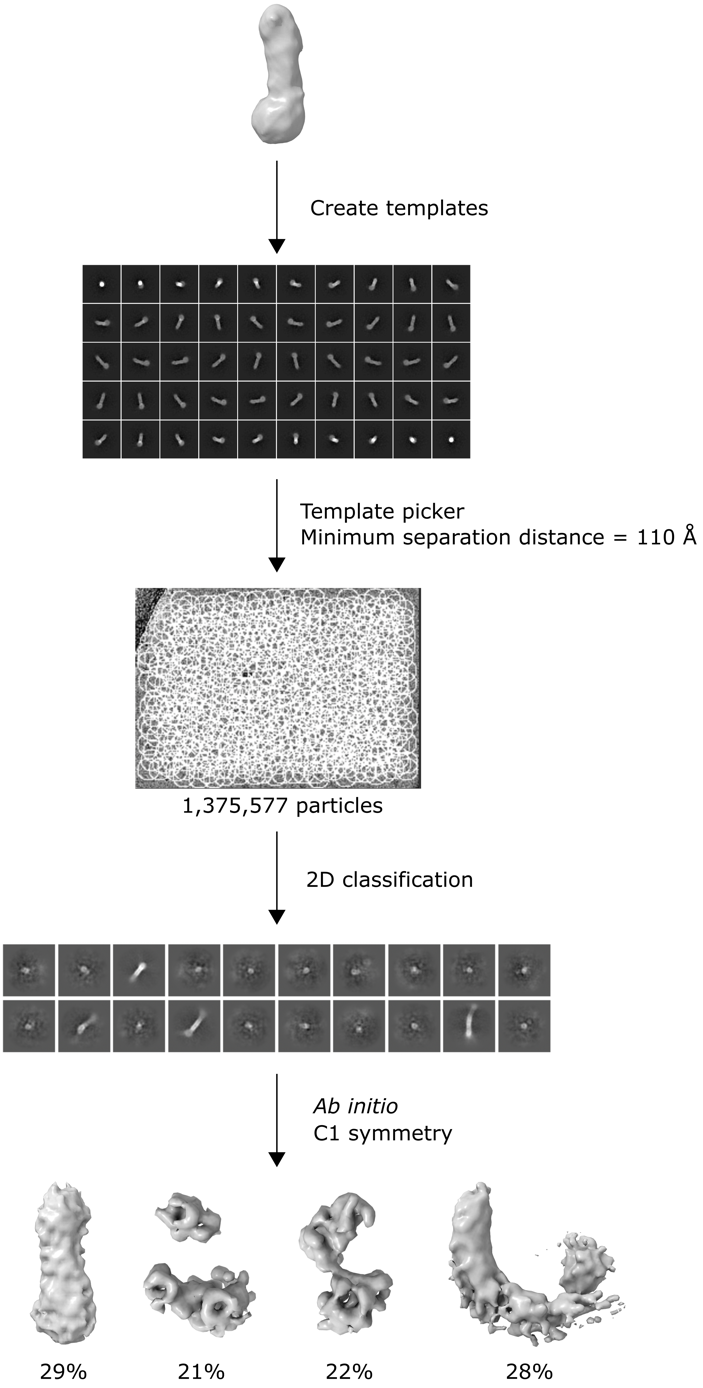


Figure 5.16. Cryo-EM data processing pipeline for the template picks with 0.5 diameter separation distance of YhjG in LDAO. Manual picks of a subset of micrographs were used to train a Topaz model that was used to pick from 2636 micrographs, then subject to multiple rounds of 2D classification and *ab initio* model generation. The best model, shown above, was used to create 2D templates to use the template picker, with a minimum separation distance of 0.5 * picking diameter = 110 Å utilised. ~1.4 million particles were extracted and 2D classification was performed. *Ab initio* modelling of the best 2D classes generated the above four models. Processing was performed in cryoSPARC v4.4.1 (Punjani et al., 2017).

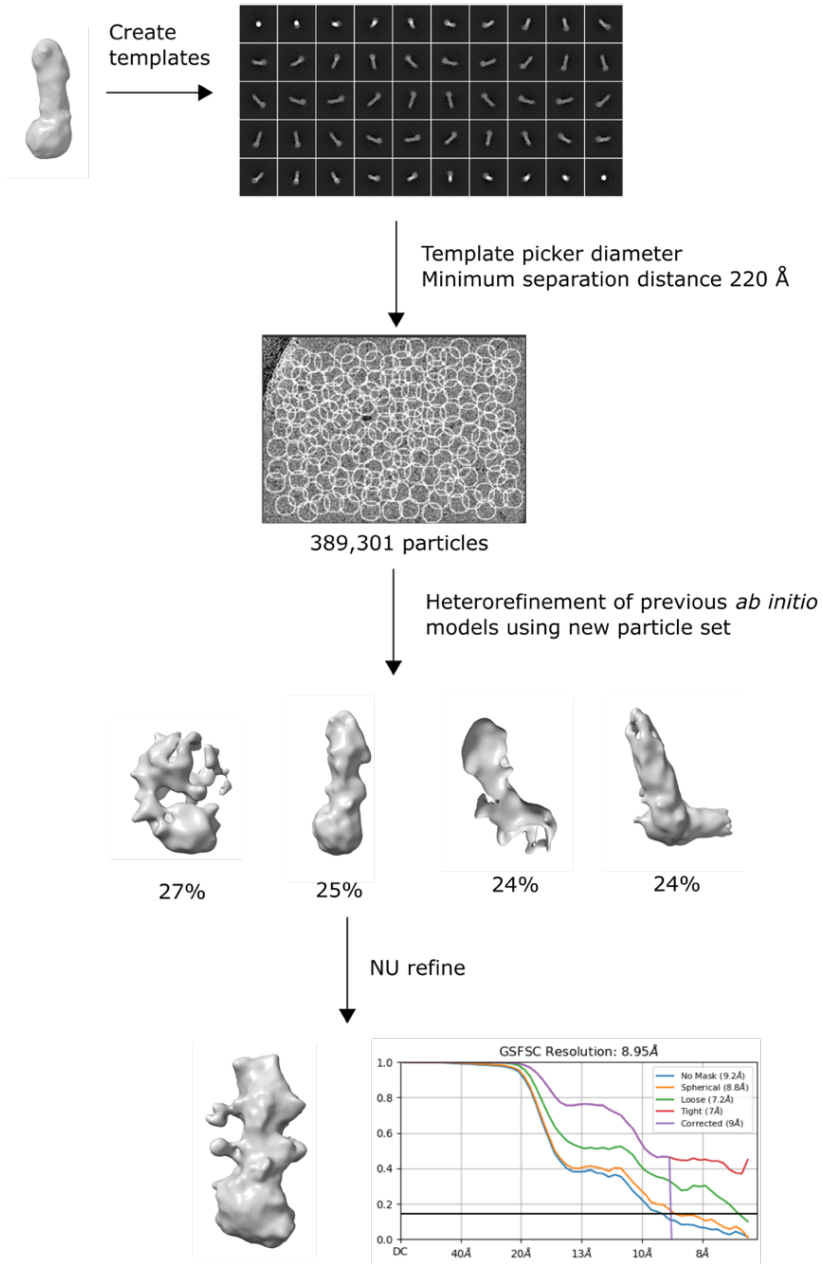


Figure 5.17. Cryo-EM data processing pipeline for the template picks with 1 diameter separation distance of YhjG in LDAO. Manual picks of a subset of micrographs were used to train a Topaz model that was used to pick from 2636 micrographs, then subject to multiple rounds of 2D classification and *ab initio* model generation. The best model, shown above, was used to create 2D templates for template picking, with a minimum separation distance of 1 picking diameter (220 Å) utilised. 389,301 particles were extracted, and heterogeneous refinement was performed using the previous *ab initio* models low pass filtered as input volumes. Non-uniform refinement of the best class was performed with C1 symmetry. The refined map is shown along with the gold-standard Fourier shell correlation (GCFSC) curve, showing the corrected map has a resolution of 8.95 Å at the gold standard cut-off of 0.143 (black line). Processing was performed in cryoSPARC v4.4.1 (Punjani et al., 2017).

The 2D classification of the trained Topaz model picks not only generated classes representative of a single YhjG monomer, but some classes displayed density for multiple YhjG monomers. Many of these displayed “V”-shaped density, where it seemed two monomers were joined at one end, as if hinged together. This class could indicate the presence of a dimer of YhjG, it could be a result of particle overcrowding, or it could perhaps be two monomers embedded into the same detergent micelle. With an inability to obtain high-resolution maps of YhjG from the 2D classes representing the “monomer” particle, it was decided to trial utilising this larger “hinged” particle to pick more robustly. It was possible that this hinged conformation could be more rigid than the “monomer” particle, and so would be more amenable to particle picking, averaging and alignment.

2D classes that represented the hinged conformation were selected and 4 *ab initio* models were generated (Figure 5.18). Templates were then created from the 3D class resembling a hinged conformation and used for template picking. The resulting 107,362 particles were subject to 2D classification and used to generate 4 *ab initio* models, the most populous of which resembled the hinged conformation (42%) and the next most populous class resembling the monomer conformation (26%). Both volumes were subject to non-uniform refinement, resulting in an 8.04 Å map for the hinged particle and a 7.38 Å map for the monomer. The GSFSC curves are also shown. The hinged refined volume revealed good density for one monomer, whilst the other was ambiguous. The density for the single monomer class was also ambiguous, with some overfitting artefacts present, and a volume that was shorter than the anticipated dimensions of YhjG.

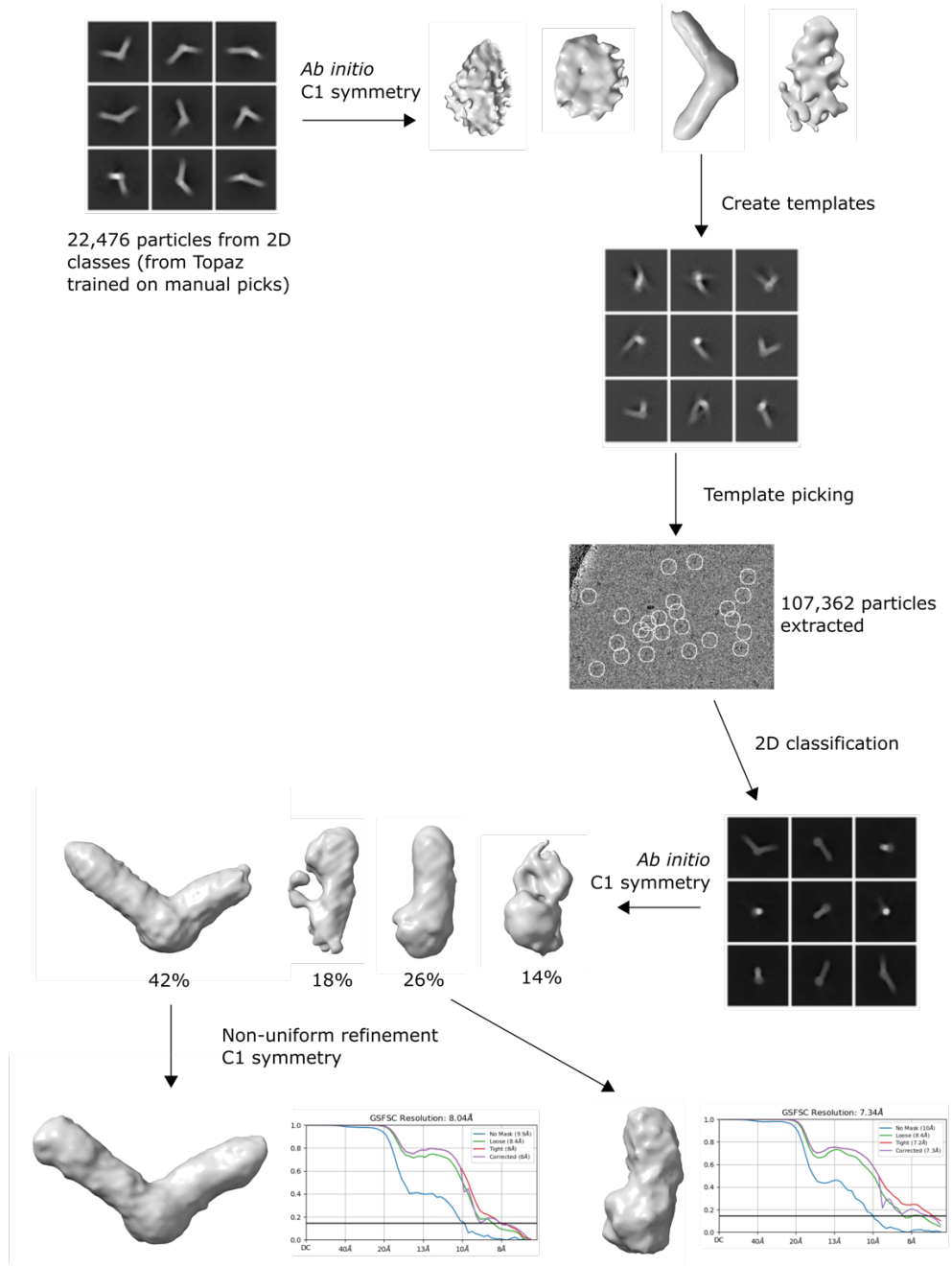


Figure 5.18. Cryo-EM data processing pipeline for the "hinged" class of YhjG in LDAO. The previously trained Topaz model was used to pick particles from 2636 micrographs, and 2D classes with two particles present were selected. The resulting 22,476 particles were used to generate *ab initio* models. One resembling the hinged particle was used to create templates for template picking of the full dataset. The resulting 107,362 particles were subject to 2D classification and *ab initio* modelling with C1 symmetry. The most populous classes were then subject to non-uniform refinements, resulting in an 8.04 Å hinged map and a 7.34 Å monomer map. The maps and gold-standard Fourier shell correlation (GCFSC) curves for each refinement are shown. Resolutions were calculated at the gold standard cut-off of 0.143 (black line on GCFSC plots). Processing was performed in cryoSPARC v4.4.1 (Punjani et al., 2017).

Further refinement of the hinged volume was attempted, including masking strategies with a mask covering the "second", more ambiguous monomer and with a cryoSPARC-implemented local refinement tight mask to see if the density for one of the monomers could be improved upon. Unfortunately, both efforts resulted in maps displaying many artefacts and poor resolution estimates (data not shown).

An improved particle picking approach was found when using the pre-trained Topaz default ResNet16 model to pick particles from the full dataset (Bepler et al., 2019). This resulted in 415,706 particles and promising class averages after 2D classification, with mostly various orientations of a single monomer with some classes showing two monomers present again (Figure 5.19). Further processing was performed using multiple rounds of 2D classification to remove junk particles and the selection of multiple views of the monomer conformation to generate 3 *ab initio* models. The most populous model showed the anticipated features of YhjG, though it exhibited wider density, especially in the "micelle-like" region, than had previously been seen. The hinged class was also present from this particle set.

Both the monomer and hinged volumes were then further refined using cryoSPARC's non-uniform refinement with C1 symmetry to 6.97 Å and 8.88 Å, respectively (Figure 5.19) (Punjani et al., 2017). The GSFSC plots for the maps are also shown. The GSFSC curve for the monomer appeared to show some evidence of overfitting, but also indicated that the map could go to a higher resolution, due to the curves not reaching 0 on the y-axis. The Nyquist limit for this processing is ~7 Å here, due to the original particle size of 0.835 Å and the use of 4x binned particles for processing. Therefore, to improve this resolution, particles were re-extracted with 2x binning, and non-uniform refinement was performed on the new particle set. Unfortunately, the resolution estimate increased to 7.38 Å, possibly due to more noise being introduced, and a higher resolution map was not obtained.

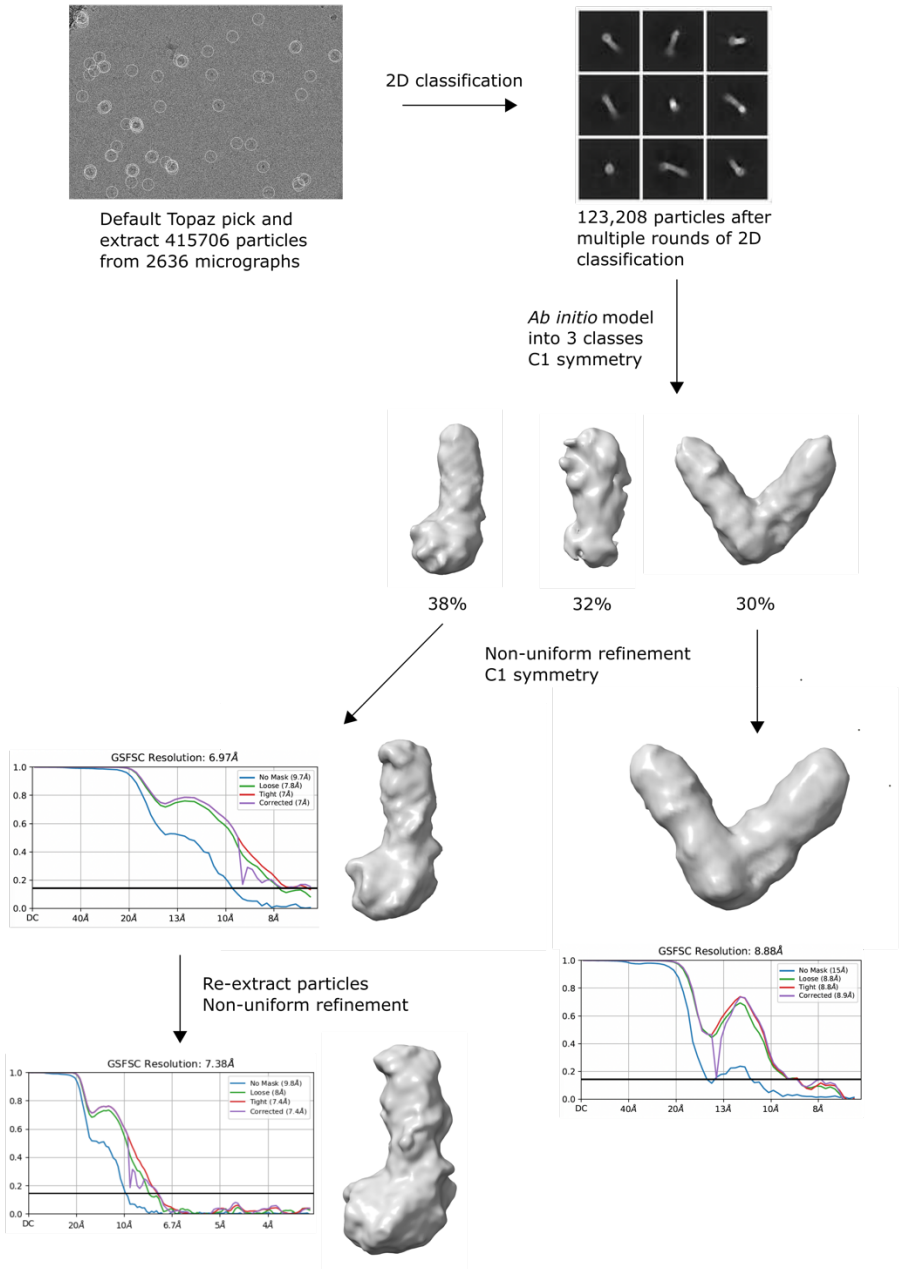


Figure 5.19. Cryo-EM data processing pipeline for the pre-trained Topaz picks of YhjG in LDAO. The pre-trained default Topaz ResNet16 model was used to pick particles from 2636 micrographs. 415,706 particles were extracted with 4x binning and subjected to multiple rounds of 2D classification to get rid of junk particles. *Ab initio* model generation into 3 classes was performed. Non-uniform refinement of the best class was performed with C1 symmetry, yielding a 6.97 Å map. The particles were re-extracted with 2x binning and the non-uniform refinement was performed, now yielding a 7.38 Å map. Non-uniform refinement with C1 symmetry was also performed on the hinged class, revealing an 8.88 Å map. The gold-standard Fourier shell correlation (GCFSC) curves for each refinement along with the maps are shown. Resolutions were calculated at the gold standard cut-off of 0.143 (black line on GCFSC plots). Processing was performed in cryoSPARC v4.4.1 (Punjani et al., 2017).

As an alternative to re-extracting the particles from one 3D class and performing the refinement, particles in the best 2D classes from the pre-trained Topaz model picks were re-extracted with no binning and subject to the same processing (Figure 5.20). 2D classification was performed to check for the presence of any junk particles, and *ab initio* model generation was performed to generate 3 classes. The most populous class was representative of the hinged configuration and yielded a 7.66 Å map after non-uniform refinement. The density for one of the YhjG monomers was poor, with a shorter tube-like structure than the other and wider, more ambiguous density. The other YhjG monomer was better refined, but still no high-resolution information could be seen.

The next most populous class represented the monomer class and yielded a 6.84 Å map after non-uniform refinement, the highest global map resolution achieved out of all the processing strategies utilised. The map showed similar features to the other final maps produced, with the tube-like structure extending from the "micelle-like" density. The GSFSC curves shown also had fewer signs of over-fitting than some of the other processing pipelines that had been previously tested. Interestingly, though the resolution estimate improved, the map was shorter and wider than some of the other final maps achieved.

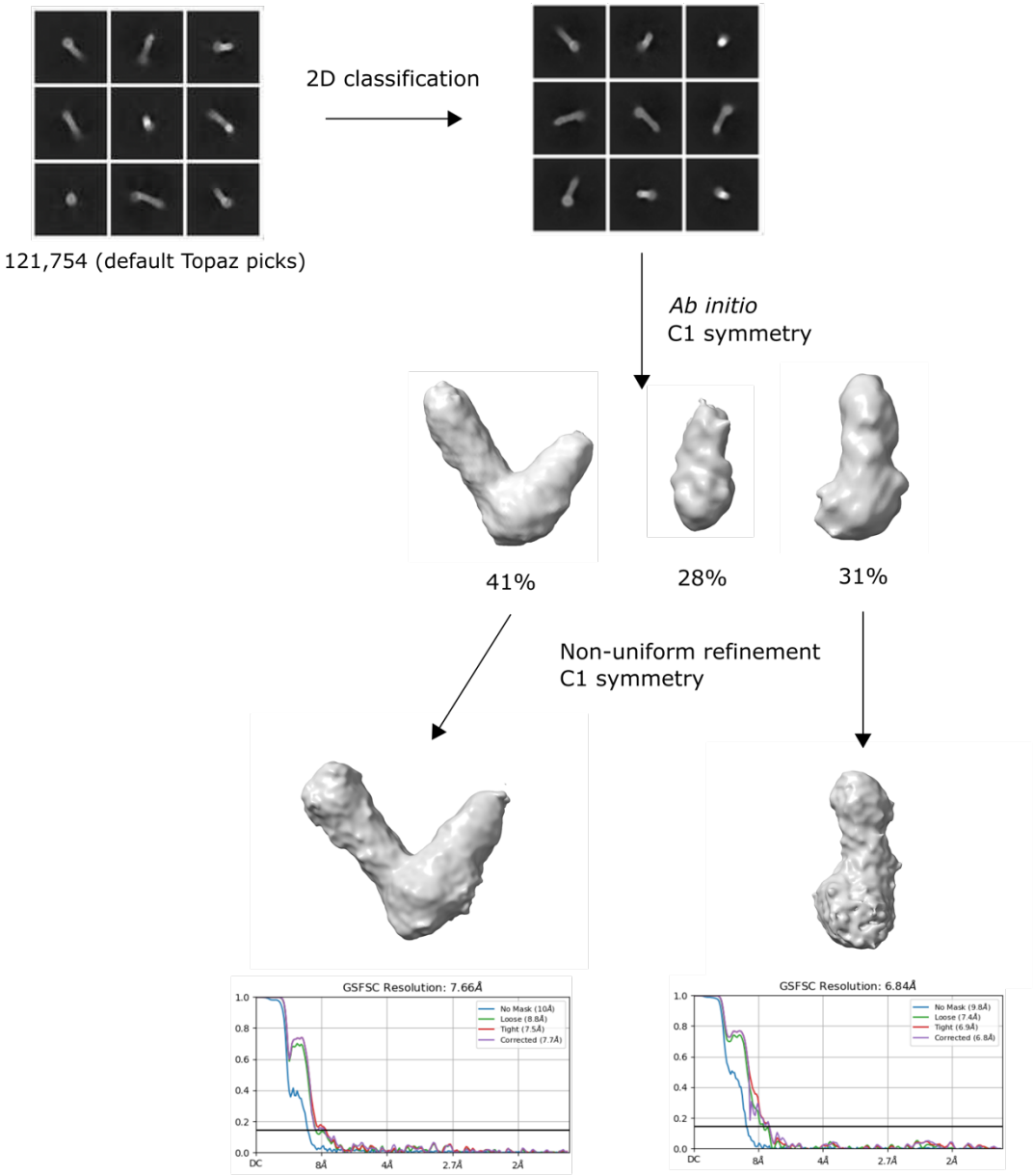


Figure 5.20. Cryo-EM data processing pipeline for the default Topaz picks of YhjG in LDAO with no binning. The default Topaz ResNet16 model was used to pick particles from 2636 micrographs. 415,706 particles were initially extracted with 4x binning and subjected to multiple rounds of 2D classification to get rid of junk particles. Particles from the resulting 2D classes were re-extracted with no binning and subject to further 2D classification and *ab initio* model generation into 3 classes. Non-uniform refinement of the hinged class was performed with C1 symmetry, yielding a 7.66 Å map. Non-uniform refinement with C1 symmetry was also performed on the "monomer" class, revealing a 6.84 Å map. The gold-standard Fourier shell correlation (GCFSC) curves for each refinement along with the maps are shown. Resolutions were calculated at the gold standard cut-off of 0.143 (black line on GCFSC plots). Processing was performed in cryoSPARC v4.4.1 (Punjani et al., 2017).

Despite multiple particle-picking strategies and processing pipelines tested, a high-resolution map was not obtained for YhjG in LDAO. The modest resolution maps that were obtained also showed signs of over-fitting, so the reported resolution of the final maps may not reflect the actual resolution. Obtaining a higher resolution map may have been hindered by the relatively low number of particles contributing to the final EM maps and/or by particle heterogeneity. This may have been caused by inherent flexibility of the protein or by the presence of two particles hinging at a single point, with various angles between the two particles adopted. The background noise of free detergent micelles from the presence of LDAO may also have affected the generation of a higher resolution map.

Nevertheless, local resolution estimations of the highest quality final maps from the various processing strategies were generated, and are shown in Figure 5.21, along with a fit of the AlphaFold structure of YhjG to the density. The local resolution fluctuated across each map, ranging from around 7-15 Å for each final map. The middle region of the tube-like structure consistently showed the highest local resolutions, indicating regions either side of this were more flexible and information was lost during particle averaging. It is also interesting to note that the hinged conformations, whilst having poor density for the "second" monomer, showed more consistent local resolution estimates across a greater portion of the tube-like structure of one of the monomers compared to the single monomer conformation maps (Figure 5.21A and D).

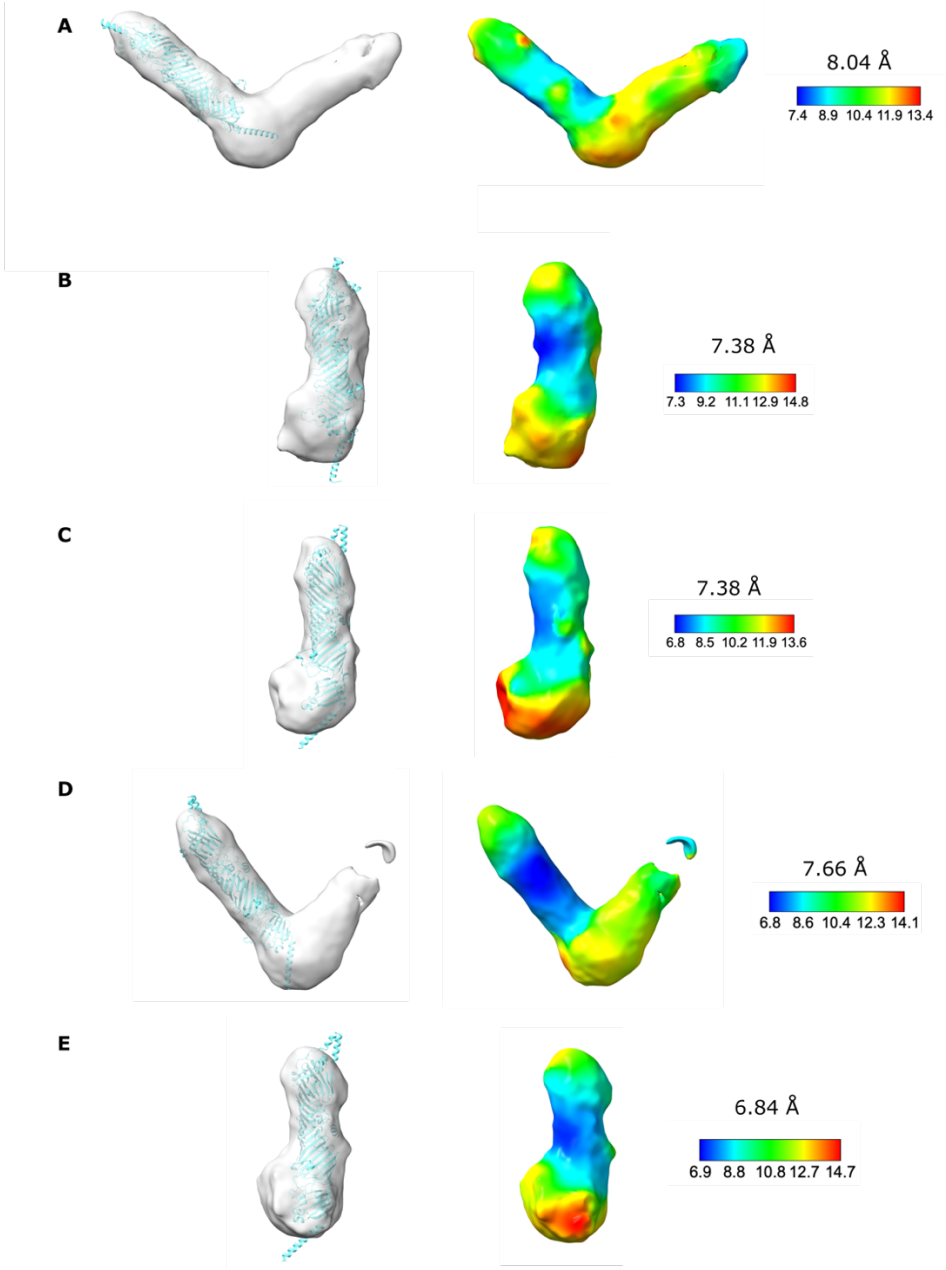


Figure 5.21. Final EM maps of YhjG in LDAO from various processing pipelines. A transparent volume with the AlphaFold model of YhjG (in blue) fit into the map is shown on the left, and the local resolution estimates are shown on the right. A key for each mapping is shown, along with the global resolution estimate. A – The hinged conformation final map of 8.04 Å after template picking using a hinged *ab initio* model. B – The monomer conformation final map of 7.38 Å after template picking using a hinged *ab initio* model. C – The monomer conformation final map of 7.38 Å after the Topaz ResNet16 model was used to pick particles. D – The hinged conformation final map of 7.66 Å from the processing of unbinned particles. E – The monomer conformation final map of 6.84 Å from the processing of the unbinned particles. All processing was performed in cryoSPARC v4.4.1, with figures generated in ChimeraX-1.7.1 (Punjani et al., 2017; Pettersen et al., 2021).

The fit of the YhjG AlphaFold model also seemed to agree more with density from the higher resolution "monomer" in the hinged conformation than with the single monomer maps. The N-terminal helix fit into the "micelle-like" density, with the periplasmic repeating β-groove (RBG) domain agreeing with the "tube-like" density, with only a few residues not explained by the density (Figure 5.21A and D). However, when assessing the fit of the model to the single monomer conformation map, the N-terminal helix and the C-terminal helix-turn-helix extended out of the density (Figure 5.21B, C and E). However, at this resolution, the model fitting was ambiguous, and it is more likely that the N-terminal helix does fit in this larger region of "micelle-like density", but that density information is lost towards the C-terminal region.

Assessment of the maps revealed that they all exhibited a cavity within the "tube-like" density. To demonstrate this, the best hinged conformation (Figure 5.21D) and single conformation (Figure 5.21E) maps were used to assess the density more closely. Figure 5.22 displays these maps at a higher contour level, which only shows the highest resolution densities from the map and allows closer inspection of density features.

In the hinged conformation map, the "second" monomer was essentially missing from the map at a high contour level, and the "micelle-like" region and the tip of the tube-like structure were also less defined (Figure 5.22A and B). This suggests that these regions were flexible, and that averaging and alignment of different positions of these regions had removed these regions from the final refined map. Focussing on the well-defined monomer, an agreement between the YhjG AlphaFold model and the density could be seen, with a cavity of diameter ~20 Å running throughout the periplasmic domain seen in both a spliced side view and top view.

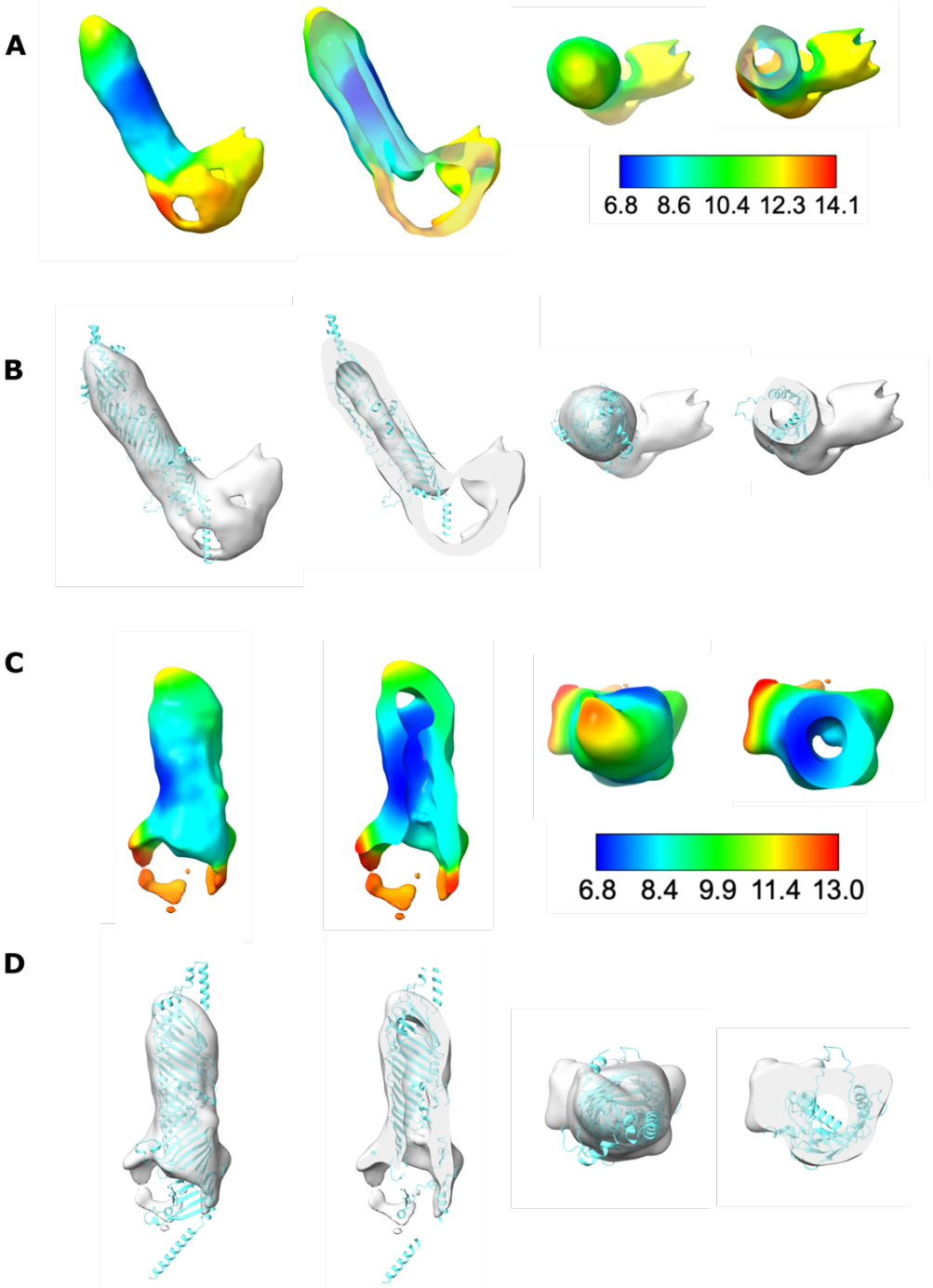


Figure 5.22. Hinged and single conformation final EM maps of YhjG in LDAO. A – Final map of the hinged conformation at a higher contour level with local resolution estimates shown. B – Final map of the hinged conformation at a higher contour level, with the AlphaFold model of YhjG overlayed. C – Final map of the monomer conformation at a higher contour level with local resolution estimates shown. D = Final map of the monomer conformation at a higher contour level with the AlphaFold model of YhjG overlayed. Each depiction shows the side view, a side view slice down the middle of the map, a top view, and a top view slice down the middle of the map. Local resolution estimate keys are shown.

The fit of the predicted AlphaFold model of YhjG into the single conformation map at a higher contour also revealed a lack of density for the N-terminal helix and the C-terminal helix-turn-helix, however a general agreement with the tube-like structure of the predicted model could still be seen (Figure 5.22C and D). The hydrophobic cavity was visible, with a width of 22 Å, slighter wider than the cavity in the hinged conformation. The fit of YhjG in the density suggested that a large N-terminal β-taco or repeating β-groove (RBG) like region sits in the wider, micelle-like density. If this is true, it could imply this N-terminal region is capable of various conformations, giving rise to ambiguous, low-resolution density.

It must be noted that any conclusions made from this model fitting are presumptive as the resolution is too low for any definitive model placement within the EM map. It is likely that the N-terminal helix should fit into the wider detergent micelle-like density, and that information towards the C-terminal region of YhjG was lost due to inherent flexibility, causing issues with particle averaging and alignment. It could also be possible that the AlphaFold model is longer than the real structure, or that YhjG experienced protein breakdown, and what the map revealed is only a partial structure of YhjG. A higher resolution map where a *de novo* atomic model can be generated would be needed to reveal this information.

YhjG in A8-35

Having experienced challenges with cryo-EM of YhjG due to various issues in grid preparation, particle distribution and data processing, an alternative grid preparation was tested as part of a proposal to work at the Membrane Protein Laboratory. Previous work as part of this proposal had identified improved particle distribution on negative stain EM for another AsmA-like protein, YdbH, when the amphipol A8-35 was used (Chapter 4). With the predicted structural similarity between the AsmA-like proteins, a grid was prepared with YhjG in A8-35 to encourage similar results. Using amphipols also removes background noise as there are no free

detergent micelles. Previous data processing results also showed a potential issue of two YhjG monomers existing within the same detergent micelle, therefore the use of A8-35 should, in theory, limit this.

Protein purification of YhjG in A8-35 was performed by exchange of purified protein from DDM into A8-35. The gel filtration yielded two main peaks (Figure 5.23A), unlike previous purifications of YhjG in detergents that normally show a single peak with a shoulder. Both peaks contained a band at the expected size for YhjG (75 kDa) on an SDS-PAGE gel, with the later peak also containing some smearing towards the bottom of the gel lanes (Figure 5.23B). This was assumed to be from excess amphipol in the sample. Fractions from the first peak were therefore taken forward for grid freezing to avoid the excess amphipol and because of the higher yield of protein.

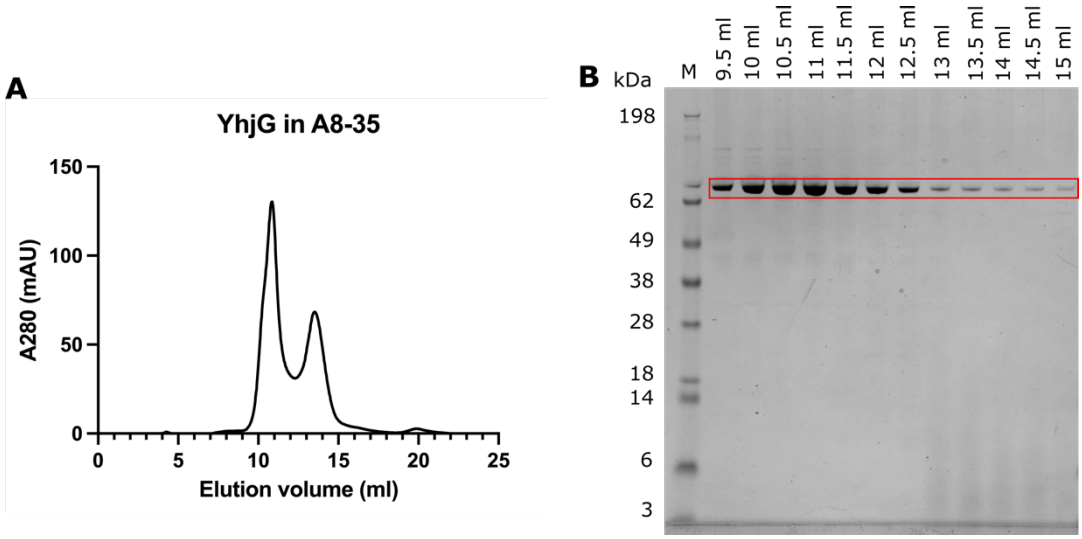


Figure 5.23. Purification of YhjG in the amphipol A8-35. A – Size exclusion chromatography trace after purified YhjG in DDM was incubated overnight with amphipol A8-35, mixed with Biobeads SM-2 for 90 minutes, then filtered and loaded onto a S200 10 300 increase column. B – SDS-PAGE analysis of peak fractions after SEC.

Multiple concentrations of YhjG in A8-35 were frozen on Cu 200 mesh R1.2/1.3 grids before screening was performed at the national eBIC cryo-EM facility. Three grids that were screened

had even ice thickness and showed promising particle distributions within grid holes and so, a full dataset was collected for each: 1.2 mg/mL, 0.9 mg/mL and 0.4 mg/mL YhjG in A8-35. Data collection parameters for each dataset are shown in Table 5.3. Representative micrographs from each of the data collections performed are shown in Figure 5.24. The protein was easily identifiable, with a similar size and shape to particles identified in the previous cryo-EM studies of YhjG in detergent. Noticeably, an electron-dense, black circle was visible at the "end" of each particle, which was assumed to be the amphipol "head" of the protein, as this was not evident in previous EM results using detergent.

Table 5.3. YhjG in A8-35 data collection parameters.

Parameter	YhjG 1.2 mg/mL	YhjG 0.9 mg/mL	YhjG 0.4 mg/mL
Detector	Falcon 4	Falcon 4	Falcon 4
Detector mode	Normal resolution	Normal resolution	Normal resolution
Voltage (kV)	200	200	200
Pixel size (Å)	1.192	1.192	1.192
Defocus (μm)	-2.5 to -1.0 in 0.3 μm intervals	-3.5 to -2.0 in 0.3 μm intervals	-2.5 to -1.0 in 0.3 μm intervals
Dose rate on specimen (e^-/px/s)	7.11	7.45	7.45
No. of frames	50	50	50
Exposure time (s)	12	11.4	9.5
No. of micrographs	7490	12727	8555

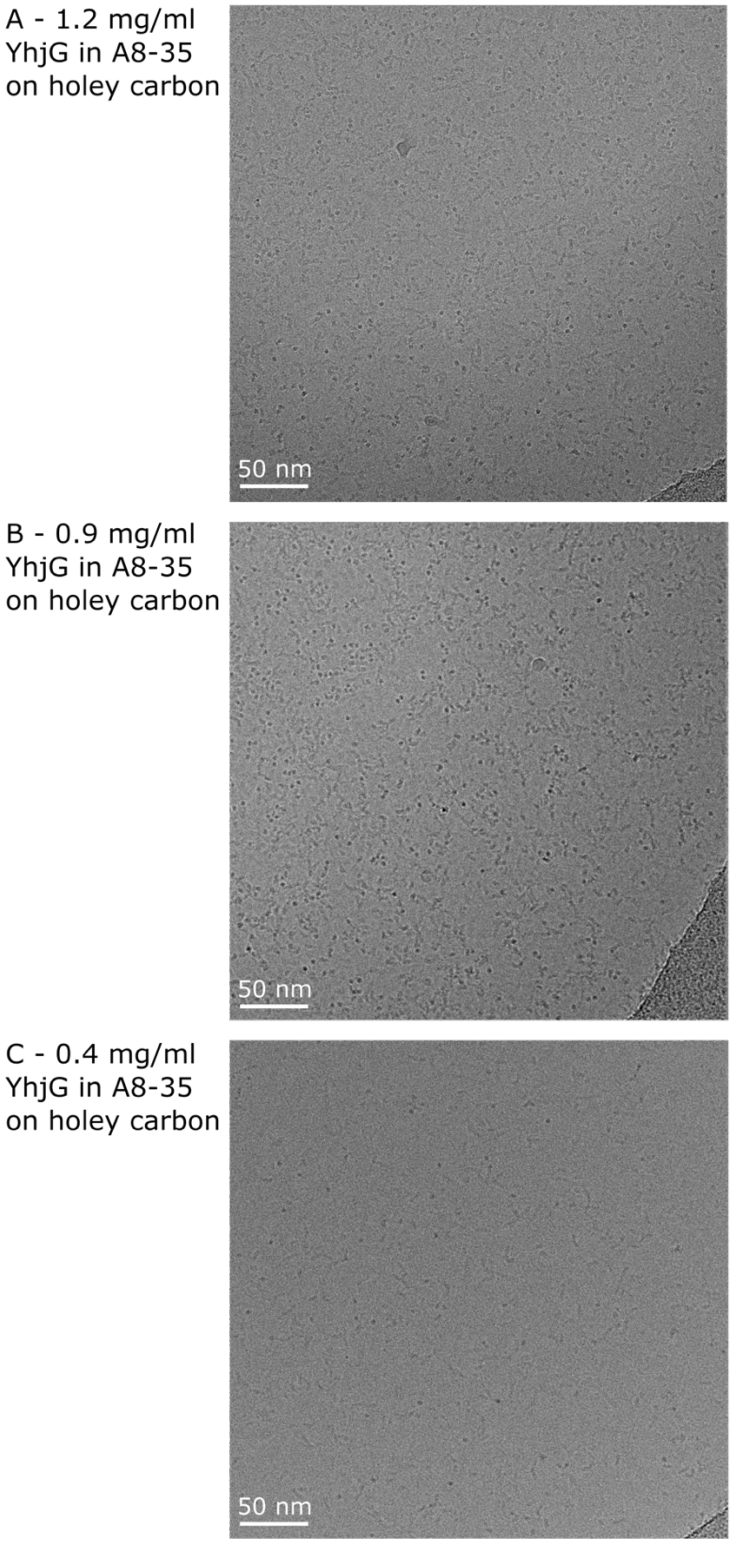


Figure 5.24. Micrographs from the three data collections of YhjG in A8-35. A – YhjG in A8-35 at 1.2 mg/mL. B – YhjG in A8-35 at 0.9 mg/mL. C – YhjG in A8-35 at 0.4 mg/mL. All three data collections were performed at eBIC cryo-EM facility on a Thermo Scientific Glacios II electron microscope.

Data collections were then automatically processed via eBIC's on-the-fly AutoProc pipeline using Relion v4.0 (Scheres, 2012). MotionCor2 was used to perform motion correction and CtfFind4 was used for CTF estimation, before automatic particle picking was performed using crYOLO picker and 2D classes were generated (Rohou and Grigorieff, 2015; Zheng et al., 2017; Wagner et al., 2019). An initial 3D *ab initio* model was also generated through this automatic processing before 3D classification into 4 classes.

Whilst visual inspection of the micrographs initially seemed promising, processing of the data was challenging. The automatic processing pipelines resulted in many 2D classes for each collection, some of which resembled a tube-like structure and some resembled top views of the structure, similar to the previous YhjG in LDAO dataset. However, automatic 3D model generation did not result in any models that resembled the anticipated structure (data not shown). It was therefore decided that manual data processing would be performed to improve initial particle picking and 2D class generation. As this work was performed at the national eBIC cryo-EM facility where the use of the data processing cluster is limited to internal users, any further manual processing of data collected was performed by Dr Peter Harrison at the Membrane Protein Laboratory using cryoSPARC v4.4.1 (Punjani et al., 2017).

Multiple particle-picking strategies were employed to improve the particle set used for downstream processing. These included cryoSPARC's blob picker – which, similar to the processing of YhjG in LDAO resulted in many false-positive picks; template picking with back-modelled 2D classes from the AlphaFold structure – which resulted in poor 2D classes; and the use of a Topaz model trained with manually picked particles – which saw improved results and was then used for further processing of each dataset.

The 2D classes generated from each dataset are shown in Figure 5.25. The presence of multiple particles in a single 2D class for all three datasets indicated that overcrowding was again an issue, even at the lowest concentration. Most classes resembled stick-like structures, but it was evident that the classes were longer than expected and exhibited a “bend” in the density, particularly prominent in the 2D classes for the lowest concentration dataset in Figure 5.27C. There were likely two particles positioned either end-to-end, resulting in density that was longer than the anticipated dimensions, or positioned similar to the “hinged” configuration of the YhjG in LDAO dataset, where two particles extend from a single point at various angles. Whether this “hinging” is then biologically relevant is unknown.

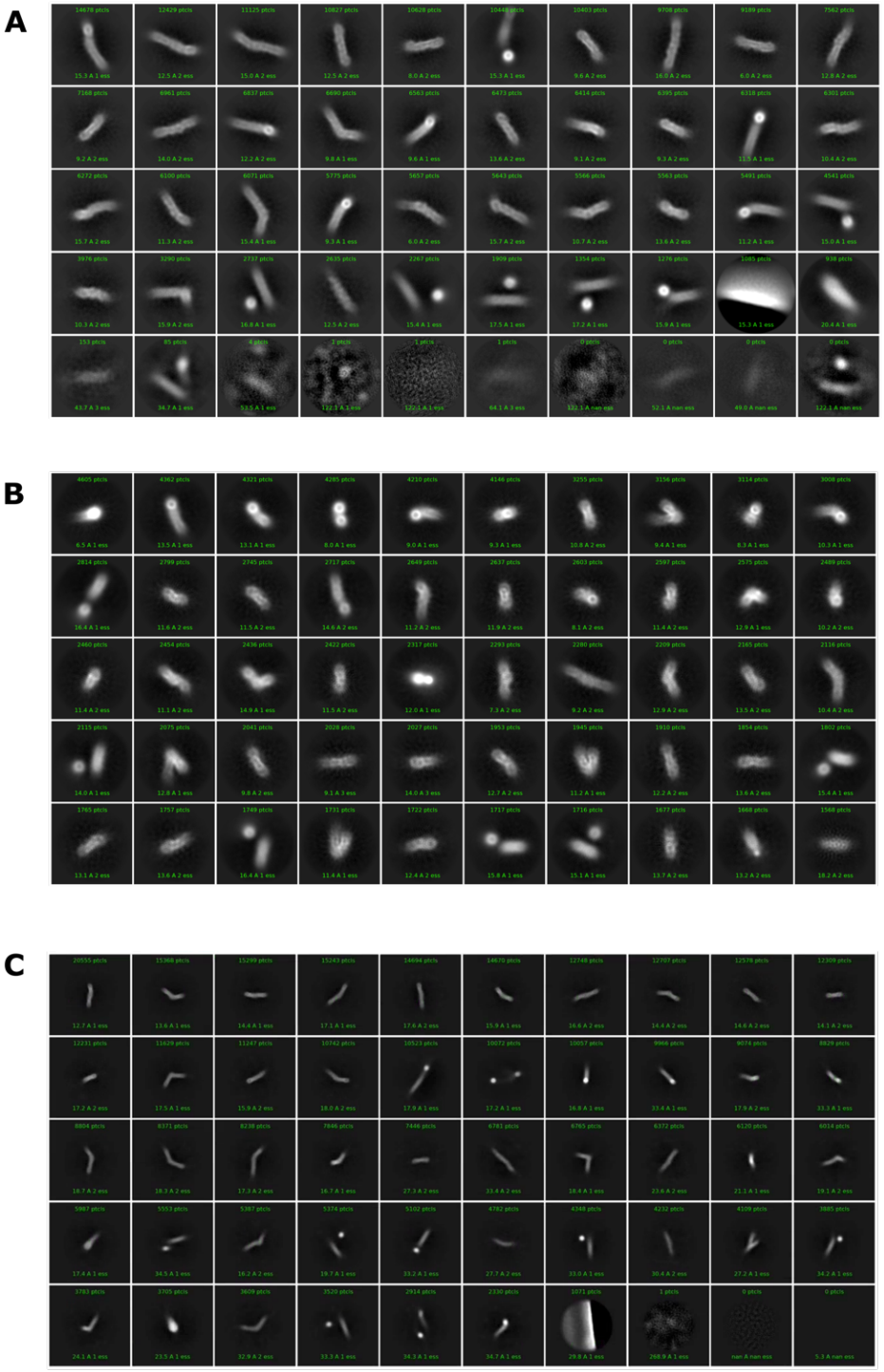


Figure 5.25. 2D classes from the three cryo-EM data collections of YhjG in amphipol A8-35. A – YhjG in A8-35 at 1.2 mg/mL. B – YhjG in A8-35 at 0.9 mg/mL. C – YhjG in A8-35 at 0.4 mg/mL. Data processing was performed by Dr Peter Harrison at the Membrane Protein Laboratory in cryoSPARC v4.4.1 (Punjani et al., 2017).

One 2D class was very promising, with the β-sheet structure and a micelle-like region visible, and dimensions that mostly agree with the AlphaFold model (Figure 5.26). The length calculated here was ~140 Å whereas the length of YhjG is predicted to be 170 Å. Like the YhjG in LDAO dataset, this could arise from the C-terminal helix-turn-helix or the N-terminal helix being flexible. Whilst further data processing was not performed here due to the inherent particle flexibility and particle heterogeneity of the rest of the 2D classes, the presence of this 2D class indicates that an atomic resolution structure solution using SPA may indeed be possible for YhjG, but that a higher quality input dataset is likely needed.

Figure 5.26. Promising 2D class of YhjG in A8-35. A 2D class from the initial processing of YhjG in A8-35 at 0.4 mg/mL of length ~140 Å and width ~30 Å.

5.3 Discussion

5.3.1 Structural elucidation of YicH

5.3.1.1 YicH has tube-like density

X-ray crystallography, SEC-SAXS, negative stain EM and cryo-EM were all employed to gain information on the structural properties of YicH. Unfortunately, no attempts at protein crystallisation were successful, and so this method was not discussed in this chapter. Likewise, SEC-SAXS of YicH was performed using both DDM-solubilised and SMA-solubilised YicH, but with poor buffer subtraction in the DDM sample and poor signal in the SMA sample, both were unable to be analysed and so were not discussed.

Negative stain EM and cryo-EM of YicH were more successful. Though initially there were concerns that the protein would be too small to be easily visible and identifiable on EM grids, stick-like particles of roughly the correct dimensions were easily seen via both negative stain and cryo-EM. At the time of collection, this was the first structural evidence of any of the full-length AsmA-like proteins. Notably, some particles seemed to display various curvatures and lengths when compared to the anticipated AlphaFold structure. Additionally, particle overcrowding was a recurring issue, even with various concentrations being tested, and caused issues with particle picking and downstream processing.

Subsequent data analysis revealed modest EM maps, depicting thin, elongated structures. Fitting the AlphaFold model of YicH into the final 9 Å cryo-EM map showed modest agreement between the predicted structure and experimental data, with density artefacts appearing from noise and the appearance of two defined regions of density where tube-like density was expected, which is not explained by the AlphaFold model. Due to issues with particle overcrowding on this grid, it could be possible that the second region of density is from an

overlapping particle within the extraction box, and that the fit of the AlphaFold model within the low-resolution map is false. As the fit of the model was ambiguous, definitive conclusions about the size and shape of YicH could not be drawn. However, what was interesting to see was the appearance of a cavity within the tube-like density.

Further inspection of this cavity is shown in Figure 5.27. The pore was calculated from the YicH AlphaFold model using the MoleOnline server, revealing a general agreement with the cavity in the YicH EM map (Pravda et al., 2018). There was slightly more variation in the pore diameter in the calculated pore compared to the EM cavity, but this was likely due to the low map resolution. The agreement of a tube-like structure validates the AlphaFold structural prediction. The cavity dimensions determined here would likely be able to accommodate phospholipids, and together with the predicted highly hydrophobic cavity lining, structural evidence towards YicH functioning as a phospholipid transporter can be seen. Utilising the ability to predict the joint structure of a protein and ligands on the AlphaFold3 server, the structure of YicH with 20 palmitate molecules, which would represent a phospholipid acyl chain, was predicted (Abramson et al., 2024). The resulting accommodation of lipid binding sites throughout the cavity provides further evidence of a capability of the YicH cavity to accommodate phospholipids (Figure 5.27).

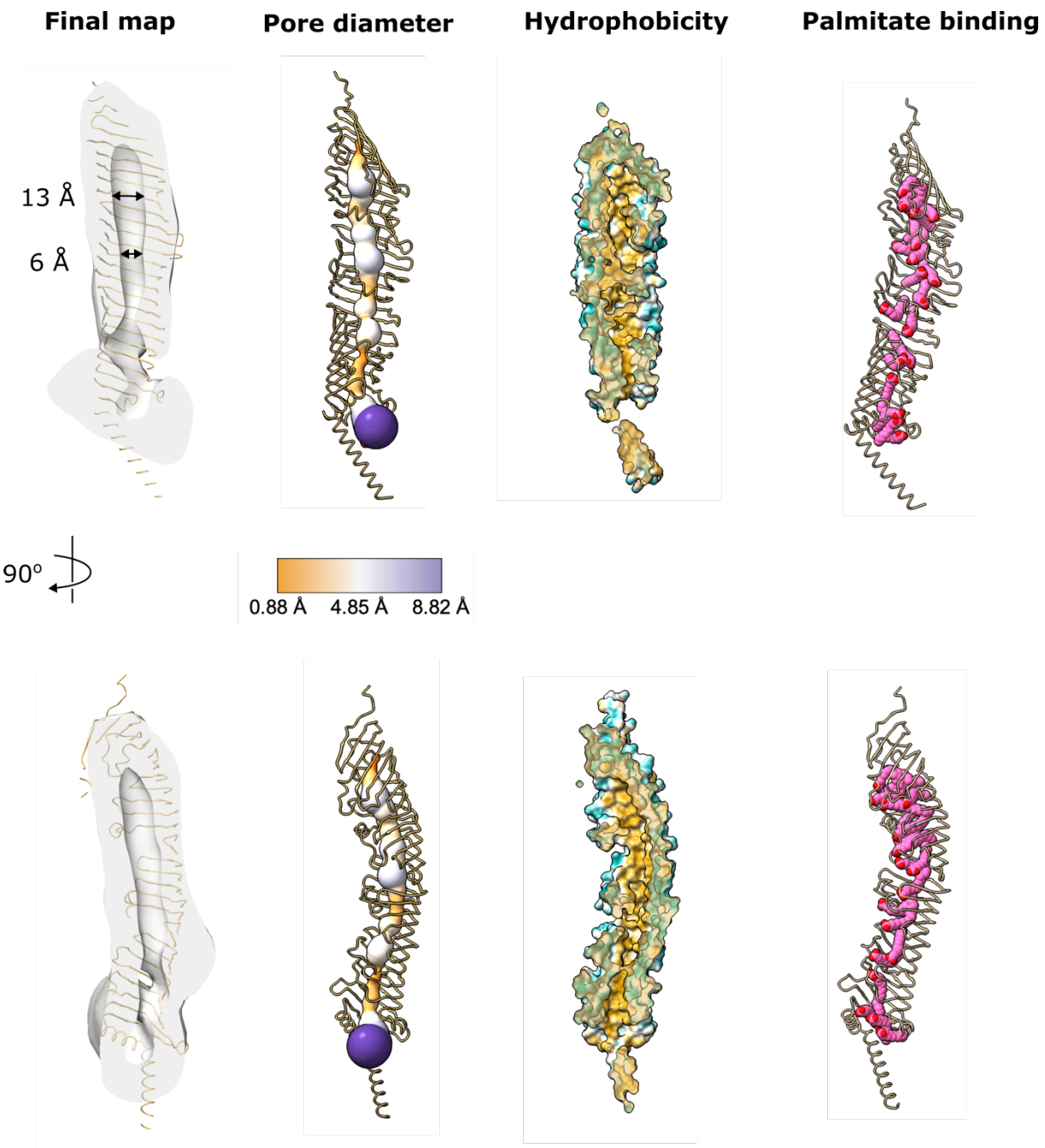


Figure 5.27. Evaluation of the final cryo-EM map of YicH and comparison to the AlphaFold model. Final cryo-EM map of YicH is shown in grey. AlphaFold structure shown as a ribbon in gold. Pore calculated using the MoleOnline server, coloured by pore diameter with key shown. Hydrophobicity surface of the AlphaFold model, from yellow to blue, hydrophobic to hydrophilic. A cavity is seen in final EM maps, potentially capable of carrying phospholipid cargo. The AlphaFold3 server was used to model the binding of 20 palmitate ligands.

5.3.2 Structural elucidation of YhjG

5.3.2.1 YhjG has tube-like density and is inherently flexible

Structural characterisation of YhjG was attempted by X-ray crystallography, EM and SEC-SAXS. Protein crystallisation was not successful, and like YicH, poor data quality was an issue with SEC-SAXS data in both DDM-solubilised and SMA-solubilised YhjG. More promising results were seen with investigation by EM. The negative stain and cryo-EM data showed that YhjG was easily identifiable on grids, with stick-like particles of varying lengths and curvatures, that closely resembled the YicH particles. The results of these investigations determined optimal particle-picking strategies and revealed that YhjG was inherently flexible.

A dataset collected of YhjG that had been purified in the detergent LDAO resulted in modest resolution EM maps. Multiple data processing strategies revealed two prominent conformations: a single monomer particle and a "hinged" particle of likely two YhjG monomers. The final monomer EM map of 6.84 Å resolution displayed a short tube-like structure, with larger density at one end. Whilst this was initially attributed to the detergent micelle, fitting of the AlphaFold model of YhjG into the map revealed that this larger region of density could be coming from flexibility of an N-terminal region. With this fit, density for the N-terminal helix and the C-terminal helix-turn-helix was absent. This could mean these regions were too flexible for density to be present in the final map, but it is more likely that the model fitting at this resolution is poor. Due to this ambiguity, it was difficult to accurately model YhjG into the density and to determine exactly which region of YhjG could be causing inherent flexibility.

The hinged conformation showed improved density for one of the YhjG monomers making up its "V" conformation. This revealed density of the anticipated dimensions for YhjG, with a

narrow cavity spanning the whole repeating β-groove (RBG) periplasmic structure. Interestingly, this hinged conformation allowed better determination of the structure of a monomer of YhjG, likely due to improved particle alignment with a larger particle, as YhjG at 74 kDa is relatively small for cryo-EM investigation.

Inherent continuous flexibility represented a major challenge in determining the atomic structure of YhjG. The lack of distinct conformations would require an infinite number of 3D classes to obtain an atomic resolution structure. Nevertheless, a low-resolution density map was obtained that to some extent validated the AlphaFold structural prediction, with both structures containing a cavity throughout the tube-like structure.

Further inspection of this cavity is shown in Figure 5.28. The pore was determined from the YhjG AlphaFold model using the MoleOnline server, revealing a pore of maximum dimension 13.6 Å (Pravda et al., 2018). Comparing this pore with the cavity in the "monomer" EM map reveals a similar pinching at the mid-point of the tube, with diameters of 1-4 Å in the calculated pore and 8 Å in the map. Interestingly, density just below this pinch point was present in this map which is unaccounted for in the AlphaFold structure. Perhaps this is a "closed" conformation, compared to the predicted "open" structure. Evaluation of the "hinged" map revealed an "open" conformation, with a continuously ~20 Å wide cavity throughout the tube. Whether this is an effect of data processing and model fitting or a mechanistically relevant finding cannot be determined at this resolution. Nevertheless, the presence of this cavity in both final maps and the predicted highly hydrophobic lining (Figure 5.28) provides evidence that the cavity of YhjG would be able to accommodate phospholipids. Like YicH, further evidence of a capability of the YhjG cavity to bind phospholipids can be seen with the placement of 20 palmitate molecules throughout the length of the cavity when modelled with the AlphaFold3 server (Figure 5.28) (Abramson et al., 2024).

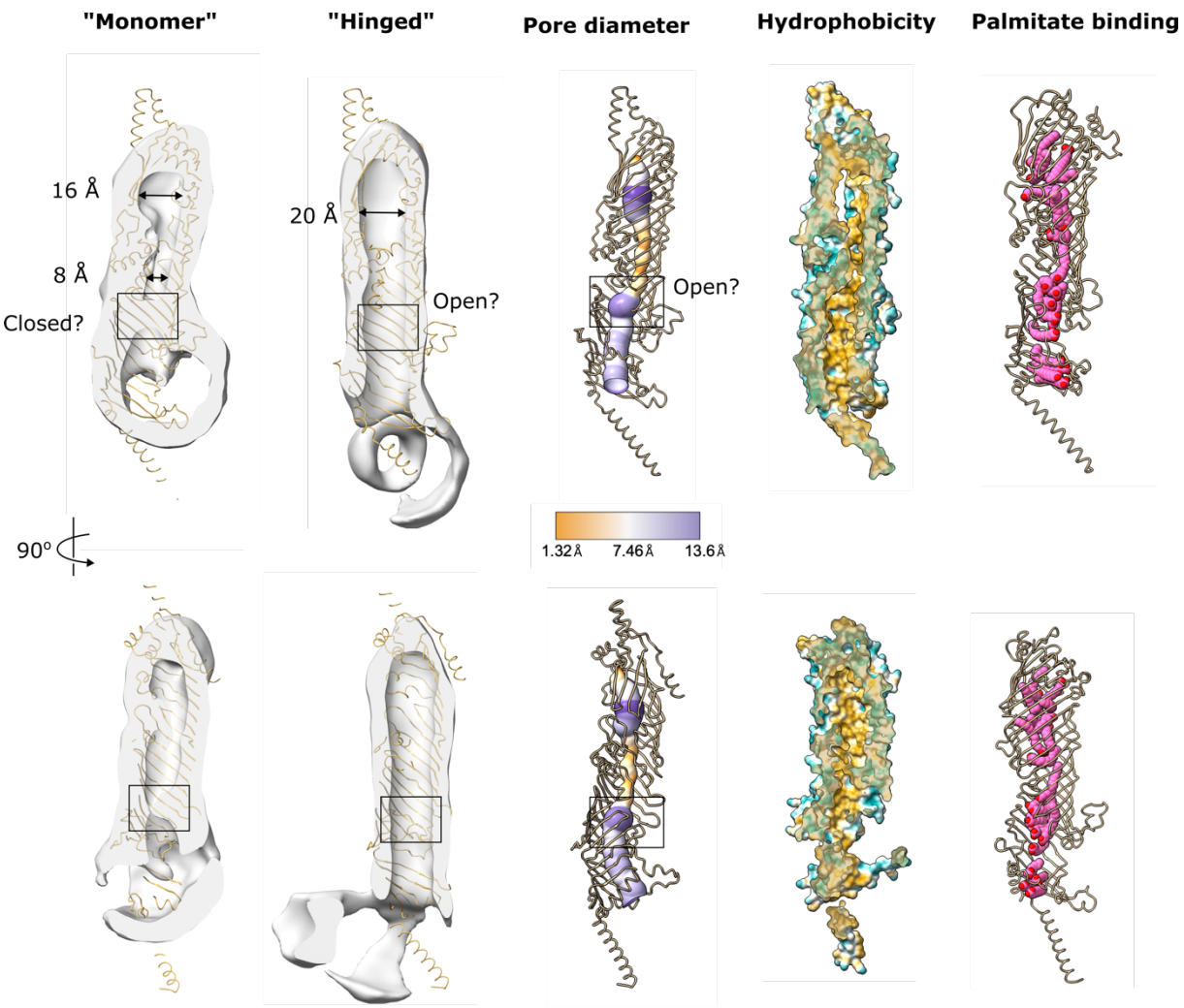


Figure 5.28. Evaluation of the final cryo-EM maps of YhjG and comparison to the AlphaFold model. "Monomer" and "hinged" final cryo-EM maps of YhjG are shown in grey. AlphaFold structure shown as a ribbon in gold. Pore calculated using the MoleOnline server, coloured in pore diameter. Hydrophobicity surface of the AlphaFold model, from yellow to blue, hydrophobic to hydrophilic. A cavity is seen in both the "monomer" and "hinged" final maps, of varying diameters, adopting a potential open and closed state. The AlphaFold3 server was used to model the binding of 20 palmitate ligands.

5.3.2.2 Can YhjG adopt a dimer conformation?

The cryo-EM of YhjG provided evidence suggesting the existence of two YhjG monomers that can pack end-to-end with varying angles between them. This led to the question of how this "hinged" conformation was occurring: was it an effect of the purification process and/or grid freezing or can YhjG exist as a dimer in the cell? Other lipid transporters are often dimers,

including those from the tubular lipid-binding (TULIP) superfamily (Wong and Levine, 2017). In the elucidated structure of the SMP domain (a member of the TULIP superfamily) homodimer from human E-Syt2 (extended-synaptotagmin 2), there is a head-to-head configuration of two SMP domains (Figure 5.29A). Here, a cavity capable of hosting hydrophobic ligands extends across the resulting dimer (Schauder et al., 2014). However, these lipid transfer proteins (LTPs) are soluble and often need to dimerise to span inter-membrane distances. YhjG on the other hand is expressed into the inner membrane and is predicted to be ~170 Å long. Therefore, a dimer configuration of end-to-end YhjG monomers would not make biological sense. AlphaFold multimer was used to predict the fold of a YhjG dimer, to assess whether the predicted conformation resembled that of the "hinged" conformation seen on cryo-EM grids. Figure 5.29B shows the resulting unlikely dimer conformation, with the PAE matrix indicating very low confidence in the conformation. The dimer seen is likely then a result of the purification or EM strategy employed. Whilst not biologically relevant, the hinged configuration does appear to aid with structure determination.

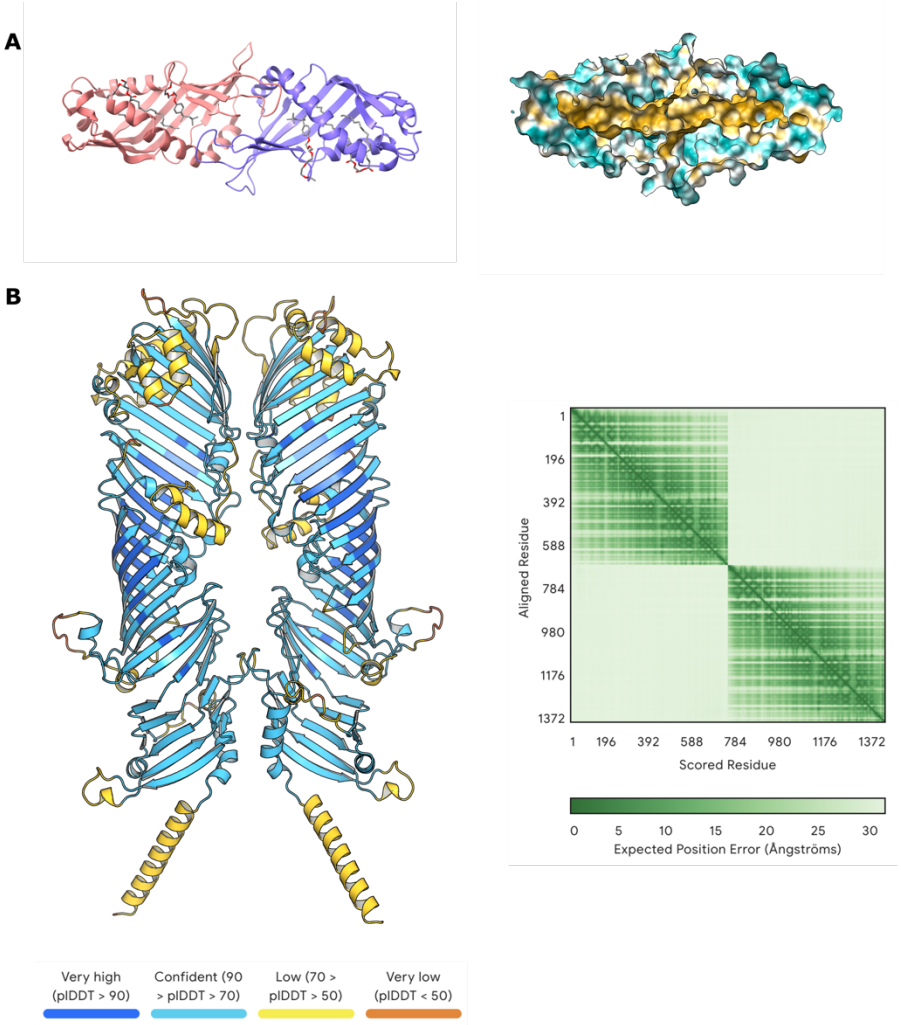


Figure 5.29. Lipid transfer protein dimers. A – The homodimer conformation of the SMP domains from human E-Syt2, shown in cartoon coloured by SMP domain with ligands present as sticks and shown as a sliced hydrophobic surface, with the hydrophobic cavity visible (Schauder et al., 2014). B - AlphaFold model of two copies of YhjG. Model coloured from orange to dark blue from low to high pLDDT scores as depicted in the key. The predicted aligned error (PAE) matrix is shown for the model, indicating low confidence in this dimeric conformation. Figures made from PDB: 4P42, AlphaFold3 server, UCSF ChimeraX and PyMOL (Schauder et al., 2014; Schrodinger, 2015; Jumper et al., 2021; Pettersen et al., 2021).

5.3.3 Implication of study for the whole AsmA-like protein family

This investigation provided valuable preliminary structural data and a promising starting point for future structural studies of the AsmA-like protein family. Some reasonable assumptions can be made about the amenability of the other four AsmA-like proteins to certain structural

techniques based on these findings. With the lack of success of crystallisation of the smaller AsmA-like proteins YhjG and YicH, it is probable that the larger proteins (YhdP, TamB, YdbH) will also not successfully crystallise. Of course, there are an indefinite number of conditions that can be tested for crystallisation, and it may well be that the right conditions were simply not trialled here.

Negative stain EM and cryo-EM were employed more successfully for YicH and YhjG. Low-resolution structures were obtained for both proteins, representing the first EM maps of any the full-length AsmA-like proteins. Many protein concentrations and grid preparations were tested with varying success, and it is likely that the other proteins would behave similarly on grids. Therefore, with the most successful cryo-EM grids being from the preparation of YhjG in the detergent LDAO or the amphipol A8-35 and without any graphene oxide coating, these could be good starting points for the other AsmA-like proteins. The varying success of particle picking strategies should also be used to influence and speed up data processing of future datasets, with Topaz yielding the best particle picking results compared to the standard blob picker or template picking.

5.3.4 Flexibility and lengths of Gram-negative envelope spanning structures

The experimental evidence of flexibility within YicH and YhjG not only explained some of the complications with obtaining high-resolution EM maps but could reveal hints about protein function. Perhaps the flexibility of the protein is integral to its function within the cell, with the protein requiring various degrees of curvature. Flexibility in Gram-negative trans-envelope protein complexes has been seen previously. Our lab has identified large dynamic radial movements of the top of the PqiB needle, part of the trans-envelope Paraquat-inducible (Pqi) putative lipid transport pathway (Cooper, 2022). We also identified that PqiB most likely needs

the outer membrane PqiC component to stabilise the needle and found that this interaction was necessary for protein function (Cooper et al., 2024).

It could be the case that the AsmA-like proteins need an OM component to stabilise the long periplasmic domain. It is known that TamB interacts with the TamA OM component and that YdbH interacts with the OM lipoprotein YnbE (Shen et al., 2014; Kumar et al., 2024). No other OM protein partners have yet been identified for the other AsmA-like proteins, though attachment to the OM itself could provide stability. YhjG, AsmA and YicH are not long enough theoretically to span the whole 200 Å wide periplasm (Matias et al., 2003). The predicted lengths of each of the proteins are shown in Figure 5.30. The potential of the C-terminal helix-turn-helix to be OM bound also reduces the predicted length of the periplasmic domain for YhjG and AsmA. Whether these smaller proteins can span the periplasm to the OM or just extend partially into the periplasm is unknown.

There are occasions where the periplasm will likely be shorter, perhaps during growth and upon certain stress conditions. Other studies have identified trans-envelope protein complexes of lengths ~125-170 Å, indicating regions of the periplasm may indeed take on a narrower conformation (Collins et al., 2007; Marlovits et al., 2004). One of these complexes, Wza-Wzc, is responsible for capsular polysaccharide (CPS) export from the cell, which are excreted polymers used in cellular defence (Collins et al., 2007). The structure of this complex revealed a periplasmic region of 125-145 Å. Also, the Type III secretion system, responsible for the injection of effector proteins into host cells, is a trans-envelope spanning complex with a periplasmic region of only 170 Å (Figure 5.34) (Marlovits et al., 2004; Burkinshaw et al., 2015). This evidence suggests that the periplasm may take on a range of widths, perhaps to allow the adoption of a trans-envelope complex. Bayer's sites of adhesions, which were initially discredited due to being microscope artefacts, may also provide evidence of regions of contact

between the two membranes, which would result in narrower distances between the two membranes either side of the adhesion site (Bayer, 1968). Variations in the *E. coli* periplasm width have also been suggested by recent evidence that found Braun's lipoprotein (Lpp) to be responsible for creating an enlarged periplasm (Cohen et al., 2017). As well as the OM and OM binding partners, other periplasmic proteins or the peptidoglycan layer could provide structural stability to the proteins within the cell.

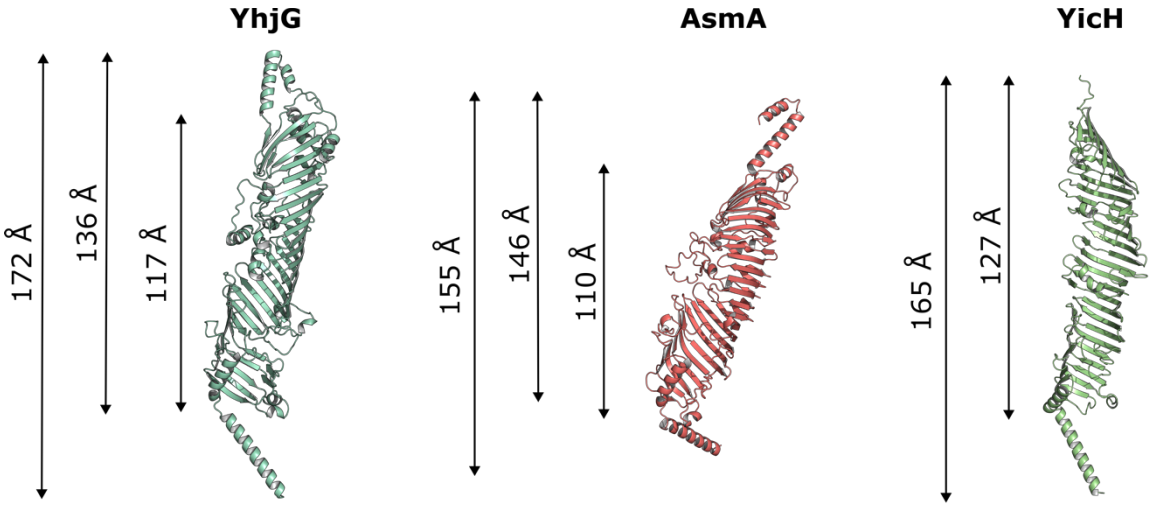


Figure 5.30. Lengths of smallest AsmA-like proteins. The calculated lengths are shown for the AlphaFold models of YhjG, AsmA and YicH. The periplasm is normally 200-250 Å.

5.3.5 Advances in the structural elucidation of bridge-like lipid transport proteins

As the AsmA-like proteins share structural homology with eukaryotic bridge-like lipid transport proteins (BLTP), the advances in the structural elucidation of these proteins can influence the design of future experiments for the AsmA-like proteins (Neuman et al., 2022).

One BLTP that has been recently characterised is the autophagy-related protein ATG2A (Wang et al., 2024b). This protein is responsible for phospholipid transport between the ER and autophagosome membranes. The structure of ATG2A was elucidated as a complex with one or both of its partner proteins (ATG2A-WIPI4 to 3.2 Å, ATG2-ATG9-WIPI4-GABARAP to 7.0 Å) (Figure 5.31A). Here, the authors also describe difficulty in data processing due to flexibility

and were unable to model the N-terminal domain responsible for lipid extraction in both maps (Wang et al., 2024b). Additional structures of BLTPs include an N-terminal fragment of Vps13 to 3.75 Å and SHIP164 (a member of the Vps family) to 8.3 Å (Figure 5.31B and C) (Li et al., 2020; Hanna et al., 2022). It is interesting to note that the authors found that Vps13 forms tail-to-tail dimers on the EM grid, and only used one monomer for processing. This is reminiscent of the longer-than-expected particles seen for both YicH and YhjG on EM grids, and of the "hinged" conformation of YhjG. These maps show remarkable similarity to the EM maps obtained for YhjG, particularly the map of SHIP164, which was only resolved to 8.3 Å.

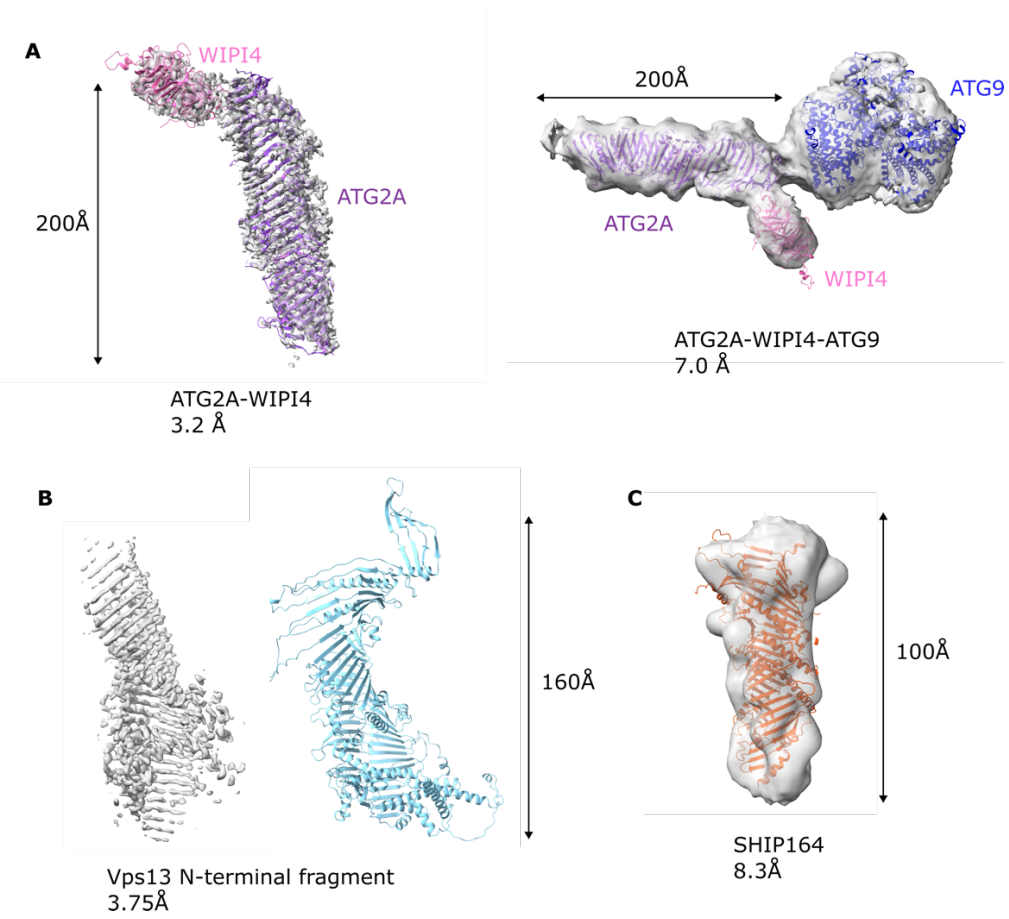


Figure 5.31. Published cryo-EM maps of structurally similar BLTPs. A – The map of ATG2A-WIPI4, solved to 3.2 Å, and ATG2A-WIPI4-ATG9, solved to 7.0 Å. B – The N-terminal fragment of Vps13, solved to 3.75 Å. C – The N-terminal fragment of SHIP164, solved to 8.3 Å. Figures adapted from (Wang et al., 2024b), (Li et al., 2020) and (Hanna et al., 2022), respectively.

A recent paper in BioRxiv determined the structure of the N-terminal domain of a BLTP called LPD-3, which revealed the consistent repeating β-groove (RBG) domain and identified 27 phospholipids inside the hydrophobic cavity (Kang et al., 2024). Two partner proteins were also characterised. One bound the BLTP to the membrane and bordered the entrance to the tunnel, which they named "Intake", and the other called "Spigot" also had one TM helix and extended 80 Å along the structure, which they proposed helps with contact site formation. There is some similarity here to the putative complex formation between YdbH, YnbE and YdbL. The EM map of the full-length protein was remarkably similar to the EM maps obtained from the YhjG in LDAO dataset collected here, with a tube-like structure, density for a detergent micelle, and a cavity throughout the RBG domains present.

Utilising the cryo-EM methods used in these studies could be applied to the determination of AsmA-like protein structures. For ATG2A, a crosslinker DSBU was used to cross-link ATG2A with its partner, WIPI4, which may have enforced protein stability (Wang et al., 2024b). They also used the detergent LMNG to solubilise the membrane-bound partner protein ATG9 for the larger complex structure. For Vps13 and SHIP164, the authors spiked grids with the detergent β-OG (Li et al., 2020; Hanna et al., 2022). The determination of the N-terminal domain of LPD-3 utilised glycol-diosgenin as the detergent used to solubilise and stabilise LPD-3 (Kang et al., 2024). These protein purification and grid preparation conditions should be trialled for the AsmA-like proteins, and other long, flexible proteins, in future studies. Although, it should be noted that the best resolutions achieved for these BLTPs were for soluble fragments of the full-length proteins, with comparable resolutions to the YhjG in LDAO dataset achieved for membrane-bound complexes and full-length BLTPs.

5.3.6 Future research recommendations

As mentioned previously, trialling previously successful cryo-EM preparations and processing pipelines for other BLTP proteins should be trialled on the AsmA-like proteins, as well as utilising the most promising grid preparation and particle picking strategy found in this study.

Additionally, as protein flexibility seems to be an inevitable problem when investigating structures of BLTPs, this should be further investigated. For cryo-EM in particular, particle flexibility can be assessed via advanced data processing. Programs like multi-body refinement have previously been used to identify different molecular conformations present, however, a rigid body must first be defined, which introduces data bias and would be difficult for a small, monomeric protein like YhjG where it is difficult to split into 'domains' (Nakane and Scheres, 2021). Other programs such as cryoDRGN and Zernike3D have recently been introduced that describe a particle's continuous flexibility using deep-learning algorithms and have shown promising results for other flexible proteins (Herreros et al., 2023; Zhong et al., 2021). Future experiments should look at utilising these programs to investigate the potential conformations adopted by YhjG, and for future cryo-EM datasets of other AsmA-like proteins.

Otherwise, to improve the likelihood of higher resolution EM structures and to improve the likelihood of protein crystallisation, a step back to protein design could be necessary. As particle misalignment is likely a factor that is limiting the resolution of the data presented in this study, methods to improve particle alignment could be trialled. One way to improve this would be to make the protein have a large domain on only one side of the tube-like structure, allowing unambiguous particle averaging and alignment. This could be achieved with the addition of nanobodies or megabodies (a larger version of nanobodies), which have seen improved results for other cell envelope proteins, or by engineering a rigid chimera with a protein of known structure (e.g. β1-adrenoreceptor fused with AmpC β-lactamase), with both strategies having

the advantage of making these sub-100 kDa proteins much larger (Collu et al., 2021; Botte et al., 2022; Wentinck et al., 2022). These techniques could also benefit protein crystallisation, with the potential to stabilise the protein and influence crystal packing.

Another idea generated during this project to promote crystal formation was to exchange the N-terminal helix with another single TM helix from a PDB-deposited crystal structure. The aim here was to encourage the crystal packing that was present in the deposited structures. Whilst this project never came to fruition during the timescale of this project, this would be an interesting avenue to take for future crystallisation trials of these proteins. Similarly, efforts to create a soluble form of YicH and YhjG were made during this project, but as previously discussed, these were not successful (Chapter 4). The creation and production of soluble protein would greatly increase the chance of crystallisation success due to not needing detergents or other membrane encapsulation agents, and therefore increasing sample homogeneity.

Additionally, whilst the SEC-SAXS data collected during this study was of too poor quality to analyse confidently, future experiments should re-attempt this analysis on YhjG and YicH and perform this on the other AsmA-like proteins as it is a relatively easy technique which can produce useful information about protein flexibility and protein conformations in solution. Ideally, soluble versions of the proteins would be utilised to minimise the detergent background, which was largely responsible for the poor data quality in the DDM-solubilised samples tested in this study. Alternative membrane mimetics could also be used, either re-attempting SMA-solubilised protein samples or using A8-35, as this had seen improved monodispersity on cryo-EM grids for YhjG and YicH.

5.3.7 Conclusion

In this chapter, determining the structures of two AsmA-like proteins YicH and YhjG was attempted. There is limited information in the literature about these two proteins and previous genetic studies did not reveal a potential role, therefore structural determination was desirable. For the first time, we reveal the cryo-EM maps of YicH and YhjG to a resolution of 9 Å and 6.8 Å, respectively. These low-resolution maps revealed that the AlphaFold predicted models generally agree with the experimentally determined data, with a clear cavity shown that could be used to transport biomolecules. Additionally, information was revealed about the likely inherent flexibility exhibited by the two proteins through issues in cryo-EM data processing. With the predicted structural similarity between all six AsmA-like proteins, and inherent flexibility seen in other cell envelope proteins and BLTPs, they all probably exhibit similar flexibility. The results and literature findings discussed in this chapter can be used to influence the design of future experiments for the structural determination of the whole AsmA-like protein family.

CHAPTER 6. FUNCTIONAL CHARACTERISATION OF YICH, YHJG AND YDBH

6.1 Introduction

6.1.1 The proposed phospholipid transport function of the AsmA-like proteins

The current evidence for the AsmA-like proteins having a role in phospholipid transport between the inner and outer membrane of Gram-negative bacteria mostly comes from genetic investigations and inferences from the predicted protein structures (Grimm et al., 2020; Douglass et al., 2022; Ruiz et al., 2021). A recent study in BioRxiv starts to address this by showing that ^{32}P-containing substrates, likely phospholipids, are found at various locations along the channel of YhdP through benzoyl-phenylalanine (BPA) cross-linking experiments (Cooper et al., 2023). They also perform molecular dynamics simulations on YhdP, identifying diffusive transport of phospholipids into the hydrophobic cavity. Similar studies of the other AsmA-like proteins have not yet been reported. Therefore, further direct biochemical evidence showing that phospholipids can bind and be channelled through these proteins is necessary.

The research presented here aims to address the lack of biochemical evidence of phospholipid transport for three of the AsmA-like proteins: YhjG, YicH and YdbH. With these three proteins being more amenable to purification compared to YhdP, TamB and AsmA, they were utilised to develop a phospholipid transport assay suitable for bridge-like membrane proteins. This assay should in theory be applicable to the whole AsmA-like family, and other membrane proteins of putative lipid transport function. Molecular dynamics simulations were then employed on these three proteins to investigate potential mechanisms of phospholipid transport.

6.1.2 Theory behind phospholipid transfer FRET-based assay

6.1.2.1 Introduction to FRET

Förster resonance energy transfer (FRET), is a phenomenon where energy is transferred from one molecule (a donor) to another (the acceptor) when the two are in within nanometre distances of each other (Hochreiter et al., 2015). The Jablonski diagram in Figure 6.1 demonstrates this effect. When a fluorophore absorbs a photon, an electron is raised from a ground state, S_0, to an excited energy state, S_1. Fluorescence occurs when the vibrational relaxation of this unstable state drops the electron back to its ground level, releasing a photon of longer wavelength within nanoseconds. FRET, on the other hand, occurs when the electron drops to the ground level, and the quantum energy is instead transferred to an acceptor through dipole-dipole coupling, exciting an electron in the acceptor molecule. The relaxation of this electron to the acceptor's ground state either releases a photon, fluorescing, or quenches the fluorescence.

The two energy states must be compatible for this to occur. This means an overlap of the emission wavelength range of the donor and the excitation wavelength range of the acceptor. The two chromophores must also be within 10-100 Å of each other. The efficiency of FRET decreases with increased distance between the donor and the acceptor, making FRET an excellent technique for measuring intermolecular distance with high sensitivity (Hochreiter et al., 2015).

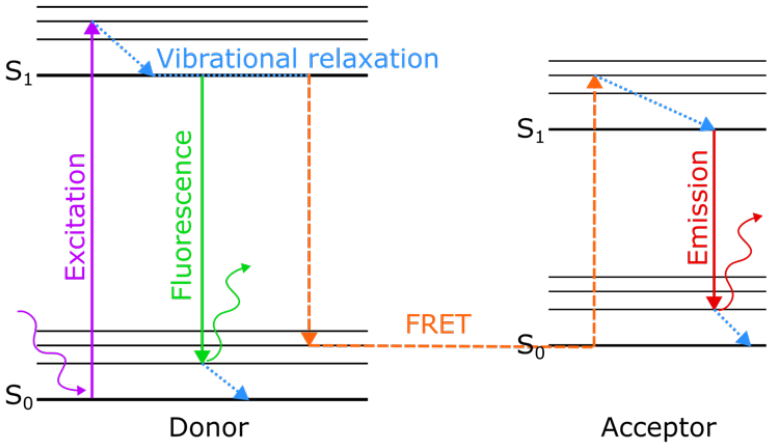


Figure 6.1. Jablonski diagram showing fluorescence and FRET. Fluorescence occurs when a chromophore absorbs a photon of light (purple), exciting an electron to the excited state S_1. Vibrational relaxation (blue) of the unstable state causes the electron to drop back to the ground state S_0, releasing a photon (green). FRET occurs if two chromophores are within 10-100 Å of each other, have overlapping excitation and emission spectra, and exhibit dipole-dipole coupling. In this case, energy is transferred from the donor chromophore to the acceptor, where an electron is excited (orange). When the excited electron in the acceptor chromophore drops to the ground state, a photon is released (red). Figure adapted from (Hochreiter et al., 2015).

6.1.2.2 Design of FRET-based assay

Lipid transfer has previously been measured multiple times *in vitro* using FRET-based assays for a wide range of proteins and lipids. Many studies have looked at the lipid transfer capabilities of tubular lipid-binding (TULIP) domain-containing proteins and other lipid transfer proteins (LTPs) through measuring changes in FRET of fluorescently labelled lipids between sites of transfer (Nichols and Pagano, 1983; Osawa et al., 2019; Bian and De Camilli, 2019; Wang et al., 2023).

A FRET-based assay was developed based on these previously published assays to assess the lipid transfer capabilities of the AsmA-like proteins that could be expressed and purified. The assay set-up is shown in Figure 6.2. Within the assay mixture, there are two populations of liposomes; a non-fluorescent "acceptor" liposome formed from 100% POPC lipids, a

commonly used model lipid, and a fluorescent "donor" liposome formed from 97% POPC, 1.5% NBD-PE and 1.5% Rhodamine-PE. NBD-PE and Rhodamine-PE are the most frequently used fluorescent probe-tagged lipids for FRET-based lipid assays, with an overlapping emission spectrum of NBD and excitation spectrum of Rhodamine and a high sensitivity even at low probe concentrations (Malinin et al., 2001). The addition of a lipid transporter capable of moving phospholipids between the two liposome populations should change the distribution of fluorescent lipids from existing in just the donor liposomes, to both the donor and acceptor liposomes. This will have the effect of diluting the fluorescent phospholipids amongst the unlabelled lipids, decreasing the occurrence of FRET (Bian and De Camilli, 2019). As FRET reduces, the fluorescent emission of NBD will increase, and this can be measured continuously with high sensitivity.

Previously reported FRET-based lipid transport assays use soluble lipid transporter proteins (Bian and De Camilli, 2019). Attempts at purifying non-membrane tethered constructs of YhjG, YicH and YdbH were unsuccessful, and removing the N-terminal helix could hinder any phospholipid transporter function, and so being able to monitor transport whilst accounting for the membrane bound region of the protein was important. Formation of proteoliposomes was therefore desirable, as this both models the protein inside a membrane environment and revokes the need for detergents. Not only do detergents place the protein in a non-native environment, but they would also likely interfere with the liposomes and impact protein function. It has been previously reported that the rapid dilution of a protein-detergent micelle complex into small unilamellar vesicles (SUVs) in a non-detergent buffer allows reconstitution of the protein into proteoliposomes (Ambudkar et al., 1986; Goddard et al., 2015; Godoy-Hernandez et al., 2023). This occurs because rapid dilution lowers the concentration of detergent below the critical micelle concentration (CMC), causing proteins to transfer into the liposomes. This method was

employed in this assay to produce a mixed population of non-fluorescent and fluorescent proteoliposomes, with minimal detergent concentrations present in the final assay mix.

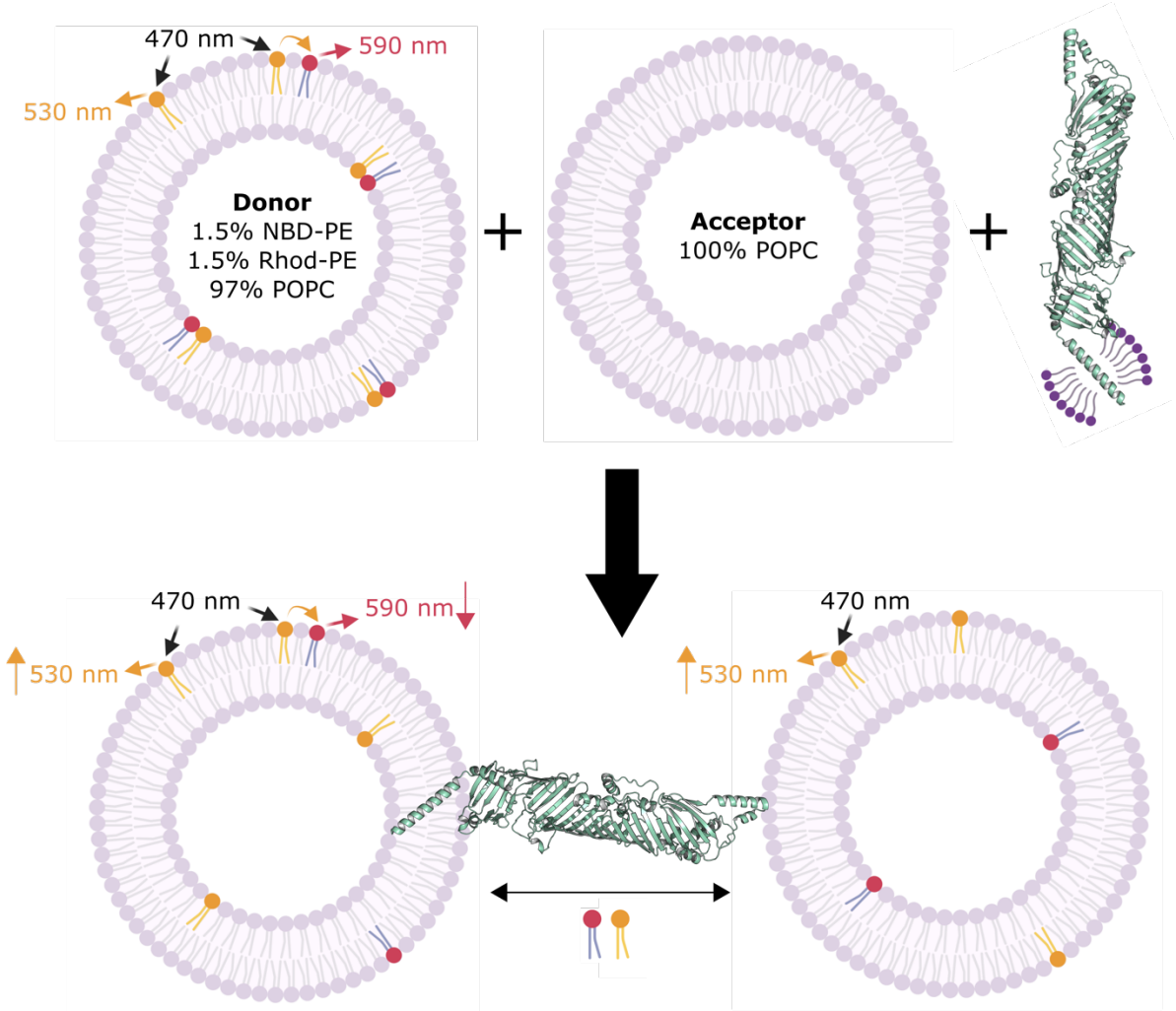


Figure 6.2. FRET-based phospholipid transfer assay set up. A "donor" liposome containing 1.5% NBD-PE (yellow), 1.5% Rhodamine-PE (red), 97% POPC (purple) was mixed with 100% POPC "acceptor" liposomes. The DDM-solubilised protein is then added to the mix, such that the DDM concentration goes below the CMC, allowing reconstitution of the protein into proteoliposomes. Excitation of the NBD-PE fluorescent lipid at 470 nm causes fluorescence emission of NBD at 530 nm. If a Rhod-PE fluorescent lipid is within FRET distance of the NBD-PE, FRET will occur, causing Rhodamine emission at 590 nm instead. Within the same "donor" liposome, FRET will occur, causing quenching of the NBD fluorescence. If the two lipid populations mix, the FRET will decrease due to dilution of the fluorophores, allowing NBD to fluoresce. The NBD fluorescence can therefore be used to assess phospholipid mixing.

6.1.3 Theory behind molecular dynamics simulations

6.1.3.1 General molecular dynamics introduction

Molecular dynamics (MD) simulations are a powerful tool for modelling the behaviour of a molecular system (Kumar et al., 2022). By solving Newton's second law of motion for each particle within a system, MD simulations predict how atoms move and interact over time:

$$\boldsymbol{F}_i(t) = m_i\boldsymbol{a}_i(t)$$

where $\boldsymbol{F}_i(t)$ is the force produced on particle i at time t, m_i is the mass of particle i and $\boldsymbol{a}_i(t)$ is the acceleration of particle i at time t.

The interactions between each particle in a system are modelled by defined forcefields, which describe the potential energy surfaces of an interaction. Interactions are categorised into bonded forces (such as bonds, angles and torsions) and non-bonded forces (such as electrostatic and Van der Waals interactions) (Kumar et al., 2022). Parameters describing these interactions are derived from physics-based experiments or quantum mechanics calculations to represent real-world conditions. MD simulations also use different types of ensembles, where certain physical conditions are kept constant. The most common ensembles are NVT and NPT ensembles, where N is fixed atom number, V is fixed volume, T is fixed temperature and P is fixed pressure. This study used an NPT ensemble for the final (production) simulations, as this represents *in vivo* conditions more accurately than NVT ensembles. Once the potential energy is calculated for a system, the force on each atom can be determined, which can then be used to calculate the acceleration of a particle of known mass. The coordinates and velocity of the particle are then updated using this acceleration.

Solving the equations of motion over time is done by numerically integrating them over small time intervals, allowing the system to evolve step by step. Integration algorithms estimate the

new atomic coordinates, velocities and accelerations for each timestep and keep the force constant (Van Gunsteren and Berendsen, 1990). The integrator used for MD simulations in this study was the GROMACS leap-frog algorithm, which calculates the positions and velocities at alternating time steps.

Simulating the potential energy of the system and integrating Newton's law of motion for each particle over time is computationally expensive, and so numerous strategies are employed to increase the computational efficiency. Enforcing distance cut-offs to only simulate neighbouring non-bonded interactions speeds up the calculations drastically, but the treatment of long-range electrostatic interactions must be accounted for via various algorithms such as the Particle Mesh Ewald (PME) algorithm (Braun et al., 2019). Additionally, the smaller the timestep, the more accurate representation of interactions and collisions but the more computationally expensive. A larger timestep can speed up the calculations, however, too large a timestep can cause atoms to move too far in a single step, resulting in artificially high energies between atoms and system instability (Kumar et al., 2022). Choosing an appropriate timestep is therefore important and is a trade-off between computational expense and simulation accuracy. Additional techniques to speed up the computation are using enhanced sampling algorithms, using specialised hardware and software capable of utilising graphical processors (e.g. GROMACS, utilising high-performance computing clusters), and coarse-graining the simulation.

6.1.3.2 Coarse-grained simulations

Whilst typical all-atom (or atomistic) MD (AAMD) simulations calculate the forces acting on every atom within a system, coarse-grained MD (CGMD) simulations instead group multiple atoms into a single "bead", simplifying the simulation by reducing the degrees of freedom and

therefore decreasing the computational expense (Figure 6.3) (Liwo et al., 2021). AAMD simulations allow accurate interpretation of a simulation at the atomic level whereas CGMD simulations are limited in spatial resolution, but the benefits of simulating larger systems over longer timescales can outweigh this limit. For studying membrane-protein interactions, the ability to simulate a complete phospholipid bilayer and protein is often more beneficial than simulating at atomic resolution.

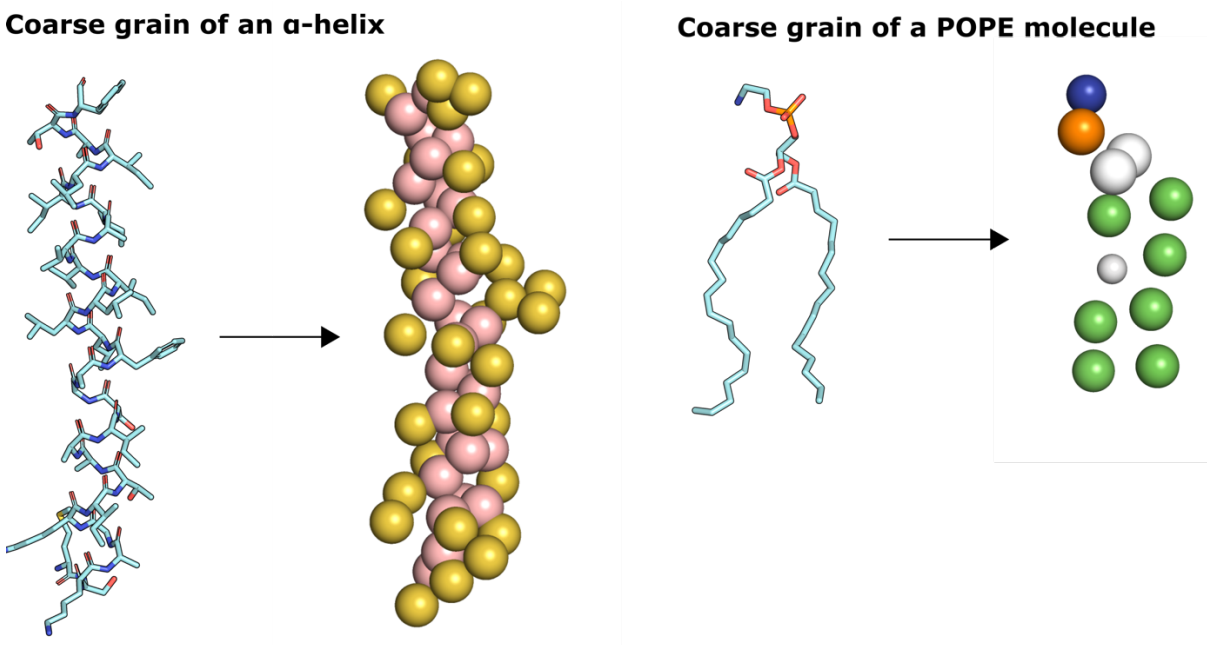


Figure 6.3. Coarse-graining of an α-helix and phospholipid. The N-terminal helix of YhjG and a POPE molecule were subject to coarse-graining utilising the Martini3 CG forcefield (Souza et al., 2021).

With macromolecules like proteins and membranes, a building block principle of coarse graining is often used. Here, coarse grained beads have well-defined parameters that maintain the charge and polarity of the atoms that make them up. The CG forcefield used in this study was the Martini forcefield, a popular forcefield used to study biomolecules (Souza et al., 2021). This groups four atoms into one bead, with well-characterised treatment of amino acids and phospholipids.

Methods of converting the CG representation back to atomic resolution are available, including the CG2AT algorithm used in this study (Vickery and Stansfeld, 2021). Here, each CG bead is treated individually, with the atoms fit into the bead as fragments. These fragments are then minimised, merged and equilibrated, producing a final *de novo* atomic structure. The structure can also be aligned to an input atomic structure (prior to coarse graining) for a final aligned structure. This is useful to identify atomic interactions that would otherwise not be visible at CG resolution, although an AAMD simulation is still preferable for detailed atomic resolution information.

6.1.3.3 Workflow of the CGMD simulations

The CGMD simulations performed in this study followed a standard workflow, shown in Figure 6.4. The AlphaFold model of each protein was oriented in the membrane and converted to a CG representation with Martini3 (Souza et al., 2021). A pre-formed membrane was then built around the orientated protein using Insane at a POPG:POPE ratio of 1:4, a simple representation of the Gram-negative cell envelope IM (Murzyn et al., 2005; Wassenaar et al., 2015).

To represent *in vivo* conditions, water was added to the system and any charge in the system was neutralised with a physiologically relevant concentration of 0.15 M NaCl. The system was then energy minimised using the steepest descent algorithm in GROMACS, which finds the lowest energy minima for the system (Abraham et al., 2015). The system was equilibrated for 20 ns before a 1 μs production simulation was performed. When further inspection of a state was required, CG2AT was used to convert CG representations into atomic representations so that putative atomic interactions could be visualised (Vickery and Stansfeld, 2021).

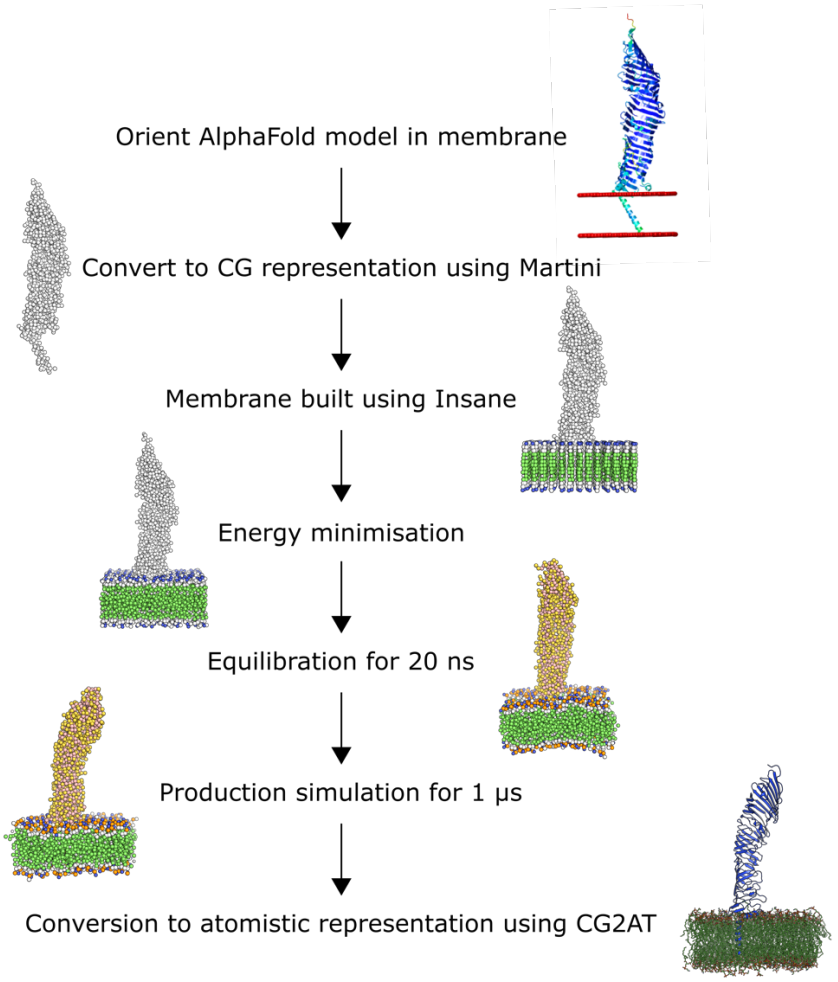


Figure 6.4. General workflow of the coarse-grained molecular dynamic simulations. The AlphaFold model of the protein was oriented in the membrane and converted to coarse-grained (CG) representation with Martini (Souza et al., 2021). A pre-formed membrane was built using Insane, at a POPE:POPG ratio of 1:4, and the whole system was energy minimised and equilibrated for 20 ns using GROMACS (Wassenaar et al., 2015; Abraham et al., 2015). The system was then simulated for 1 µs (also using GROMACS) and any states of interest were converted back to atomistic representation using CG2AT (Vickery and Stansfeld, 2021).

6.1.4 Chapter aims

This chapter aimed to investigate the phospholipid transport capacity of three of the AsmA-like proteins: YhjG, YicH and YdbH. The work in this chapter addressed two main aims:

1. To adapt *in vitro* FRET-based phospholipid transport assays for investigation of the phospholipid transport activity of the AsmA-like proteins.
2. To investigate the mechanisms of phospholipid transport in these proteins using *in silico* coarse-grained molecular dynamics simulations.

6.2 Results

6.2.1 An *in vitro* FRET-based phospholipid transfer assay

6.2.1.1 Phospholipid transfer activity of YhjG

The FRET-based phospholipid transfer assay used in this study was developed using YhjG. This was due to its relative stability and the high yield of protein that could be achieved. The theory of the assay was that YhjG would allow the transfer of phospholipids between liposomes, resulting in a dilution of the two chromophores and reducing the amount of FRET between NBD and Rhodamine. The rate at which phospholipid transport occurs is unknown, although *in vivo* studies on phospholipid transport flux in *E. coli* indicate a rapid flux of phospholipid transport, with the deletion of *yhdP* causing inner membrane shrinkage of 20% in ~20 minutes (Grimm et al., 2020). Due to this, measurements were taken rapidly upon addition of YhjG, anticipating an initial sudden increase in fluorescence, and a time course of >1 hour was used to track further changes in fluorescence.

The assay was performed, in triplicate, with varying concentrations of YhjG, with the data plotted as a percentage of the total NBD fluorescence to show the de-quenching of NBD fluorescence over time (Figure 6.5). The data show that as the concentration of YhjG increased, the higher the percentage total NBD fluorescence that was reached. Interestingly, there seemed to be an initial burst phase of a rapid increase in NBD fluorescence between 0 - 5 minutes before reaching a steady state. This was more noticeable in the higher concentrations of YhjG. This could potentially be due to the insertion of the protein into the liposomes causing some immediate separation of fluorophores and/or an initial rapid movement of phospholipids into the YhjG cavity, before slower transport along the cavity or into acceptor liposomes limits the

rate. It could also be an effect of the detergent, although as the burst phase is not seen in the buffer control and appears to be concentration-dependent, this explanation is unlikely.

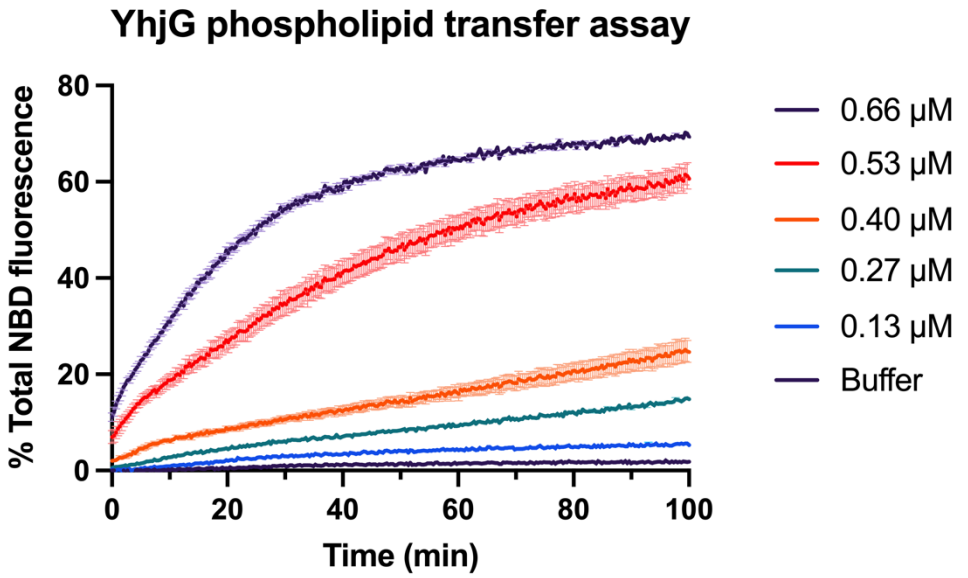


Figure 6.5. Phospholipid transport FRET-based assay of YhjG. Time course assay of phospholipid transport via YhjG, with addition of buffer as the negative control. The assay was performed in triplicate, with the mean ± S.E.M. shown.

To evaluate rates of phospholipid transport between each concentration, initial rates were calculated from 7 - 14 minutes, which was after the initial burst phase and prior to the plateau in the higher concentrations tested. The initial rates are shown in Table 6.1 along with the maximum fluorescence reached for each concentration. The highest concentration tested of 0.66 μM YhjG showed a high initial rate of NBD fluorescence increase of 1.62 ± 0.07% total NBD fluorescence min^{-1} which then reached a plateau of ~70% total NBD fluorescence after ~1 hour. This plateau likely represents a state of saturation, where the fluorescent lipids become maximally distributed across the liposomes. However, the plateau does not reach 100% total NBD fluorescence, indicating that there must remain some level of FRET occurring. This could be due to the equilibrated fluorophores still being within FRET distance in each liposome, as compared to fully separated within DDM micelles.

Table 6.1. Initial rates of YhjG phospholipid transport. Initial rate measured as a percentage of the total NBD fluorescence per minute. The gradient of the linear curve from 7 - 14 mins was used to calculate rate, with the standard error shown. Maximum percentage of the total NBD fluorescence is also reported with the S.E.M. shown. Data reported to 2 d.p.

Concentration of YhjG (μM)	Initial rate (% Total NBD fluorescence min^{-1})	Maximum fluorescence (% Total NBD fluorescence)
0	0.04 ± 0.01	1.80 ± 0.09
0.13	0.09 ± 0.01	5.47 ± 0.22
0.27	0.23 ± 0.01	14.66 ± 0.24
0.40	0.26 ± 0.04	24.71 ± 2.11
0.53	0.87 ± 0.11	60.85 ± 2.24
0.66	1.62 ± 0.07	69.60 ± 0.14

The lower concentrations of YhjG exhibited a linear increase in fluorescence, continuing steadily towards the endpoint of the assay and failing to reach a plateau. The equivalent addition of buffer showed stable baseline fluorescence throughout the duration of the measurement. This result was very promising, indicating that the addition of YhjG was indeed causing an increase in NBD fluorescence, that the effect was concentration-dependent, and demonstrated that this assay could be used to monitor phospholipid transport activity. The data provided the first evidence of YhjG functioning as a phospholipid transporter and is the first biochemical evidence of phospholipid transport for any of the AsmA-like proteins.

6.2.1.2 Phospholipid transfer activity of YicH

The phospholipid transfer activity of YicH was then tested at the same concentrations as YhjG (Figure 6.6). The increase in NBD fluorescence lessened compared to YhjG, with slower rates and a lower maximum fluorescence reached. None of the concentrations tested for YicH reached a plateau of NBD fluorescence, indicating that saturation of the system was not achieved. A higher concentration of YicH or longer time may be needed to reach saturation. Alternatively, this could indicate that YicH was less efficient at depositing lipids into an

acceptor liposome and was instead stalling the lipids within the cavity of YicH, though with fluorescence still increasing at the end of the assay, this is unlikely. Higher concentrations of YicH could not be tested here due to the need to concentrate the protein sample, which had the adverse effect of concentrating the detergent present in the buffer. When concentrated proteins were tested in this system, an immediate popping of liposomes was seen due to the higher detergent concentration, likely above the CMC, resulting in 100% NBD fluorescence immediately and an inability to follow rates of phospholipid transfer.

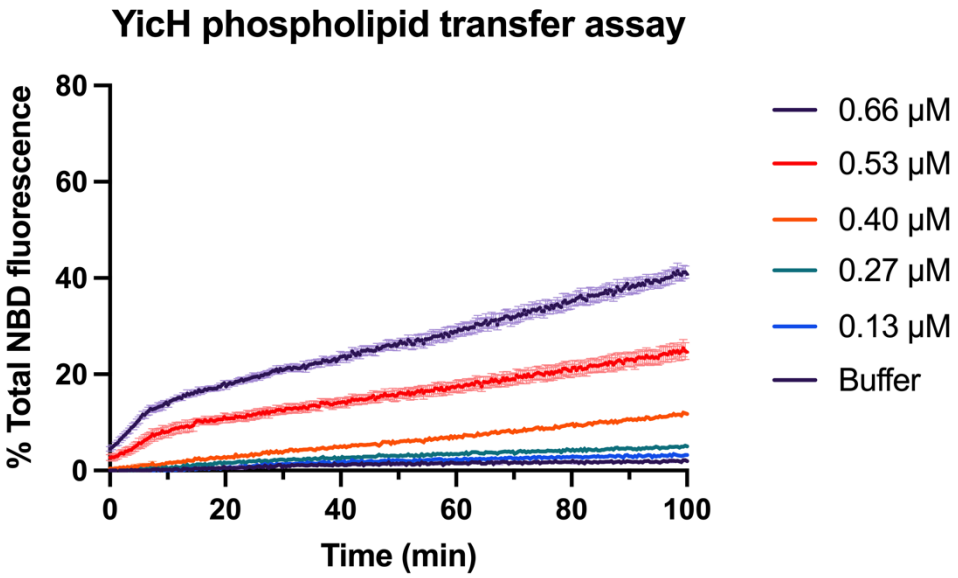


Figure 6.6. Phospholipid transport FRET-based assay of YicH. Time course assay of phospholipid transport via YicH, with addition of buffer as the negative control. The assay was performed in triplicate, with the mean ± S.E.M. shown.

Nevertheless, a rise in the rate of phospholipid transport can be seen as the concentration of YicH increases. Again, there appears to be two distinct linear rates over the time course of the assay: the burst phase rate (0-6 mins) and a slower steady rate (>6 mins). Again, the addition of YicH may allow immediate transport of phospholipids to its cavity, giving rise to the higher burst rate, but then subsequent transport throughout the cavity or into acceptor liposomes is slower, and limits further separation of lipids. At 0.66 µM YicH, a maximum percentage total

NBD fluorescence was measured at ~40%, with the likelihood of further increasing over time. Initial rates were calculated using the same time interval as YhjG to allow direct comparison and are shown in Table 6.2 along with the maximum fluorescence reached. The initial rates for YicH were lower than the respective YhjG concentrations but remained within the same range, with a maximum initial rate of 0.50 ± 0.05% total NBD fluorescence min^{-1} at 0.66 μM YicH. The data indicate that YicH is also capable of transporting phospholipids, but at a slower rate than YhjG, and failing to reach saturation.

Table 6.2. Initial rates of YicH phospholipid transport. Measured as a percentage of the total NBD fluorescence increase per minute. The gradient of the linear curve from 7 - 14 min was used to calculate rate, with the standard error shown. Maximum percentage of the total NBD fluorescence is also reported with the S.E.M. shown. Data reported to 2 d.p.

Concentration of YicH (μM)	Initial rate (% Total NBD fluorescence min^{-1})	Maximum fluorescence (% Total NBD fluorescence)
0	0.02 ± 0.02	1.86 ± 0.28
0.13	0.05 ± 0.02	3.01 ± 0.11
0.27	0.08 ± 0.01	4.86 ± 0.11
0.40	0.14 ± 0.02	11.45 ± 0.15
0.53	0.31 ± 0.09	24.88 ± 1.51
0.66	0.50 ± 0.05	41.05 ± 1.23

6.2.1.3 Phospholipid transfer activity of YdbH

Preliminary data for the phospholipid transport capabilities of YdbH using this assay were also obtained (Figure 6.7). Due to issues in sample concentration and protein stability during the assay set-up, samples above a final concentration of 0.13 μM could not be tested. Additionally, the YdbH assay was only able to be performed in duplicate due to limited protein sample. Nevertheless, the preliminary data reported here indicates that YdbH is also capable of transporting phospholipids, even at the low concentration of 0.13 μM, where a plateau of ~50% total NBD fluorescence was achieved after ~50 minutes (Figure 6.7). The initial rate of

phospholipid transport at 0.13 µM YdbH was calculated to be 1.30 ± 0.08% total NBD fluorescence min^{-1}, again using the same timepoints as YhjG (Table 6.3). This indicates that YdbH transports phospholipids at a faster rate than both YhjG and YicH at the equivalent concentrations, which had initial rates of 0.09 ± 0.01% total NBD fluorescence min^{-1} and 0.05 ± 0.02% total NBD fluorescence min^{-1}, respectively. There was also less indication of a burst phase for YdbH than YhjG and YicH. If the burst phase is indeed caused by the rapid entry of phospholipids into the hydrophobic cavity, then it could be possible that YdbH is more efficient at transporting the phospholipids through its cavity and into acceptor liposomes than YhjG and YicH. This could mean the initial rate for YdbH should instead be calculated at an earlier time interval, but to allow direct comparison between the three proteins, the same timepoints were used here. It also may be that a higher concentration of YdbH is needed to see this burst phase, as the burst phase was more noticeable in the higher concentrations of YhjG and YicH.

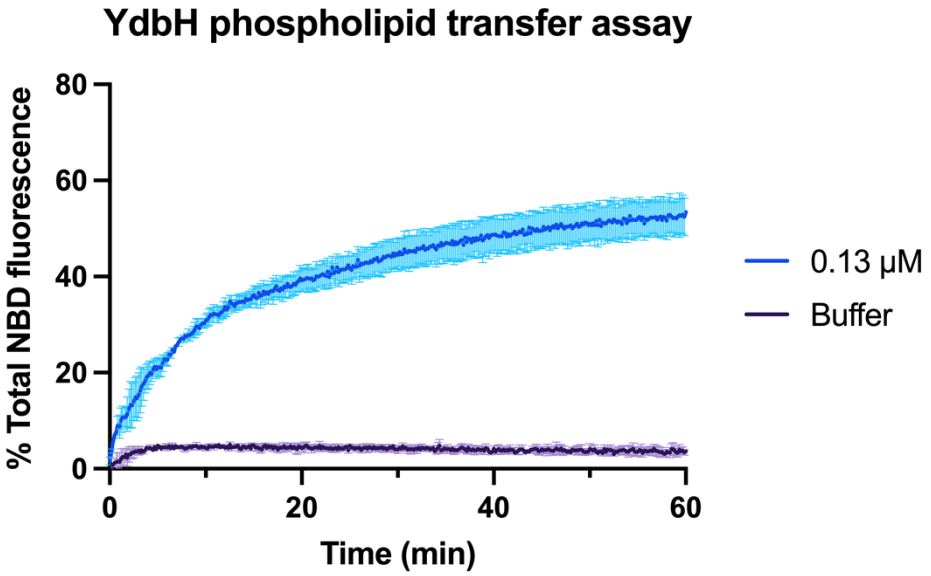


Figure 6.7. Phospholipid transport FRET-based assay of YdbH. Time course assay of phospholipid transport via YdbH, with addition of buffer as the negative control. The assay was performed in duplicate, with the mean ± S.E.M. shown.

Table 6.3. Initial rates of YdbH phospholipid transport. Measured as a percentage of the total NBD fluorescence increase per minute. The gradient of the linear curve from 7 - 14 min was used to calculate rate, with the standard error shown. Maximum percentage of the total NBD fluorescence is also reported with the S.E.M. shown. Data reported to 2 d.p.

Concentration of YdbH (μM)	Initial rate (% Total NBD fluorescence min^{-1})	Maximum fluorescence (% Total NBD fluorescence)
0	0.01 ± 0.03	3.60 ± 1.14
0.13	1.30 ± 0.08	53.43 ± 3.87

These assays demonstrate, for the first time, that YdbH, YhjG and YicH exhibit phospholipid transport activity. With no energy source available in the system, this transport must also be diffusive. The observed differences in transport rates and total fluorescence recovery between the proteins suggest varying transport efficiency, providing valuable insights into their functional roles in the cell.

6.2.2 Molecular dynamics simulations

With evidence of diffusive phospholipid transport activity in YhjG, YicH and YdbH via the FRET-based lipid transfer assay, further investigation was required to determine regions of importance within the protein for lipid transport and to elucidate potential mechanisms of action. CGMD simulations to recreate the proteins in a phospholipid bilayer were used to investigate this transport, with the aim of identifying spontaneous phospholipid movement into and throughout the cavity of these proteins.

6.2.2.1 Coarse-grained MD simulation of YhjG in bilayer

The AlphaFold model of YhjG was embedded into a phospholipid bilayer via its N-terminal helix. This was predicted to be between residues 7-29, based on the transmembrane helix location predictor DeepTMHMM program (Hallgren et al., 2022). The same N-terminal helix was predicted for all AsmA-like proteins. However, it must be noted that the Orientation of Proteins in Membranes (OPM) server did not predict this region but rather the C-terminal helix-

turn-helix (residues 641-666) (Figure 6.8A) (Lomize et al., 2022). This region was also predicted to interact with the membrane when the MemSat server was used (membrane interaction: residues 8-28 and residues 644-661) (Jones, 2007). Whilst this is encouraging in terms of our theory that the C-terminal of YhjG may interact with the OM in the cell, it was not useful for the intention of simulating lipid transport at the IM, and so manual placement of the N-terminal helix into the membrane was performed.

The simulated system is shown in Figure 6.8B. The phospholipids remained in a single bilayer and the protein remained embedded via its N-terminal helix, with the periplasmic domain flexing side-to-side but mostly remaining perpendicular to the bilayer plane. Proteins are dynamic in solution, but the flexibility seen here is particularly encouraging due to the evidence of YhjG being a conformationally dynamic protein from prior EM studies (Chapter 5).

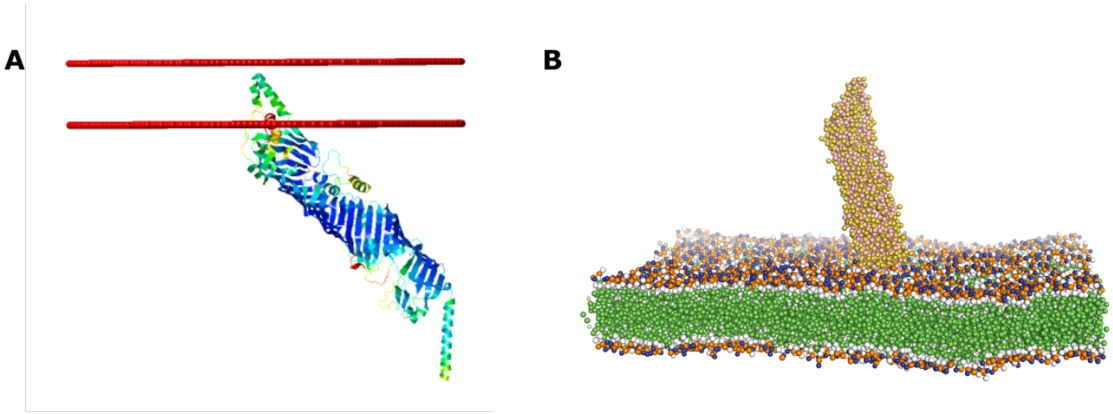


Figure 6.8. YhjG insertion into simulated phospholipid bilayer. A – The result of the OPM server preferentially inserts YhjG into a bilayer via its C-terminal helix-turn-helix. OPM colouring uses the AlphaFold pLDDT scores. B – The coarse-grained equilibrated system showing YhjG inserted into a phospholipid bilayer of 1:4 POPG:POPE via its N-terminal helix. The protein is shown in yellow and pink spheres, with the lipids shown in green spheres (acyl chains) and white, blue and orange spheres (head groups). Figures made with OPM and PyMOL 2.4.1 (Lomize et al., 2022; Schrodinger, 2015).

Throughout the 1 μs simulation, spontaneous movement of phospholipids from the upper leaflet of the bilayer into the cavity entrance was observed multiple times (Figure 6.9). Many

phospholipids attempted this entry, with most diffusing back into the bilayer. In particular, only POPE molecules were seen entering the cavity. This could imply that POPE is the preferred or only substrate of YhjG, however, with 4x as many POPE than POPG in the simulated bilayer, this was likely an effect of the higher probability of POPE diffusing near the YhjG cavity rather than POPG.

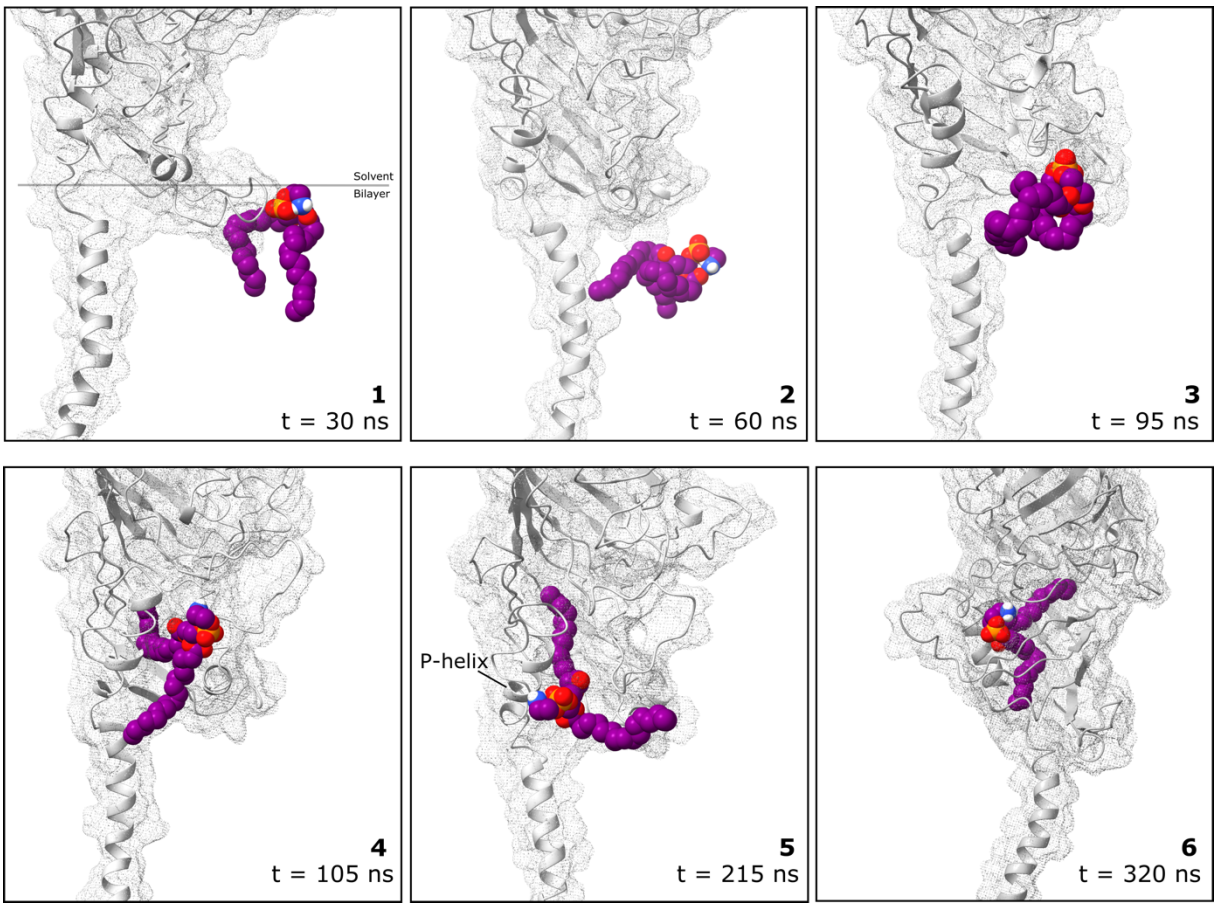


Figure 6.9. Snapshots of POPE moving into the hydrophobic cavity of YhjG. YhjG was simulated using CGMD in a POPG:POPE 1:4 phospholipid bilayer for 1 µs, and phospholipids were seen entering the hydrophobic cavity. States of interest depicting 6 key steps of POPE entry were converted from CG representation to atomistic representation using CG2AT (Vickery and Stansfeld, 2021). Steps 1-3 show the POPE nearing the cavity and tilting its acyl chains into the cavity. Steps 4-6 show the splaying of the POPE acyl chains as one tail finds entry into the cavity. Step 6 depicts the highest position that the POPE reaches within the cavity. YhjG depicted as a cartoon and dot surface coloured grey. POPE depicted as purple spheres, with coloured headgroups. Figures made in ChimeraX-1.7.1.

Figure 6.9 shows snapshots of the simulation that were converted from a coarse-grain representation to an atomistic representation using the *de novo* conversion method from the CG2AT program, focussing on one POPE molecule which spontaneously moved into the hydrophobic cavity. Whilst only one phospholipid molecule is depicted for clarity, it is important to note that there were many phospholipids packing around the N-terminal helix and the POPE molecule shown here.

The POPE molecule diffuses near the hydrophobic cavity of YhjG (Figure 6.9, step 1) and begins to tilt such that the acyl chains are parallel to the bilayer plane (Figure 6.9, step 2). Various conformations are then adopted where the acyl chains point towards the cavity, whilst the headgroup remains solvent exposed (Figure 6.9, step 3). The POPE lipid tails then adopt a splayed conformation, with one tail starting to feed into the hydrophobic cavity (Figure 6.9, step 4). The headgroup finds its way into a solvent exposed pocket. As the simulation was coarse grained, identification of what allows this movement is ambiguous, however, the atomistic snapshots indicate that the β3-β4 and β4-β5 loops may play an important role (discussed later). Figure 6.9, step 5 depicts one of the various conformations that the POPE molecule then takes once it has entered the cavity. The headgroup moved towards the small helix parallel to the bilayer, labelled 'P-helix', still solvent exposed. It is important to note that in the simulation, multiple phospholipids at this point returned to the bilayer, and when they didn't, other phospholipids began to follow the phospholipid into the cavity. The conformation observed in Figure 6.9, step 6 represents the furthest any single phospholipid molecule travelled within the YhjG cavity, with the headgroup again solvent exposed above the flexible loop.

Further investigation into the interactions between the β3-β4 and β4-β5 loops and POPE to determine what may have allowed the movement of the phospholipid further into the cavity between steps 3 and 4 and the highest POPE conformation adopted in step 6 was performed,

with details shown in Figure 6.10. In step 3, the β3-β4 helix/loop (in blue) sits adjacent to and forms H-bonds with the POPE headgroup. The β4-β5 loop (in orange) also interacts with the POPE molecule, particularly between K154 and the headgroup. In step 4, the β4-β5 loop changes conformation dramatically. The lipid headgroup sits above the β3-β4 loop, with various new interactions formed, including H-bonds with E44, K113, N149 and R161. The POPE conformation in step 6 then shows how further movement of these loops is likely required for the headgroup to progress up the cavity. The surface representation shows how the headgroup now sits above the β4-β5 loop, in a solvent exposed gap. New interactions are again made between the POPE molecule and residues within these loops and the hydrophobic cavity. The headgroup makes a direct interaction with E244 in the β8-β9 loop (in pink), which appears to be blocking the progression of the phospholipid and may have to undergo a similar conformational change to allow further phospholipid movement.

These flexible loops likely act as gates to allow or prevent progress of the phospholipid through the hydrophobic cavity. The polar residues in these loops often interact with the lipid headgroups which are positioned to remain solvent accessible, whilst the hydrophobic interior protects the lipid tails from solvent exposure.

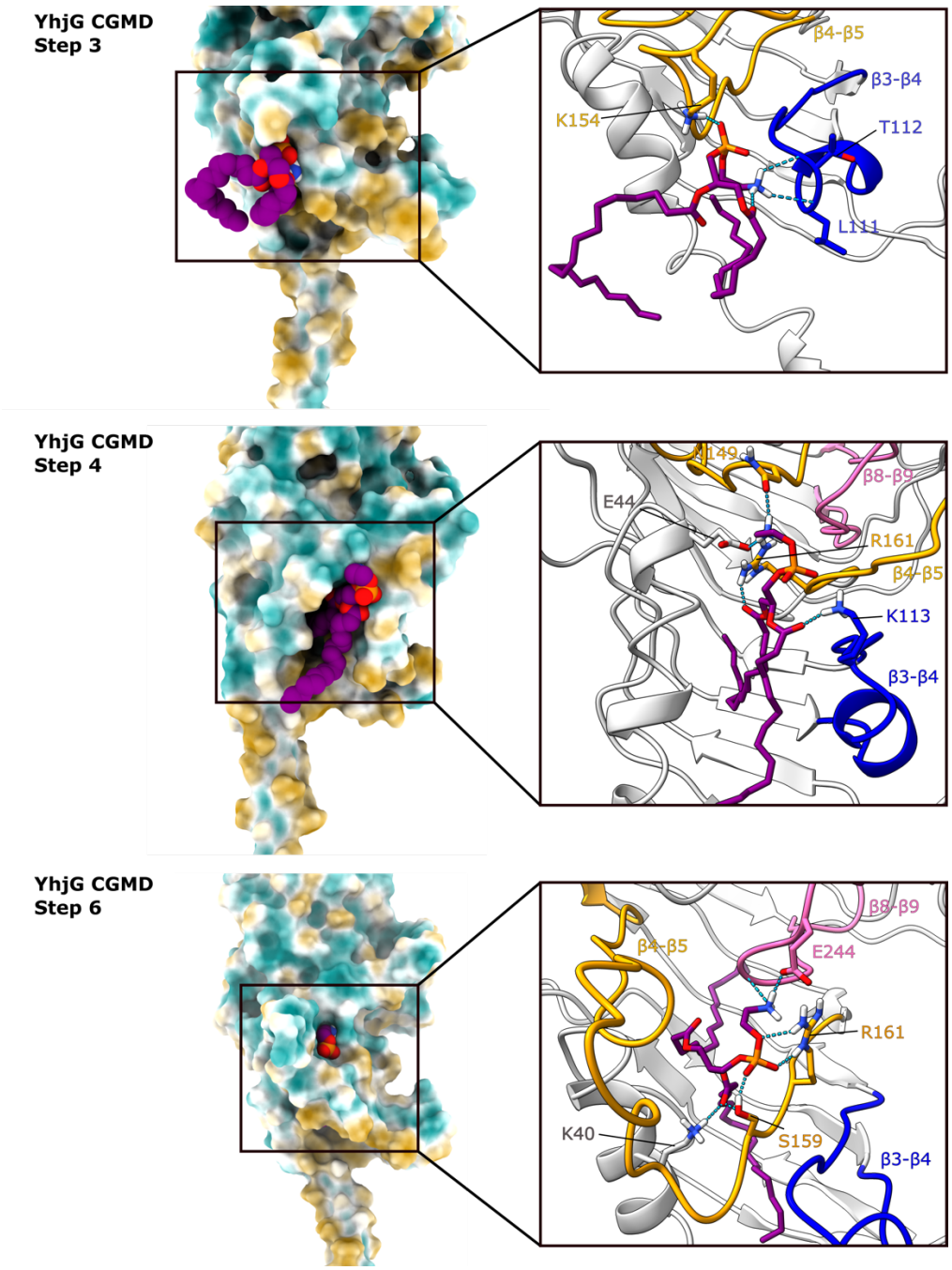


Figure 6.10. Close up of interactions in steps 3, 4 and 6 of the YhjG CGMD simulation. Step 3 shows the POPE molecule before entry into the hydrophobic cavity. Step 4 shows the tail-first entry of the POPE molecule, adopting a splayed conformation. Step 6 shows the conformation at which POPE had travelled the furthest into the cavity. The headgroup remains solvent accessible in all steps. Surface representation coloured by hydrophobicity (from hydrophilic in blue to hydrophobic in yellow). Cartoon representation in grey, with the β3-β4 loop/helix in blue, the β4-β5 loop in orange and the β8-β9 loop in pink. The POPE molecule is depicted as purple spheres or sticks, with headgroups coloured. Side chains of nearby residues are shown as sticks and labelled. Figures made in ChimeraX-1.7.1.

6.2.2.2 Coarse-grained MD simulation of YicH in bilayer

CGMD simulations were performed using the AlphaFold model of YicH due to its phospholipid transport activity in the FRET-based assay. The OPM server predicted an orientation of YicH into the membrane via its N-terminal helix (Figure 6.11A), and so this was used to place YicH into a preformed bilayer of 1:4 POPG:POPE. The resulting system at various points throughout the simulation is shown in Figure 6.11B. Flexing of the CG YicH model was seen throughout the simulation, with varying degrees of curvature adopted. Like YhjG, this was encouraging after evidence of protein flexibility was seen during negative stain and cryo-EM investigations (Chapter 5).

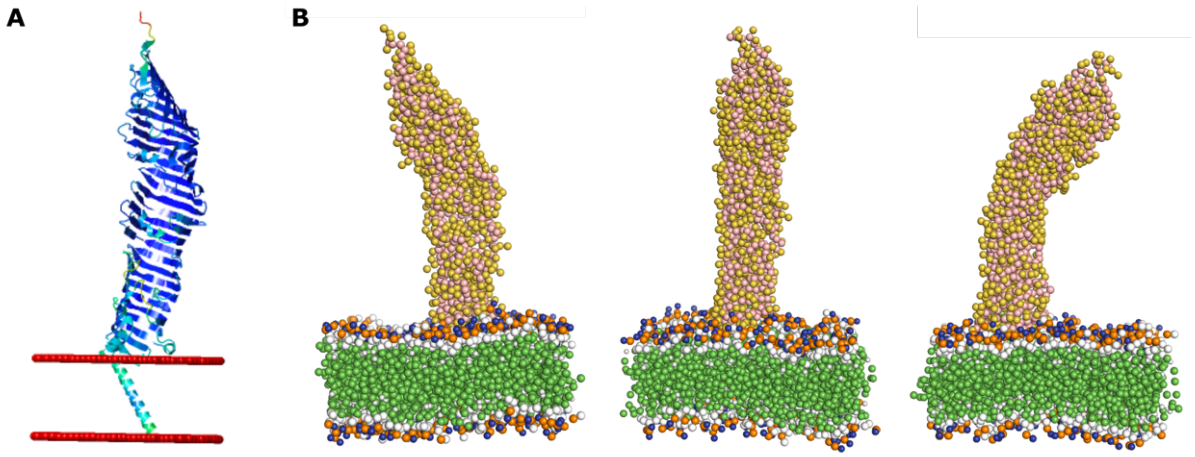


Figure 6.11. YicH insertion into simulated phospholipid bilayer. A – The result of the OPM server preferentially inserts YicH into a bilayer via its N-terminal helix. B – The coarse-grained equilibrated system showing YicH inserted into the bilayer via its N-terminal helix at three timepoints throughout the simulation to demonstrate the flexibility of the protein. YicH is shown in yellow and pink spheres, with the lipids shown in green spheres (acyl chains) and white, blue and orange spheres (head groups). Figures made with OPM and PyMOL 2.4.1 (Lomize et al., 2022; Schrodinger, 2015).

Like YhjG, multiple phospholipids from the upper leaflet of the bilayer spontaneously entered the hydrophobic cavity of YicH via a splayed conformation of the phospholipid tails (Figure 6.12). This tail-first entry was very similar to the phospholipid entry seen in YhjG, which is not

surprising considering the high structural similarity, particularly in the N-terminal Chorein-N like region. The phospholipid tilted parallel to the membrane plane, before a splayed conformation allowed the entry of one acyl chain into the cavity. Figure 6.12 steps 1-3 show the atomistic representations of a POPE molecule (in purple) that spontaneously entered the YicH hydrophobic cavity following this mechanism.

In this CGMD simulation of YicH, the highlighted POPE molecule then failed to make any further progress up the hydrophobic cavity, potentially being prevented by the β4-β5 loop. During the simulation, the POPE molecule was eventually replaced by the entry of a POPG molecule (steps 4-6 in Figure 6.12). At first, the POPG molecule (in turquoise) packed tightly against the POPE molecule near the entrance to the cavity (step 4), eventually adopting a conformation where both acyl chains were further into the cavity than POPE, having pushed the single acyl chain of the POPE molecule out (step 5). The POPG molecule then proceeded to adopt a splayed conformation, with one tail reaching the furthest position into the cavity of any of the phospholipids across the simulation (step 6). The headgroup remained in the same position as the POPE molecule in step 3, with the β4-β5 loop potentially preventing further progression. The close packing of the POPE tail between the two POPG tails appeared to allow a wider splayed conformation.

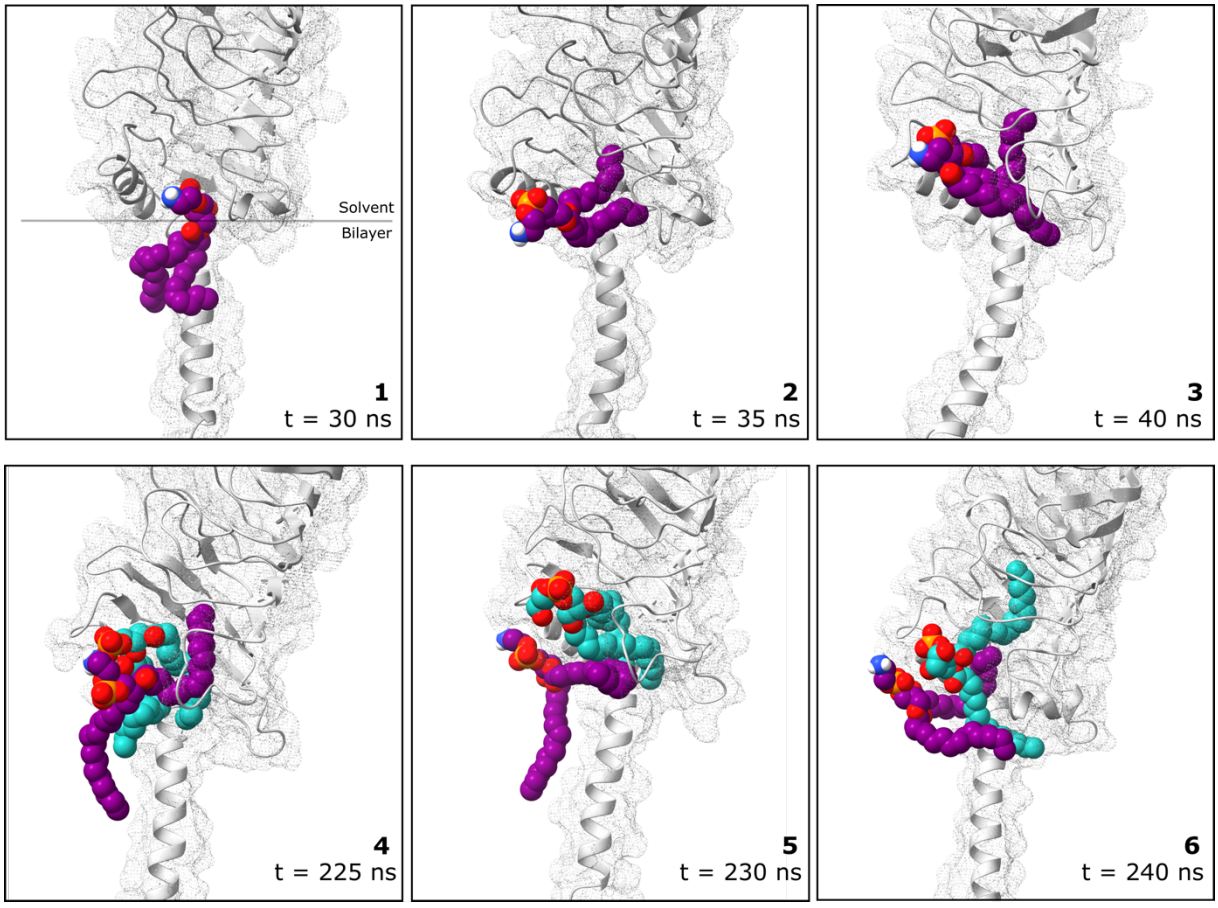


Figure 6.12. Snapshots of a POPE molecule and a POPG molecule moving into the hydrophobic cavity of YicH. YicH was simulated using CGMD in a POPG:POPE 1:4 phospholipid bilayer for 1 μs, and phospholipids were seen entering the hydrophobic cavity. States of interest depicting 6 key steps of POPE and POPG entry were converted from CG representation to atomistic representation using CG2AT (Vickery and Stansfeld, 2021). Steps 1-3 show the POPE molecule nearing the cavity entrance, tilting its acyl chains and adopting a splayed conformation for tail-first entry into the cavity. Steps 4-6 show the replacement of this POPE molecule with a POPG molecule, where the acyl chains closely pack together. The POPG molecule progresses further up the cavity than the POPE molecule, likely due to the further splaying of the POPG tails by the POPE tail in step 6. YicH is depicted as a cartoon and dot surface and is coloured in grey. POPE and POPG are depicted as purple and turquoise spheres, respectively, with coloured headgroups. Figures made in ChimeraX-1.7.1.

More detailed inspection of the positioning of POPE and POPG in steps 3 and 6 was then performed to determine what was preventing further transport, shown in Figure 6.13. Interactions between either POPE and POPG headgroups were seen with S42 in the small P-helix-β1 loop and R95 in the β3-β4 loop (in blue). These interactions may be trapping the lipid

from further progression up the cavity. The large β4-β5 loop (in orange) does not appear to make any direct interactions with either POPE or POPG in these steps, unlike the equivalent step in the YhjG CGMD (Figure 6.10, step 3) where a lysine interacts with the POPE headgroup. This loop is also shorter in YicH than YhjG and has many nonpolar residues, e.g. L116, F118. The large conformational change of the β4-β5 loop that was seemingly needed for phospholipid progression in the YhjG CGMD simulations may be restricted because of this, preventing further phospholipid movement.

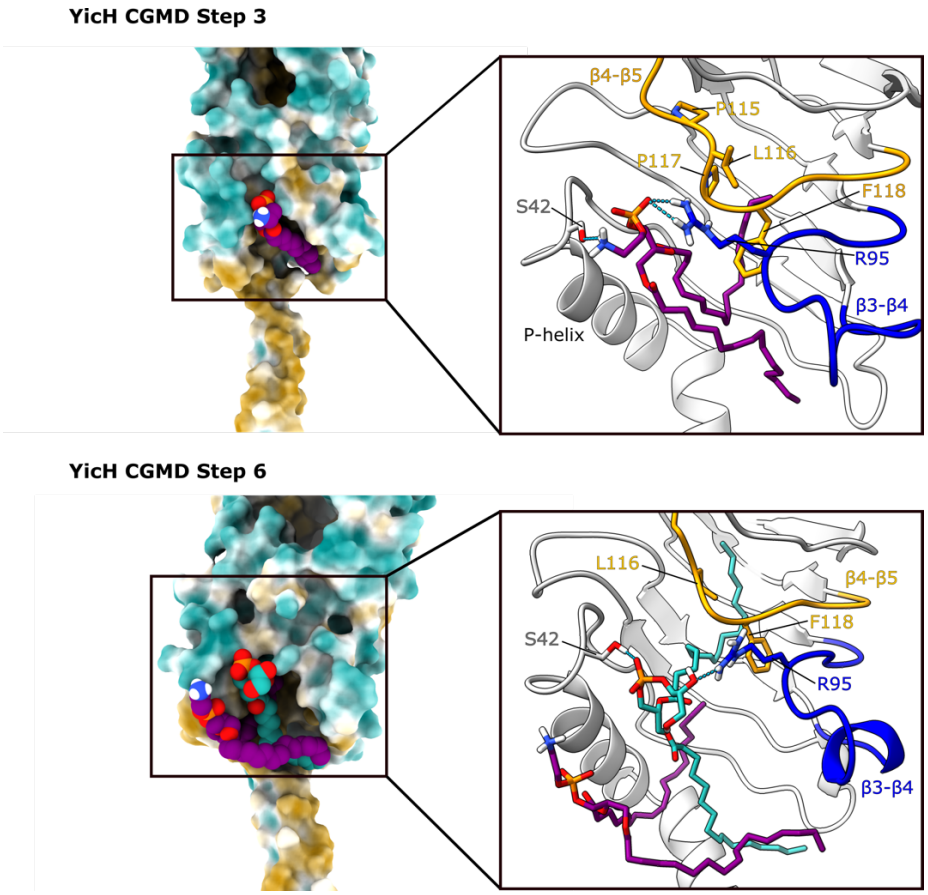


Figure 6.13. Close up of interactions in steps 3 and 6 of the YicH CGMD simulation. Step 3 shows the tail-first entry of the POPE molecule, adopting a splayed conformation. Step 6 shows the replacement of the POPE molecule by a POPG molecule. The headgroups remains solvent accessible in all steps. Surface representation coloured by hydrophobicity (from hydrophilic in blue to hydrophobic in yellow). Cartoon representation of YicH in grey, with the β3-β4 loop/helix in blue and the β4-β5 loop in orange. The POPE and POPG molecules are depicted as purple or turquoise spheres/sticks, respectively, with coloured head groups. Nearby residues are shown as sticks and labelled. Figures made in ChimeraX-1.7.1.

6.2.2.3 Coarse-grained MD simulation of YdbH in bilayer

CGMD simulations of YdbH in a phospholipid bilayer were also performed based on phospholipid transport evidence through both the preliminary FRET-based assay data presented in this study and from recent literature findings showing a genetic functional redundancy with the putative phospholipid transporter YhdP (Ruiz et al., 2021; Douglass et al., 2022). Both CGMD simulations with the AlphaFold model of YdbH and the AlphaFold-multimer of the putative YdbH complex with YnbE and YdbL were performed (Kumar et al., 2024). As movement of phospholipids was observed in only the N-terminal region of YdbH in both simulations, the addition of YnbE and YdbL did not alter any of the general observations found with YdbH alone and so for clarity, only the CGMD simulation of YdbH alone will be discussed. The same 1:4 POPG:POPE bilayer was used for YdbH CGMD simulations, with an insertion into the bilayer via the N-terminal helix based on the results of the OPM server, shown in Figure 6.14A (Lomize et al., 2022). The CG system is shown in Figure 6.14B. Like YhjG and YicH, YdbH exhibited flexibility throughout the 1 μs simulation, with large radial motion seen from a hinge point about one third of the way up the protein from the membrane (indicated with a black line in Figure 6.14B). The C-terminal tail of YdbH is very dynamic in the simulation, however, the confidence in the AlphaFold prediction for this region is low and it is putatively involved in complex formation with YnbE, so any simulated movement of this tail is likely not physiologically relevant.

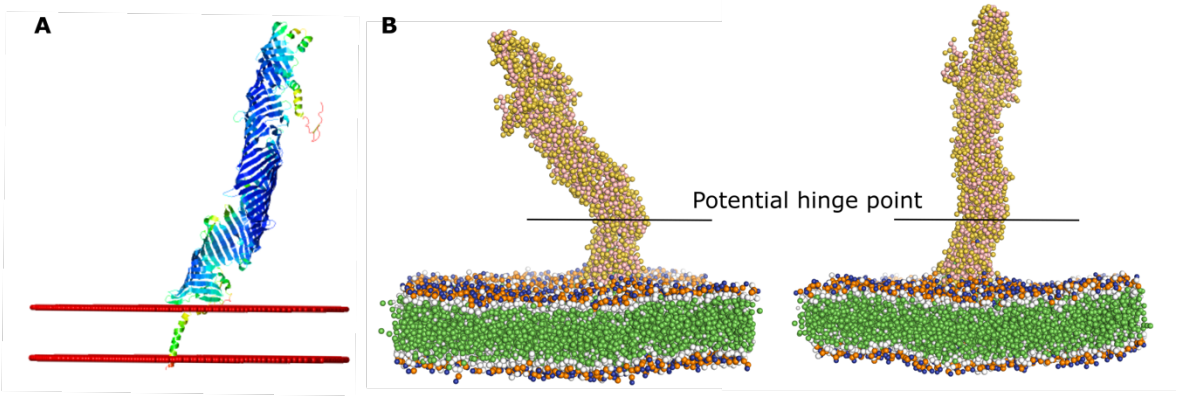


Figure 6.14. YdbH insertion into simulated phospholipid bilayer. A – The result of the OPM server preferentially inserts YdbH into a bilayer via its N-terminal helix. B – The coarse-grained equilibrated system at two timepoints showing YdbH inserted into the bilayer via its N-terminal helix and that YdbH was flexible, particularly from a potential hinge point (black line). The protein is shown in yellow and pink spheres, with the lipids shown in green spheres (acyl chains) and white, blue and orange spheres (head groups). Figures made with OPM and PyMOL 2.4.1 (Lomize et al., 2022; Schrodinger, 2015).

Multiple phospholipids from the upper leaflet of the bilayer spontaneously entered the hydrophobic cavity of YdbH via a splayed conformation of the phospholipid tails, much like YhjG and YicH (Figure 6.15). The lipids progressed further up the hydrophobic cavity of YdbH in the simulation compared to YhjG and YicH, however. Atomistic snapshots of a POPE molecule that entered the cavity, closely followed by other phospholipid species, are shown in Figure 6.15. Step 1 shows the POPE (purple) near the YdbH cavity entrance. In step 2, the POPE acyl chain has further entered the cavity via the familiar splayed tail conformation. Additional POPE (maroon) and POPG (turquoise) lipids packed in tightly with the POPE molecule at the entrance to the cavity. As in the previous simulations, the presence of the acyl chains of these lipids appears to allow the wide splaying conformation of the purple POPE. Step 3 shows the POPE molecule then able to tilt both acyl chains up into the hydrophobic cavity, with the aid of the POPG tails packed in below. Multiple phospholipids then exchanged with the bilayer for this space, with step 4 showing another POPE molecule (maroon) in the place of

the POPG molecule. The POPE in step 4 progressed further, with the headgroup now buried within the hydrophobic cavity. Step 5 then shows the acyl chains again splayed, with one tail rising further in the cavity and the headgroup remaining buried. The final step shown depicts the highest positioning of the POPE molecule seen during the simulation, with a similar splayed tail conformation. The headgroup was situated within the cavity again but seems to be solvent accessible through a small pocket, with an opening just above the small helix immediately after β4 (residues 96-101, labelled in step 6).

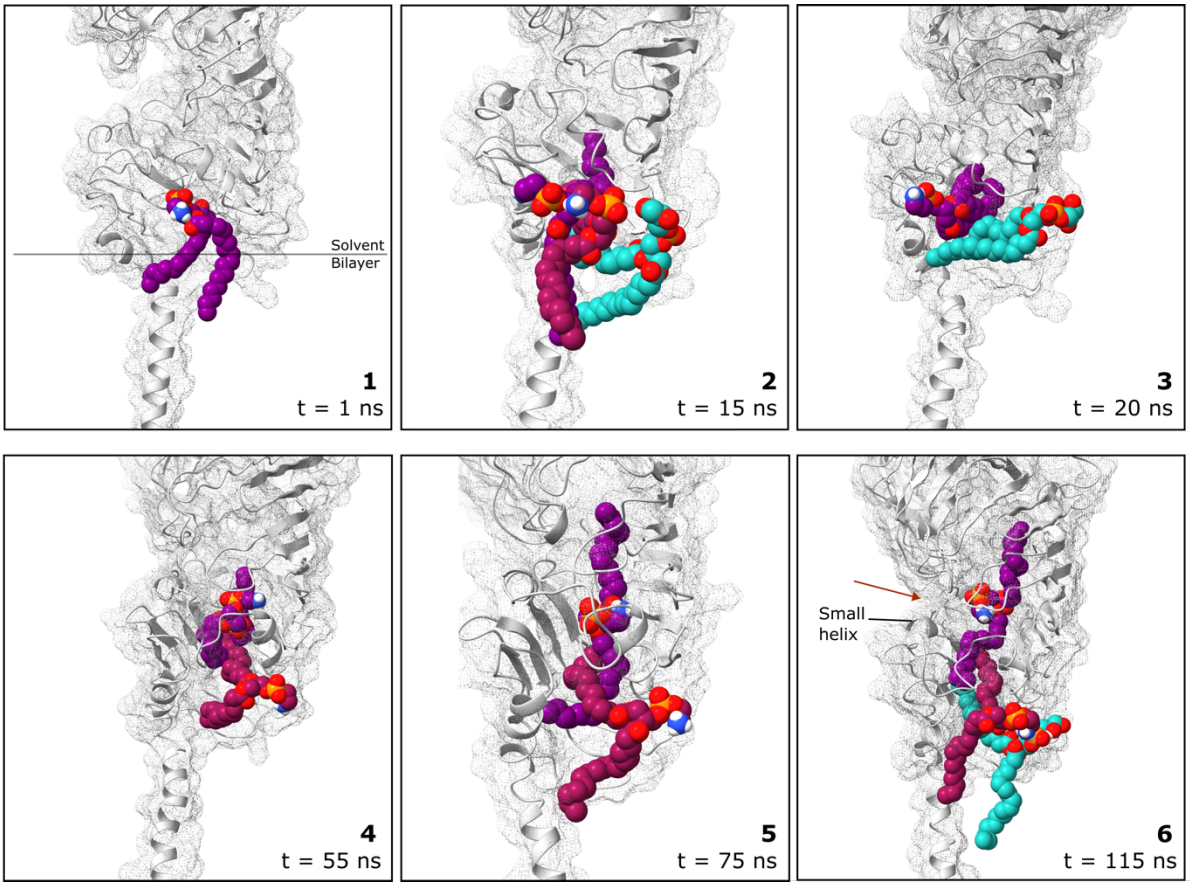


Figure 6.15. Snapshots of POPE and POPG molecules moving into the hydrophobic cavity of YdbH. YdbH was simulated using CGMD in a POPG:POPE 1:4 phospholipid bilayer for 1 µs, and phospholipids were seen entering the hydrophobic cavity. States of interest depicting 6 key steps of POPE (purple/maroon) and POPG (turquoise) entry were converted from CG representation to atomistic representation using CG2AT (Vickery and Stansfeld, 2021). Steps 1 shows the POPE molecule nearing the cavity entrance, step 2 shows the POPE tilting its acyl chains and adopting a splayed conformation with the help of an additional POPE and POPG. Step 3 shows the tilting of both acyl chains of POPE upward in the cavity, with lipid tail packing from below the POPG. Step 4 shows the replacement of the lower POPG lipid with another POPE, and further progression of the POPE, which now is completely buried in the cavity. Steps 5 and 6 show the POPE in a splayed conformation, making further progress up the cavity with additional POPE and POPG molecules following closely behind. Channel depicted with red arrow above the labelled small helix in step 6. YdbH is depicted as a cartoon and dot surface and is coloured in grey. POPE and POPG are depicted as purple and turquoise spheres, respectively, with coloured headgroups. Figures made in ChimeraX-1.7.1.

To investigate what allows phospholipid transport to progress further in the YdbH CGMD simulations compared to the YhjG and YicH simulations, investigation into the potential interactions formed between the phospholipid headgroups and YdbH at steps 3, 4 and 6 was

performed, shown in Figure 6.16. The β3-β4 loop (blue) seems to play a less important role for YdbH than YhjG, only forming interactions with lipids that are still mostly embedded in the bilayer. Movement of the β4-β5 loop (orange) again appears to play a role in allowing phospholipid transport, with the POPE headgroup situated below the loop in step 3 and above the loop in step 4. No direct interactions with the headgroups can be seen from residues in the loops in these steps, so instead, nearby residues are highlighted to depict the local environment (Figure 6.16). The lack of a polar interface for the POPE headgroup in step 4 is surprising and suggests the existence of intermediate states where this ionic headgroup can exist within the hydrophobic cavity.

Step 6 shows the furthest positioning of the POPE molecule within the cavity, with the tip of the highest tail spanning up to the 11th β-strand. This also seems to be where the flexible motion seen throughout the simulation originates. Hinging of the protein at this point can be more clearly seen in the zoomed-out surface representation of Figure 6.16 step 6, with the structure curving by ~45^{o} from the axis normal to the bilayer plane. Additionally, the burying of the anionic headgroup within the hydrophobic cavity may, in this step, be stabilised by a small pocket (Figure 6.16, step 6, red arrow) that is formed above the small helix in the β4-β5 loop (orange) with D201 from the β10-β11 loop (green) at its entrance. A hydrogen bond may also form between the backbone of D201 and the POPE headgroup, which will help maintain the hydrophilic head in this environment. The β10-β11 loop may again be acting as a gate to allow phospholipid access. Another POPE and POPG molecule follow this lipid in at the entrance to the cavity, packing in the hydrophobic space and following similar tail-first mechanisms of entry to the previously described phospholipids.

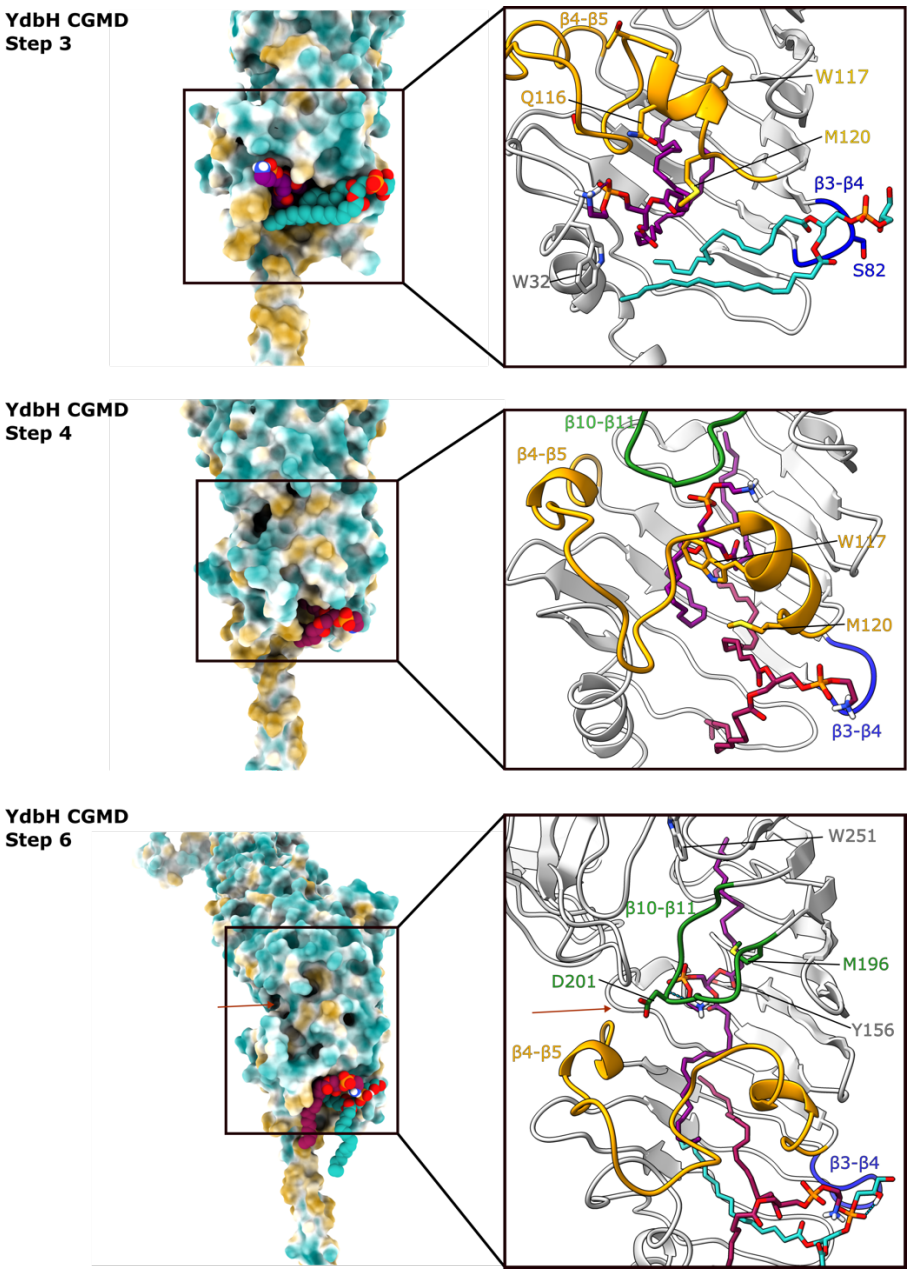


Figure 6.16. Close up of interactions in steps 3, 4 and 6 of the YdbH CGMD simulation. Step 3 shows the POPE molecule after entry into the hydrophobic cavity of YdbH, with the headgroup situated below β4-β5. Step 4 shows the POPE molecule after conformational change of this loop so that the headgroup now sits above it. Step 6 shows the highest positioning of the POPE lipid within the cavity, with the β10-β11 loop interacting with the headgroup. Red arrow indicates channel opening. Surface representation coloured by hydrophobicity (from hydrophilic in blue to hydrophobic in yellow). Cartoon representation of YdbH in grey, with the β3-β4 loop in blue, the β4-β5 loop in orange and the β10-β11 loop in green. The POPE and POPG molecules are depicted as purple/maroon or turquoise spheres/sticks, respectively, with coloured head groups. Side chains of nearby residues are shown as sticks, with residues of importance labelled. Figures made in ChimeraX-1.7.1.

6.3 Discussion

6.3.1 Biochemical evidence of diffusive phospholipid transport

The FRET-based phospholipid transport assay data presented here is the first biochemical evidence of phospholipid transport of any of the AsmA-like proteins. The increase in NBD fluorescence upon addition of any of the three proteins tested here, YicH, YhjG and YdbH, indicates a phospholipid mixing event between the "donor" and "acceptor" liposome populations. The lack of increase when the equivalent amount of buffer to control for any detergent-mediated effects further solidifies that the proteins are responsible for the phospholipid mixing. Furthermore, this transport was diffusive, as there was no energy source, such as ATP or an induced proton motive force gradient, utilised in this assay.

Differences in rates between the three proteins tested here reveal potentially interesting information about their respective phospholipid transport functions. YdbH had the highest initial rate of phospholipid transport and appeared to reach saturation, even at the low concentration of 0.13 μM YdbH. At the equivalent concentration, YhjG and YicH had much lower transport rates and failed to reach a plateau. Furthermore, even at the highest concentration of 0.66 μM tested, YicH was unable to reach saturation, whereas YhjG then showed a similar response to 0.13 μM YdbH. From this data, it can be interpreted that YdbH exhibits the fastest rate of phospholipid transport, with YicH exhibiting the slowest rate of phospholipid transport.

One possible hypothesis for the ability of YdbH and YhjG to reach a plateau and not YicH is the presence of the C-terminal helix-turn-helix that is present in both YhjG and YdbH, but not YicH (Figure 6.17). This hydrophobic region may interact with the membrane, as was suggested for YhjG when using the OPM server to determine MD simulation protein orientation. Perhaps

then, YdbH and YhjG can more effectively deposit the phospholipids into the acceptor liposomes because of a stronger anchor, whereas YicH does not form a stable interaction and stalls the phospholipids in its cavity. Removing the C-terminal helix-turn-helix from YhjG and YdbH and repeating the assay would be the next step in investigating this hypothesis. If the region were important in transporting phospholipids between the liposome populations, then there should be a reduction in the initial rate and maximal fluorescence observed.

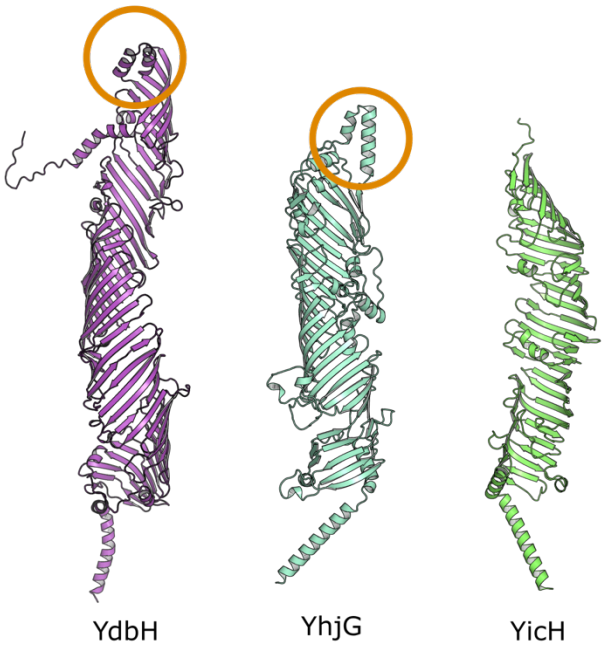


Figure 6.17. AlphaFold structures of YdbH, YhjG and YicH, highlighting a key structural difference. YdbH and YhjG are predicted to have a hydrophobic C-terminal helix-turn-helix motif (in orange circles) which may interact with membranes, whereas YicH does not.

Further mutations could also be made to potentially disrupt phospholipid transport function, especially considering the novel findings from the MD simulations. Removal of putatively important loops at the N-terminal entrance to the hydrophobic cavity, mutation of headgroup interacting residues, and mutation of hydrophobic residues within the cavity to hydrophilic residues, as had slightly abolished function of YhdP and moreso TamB in *in vivo* complementation studies (Chapter 3), could all further elucidate residues of importance using

the direct output of phospholipid transport via this assay. Identifying non-functional mutants that abolish transport would also strengthen the validity of the assay in measuring phospholipid transport activity of other membrane proteins.

6.3.2 Similar FRET-based liposome phospholipid transfer assays

Similar assays have been reported for multiple other LTPs, including shuttle LTPs such as MlaC, STARD4 and PITPs, and BLTPs such as E-Syt1, the ERMES complex and Atg2 (Kawano et al., 2018; Panagabko et al., 2019; Osawa et al., 2019; Zhang et al., 2022; Wotherspoon et al., 2024; Wang et al., 2024b). These assays often use a different mixture of phospholipids to replicate the membrane environments of the donor and acceptor membranes for the various proteins. They all have in common the use of NBD- and Rhodamine-labelled phospholipids to measure the degree of lipid mixing. Particularly relevant to this study, however, were the assays performed on the eukaryotic BLTPs.

The assay developed in this study was based primarily on the assay for characterising the phospholipid transport activity of the BLTP extended synaptotagmin 1 (E-Syt1), a protein that transports lipids from the ER to the plasma membrane in eukaryotes (Bian and De Camilli, 2019). The main difference in assay setup being the ability to use detergent-solubilised membrane proteins and reconstituting proteoliposomes in this assay, as compared to the previously reported assay which only applies to soluble proteins with membrane tethers, as opposed to membrane anchors. In the E-Syt1 assay, the authors reported that the total NBD fluorescence initially increased rapidly before reaching a plateau after ~10 minutes when E-Syt1 and a necessary Ca^{2+} ion were present in the assay mixture, similar to the response of 0.13 μM YdbH and 0.66 μM YhjG. However, without the Ca^{2+}, NBD increase follows a slow linear rate, similar to the curves seen for the lower concentrations of YhjG and for YicH. They reason that the lower rate implies that basal lipid transfer still occurs by chance, but that the Ca^{2+}

enhances the lipid transfer by allowing tethering to the acceptor liposome, which they back up with other methods (Bian et al., 2018). This aligns with the theory that YicH may not be able to tether to another liposome, and that the slow rate of NBD fluorescence increase could represent basal lipid transfer through chance interaction of the C-terminal region with a liposome. The C-terminal regions of YdbH and YhjG may stabilise this chance interaction, allowing the transfer of more phospholipids and achieving saturation more rapidly.

6.3.3 Limitations of the FRET-based phospholipid transport assay

The main limitation to the phospholipid transport assay presented here is that the directionality of phospholipid transport is impossible to determine. With the addition of protein-detergent micelles into the SUV population, there is an equal chance of the protein inserting into the donor and acceptor liposome. With the assay set up to detect a decrease in FRET between NBD and Rhodamine in the donor liposomes, this doesn't affect the ability to track phospholipid transport, as whichever direction phospholipids flow, there will be a mixing of the lipids and a separation of the two fluorophores.

Similar assays performed on membrane proteins have mixed reconstituted proteoliposomes of a donor liposome with one component of a phospholipid transport pathway (i.e. MlaA) and an acceptor liposome with another component of the pathway (i.e. MlaFEDB), allowing directionality to be assessed (Tang et al., 2021). However, with the presented evidence showing that at least some of the AsmA-like proteins are likely embedding into liposomes at both the N-terminal and C-terminal regions, this assay system cannot be used. It may be possible to cleave the protein into two, with the N-terminal and C-terminal halves reconstituted into separate liposomes. However, this would likely expose many hydrophobic residues within the cavity leading to instability and relies on the successful binding of the two halves in solution. Alternatively, a fluorescent tag capable of FRET with a fluorescent lipid could be engineered

into the protein. If this tag were at the C-terminal region for example, then movement of a fluorescent lipid up the hydrophobic cavity may position the lipid fluorophore within FRET distance of the conjugated fluorophore. An increase in FRET would then indicate phospholipid transport, but issues with directionality would still arise if the C-terminal region can embed into a liposome and retrograde transport is possible. Development of either alternative assay system suggested would require much further investigation but would strengthen the evidence presented in this study.

Another limitation of the assay system used here is the choice of phospholipids within the liposomes. Whilst PC is a good mimic of the bacterial membrane and forms stable liposomes, it is worth noting that bacteria do not contain PC (Carey et al., 2022). It is possible that the presence of PC interferes with the putative lipid transport function of YhjG, YicH and YdbH, and could overestimate or underestimate physiological phospholipid transport rates depending on its effect. There is also the addition of large fluorescent probes on the headgroups of the PE molecules within the assay, which may also interfere with any normal phospholipid transport function of the proteins.

It should be possible to perform the assay with liposomes prepared from *E. coli* polar lipids instead of PC, which would more accurately model the bacterial inner membrane. The use of *E. coli* polar lipids in a similar system has recently been reported and could be used to design future follow-up experiments (Tang et al., 2021; Wotherspoon et al., 2024). Furthermore, tail-labelled NBD-PE could be trialled to test whether the location of the fluorescent probe affects phospholipid transport, which could also reveal information about the substrate specificity for the headgroup vs the tails.

One final caveat to note is the inability to confidently claim that the increase in NBD fluorescence seen is from phospholipids shuttling between liposomes via the protein, as the phospholipid mixing event could in theory be caused by liposome fusion (Zhang et al., 2022). Further experiments to confirm that the results are not due to liposome fusion should be performed. With the investigation of non-functional mutants, this caveat can be easily addressed, as a single-point mutant would likely not affect the ability of the protein to fuse two liposomes, should this occur. Additionally, a membrane protein with a known function that is not phospholipid transport should also be tested, as any detergent-mediated effects or liposome fusion events would likely also be present here. Therefore, a priority for validating this assay further would be to test mutants and perform the membrane protein negative control to eliminate liposome fusion events from being responsible for the increase in NBD fluorescence.

6.3.4 Spontaneous phospholipid entry into the hydrophobic cavities of YhjG, YicH and YdbH is seen in CGMD simulations

The CGMD simulation data presented in this study provides evidence to support the putative role of phospholipid transport for YhjG, YicH and YdbH. The tail-first entry of phospholipids into the hydrophobic cavity was seen multiple times throughout the CGMD simulations for each protein, resulting in a splayed tail conformation of the lipid. This mechanism of phospholipid entry has been reported in MD simulations of other LTPs, including YhdP, the phospholipid exchange between MlaD and MlaC, and a fragment of Atg2 (Cooper et al., 2023; Wotherspoon et al., 2024; Wang et al., 2024b).

After initial phospholipid entry, various residues within each protein, particularly in the flexible loops between β strands, made direct interactions with the lipid headgroups. In YhjG and YicH, the β3-β4 loop appears to be important in the acceptance of the lipid, whereas in YdbH, this loop is smaller and doesn't make as many direct interactions. The β4-β5 loop appears to be the

equivalent of this loop in YdbH. Further progression of the phospholipids in YhjG and YdbH identified direct interactions between the lipid headgroup and the β4-β5 loop in YhjG and the β10-β11 loop in YdbH. The acyl chains remain protected in the hydrophobic cavity throughout the simulation, whilst the headgroup stays mostly solvent exposed.

It is particularly interesting to note that all three proteins form a protected hydrophobic space between the acyl chains of the lipids and the entrance to the hydrophobic cavity due to the embedding of the protein in the bilayer. Phospholipid tails fill this space, with many tilting sideways within the bilayer between the N-terminal helix and the cavity opening. This deep embedding of the protein within the bilayer provides a path for the hydrophobic tails of the phospholipid to enter the cavity, as other phospholipid headgroups are excluded from the space. Also, multiple phospholipid tails can exist within the cavity entrance, with further phospholipid progression through the cavity seemingly relying on the presence of the following phospholipids. This was seen most clearly in the last step investigated in the YdbH simulation.

A common finding between the simulations was that the flexible loops appeared to be acting as gates for phospholipid transport. Significant conformational changes in these loops were seen between states of phospholipid progression up the cavity. This finding is reminiscent of shuttle LTPs that form box-like cavities with lids, where a conformational shift of the lid is required for lipid exchange (Ryan et al., 2007). The association of the LTP with the membrane causes the α-helix to turn its hydrophobic face to the membrane, which in turn opens the cavity, allowing lipids to enter. This lid-like movement could be happening with the loops in the AsmA-like proteins, firstly to allow phospholipid entry, and then to allow phospholipid movement progressively up the cavity.

It would also be interesting to see if the restriction of phospholipid movement to the N-terminal region is due to the lack of an OM. It may be possible that an interaction at the C-terminal region is needed to provide stability within the structure and/or to provide some sort of mechanical force that encourages or allows conformational changes within the loops, allowing phospholipids to progress. However, the recent MD simulations on YhdP also found a restriction of phospholipids to the N-terminal region, even with both an IM and OM present (Cooper et al., 2023). Nevertheless, including an OM in the simulation should still be investigated in future studies, if only to assess the ability of the proteins to embed within the OM. Alternatively, it may be that the loop conformational change is a rare event and the timeframe of CGMD simulations is simply not long enough to see full phospholipid transport.

6.3.5 Limitations of the MD simulations

Whilst the CGMD simulations show exciting spontaneous phospholipid entry in all three proteins studied here, it is important to remember that the initial protein structures used were AlphaFold structural predictions, and not experimental structural data. This could mean that the atomic coordinates may be different from the true structure and would have an impact on the modelled interaction with the phospholipid bilayer. It is worth noting though that the regions of lowest confidence in the AlphaFold models of each of YhjG, YicH and YdbH are the flexible loops and C-terminal regions. Experimental structural data would also likely produce ambiguous density for these flexible regions, as we have observed in the cryo-EM of YhjG (Chapter 5). Therefore, even when an existing experimental structure is available, MD simulations often must be performed with assumptions about such flexible regions. Nevertheless, an experimental, high-resolution structure would be the ideal starting point for MD simulations, especially as density for a lipid/substrate may be present within the structure.

Similarly, the interpretation of the atomistic conversions of the CG snapshots revealed potentially important residues that make interactions with the lipid headgroups and revealed that flexible loops between β-strands may be acting as a gate. However, these atomistic representations were built into the CG bead model through fragment-based reconstruction, and whilst highly accurate, the reconstruction is not without error (Vickery and Stansfeld, 2021). It could model residues with an incorrect side chain orientation for example, which would greatly alter any protein-lipid interactions and our interpretation of the data. Exact location of residues therefore is not confidently known from these simulations. AAMD simulations would solve this issue but at the cost of computational expense. Unfortunately, the time constraints of this project prevented AAMD simulations from being performed with the available computing power. Confirming the protein-lipid interactions discussed for each protein in this chapter through AAMD would greatly increase the confidence in the conclusions made here.

Another limitation of the MD simulations performed in this study that should be noted is the choice of phospholipids. The 1:4 ratio of POPG:POPE is a simplistic representation of the Gram-negative inner membrane. Physiological composition of the IM is around 70-80% PE, 15-20% PG, and 5-10% CL (Yasuhiro et al., 1967). Whilst the main components are represented in the simulations used in this study, the absence of CL and varying lengths and saturations of acyl chains could affect the behaviour of the bilayer in the presence of these proteins. It would also be interesting to test whether CL can enter the hydrophobic cavities of the AsmA-like proteins during the simulation, as there are four acyl chains to account for instead of the two found in PG and PE. There is also some evidence of phospholipid asymmetry within the IM of Gram-negative bacteria, with different distributions depending on cell shape (Bogdanov et al., 2020). It was found that PE predominantly resides in the periplasmic leaflet of the IM in filamentous *E. coli* cells whereas normal rod-shaped *E. coli* cells show that the majority of PE

resides within the cytoplasmic-facing leaflet. This putative asymmetry was not accounted for within the bilayer model in these simulations due to maintaining model simplicity and the lack of further evidence for IM lipid asymmetry.

6.3.6 Conclusion

The FRET-based phospholipid transfer assay and CGMD simulations presented here provide the first biochemical evidence of diffusive phospholipid transport for YdbH, YhjG and YicH. This is also the first biochemical evidence of phospholipid transport for any of the AsmA-like proteins. The data suggest that YdbH transports phospholipids at a higher rate than YhjG and YicH, with YicH potentially not as capable of depositing phospholipids into an acceptor liposome, though this is very much still speculation. YdbH also saw the furthest progression of phospholipids into its cavity during MD simulations. Whilst the research presented here is only preliminary, much of the data and conclusions made can be used to influence the design of follow-up experiments to further validate the assay and to confirm the potential protein-lipid interactions suggested in the CGMD simulations. The priorities for further research should therefore be to design and test the phospholipid transport capabilities of a series of non-functional mutants and to perform AAMD simulations, ideally including an OM and using an experimentally derived input structure.

CHAPTER 7. CONCLUDING REMARKS

7.1 Summary of findings

The broad aim of this study was to further characterise the roles of the AsmA-like superfamily in *E. coli* using a genetic, biochemical and structural approach. This protein family have only recently been implicated in the fundamental process of GPL trafficking to the OM of the cell envelope, and hence, many questions about their exact role and how they may transport phospholipids are yet to be addressed.

The phenotypic profiling screen presented in Chapter 3 represents, to the best of our knowledge, the most varied set of conditions that the phenotypes of both the single and double knockouts of the *asmA*-like genes have been assessed on. The severe fitness defects seen in membrane stressing conditions for Δ*yhdP* and the double knockout of Δ*yhdP*Δ*tamB* support the current supposition of an important and redundant role for the two largest AsmA-like proteins in OM biogenesis and/or maintenance (Ruiz et al., 2021; Douglass et al., 2022). Furthermore, we reveal that loss of *asmA* partially recovers the fitness defect seen for Δ*yhdP* mutants in the presence of the OM stressors SDS and EDTA, potentially due to an indirect result of the lower LPS levels seen in Δ*asmA* mutants or perhaps indicating that AsmA has an opposing function to YhdP (Deng and Misra, 1996). We also present a novel gain-of-fitness phenotype for the Δ*yhdP*Δ*tamB* mutant in the presence of the MreB inhibitor A22, which is likely a result from the slower OM biogenesis minimising the lethal consequence of an impaired peptidoglycan cell wall.

Chapter 4 then summarises the expression and purification trials that were performed to produce pure, stable protein for further functional and structural study. Despite trialling different expression and purification conditions and performing thorough solubilisation screens,

purification of YhdP, TamB and AsmA was unsuccessful, often producing a low yield of impure, unstable protein. We were able to purify YdbH, YhjG and YicH to a sufficient quantity and quality for further study, though further work is needed to stabilise YdbH in solution, likely through co-expression of its YnbE partner protein. In general, this protein family were difficult to work with *in vitro*, and it is likely due to this that very few structural and functional studies have been reported on these proteins so far (Josts et al., 2017; Cooper et al., 2023; Wang et al., 2024a).

We then report the first cryo-EM maps of two of the AsmA-like proteins, YicH and YhjG, in Chapter 5. The experimental structural data show a general agreement with the AlphaFold predicted structures, with a hollow cavity throughout the periplasmic regions of both proteins, likely capable of accommodating phospholipids. Intrinsic flexibility within the proteins limited the resolution that could be achieved, particularly in the N-terminal and C-terminal regions of each protein. This flexibility is likely integral to the function of the protein, and further strategies to limit this would be needed to improve the resolution. Nevertheless, the thorough optimisation of grid preparation conditions and particle picking strategies that were performed to obtain these EM maps can be used to influence future cryo-EM experiments of all six of the AsmA-like proteins.

Finally, we report the first phospholipid transfer assay designed to be used with bridge-like membrane proteins in Chapter 6. Utilising this assay, we provide evidence of diffusive phospholipid transport for YhjG, YicH and YdbH, which is the first reported biochemical evidence of phospholipid transport for any of the AsmA-like proteins. This assay can be utilised for investigation of the phospholipid transport capabilities of YhdP, TamB and AsmA, and other membrane-bound phospholipid transport proteins. We then investigate a mechanism of transport using MD simulations, in which we identify spontaneous tail-first entry of

phospholipids into the hydrophobic cavities of YdbH, YhjG and YicH. Progression of phospholipids is limited to the N-terminal regions of each protein, with flexible loops that span across the cavity opening between β-strands spaced throughout the length of the structure seemingly important in gating the movement of phospholipids.

7.2 Redundancy of phospholipid transport

The assay and CGMD simulations presented in Chapter 6 provide biochemical and *in silico* evidence for a phospholipid transport function for three of the AsmA-like proteins, YdbH, YhjG and YicH. The questions that remain are why this is the case, if/when these proteins are utilised within the cell, and why there is such an apparent redundancy for this function? If they are needed for fundamental GPL transport to the OM, then why are they not essential, based on the genetics studies (Chapter 3)? Our genetic evidence and other published studies propose that YhdP and TamB are the most utilised GPL transporters with YdbH also capable of *in vivo* GPL transport but seemingly less utilised, as a triple knockout of these genes was synthetically lethal but only Δ*yhdP*Δ*tamB* knockouts show the most severe fitness defects (Ruiz et al., 2021; Douglass et al., 2022). If YhjG and YicH are capable of transporting phospholipids *in vitro*, then why do they not allow this knockout to be viable by providing further alternative GPL transport routes in the cell? It could be possible that they are evolutionary relics, either with a historic function that is no longer needed or a result of gene duplication and subsequent divergence/truncation of the other AsmA-like genes (Bolotin and Hershberg, 2016). Alternatively, perhaps these smaller proteins are involved in upkeep of the peptidoglycan somehow or have a completely different undetermined function, with their ability to bind and transport phospholipids *in vitro* and *in silico* not relevant to their physiological function.

As previously discussed in Chapter 5, YhjG and YicH are also too short to span the normally 200-250 Å wide periplasm, and it could be possible that these proteins are involved in GPL trafficking only at sites of closer contact between the membranes. Based on the presented and previously reported data, we hypothesise that whilst the larger YhdP, TamB and YdbH are the main transporters of GPLs from the IM to the OM, the smaller AsmA-like paralogues are only utilised when the periplasm is likely smaller, either through external environmental pressures or at different stages of the cell cycle. Much further investigation to confirm this will be needed.

7.3 Priorities for future study

The genetics studies supported the evidence of *yhdP* and *tamB* having the most essential role in the cell, but also revealed many new phenotypes and interactions that require further investigation, including determining the link between YhdP, TamB and the MreBCD elongasome complex, and the genetic interaction between *yhdP* and *asmA*. Further conditions should be tested to find a more obvious loss- or gain-of-fitness phenotype for the other *asmA*-like genes, which will potentially reveal a particular environmental condition in which the smaller *asmA* paralogues are utilised under and will also allow the determination of important residues through phenotypic complementation assays.

With regards to protein purification, the priority for future study is to continue the trialling of different purification tags and conditions, and to design soluble constructs. Use of alternative tags may improve the purify and limit any protein breakdown occurring by metalloproteases, which are not inhibited by the EDTA-free protease inhibitors used in the Ni-NTA purifications presented in this study. Use of the novel ML-based algorithms described in Chapter 4 may aid soluble construct design which would be useful for further structural studies. Assessment of soluble constructs in the *in vitro* assay may also provide insight into the importance of the

membrane tether for protein function, as the removal of the TM helix did not abolish YhdP function in the *in vivo* phenotypic complementation assays.

For improved structural determination of YhjG and YicH, a higher number of particles will likely be needed to allow use of tailored algorithms designed for the processing of intrinsically flexible structures with multiple conformations. Alternatively, a step-back to protein design may be required to limit structural flexibility and/or increase the particle size for improved particle averaging. The recommendations presented in this study will also be useful for the structural determination of the larger AsmA-like proteins, which will likely be more flexible and pose even greater problems than YhjG and YicH in data processing.

The next steps with regards to the FRET-based phospholipid transport assay would be to further validate its use with non-functional mutants and a non-phospholipid transporter membrane protein control. The determination of important residues or regions (e.g. the C-terminal region) will also provide detailed insight into the mechanism of phospholipid transport via these proteins. Ultimately, the priority is to design an assay that can also elucidate directionality of transport. Further studies using MD simulations should also test non-functional mutants of these proteins and should assess whether the addition of the OM is needed for further phospholipid progression up the cavity. Ideally, an experimentally determined input structure and atomistic MD simulations would be used.

Whilst there is growing evidence of the role for the AsmA-like proteins in fundamental GPL transport, there is a clear redundancy between the proteins in the family and potentially with other protein complexes within the cell, e.g. Pqi and Let. With only the largest family members having a combined essential function in the cell, questions remain about when or if the smaller AsmA-like proteins are used. Therefore, a priority of future study should be the determination

of expression levels of the *asmA*-like genes, to determine why and when each of the genes are being utilised. Furthermore, investigation into whether there are any interactions or links with the other putative phospholipid transport pathways in *E. coli* is needed, as it is unclear why the cell needs so many pathways to perform this transport, if indeed that is their function.

There also needs to be further investigation regarding the role of TamB. With evidence supporting both a role in phospholipid transport, and a role in OMP transport and folding, determining if these functions are physiologically relevant is needed (Selkrig et al., 2012; Ruiz et al., 2021; Wang et al., 2024a). We hypothesise that TamB is capable of transporting both substrates, acting as a generic transporter of hydrophobic molecules to the OM. The regulation of its use in the cell would need to be investigated if this were the case.

Finally, IM or OM binding partners are yet to be identified for YhdP, AsmA, YhjG and YicH. With TamB binding the OMP TamA, and YdbH binding the OM lipoprotein YnbE and the chaperone YdbL, it is not implausible to hypothesise that the other AsmA-like proteins have IM/OM binding partners (Shen et al., 2014; Kumar et al., 2024). However, there is very little evidence for the identity of possible interactors. The structurally homologous BLTP ATG2 is known to interact with the lipid scramblases ATG9, VMP1 and TMEM41 (Ghanbarpour et al., 2021). Two of these share a domain seen in the DedA superfamily (Morita et al., 2018). With the recent evidence of the first bacterial phospholipid flippase being a DedA family member, this superfamily has been suggested to be IM binding partners of the AsmA-like proteins in bacteria (Roney and Rudner, 2023). The MD simulations revealed that it was only phospholipids from the periplasmic-facing leaflet that gained entry into the cavity, and so it is logical to conclude that transbilayer movement of phospholipids from the inner leaflet to the periplasmic leaflet may be coupled with transport along the hydrophobic conduit of the AsmA-

like protein to the OM. *In vitro* pull-down experiments or *in vivo* BPA cross-linking should be utilised to identify possible interactors in future studies.

7.4 Conclusion

Overall, the data presented in this study further characterises the structures and functions of members of the AsmA-like protein family in *E. coli*. With further evidence of a phospholipid transport role through genetic, biochemical and structural investigation, we hypothesise that all six members of the family are likely capable of GPL transport, but that their utilisation within the cell varies. Whether the smaller AsmA-like proteins are evolutionary relics or have a role at particular points of the cell cycle or under certain environmental conditions, is yet to be ascertained. We also provide recommendations for future work based off the data presented, with determination of the phenotypes of the smaller *asmA*-like genes and the structures and *in vitro* functional characterisation of the larger AsmA-like proteins a priority of future work.

BIBLIOGRAPHY

Abeyrathne, P.D. and Grigorieff, N. (2017) Expression, purification, and contaminant detection for structural studies of Ralstonia metallidurance ClC protein rm1. *PLoS ONE*, 12 (7): e0180163. doi:10.1371/journal.pone.0180163.

Abraham, M.J., Murtola, T., Schulz, R., et al. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2: 19–25. doi:10.1016/j.softx.2015.06.001.

Abramson, J., Adler, J., Dunger, J., et al. (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630 (8016): 493–500. doi:10.1038/s41586-024-07487-w.

Alcock, F., Stansfeld, P.J., Basit, H., et al. (2016) Assembling the Tat protein translocase Ben-Tal, N. (ed.). *eLife*, 5: e20718. doi:10.7554/eLife.20718.

Allen, W.J., Watkins, D.W., Dillingham, M.S., et al. (2020) Refined measurement of SecA-driven protein secretion reveals that translocation is indirectly coupled to ATP turnover. *Proceedings of the National Academy of Sciences*, 117 (50): 31808–31816. doi:10.1073/pnas.2010906117.

Ambudkar, S.V., Lynn, A.R., Maloney, P.C., et al. (1986) Reconstitution of ATP-dependent calcium transport from streptococci. *Journal of Biological Chemistry*, 261 (33): 15596–15600. doi:10.1016/S0021-9258(18)66756-0.

Awuni, E. and Mu, Y. (2019) Effect of A22 on the Conformation of Bacterial Actin MreB. *International Journal of Molecular Sciences*, 20 (6): 1304. doi:10.3390/ijms20061304.

Bai, J. and Pagano, R.E. (1997) Measurement of spontaneous transfer and transbilayer movement of BODIPY-labeled lipids in lipid vesicles. *Biochemistry*, 36 (29): 8840–8848. doi:10.1021/bi970145r.

Baldwin, P.R. and Penczek, P.A. (2007) The Transform Class in SPARX and EMAN2. *Journal of Structural Biology*, 157 (1): 250–261. doi:10.1016/j.jsb.2006.06.002.

Bammes, B.E., Jakana, J., Schmid, M.F., et al. (2010) Radiation damage effects at four specimen temperatures from 4 K to 100 K. *Journal of structural biology*, 169 (3): 331–341. doi:10.1016/j.jsb.2009.11.001.

Barton, B., Grinnell, A. and Morgenstein, R.M. (2021) Disruption of the MreB Elongasome Is Overcome by Mutations in the Tricarboxylic Acid Cycle. *Frontiers in Microbiology*, 12: 664281. doi:10.3389/fmicb.2021.664281.

Bayer, M.E. (1968) Areas of adhesion between wall and membrane of Escherichia coli. *Journal of General Microbiology*, 53 (3): 395–404. doi:10.1099/00221287-53-3-395.

Bendezú, F.O. and de Boer, P.A.J. (2008) Conditional lethality, division defects, membrane involution, and endocytosis in mre and mrd shape mutants of Escherichia coli. *Journal of Bacteriology*, 190 (5): 1792–1811. doi:10.1128/JB.01322-07.

Bepler, T., Morin, A., Rapp, M., et al. (2019) Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods*, 16 (11): 1153–1160. doi:10.1038/s41592-019-0575-8.

Bernadac, A., Gavioli, M., Lazzaroni, J.C., et al. (1998) Escherichia coli tol-pal mutants form outer membrane vesicles. *Journal of Bacteriology*, 180 (18): 4872–4878. doi:10.1128/JB.180.18.4872-4878.1998.

Bhella, D. (2019) Cryo-electron microscopy: an introduction to the technique, and considerations when working to establish a national facility. *Biophysical Reviews*, 11 (4): 515–519. doi:10.1007/s12551-019-00571-w.

Bian, X. and De Camilli, P. (2019) In Vitro Assays to Measure the Membrane Tethering and Lipid Transport Activities of the Extended Synaptotagmins. *Methods in molecular biology (Clifton, N.J.)*, 1949: 201–212. doi:10.1007/978-1-4939-9136-5_15.

Bian, X., Saheki, Y. and De Camilli, P. (2018) Ca2+ releases E-Syt1 autoinhibition to couple ER-plasma membrane tethering with lipid transport. *The EMBO journal*, 37 (2): 219–234. doi:10.15252/embj.201797359.

Biou, V. (2023) Lipid-membrane protein interaction visualised by cryo-EM: A review. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1865 (1): 184068. doi:10.1016/j.bbamem.2022.184068.

Birch, J., Cheruvara, H., Gamage, N., et al. (2020) Changes in Membrane Protein Structural Biology. *Biology*, 9 (11): 401. doi:10.3390/biology9110401.

Bishop, R.E., Gibbons, H.S., Guina, T., et al. (2000) Transfer of palmitate from phospholipids to lipid A in outer membranes of Gram-negative bacteria. *The EMBO Journal*, 19 (19): 5071–5080. doi:10.1093/emboj/cdd507.

Bogdanov, M., Pyrshev, K., Yesylevskyy, S., et al. (2020) Phospholipid distribution in the cytoplasmic membrane of Gram-negative bacteria is highly asymmetric, dynamic, and cell shape-dependent. *Science Advances*, 6 (23): eaaz6333. doi:10.1126/sciadv.aaz6333.

Bolotin, E. and Hershberg, R. (2016) Bacterial intra-species gene loss occurs in a largely clocklike manner mostly within a pool of less conserved and constrained genes. *Scientific Reports*, 6 (1): 35168. doi:10.1038/srep35168.

Botte, M., Ni, D., Schenck, S., et al. (2022) Cryo-EM structures of a LptDE transporter in complex with Pro-macrobodies offer insight into lipopolysaccharide translocation. *Nature Communications*, 13 (1): 1826. doi:10.1038/s41467-022-29459-2.

Braun, E., Gilmer, J., Mayes, H.B., et al. (2019) Best Practices for Foundations in Molecular Simulations [Article v1.0]. *Living journal of computational molecular science*, 1 (1): 5957. doi:10.33011/livecoms.1.1.5957.

Burkinshaw, B.J., Deng, W., Lameignère, E., et al. (2015) Structural Analysis of a Specialized Type III Secretion System Peptidoglycan-cleaving Enzyme. *The Journal of Biological Chemistry*, 290 (16): 10406–10417. doi:10.1074/jbc.M115.639013.

Carey, A.B., Ashenden, A. and Köper, I. (2022) Model architectures for bacterial membranes. *Biophysical Reviews*, 14 (1): 111–143. doi:10.1007/s12551-021-00913-7.

Carpenter, E.P., Beis, K., Cameron, A.D., et al. (2008) Overcoming the challenges of membrane protein crystallography. *Current Opinion in Structural Biology*, 18 (5): 581–586. doi:10.1016/j.sbi.2008.07.001.

Chen, J., Fruhauf, A., Fan, C., et al. (2023) Structure of an endogenous mycobacterial MCE lipid transporter. *Nature*, 620 (7973): 445–452. doi:10.1038/s41586-023-06366-0.

Chi, X., Fan, Q., Zhang, Y., et al. (2020) Structural mechanism of phospholipids translocation by MlaFEDB complex. *Cell Research*, 30 (12): 1127–1135. doi:10.1038/s41422-020-00404-6.

Chong, Z.-S., Woo, W.-F. and Chng, S.-S. (2015) Osmoporin OmpC forms a complex with MlaA to maintain outer membrane lipid asymmetry in Escherichia coli. *Molecular Microbiology*, 98 (6): 1133–1146. doi:10.1111/mmi.13202.

Cohen, E.J., Ferreira, J.L., Ladinsky, M.S., et al. (2017) Nanoscale-length control of the flagellar driveshaft requires hitting the tethered outer membrane. *Science*, 356 (6334): 197–200. doi:10.1126/science.aam6512.

Collins, R.F., Beis, K., Dong, C., et al. (2007) The 3D structure of a periplasm-spanning platform required for assembly of group 1 capsular polysaccharides in Escherichia coli.

Proceedings of the National Academy of Sciences of the United States of America, 104 (7): 2390–2395. doi:10.1073/pnas.0607763104.

Collu, G., Mohammed, I., Lafita, A., et al. (2021) Cryo-EM structure of a single-chain β1-adrenoceptor – AmpC β-lactamase fusion protein. *bioRxiv*, p. 2021.09.25.461805. doi:10.1101/2021.09.25.461805.

Cooper, B.F. (2022) *Structural and function investigation of the Paraquat-inducible pathway.*

Cooper, B.F., Clark, R., Kudhail, A., et al. (2023) Phospholipid transport to the bacterial outer membrane through an envelope-spanning bridge. *bioRxiv*, p. 2023.10.05.561070. doi:10.1101/2023.10.05.561070.

Cooper, B.F., Ratkevičiūtė, G., Clifton, L.A., et al. (2024) An octameric PqiC toroid stabilises the outer-membrane interaction of the PqiABC transport system. *EMBO reports*, 25 (1): 82–101. doi:10.1038/s44319-023-00014-4.

Cotter, P.A., Chepuri, V., Gennis, R.B., et al. (1990) Cytochrome o (cyoABCDE) and d (cydAB) oxidase gene expression in Escherichia coli is regulated by oxygen, pH, and the fnr gene product. *Journal of Bacteriology*, 172 (11): 6333–6338. doi:10.1128/jb.172.11.6333-6338.1990.

Coudray, N., Isom, G.L., MacRae, M.R., et al. (2020) Structure of bacterial phospholipid transporter MlaFEDB with substrate bound Boudker, O., Frost, A., van den Berg, B., et al. (eds.). *eLife*, 9: e62518. doi:10.7554/eLife.62518.

Dekker, N. (2000) Outer-membrane phospholipase A: known structure, unknown biological function. *Molecular Microbiology*, 35 (4): 711–717. doi:10.1046/j.1365-2958.2000.01775.x.

Deng, M. and Misra, R. (1996) Examination of AsmA and its effect on the assembly of Escherichia coli outer membrane proteins. *Molecular Microbiology*, 21 (3): 605–612. doi:10.1111/j.1365-2958.1996.tb02568.x.

Doerrler, W.T., Reedy, M.C. and Raetz, C.R. (2001) An Escherichia coli mutant defective in lipid export. *The Journal of Biological Chemistry*, 276 (15): 11461–11464. doi:10.1074/jbc.C100091200.

Doherty, H.M., Kritikos, G., Galardini, M., et al. (2023) ChemGAPP: a tool for chemical genomics analysis and phenotypic profiling. *Bioinformatics (Oxford, England)*, 39 (4): btad171. doi:10.1093/bioinformatics/btad171.

Donohue-Rolfe, A.M. and Schaechter, M. (1980) Translocation of phospholipids from the inner to the outer membrane of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 77 (4): 1867–1871. doi:10.1073/pnas.77.4.1867.

Douglass, M.V., McLean, A.B. and Trent, M.S. (2022) Absence of YhdP, TamB, and YdbH leads to defects in glycerophospholipid transport and cell morphology in Gram-negative bacteria. *PLOS Genetics*, 18 (2): e1010096. doi:10.1371/journal.pgen.1010096.

Doyle, M.T., Jimah, J.R., Dowdy, T., et al. (2022) Cryo-EM structures reveal multiple stages of bacterial outer membrane protein folding. *Cell*, 185 (7): 1143-1156.e13. doi:10.1016/j.cell.2022.02.016.

Drulyte, I., Johnson, R.M., Hesketh, E.L., et al. (2018) Approaches to altering particle distributions in cryo-electron microscopy sample preparation. *Acta Crystallographica. Section D, Structural Biology*, 74 (Pt 6): 560–571. doi:10.1107/S2059798318006496.

Dubochet, J., Adrian, M., Chang, J.J., et al. (1988) Cryo-electron microscopy of vitrified specimens. *Quarterly Reviews of Biophysics*, 21 (2): 129–228. doi:10.1017/s0033583500004297.

Ducret, A., Quardokus, E.M. and Brun, Y.V. (2016) MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. *Nature Microbiology*, 1 (7): 1–7. doi:10.1038/nmicrobiol.2016.77.

Duong, F., Eichler, J., Price, A., et al. (1997) Biogenesis of the Gram-Negative Bacterial Envelope. *Cell*, 91 (5): 567–573. doi:10.1016/S0092-8674(00)80444-4.

Ekiert, D.C., Bhabha, G., Isom, G.L., et al. (2017) Architectures of Lipid Transport Systems for the Bacterial Outer Membrane. *Cell*, 169 (2): 273-285.e17. doi:10.1016/j.cell.2017.03.019.

Ernst, C.M., Kuhn, S., Slavetinsky, C.J., et al. (2015) The Lipid-Modifying Multiple Peptide Resistance Factor Is an Oligomer Consisting of Distinct Interacting Synthase and Flippase Subunits. *mBio*, 6 (1): 10.1128/mbio.02340-14. doi:10.1128/mbio.02340-14.

Fenn, K.L., Horne, J.E., Crossley, J.A., et al. (2024) Outer membrane protein assembly mediated by BAM-SurA complexes. *Nature Communications*, 15 (1): 7612. doi:10.1038/s41467-024-51358-x.

Franken, L.E., Grünewald, K., Boekema, E.J., et al. (2020) A Technical Introduction to Transmission Electron Microscopy for Soft-Matter: Imaging, Possibilities, Choices, and Technical Developments. *Small*, 16 (14): 1906198. doi:10.1002/smll.201906198.

Gallagher, J.R., Kim, A.J., Gulati, N.M., et al. (2019) Negative-Stain Transmission Electron Microscopy of Molecular Complexes for Image Analysis by 2D Class Averaging. *Current Protocols in Microbiology*, 54 (1): e90. doi:10.1002/cpmc.90.

Gamage, N., Cheruvara, H., Harrison, P.J., et al. (2023) "High-Throughput Production and Optimization of Membrane Proteins After Expression in Mammalian Cells." In Sousa, Â. and Passarinha, L. (eds.) *Advanced Methods in Structural Biology*. New York, NY: Springer US. pp. 79–118. doi:10.1007/978-1-0716-3147-8_5.

Garde, S., Chodisetti, P.K. and Reddy, M. (2021) Peptidoglycan: Structure, Synthesis, and Regulation. *EcoSal Plus*, 9 (2). doi:10.1128/ecosalplus.ESP-0010-2020.

Gerding, M.A., Ogata, Y., Pecora, N.D., et al. (2007) The trans-envelope Tol-Pal complex is part of the cell division machinery and required for proper outer-membrane invagination during cell constriction in E. coli. *Molecular Microbiology*, 63 (4): 1008–1025. doi:10.1111/j.1365-2958.2006.05571.x.

Ghanbarpour, A., Valverde, D.P., Melia, T.J., et al. (2021) A model for a partnership of lipid transfer proteins and scramblases in membrane expansion and organelle biogenesis. *Proceedings of the National Academy of Sciences*, 118 (16): e2101562118. doi:10.1073/pnas.2101562118.

Gitai, Z., Dye, N.A., Reisenauer, A., et al. (2005) MreB Actin-Mediated Segregation of a Specific Region of a Bacterial Chromosome. *Cell*, 120 (3): 329–341. doi:10.1016/j.cell.2005.01.007.

Goddard, A.D., Dijkman, P.M., Adamson, R.J., et al. (2015) "Chapter Nineteen - Reconstitution of Membrane Proteins: A GPCR as an Example." In Shukla, A.K. (ed.) *Methods in Enzymology*. Membrane Proteins—Production and Functional Characterization. Academic Press. pp. 405–424. doi:10.1016/bs.mie.2015.01.004.

Godoy-Hernandez, A., Asseri, A.H., Purugganan, A.J., et al. (2023) Rapid and Highly Stable Membrane Reconstitution by LAiR Enables the Study of Physiological Integral Membrane Protein Functions. *ACS Central Science*, 9 (3): 494–507. doi:10.1021/acscentsci.2c01170.

Goverde, C.A., Pacesa, M., Goldbach, N., et al. (2024) Computational design of soluble and functional membrane protein analogues. *Nature*, 631 (8020): 449–458. doi:10.1038/s41586-024-07601-y.

Grasekamp, K.P., Beaud Benyahia, B., Taib, N., et al. (2023) The Mla system of diderm Firmicute Veillonella parvula reveals an ancestral transenvelope bridge for phospholipid trafficking. *Nature Communications*, 14 (1): 7642. doi:10.1038/s41467-023-43411-y.

Grimm, J., Shi, H., Wang, W., et al. (2020) The inner membrane protein YhdP modulates the rate of anterograde phospholipid flow in Escherichia coli. *Proceedings of the National Academy of Sciences*, 117 (43): 26907–26914.

Guo, D., Singh, H., Shimoyama, A., et al. (2021) Energetics of lipid transport by the ABC transporter MsbA is lipid dependent. *Communications Biology*, 4 (1): 1–10. doi:10.1038/s42003-021-02902-8.

Guo, L., Lim, K.B., Poduje, C.M., et al. (1998) Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides. *Cell*, 95 (2): 189–198. doi:10.1016/s0092-8674(00)81750-x.

Hallgren, J., Tsirigos, K., Pedersen, M.D., et al. (2022) *DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural network*. doi:10.1101/2022.04.08.487609.

Hanna, M.G., Suen, P.H., Wu, Y., et al. (2022) SHIP164 is a chorein motif lipid transfer protein that controls endosome–Golgi membrane traffic. *Journal of Cell Biology*, 221 (6): e202111018. doi:10.1083/jcb.202111018.

Heinrichs, D.E., Monteiro, M.A., Perry, M.B., et al. (1998) The Assembly System for the Lipopolysaccharide R2 Core-type of Escherichia coli Is a Hybrid of Those Found in Escherichia coli K-12 and Salmonella enterica. *Journal of Biological Chemistry*, 273 (15): 8849–8859. doi:10.1074/jbc.273.15.8849.

Heinz, E., Selkrig, J., Belousoff, M.J., et al. (2015) Evolution of the Translocation and Assembly Module (TAM). *Genome Biology and Evolution*, 7 (6): 1628–1643. doi:10.1093/gbe/evv097.

Herreros, D., Lederman, R.R., Krieger, J.M., et al. (2023) Estimating conformational landscapes from Cryo-EM particles by 3D Zernike polynomials. *Nature Communications*, 14 (1): 154. doi:10.1038/s41467-023-35791-y.

Hess, B., Bekker, H., Berendsen, H.J.C., et al. (1997) LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18 (12): 1463–1472.

Hochreiter, B., Pardo-Garcia, A. and Schmid, J.A. (2015) Fluorescent Proteins as Genetically Encoded FRET Biosensors in Life Sciences. *Sensors*, 15 (10): 26281–26314. doi:10.3390/s151026281.

Huang, Y.-M.M., Miao, Y., Munguia, J., et al. (2016) Molecular dynamic study of MlaC protein in Gram-negative bacteria: conformational flexibility, solvent effect and protein-phospholipid binding. *Protein Science: A Publication of the Protein Society*, 25 (8): 1430–1437. doi:10.1002/pro.2939.

Hughes, G.W., Hall, S.C.L., Laxton, C.S., et al. (2019) Evidence for phospholipid export from the bacterial inner membrane by the Mla ABC transport system. *Nature Microbiology*, 4 (10): 1692–1705. doi:10.1038/s41564-019-0481-y.

Hummels, K.R., Berry, S.P., Li, Z., et al. (2023) Coordination of bacterial cell wall and outer membrane biosynthesis. *Nature*, 615 (7951): 300–304. doi:10.1038/s41586-023-05750-0.

Isom, G.L., Coudray, N., MacRae, M.R., et al. (2020) LetB structure reveals a tunnel for lipid transport across the bacterial envelope. *Cell*, 181 (3): 653-664.e19. doi:10.1016/j.cell.2020.03.030.

Isom, G.L., Davies, N.J., Chong, Z.-S., et al. (2017) MCE domain proteins: conserved inner membrane lipid-binding proteins required for outer membrane homeostasis. *Scientific Reports*, 7 (1): 8608. doi:10.1038/s41598-017-09111-6.

Iwai, N., Nagai, K. and Wachi, M. (2002) Novel S-Benzylisothiourea Compound That Induces Spherical Cells in Escherichia coli Probably by Acting on a Rod-shape-determining Protein(s) Other Than Penicillin-binding Protein 2. *Bioscience, Biotechnology, and Biochemistry*, 66 (12): 2658–2662. doi:10.1271/bbb.66.2658.

Janausch, I.G., Zientz, E., Tran, Q.H., et al. (2002) C4-dicarboxylate carriers and sensors in bacteria. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1553 (1): 39–56. doi:10.1016/S0005-2728(01)00233-X.

Jiang, S., Çelen, G., Glatter, T., et al. (2024) A cell-free system for functional studies of small membrane proteins. *bioRxiv*, p. 2023.12.22.573026. doi:10.1101/2023.12.22.573026.

Jin, J., Wu, R., Zhu, J., et al. (2015) Insights into the Cellular Function of YhdE, a Nucleotide Pyrophosphatase from Escherichia coli. *PLoS ONE*, 10 (2): e0117823. doi:10.1371/journal.pone.0117823.

Jones, D.T. (2007) Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics (Oxford, England)*, 23 (5): 538–544. doi:10.1093/bioinformatics/btl677.

Jones, N.C. and Osborn, M.J. (1977) Translocation of phospholipids between the outer and inner membranes of Salmonella typhimurium. *The Journal of Biological Chemistry*, 252 (20): 7405–7412.

Joseph, B.C., Pichaimuthu, S. and Srimeenakshi, S. (2015) An Overview of the Parameters for Recombinant Protein Expression in Escherichia coli. *Journal of Cell Science & Therapy*, 06 (05). doi:10.4172/2157-7013.1000221.

Josts, I., Stubenrauch, C.J., Vadlamani, G., et al. (2017) The Structure of a Conserved Domain of TamB Reveals a Hydrophobic β Taco Fold. *Structure (London, England: 1993)*, 25 (12): 1898-1906.e5. doi:10.1016/j.str.2017.10.002.

Jumper, J., Evans, R., Pritzel, A., et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, 596 (7873): 583–589. doi:10.1038/s41586-021-03819-2.

Kanfer, J. and Kennedy, E.P. (1964) Metabolism and Function of Bacterial Lipids. *Journal of Biological Chemistry*, 239 (6): 1720–1726. doi:10.1016/S0021-9258(18)91247-0.

Kang, Y., Lehmann, K.S., Vanegas, J., et al. (2024) Structural basis of bulk lipid transfer by bridge-like lipid transfer protein LPD-3. *bioRxiv*, p. 2024.06.21.600134. doi:10.1101/2024.06.21.600134.

Kaur, H., Jakob, R.P., Marzinek, J.K., et al. (2021) The antibiotic darobactin mimics a β-strand to inhibit outer membrane insertase. *Nature*, 593 (7857): 125–129. doi:10.1038/s41586-021-03455-w.

Kaushik, S., He, H. and Dalbey, R.E. (2022) Bacterial Signal Peptides- Navigating the Journey of Proteins. *Frontiers in Physiology*, 13. doi:10.3389/fphys.2022.933153.

Kawano, S., Tamura, Y., Kojima, R., et al. (2018) Structure-function insights into direct lipid transfer between membranes by Mmm1-Mdm12 of ERMES. *The Journal of Cell Biology*, 217 (3): 959–974. doi:10.1083/jcb.201704119.

Kellenberger, E. (1990) The “Bayer bridges” confronted with results from improved electron microscopy methods. *Molecular Microbiology*, 4 (5): 697–705. doi:10.1111/j.1365-2958.1990.tb00640.x.

Kim, S.H., Yoo, H.J., Park, E.J., et al. (2021) Nano Differential Scanning Fluorimetry-Based Thermal Stability Screening and Optimal Buffer Selection for Immunoglobulin G. *Pharmaceuticals*, 15 (1): 29. doi:10.3390/ph15010029.

Kol, M.A., van Dalen, A., de Kroon, A.I.P.M., et al. (2003) Translocation of phospholipids is facilitated by a subset of membrane-spanning proteins of the bacterial cytoplasmic membrane. *The Journal of Biological Chemistry*, 278 (27): 24586–24593. doi:10.1074/jbc.M301875200.

Kolich, L.R., Chang, Y.-T., Coudray, N., et al. (2020) Structure of MlaFB uncovers novel mechanisms of ABC transporter regulation. *eLife*, 9: e60030. doi:10.7554/eLife.60030.

Kotzialampou, A., Protonotariou, E., Skoura, L., et al. (2021) Synergistic Antibacterial and Antibiofilm Activity of the MreB Inhibitor A22 Hydrochloride in Combination with Conventional Antibiotics against Pseudomonas aeruginosa and Escherichia coli Clinical Isolates. *International Journal of Microbiology*, 2021: 3057754. doi:10.1155/2021/3057754.

Kritikos, G., Banzhaf, M., Herrera-Dominguez, L., et al. (2017) A tool named Iris for versatile high-throughput phenotyping in microorganisms. *Nature Microbiology*, 2: 17014. doi:10.1038/nmicrobiol.2017.14.

Krogh, A., Larsson, B., von Heijne, G., et al. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*, 305 (3): 567–580. doi:10.1006/jmbi.2000.4315.

Kruse, T., Bork-Jensen, J. and Gerdes, K. (2005) The morphogenetic MreBCD proteins of Escherichia coli form an essential membrane-bound complex. *Molecular Microbiology*, 55 (1): 78–89. doi:10.1111/j.1365-2958.2004.04367.x.

Kuk, A.C.Y., Hao, A. and Lee, S.-Y. (2022) Structure and Mechanism of the Lipid Flippase MurJ. *Annual review of biochemistry*, 91: 705–729. doi:10.1146/annurev-biochem-040320-105145.

Kumar, G., Mishra, R.R. and Verma, A. (2022) "Introduction to Molecular Dynamics Simulations." In Verma, A., Mavinkere Rangappa, S., Ogata, S., et al. (eds.) *Forcefields for Atomistic-Scale Simulations: Materials and Applications*. Singapore: Springer Nature. pp. 1–19. doi:10.1007/978-981-19-3092-8_1.

Kumar, N., Leonzino, M., Hancock-Cerutti, W., et al. (2018) VPS13A and VPS13C are lipid transport proteins differentially localized at ER contact sites. *The Journal of Cell Biology*, 217 (10): 3625–3639. doi:10.1083/jcb.201807019.

Kumar, S., Davis, R.M. and Ruiz, N. (2024) YdbH and YnbE form an intermembrane bridge to maintain lipid homeostasis in the outer membrane of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 121 (21): e2321512121. doi:10.1073/pnas.2321512121.

Kumar, S. and Ruiz, N. (2023) Bacterial AsmA-Like Proteins: Bridging the Gap in Intermembrane Phospholipid Transport. *Contact*, 6: 25152564231185931. doi:10.1177/25152564231185931.

Langley, K.E., Hawrot, E. and Kennedy, E.P. (1982) Membrane assembly: movement of phosphatidylserine between the cytoplasmic and outer membranes of Escherichia coli. *Journal of Bacteriology*, 152 (3): 1033–1041.

Laue, T., Shah, B., Ridgeway, T., et al. (1992) Computer-aided Interpretation of Sedimentation Data for Proteins. *Analytical ultracentrifugation in biochemistry and polymer science*. Available at: https://scholars.unh.edu/faculty_pubs/801.

Li, C., Tan, B.K., Zhao, J., et al. (2016) In Vivo and in Vitro Synthesis of Phosphatidylglycerol by an Escherichia coli Cardiolipin Synthase. *The Journal of Biological Chemistry*, 291 (48): 25144–25153. doi:10.1074/jbc.M116.762070.

Li, M., Tang, H., Qing, R., et al. (2024) Design of a water-soluble transmembrane receptor kinase with intact molecular function by QTY code. *Nature Communications*, 15 (1): 4293. doi:10.1038/s41467-024-48513-9.

Li, P., Lees, J.A., Lusk, C.P., et al. (2020) Cryo-EM reconstruction of a VPS13 fragment reveals a long groove to channel lipids between membranes. *Journal of Cell Biology*, 219 (5): e202001161. doi:10.1083/jcb.202001161.

Liu, A., Tran, L., Becket, E., et al. (2010) Antibiotic Sensitivity Profiles Determined with an Escherichia coli Gene Knockout Collection: Generating an Antibiotic Bar Code. *Antimicrobial Agents and Chemotherapy*, 54 (4): 1393–1403. doi:10.1128/aac.00906-09.

Liu, B., Furevi, A., Perepelov, A.V., et al. (2019) Structure and genetics of Escherichia coli O antigens. *FEMS Microbiology Reviews*, 44 (6): 655–683. doi:10.1093/femsre/fuz028.

Liwo, A., Czaplewski, C., Sieradzan, A.K., et al. (2021) Theory and Practice of Coarse-Grained Molecular Dynamics of Biologically Important Systems. *Biomolecules*, 11 (9): 1347. doi:10.3390/biom11091347.

Lomize, A.L., Todd, S.C. and Pogozheva, I.D. (2022) Spatial arrangement of proteins in planar and curved membranes by PPM 3.0. *Protein Science: A Publication of the Protein Society*, 31 (1): 209–220. doi:10.1002/pro.4219.

Low, W.-Y., Thong, S. and Chng, S.-S. (2021) ATP disrupts lipid-binding equilibrium to drive retrograde transport critical for bacterial outer membrane asymmetry. *Proceedings of the National Academy of Sciences*, 118 (50): e2110055118. doi:10.1073/pnas.2110055118.

Malinin, V.S., Haque, Md.E. and Lentz, B.R. (2001) The Rate of Lipid Transfer during Fusion Depends on the Structure of Fluorescent Lipid Probes: A New Chain-Labeled Lipid Transfer Probe Pair. *Biochemistry*, 40 (28): 8292–8299. doi:10.1021/bi010570r.

Malinverni, J.C. and Silhavy, T.J. (2009) 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*, 106 (19): 8009–8014. doi:10.1073/pnas.0903229106.

Malinverni, J.C., Werner, J., Kim, S., et al. (2006) YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in Escherichia coli. *Molecular Microbiology*, 61 (1): 151–164. doi:10.1111/j.1365-2958.2006.05211.x.

Mann, D., Fan, J., Somboon, K., et al. (2021) Structure and lipid dynamics in the maintenance of lipid asymmetry inner membrane complex of A. baumannii. *Communications Biology*, 4 (1): 1–9. doi:10.1038/s42003-021-02318-4.

Marciano, D.C., Wang, C., Hsu, T.-K., et al. (2022) Evolutionary action of mutations reveals antimicrobial resistance genes in Escherichia coli. *Nature Communications*, 13 (1): 3189. doi:10.1038/s41467-022-30889-1.

Marlovits, T.C., Kubori, T., Sukhan, A., et al. (2004) Structural Insights into the Assembly of the Type III Secretion Needle Complex. *Science (New York, N.Y.)*, 306 (5698): 1040–1042. doi:10.1126/science.1102610.

Masuda, K., Matsuyama, S. and Tokuda, H. (2002) Elucidation of the function of lipoprotein-sorting signals that determine membrane localization. *Proceedings of the National Academy of Sciences of the United States of America*, 99 (11): 7390–7395. doi:10.1073/pnas.112085599.

Matias, V.R.F., Al-Amoudi, A., Dubochet, J., et al. (2003) Cryo-Transmission Electron Microscopy of Frozen-Hydrated Sections of Escherichia coli and Pseudomonas

aeruginosa. *Journal of Bacteriology*, 185 (20): 6112–6118. doi:10.1128/jb.185.20.6112-6118.2003.
Miroux, B. and Walker, J.E. (1996) Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology*, 260 (3): 289–298. doi:10.1006/jmbi.1996.0399.
Misra, R. and Miao, Y. (1995) Molecular analysis of asmA, a locus identified as the suppressor of OmpF assembly mutants of Escherichia coli K-12. *Molecular Microbiology*, 16 (4): 779–788. doi:10.1111/j.1365-2958.1995.tb02439.x.
Mitchell, A.M., Srikumar, T. and Silhavy, T.J. (2018) Cyclic Enterobacterial Common Antigen Maintains the Outer Membrane Permeability Barrier of Escherichia coli in a Manner Controlled by YhdP. *mBio*, 9 (4): e01321-18. doi:10.1128/mBio.01321-18.
Mitchell, A.M., Wang, W. and Silhavy, T.J. (2017) Novel RpoS-Dependent Mechanisms Strengthen the Envelope Permeability Barrier during Stationary Phase. *Journal of Bacteriology*, 199 (2): e00708-16. doi:10.1128/JB.00708-16.
Morita, K., Hama, Y., Izume, T., et al. (2018) Genome-wide CRISPR screen identifies TMEM41B as a gene required for autophagosome formation. *The Journal of Cell Biology*, 217 (11): 3817–3828. doi:10.1083/jcb.201804132.
Murzyn, K., Róg, T. and Pasenkiewicz-Gierula, M. (2005) Phosphatidylethanolamine-Phosphatidylglycerol Bilayer as a Model of the Inner Bacterial Membrane. *Biophysical Journal*, 88 (2): 1091–1103. doi:10.1529/biophysj.104.048835.
Nakane, T. and Scheres, S.H.W. (2021) Multi-body Refinement of Cryo-EM Images in RELION. *Methods in Molecular Biology (Clifton, N.J.)*, 2215: 145–160. doi:10.1007/978-1-0716-0966-8_7.
Nakayama, T. and Zhang-Akiyama, Q.-M. (2016) pqiABC and yebST, Putative mce Operons of Escherichia coli, Encode Transport Pathways and Contribute to Membrane Integrity. *Journal of Bacteriology*, 199 (1): 10.1128/jb.00606-16. doi:10.1128/jb.00606-16.
Neuman, S.D., Levine, T.P. and Bashirullah, A. (2022) A novel superfamily of bridge-like lipid transfer proteins. *Trends in Cell Biology*, 32 (11): 962–974. doi:10.1016/j.tcb.2022.03.011.
Neumann, P., Dickmanns, A. and Ficner, R. (2018) Validating Resolution Revolution. *Structure*, 26 (5): 785-795.e4. doi:10.1016/j.str.2018.03.004.
Nichols, J.W. and Pagano, R.E. (1983) Resonance energy transfer assay of protein-mediated lipid transfer between vesicles. *The Journal of Biological Chemistry*, 258 (9): 5368–5371.
Nichols, R.J., Sen, S., Choo, Y.J., et al. (2011) Phenotypic Landscape of a Bacterial Cell. *Cell*, 144 (1): 143–156. doi:10.1016/j.cell.2010.11.052.
Nikaido, H. (2003) Molecular Basis of Bacterial Outer Membrane Permeability Revisited. *Microbiology and Molecular Biology Reviews*, 67 (4): 593–656. doi:10.1128/MMBR.67.4.593-656.2003.
Nikaido, H. and Zgurskaya, H.I. (2001) AcrAB and related multidrug efflux pumps of Escherichia coli. *Journal of Molecular Microbiology and Biotechnology*, 3 (2): 215–218.
Okuda, S., Sherman, D.J., Silhavy, T.J., et al. (2016) Lipopolysaccharide transport and assembly at the outer membrane: the PEZ model. *Nature reviews. Microbiology*, 14 (6): 337–345. doi:10.1038/nrmicro.2016.25.
Okuda, S. and Tokuda, H. (2011) Lipoprotein sorting in bacteria. *Annual Review of Microbiology*, 65: 239–259. doi:10.1146/annurev-micro-090110-102859.

O'Neill, J. (2016) *The Review on Antimicrobial Resistance*. Wellcome Trust and HM Government. Available at: https://amr-review.org/ (Accessed: 25 July 2024).

Osawa, T., Kotani, T., Kawaoka, T., et al. (2019) Atg2 mediates direct lipid transfer between membranes for autophagosome formation. *Nature Structural & Molecular Biology*, 26 (4): 281–288. doi:10.1038/s41594-019-0203-4.

Oswald, J., Njenga, R., Natriashvili, A., et al. (2021) The Dynamic SecYEG Translocon. *Frontiers in Molecular Biosciences*, 8. Available at: https://www.frontiersin.org/article/10.3389/fmolb.2021.664241 (Accessed: 24 May 2022).

Pan, K.-L., Hsiao, H.-C., Weng, C.-L., et al. (2003) Roles of DegP in Prevention of Protein Misfolding in the Periplasm upon Overexpression of Penicillin Acylase in Escherichia coli. *Journal of Bacteriology*, 185 (10): 3020–3030. doi:10.1128/JB.185.10.3020-3030.2003.

Panagabko, C., Baptist, M. and Atkinson, J. (2019) In vitro lipid transfer assays of phosphatidylinositol transfer proteins provide insight into the in vivo mechanism of ligand transfer. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1861 (3): 619–630. doi:10.1016/j.bbamem.2018.12.003.

Parsons, J.B. and Rock, C.O. (2013) Bacterial lipids: metabolism and membrane homeostasis. *Progress in Lipid Research*, 52 (3): 249–276. doi:10.1016/j.plipres.2013.02.002.

Pettersen, E.F., Goddard, T.D., Huang, C.C., et al. (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science : A Publication of the Protein Society*, 30 (1): 70–82. doi:10.1002/pro.3943.

Pravda, L., Sehnal, D., Toušek, D., et al. (2018) MOLEonline: a web-based tool for analyzing channels, tunnels and pores (2018 update). *Nucleic Acids Research*, 46 (W1): W368–W373. doi:10.1093/nar/gky309.

Price, M.N., Wetmore, K.M., Waters, R.J., et al. (2018) Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature*, 557 (7706): 503–509. doi:10.1038/s41586-018-0124-0.

Punjani, A., Rubinstein, J.L., Fleet, D.J., et al. (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, 14 (3): 290–296. doi:10.1038/nmeth.4169.

Quan, S., Schneider, I., Pan, J., et al. (2007) The CXXC motif is more than a redox rheostat. *The Journal of Biological Chemistry*, 282 (39): 28823–28833. doi:10.1074/jbc.M705291200.

Raetz, C.R. and Dowhan, W. (1990) Biosynthesis and function of phospholipids in Escherichia coli. *The Journal of Biological Chemistry*, 265 (3): 1235–1238.

Rai, A.K., Sawasato, K., Bennett, H.C., et al. (2024) Genetic evidence for functional diversification of gram-negative intermembrane phospholipid transporters. *PLOS Genetics*, 20 (6): e1011335. doi:10.1371/journal.pgen.1011335.

Ratkevičiūtė, G. (2023) *Structural investigation of putative lipid transport systems in Bdellovibrio bacteriovorus HD100*.

Ratkeviciute, G., Cooper, B.F. and Knowles, T.J. (2021) Methods for the solubilisation of membrane proteins: the micelle-aneous world of membrane protein solubilisation. *Biochemical Society Transactions*, 49 (4): 1763–1777. doi:10.1042/BST20210181.

Riu, F., Ruda, A., Ibba, R., et al. (2022) Antibiotics and Carbohydrate-Containing Drugs Targeting Bacterial Cell Envelopes: An Overview. *Pharmaceuticals*, 15 (8): 942. doi:10.3390/ph15080942.

Robichon, C., Luo, J., Causey, T.B., et al. (2011) Engineering Escherichia coli BL21(DE3) Derivative Strains To Minimize E. coli Protein Contamination after Purification by Immobilized Metal Affinity Chromatography. *Applied and Environmental Microbiology*, 77 (13): 4634–4646. doi:10.1128/AEM.00119-11.

Rohou, A. and Grigorieff, N. (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*, 192 (2): 216–221. doi:10.1016/j.jsb.2015.08.008.

Roney, I.J. and Rudner, D.Z. (2023) The DedA superfamily member PetA is required for the transbilayer distribution of phosphatidylethanolamine in bacterial membranes. *Proceedings of the National Academy of Sciences*, 120 (20): e2301979120. doi:10.1073/pnas.2301979120.

Rosenthal, P.B. and Henderson, R. (2003) Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *Journal of Molecular Biology*, 333 (4): 721–745. doi:10.1016/j.jmb.2003.07.013.

Rowlett, V.W., Mallampalli, V.K.P.S., Karlstaedt, A., et al. (2017) Impact of Membrane Phospholipid Alterations in Escherichia coli on Cellular Function and Bacterial Stress Adaptation. *Journal of Bacteriology*, 199 (13): e00849-16. doi:10.1128/JB.00849-16.

Ruiz, N., Davis, R.M. and Kumar, S. (2021) YhdP, TamB, and YdbH Are Redundant but Essential for Growth and Lipid Homeostasis of the Gram-Negative Outer Membrane. *mBio*, 12 (6): e0271421. doi:10.1128/mBio.02714-21.

Ryan, M.M., Temple, B.R.S., Phillips, S.E., et al. (2007) Conformational Dynamics of the Major Yeast Phosphatidylinositol Transfer Protein Sec14p: Insight into the Mechanisms of Phospholipid Exchange and Diseases of Sec14p-Like Protein Deficiencies. *Molecular Biology of the Cell*, 18 (5): 1928–1942. doi:10.1091/mbc.E06-11-1024.

Saha, S., Lach, S.R. and Konovalova, A. (2021) Homeostasis of the Gram-negative cell envelope. *Current opinion in microbiology*, 61: 99–106. doi:10.1016/j.mib.2021.03.008.

Schauder, C.M., Wu, X., Saheki, Y., et al. (2014) Structure of a lipid-bound extended synaptotagmin indicates a role in lipid transfer. *Nature*, 510 (7506): 552–555. doi:10.1038/nature13269.

Scheres, S.H.W. (2012) RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, 180 (3): 519–530. doi:10.1016/j.jsb.2012.09.006.

Schrodinger (2015) *The PyMOL Molecular Graphics System, Version 1.8*.

Schuck, P. (2000) Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophysical Journal*, 78 (3): 1606–1619. doi:10.1016/S0006-3495(00)76713-0.

Selkrig, J., Mosbahi, K., Webb, C.T., et al. (2012) Discovery of an archetypal protein transport system in bacterial outer membranes. *Nature Structural & Molecular Biology*, 19 (5): 506–510, S1. doi:10.1038/nsmb.2261.

Shen, H.-H., Leyton, D.L., Shiota, T., et al. (2014) Reconstitution of a nanomachine driving the assembly of proteins into bacterial outer membranes. *Nature Communications*, 5 (1): 5078. doi:10.1038/ncomms6078.

Sherman, D.J., Xie, R., Taylor, R.J., et al. (2018) Lipopolysaccharide is transported to the cell surface by a membrane-to-membrane protein bridge. *Science (New York, N.Y.)*, 359 (6377): 798–801. doi:10.1126/science.aar1886.

Shiver, A.L., Osadnik, H., Kritikos, G., et al. (2016) A Chemical-Genomic Screen of Neglected Antibiotics Reveals Illicit Transport of Kasugamycin and Blasticidin S. *PLOS Genetics*, 12 (6): e1006124. doi:10.1371/journal.pgen.1006124.

Shrivastava, R. and Chng, S.-S. (2019) Lipid trafficking across the Gram-negative cell envelope. *The Journal of Biological Chemistry*, 294 (39): 14175–14184. doi:10.1074/jbc.AW119.008139.

Shrivastava, R., Jiang, X. and Chng, S.-S. (2017) Outer membrane lipid homeostasis via retrograde phospholipid transport in Escherichia coli. *Molecular Microbiology*, 106 (3): 395–408. doi:10.1111/mmi.13772.

Sigworth, F.J. (2016) Principles of cryo-EM single-particle image processing. *Microscopy*, 65 (1): 57–67. doi:10.1093/jmicro/dfv370.

Silhavy, T.J., Kahne, D. and Walker, S. (2010) The Bacterial Cell Envelope. *Cold Spring Harbor Perspectives in Biology*, 2 (5): a000414. doi:10.1101/cshperspect.a000414.

Souza, P.C.T., Alessandri, R., Barnoud, J., et al. (2021) Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nature Methods*, 18 (4): 382–388. doi:10.1038/s41592-021-01098-3.

Sperandeo, P., Lau, F.K., Carpentieri, A., et al. (2008) Functional analysis of the protein machinery required for transport of lipopolysaccharide to the outer membrane of Escherichia coli. *Journal of Bacteriology*, 190 (13): 4460–4469. doi:10.1128/JB.00270-08.

Sposato, D., Mercolino, J., Torrini, L., et al. (2024) Redundant essentiality of AsmA-like proteins in Pseudomonas aeruginosa. *mSphere*, 9 (2): e00677-23. doi:10.1128/msphere.00677-23.

Steinberg, R., Origi, A., Natriashvili, A., et al. (2020) Posttranslational insertion of small membrane proteins by the bacterial signal recognition particle. *PLoS biology*, 18 (9): e3000874. doi:10.1371/journal.pbio.3000874.

Steinkühler, J., Peruzzi, J.A., Krüger, A., et al. (2024) Improving Cell-Free Expression of Model Membrane Proteins by Tuning Ribosome Cotranslational Membrane Association and Nascent Chain Aggregation. *ACS Synthetic Biology*, 13 (1): 129–140. doi:10.1021/acssynbio.3c00357.

Studier, F.W. (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *Journal of Molecular Biology*, 219 (1): 37–44. doi:10.1016/0022-2836(91)90855-z.

Suits, M.D.L., Sperandeo, P., Dehò, G., et al. (2008) Novel structure of the conserved gram-negative lipopolysaccharide transport protein A and mutagenesis analysis. *Journal of Molecular Biology*, 380 (3): 476–488. doi:10.1016/j.jmb.2008.04.045.

Sutterlin, H.A., Shi, H., May, K.L., et al. (2016) Disruption of lipid homeostasis in the Gram-negative cell envelope activates a novel cell death pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 113 (11): E1565-1574. doi:10.1073/pnas.1601375113.

Szklarczyk, D., Gable, A.L., Nastou, K.C., et al. (2021) The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*, 49 (D1): D605–D612. doi:10.1093/nar/gkaa1074.

Tan, W.B. and Chng, S.-S. (2024) Primary role of the Tol-Pal complex in bacterial outer membrane lipid homeostasis. *bioRxiv*. doi:10.1101/2024.05.08.593160.

Tang, X., Chang, S., Qiao, W., et al. (2021) Structural insights into outer membrane asymmetry maintenance in Gram-negative bacteria by MlaFEDB. *Nature Structural & Molecular Biology*, 28 (1): 81–91. doi:10.1038/s41594-020-00532-y.

Theuretzbacher, U., Outterson, K., Engel, A., et al. (2020) The global preclinical antibacterial pipeline. *Nature Reviews Microbiology*, 18 (5): 275–285. doi:10.1038/s41579-019-0288-0.

Thompson, R.F., Walker, M., Siebert, C.A., et al. (2016) An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods (San Diego, Calif.)*, 100: 3–15. doi:10.1016/j.ymeth.2016.02.017.

Thong, S., Ercan, B., Torta, F., et al. (2016) Defining key roles for auxiliary proteins in an ABC transporter that maintains bacterial outer membrane lipid asymmetry Storz, G. (ed.). *eLife*, 5: e19042. doi:10.7554/eLife.19042.

Trent, M.S. (2004) Biosynthesis, transport, and modification of lipid A. *Biochemistry and Cell Biology*, 82 (1): 71–86. doi:10.1139/o03-070.

Van Gunsteren, W.F. and Berendsen, H.J.C. (1990) Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. *Angewandte Chemie International Edition in English*, 29 (9): 992–1023. doi:10.1002/anie.199009921.

Veesler, D., Blangy, S., Cambillau, C., et al. (2008) There is a baby in the bath water: AcrB contamination is a major problem in membrane-protein crystallization. *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications*, 64 (Pt 10): 880–885. doi:10.1107/S1744309108028248.

Vickery, O.N. and Stansfeld, P.J. (2021) CG2AT2: an Enhanced Fragment-Based Approach for Serial Multi-scale Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation*, 17 (10): 6472–6482. doi:10.1021/acs.jctc.1c00295.

Vieni, C., Coudray, N., Isom, G.L., et al. (2022) Role of Ring6 in the Function of the *E. coli* MCE Protein LetB. *Journal of Molecular Biology*, 434 (7): 167463. doi:10.1016/j.jmb.2022.167463.

Vijgenboom, E., Nilsson, L. and Bosch, L. (1988) The elongation factor EF-Tu from E. coli binds to the upstream activator region of the tRNA-tufB operon. *Nucleic Acids Research*, 16 (21): 10183–10197.

Vogel, J. and Luisi, B.F. (2011) Hfq and its constellation of RNA. *Nature reviews. Microbiology*, 9 (8): 578–589. doi:10.1038/nrmicro2615.

Vollmer, W., Blanot, D. and De Pedro, M.A. (2008) Peptidoglycan structure and architecture. *FEMS Microbiology Reviews*, 32 (2): 149–167. doi:10.1111/j.1574-6976.2007.00094.x.

Wagner, S., Klepsch, M.M., Schlegel, S., et al. (2008) Tuning Escherichia coli for membrane protein overexpression. *Proceedings of the National Academy of Sciences of the United States of America*, 105 (38): 14371–14376. doi:10.1073/pnas.0804090105.

Wagner, T., Merino, F., Stabrin, M., et al. (2019) SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Communications Biology*, 2 (1): 1–13. doi:10.1038/s42003-019-0437-z.

Wang, X., Nyenhuis, S.B. and Bernstein, H.D. (2024a) The translocation assembly module (TAM) catalyzes the assembly of bacterial outer membrane proteins in vitro. *Nature Communications*, 15 (1): 7246. doi:10.1038/s41467-024-51628-8.

Wang, Y., Dahmane, S., Ti, R., et al. (2024b) Structural basis for lipid transfer by the ATG2A–ATG9A complex. *Nature Structural & Molecular Biology*, pp. 1–13. doi:10.1038/s41594-024-01376-6.

Wang, Y., Li, Z., Wang, X., et al. (2023) Insights into membrane association of the SMP domain of extended synaptotagmin. *Nature Communications*, 14 (1): 1504. doi:10.1038/s41467-023-37202-8.

Wassenaar, T.A., Ingólfsson, H.I., Böckmann, R.A., et al. (2015) Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *Journal of Chemical Theory and Computation*, 11 (5): 2144–2155. doi:10.1021/acs.jctc.5b00209.

Watson, J.L., Juergens, D., Bennett, N.R., et al. (2023) De novo design of protein structure and function with RFdiffusion. *Nature*, 620 (7976): 1089–1100. doi:10.1038/s41586-023-06415-8.

Webby, M.N., Williams-Jones, D.P., Press, C., et al. (2022) Force-Generation by the Trans-Envelope Tol-Pal System. *Frontiers in Microbiology*, 13: 852176. doi:10.3389/fmicb.2022.852176.

Wentinck, K., Gogou, C. and Meijer, D.H. (2022) Putting on molecular weight: Enabling cryo-EM structure determination of sub-100-kDa proteins. *Current Research in Structural Biology*, 4: 332–337. doi:10.1016/j.crstbi.2022.09.005.

WHO (2024) *WHO bacterial priority pathogens list, 2024*. Available at: https://www.who.int/publications/i/item/9789240093461 (Accessed: 27 July 2024).

Wong, L.H., Gatta, A.T. and Levine, T.P. (2019) Lipid transfer proteins: the lipid commute via shuttles, bridges and tubes. *Nature Reviews Molecular Cell Biology*, 20 (2): 85–101. doi:10.1038/s41580-018-0071-5.

Wong, L.H. and Levine, T.P. (2017) Tubular lipid binding proteins (TULIPs) growing everywhere. *Biochimica et Biophysica Acta*, 1864 (9): 1439–1449. doi:10.1016/j.bbamcr.2017.05.019.

Wotherspoon, P. (2022) *An investigation into lipid exchange in the maintenance of lipid asymmetry pathway*.

Wotherspoon, P., Johnston, H., Hardy, D.J., et al. (2024) Structure of the MlaC-MlaD complex reveals molecular basis of periplasmic phospholipid transport. *Nature Communications*, 15 (1): 6394. doi:10.1038/s41467-024-50615-3.

Xiong, X., Deeter, J.N. and Misra, R. (1996) Assembly-defective OmpC mutants of Escherichia coli K-12. *Journal of Bacteriology*, 178 (4): 1213–1215. doi:10.1128/jb.178.4.1213-1215.1996.

Yasuhiro, K., Yuzuru, A. and Shoshichi, N. (1967) Composition and turnover of the phospholipids in *Escherichia coli*. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism*, 144 (2): 382–390. doi:10.1016/0005-2760(67)90167-1.

Yeow, J. and Chng, S.-S. (2022) Of zones, bridges and chaperones – phospholipid transport in bacterial outer membrane assembly and homeostasis. *Microbiology*, 168 (4): 001177. doi:10.1099/mic.0.001177.

Yeow, J., Luo, M. and Chng, S.-S. (2023) Molecular mechanism of phospholipid transport at the bacterial outer membrane interface. *Nature Communications*, 14 (1): 8285. doi:10.1038/s41467-023-44144-8.

Yeow, J., Tan, K.W., Holdbrook, D.A., et al. (2018) The architecture of the OmpC-MlaA complex sheds light on the maintenance of outer membrane lipid asymmetry in Escherichia coli. *The Journal of Biological Chemistry*, 293 (29): 11325–11340. doi:10.1074/jbc.RA118.002441.

Yip, K.M., Fischer, N., Paknia, E., et al. (2020) Atomic-resolution protein structure determination by cryo-EM. *Nature*, 587 (7832): 157–161. doi:10.1038/s41586-020-2833-4.

Zhang, X. and Hong Zhou, Z. (2011) Limiting factors in atomic resolution cryo electron microscopy: No simple tricks. *Journal of Structural Biology*, 175 (3): 253–263. doi:10.1016/j.jsb.2011.05.004.

Zhang, X., Xie, H., Iaea, D., et al. (2022) Phosphatidylinositol phosphates modulate interactions between the StarD4 sterol trafficking protein and lipid membranes. *The Journal of Biological Chemistry*, 298 (7): 102058. doi:10.1016/j.jbc.2022.102058.

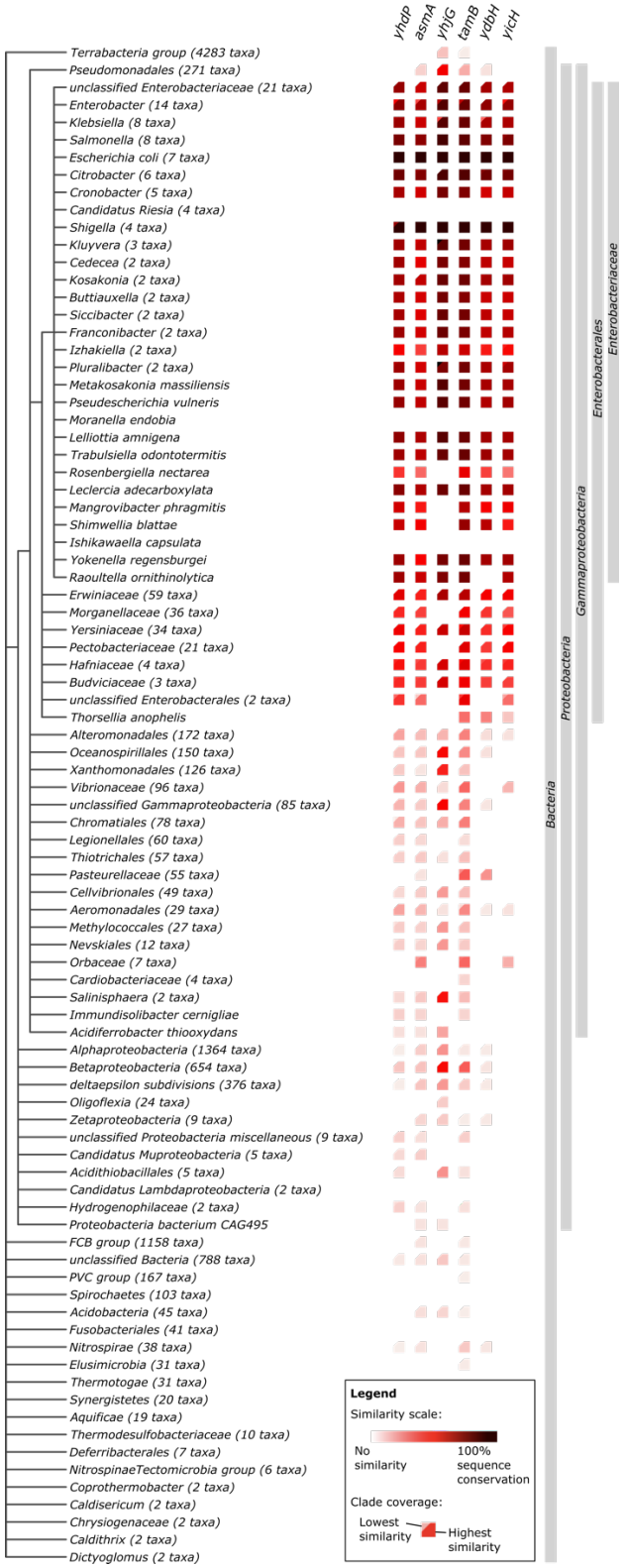
Zheng, S.Q., Palovcak, E., Armache, J.-P., et al. (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods*, 14 (4): 331–332. doi:10.1038/nmeth.4193.

Zhong, E.D., Bepler, T., Berger, B., et al. (2021) CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods*, 18 (2): 176–185. doi:10.1038/s41592-020-01049-4.

Zückert, W.R. (2014) Secretion of Bacterial Lipoproteins: Through the Cytoplasmic Membrane, the Periplasm and Beyond. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1843 (8): 1509–1516. doi:10.1016/j.bbamcr.2014.04.022.

APPENDIX

Appendix 1. Sequence conservation of the AsmA-like protein family across bacteria. Gene cooccurrence data from the STRING database (Szklarczyk et al., 2021). Sequence recovery of homologues across species shown by colour of clade.



Appendix 2. Fitness scores of the *asmA*-like genes from high-throughput screens. Data curated from online phenotypic screen databases. Ordered from greatest fitness defect to greatest fitness increase (Nichols et al., 2011; Shiver et al., 2016; Price et al., 2018).

Knockout	Condition	Score	q values
yhdP	SDS 0.5% (w/v) EDTA 0.5 mM	-28.02217	7.03E-174
yhdP	SDS 1.0% (w/v) EDTA 0.5 mM	-26.93844	1.63E-159
yhdP	SDS 0.5% (w/v) EDTA 0.5 mM	-18.89192	5.49E-77
yhdP	Bacitracin 0.5 mg/mL	-17.69054	2.58E-71
asmA	Bile salts 2% (w/v)	-12.9742	1.30E-36
asmA	Spiramycin 20	-12.53253	2.27E-34
asmA	Deoxycholate 2	-11.97874	9.35E-31
yhdP	Bacitracin 1 mg/mL	-11.1366	5.12E-28
asmA	Spiramycin 5	-10.34281	1.08E-22
yhdP	Vancomycin 40 μg/mL	-10.029	3.52E-21
yhdP	Vancomycin 50 μ/mL	-9.882008	5.03E-21
asmA	Clarythromycin 5	-9.79487	2.76E-20
asmA	Rifampicin 4 μg/mL	-9.794287	7.56E-20
asmA	Clarythromycin 10	-9.505314	4.76E-19
yhdP	Rifampicin 4 μg/mL	-8.427201	1.18E-14
asmA	Chloramphenicol Erythromycin	-8.39463	1.79E-14
asmA	Casaminos N	-8.332143	2.47E-17
asmA	Benzalkonium Minocycline	-8.279666	2.80E-14
yhdP	EDTA 1 mM	-8.207705	1.30E-13
asmA	Clarythromycin 1	-7.881004	9.36E-13
asmA	Streptonigrin 500 ng/mL	-7.748191	5.78E-12
asmA	Tylosin 50	-7.620051	5.25E-12
asmA	Deoxycholate 0.5	-7.432267	1.76E-11
asmA	Deoxycholate 1% (w/v)	-7.42501	1.35E-11
asmA	Actinomycin D 15	-7.047523	2.14E-10
asmA	Erythromycin 5	-7.037306	2.80E-10
asmA	Fusidic acid 50	-6.999065	1.10E-09
asmA	Bile salts 1% (w/v)	-6.899365	6.76E-10
asmA	Fusidic 0.03 mg/mL	-6.893182	3.42E-11
asmA	Dibucaine 1.2	-6.808958	9.12E-10
asmA	Vancomycin 20	-6.756813	2.74E-09
yhdP	EDTA 1 mM	-6.71004	6.41E-14
yhdP	Benzalkonium chloride 0.004 mg/mL	-6.661771	2.16E-10
asmA	Minocyclin E 0.2	-6.584929	2.14E-08
asmA	L-Glutamine N	-6.4782	1.70E-10

asmA	Vancomycin 50 μg/mL	-6.426853	1.26E-08
asmA	Bile salts 2% (w/v)	-6.412535	2.04E-08
yhdP	SDS 1% (w/v)	-6.380611	6.28E-13
yhdP	Benzalkonium 25	-6.285009	3.10E-08
asmA	Azithromycin 0.1	-6.148607	1.99E-07
asmA	Triclosan 0.05	-6.106946	5.06E-08
asmA	Vancomycin 10	-6.098681	1.11E-06
asmA	Doxycycline clarithromycin	-6.080601	1.05E-07
asmA	Streptonigrin 0.5	-6.071781	2.03E-07
yhdP	Deoxycholate 1% (w/v)	-6.041925	1.25E-07
asmA	Dibucaine 0.8	-6.011623	4.35E-07
yhdP	Clarythromycin 10	-5.982379	3.49E-07
yhdP	EGTA 2 mM	-5.934383	5.78E-07
asmA	Doxycycline hyclate 0.001 mg/mL	-5.842035	1.06E-07
asmA	Bile salts 0.5% (w/v)	-5.824033	9.67E-07
asmA	Chloramphenicol 0.001 mg/mL	-5.790682	8.49E-08
yhdP	Bile salts 2% (w/v)	-5.770162	3.69E-07
asmA	Erythromycin.1	-5.548909	3.68E-06
asmA	Spectinomycin 0.0125 mg/mL	-5.540356	1.96E-07
yhdP	EGTA 1% (w/v)	-5.239271	3.31E-05
asmA	Pseudomonic acid A 36 μg/mL	-5.223065	2.65E-05
yhdP	Azithromycin 1	-5.216689	2.95E-05
yhdP	Chloramphenicol Rifampicin	-5.215114	6.38E-06
asmA	NACl 171 mM	-5.186219	4.65E-06
asmA	Minocyclin E 0.2	-5.155804	4.12E-05
asmA	Erythromycin 1	-5.154951	4.26E-05
asmA	Actinomycin D 10	-5.117147	5.60E-05
asmA	Dibucaine 0.4	-5.092055	4.38E-05
asmA	Oxacillin 40	-5.088078	0.000102
asmA	CCCP 2	-5.080862	0.000196
asmA	L-Asparagine N	-5.004795	5.29E-07
yhdP	SDS 4% (w/v)	-4.933153	8.24E-05
asmA	Indole T40C	-4.901516	2.56E-05
asmA	Minocyclin E 0.5	-4.888724	0.000171
yhdP	A-Ketoglutaric C	-4.694995	8.62E-06
asmA	Methylglyoxal 0.008	-4.693421	1.98E-05
asmA	Bacitracin 300	-4.620436	0.000409
asmA	Gly Glu N	-4.58626	2.56E-05
asmA	Bacitracin.100	-4.5387	0.002876
asmA	Spectinomycin 0.00625 mg/mL	-4.536677	5.57E-05

asmA	D-Glucose C	-4.534024	2.66E-05
asmA	Bleomycin 1	-4.460208	0.002066
asmA	NAG C	-4.391957	2.55E-05
asmA	Sucrose C	-4.359312	2.64E-05
asmA	Bacitracin 0.5 mg/mL	-4.318297	0.000154
yhdP	A22 0.08 mg/mL	-4.300599	0.000353
asmA	Bacitracin 200	-4.250303	0.005245
asmA	Erythromycin 10	-4.247379	0.002724
asmA	Azithromycin 1	-4.245026	0.002388
asmA	Benzalkonium 25	-4.211768	0.001382
asmA	Fusidic acid 0.1 mg/mL	-4.141525	0.000607
asmA	Nigericin 5	-4.060162	0.034669
asmA	Puromycin 25	-4.051793	0.013687
ytfN	Chloramphenicol Rifampicin	-4.022819	0.001413
asmA	A-Ketoglutaric xC	-3.941093	0.000324
asmA	Bleomycin 2	-3.899211	0.014832
asmA	Chloramphenicol 0.0025 mg/mL	-3.877022	0.001037
yhdP	SDS 3% (w/v)	-3.81407	0.016461
asmA	Vancomycin 40 µg/mL	-3.80742	0.003837
yicH	M63 no sulfur 5 mM L-Cys	-3.802041	0.000924
yhdP	Bacitracin 300	-3.799139	0.011355
ydbH	M9min acetate 0.6% (w/v)	-3.738432	0.003334
ytfN	Bacitracin 1 mg/mL	-3.734257	0.001446
asmA	L-Aspartic acid N	-3.719118	0.001127
ytfN	Rifampicin 4 µg/mL	-3.678598	0.007328
yicH	M63 no sulfur 5 mM sulfate	-3.664147	0.001119
yhdP	Vancomycin 20	-3.661653	0.026037
asmA	Bleomycin 0.1	-3.597866	0.04161
asmA	CHIR090 Cefsulodin B155din	-3.594977	0.011734
yhdP	Erythromycin 5	-3.592291	0.023304
yicH	Norfloxacin 0.04	-3.527352	0.027916
ytfN	Aluminum chloride 5 mM	-3.504	0.00657
yicH	Urea 750 mM	-3.488737	0.028855
asmA	Spiramycin Trimethoprim	-3.432954	0.022733
asmA	Chloramphenicol 0.005 mg/mL	-3.409091	0.010136
asmA	Ethanol 6% (w/v)	-3.405148	0.022286
asmA	NACl RT	-3.318988	0.020076
yhdP	Indole 2 mM	-3.292412	0.025707
asmA	Triclosan 0.05	-3.263049	0.04663
asmA	L-Malic C	-3.262604	0.005356

ydbH	A22 Pyocyanin	-3.241249	0.024875
ytfN	Doxycycline Clarithromycin	-3.224438	0.021009
asmA	Tetracycline 0.0008 mg/mL	-3.220533	0.011299
asmA	No salt UV	-3.206992	0.036851
asmA	No salt RT	-3.191469	0.021113
asmA	No salt Bile salts	-3.190418	0.01723
asmA	Nalidixic 0.006 mg/mL	-3.185573	0.010555
asmA	Clindamycin 64 μg/mL	-3.185321	0.032388
asmA	D-Alanine N	-3.184195	0.005187
asmA	D-Gly L-Asp N	-3.175758	0.005785
asmA	NACl UV	-3.174552	0.020517
asmA	Minocyclin E 0.5	-3.167172	0.029024
ytfN	Doxycycline Clarithromycin	-3.148263	0.036006
asmA	Trimethoprim Cephalexin	-3.144918	0.02992
asmA	NACl 171 mM	-3.095874	0.033681
asmA	Con Media PAP 14	-3.068363	0.021789
yhdP	Ampicillin 4 μg/mL	-3.057514	0.008168
ytfN	holomycin 2.5 μg/mL	-3.056726	0.034891
asmA	Tetracycline 0.0004 mg/mL	-3.055565	0.024436
ydbH	Pyocyanin 10	-3.022701	0.012997
asmA	Doxycycline Clarithromycin	-2.9877	0.040061
asmA	D-Lactate C	-2.966129	0.016657
yhdP	D-Ribose C	-2.959157	0.011303
yhdP	Syringaldehyde 10 mM	-2.940909	0.025272
yhdP	D-Lactate C	-2.893548	0.020897
asmA	Tigecycline 10	-2.88513	0.044345
asmA	D-Serine C	-2.875639	0.023507
asmA	Succinate C	-2.843296	0.019518
ytfN	Tigecycline 10	-2.840961	0.049795
asmA	D-Glucose-6-phosphate C	-2.827241	0.024407
asmA	Casaminos C	-2.825957	0.027053
yhdP	D-Trehalose C	-2.726892	0.024688
asmA	Acetate C	-2.70653	0.029676
asmA	Ammonium chloride N	-2.66149	0.048679
asmA	D-Gluconic acid C	-2.627956	0.043405
asmA	D-Galacturonic acid C MOPS	-2.608721	0.046522
asmA	D-Galacturonic acid C	-2.53125	0.046552
yhdP	Glycolic acid C	-2.471686	0.043904
ydbH	Ramoplanin Cefsulodin	2.611641	0.049677
yicH	M9 Glycerol	2.87257	0.021254

yicH	MOPS L-Glutamine	2.926992	0.04127
ytfN	Chloramphenicol 0.001 mg/mL	2.951591	0.024663
yhdP	No salt Fumarate	2.955223	0.042558
asmA	MOPS Fumaric acid	2.980436	0.037588
yhdP	pH 9 Cefsulodin	3.033534	0.039268
asmA	Butanol 5% (w/v)	3.036414	0.048628
asmA	Urea 750 mM	3.051284	0.046184
yicH	Gliotoxin 10 μg/mL	3.148041	0.008344
ydbH	Trimethoprim Chloramphenicol	3.15324	0.024813
asmA	M9 Glycerol 37C	3.243021	0.027061
yhdP	HNP1 37C	3.24772	0.028555
yhdP	Curcumin 200	3.278324	0.028254
yhdP	Glucose	3.286966	0.020474
yicH	MOPS AAAB	3.292939	0.018957
asmA	pH 6	3.316234	0.049985
ydbH	Paraquat 10 uM	3.33999	0.032766
ydbH	Mecillinam Cephalexin	3.418008	0.003239
yhdP	M9 Glycerol 37C	3.4356	0.015265
yhdP	A22 5 μg/mL	3.436617	0.007067
asmA	A22 2 5 μg/mL	3.447533	0.024157
ydbH	Benzalkonium Polymixin B	3.472867	0.014786
ydbH	Berberine 200	3.49576	0.007347
asmA	Indole 40C	3.498221	0.004376
yhdP	M9 Maltose 37C	3.521582	0.010207
yhdP	HNP2 37C	3.539251	0.016346
yicH	37C	3.540062	0.04998
yhdP	Fumarate 37C	3.591808	0.007696
yicH	MOPS L-Serine	3.664364	0.00496
asmA	A22 5 μg/mL	3.68254	0.00321
yicH	MOPS Glycine	3.701299	0.002344
asmA	40C	3.707597	0.009151
asmA	Bile salts 37C	3.732489	0.006418
asmA	M9 Glucose 37C	3.741828	0.004513
asmA	Chloramphenicol Rifampicin	3.746762	0.005635
yhdP	MOPS NACl	3.845082	0.001842
asmA	A22 2 μg/mL	3.874271	0.005641
yicH	M9 Maltose 37C	3.939794	0.002153
asmA	M9 Fructose	3.947322	0.000931
asmA	37C	3.956749	0.003848
yicH	M9 Fructose	4.03112	0.000649

asmA	MOPS L-Alanine	4.076895	0.000887
yicH	MOPS AAAF	4.081061	0.000873
asmA	Triclosan 0.05	4.1279	0.000832
ydbH	A22 5 μg/mL	4.13982	0.000632
asmA	Furanone 37C	4.154135	0.001651
asmA	X5 Fluorouracil 37C	4.186283	0.001809
asmA	Bipyridyl 37C	4.210664	0.001374
asmA	pH 5	4.264168	0.002456
yhdP	TMAO 37C	4.326636	0.000897
ytfN	Chloramphenicol 0.0025 mg/mL	4.329277	0.00016
asmA	Tylosin 50	4.390798	0.000433
yicH	M9 Fructose 37C	4.41489	0.000139
yicH	MOPS AAFB	4.466202	8.91E-05
yicH	M9 Glucose	4.584213	6.42E-05
asmA	M9 Fructose 37C	4.699826	3.62E-05
yhdP	M9 Glucose 37C	4.70808	9.79E-05
yicH	M9 GlcNac 37C	5.110398	3.96E-06
asmA	M9 Glucose	5.161538	3.64E-06
asmA	PAPA01 37C	5.253183	8.91E-06
yicH	MOPS Glucose	5.285522	3.33E-06
yicH	MOPS pH 5	5.435054	3.10E-07
yhdP	Chloramphenicol Rifampicin	5.516493	2.70E-06
yhdP	M9 Maltose 37C	5.646691	4.59E-07
yhdP	Indole 2 mM	5.861027	1.15E-07
yhdP	Metformin 80 mM	6.128391	2.46E-08
asmA	Cephalothin 0.01 mg/mL	6.324747	6.23E-09
yhdP	MOPS Fumaric acid	6.825061	1.20E-10
asmA	Indole 37C	7.379211	8.33E-12
yhdP	M9 Fructose 37C	8.104671	1.06E-14
yhdP	PMS Procaine	9.07205	7.14E-20
yhdP	M9 GlcNac 37C	9.296343	1.28E-19
yhdP	MOPS pH 5	9.466298	4.71E-21
asmA	pH 9 40C	9.84831	2.88E-22
asmA	MOPS pH 9	10.68569	3.44E-26