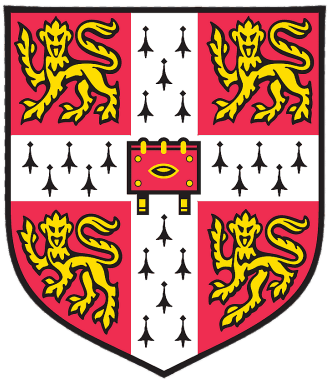
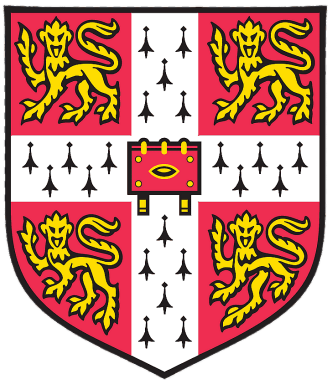
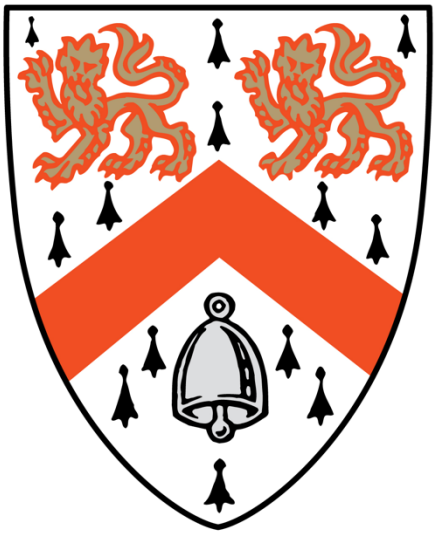


University of Cambridge



Structure & Function of Bacterial Transport Machines in their Cellular Context

Angela Mary Kirykowicz



Wolfson College

September 2023

This thesis is submitted for the degree of Doctor of Philosophy

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the Faculty of Biology Degree Committee.

Abstract: Structure & Function of Bacterial Transport Machines in their Cellular Context

Angela Mary Kirykowicz

Membranes confer cells with individual identity and capacity to regulate their response to their environment. A critical aspect of having a membranous partition is the ability to transport substances into and out of cells as part of life-sustaining functions. In pathogenic bacteria, transporters aid infection and survival in the host. Two such transporters in Gram-negative bacterial species are the MacA-MacB-TolC (MacAB-TolC) antibiotic efflux pump and the Type I Secretion System (T1SS), responsible respectively for antibiotic resistance and export of protein virulence factors. To pass the Gram-negative envelope in a one-step translocation process, both machines use a tripartite system, consisting of outer membrane protein TolC, a periplasmic adapter protein (MacA or haemolysin D (HlyD) in the T1SS), and an inner membrane protein (MacB or haemolysin B (HlyB) in the T1SS). Both use the power of ATP-hydrolysis to export their substrates. Here, I utilise computational and experimental approaches to elucidate the mechanism of function for both machines. I conduct molecular dynamics (MD) simulations of membrane embedded HlyB component of the T1SS with and without its haemolysin A (HlyA) substrate as *in silico* experiments. I also conduct MD simulations with and without substrate for a related peptidase. I show that substrate recognition is via conserved charge-charge interactions. I also show that HlyB has an asymmetric preferential interaction with cardiolipin when its substrate is present, which is not seen in the peptidase simulations. I propose that this preference is part of the mechanism of transport, with cardiolipin providing energy via the proton-motive force. I test this hypothesis through flow cytometry detection of labelled substrate trapped T1SS in a mixed population of cells, by comparing parental MG1655 *Escherichia coli* with a cardiolipin deficient MG1655 strain. I found that the cardiolipin deficient strain has reduced T1SS levels compared to its parent. To aid structural studies, I optimise the expression of the T1SS using a flow cytometry based sequential design strategy where conditions are iteratively tested via detection of substrate trapped T1SS and updated until no more improvement can be made. I also test purification strategies for single-particle cryo-electron microscopy studies. Finally, I apply further bioinformatic approaches and synthesise my computational and experimental

results to propose a mechanism of transport and suggest future experimental tests. I conduct MD simulations of MacB in membrane with and without a trapped lipid. I show that this trapped lipid locks MacB into an open state, allowing for substrate entry into the pump. I contextualise the results by comparing MD simulations to MacB-like structures and propose a revised mechanism of transport as a function of its free-energy landscape. Lastly, I explore the use of cryo-electron tomography (cryo-ET) as a method to obtain *in vivo* structural insights. I show that the use of "ghost" partially lysed *E. coli* can produce high-contrast specimens for tomography. I collect a tomographic dataset of "ghost" MacAB-TolC containing cells and apply subtomogram averaging. Preliminary results suggest that MacAB-TolC forms an array in cells, and that MacB is structurally flexible, likely in its nucleotide-binding domain. Together, these studies of the MacAB-TolC efflux pump and the T1SS shed light on their function and suggest new avenues of research to explore in order to fulfil the goal of finding novel inhibitors.

Table of Contents

Acknowledgements

This has been a long four years and there are many people to thank for taking me to the finish line. Firstly, I would like to thank my primary supervisor, Ben Luisi, for providing me with the encouragement to pursue my interests – I am grateful that you were always willing to listen and be available. Thanks to members of the Luisi lab, many of which have since left: Saiful Islam, Laura Shen, Kasia Bandyra, Tai Yang, Tom Dendooven, Giulia Paris, Kai Katsuya-Gaviria, Miao Ma, Emmanouela Petsolari, Andrzej Harris, Yvette Ntsogo, and Kotryna Bloznelyte. Thanks Emma for reading some chapters and offering helpful comments! I am sorry I was not around for many of your leaving parties! Thanks as well to Joanna Cerveira at the Department of Pathology for showing me how to use the flow cytometer. Thanks as well to our collaborator Lutz Schmitt at the University of Düsseldorf, and Schmitt lab members, Sakshi Khosa and Eymen Hachani, who took the time to read this thesis and offer helpful comments. Also many thanks to Syma Khalid for providing advice at the start of my MD journey, and for providing helpful comments on my MD chapters.

I would also like to thank members of the electron Bioimaging Unit at the Diamond Light Source, where I spent the last two years of this journey. In particular, thanks to Craig McGregor-Chatwin for always having time to show me the ropes and navigate the bureaucracy – I would have been completely lost without your kindness. Also thank you to my second supervisor, Peijun Zhang, for having me in the unit, and to other team members at eBIC and the Zhang lab at the University of Oxford.

Lastly, I wouldn't be here without the loving support of my parents and partner, Jack, and his family. My parents have always supported my dreams and helped me move countries (more than once!) – this thesis is dedicated to you. I am so grateful, Jack, that you (literally) walked into my life – I've had the best two years of my PhD with you, and we've had so many adventures. Here's to more adventures to come! Finally, thanks to my landlord, Andrew, for keeping me entertained with his yarns about beekeeping, and the many friends I've made since moving to Didcot.

I would also like to acknowledge my funders, the Diamond Light Source and European Research Council. Thanks as well to Cambridge Spark for sponsoring my courses in data science and machine learning, without which much of this thesis wouldn’t exist!

List of Abbreviations

2D	Two-Dimensional
2xYT	Yeast Extract Tryptone
3D	Three-Dimensional
ABC	ATP-Binding Cassette
AcrA-AcrB-TolC	AcrAB-TolC
ADP	Adenosine Diphosphate
AMPPNP	Adenylyl-Imidodiphosphate
ATP	Adenosine Triphosphate
BLAST	Basic Local Alignment Search Tool
CCD	Charge Coupled Device
CDL2	Cardiolipin
CLD	C39-like Domain
CLEM	Correlative Light Electron Microscopy
Cryo	Cryogenic
Cryo-EM	Cryo-Electron Microscopy
CTF	Contrast Transfer Function
β-DDM	n-Dodecyl-beta-Maltoside
α-DMP	n-Dodecyl-alpha-D-Maltopyranoside
DNA	Deoxyribose Nucleic Acid
DPPC	Dipalmitoyl-phosphatidyl-choline
DQE	Detective Quantum Efficiency
E. coli	*Escherichia coli*
eGFP	engineered Green Fluorescent Protein
ET	Electron Tomography
FIB	Focused Ion Beam
FSC	Fourier Shell Correlation
FT	Fourier Transform
FT^{-1}	Inverse Fourier Transform
GPU	Graphics Processing Unit

Hly	Haemolytic Factor
HlyA	Haemolysin A
HlyB	Haemolysin B
HlyC	Haemolysin C
HlyD	Haemolysin D
IM	Inner Membrane
kDa	kilo Dalton
LB	Luria-Bertani
LP	Lipolyl
Lpp	Braun's Lipoprotein
MacAB-TolC	MacA-MacB-TolC
MD	Molecular Dynamics
MSA	Multiple Sequence Alignment
MW	Molecular Weight
NBD	Nucleotide Binding Domain
NMR	Nuclear Magnetic Resonance
NN	Neural Network
NPT	Number of Particles, Pressure, and Temperature
NVT	Number of Particles, Volume, and Temperature
OM	Outer Membrane
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PCA	Principal Component Analysis
PCAT	Peptidase-Containing ATP-Binding Cassette Transporter
PDB	Protein Data Bank
PE	Phosphatidylethanolamine
PEP	C39 Cysteine Peptidase
PG	Peptidoglycan
P_i	Inorganic Phosphate
POPE	Palmitoyloleoyl-phosphatidyl-ethanolamine
POPG	Palmitoyloleoyl-phosphatidyl-glycerol
PrtD	Protease Transporter

PSF	Point Spread Function
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RTX	Repeat in Toxin
T1SS	Type I Secretion System
TBS	Tris Buffered Saline
TEM	Transmission Electron Microscopy
TMD	Transmembrane Domain
SDS	Sodium Dodecyl Sulphate
SGLT	Sodium-Glucose Transporters
SNR	Signal-to-Noise Ratio

Chapter I: Uncovering Functional Dynamics in Membrane Protein Transporters

1. Membranes as the Functional Unit of Life

One of the fundamental requirements of life is the ability to create a partition to sustain individual identity and to regulate exchange between the living organism and the Universe it inhabits – in cells, this is achieved through the biological membrane (Watson, 2015). Membranes are typically composed of phospholipids, which through their amphiphilic chemical nature spontaneously form self-assemblies in water which can act to encapsulate metabolic reactions – this is achieved through the separation of polar ("water-loving") phosphate head groups from hydrophobic ("water-hating") fatty acid tails by a physical process known as the hydrophobic effect (Tanford, 1978). Although a precise and complete definition of life remains elusive (Gómez-Márquez, 2021), it seems reasonable to suggest that self-assembly and the ability to regulate what goes in and out of a cell is a prerequisite for carrying out life-defining functions such as metabolism, reproduction, and mechanical/chemical work (Fernandez-de-Cossio-Diaz & Vazquez, 2018).

The ability to shuttle molecules through a cell membrane is carried out by molecular machines known as transporters (*e.g.*, Drew & Boudker, 2016). Before reviewing specific transporters, it is worth considering how, in principle, such transporter nanomachines could operate. Firstly, due to the nature of the biological membrane, the transporter must at least be able to interact with the hydrophobic fatty-acid tails. Secondly, the nanomachine must be able to form a pore or channel through the membrane to allow the shuttling of molecules. By considering these two principles, we can see that at a minimum the transporter must be able to form a small channel with a hydrophobic shell (to interact with the fatty-acid tails) and a polar core. In this case, although diffusive transport of small polar molecules would be possible, there is no discrimination by the channel in the direction of transport. To have a direction of transport, the channel must undergo conformational changes to allow for directional substrate passage. Jardetzky (1966) first proposed a simple model for sodium ion

transport, in what is now known as the alternating access model of transport. Jardetzky (1966) proposed that directional transport occurs as long as three conditions are met: 1) the transporter has a cavity large enough to admit the target substrate, 2) the transporter has the ability to assume two different conformations, conforming to an open or a closed state, and 3) the transporter has a binding site for the target substrate, the affinity for which is different in the two conformations. These ideas were expanded upon by Klingenberg (1981) to consider the role of oligomeric structure in the transportation process. The alternating access model is supported by structural evidence, for example galactoside transporter LacY (Smirnova, Kasho & Kaback, 2018), zinc transporter Yiip (Lopez-Redondo *et al*, 2018), and the SemiSWEET glucose transporter (Latorraca *et al*, 2017). The alternating access model is also known as a rocker-switch, such as the case with LacY transport, however an updated case has been included and is known as the elevator mechanism, such as the case for the glutamate transporter homologue Glt_{Ph}. In the latter case, the transporter undergoes lateral movement through the membrane in addition to conformational changes to allow for directional transport (Ryan & Vandenberg, 2016).

There are many structural examples for a variety of transporters, with different sizes and conformation dynamics (**Figure 1.1-1**). For ease of reference, I have separated transporters into five different examples (**I – V**) which we will now discuss in detail.

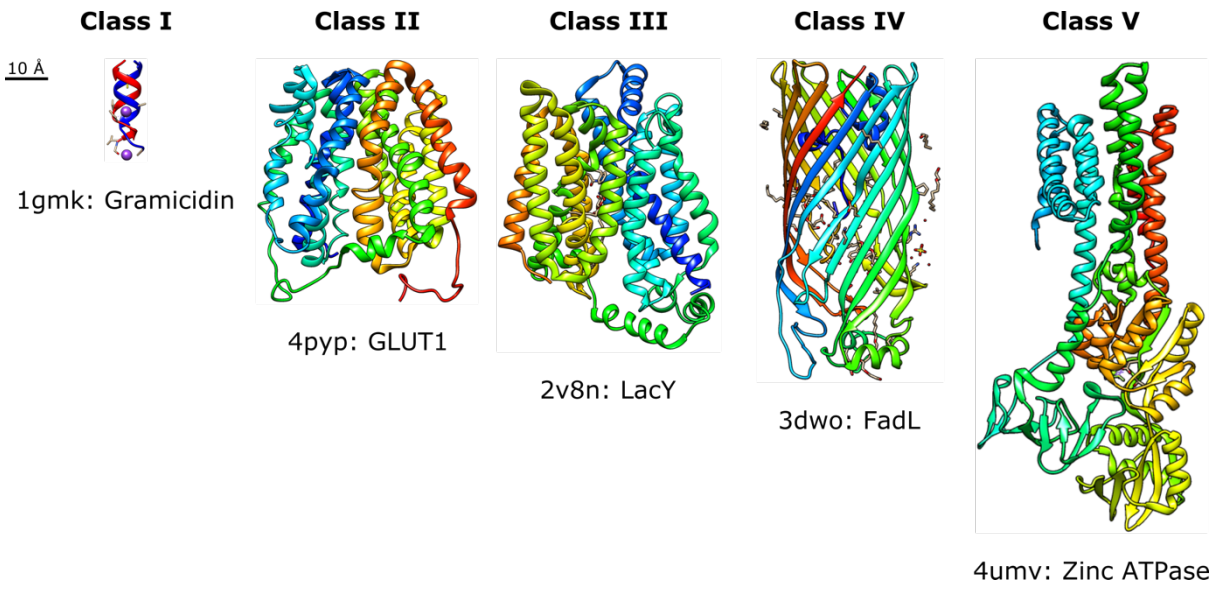


Figure 1.1-1. Transporter example structures. Example I transports are the simplest case, with the ability to transport small molecules: *Brevibacillus brevis* antibiotic Gramicidin is shown as an example (pdb ID 1gmk) (Doyle & Wallace, 1997). Example II transporters can also transport small molecules, although they are larger and can undergo conformational changes to alter selectivity: the human glucose transporter GLUT1 is shown as an example (pdb ID 4pyp) (Deng *et al*, 2014). Example III transporters can also undergo conformational changes during transport, in addition to being able to set up gradients to power transport: the *Escherichia coli* lactose permease LacY is given as an example (pdb ID 2v8n) (Guan *et al*, 2007). Example IV transporters can transport hydrophobic molecules, as given by the bacterial long-chain fatty acid transporter FadL from *Pseudomonas aeruginosa* (pdb ID 3dwo) (Hearn *et al*, 2009). Finally, Example V transporters can adopt multiple conformational states during the transportation process, with transport powered by ATP-hydrolysis: the E2P state of the Zinc transporter P-type ATPase from *Shigella sonnei* is given as an example (pdb ID 4umv) (Wang *et al*, 2014).

2. Examples of Transporters

2.1 Example I: Transport of Small Molecules

At the basic level of organisation, a transporter can be thought of as a conduit which allows diffusion of polar substrates. Aquaporins are a natural example of such a transporter; they assemble to form a channel through the membrane. Their ubiquitous occurrence in species from the bacteria, plant, and animal kingdoms underscores their functional and biological importance. Aquaporins usually transport water molecules with low-affinity but can also transport small charged ions and glycerol (Verkman, 2013). Due to their importance in controlling cell water homeostasis, many different types of aquaporins are expressed in different mammalian tissues (Takata, Matsuzaki & Tajika, 2004). Transporter selectivity is achieved through pore diameter as well as stereo- and electrochemical restraints which allows for the passage of water without proton shuttling, thus preserving the cell's electrochemical gradient (Tajkhorshid *et al*, 2002).

Another example are bacterial ionophores which passively allow the transport of ions to maintain electrochemical homeostasis. Ionophores consist of a hydrophobic shell and hydrophilic core which allows for the free passage of cations (Freedman, 2012). Several ionophores can act as antibiotics – in its natural role, valinomycin shuttles K^+ and H^+ ions to the inside of the bacterial cell and exudes its antimicrobial function by dissipating essential transmembrane electrochemical gradients (Stillwell, 2016). Gramicidin is another antimicrobial ionophore, a natural product of *Bacillus brevis* against Gram-positive bacterial species; it forms a restrictive 4 Å pore with a selectivity for monovalent cations (see **Figure 1.1-1**). Due to the pore size, cations can only move through the pore in a single-file mode of transport. There is a continuous flow of hydrogen bonded water molecules through the pore, and the cations are transported by hitching a ride on the column (Freedman, 2012).

Another example is urea transporters which are members of the solute carrier family 14 (SLC14) involved in passive transport of urea across cell membranes. The solved structure of a bacterial homologue of this transporter type from *Desulfovibrio vulgaris* showed that it is

trimeric and transports dehydrated urea in a single-file manner. Selectivity is imparted by the 16 Å central pore which contains highly conserved amino acids which form an oxygen ladder through the pore, with nitrogen-containing amino acids at either end. This could allow dehydrated urea to hydrogen bond as it moves through the pore and also aids in selectivity in this otherwise continuously open channel (Levin, Quick & Zhou, 2009).

2.2 Example II: Transport of Polar or Charged Molecules

In the next class, we have transporters which work to move polar or charged molecules across the cell membrane, and which often involve conformational changes to direct transport. Glucose transporters are an example of this class, particularly the facilitative glucose transporters (see **Figure 1.1-1**) (Navale & Paranjape, 2016). Affinities for glucose/fructose differ depending on the transporter type, although they tend to contain conserved charged/polar/hydrophobic residues in their core for selective substrate uptake (Barrett, Walmsley & Gould, 1999; Gorovits & Charron, 2006).

Neurotransmitter Sodium Symporters (NSS) are another example; they function to selectively transport substrate and Na^+ ions into the cell and are crucial for removing neurotransmitters from the synapse. Studying a bacterial homologue of this transporter, LeuT, which functions to import the amino acid leucine, established that Na^+ binding allows for recruitment of substrate; once substrate binds this causes domain movements to allow both Na^+ and substrate to enter the cell. Two charged residues (aspartic acid and arginine) act as a gate to the transporter; upon Na^+ and substrate binding they form a salt bridge to prevent reverse transport, along with exit occlusion by two hydrophobic amino acids (tyrosine and phenylalanine). Domain movements then likely positions the charge gate to allow for the funnelling out of Na^+ and substrate (Yamashita *et al*, 2005; Krishnamurthy & Gouaux, 2012).

2.3 Example III: Charged Gradients and Importers/Exporters

In Example III, a gradient can be utilised to import/export different substrates. The bacterial lactose permease is an excellent example; LacY is a Major Facilitator Superfamily member which transports lactose by using a proton gradient (see **Figure 1.1-1**). There is a limited set

of conserved charged residues for transport, which act to facilitate dual transport of galactosides and protons, with the proton gradient providing energy for transport against the substrate concentration gradient. Substrate transport directionality is controlled by a transporter conformational change after substrate binding (Kaback, 2005).

Another example is the arginine:agmatine antiporter, AdiC, which allows bacteria to survive acidic environments by decarboxylating arginine and glutamate and exchanging these intracellular reaction products (γ-aminobutyric acid and agmatine) with extracellular arginine and glutamate. The central pore is the most conserved region consisting of hydrophobic, polar, and charged residues. Substrate binding induces a conformational change from outward-facing to inward-facing; in this state, agmatine can outcompete arginine for binding and leads to the substrate-bound occluded conformation where agmatine can be released into the periplasm. Differences in arginine and agmatine binding affinities is thought to be due to a glutamic acid sensor residue on AdiC which undergoes deprotonation in the switch from periplasm to cytoplasm pH (Gao *et al*, 2009). Later work proposed that the structure undergoes a symmetrical switch between inward- and outward- facing conformations (Kowalczyk *et al*, 2011).

Sodium-glucose transporters (SGLT) are an example of utilising dual substrate import to actively move glucose into cells. First, sodium is extruded by a sodium/potassium ATPase which creates an electrochemical gradient of sodium ions. Concomitant downhill movement of sodium drives uphill movement of glucose against its concentration gradient via co-transport using the sodium-glucose transporter. If the imported glucose is not metabolised or moved from intestinal tissues, it is then exported by facilitative glucose exporters mentioned in **Example II**. This allows for efficient uptake of dietary glucose and retention in kidney tissues to limit urinary loss (Brown, 2000). Recently, a thermostable SGLT1 was engineered, and its structure determined by cryo-electron microscopy (cryo-EM); the transport pathway is highly conserved and hydrophilic; a presumed glucose-binding pocket was found on the cytosolic side of the protein, consisting of polar, charged, and hydrophobic amino acids. Interestingly, the authors of the study found that water could traverse the same path as glucose, although the water transport activity is not linked to glucose transport, with

extracellular gate residues allowing the passive passage of water but not glucose (Han *et al*, 2022).

2.4 Example IV: Hydrophobic Molecule Transporters

Although it is often assumed that small hydrophobic molecules can cross the cell membrane through diffusive processes, specific transporters of hydrophobic substrates do exist. For transporting lipids, lipid carrier proteins have a range of mechanisms for transport, including forming box-like shuttles, tubes, and bridges to cross different membranes (Wong, Gatta & Levine, 2019). However, due to the nature of the substrates transported and inherent difficulties in isolating and characterising membrane proteins, exact mechanisms of transport often remain elusive (Claus, Jezierska & van Bogaert, 2019).

In Gram-negative bacteria, long-chain fatty acids are transported by FadL which forms a barrel-like pore through the membrane. Intriguingly, the structure of FadL has an N-terminal extracellular domain which resembles a lid (see **Figure 1.1-1**). The hydrophobic substrates are channelled via an extracellular groove towards a high-affinity binding pocket consisting of conserved charged and hydrophobic residues to interact respectively with the head and tail regions (van den Berg *et al*, 2004). Structurally, FadL is similar to siderophores and outer-membrane porins which associate with proteins that provide transport energy via the proton-motive force and/or ATP-hydrolysis. Siderophores transport chelated metals by shuttling metals from the outer-membrane protein barrel-like structures to periplasmic partners which use a proton-motive force to provide energy in the first step. Metals are then shuttled from the periplasmic partners to awaiting ABC-transporters on the cytoplasmic side which then use the hydrolysis of ATP to import the metals (Ferguson & Deisenhofer, 2004). Association of porins with periplasmic and ABC-transporter partners also allows for the import/export of a diversity of substrates, under the power of ATP-hydrolysis, including antibiotics (Zeth & Thein, 2010; Prajapati, Kleinekathöfer & Winterhalter, 2021). This highlights how transporters adapted for transport by diffusion can easily alter function through association with ATP-powered protein modules, a defining feature in **Example IV**.

Another example are transporters which facilitate energetically unfavourable reactions. The class of F and V type ATPase enzymes, which couple proton transport to ATP-synthesis is a good example. F and V ATPases form a stalk-like structure composed of multiple proteins; proton shuttling down its electrochemical gradient through one part of the stalk drives a rotation mechanism of the head region which synthesises ATP. Hydrolysis of ATP therefore rotates the head in the opposite direction and shuttles protons against their concentration gradient. Hence, these ATPases neatly show the power of transport in facilitating metabolic reactions (Nakanishi-Matsui *et al*, 2010).

2.5 Example V: Powered Transporters

In Example V, the transporter must use power from ATP-hydrolysis in order to move substrates across the cell. This is the case for Na^+/K^+ pumps which require energy from ATP in order to move those ions across their concentration gradients. These pumps have specific Na^+ and K^+ binding sites at either end of the transporter, allowing for selective export of Na^+ in conjunction with selective import of K^+. This is vital for creating a voltage across the membrane which allows for formation of action potentials (Ratheal *et al*, 2010). Na^+/K^+ pumps are part of a larger family of transporters known as the P-type ATPases which function to pump ions or toxins outside of cells. They consist of highly conserved cytoplasmic core and six to ten hydrophobic membrane spanning helices (Kühlbrandt, 2004). The cytoplasmic core consists of phosphorylation, nucleotide-binding, and actuator domains which carry out ATP-hydrolysis and autophosphatase activity. During their functional cycle, they exhibit two enzymatic states (E1 and E2) which have different affinities for the respective transported ions, Na^+ and K^+. These states are cycled through under active transport, requiring a high ATP to ADP ratio in order to cycle from the E1 substrate binding state to the E2 export state. In the E1 state, the transporter has a high-affinity for the exported ion and autophosphorylation after ATP-hydrolysis then allows for a switch to the E2 state where the ion is released under low binding affinity. Dephosphorylation is then coupled to counterion binding and a switch back to the E1 state where counterions are released into the cytosol (Dyla *et al*, 2020). See **Figure 1.1-1** for an example structure of a P-type ATPase.

As mentioned in the case of **Example IV** porins, ABC-transporters are ubiquitous ATP-powered exporters/importers. Structurally, they form dimers consisting of a conserved ATP-binding cassette (ABC) nucleotide-binding domains (NBD) and transmembrane domains (TMD). Other modules can be fused to this basic architecture in order to alter function. The NBD consists of several conserved signatures – a P-loop/Walker A motif (GXXGXGK(S/T)), a Walker B motif (ϕϕϕϕD, where ϕ is hydrophobic), a Q-loop, a H-motif/switch region, and an LSGGQ alpha-helical motif. ATP-binding occurs between the dimers, and thus NBD dimerisation is prerequisite for hydrolysis. The TMD consists of six helices per protomer which adopt different folds depending on their function. From this base architecture, a range of substrates can be transported including nutrients, amino acids, sugars, metals, lipids, fatty acids, and cholesterol (Rees, Johnson & Lewinson, 2009). As they are found in all domains of life, ABC-transporters show a range of different types; supposition based on architecture would suggest a similar mechanism of transport (depending on transport directionality), although exact mechanisms are under continual investigation (Thomas & Tampé, 2020).

3. Transporters in Disease

As we have seen from above, there are many different types of transporters involved in a range of cellular processes. It also shows the inherent difficulties in classification: should we classify transporters by structural similarity (as is done for ABC-transporters and P-type ATPases), or by functional similarity (as is done for amino acid transporters such as LeuT and AdiC)? The above exercise allows us to view these flavours of transporters without imposing a particular bias, and already allows us to see how different elements can be combined to transport a variety of substrates.

The next natural question to ask is: how is transport related to disease? Evidently, it can be seen that a dysregulation in transport is likely to be linked to disease states, and this is the case for urea transporters (Klein, Blount & Sands, 2012), sodium-glucose transporters (Poulsen, Fenton & Rieg, 2015), amino acid transporters (Yahyaoui & Pérez-Frías, 2020), P-type ATPases (Bublitz, Morth & Nissen, 2011), and ABC-transporters (Vasiliou, Vasiliou & Nebert, 2009) (see **Table 1.3-1**). Transporters can also aid in bacterial disease pathogenesis, by exporting virulence factors and allowing for nutrient uptake – the ABC-transporter class is a well-known example (Akhtar & Turner, 2022).

Table 1.3-1. Transporters with roles in human disease.

Transporter	Disease Role in Humans	Reference
Urea Transporter (UT or SLC14A)	Role in diabetes in humans. Increase in urea excretion upon development of Type I *diabetes mellitus* in rats, with a decrease in UT-A1 abundance.	Klein, Blount & Sands (2012)
Sodium-Glucose Transporter (SGLT)	SGLT is involved in glucose absorption in the intestine. Inhibition of SGLT used as a treatment for Type I and II *diabetes mellitus* to improve glycaemic control.	Poulsen, Fenton & Rieg (2015)
Amino Acid Transporter (SLC)	Associated with inherited metabolic disease, including early infantile epileptic encephalopathy, deafness (autosomal 25), and cystinuria.	Yahyaoui & Pérez-Frías (2020)
P-Type ATPases	Dysfunction of SERCA, NKA, or Cu^{+} ATPases involved respectively in heart failure, rapid-onset dystonia parkinsonism or Wilson disease.	Bublitz, Morth & Nissen (2011)
ABC-Transporter	Mutations in 11 ABC transporters linked to disease *e.g.* cystic fibrosis, Stargardt disease, X-linked adrenoleukodystrophy, Dubin-Johnson syndrome.	Vasiliou, Vasiliou & Nebert (2009)

Bacterial secretion systems are a diverse set of transporters which export various substrates; they are broadly classified into one- or two-step translocons across nine different types. In Gram-negative bacteria, two-step translocation occurs via periplasmic intermediates while in one-step translocation the substrate is exported across the double membrane (Christie, 2019). Secretion systems have been implicated in virulence and disease. For example, the Type III secretion system acts to inject virulence factors directly into host cells, resembling a tiny molecular syringe (Deng *et al*, 2017). Mycobacterial infection of human hosts is facilitated

by the Type VII secretion system, which forms a unique architecture to secrete folded effector proteins (Famelis *et al*, 2019; Bunduc *et al*, 2021). The Type VI secretion system facilitates bacterial competition by allowing bacteria to inject effectors which eliminate competitors; it has been implicated in allowing the El Tor strain of *Vibrio cholerae* to become the dominant strain in cholera pandemics since the 1960s (Kostiuk *et al*, 2021). The Repeat in Toxin (RTX) family is a class of secreted protein effectors which aid in bacterial virulence (Linhartová *et al*, 2010). They are of particular interest as they are found to be secreted by human pathogens in a range of diseases, for example by *Escherichia coli* (Menestrina *et al*, 1994), *Vibrio cholerae* (Lin *et al*, 1999), *Vibrio parahaemolyticus* (Fabbri *et al*, 1999), and *Bordetella pertussis* (Benz *et al*, 1994) which respectively cause urinary-tract and peritoneum infections, cholera, gastrointestinal infections, and whooping cough. RTX members are secreted by a common mechanism, consisting of an inner membrane ABC-transporter, a periplasmic adapter protein, and outer membrane protein; together they form a membrane-spanning channel known as the Type I Secretion System (T1SS) (Hodges *et al*, 2022).

We will now focus on reviewing our current knowledge of the T1SS and how it functions to secrete its substrates.

4. The Type I Secretion System – A History

In the 1950s, it was noticed that bacteria-free fluid isolated from *Escherichia coli* that were harvested from human urinary tract infections or faeces could lyse red blood cells and promote illness when injected into mice (Robinson, 1951). Later it was found that this haemolytic factor (Hly) resided in genetic elements which could be transferred to non-pathogenic strains (Smith & Halls, 1967; Goebel & Schrempf, 1971). In the 1980s, DNA encoding the Hly factor was isolated and added recombinantly to avirulent *E. coli* strains, which were able to cause rat death in a peritonitis infection model (Welch *et al*, 1981; Welch & Falkow, 1984). For successful transport of the Hly factor, named HlyA, it was found that two genetic elements on the same Hly plasmid were required, named $HlyB_a$ (later renamed HlyB) and $HlyB_b$ (later renamed HlyD), with defects causing blocked secretion (Wagner, Vogel & Goebel, 1983; Härtlein *et al*, 1983). It was also found that HlyA required modification by HlyC to become haemolytically active, with suggestions that HlyC proteolytically processes HlyA (Härtlein *et al*, 1983). However, work by Nicaud *et al* (1985) found that HlyA does not undergo proteolysis to become active; rather HlyC post-translationally modifies HlyA, with both active and inactive forms being secretion competent by HlyB and HlyD. Secretion competence was then isolated to the C-terminal domain of HlyA (Nicaud *et al*, 1986). Isolation of HlyB and HlyD components (Mackman *et al*, 1985) in the context of unfolded HlyA C-terminal secretion established that HlyB couples ATP-hydrolysis to translocation while HlyD resides in the periplasm (Gray *et al*, 1989). It was later found that outer membrane protein TolC is also required for HlyA export (Wandersman & Delepelaire, 1990). An export model was proposed, based on *in vivo* cross-linking experiments, whereby TolC forms a transient complex with HlyB and HlyD for one-step translocation of HlyA (Thanabalu *et al*, 1998) (**Figure 1.4-1a**).

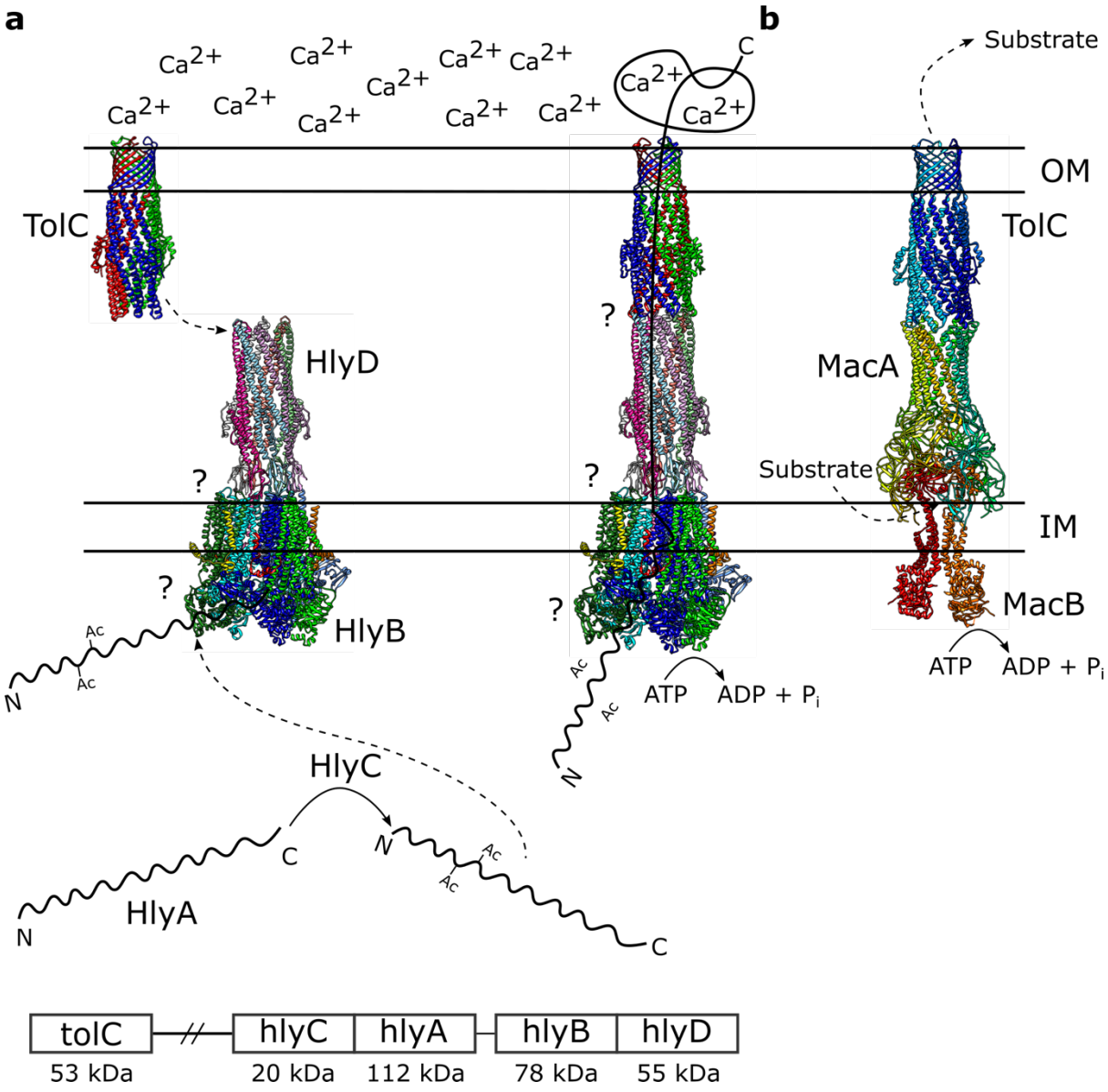


Figure 1.4-1 (previous page). Model of export of haemolysin A. a) Activation of toxic haemolysin A (HlyA) requires acylation of K564 and K690 by HlyC. HlyA then interacts with HlyB, probably in the C39-like domain (Lecher *et al*, 2012). Interaction of HlyB and HlyD with HlyA triggers recruitment of TolC (Balakrishnan, Hughes & Koronakis, 2001), allowing for full secretion assembly. HlyA is then secreted by action of ATP-hydrolysis in the unfolded state (Benabdelhak *et al*, 2003). High extracellular concentration of Ca^{2+} aids correct folding of HlyA upon export through binding to GG loops in β-sheet forming GG repeat motifs (Sánchez-Magraner *et al*, 2010). *In situ* structures for HlyC, HlyD, HlyB, and HlyA are not known (see **Table 1.5-1**). HlyD forms a hexameric arrangement (Lee *et al*, 2012; Kim *et al*, 2016) with three dimer-pairs of HlyB (Zhao, Lee & Chen, 2022) to form a bridge with TolC for export (Pimenta *et al*, 1999); missing membrane-bound N and C-terminal amino acids likely interact with HlyB. HlyB likely undergoes a conformational rearrangement to "open-up" and allow for HlyA entry into the channel in the unfolded state. As in other transport systems, TolC likely changes to "open" conformation upon interaction with HlyD (Andersen, Hughes & Koronakis, 2000). Note the gene products for

each component is not to scale. Mass of protein products are given. PDB structures used: TolC closed (1ek9), TolC open (5nik), periplasmic HlyD (5c22), and HlyB with partial HlyD (8dck). **b)** An analogous ABC-transporter is the MacA-MacB-TolC system, which transports various substrates. MacA forms a hexameric arrangement allowing interaction with TolC. MacA interacts with MacB via its β-barrel membrane proximal domains. The structured part of MacB is dimeric, although four other disordered subunits can be present during transport (Fitzpatrick *et al*, 2017). Cryo-EM structure is shown (pdb 5nik).

From this model, we can see the steps required to achieve export of haemolytically active HlyA: 1) before export, HlyA is activated by HlyC acylation of two N-terminal lysine residues, K564 and K690, with acyl carrier protein as the donor (Stanley *et al*, 1994), 2) HlyA then interacts with HlyB and HlyD via its C-terminal to initiate export (Nicaud *et al*, 1986; Mackman *et al*, 1987; Gray *et al*, 1986; Koronakis, Koronakis & Hughes, 1989; Jarchau *et al*, 1994; Lenders *et al*, 2015), 3) TolC is recruited to the complex to form the full tripartite secretion competent assembly (Wandersman & Delepelaire, 1990; Thanabalu *et al*, 1998), 4) HlyA is exported under the power of ATP-hydrolysis by ABC-transporter (Higgins *et al*, 1986) HlyB (Härtlein *et al*, 1983; Benabdelhak *et al*, 2003), 5) Upon exit from the cell, HlyA experiences an increased Ca^{2+} concentration which aids in folding through its RTX GG-motif (Felmlee & Welch, 1988; Holland *et al*, 2016; Spitz *et al*, 2019), 6) After export, the complex dissociates (Balakrishnan, Hughes & Koronakis 2001).

Comparing the T1SS to a similar transporter, the MacA-MacB-TolC efflux pump (Fitzpatrick *et al*, 2017) (**Figure 1.4-1b**), we can see some similarities in component interaction and mechanism: 1) HlyD is likely to interact with TolC in a tip-to-tip fashion analogous to the MacA-TolC interaction, 2) HlyD membrane-bound lipolyl domains interact with HlyB, as MacA β-barrel and membrane proximal domains interact with the periplasmic head of MacB, and 3) Both ABC inner membrane proteins HlyB and MacB use the power of ATP-hydrolysis to export substrates. However, we can already see some notable differences: 1) Dimers of HlyB form a trimer to interact with hexameric HlyD, while dimeric MacB interacts with hexameric MacA, 2) HlyA substrate entry route is between the transmembrane domains of one dimer (Zhao, Lee & Chen, 2022), while the substrate entry route in MacB is the periplasmic port (Crow *et al*, 2017), and 3) Full HlyB-HlyD interaction is not known, as the periplasmic region of HlyD is likely unstructured until substrate-triggered recruitment of TolC (Zhao, Lee & Chen, 2022).

5. Specifics of Export – Filling In the (Structural) Gaps

Currently, we do not have an experimental structure of the full T1SS complex. However, different components of the complex have been solved by various methods (**Table 1.5-1**). Before the HlyB/D complex was published in 2022 (Zhao, Lee & Chen 2022), only certain domains of HlyB and HlyD had been solved: the C39-like domain (CLD) and NBD of HlyB, and the periplasmic α-helical domain with the lipolyl domain of HlyD (**Table 1.5-1**). There is no solved experimental structure of the toxin HlyA, although based on other haemolytic toxins (**Table 1.5-1**) it is likely to form a higher-order oligomer in order to exert its pore-forming lytic effects (Menestrina *et al*, 1994).

Table 1.5-1. Structures of Type I Secretion System Components

Component	Organism	PDB Code	Method	Resolution (Å)	Length (Total)	Symmetry[a]
TolC	*E. coli*	1EK9	X-ray Diffraction	2.1	428 (493)	C3
HlyB NBD	*E. coli*	1XEF	X-ray Diffraction	2.5	241 (707)	C2
HlyB NBD	*E. coli*	2FF7	X-ray Diffraction	1.6	247 (707)	C1
HlyB CLD	*E. coli*	3ZUA	Solution NMR	N/A	142 (707)	C1
HlyD	*E. coli*	5C21	X-ray Diffraction	2.5	279 (478)	C2
HlyD	*E. coli*	5C22	X-ray Diffraction	2.3	279 (478)	C1
HlyB/D	*E. coli CFT073*	7SGR	Cryo-Electron Microscopy	2.90	HlyB: 707 (707); HlyD[a]: 356 (478)	C3
HlyB/D	*E. coli CFT073*	8DCK	Cryo-Electron Microscopy	3.40	HlyB: 707 (707); HlyD[c]: 478 (478)	C3
HlyA	*Staphylococcus aureus subsp. aureus Mu50*[b]	3ANZ	X-ray Diffraction	2.3	302 (319)	C7
Haemolysin	*Vibrio parahaemolyticus*	3A57	X-ray Diffraction	1.5	154 (165)	C4

[a]Based on PDB entry and does not necessarily reflect biological symmetry in full complex.

[b]There is currently no *E. coli* HlyA structure. The closest homologue with a solved structure is from Gram-positive *S. aureus*. Note there is significant divergence between the sequences, with *E. coli* HlyA significantly longer at 1024 amino acids.

[c]Note that most of the HlyD complex is unmodeled in these entries – coordinates provided for approximately residue 10–80.

The N- and C-termini of HlyA are structurally distinct: the N-terminal domain, where the lytic activity in activated HlyA resides, has a predominantly α-helical content (Valeva *et al*, 2008), whereas the C-terminal domain has mostly β-sheet content which aids folding upon export by binding Ca^{2+} ions (Bumba *et al*, 2016). AlphaFold2 structure prediction (Jumper *et al*, 2021) of *E. coli* HlyA places lower confidence of the model in the N-terminal domain (**Figure 1.5-1a**). Although it is known that the HlyA substrate is exported via its C-terminal domain, with the HlyB CLD likely playing a role in the recognition, the specifics of interaction between HlyB and HlyD are unknown (Lecher *et al*, 2012; Zhao, Lee & Chen, 2022). The CLD is not a strict requirement in RTX family member secretion; the protease transporter (PrtD) from *Aquifex aeolicus* lacks an N-terminal CLD and substrate recruitment and transport likely occurs through the transmembrane helices (Morgan, Acheson & Zimmer, 2017). In addition, testing of cross-species components of the T1SS established that HlyA does not interact with PrtD although toxin HasA from *Serratia marcescens* can. This may be related to the fact that inner membrane ABC-transporter HasB also lacks a CLD (Létoffé, Delepelaire & Wandersman, 1996). Functionally, it is not clear what role the CLD plays in substrate export (Kanonenberg, Schwarz & Schmitt, 2013).

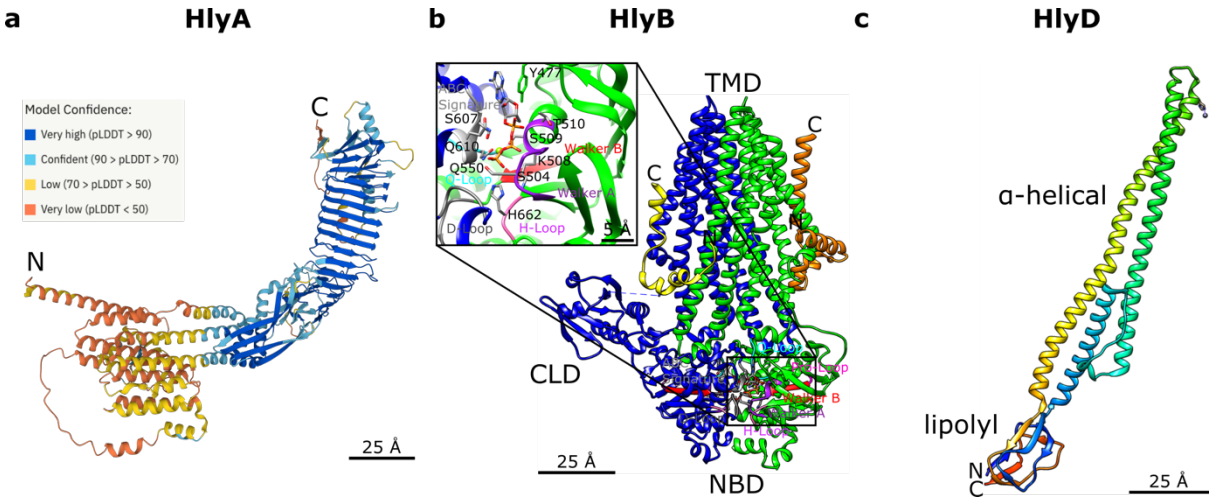


Figure 1.5-1. Gallery of Structures of the T1SS Components. **a)** AlphaFold2 prediction (Jumper *et al*, 2021) for HlyA structure shows high-confidence for the C-terminal domain containing β-sheets with Ca^{2+} binding GG-motif loops, while the α-helical N-terminal domain has less confidence. **b)** HlyB structure with bound ATP (pdb ID 8dck) (Zhao, Lee & Chen, 2022) shows ATP-binding occurs between the nucleotide-binding domain (NBD) dimers. NBD regions are highlighted: residues 502–510 (Walker A, purple), 549–556 (Q-loop, cyan), 606–610 (ABC-signature, grey), 623–625 (Pro-loop, bright pink), 626–630 (Walker B, red), 634–637 (D-loop, dark grey), and 661–663 (H-loop, pink) (Zaitseva *et al*, 2006). The two arms of HlyD are present (yellow, orange). **c)** HlyD monomer with α-helical and lipolyl domains (pdb ID 5c22) (Kim *et al*, 2016). In the HlyB/D complex, missing N- and C-terminal regions from the crystal structure wrap around HlyB (yellow, orange in **b**) with the lipolyl domain positioned on top of the HlyB TMD. Six α-helical domains interact with each other to form a substrate "basket", with TolC likely stabilising the interactions, as depicted in the model shown in **Figure 1.4-1a**.

Isolation and reconstruction by X-ray crystallography of HlyB NBDs established a Mg^{2+} enhanced asymmetry between the two, with one open and one closed phosphate exit tunnel. Like other ABC-transporters (Rees, Johnson & Lewinson, 2009), ATP-induced NBD dimerisation is a prerequisite to hydrolysis (Zaitseva *et al*, 2006). This opens the possibility of sequential ATP-hydrolysis being a requirement for function (Zaitseva *et al*, 2006), in a similar mechanism to the *Thermus thermophilus* multi-drug resistance proteins A and B (TmrAB) (Hofmann *et al*, 2019). However, it is not known how the NBDs communicate substrate export to the transmembrane helices (**Figure 1.5-1b**). One means of communication could be through the hinge-like region of the TMD at the protomer pairs through the Q-loop of the NBD. Hydrolysis of ATP in the central NBD, involving residues E631 and H662 which form a catalytic dyad, could trigger conformational changes through the Q-loop to the interface (Zaitseva *et al*, 2006). A substrate export pathway through the helices has been established.

HlyD is an essential requirement for transport; it is composed of a long α-helical periplasmic domain and a short lipolyl domain (**Figure 1.5-1c**). The lipolyl domains connect with the arm regions which wrap around HlyB (**Figure 1.5-1b**) and are missing in the crystal structure (**Figure 1.5-1c**). The α-helical domains oligomerise when TolC is present to form a substrate channel (Zhao, Lee & Chen, 2022); however, as there is no full T1SS structure, it is unknown whether a periplasmic gate exists in HlyD which aids to control substrate exit, as is the case for MacA (Fitzpatrick *et al*, 2017).

6. Outline for This Thesis

This thesis aims to elucidate the structural mechanisms of transport for the T1SS and MacAB-TolC efflux pump. Due to the broad nature of the research question, I have split the investigation into three themes: in the first theme, I utilise simulation to uncover the dynamics of transport for the T1SS (**Chapter II**) and the MacAB-TolC efflux pump (**Chapter V**); in the second theme, I utilise the results of dynamics to experimentally test hypotheses related to transport for the T1SS (**Chapter III**) and MacAB-TolC (**Chapter IV**), as well as uncover the factors which influence T1SS production in *E. coli* (**Chapter III**); finally, in the third theme, I utilise bioinformatics to further inform the results from the last two themes to propose a mechanism of transport for the T1SS as well as propose experiments to test this model (**Chapter VI**).

The research questions that guided the work of this thesis are as follows:

1) How does the T1SS recognise its substrates? (**Chapter II**)
2) What factors affect T1SS production in *E. coli* and how can I test the results from Aim 1) (**Chapter III**)
3) How can I determine an *in situ* structure for MacAB-TolC? (**Chapter IV**)
4) How does MacB transport its substrates? (**Chapter V**)
5) Incorporating all available information from this thesis, what plausible mechanism can I determine for T1SS substrate transport and how can I test this model? (**Chapter VI**)

This will be achieved with the following Objectives:

1) Conduct *in silico* experiments by utilising molecular dynamics (MD) simulations to uncover structural dynamics of transport in the T1SS and MacAB-TolC (**Chapter II** and **Chapter V**)
2) Test different expression conditions for T1SS production in *E. coli* and use flow cytometry to sort out the populations to gather relevant statistics. Correlate the results using cryo-correlative light microscopy (cryo-CLEM) and conduct purification tests to obtain *in vitro* T1SS (**Chapter III**)

3) Collect cryo-electron tomography (cryo-ET) data for MacAB-TolC and test different reconstruction strategies for high-throughput characterisation of *in situ* MacAB-TolC (Chapter **IV**)
4) Conduct a bioinformatics analyses for the T1SS to synthesize results from Objectives 1) and 2) in order to make a plausible mechanism for T1SS substrate transport. Propose some experimental tests for the model (**Chapter VI**)

By combining computational methods and experimental tests, my thesis advances knowledge on both the T1SS and the MacAB-TolC efflux pump.

Chapter II: Molecular Dynamics to Investigate Transport Mechanisms in the Type I Secretion System

1. Introduction

1.1 What is Molecular Dynamics?

Molecular dynamics (MD) is the computation of forces between all the atoms in a simulated system in order to plot a resulting trajectory (Karplus & McCammon, 2002). The basic procedure follows an iterative cycle (**Figure 2.1.1-1**), stepping through small time increments until the desired length of the simulation is complete. MD simulations are powerful methods for discovering both equilibrium and non-equilibrium properties of the molecule(s) under investigation. The first reported MD simulation was published in 1957 (Alder & Wainwright, 1957) and consisted of different system sizes of particles placed in a rectangular box with periodic boundary conditions applied (see below). For systems of biological significance, the first published MD simulation was reported nearly half a century ago (McCammon, Gelin & Karplus, 1977), which showed that it was possible to derive useful insights into biological processes *in silico*. MD simulations are now routinely used to investigate a wide-range of biological processes, including selecting lead drug candidates for a particular target (De Vivo *et al*, 2016), or exploring protein behaviour in a lipid environment (Gumbart *et al*, 2005), or the dynamics of fast folding proteins (Scheraga, Khalili & Liwo, 2007), or changes in protein conformations (Hollingsworth & Dror, 2018).

Although MD simulations are versatile in the types of scientific questions they can address, timescale limitations place upper limits on biological processes which can be simulated. Many biological processes occur in the millisecond to second timescale (Shamir *et al*, 2016), which is out of reach for typical MD simulation experiments (Sweet *et al*, 2013). A major breakthrough occurred with the use of GPU (Graphics Processing Unit) acceleration and efficient parallelisation of calculations (Stone *et al*, 2010), meaning that simulations can now routinely be conducted on the nanosecond to microsecond timescale (Kutzner *et al*, 2015; Lee *et al*, 2018; Kutzner *et al*, 2019; Phillips *et al*, 2020) rather than the previous picosecond

timescale limitation (*e.g.,* McCammon, Wolynes & Karplus, 1979; Aqvist *et al*, 1985; Paulsen, Bass & Ornstein, 1991).

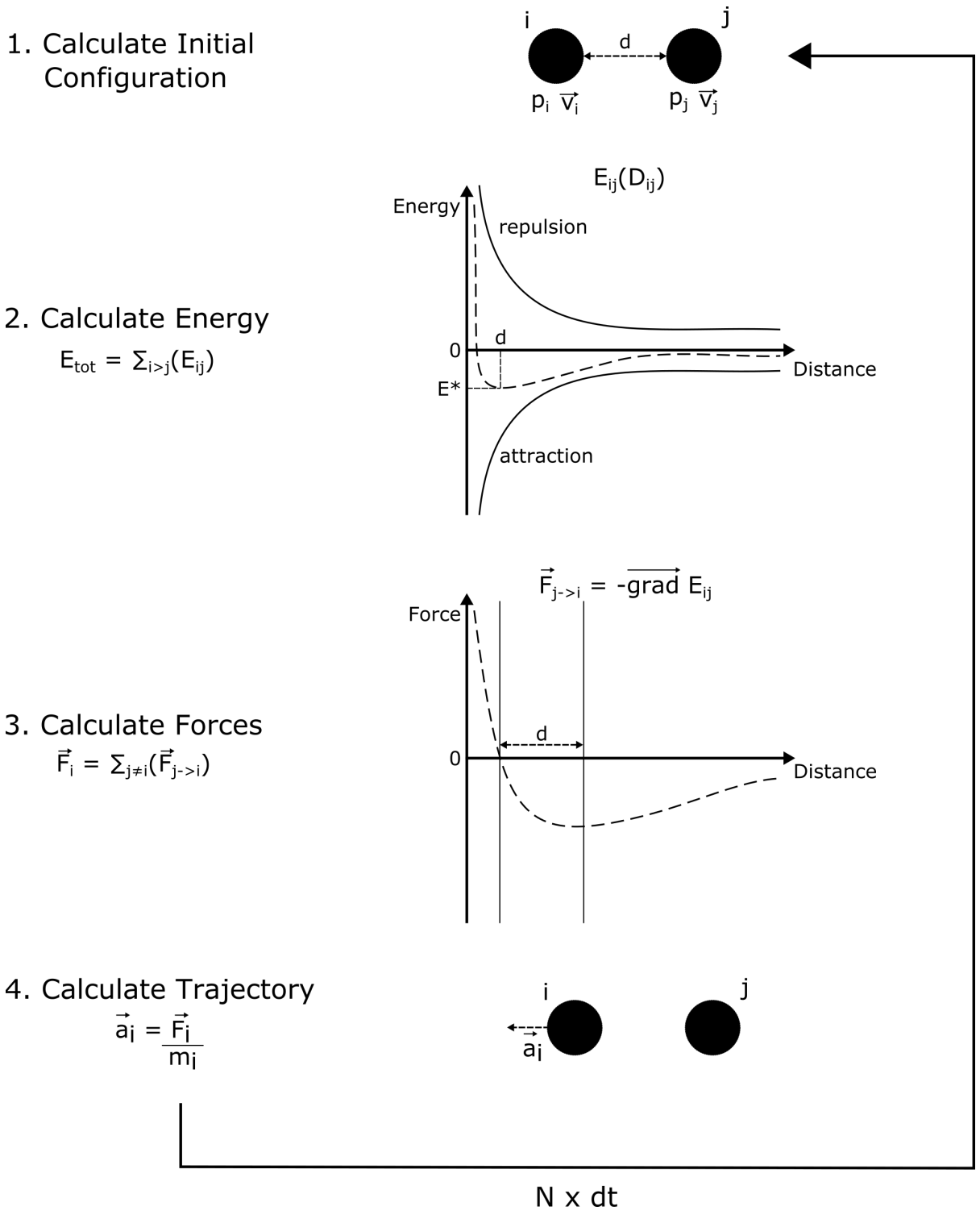


Figure 2.1.1-1. The Molecular Dynamics (MD) Cycle. An MD cycle follows four stages: 1) Initial positions (p) and velocities (v) for each atom pair (i, j) are inputted and from that, 2) the total energy (E_{tot}) for each atom is calculated based on attractive (Coulombic and van der Waals interactions) and repulsive (electronic cloud) interactions at different distances ($E_{ij}(D_{ij})$), 3) the total force (F) on each atom is calculated based on 2). For example, the force of atom j on atom i ($F_{j->i}$) is calculated by finding the distance (d) between the atoms which gives the lowest energy (E*). This is found by calculating the gradient of the energy curve for the atom pair, and

4) the final trajectory (acceleration, a) is calculated based on 3) taking into account the mass (m) of the atom. The new position and velocity of each atom then acts as the initial configuration for the next round of the cycle. The cycle proceeds iteratively for a small increment of time (dt) for N steps. Explanation adapted from Tuckerman & Martyna (2000).

Another limitation is the use of classical physics to calculate the overall forces in a molecule to plot the trajectory (Braun *et al*, 2019); this means accurate simulations of events depending on quantum effects are currently out of scope for a typical MD *in silico* experiment, ruling out processes such as enzyme catalysis (*e.g.,* Ditzler *et al*, 2010) or protein misfolding (*e.g.,* Gillet, 2022). Nevertheless, the insights gained from well-planned simulations can greatly aid experimental design and testing of the system under investigation.

Another approach that has made MD calculations tractable is the use of coarse graining, which is a method for smoothing out the system's energy landscape. Here, atoms are grouped together in the molecular representation and calculation of resulting forces (Smit *et al*, 1990) (**Figure 2.1.1-2**). The main benefit is that longer trajectory timescales are allowed, because the integration step can be ten times larger than that required for accurate atomistic simulations (Liwo *et al*, 2021). Coarse graining methods are particularly useful for simulating proteins embedded in a lipid environment, as determining accurate protein-lipid interactions requires long microsecond timescales which is still difficult to achieve in a reasonable timeframe for atomistic simulations (Jeffries & Khalid, 2021). In addition, since the energy landscape is much smoother, the protein is more stable and there is less chance of encountering issues such as a protein becoming trapped in an unfavourable conformation. A powerful method is to combine atomistic and coarse grain simulations to understand the function of a system, exploiting the longer timescales allowed in coarse graining to calculate average interactions, especially in a lipid system, while gaining functional insight into atom-atom interactions in an atomistic system, such as specific hydrogen bonding or charge-charge interactions (Jeffries & Khalid, 2021).

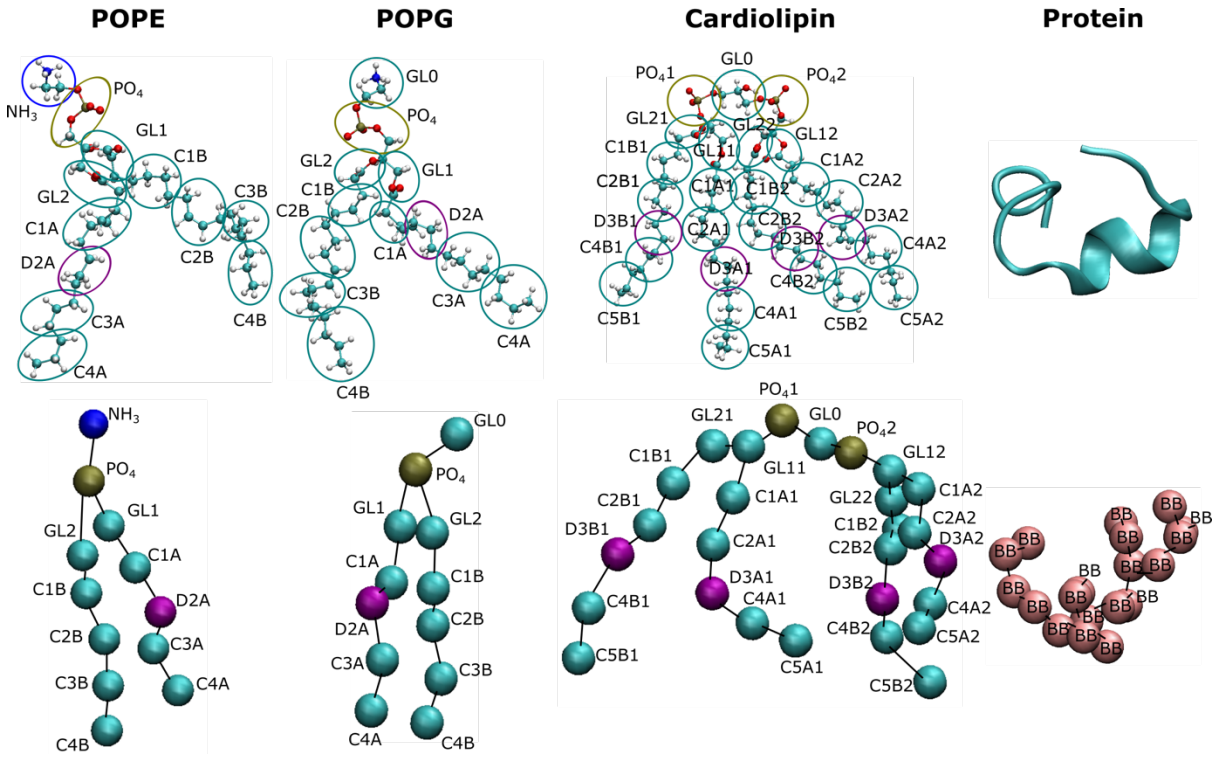
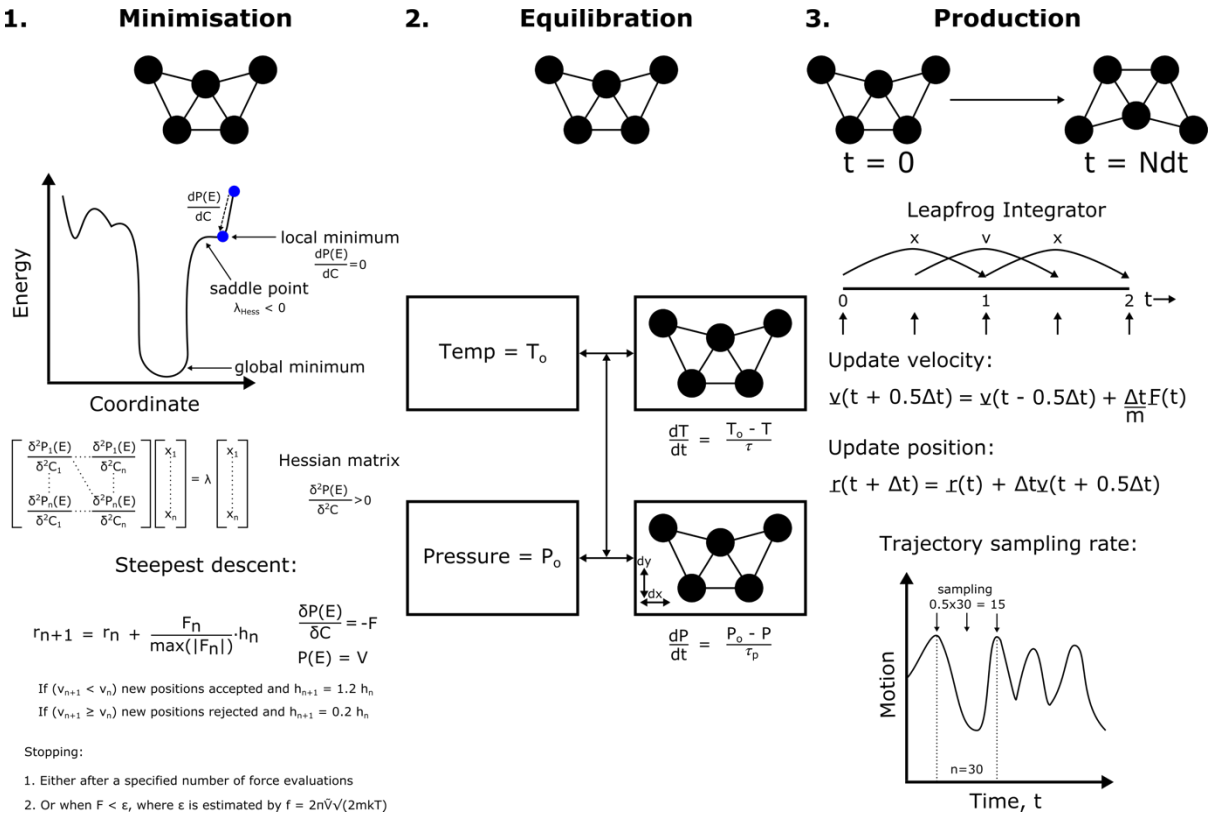


Figure 2.1.1-2. Representations and resolutions in MD simulations. Biological material, such as lipid (*e.g.* POPE, POPG, and cardiolipin) and protein can be represented as either a full atomistic (top panel) or coarse grain (bottom panel) model. In coarse grain models, groups of atoms (circled) are represented by beads which results in a smoother energy landscape. Note that for the protein model only backbone atoms are shown for clarity. Named group of atoms for the coarse grain representation are based on CHARMM names for the atom groups (Jo *et al*, 2008). NH_3: amide group, PO_4x: phosphate head, GLx(y): glycerol, Cn{A/B}x: carbon tails, Dn{A/B}x: carbon double bonds, BB: protein backbone.

1.2 Molecular Dynamics Workflow

A typical MD simulation set-up of the prepared system follows the steps of energy minimisation, equilibration, and production (**Figure 2.1.2-1**). During energy minimisation, the kinetic energy is removed from the system to reduce the effects of thermal noise. This works by minimising the potential energy function over coordinate space for the system which removes internal strains within molecules and intermolecular clashes (**Figure 2.1.2-1**). This stabilises the system before a production run, to avoid simulation failure after applying large forces on a system far from equilibrium. The steepest descent or the conjugate gradient method are typically used for this step (Adcock & McCammon, 2006); steepest descent has

the advantage of approaching the energy minimum very quickly, but it does not consider the previous gradient information as the conjugate gradient method does (Lindahl *et al*, 2021).



<u>Figure 2.1.2-1</u>. Molecular Dynamics (MD) Simulation Set-Up. A typical MD simulation step-up will follow the steps of minimisation, equilibration, and production. **1) Minimisation**: the potential energy of the system is minimised by calculating first and second order derivatives of the energy landscape. Minima will exist where the first derivate is equal to zero. Most algorithms can only find the nearest local minimum (not necessarily geometrically near), due to the presence of saddle points. The Hessian matrix describes the change in coordinates of the system as the energy changes in the landscape; to find local minima, we only need to consider where the second-order partial derivate of the potential function is greater than zero. This is done by solving for eigenvectors and eigenvalues for the Hessian matrix over the coordinate space. Note that at a saddle point, only a single negative eigenvalue exists. Due to the high-dimensionality of the energy space, it may not be feasible to find the global minimum in a reasonable length of computational time. Instead, the search is restricted to local coordinate space. For example, in the steepest descent algorithm, new positions (r_{n+1}) are found by taking a small random incremental step (h_n) in the landscape; if the updated potential (v_{n+1}) is less than the previous potential (v_n), the step is accepted and a new step (h_{n+1}) is taken at 1.2 times h_n. Otherwise, a new random step is taken at a reduced step size (0.2 times h_n) and a new potential is calculated. This procedure proceeds iteratively until either a specific number of force evaluations have been completed, or after the force has fallen below a specified epsilon value (ε), where epsilon is estimated by using the force harmonic oscillator (f),

calculated by taking into account the oscillator frequency ($\tilde{\nu}$), the Boltzmann constant (k), and the temperature (T). **2) Equilibration:** after minimisation, the system is equilibrated to the correct temperature and pressure by coupling the system to thermostats and barostats at the specified reference temperature (T_o) and pressure (P_o). Exchange is allowed between the two systems until the system under investigation is equilibrated. This can be done using a variety of algorithms, depending on the ensemble. Here, is shown the NPT (number of particles, pressure, and temperature) ensemble. Note that the figure gives the Berendsen temperature and pressure coupler (Berendsen *et al*, 1984) (other coupling energy terms are not shown), which should not be used in a production run as it does not give the correct NPT ensemble, but can be used for equilibrations (Berendsen, van der Spoel & Drunen, 1995). Also note that equilibrating for a specific pressure also means that the box size (x, y) will change by a small amount (dx, dy). **3) Production:** after the system is fully equilibrated, a production run can be generated to capture molecular motion in the system (typically without the molecular restraints present in the minimisation and equilibration steps). New velocities ($\underline{v}$) and positions ($\underline{r}$) can be calculated by using the leapfrog integrator, where new positions are calculated by using velocities calculated at half-step increments relative to the positions (the velocities "leap over" the positions). Note that an integrator is used for any step requiring a time component, which also applies to equilibration. The most important factor in a production run is to determine the sampling rate of the molecular motion (how often to write the energy and trajectory output files). For a complete description, energy components must be saved at half the sampling rate of the smallest molecular motion; if a full motion occurs every 30 steps then energies must be saved every 15 steps. Note that the sampling rate for the output trajectory does not need to match the sampling rate of the energy file. Typically, energies need to be sampled more to give an accurate output trajectory, but the sampling rate for the final saved trajectory outputs (positions and velocities) can be much sparser to make the file sizes more manageable. The leapfrog integrator drawing is adapted from the Gromacs Reference Manual 2021 (Lindahl *et al*, 2021). Information presented here is also adapted from the manual.

After energy minimisation, the system is equilibrated to the target temperature and pressure (or volume). In an NPT (number of particles, pressure, and temperature) ensemble simulation, the pressure is equilibrated to a target meaning that the simulation box size must change to maintain the pressure. This differs from an NVT ensemble simulation, where the volume is fixed. Simulation of biological processes will usually follow an NPT ensemble (**Figure 2.1.2-1**). Equilibration is done by coupling the system to a temperature bath and a barostat at the target temperature and pressure (Lindahl *et al*, 2021). Exchange between the systems is done on a sufficient timescale to allow the system to be adequately equilibrated (typically 5–20 ns) (Walton & Vanvliet, 2006) (**Figure 2.1.2-1**). Box sizes (volume) will fluctuate slightly (<0.1 nm on average) in order to maintain the desired pressure. Too much fluctuation in the box size may indicate that the system is unstable (Braun *et al*, 2019). In a protein-membrane system, long equilibration times are necessary to allow the lipid to adjust to the protein (*e.g.*,

Ingólfsson *et al*, 2013). In practice, rules of thumb generally govern the length of simulation time required for equilibration due to the empirical uncertainties in determining if a system is actually at equilibrium (Genheden & Ryde, 2012). Generally, it is better to have an equilibration time which is too long than one which is too short before going into a production run. After equilibration, the system should be stable enough to complete a production run for a desired length of time.

Production runs output the simulation positions, velocities, and energies to be used in the final analyses. Typically, the number of steps before writing to the output file in a production run can be more than for equilibration as the system is more stable and the longer run timescales means the file sizes are much larger. A step number is chosen based on the scientific question which is being addressed; for example, shorter step sizes are used for short simulations which need frequent outputs (*e.g*. hydrogen bonding lifetimes), while longer step sizes are used for long simulations which do not need frequent outputs (*e.g.* diffusion over a large area). The resulting file size should also be considered; if the file size is too big this could limit the use of repeat simulations and make analyses difficult and time-consuming. If output frames need to be skipped to conduct most analyses in the production run, this indicates that the output step size is too small and could be increased (Lindahl *et al*, 2021). Ideally, production runs should be long enough to adequately sample the protein conformational space. In practice, determining if the system has sampled all possible protein conformations which should occur at equilibrium is difficult. Indeed, it has been suggested that most, if not all, MD simulations are not fully converged systems (in terms of protein conformational sampling) due to the short run times imposed by computational limits in a single study (*e.g.,* Caves, Evanseck & Karplus, 1998; Grossfield, Feller & Pittman, 2007; Sawle & Gosh, 2016). Some have suggested the use of shorter trajectories with more repeats rather than relying on single, longer trajectories (Knapp, Ospina & Deane, 2018). I would argue that such an approach would depend on the scientific questions under investigation – for protein-lipid interactions, longer simulation times are required to adequately assess lipid contact with the protein. At least two simulations should be conducted under each simulation condition to evaluate simulation variability in the target analyses.

1.3 Other Considerations

There are a number of additional considerations to take into account during a simulation set-up (**Figure 2.1.2-1**). The most important is the use of periodic boundary conditions, which is applied to the system in order to avoid the production of artefacts caused by the unnatural phase boundaries between what is considered inside and outside the system (*e.g.* protein in water vs vacuum). This is done by placing atoms in a box with translated copies; effectively, this replaces the phase boundary artefacts with periodic boundary condition artefacts. It should be noted though that for large systems (>10,000 particles) the error introduced is small. The shape of the system box used can vary (*e.g.* cubic, rhombic dodecahedron, truncated octahedron) and has an impact on the computational efficiency of the simulation. Box sizes are chosen to be large enough so that the protein does not interaction with copies of itself (Wassenaar & Mark, 2005). In addition, the shape of the box is chosen to fit molecular motions which occur across the periodic boundary; thus, molecules and its translated copies can only be stacked in certain shapes to form a space-filling model (Bekker *et al*, 1995).

Another important consideration is picking the right force field for the system under investigation (**Figure 2.1.3-1**). Force fields are an approximation of intra- and inter-molecular interactions in the system which can be split into short-range and long-range interactions (Villa *et al*, 2007; Oakes & Domene, 2016). The force-field equation takes into consideration both bonded and non-bonded interactions between all the atoms in the simulation:

$$E_{tot} = \sum_{bonds} f(r) + \sum_{angles} g(\theta) + \sum_{dihedrals} h(\phi) + \sum_{i<j} l\left(R_{ij}\right) \quad \text{(equation 2.1.3-1)}$$

where the first three terms of equation (2.1.3.1) describe the bonded interactions and the last the non-bonded interactions.

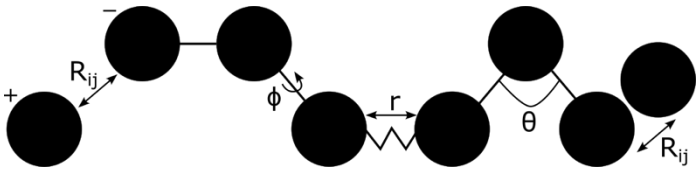


Figure 2.1.3-1. Molecular dynamics force fields. Force fields utilise a force field equation (*e.g.* equation 2.1.3-1) to describe all bonded and non-bonded interactions in the atomic system. Bonded interactions describe bond distance (r), bond angle (θ), and dihedral angles for bond-pairs (ϕ). Non-bonded interactions describe both electrostatic interactions and Leonard-Jones potential within a given distance (R_{ij}). See Durrant & McCammon (2011) for more details on the specifics of the functions in the force-field equation described.

There are a number of force field options to consider: AMBER (Hornak *et al*, 2006), CHARMM (Best *et al*, 2012), GROMOS (Oostenbrink *et al*, 2004), OPLS (Harder *et al*, 2016), and MARTINI (Marrink *et al*, 2007) are popular options. The choice of force field will depend on the molecules(s) being simulated and whether the system is an atomistic or coarse grain representation (Lopes, Guvench & MacKerell, 2015). Short-range interactions typically encompass atom pairs, split by bonded and non-bonded interactions. For this calculation, a minimum image convention is applied whereby the nearest image for each particle pair is considered and the cut-off radius cannot exceed half the box size. Pairs are found by a neighbourhood search updated after a target number of steps with a supplied interaction cut-off distance (*e.g.* Verlet buffer). Self-interactions are not included (Lindahl *et al*, 2021). For long-range electrostatic interactions encompassing N particles and their periodic images, Ewald summation or Particle-Mesh Ewald is performed (Lindahl *et al*, 2021). Long-range interaction cut-off radius is usually determined by the Leonard-Jones potential (for between atom interactions) and sometimes by Coulomb interactions for calculating distance and charge (Lindahl *et al*, 2021).

Lastly, one should also consider the solvent model to be applied to the system. The choice is usually between an implicit and explicit solvent (Pechlaner *et al*, 2022). Implicit solvents do not add water molecules to the system, but instead rely on an implicit representation of water interactions with the molecule(s) in the system (Kleinjung *et al*, 2012). The main benefit is that the computational load is greatly reduced as water molecules make up the majority of the atoms in the simulated system. However, the accuracy of the simulation is reduced and

may affect the output results, especially if a hydration shell or specific interactions with water are important to the function of the system. Explicit solvent will model water molecules and add them to the system. There are a number of choices, based on the forcefield used, but one of most popular is the TIP3P or SPC representations, respectively under CHARMM and GROMOS forcefields (Brini *et al*, 2017). In these cases, the charge of the water molecules is fixed. However, polarisable water models can be used under the CHARMM/Drude (Lemkul *et al*, 2016) or AMOEBA (Ponder *et al*, 2010) forcefields.

1.4 Molecular Dynamics as "*In Silico*" Experiments

MD simulations allows us to conduct *in silico* experiments which would be difficult to do experimentally. Furthermore, the results can be used to generate hypotheses which can be tested to improve our understanding of the system under investigation. Specifically, I want to investigate the use of MD simulations to answer questions and develop hypotheses for the transport of haemolysin A in the type I secretion system (T1SS). For tractability, I utilise a model of haemolysin B (HlyB), the inner-membrane ABC-transporter protein responsible for recognizing HlyA substrate and providing the energy by ATP-hydrolysis. As the study was conducted before the publication of AlphaFold2 in 2021 (Jumper *et al,* 2021), this model is based on homology modelling.

By conducting *in silico* experiments, whereby HlyB is simulated in a simplified membrane system with and without its HlyA substrate, I can determine specific protein-substrate interactions which aid recognition. Furthermore, this will yield insight into specific protein-lipid interactions which may also aid substrate translocation (see **Methodology Section 2.1**). Additionally, there is a peptidase ABC-transporter homologue which is functionally the closest to HlyB with an available high-resolution structure (pdb ID 6v9z). This structure has peptide substrate trapped in two different states: an active translocating state with access to the channel opening in the transmembrane helices, and a non-translocating state (Kieuvongngam *et al*, 2020). By conducting MD simulations with and without its bound substrate, I can compare the output of this system with the HlyB system to determine if there are conserved protein-substrate and protein-lipid interactions, as well as any differences between translocating and non-translocating forms. Lastly, there is an available database,

MemProtMD, which automatically embeds membrane proteins deposited to the Protein Data Bank (PDB) (Berman *et al*, 2000) in DPPC lipid and conducts microsecond long simulations (Newport, Sansom & Stansfeld, 2019). Protein-lipid interaction outputs are available, so that I can compare both the HlyB and peptidase protein-lipid interactions to that of similar ABC-transporters. These predictions will reinforce observations of potential conserved features of protein-lipid interactions.

The Aims of this Chapter involves answering the following questions:

1) How does haemolysin B (HlyB) interact with its substrate haemolysin A (HlyA)?
2) Is the lipid environment important for substrate recognition/interaction?
3) Are there conserved elements in protein-substrate and protein-lipid interactions?

These Aims will be addressed with the following Objectives:

1) Produce a homology model of HlyB bound with ~20 C-terminal amino acids of HlyA
2) Prepare the system(s) by placing the model in simplified membrane
3) Conduct MD simulations with and without bound substrate
4) Analyse protein-substrate and protein-lipid interactions over the course of the simulations
5) Compare these results to MD simulations of the closest ABC-transporter homologue with an available high-resolution structure (peptidase, pdb ID 6v9z)
6) Compare these results to protein-lipid interactions of all ABC-transporters mined from MemProtMD

2. Methodology

2.1 Overview of the Simulation Pipeline

The general method is outlined in **Figure 2.2.1-1** and an overview of the simulations conducted in this study is given in **Table 2.2.1-1**.

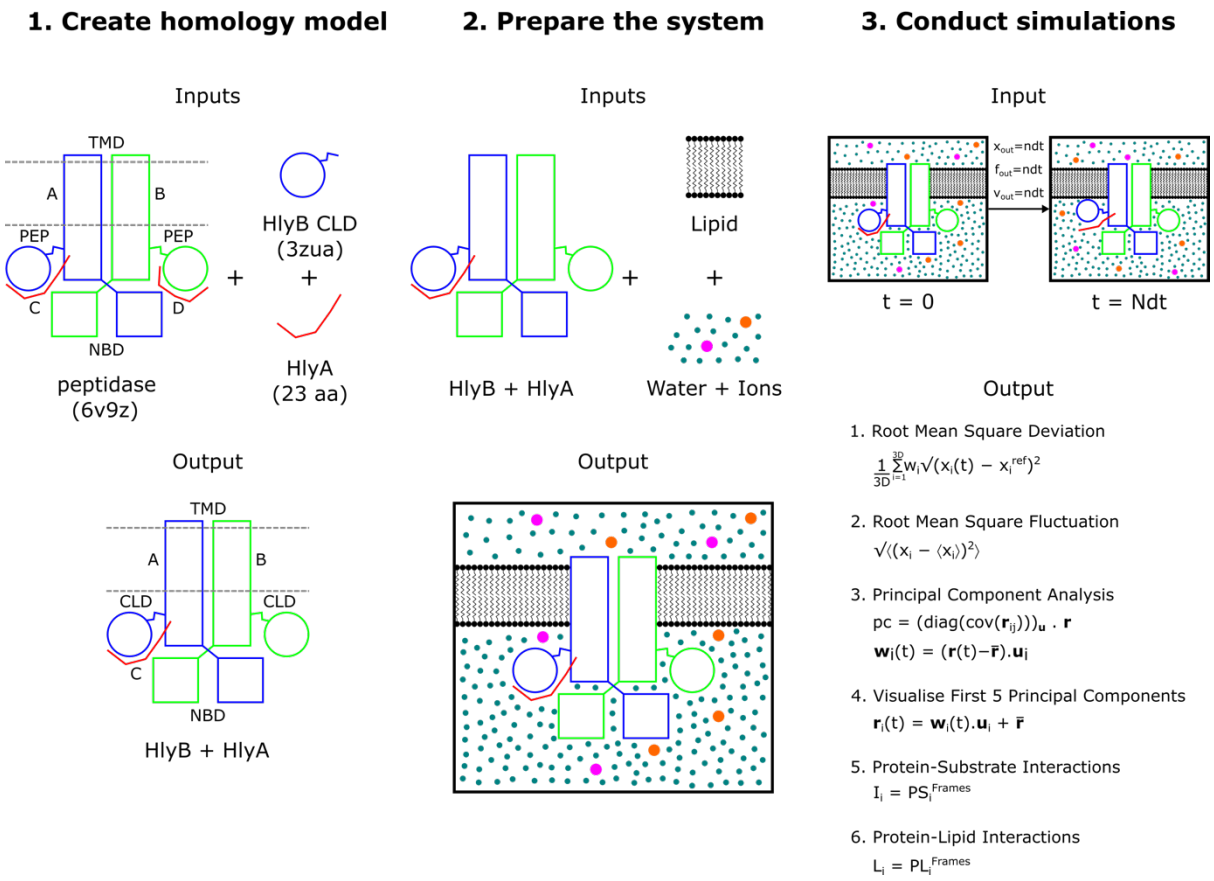


Figure 2.2.1-1. General methodology. Methods conducted for this chapter followed three stages: **1)** A homology model for haemolysin B (HlyB) with 23 C-terminal amino acids for its substrate haemolysin A (HlyA) was made using the closest structural homologue (peptidase, pdb ID 6v9z) as a backbone for the transmembrane domain (TMD) and nucleotide-binding domain (NBD). This was combined with the structure of HlyB C39-like domain (CLD) to make a full model of HlyB with bound HlyA. **2)** The HlyB model with bound HlyA was embedded in membrane and added to a box with water and ions. **3)** Atomistic and coarse grain simulations were conducted for the desired length of time. The outputs of the simulations were then analysed according to a pipeline.

Table 2.2.1-1. Overview of simulations conducted for this study.

Name (identifier(s)[a])	Temperature (K)	# Repeats	Atomistic/Coarse Grain?	Simulation Length (ns)
HlyB with substrate (u1, u2, u3)	303.15	3	Atomistic	750
HlyB with substrate (u1, u2)	303.15	2	Coarse Grain	5400
HlyB with substrate (u3, u4)	310.15	2	Coarse Grain	3240
HlyB without substrate (u5, u6)	303.15	2	Coarse Grain	5400
HlyB without substrate (u7, u8)	310.15	2	Coarse Grain	3240
Peptidase (6v9z) with substrate (u1, u2)	303.15	2	Coarse Grain	5400
Peptidase (6v9z) with substrate (u3, u4)	310.15	2	Coarse Grain	3240
Peptidase (6v9z) without substrate (u5, u6)	303.15	2	Coarse Grain	5400
Peptidase (6v9z) without substrate (u7, u8)	310.15	2	Coarse Grain	3240

[a]Identifiers correspond to a unique simulation for each study. For simulation reference to each identifier see the analysis code github repository available at :https://github.com/AMKCam [last accessed 8th October 2023]. An explanation of each analysis Notebook can be found in Table B in Appendix B.

2.2 Methods

2.2.1 Homology Modelling

Homology modelling for HlyB was completed using MODELLER 9.23 (Sali & Blundell, 1993). The peptidase/bacteriocin transporter (peptidase-containing ATP-binding cassette transporter, PCAT) (pdb ID 6v9z Chain A/B) (~30% sequence identity) (Kieuvongngam *et al*, 2020) and NMR structure of the HlyB CLD domain (pdb ID 3zua) (~98% sequence identity) (Lecher *et al*, 2012) were used as templates. C2 symmetry constraints based on the 6v9z model were imposed. The first 10 amino acids from 3zua were removed to more accurately position the N-terminal loop. For HlyA, 23 C-terminal amino acids were built using the PCAT peptide (6v9z Chain C) (~16% sequence identity) as the template. Sequence alignments between target and templates were completed and refined using UCSF-Chimera (Petterson *et al*, 2004) and exported for use in MODELLER. HlyA was positioned into the HlyB CLD using distance restraints along the sequence. Residues in the loop 128–138 were refined and assessed using DOPE (Shen & Sali, 2006). The best model was chosen based on DOPE score (Shen & Sali, 2006) and visual inspection of the output structure compared to input templates. Side chain rotamers were then refined using Scwrl 4 (Krivov *et al*, 2009).

2.2.2 Molecular Dynamics Simulations

Simulations were performed using GROMACS 2020.4 (Lindahl *et al*, 2020) or GROMACS 2021 (Lindahl *et al*, 2021). For atomistic simulations, hydrogens were first added using PDB2PQR server (http://apbs.poissonboltzmann.org/) (Dolinsky *et al*, 2004) at pH 7.0 under PROPKA to assign protonation states; AMBER naming convention was selected. The orientation of the protein in the lipid membrane was computed using the OPM server (Lomize *et al*, 2011). The resulting file was uploaded either into Martini Maker Bilayer Builder (Qi *et al*, 2015; Hsu *et al*, 2017) (coarse grain) or Membrane Builder Bilayer Builder (Jo *et al*, 2007; Wu *et al*, 2014)

(atomistic) accessed via the CHARMM-GUI web-interface (http://www.charmm-gui.org/) (Jo *et al*, 2008).

The CHARMM 36 force-field (Brooks *et al*, 2009; Lee *et al*, 2016) and TIP3 water model was used for atomistic simulations. For coarse grain simulations, the elnedynp model was used. This model uses an elastic network for coarse graining a protein structure and combines it with a MARTINI force field. The elastic network is useful for maintaining protein scaffolds, although bias towards the reference structure can occur. To inject a degree of physical reality to this model, use of the MARTINI framework is particularly useful, as it is based on experimental thermodynamic data for a range of molecule types (Periole *et al*, 2009). For HlyB, membrane of system size 140 Å was built using 75% POPE, 25% POPG, and 5% cardiolipin (Raetz & Downhan, 1990) in a 1:1 ratio between top and bottom leaflet. For PCAT, the same system size was used with 50% POPE, 20% POPG, and 30% cardiolipin (Bolobova, Zhukov & Klyosov, 1994; Timmons *et al*, 2009). Protein was equilibrated with the membrane (see below). The system was neutralised with either K^+ or Na^+ ions added using the Distance method. For minimisation, one or two rounds of steepest descent were performed for 5000 steps under a tolerance maximum force less than 1000 kJ/mol/nm. For coarse grain, the Verlet scheme (Verlet, 1967) with tolerance 0.005 kJ/mol/ps was used for computing pair list interactions and updated every 20 steps using a grid search. For atomistic, the Verlet scheme (Verlet, 1967) with a cut-off distance of 1.2 nm was used and pairs updated every 10 steps.

For coarse grain simulations, reaction-field electrostatics was used under relative dielectric constant 2.5 and distance cut-off 1.1 nm. Potential-shift-verlet van der Waals modifier was used with distance cut-off 1.1 nm. For atomistic, Particle Mesh Ewald electrostatics (Darden, York & Pedersen, 1993) was used with Coulomb cut-off of 1.2 nm. Force Switch van der Waals modifier was used with cut-off between 1.0 and 1.2 nm. Hydrogen bonds were used as constraints with the LINCS algorithm (Hess *et al*, 1998).

Temperature was maintained using velocity-rescaling temperature coupling (Bussi, Donadio & Parrinello, 2007) with a reference temperature of either 303.15 K or 310.15 K and tau_t of 1.0 ps in xyz directions. Pressure was maintained using Berendsen pressure coupling (Berendsen *et al*, 1984) with tau_p of 5.0 ps under semi-isotropic conditions, with a

compressibility of $3x10^{-4}$ bar^{-1} and reference pressure 1 bar in x-y directions. For atomistic simulations, Berendsen temperature (Berendsen *et al*, 1984) coupling was used with tau_p of 1 ps and reference temperature of 303.15 K.

All reference coordinates were scaled. Velocities and temperature were generated randomly at the start of the run using seeds under a Maxwell temperature distribution. For coarse grain equilibration, 4 rounds were performed at increasing integration times (2 fs, 5 fs, 10 fs, 15 fs) under leap-frog integrator for 1000 ps each round. A final equilibration step was performed for 18.75 ns. All other inputs were the same as for minimisation, except velocities and temperature were not generated at the start of the run. Outputs were saved every 1000 steps. For atomistic equilibration, 5 rounds were performed. Rounds 1−2 were performed under temperature coupling (fixed box size, NVT ensemble) with 1 fs integration, for a total time of 250 ps. Rounds 3−5 applied Berendsen pressure coupling (NPT ensemble) (Berendsen *et al*, 1984) with compressibility $4.5x10^{-5}$ bar^{-1} in x-y directions, for a total time of ~15 ns. Integration was increased to 2 fs from round 4. Velocity, positions, and forces were written every 5000 steps, while energies were calculated every 100 steps and written every 1000 steps. Centre of mass translational velocity was removed every 100 steps. Reference coordinates were scaled by their centre of mass using the scaling matrix of pressure coupling.

For coarse grain production, pressure was maintained using Parrinello-Rahman pressure coupling (Parrinello & Rahman, 1981) with a tau_p of 12 ps. Production was run for either 5.4 µs (303.15 K temperature) or 3.24 µs (310.15 K temperature) and outputs written every 5000 steps (75 ps). Repeat simulations were independently equilibrated to counter any biases in the velocities at the start of the production runs. For atomistic production, position restraints were removed, and Nosé -Hoover temperature coupling (Nosé, 1984) and Parrinello-Rahman pressure coupling (Parrinello & Rahman, 1981) were used with same parameters as before. Positions, velocities, and forces were written every 50000 steps, while energies were calculated every 100 steps and written every 1000 steps.

2.2.3 Analysis Pipeline

Analysis was performed on periodic boundary corrected trajectories. Initial protein-lipid and protein-substrate interactions were checked using GROMACS 2021 (Lindahl *et al*, 2021), with contacts queried within the cut-off radius (3.5 Å). Lipid density was also checked, with densities calculated for the phosphate head, ester group, or acyl tail in the Z-direction (normal to the membrane) and centered on the protein (Lindahl *et al*, 2020). Further analysis was performed using home-made scripts I developed in Jupyter Notebook (Kluyver *et al*, 2016) utilising the MDAnalysis package (Michaud-Agrawal *et al*, 2011; Gowers *et al*, 2016), pandas (Reback *et al*, 2022), and python 3. The analysis Notebooks are available on GitHub (https://github.com/AMKCam) (see **Table B** in **Appendix B** for a description of each NoteBook). First, Root Mean Square Deviation (RMSD) (Theobald, 2005; Liu, Agrafiotis & Theobald, 2010) and Root Mean Square Fluctuation (RMSF) calculations were performed followed by Principal Component Analysis (PCA) to check trajectory protein conformational sampling (Hess, 2002). The first five principal components were projected onto the backbone of the structure to visualise the main protein movements during the simulation. Solvent Accessible Surface Area (SASA) analysis was conducted using GROMACS 2021 (Lindahl *et al*, 2021), with the area computed on CLD residues found to critically interact with lipid (see protein-lipid analysis below). Protein-lipid and protein-substrate contacts were calculated over the entire trajectory using a 3.5 Å radius cut-off. For protein-lipid interactions, only the lipid head group was considered. To determine significant protein-lipid interactions, either a permutation test or a Monte Carlo permutation test was performed based on the amount of data (see **Appendix A** for an explanation). For the permutation test, an empirical sampling distribution was calculated by randomly shuffling protein-lipid interaction values (fraction of frames each interaction occurred) and calculating the mean difference between groups. For the Monte Carlo permutation test, an empirical sampling distribution was calculated by randomly shuffling protein-lipid interaction values and then randomly drawing a sample of 20% for each group and calculating the mean difference. Real interaction values were then compared to the empirical sampling distribution to derive p-values (see **Appendix A**). To compensate for multiple testing, the p-values were Bonferroni corrected using an applied significance cut-off. To examine lipid enrichment, the radial distribution curve was calculated

using MDAnalysis (Michaud-Agrawal *et al*, 2011). For the protein selection, statistically significant protein-lipid interactions uncovered by the permutation tests were used. For the lipid selection, phosphate heads were considered. The data was scaled to derive a probability of interaction. Total probabilities of interaction were calculated in 5 Å bands by integration over the probability of interaction curve. Lipid enrichment was calculated as follows:

$$Enrichment\ lipid_i = \frac{1}{N-1}\sum_{n=1}^{N}\frac{prob(lipid_i)}{prob(lipid_1)} + \cdots + \frac{prob(lipid_i)}{prob(lipid_{N-1})} \quad \text{(equation 2.2.3-1)}$$

For the *i*th lipid in N total lipids. This can be understood as how much more likely, on average, are we to find $lipid_i$ compared to any other lipid in this region. For lipid clustering analysis, top and bottom lipid leaflets were defined using LeafletFinder in MDAnalysis and interactions calculated with a distance cut-off of 15 Å (Michaud-Agrawal *et al*, 2011). To speed up the computation, only amino acids with the longest interactions with lipid (>1000 frames) were used and the output values were scaled by this number. Significant differences were determined by the permutation test using all the data at an applied Bonferroni corrected significance level.

To examine trajectory similarities, ensemble cluster analysis was performed (Tiberti *et al*, 2015) using the backbone of the protein and skipping every 1000 frames to speed up the computation. K-means was performed with a range of clustering values and similarities in structures scored by Jensen-Shannon divergence (Beckstein *et al*, 2009). Errors were estimated for each K-means clustering value used.

2.2.4 MemProtMD Data Mining

Structures used to determine chain asymmetry in lipid interactions are given in **Table 2.2.2.4-1**. Data on protein-lipid interactions were downloaded from the MemProtMD database (Newport, Sansom & Stansfeld, 2019) and imported into Jupyter Notebook (Kluyver *et al*, 2016). Protein-lipid interactions were analysed using code I developed and is available on GitHub (see **Table B** in **Appendix B**). Data was cleaned to remove null interactions. Statistically

significant interactions at different amino acid sites were determined by the Monte Carlo permutation test with a Bonferroni correction for multiple tests (see **Section 2.2.3**).

Table 2.2.2.4-1. MemProtMD ABC-Transporters chosen for this study.

PDB ID	Organism	Notes	Use in this Study?	Reference
6v9z	*Acetivibrio thermocellus ATCC 27405*	PCAT substrate bound	Yes	Kieuvongngam *et al* (2020)
7t54	*Acetivibrio thermocellus ATCC 27405*	PCAT ATP + Mg^{2+}	No – not present in database	Kieuvongngam & Chen (2022)
7t55	*Acetivibrio thermocellus ATCC 27405*	PCAT substrate + ATP	Yes	Kieuvongngam & Chen (2022)
7t56	*Acetivibrio thermocellus ATCC 27405*	PCAT substrate + ATP	Yes	Kieuvongngam & Chen (2022)
7t57	*Acetivibrio thermocellus ATCC 27405*	PCAT substrate + ATP	Yes	Kieuvongngam & Chen (2022)
4s0f	*Acetivibrio thermocellus ATCC 27405*	PCAT E648Q mutant + ATP	No – side chains not built into structure	Lin, Huang, & Chen (2015)
4pl0	*Escherichia coli*	McjD ouward occluded state	Yes	Choudhury *et al* (2014)
5ofp	*Escherichia coli*	McjD inward occluded state	No – only one chain built	Bountra *et al* (2017)
5eg1	*Escherichia coli*	McjD with resolved lipid	Yes	Mehmood *et al* (2016)
5l22	*Aquifex aeolicus VF5*	PrtD	Yes	Morgan, Acheson & Zimmer (2017)

3. Results

3.1 Homology Modelling

Homology modelling results are given in **Figure 2.3.1-1**. The structure of the templates (**Figure 2.3.1-1**.**a**) were checked for errors using the Ramachandran plot; no serious errors were found (**Figure 2.3.1-1b**). The HlyB backbone bound with 23 C-terminal HlyA amino acids was built using these templates (**Figure 2.3.1-1c**). The output model was checked for errors, using both the DOPE score and Ramachandran plot; the model backbone matched well with the input templates based on the DOPE score, while the Ramachandran plot found no serious errors (**Figure 2.3.1-1d**). It should be noted that Ramachandran outliers present in the model that were not also present in the templates were in highly flexible loop structures. Each outlier was manually inspected and found not to pose a problem which could not be fixed during rotamer optimisation in Scwrl 4 and energy minimisation in Gromacs (*e.g.* they existed away from an inter-domain interface). Output geometry was inspected and found to be acceptable. HlyA substrate was built using the bound PCAT peptide in its translocating form (**Figure 2.3.1-1e**). Only 23 amino acids were available in the template structure to model; the best match was found by aligning the PCAT peptide to several HlyA-like sequences mined from a variety of Gram-negative bacterial species.

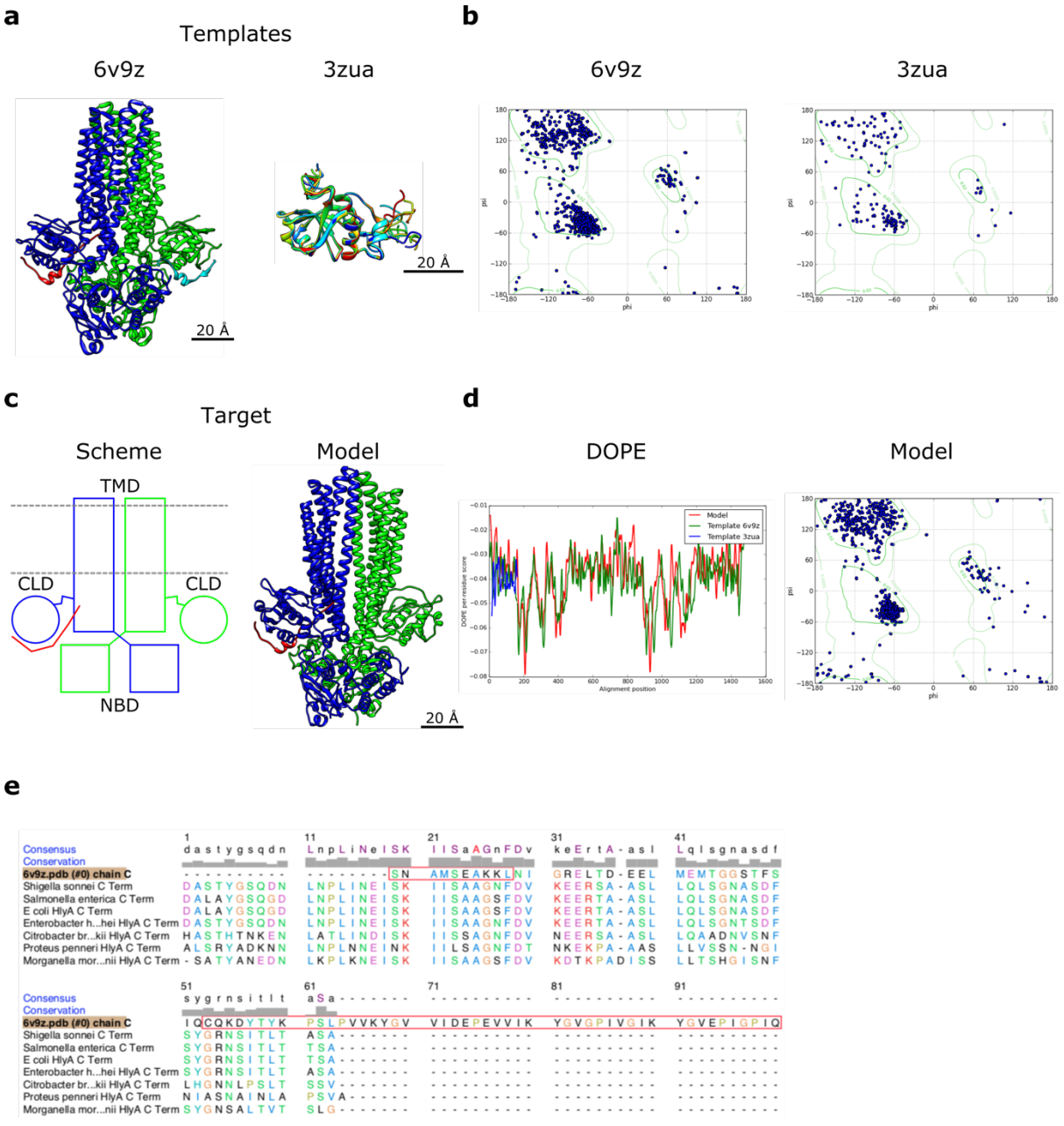


Figure 2.3.1-1. Homology model results. **a)** Two templates, PCAT transporter (6v9z) and HlyB C39-like domain (CLD) (3zua) were used to produce a full HlyB model. **b)** Ramachandran plots for the two templates, plotted using UCSF-Chimera (Petterson *et al*, 2004). **c)** HlyB homology model with bound 23 amino acids of HlyA C-terminus. HlyA enters HlyB between the CLD and nucleotide-binding domain (NBD). Approximate membrane region shown (grey, dotted line) in scheme. **d)** MODELLER DOPE score (left) and Ramachandran plot (right) for HlyB model. **e)** Sequence alignment of last 60 C-terminal amino acids of HlyA from various species with PCAT peptide (6v9z, Chain C). Amino acids in the red box could not be built into the structure. The other 23 amino acids which were built into the PEP domain pocket were used as a template for building the HlyA C-terminal tail.

3.2 Minimisation, Equilibration & Production Energy Terms

To check for successful minimisation of the system, the potential energy was plotted. Energies (potential, kinetic, enthalpy), temperature, pressure, and box size were checked after equilibration and production. In addition, lipid densities and total system density was checked after production. Example outputs are given in **Figure 2.3.2-1**. Visual inspection of the simulations after equilibration and production found no obvious structural errors or leakage of the membrane. In addition, energy terms were stable for all simulations after equilibration and production, and the box size fluctuations had stabilised during the equilibration.

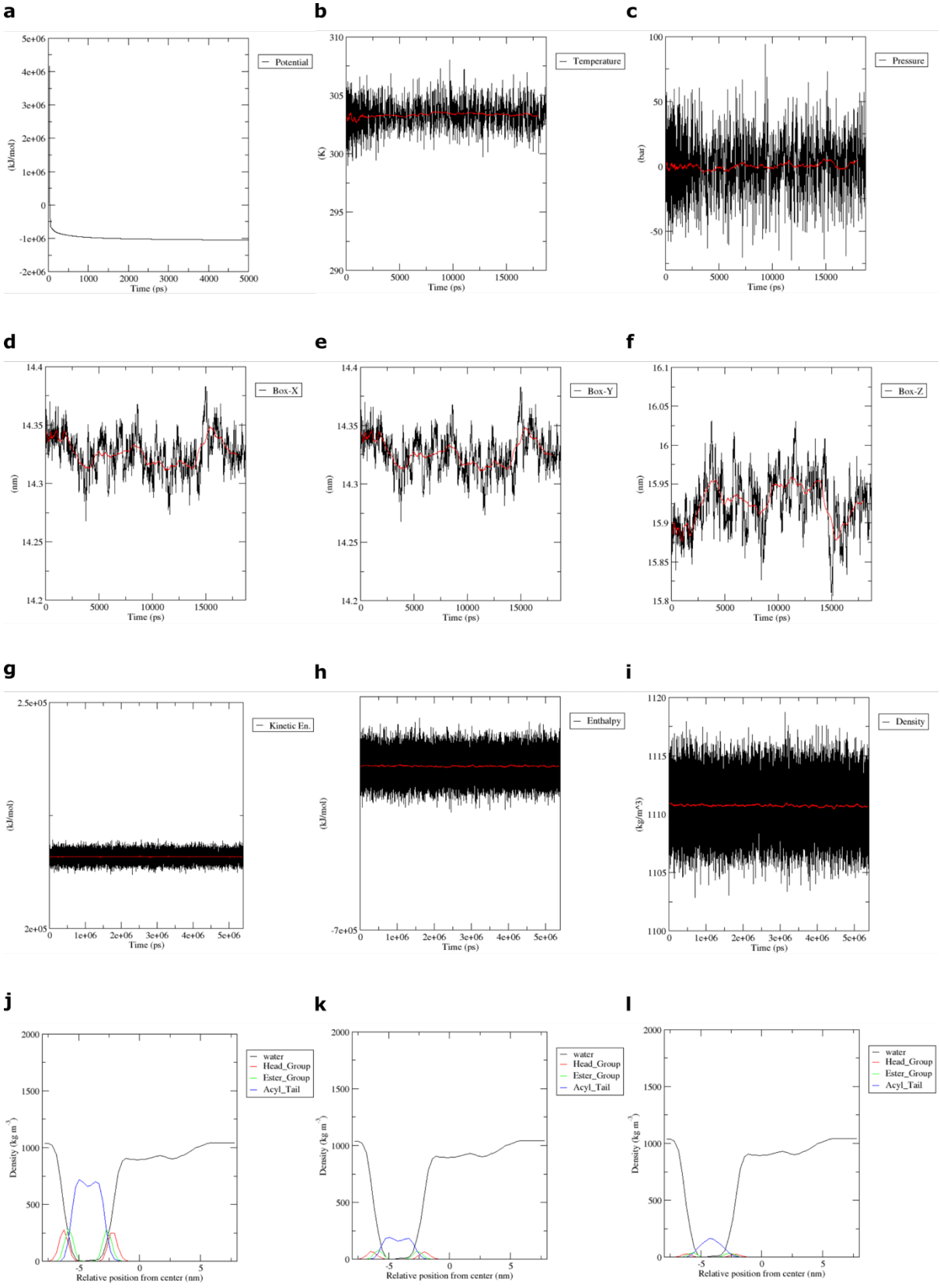


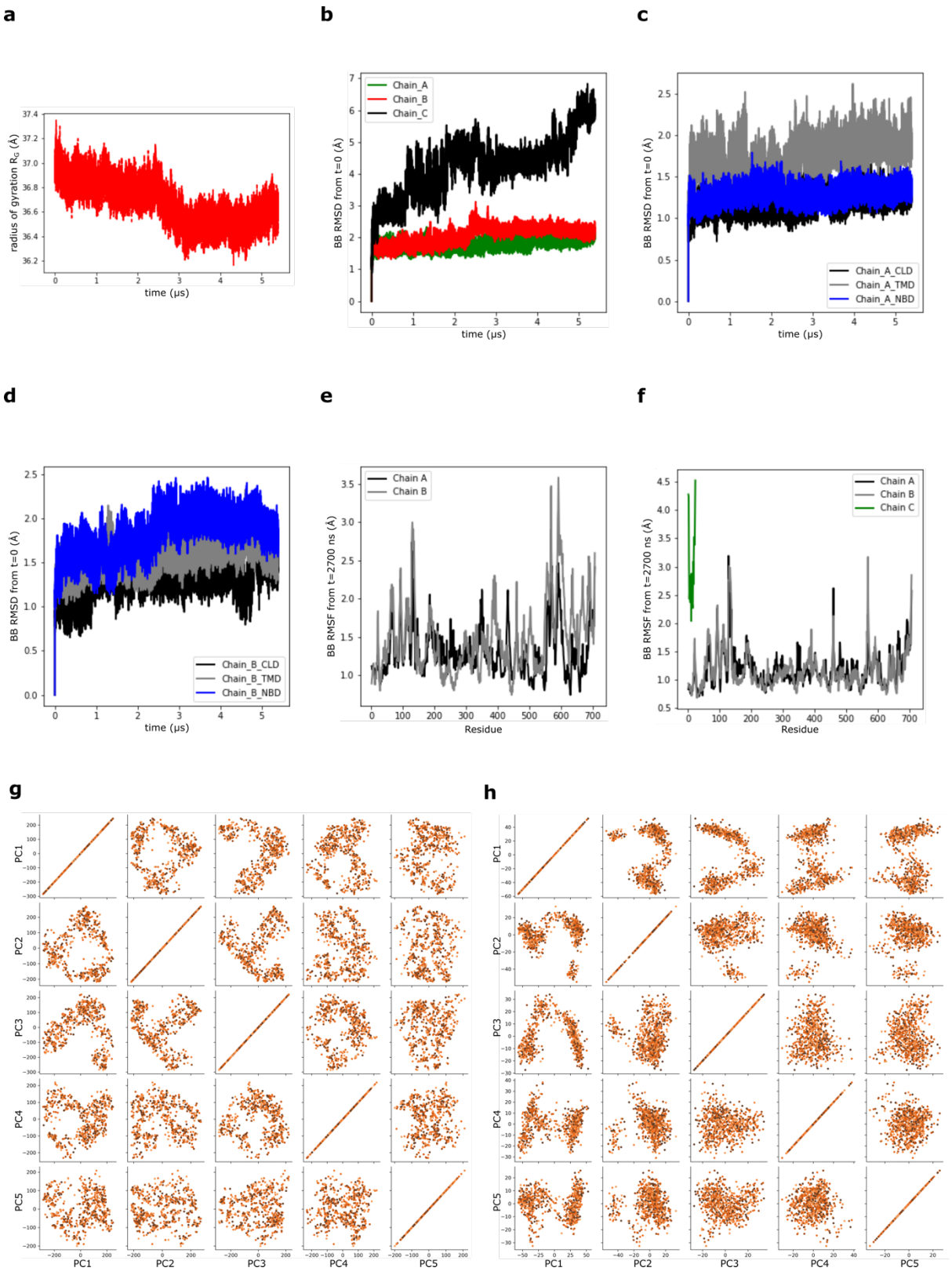
Figure 2.3.2-1. Example outputs after coarse grain minimisation, equilibration, and production. **a)** Reduction in potential energy after minimisation. **b) – f)** Stabilisation of temperature, pressure, and box size (X, Y, Z direction) after equilibration. **g) – i)** Stabilisation of kinetic energy, enthalpy, and total density after production**.**

j) – l) Lipid densities for POPE (j), POPG (k), and cardiolipin (l) were checked for water leakage across the membrane. The phosphate heads should make complete contact with the water molecules, while most of the tail density should be excluded from the water.

3.3 RMSD, RMSF, PCA, and Ensemble Similarity

Radius of gyration, Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), and Principal Component Analysis (PCA) on backbone atoms were checked after production. PCA was used to calculate the length of the production run to ensure adequate protein conformational sampling (Hess, 2002). Initial analyses were performed after 540 ns run-time and then every microsecond until the desired length of run-time had been complete. Example outputs are given in **Figure 2.3.3-1**. The radius of gyration decreased over time for all coarse grain simulations, indicating that the structures had adopted a more compact conformation and were likely more stable. There was no noticeable difference in RMSD and RMSF plots between simulations. The first five principal components were re-projected back onto the structure to visualise any large-scale movements which had occurred during the simulations. No noticeable transitions occurred in any of the simulations, even when the substrate had been removed. The main movements occurred in the binding pocket when the substrate was present, likely reflecting accommodation due to substrate flexibility. A similar result was found for the atomistic simulations (**Figure 2.3.3-2**). RMSD was similar for the CLD, transmembrane domain (TMD), and NBD in substrate-binding chain A for all three trajectories (**Figure 2.3.3-2a**). The CLD and NBD showed differences in RMSD between trajectories in non-substrate binding chain B (**Figure 2.3.3-2b**). The CLD showed much more flexibility compared to the other two domains in trajectory u2, while both the CLD and NBD were much more mobile than the transmembrane domain for trajectory u3 (**Figure 2.3.3-2b**). Comparisons of the first two principal components for all three trajectories showed similar patterns of contraction and expansion to accommodate substrate movement in the binding pocket (**Figure 2.3.3-2c**). Here, the transmembrane domain acts as a conduit for inter-domain movement, both within its own chain and also across chains (**Figure 2.3.3-2c**). To check the similarity of the trajectories, ensemble clustering was performed (**Figure 2.3.3-3**). For HlyB, two clear groupings emerge, with the trajectories with substrate showing high in-group similarity and those without substrate also showing high in-group similarity (**Figure 2.3.3-3**).

For PCAT, only the trajectories with substrate showed high in-group similarity. Interestingly, for PCAT the trajectories without substrate showed alternating patterns of similarity and dissimilarity with the trajectories with substrate (**Figure 2.3.3-3**). This many indicate a systematic bias in the repeat simulations for trajectories without substrate.



<u>Figure 2.3.3-1</u>. Coarse Grain Trajectories Example Initial Analysis. a) All simulations (u1–u8) showed a decrease in the radius of gyration. Example taken from trajectory u2. **b)** Root mean square deviation (RMSD) for each chain. For all simulations with substrate (u1–u4), the protein (Chain A and B) showed lower RMSD than the substrate (Chain C). Example taken from trajectory u2. **c) – d)** Similar RMSD values were found for all domains (C39-like domain (CLD), transmembrane domain, and nucleotide-binding domain (NBD)) for both chains in all

simulations. Example taken from trajectory u2. **e) – f)** All simulations (u1–u8) showed similar root mean square fluctuation (RMSF) for both protein only (Chain A and B) and with substrate (Chain C). Example taken from trajectories u1 and u2 for the respective figures. **g) – h)** Plotting the first five principal components gives an indication of the protein conformational sampling for each individual trajectory. At 2160 ns **(g)**, there is some indication that the trajectory has visited some conformations more sparsely than others, while the 5400 ns trajectory **(h)** shows more complete sampling. Note each point is a time frame. Also note that **h)** will look denser than **g)** as there are more frames plotted – the clustering of the points gives an indication of the trajectory sampling. More "island" clusters could potentially indicate inadequate sampling of certain conformations, as these were visited less often in the trajectory. Example taken from trajectory u5. Note that in this figure, each example comes from one trajectory only.

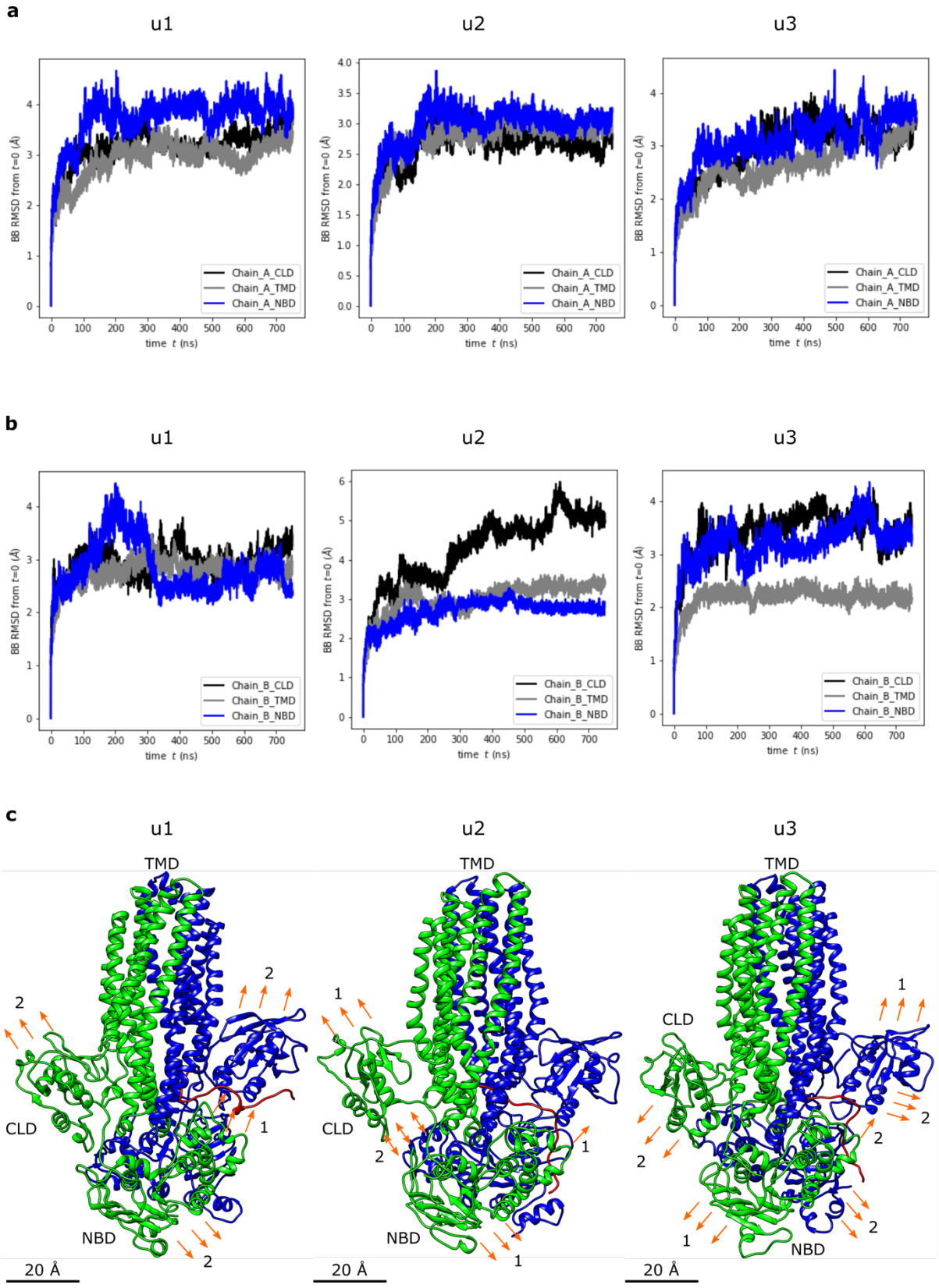


Figure 2.3.3-2. Main atomistic simulation movements of HlyB with HlyA substrate. a) Root mean square deviation (RMSD) of the protein backbone (BB) for all domains (CLD, C39-like domain; TMD, transmembrane domain; NBD, nucleotide-binding domain) in substrate-binding chain A across all three trajectories (u1–u3).

Domain movements are very similar across all domains in all three trajectories. **b)** RMSD for all domains in non-substrate binding chain B across all three trajectories. Trajectories u2 and u3 showed an increase in CLD movement compared to the u1 trajectory. **c)** The main movements for the first two re-projected principal components for all three trajectories are highlighted. The movements show accommodation of the substrate (red) within the HlyB binding pocket (chain A, blue) by simultaneous contractions and expansions (orange arrows). The main movements occurred in chain B (green) CLD and NBD and chain A CLD, while the transmembrane domain in both chains act as movement conduits for the simultaneous movement of the CLD and NBD. Structures are shown at time t=375 ns for all three trajectories. Note that the arrows are intended as a way to show the direction of movement, and hence are not vector representations.

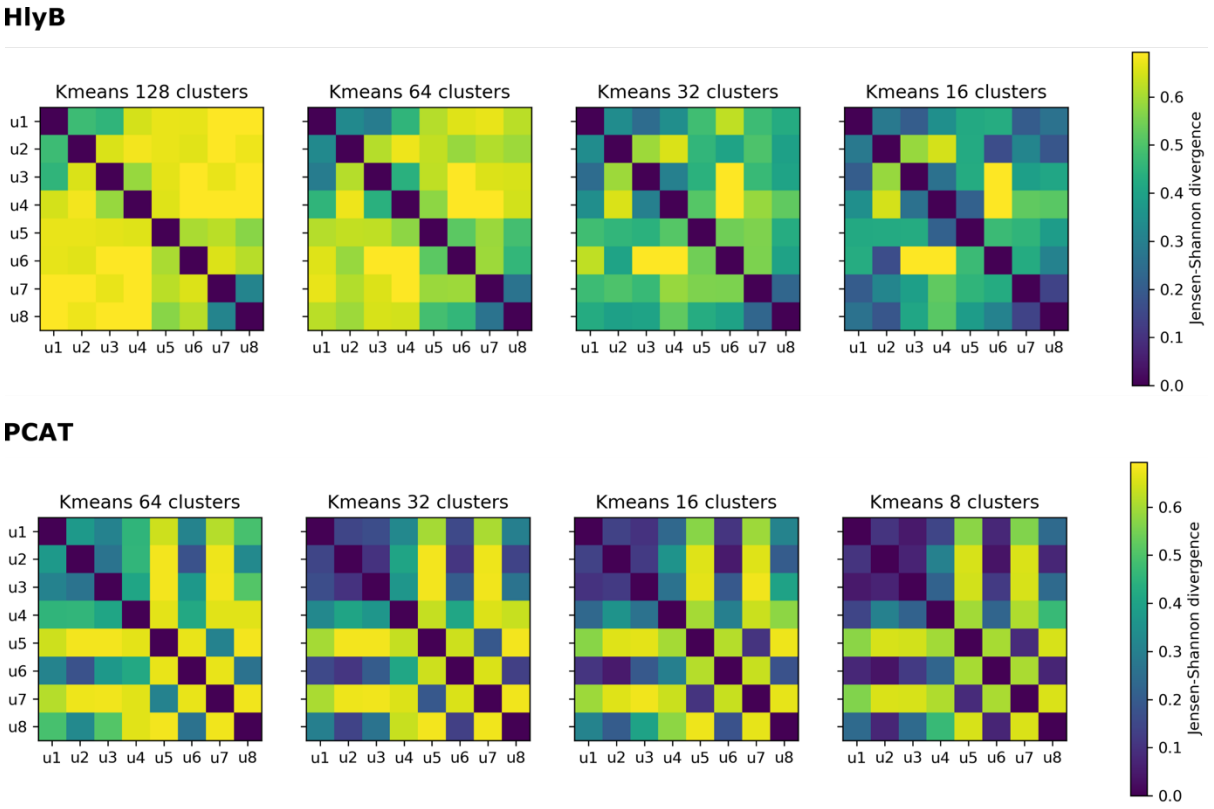


Figure 2.3.3-3. Ensemble similarity. K-mean clustering of samples of protein conformations taken every 1000 frames for all trajectories. Jensen-Shannon divergence provides a measure of conformation similarity, with 0 indicating complete similarity and 1 indicating complete dissimilarity. u1: with substrate, 303.15 K; u2: with substrate, 303.15 K repeat; u3: with substrate, 310.15 K; u4: with substrate, 310.15 K repeat; u5: without substrate, 303.15 K; u6: without substrate, 303.15 K repeat; u7: without substrate, 310.15 K; u8: without substrate, 310.15 K repeat.

3.4 Protein-Substrate Interactions

3.4.1 Coarse Grain Protein-Substrate Interactions

Protein-substrate interactions for HlyB are summarised in **Figure 2.3.4.1-1**. The majority of critical (>0.5 fraction of frames) interactions occur in the CLD. Interestingly, the HlyA substrate makes the majority of its NBD contacts with the opposite chain (**Figure 2.3.4.1-1a**). This may suggest a degree of cooperativity in HlyA binding. In addition, temperature did not seem to have a significant effect on substrate contacts (**Figure 2.3.4.1-1a**), with similar contact types and contact duration occurring at both temperatures. HlyA binding seems to be mostly mediated by charge-charge interactions in the CLD (**Figure 2.3.4.1-1b**). Inspection of structures showed that these charge-charge interactions mainly occur at the extreme N- and C-terminus of the substrate (**Figure 2.3.4.1-1c**). HlyA D20 makes critical contact with CLD residue K6 on Chain A at the C-terminus, while the N-terminal tail (D1, E4, and E5) is stabilised by charge-charge interactions with K51, K58, K61, and K62 (**Figure 2.3.4.1-1d**). Internally, the substrate is pinned by a charge-charge interaction of R6 with E79. Further polar contacts (Q13, S15, N17, and S19) with Chain A transmembrane amino acids S324 and S331, and Chain B NBD amino acids N556 and T597 help to stabilise the substrate in the pocket ready for translocation (**Figure 2.3.4.1-1d**).

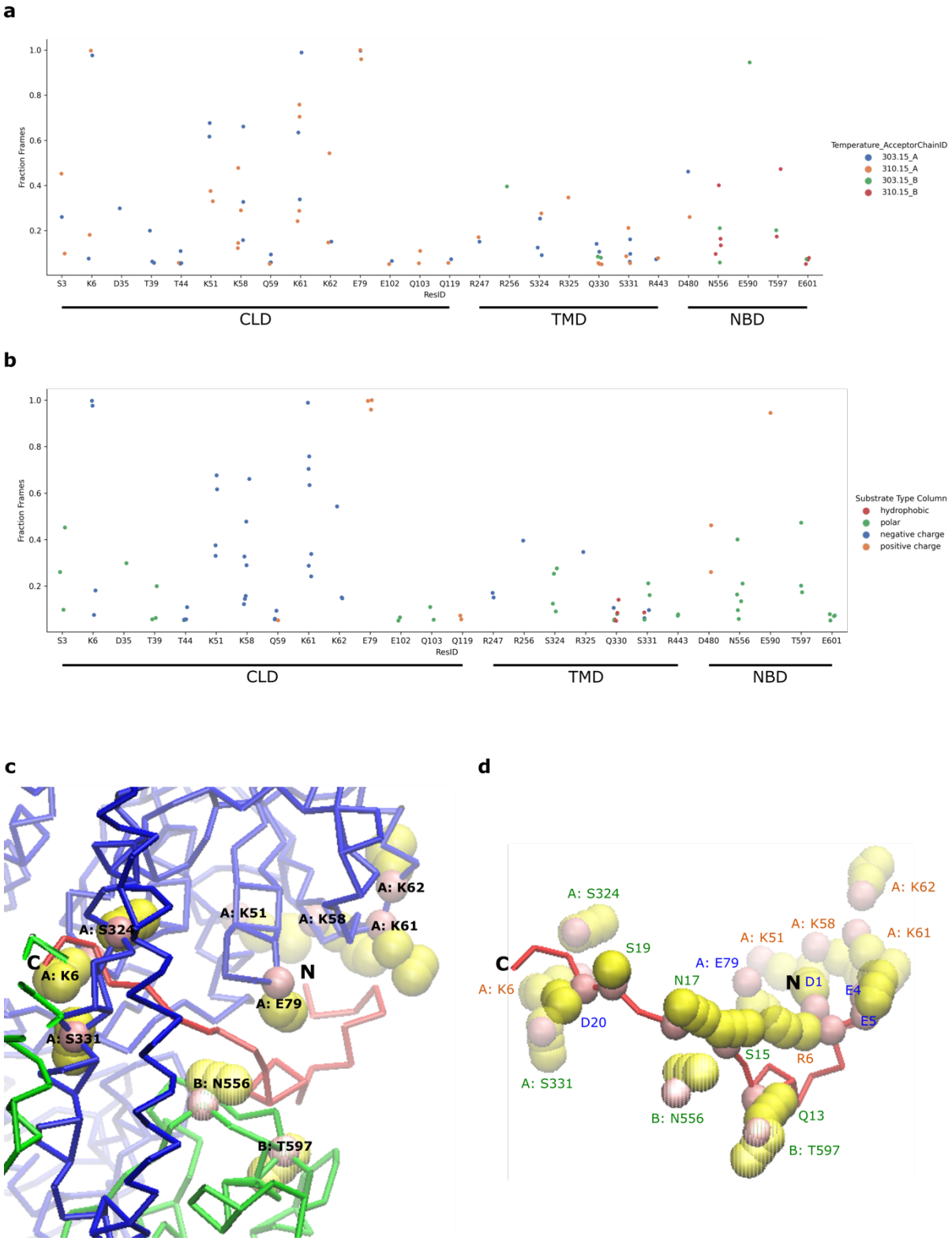


Figure 2.3.4.1-1. HlyB Average Protein-Substrate Interactions. a) Mean protein-substrate interactions labelled by chain and temperature across all HlyB coarse grain simulations with substrate (u1–u4). The majority of substrate interactions occur in the C39-like domain (CLD) (ResID 3–119), with a small number occurring in the transmembrane domain (TMD) (ResID 247–443) and nucleotide binding domain (NBD) (ResID 480–601). **b)** Mean

protein-substrate interactions labelled by substrate amino acid type across all HlyB coarse grain simulations with substrate (u1–u4). The majority of critical (>0.5 fraction of frames) protein-substrate interactions occur via charge-charge interactions or polar contacts. **c)** Still view of the substrate (red) in the binding pocket in contact with critical amino acids. Most of these critical interactions are in the CLD on the substrate binding chain (Chain A, blue), while two occur in the NBD on the opposite chain (Chain B, green). **d)** Still view of the substrate (red) in contact with the same critical amino acids. Recognition appears to be mediated mostly by charge-charge interactions, with a few polar contacts. Blue: negative charge amino acid; orange: positive charge amino acid; green: polar amino acid. Still views are from the u2 trajectory (see **Figure 2.3.3-3**), at time 2625 ns (out of 5400 ns). N- and C-terminus of the HlyA substrate are shown.

The PCAT structure has two substrates; one in the translocating form (**Figure 2.3.4.1-2**) and the other in the non-translocating form (**Figure 2.3.4.1-3**). Like the HlyB-HlyA interaction, there was no observable effect of temperature on PCAT protein-substrate interactions for either the translocating (**Figure 2.3.4.1-2a**) or non-translocating (**Figure 2.3.4.1-3a**) form. Furthermore, PCAT substrate stabilisation also occurs via charge-charge interactions and polar contacts for both the translocating (**Figure 2.3.4.1-2b**) and non-translocating (**Figure 2.3.4.1-3b**) form. Stabilisation of the N-terminal tail of the substrates occurs in an analogous manner to the HlyB-HlyA interaction; charge-charge interactions of substrate amino acids E5, D8, and E10 with PEP domain amino acids K70, K63, and K73 helps to stabilise the tail. Internal charge-charge interactions of R4 with D78 and E9 with K499 helps to pin the substrate to the pocket, while a charge-charge contact of E13 on the substrate with NBD amino acid K607 helps to stabilise the C-terminal region (**Figure 2.3.4.1-2c–d**). Interestingly, the non-translocating form of the substrate has lost most of the polar contacts of NBD amino acid N617 with S18/S21 and PEP domain amino acid Q15 with T19 (**Figure 2.3.4.1-3c–d**). In addition, there are no contacts with the transmembrane region (**Figure 2.3.4.1-3a–b**). Both translocating and non-translocating substrates form the majority of NBD contacts with the opposite chain than the one they are bound to (**Figure 2.3.4.1-2b/2.3.4.1-3b**), indicating that there is also a cooperative effect in substrate recognition and translocation.

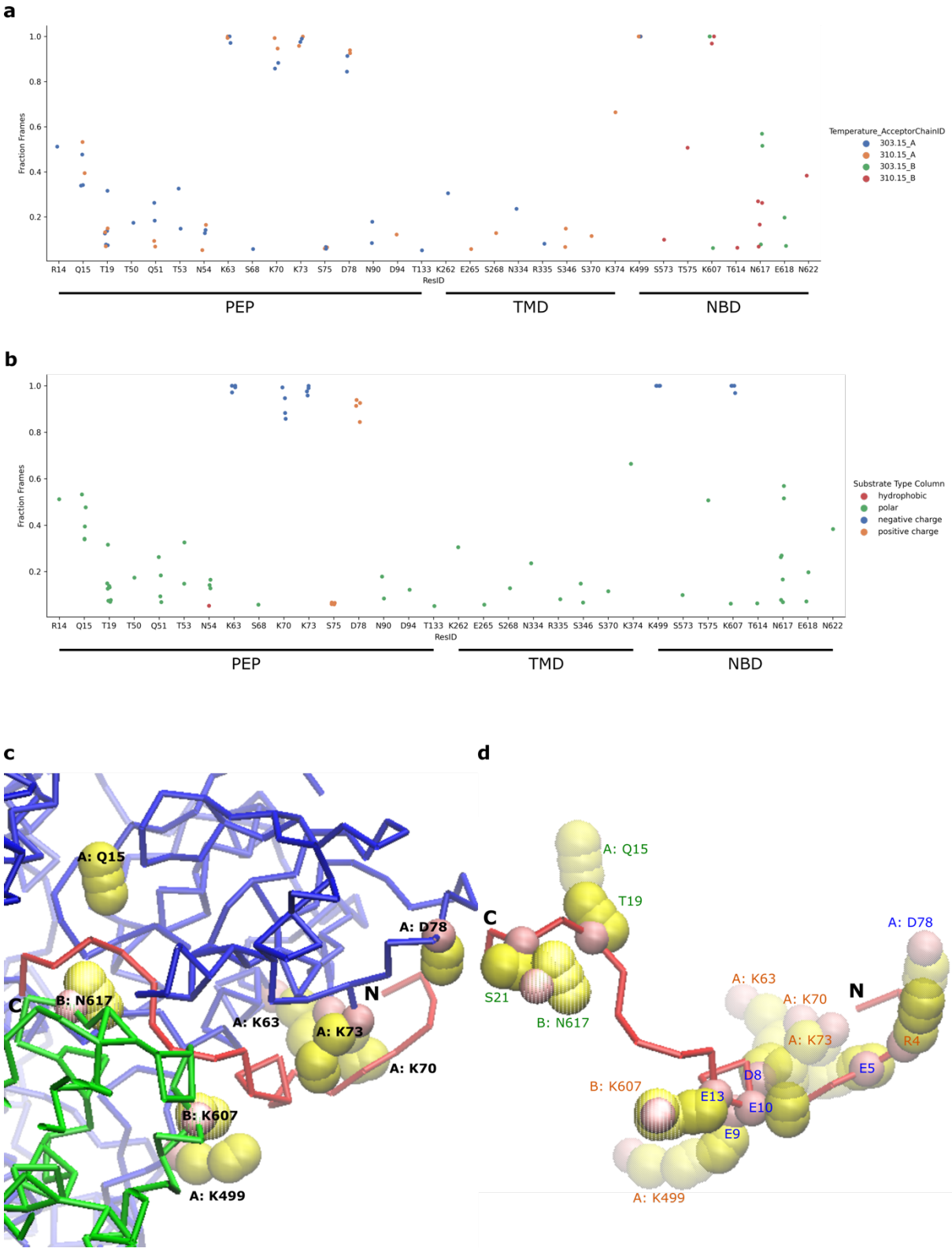


Figure 2.3.4.1-2. Average PCAT Translocating Protein-Substrate Interactions. a) Mean protein-substrate interactions labelled by chain and temperature for the translocating form of the substrate across all PCAT coarse grain simulations with substrate (u1–u4). The majority of substrate interactions occur in the PEP domain (ResID 7–126), with a small number occurring in the transmembrane domain (TMD) (ResID 255–367) and nucleotide

binding domain (NBD) (ResID 492–615). **b)** Mean translocating protein-substrate interactions labelled by substrate amino acid type across all PCAT coarse grain simulations with substrate (u1–u4). The majority of critical (>0.5 fraction of frames) protein-substrate interactions occur via charge-charge interactions or polar contacts. **c)** Still view of the substrate (red) in the binding pocket in contact with critical amino acids. Most of these critical interactions are in the PEP domain on the substrate binding chain (Chain A, blue), while two occur in the NBD on the opposite chain (Chain B, green). **d)** Still view of the substrate (red) in contact with the same critical amino acids. Like HlyB, recognition appears to be mediated mostly by charge-charge interactions, with a few polar contacts. Blue: negative charge amino acid; orange: positive charge amino acid; green: polar amino acid. Still views are from the u2 trajectory (see **Figure 2.3.3-3**), at time 2595 ns (out of 5400 ns). N- and C-terminus of the PCAT substrate are shown.

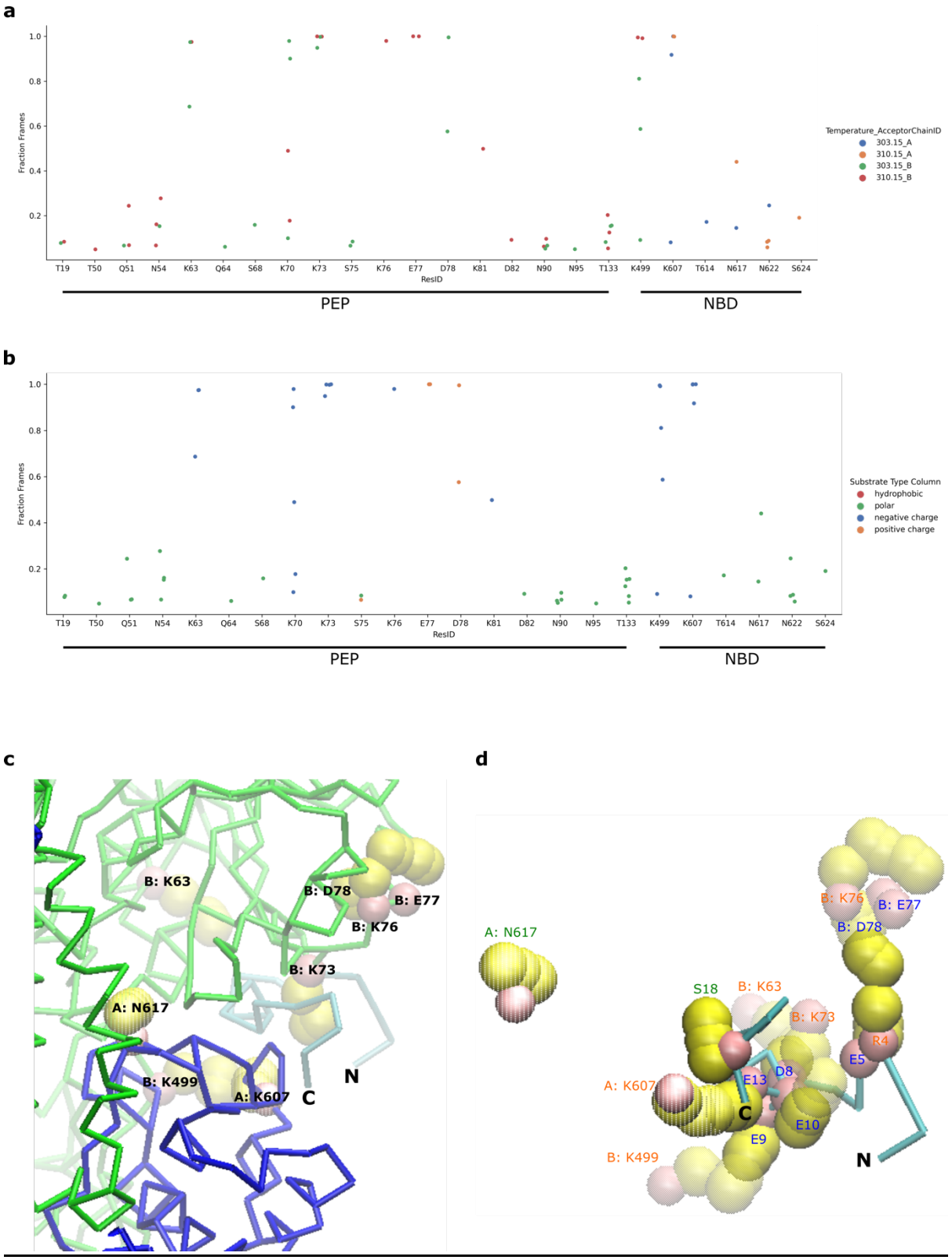


Figure 2.3.4.1-3. Average PCAT Non-Translocating Protein-Substrate Interactions. a) Mean protein-substrate interactions labelled by chain and temperature for the non-translocating form of the substrate across all PCAT coarse grain simulations with substrate (u1–u4). The majority of substrate interactions occur in the PEP domain (ResID 12–126), with a small number in the nucleotide binding domain (NBD) (ResID 492–617). No contacts are made with the transmembrane domain (TMD). **b)** Mean non-translocating protein-substrate interactions

labelled by substrate amino acid type across all PCAT coarse grain simulations with substrate (u1–u4). The majority of critical (>0.5 fraction of frames) protein-substrate interactions occur via charge-charge interactions or polar contacts. **c)** Still view of the substrate (cyan) in the binding pocket in contact with critical amino acids. Most of these critical interactions are in the PEP domain on the substrate binding chain (Chain B, green), while two occur in the NBD on the opposite chain (Chain A, blue). **d)** Still view of the non-translocating substrate (cyan) in contact with the same critical amino acids. Note the lack of contact with Q15 in the PEP domain and reduced contact with N617 in the NBD. Blue: negative charge amino acid; orange: positive charge amino acid; green: polar amino acid. Still views are from the u2 trajectory (see **Figure 2.3.3-3**), at time 2595 ns (out of 5400 ns). N- and C-terminus of the PCAT non-translocating substrate are shown.

To compare HlyB and PCAT protein-substrate interactions over the course of the simulation, critical contacts were computed as a time course (**Figure 2.3.4.1-4**). Critical contacts pinning the substrate into position were maintained throughout the simulation, with little divergence based on temperature (**Figure 2.3.4.1-4**). HlyB and PCAT have very similar protein-substrate contacts, with a series of interaction “pins” placing the substrate into a translocation-ready state (**Figure 2.3.4.1-4**). Most of these interactions are charge-charge with a few polar contacts; of particular interest are the charge-charge pins located at the C- (“top”) and N- (“tail”) termini of the substrates (**Figure 2.3.4.1-4**).

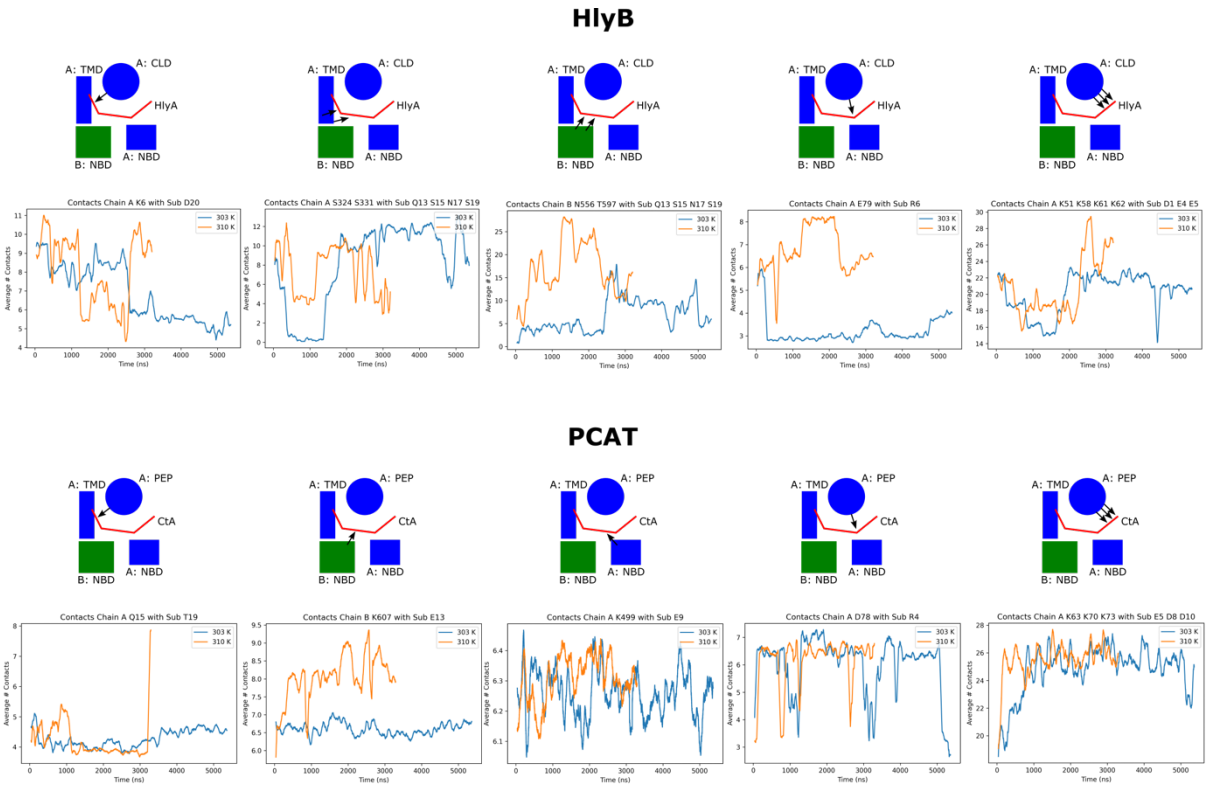
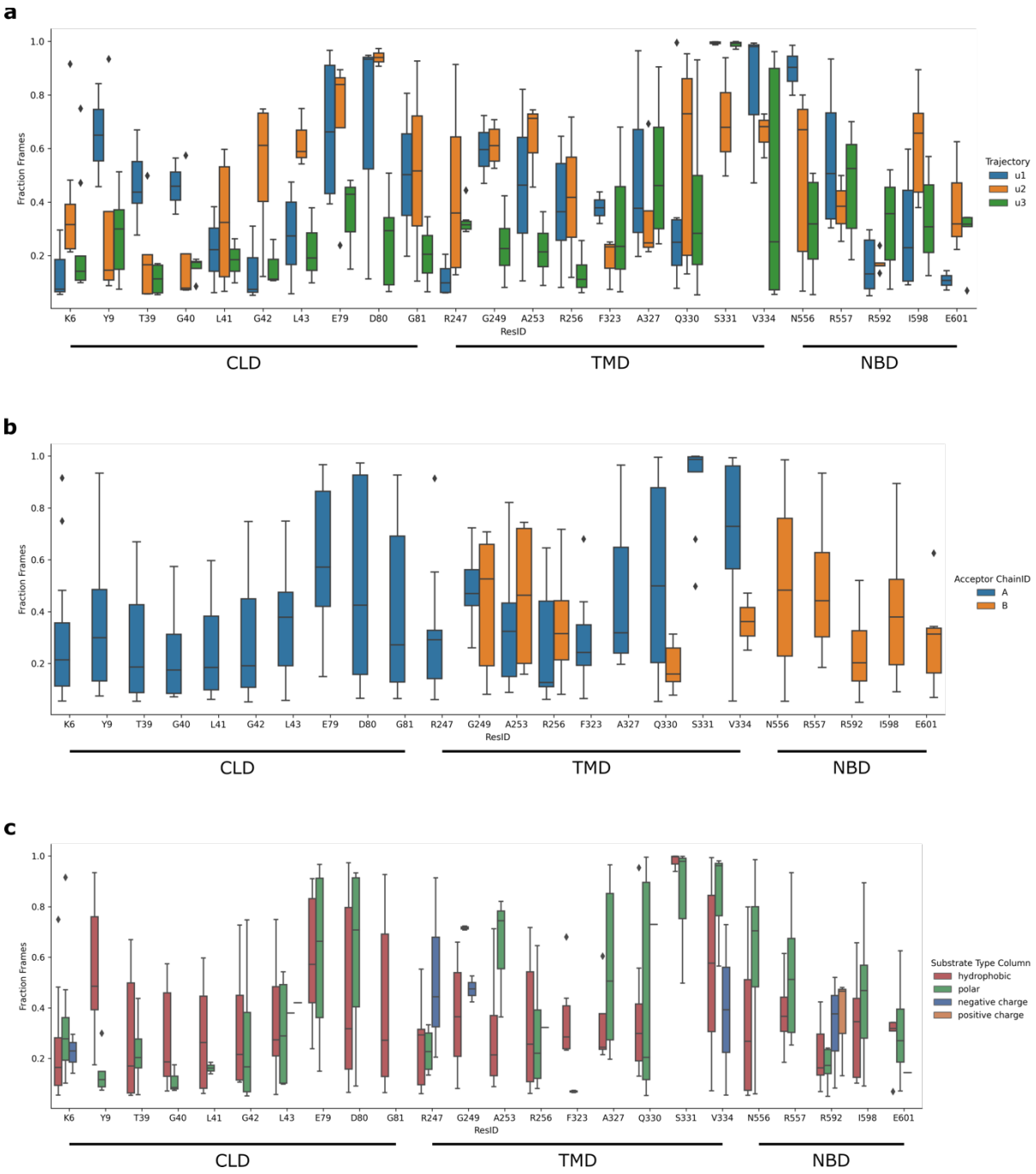


Figure 2.3.4.1-4. Time course of HlyB and PCAT Protein-Substrate Interactions. HlyB (top) makes critical contacts with its HlyA substrate as critical junctures: a "top pin" (far left panel), a "tail pin" (far right panel) and several "middle pins" (middle panels). Temperature of the simulation does not seem to affect most of the critical interactions. Analogously, PCAT (bottom) also has the same substrate-pinning interactions, with the exception that the TMD does not make any critical interactions and the role is instead filled in by chain A NBD.

3.4.2 Atomistic Protein-Substrate Interactions

Critical atomistic protein-substrate interactions are similar to the coarse grain results (see **Section 3.4.1**). Interactions across all three trajectories were examined to derive a consensus set. Interactions were visualised as a distribution across all three simulations rather than as an average to see the median, minimum, maximum, and interquartile range of interaction values. (**Figure 2.3.4.2-1a**). Examination of the consensus set by chain showed the same pattern as the coarse grain results; interactions primarily occurred in the substrate-binding chain A CLD and transmembrane domain, and chain B TMD and NBD (**Figure 2.3.4.2-1b**). Examination of interactions by amino acid type found a pattern of charge-charge interactions in the CLD (K6), TMD (R247), and NBD (R592) along with polar-polar contacts (E79, D80, S331, N556, and E601). Contact of hydrophobic HlyA substrate residues with polar or charged

residues on HlyB indicates that the substrate may be marginally stable in the binding pocket in the atomistic simulations (**Figure 2.3.4.2-1c**).



<u>Figure 2.3.4.2-1</u>. Distribution of atomistic protein-substrate interactions. Over the three atomistic simulations (u1–u3) conducted, the spread of interactions between the HlyB protein and HlyA substrate within 3 Å was visualised by**: a)** Consensus (occurs in all three simulations) interactions for all three trajectories in the C39-like domain (CLD), transmembrane domain (TMD), and nucleotide-binding domain (NBD). **b)** Consensus protein-substrate interactions for all three trajectories by chain. Like the coarse grain simulation results, main

interactions occur in substrate-binding chain A CLD and TMD, and in non-substrate binding chain B NBD. **c)** Consensus interactions for all three trajectories by substrate amino acid type. Hydrophobic contacts with polar or charged residues may indicate a degree of marginal stability of the substrate in the binding pocket. Note the box plots show the spread of the interaction data, with smaller boxes showing a tighter range of interaction at that amino acid.

Further examination of protein-substrate interactions found similar results to the coarse grain simulations (**Figure 2.3.4.2-2**). Most critical contacts were between charged/polar residues, with a few unfavourable contacts. The greatest difference between the atomistic and coarse grain simulations is the lack of tail stabilisation in the atomistic simulations (**Figure 2.3.4.2-2**), which may explain some of the unfavourable contacts. However, like the coarse grain simulations there appears to be a gate-keeping mechanism formed by the opposite chain's NBD; polar contacts with N556 and R557 helps to maintain substrate positioning within the binding pocket. A few key contacts with the TMD also appears to aid access to the transmembrane channel; R247 forms a charge-charge interaction with D20 on the substrate, while S331 forms polar contacts with N17 and S19 (**Figure 2.3.4.2-2**).

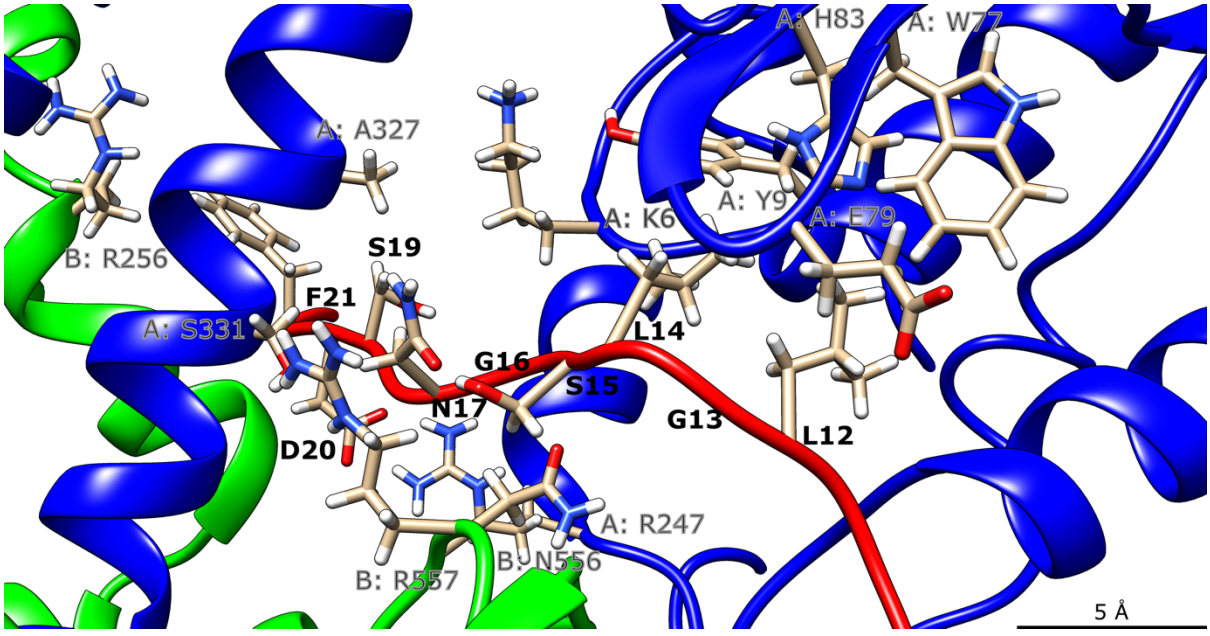


Figure 2.3.4.2-2. Atomistic substrate binding. HlyA substrate makes critical interactions via polar and charge-charge interactions in the HlyB binding pocket. There are also unfavourable polar/charge contacts with hydrophobic residues. Residues K6, Y9, and E79 in chain A (blue) C39-like domain (CLD) makes contact with

substrate (red) residues G16, N17, L14, and L12. Residues H83 and W77 in the CLD also make contact with L12 and L14. In the transmembrane domain, residue chain A residue R247 makes a charge-charge contact with D20, while chain A residue S331 makes polar contacts with N17 and S19, and A327 makes contact with S19. Chain B (green) residue R256 also makes contact with F21 on the substrate. In the nucleotide-binding domain (NBD), chain B residues N556 and R557 make polar contacts with substrate residues S15, G16, and N17. Still view is taken from trajectory u2 at time 375 ns.

3.5 Protein-Lipid Interactions

To uncover the dynamic range of protein-lipid interactions, the analysis was split into three cut-off ranges: an overview of the whole system using lipid enrichment analysis (3.5–140 $\mathrm{\AA}$), specific chain and domain interactions with lipid (up to 15 $\mathrm{\AA}$) (lipid clustering), and specific amino acid contacts with lipid (within 3.5 $\mathrm{\AA}$) (lipid preference). Lipid enrichment analysis shows a general enrichment of cardiolipin near the protein, for both HlyB and PCAT (**Figure 2.3.5-1**). Within 5 $\mathrm{\AA}$ of the protein, cardiolipin is 2–3 times more likely to be found within this distance than either POPE or POPG. While PCAT lipid enrichment does not show an effect of temperature (**Figure 2.3.5-1b**), there may be one for HlyB with trajectories at 303.15 K showing more cardiolipin enrichment than trajectories at 310.15 K (**Figure 2.3.5-1a**). In addition, PCAT does not show an effect of the presence of substrate on enrichment (**Figure 2.3.5-1b**), while for HlyB there might be the tendency for cardiolipin to be slightly more enriched when the substrate is present (**Figure 2.3.5-1a**).

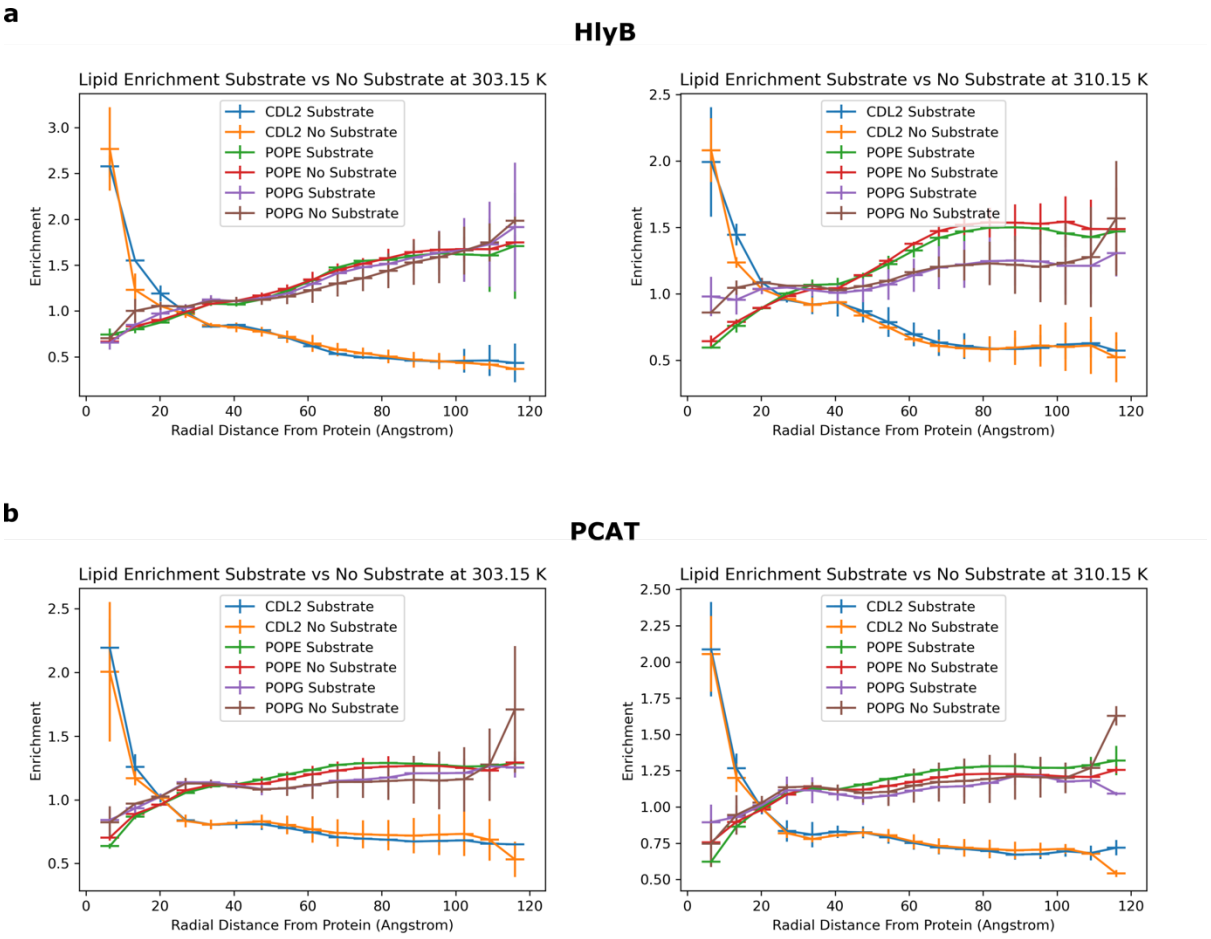


Figure 2.3.5-1. Lipid Enrichment. a) Lipid enrichment for HlyB at 303.15 K (left) and 310.15 K (right) shows a general enrichment for cardiolipin (CDL2) within 20 Å of the protein. The presence of HlyA substrate had no impact for POPE/POPG enrichment, while there might be an effect for cardiolipin within 15 Å of the protein. There may also be a tendency for greater cardiolipin enrichment within 5 Å of the protein under lower temperature. **b)** Lipid enrichment for PCAT at 303.15 K (left) and 310.15 K (right) also shows a general enrichment for cardiolipin (CDL2) within 20 Å of the protein. The presence of substrate or the temperature does not appear to affect lipid enrichment. Bars in the x-axis show each 5 Å band used in the lipid enrichment calculation, while bars in the y-direction are error bars showing standard deviation (n=2).

To untangle these relationships further, lipid clustering analysis was performed. The proteins were split into different regions to determine if there were statistically significant differences in how the lipid clustered to the protein depending on if the substrate is present or not (**Figure 2.3.5-2**). Strikingly, HlyB Chain A CLD showed a dramatic decrease in all lipid counts (cardiolipin, POPE, and POPG) when the substrate was removed. There was also a decrease in the number of POPE in the transmembrane top region and Chain B CLD when the substrate

was removed (**Figure 2.3.5-2a**). This was in contrast to the results for PCAT, where lipid counts increased in the transmembrane top region for cardiolipin and POPE when the substrate was removed (**Figure 2.3.5-2b**). All other regions had similar lipid clustering profiles.

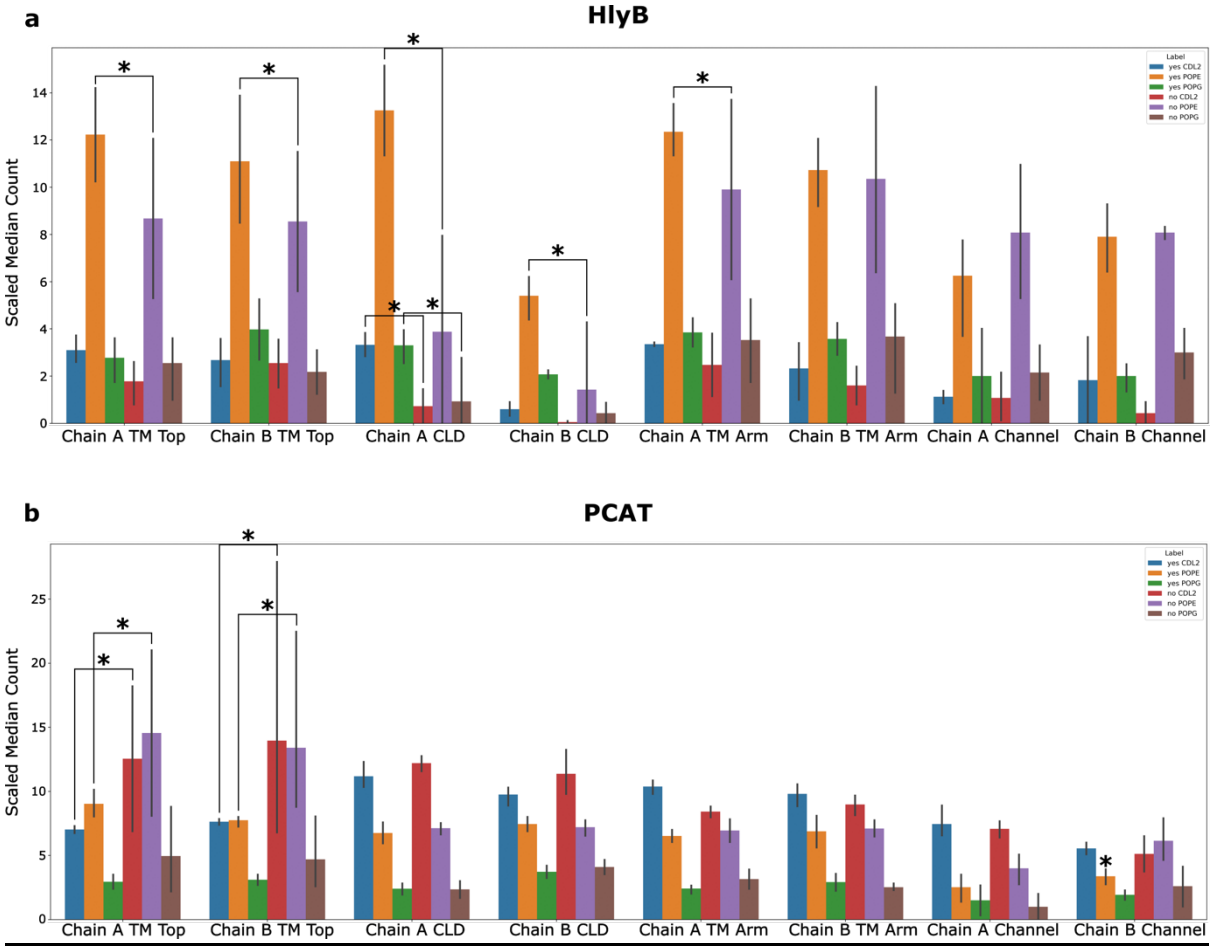


Figure 2.3.5-2. Lipid Clustering. **a)** A significant ($p < 0.01$) (asterisk) increase in lipid counts was found in Chain A C39-like domain (CLD) for cardiolipin (CDL2), POPE, and POPG and Chain B CLD for POPE when the substrate was present ("yes") compared to when the substrate is not present ("no"). Chain A and B transmembrane (TM) top and Chain A transmembrane (TM) arm also showed an increase in POPE counts when the substrate was present. The following residues were used in the analysis – CLD: Q19, N22, T87, K88, N94, Q103, N105, T126, S129, S130; transmembrane arm (TM Arm): T141, R151, R152, Q268, S272, K369; transmembrane top (TM Top): Q176, R186, S189, N192, S290, S402, S406, Q409; substrate entrance channel (Channel): S308, S312, Q429, Q432, Q435, Q436. **b)** A significant ($p < 0.01$) (asterisk) decrease in lipid counts was found in Chain A and B transmembrane (TM) top for cardiolipin (CDL2) and POPE when the substrate was present. A decrease in POPE counts was also found in Chain B substrate entrance channel (Channel) when the substrate was present. The following residues were used in the analysis – PEP: K9, R14, S27, Q31, S105, N108, T112, R122, T151, Q152; transmembrane arm (TM Arm): N153, K163, K166, K167, K168, K233, T284, S381, K384; transmembrane top (TM Top): N204, S306, S307, N414, N419; substrate entrance channel (Channel): T324, N327, K328, Q331, N332,

R335, Q446, Q450, T451, N457. The bar graphs show a median with the error bars reflecting a 90% confidence interval on that estimate from the data. Statistical significance was determined for pairwise comparisons (n=4 trajectories with substrate vs n=4 trajectories without substrate) using a permutation test with Bonferroni correction for multiple testing. Note that Chain A binds the HlyA substrate (HlyB) or the translocating form of the substrate (PCAT). Counts were found by taking the median lipid interaction values over the trajectory and scaling by the number of amino acids queried in each region.

To confirm the results of the lipid clustering analysis, a Solvent Accessible Surface Area (SASA) analysis was performed for the respective CLD and PEP region of HlyB and PCAT (**Figure 2.3.5-3**). The results corroborate what was found for the CLD/PEP regions in the lipid clustering analysis; decrease in HlyB lipid counts when substrate is not bound is confirmed in the SASA analysis, where greater solvent accessibility is seen on average in no substrate trajectories (**Figure 2.3.5-3a**). In addition, the SASA analysis confirms a chain asymmetry, where substrate-bound Chain A shows lower solvent accessibility when substrate is present, but no such pattern is detected in Chain B. In contrast, PCAT trajectories showed no difference between substrate or no substrate trajectories in terms of solvent accessibility (**Figure 2.3.5-3b**). In addition, no chain asymmetry is observed. This also fits with the lipid clustering results, where PCAT showed no differences in the PEP domain for either chain. Visualisation of structures indicates that differences may be due to how the CLD is interacting with the membrane; when substrate is bound, the Chain A CLD lies up against the membrane like a flat sheet, which is not seen for Chain B. However, for PCAT PEP interaction with membrane is the same on either chain (**Figure 2.3.5-3c**).

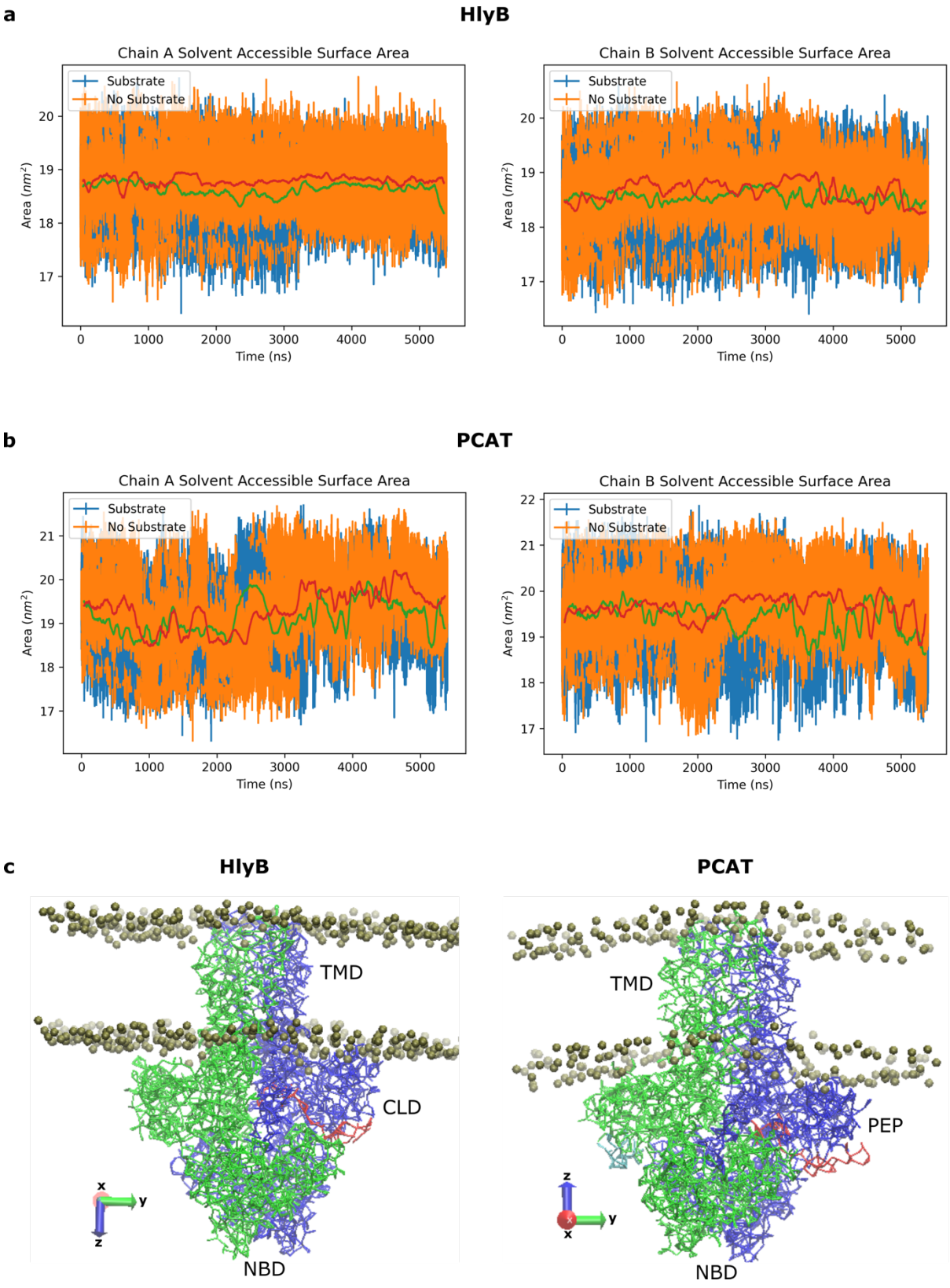


Figure 2.3.5-3. HlyB and PCAT Solvent Accessible Surface Area (SASA) Analysis. a) SASA for all coarse grain trajectories for HlyB (n=4 trajectories with substrate vs n=4 trajectories without substrate) shows that on average substrate-binding Chain A has lower solvent accessibility when substrate (red line) is bound compared

to when it is not bound (green line). The same pattern is not seen for Chain B. **b)** PCAT coarse grain trajectories show similar SASA for Chain A and Chain B whether or not substrate is bound (n=4 with substrate vs n=4 without substrate). **c)** Visualisation of membrane-bound (phosphate heads, gold beads) HlyB or PCAT in the presence of substrate indicates that the SASA results may be explained by an asymmetry in HlyB, where there is greater lipid contact in the CLD where substrate is bound. This is not found for PCAT. Example structures taken from u2 at 2700 ns. For SASA plots, running averages are shown for substrate (red) and no substrate (green) trajectories.

Since these results seemed to suggest that the presence of substrate was having an effect on lipid interaction with the protein, further lipid preference analysis was conducted. To uncover specific amino acid interactions with lipid, a smaller (<3.5 Å) cut-off radius was used. Statistical significance between substrate and no substrate trajectories in lipid preference was tested by using the Monte Carlo permutation test and tests were conducted between matched pairs (by chain and amino acid). This test was chosen as it is more robust than the permutation test for larger sample sizes (see **Appendix A**). The results agree with the lipid clustering analysis (**Figure 2.3.5-4**). For cardiolipin, amino acids N68, T87, and K88 on chain A CLD showed more interactions when the substrate was present. On the transmembrane arm, T141, Q268, and K369 on chain A showed more interaction with cardiolipin when the substrate was present. One amino acid on the transmembrane top region (S402) also showed an increase in cardiolipin interaction when the substrate was present. Amino acids on the transmembrane arm (R151), transmembrane top (N192) and substrate entrance channel (Q435) showed a decrease in cardiolipin interactions when the substrate was present (**Figure 2.3.5-4a**). Further visual inspection of cardiolipin interaction with the protein found that indeed cardiolipin was forming a specific architecture around the protein, interacting specifically with amino acids found in the CLD, transmembrane top, and transmembrane arm/substrate entrance channel regions on chain A where the substrate is bound (**Figure 2.3.5-4c**). For POPE and POPG, interaction increases were also found mostly in chain A with CLD amino acids Q19, N22, N68, T87, N94, N105, and S112, transmembrane arm amino acid T141, transmembrane top amino acid N192, and substrate entrance channel amino acids S312, Q435, and Q436 showing a preference for interaction when the substrate was present. On chain B, substrate entrance channel amino acid S312 and transmembrane top amino acid S402 also showed an interaction preference when the substrate was present. Amino acids in the CLD (chain A S90), transmembrane top region (chain A S189, Q409), transmembrane arm (chain A S272), and substrate entrance channel (chain A Q429, chain B Q435) showed a

decrease in POPE/POPG interaction when the substrate was present (**Figure 2.3.5-4b**). Visual inspection of POPE/POPG interaction with the HlyB protein found no obvious architecture forming (**Figure 2.3.5-4d**). Instead, POPE and POPG lipids would interact and diffuse away throughout the trajectory.

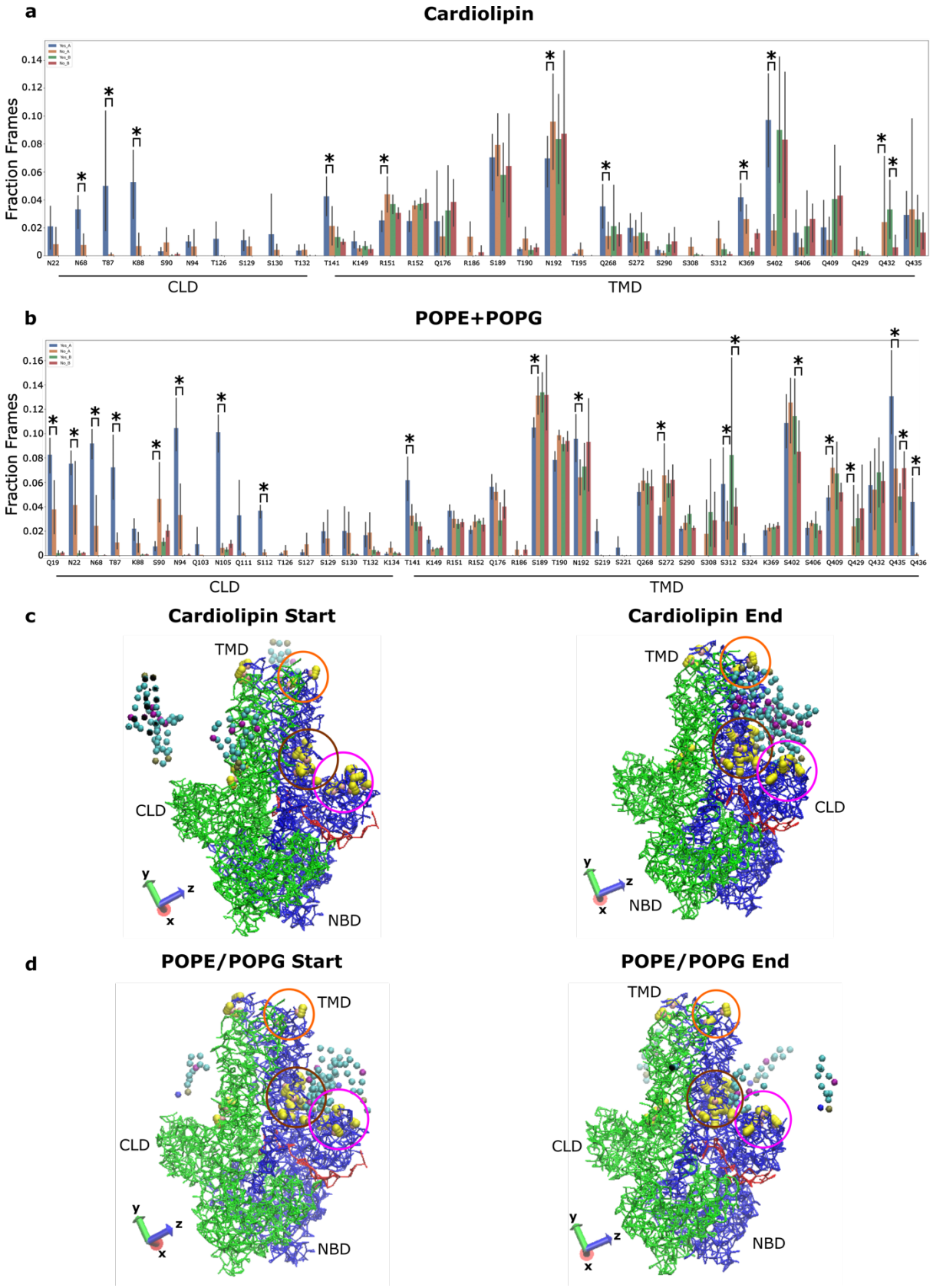


Figure 2.3.5-4. HlyB Lipid Preference. a) Amino acids which interact significantly ($p < 0.01$) with cardiolipin when the HlyA substrate (red) was present were identified. Most of the significant differences were found to be localised to chain A (blue), where the substrate is bound. **b)** Amino acids which interact significantly ($p < 0.01$)

with POPG and POPG when the HlyA substrate was present were identified. Again, most of the significant differences were found to be localised in chain A. **c)** Interaction of cardiolipin with the trajectory at the start (t=0 ns) and end (t=5400 ns) of the simulation. By the end of the simulation, the cardiolipin is localised predominantly around the CLD (pink circle) and transmembrane top (orange circle) regions of chain A (blue). Some interactions could also be seen for the transmembrane arm and substrate entrance channel regions (brown circle). Chain B (green) and HlyA substrate (red) are shown. **d)** Interaction of POPE and POPG with the trajectory at the start (t=0 ns) and end (t=5400 ns) of the simulation. There was no clustering of POPE and POPG with the protein that could be detected. Significance differences between substrate and no substrate trajectories were tested using the Monte Carlo permutation test (n=4 trajectories with substrate vs n=4 trajectories without substrate), with a Bonferroni correction for multiple tests. Error bars show the 90% CI centred around the mean fraction frames for the protein-lipid interactions. Blue: chain A, green: chain B, red: translocating substrate. Still views of protein-lipid interactions were taken from the u2 trajectory.

Lipid preference analysis was also conducted on the PCAT trajectories (**Figure 2.3.5-5**). There were more significant cardiolipin interactions with amino acids in the PEP domain compared to the HlyB trajectories, but these were confined to chain B where the non-translocating substrate is bound; there was a preference for cardiolipin when no substrate is bound for amino acids K8, K9, S105, K107, and T112. Only amino acid Q31 in chain A showed a cardiolipin preference when substrate was present. Transmembrane arm amino acids N153 and T168 (chain A) and transmembrane top amino acid S306 (chain A) showed cardiolipin preference when substrate was present. Substrate entrance channel amino acids K328 (chain B), N332 (chain A), Q450 (chain A and B), and T451, N457 (chain A) also showed a cardiolipin preference when substrate was present (**Figure 2.3.5-5a**). This curious preference for cardiolipin in the transmembrane region when substrate is present contrasts with the HlyB trajectories, which showed that cardiolipin preference occurred almost equally between the CLD and transmembrane regions when the substrate was present (**Figure 2.3.5-4a**). Visual inspection of cardiolipin interactions with the PCAT protein when substrate was present also found a particular architecture that was present at the start and end of the trajectory (**Figure 2.3.5-5c**). For POPE/POPG preference, in the PEP domain T112 (chain B) showed a preferential interaction when substrate was present while T151 (chain A) showed a preferential interaction when substrate was not present. Transmembrane arm amino acids N153, T168 (chain A), transmembrane top amino acid N204 (chain B), and substrate entrance channel amino acids T324, N327, Q446 (chain B), Q450, N457 (chain A) showed a POPE/POPG

preference when no substrate was present. Transmembrane top amino acids S307, N414 (chain A), transmembrane arm amino acid T168 (chain B), and substrate entrance channel amino acid N332 (chain A) showed a POPE/POPG preference when substrate was present (**Figure 2.3.5-5b**). Like the HlyB trajectory, no architecture for POPE/POPG could be seen when the substrate was present (**Figure 2.3.5-5d**).

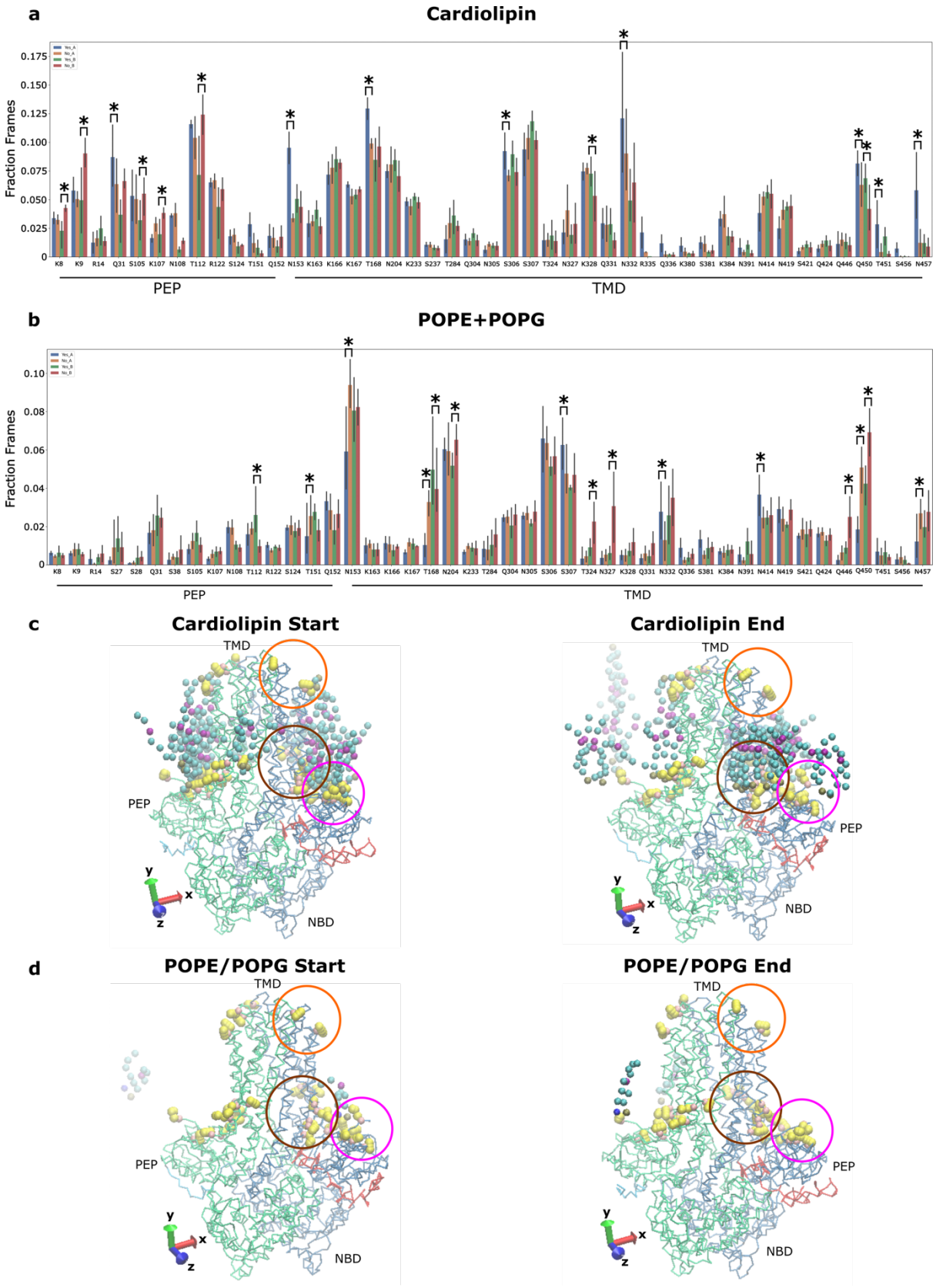


Figure 2.3.5-5. **PCAT Lipid Preference. a)** Amino acids which interact significantly ($p < 0.01$) with cardiolipin when the substrates were present were identified. In the PEP domain, most of the significant differences were found

to be localised to chain B, where the non-translocating substrate is bound. In the transmembrane domain, most of the significant differences were found to be in chain A, where the translocating substrate is bound. **b)** Amino acids which interact significantly ($p<0.01$) with POPG and POPG when the substrates were present were identified. The PEP domain did not show many significant differences between substrate and no substrate trajectories. The transmembrane domain showed significant differences, split evenly between amino acids located in chain A and chain B. **c)** Interaction of cardiolipin with the trajectory at the start (t=0 ns) and end (t=5400 ns) of the simulation. The start and end of the simulation show a similar cardiolipin architecture forming around the protein, with interactions occurring in the PEP (pink circle), transmembrane top (orange circle), and transmembrane arm/substrate entrance channel (brown circle). **d)** Interaction of POPE and POPG with the trajectory at the start (t=0 ns) and end (t=5400 ns) of the simulation. Like the HlyB trajectories (see **Figure 2.3.5-4),** there was no discernible architecture formed when POPE/POPG interact with the PCAT protein. Significance differences between substrate and no substrate trajectories were tested using the Monte Carlo permutation test (n=4 trajectories with substrate vs n=4 trajectories without substrate), with a Bonferroni correction for multiple tests. Error bars show the 90% CI centred around the mean fraction frames for the protein-lipid interactions. Blue: chain A, green: chain B, red: translocating substrate, cyan: non-translocating substrate. Still views of protein-lipid interactions were taken from the u2 trajectory.

3.6 Protein-Lipid Interactions Chain Asymmetry

The data from HlyB and PCAT protein-lipid interactions seem to suggest a curious asymmetry between chains (see **Section 3.5**). For example, HlyB CLD interacted more with cardiolipin in chain A whereas PCAT PEP domain interacted more with cardiolipin in chain B. Interaction asymmetries could arise due to the finite size of the box; in nature, any particular lipid could interact with a protein and then diffuse far enough away to not interact again. In simulation, the box size prevents long-distance diffusion and hence any particular lipid may interact many more times with the protein. The finite box size, coupled with short simulation times, could produce results where some protein-lipid interactions statistically fall on the extremes but may not be biologically meaningful. One way to test for a biological component is to analyse protein-lipid interactions across many different homologous structures in different species; preservation of protein-lipid contacts at structurally homologous amino acid sites, and the presence of an asymmetry, may then suggest a biological mechanism (*e.g.,* Stansfeld, Jefferys & Sansom, 2013).

I analysed protein-lipid interactions from a group of structures which are homologous to both HlyB and PCAT. There were several PCAT structures which were solved in different states: PCAT with bound substrate, PCAT with bound γATP, PCAT with bound substrate and γATP in the inward-facing wide conformation, PCAT with bound substrate and γATP in the inward-facing intermediate conformation, and PCAT with bound substrate and γATP in the inward-facing narrow conformation. There is also PCAT with bound Mg^{2+} in the outward facing conformation, but this structure is not in the MemProtMD database. The results for chain preference for protein-lipid interactions is shown in **Figure 2.3.6-1**. The PCAT structure with bound substrate and ATP in the inward-facing wide conformation had the most disparity with chain preference in all the PCAT structures examined (**Figure 2.3.6-1b**). To determine where the DPPC lipid was interacting with protein, and if there was any difference between the MemProtMD simulations and my results, the available structures were superimposed and amino acids which showed differences in lipid interaction preference highlighted (**Figure 2.3.6-2**). The superimposition shows that there is no asymmetry between the chains when interacting with lipid; both chain A and chain B are as equally likely to show a lipid preference. The most striking difference is between the DPPC (**Figure 2.3.6-2a**) and cardiolipin with POPE/POPG (**Figure 2.3.6-2b).** When only DPPC is present, even with substrate lipid preference is confined to transmembrane arm and top regions (**Figure 2.3.6-2a**). However, the presence of cardiolipin and POPE/POPG seems to drive interactions with the PEP and substrate entrance channel regions (**Figure 2.3.6-2b**). This may reflect a combination of substrate-driven preference coupled to the lipid environment.

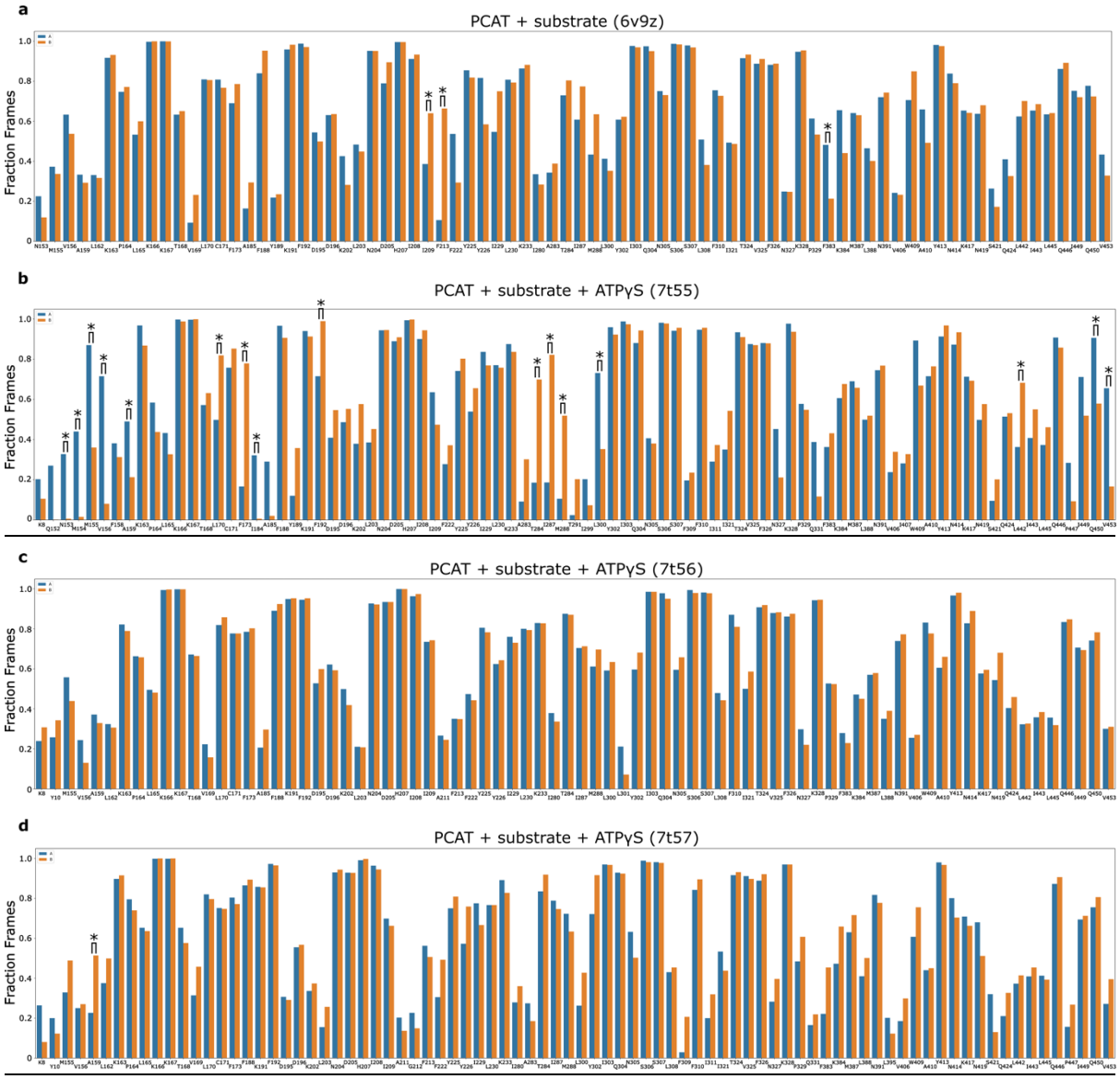


Figure 2.3.6-1. PCAT Chain Preference for DPPC. a) Chain preference results for PCAT with bound substrate (pdb ID 6v9z). **b)** Chain preference results for PCAT with bound substrate and ATP in the inward-facing wide conformation (pdb ID 7t55). **c)** Chain preference results for PCAT with bound substrate and ATP in the inward-facing intermediate conformation (pdb ID 7t56). **d)** Chain preference results for PCAT with bound substrate and ATP in the inward-facing narrow conformation (pdb ID 7t57). Statistical significance ($p<0.05$) was determined by the Monte Carlo permutation test for each individual simulation.

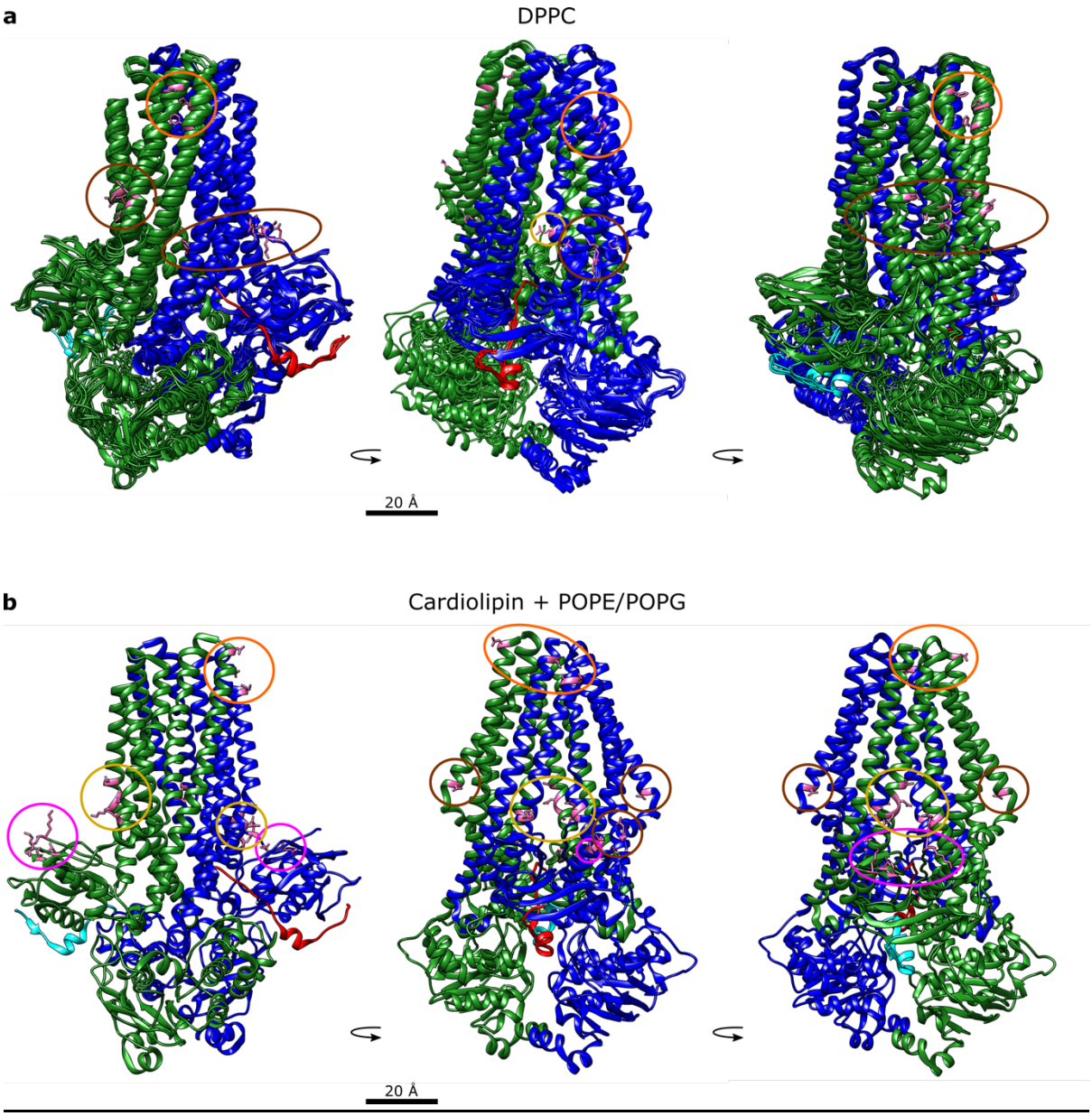


Figure 2.3.6-2. Structural Alignment of PCAT. a) PCAT structures in different conformations (pdb IDs 6v9z, 7t55, 7t56, 7t57) were superimposed and amino acids which showed a chain preference for interacting with DPPC lipid are highlighted (pink). Both translocating (red) and non-translocating (cyan) substrates are show bound to their respective chains (Chain A, blue; Chain B, green). When interacting with DPPC, amino acids in the transmembrane arm (brown circle) and transmembrane top (orange circle) drive preferential interactions**. b)** PCAT interaction with cardiolipin and POPE/POPG drives amino acid preference to include both the PEP domain (pink circle) and substrate entrance channel (gold circle). For both the MemProtMD results (a) and my simulation results (b) there is no asymmetry in the chains when interacting with lipid. Atomistic structures are shown for illustrative purposes.

Next, I analysed DPPC chain preference for structures homologues to HlyB and PCAT. This included two states of microcin antibiotic transporter McjD and the protease transporter PrtD (**Figure 2.3.6-3**). Both structures lack a CLD/PEP domain equivalent.

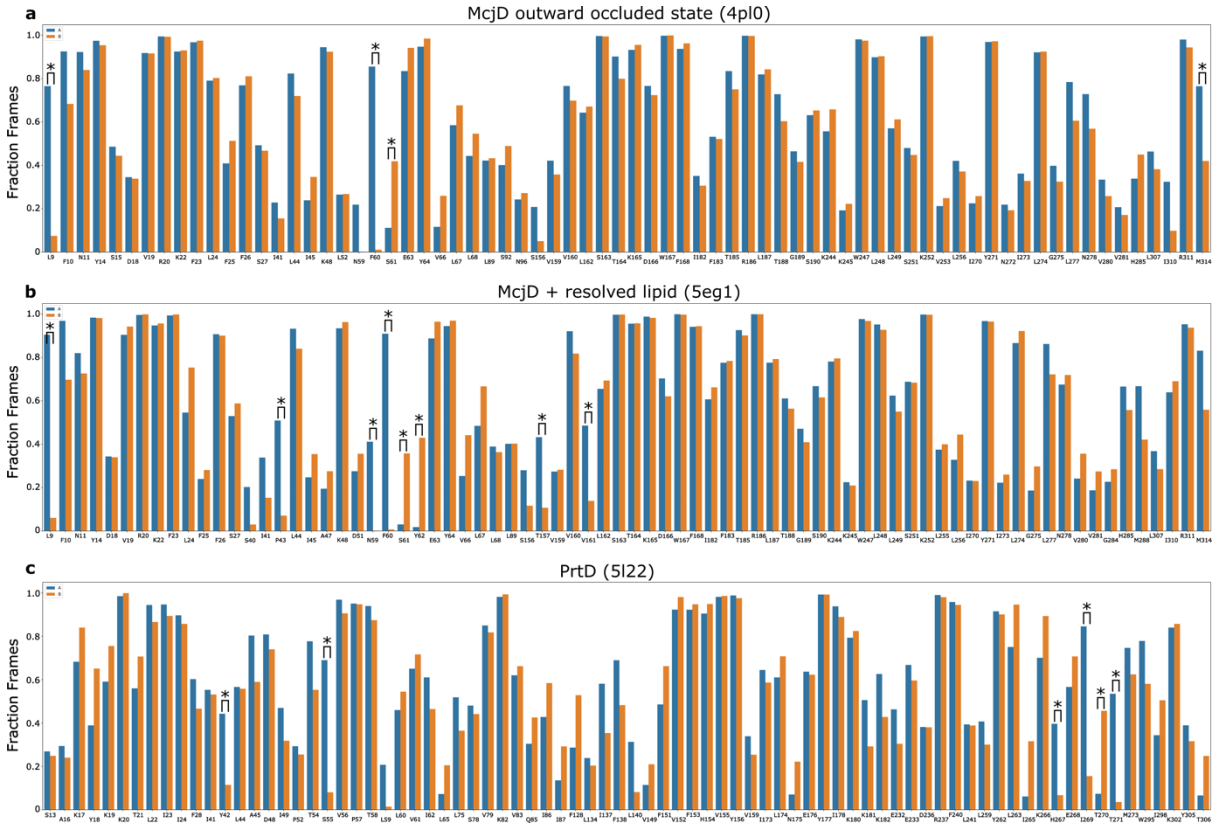


Figure 2.3.6-3. HlyB/PCAT Structural Homologues Chain Preference for DPPC. a) Chain preference results for *E. coli* McjD in the outward occluded state (pdb ID 4pl0). **b)** Chain preference results for *E. coli* McjD with a resolved lipid (pdb ID 5eg1). **c)** Chain preference results for *Aquifex aeolicus VF5* protease transporter (PrtD) (pdb ID 5l22). Statistical significance ($p<0.05$) was determined by the Monte Carlo permutation test for each individual simulation.

The chain preference results show some divergence from the PCAT structures. All structures examined show 4–8 sites where a chain preference appears (**Figure 2.3.6-3**), in contrast to PCAT where only one structure showed a large set of preferential interactions **(Figure 2.3.6-1**). Intriguingly, the ratio of chain A to chain B preferential interactions with DPPC was 12:3 suggesting a degree of asymmetry in the interactions. Structural analysis showed that in DPPC, amino acids were found predominantly in the transmembrane top region with a couple in the transmembrane arm region (**Figure 2.3.6-4a**). Like PCAT, the presence of the lipid

environment seems to drive recruitment of the CLD and substrate entrance channel regions in preferentially interacting with lipid. However, unlike in PCAT structures, HlyB shows a marked chain asymmetry (**Figure 2.3.6-4b**). HlyB chain A is 6 times more likely to interact with lipid than chain B (ratio 24:4 amino acid sites) (**Figure 2.3.6-4b**). This is a stronger asymmetry preference than was found for the structural homologues of HlyB and PCAT in DPPC (4 times more likely to interact with chain A than chain B). This stronger asymmetry interaction preference may also reflect the presence and absence of substrate, in combination with the lipid environment as a driving factor.

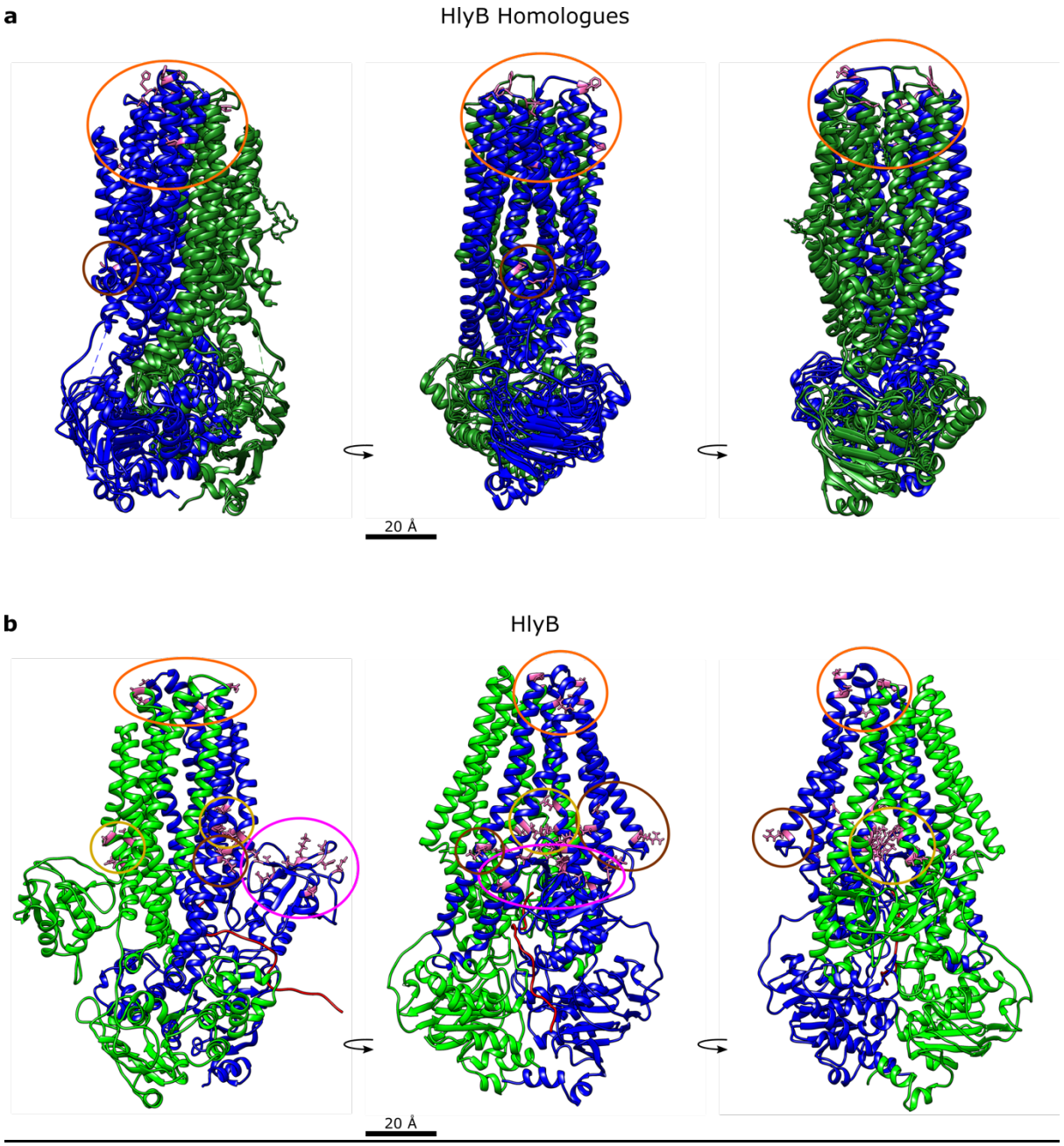


Figure 2.3.6-4. Structural Alignment of Homologues for HlyB. a) HlyB-like structures showed a strong chain A (blue) over chain B (green) preference for DPPC interaction with amino acids predominantly found in the transmembrane top (orange circle) and a few in the transmembrane arm (brown circle) regions. Note that these proteins do not have a CLD/PEP domain like HlyB/PCAT. **b)** HlyB also shows a strong chain A preference (blue) with interacting amino acids found also in the CLD (pink circle) and substrate entrance channel (gold circle) regions. For both the MemProtMD results (a) and my simulation results (b) there is a strong asymmetry in the chains when interacting with lipid. Atomistic structures are shown for illustrative purposes.

4. Discussion

4.1 Signal-Independent Secretion

The structure of the HlyB/D complex was published recently by Zhao, Lee & Chen (2022). The published structure in the apo (pdb ID 7sgr) and ATP-bound (pdb ID 8dck) states solved some questions on the transportation process. It was suggested by previous cross-linking experiments that HlyB/D acts as a complex to recruit TolC when HlyA is bound (Thanabalu *et al*, 1998; Balakrishnan, Hughes & Koronakis 2001). Zhao, Lee & Chen (2022) found that the HlyB/D complex consists of a HlyB trimer with hexameric HlyD (3:6 ratio) (**Figure 2.4.1-1**). They conducted photo-cross linking experiments to propose a substrate entrance channel through the CLD and TMD of one open HlyB monomer, rather than through the central pore. Interestingly, they also found that one CLD in each monomer was very flexible with the other CLD acting as a bridge between the trimers. The published structure confirms that my choice of simulating one unfolded HlyA monomer binding to one HlyB CLD is correct. However, since I only simulated one dimer of HlyB with and without bound HlyA substrate, the role of the other two promoters in the transportation mechanism remains elusive. In this discussion, I attempt to fill in this knowledge gap using my work from Chapter II, the published HlyB/D structure, and the available literature. In **Chapter VI**, I dive deeper into transport mechanisms for the T1SS using bioinformatic analyses in addition to the above lines of evidence.

Comparison of my HlyB-HlyA homology model with the HlyB/D complex found that they were in agreement in the TMD and NBD. The CLD, however, was different in its relative orientation. However, given the inherent flexibility in both PCAT PEPs (Lin, Huang & Chen, 2015; Kieuvongngam *et al*, 2020; Kieuvongngam & Chen, 2022) and HlyB CLD, it is possible that HlyB CLD adopts many different conformations and that there may be a specific substrate-binding induced conformation yet to be determined experimentally. Indeed, as suggested by Zhao, Lee & Chen (2022) it is feasible that the flexible CLD in one monomer acts to bind and transport HlyA while all three trimers can hydrolyse ATP (**Figure 2.4.1-1**). Models of transport will be discussed further in **Chapter VI**.

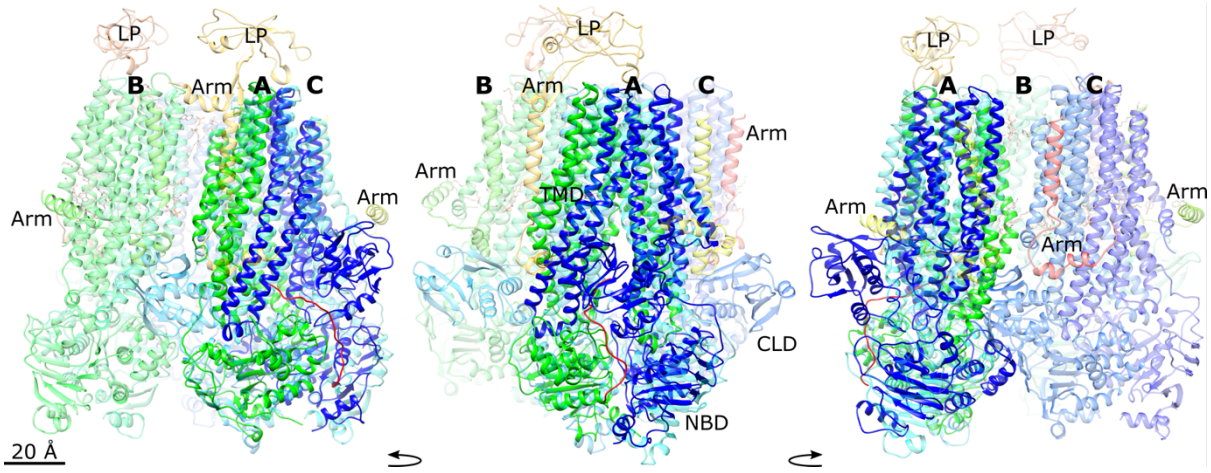


Figure 2.4.1-1. Comparison of HlyB/D complex with HlyB-HlyA homology model. HlyB homology model chain A (blue) and chain B (green) with HlyA substrate (red) overlaid with HlyB/D complex (transparent) (pdb ID 7sgr) (Zhao, Lee & Chen, 2022). HlyB/D complex structure was solved as a trimer (protomers A, B, C) with resolved domains of hexameric HlyD (LP: lipolyl domain; Arm: N-terminal cytosolic arm). The arms of HlyD wrap around the transmembrane domain (TMD) of each HlyB protomer, while the C39-like domain (CLD) acts as a bridge between protomers. Each protomer nucleotide-binding domain (NBD) also makes contact with other protomer NBDs. Note that only one CLD for each protomer was resolved. My HlyB homology model shows a possible substrate-binding conformation at the entrance of the open TMD.

Modelling HlyB with 23 C-terminal HlyA amino acids allowed me to conduct tractable MD simulations with the knowledge at the time. I found that the main driving force in HlyA binding is the interaction of charge-charge and polar residues between the substrate and HlyB protein. This was found in both HlyB and PCAT coarse grain MD simulations, suggesting a conserved mechanism. The coarse grain models suggested that charge-charge N-terminal tail stabilisation aided correct pinning of the substrate into the HlyB cavity, facilitating charge-charge and polar contacts in the C-terminal region. In the atomistic simulations, tail stabilisation was partly disrupted, leading to more unfavourable contacts of hydrophobic residues in the substrate with charged and polar residues in the binding cavity. The findings fit with previous experimental results, which show that HlyA secretion does not require a specific sequence but does require a biochemical pattern of charged and small hydrophobic residues (Mackman *et al*, 1987; Felmlee & Welch, 1988; Gray *et al*, 1989; Koronakis, Koronakis & Hughes, 1989; Jarchau *et al*, 1994; Hui *et al*, 2000). The octet GG repeat in the C-terminal domain found by Felmlee & Welch (1988) to be conserved in a range of bacterial species was later discovered to bind Ca^{2+} ions which aid folding upon export (Welch, 1991). Later work

established that a small amphiphilic helix was the only requirement for export (Hui *et al*, 2000). In PCAT peptidase, the GG-motif acts to position the sequence for cleavage by the catalytic triad (C21, H99, and D115) (Kieuvongngam *et al*, 2020). Indeed, it was found in my atomistic MD simulations that the analogous residues to C39 peptidases catalytic triad residues C21 and H99 (Y9, and H83) did not interact significantly with the substrate. This result broadly fits with previous NMR (Lecher *et al*, 2012) and recent HlyB/D-HlyA (Zhao, Lee & Chen, 2022) studies, showing that H83 forms a π- π stacking interaction with W77 (see **Figure 2.3.4.2-2**). Thus, it would appear that the lack of conserved sequence information for HlyA-like substrates is likely due to the mode of transport, which only requires a conserved physiochemical pattern. A complete description of likely mechanisms will be discussed further in **Chapter VI**.

4.2 Cooperativity in NBD Substrate Binding

Lacking in previous studies of HlyB-HlyA interaction is the role of HlyB NBD in HlyA transport. In my coarse grain and atomistic MD results for HlyB and related PCAT, the NBD seemed to act as a gatekeeper for access to the transmembrane domain forming polar or charged contacts with the substrate (see **Figures 2.3.4.1-1**, **2.3.4.1-2**, and **2.3.4.2-2**). Loss of NBD contact seemed to be a feature in the non-translocating peptide for PCAT (see **Figure 2.3.4.1-3**). Intriguingly, NBD contact to the substrate is made primarily by the opposite chain in both HlyB and PCAT (see **Figures 2.3.4.1-1**, **2.3.4.1-2**, and **2.3.4.2-2**), except when the substrate is non-translocating (see **Figure 2.3.4.1-3**). Analysis of protein movements in my coarse grain and atomistic simulations found it to be primarily in the NBD and CLD, with contractions and expansions occurring in the opposite chain NBD to accommodate substrate movement (see **Figure 2.3.3-2**). Considering that unlike in other ABC-transporters, HlyB NBD dimers are likely in close contact during ATP binding (Jones & George, 2009), this substrate binding arrangement may induce a cooperative effect in the transportation process, possibly stimulating ATP binding and hydrolysis to move the system from an equilibrium state to non-equilibrium movement under ATP turnover (Rahman & Mchaourab, 2020; Kieuvongngam & Chen, 2022).

4.3 Cardiolipin Aided Transport

I found an enrichment of cardiolipin within 20 Å of HlyB and PCAT, irrespective of the presence of substrate. Cardiolipin interaction with a range of membrane embedded proteins has been found previously in coarse grain simulations. In these cases, asymmetry in cardiolipin-protein interaction between the upper and lower membrane leaflets was found, facilitated by arginine/lysine interactions with the cardiolipin phosphate heads (Corey *et al*, 2021). In my study, I found that further analysis of lipid clustering in the HlyB trajectories found that all three lipids (POPE, POPG, and cardiolipin) are enriched in the CLD where the substrate was bound, and that this enrichment was only found when the substrate was present. This analysis was confirmed by SASA analysis of critical CLD lipid interacting residues. There was also enrichment of POPE in at the top of the transmembrane when the substrate was present. This lipid clustering result was not found in PCAT, where lipid depletion in the top of the transmembrane was observed in the presence of substrate. Again, SASA analysis confirmed this observation for critical PEP lipid interacting residues. The results for HlyB seem to suggest that lipid could potentially act as an anchor for the CLD when the substrate is bound, limiting some of its movement (see **Figure 2.3.5-3c**). Indeed, in one of the atomistic trajectories the substrate unbound CLD showed more RMSD movement (see **Figure 2.3.3-2**). Analysis of HlyB and PCAT trajectories in the membrane found that the membrane acts as a sheet on the substrate bound CLD of HlyB, while in PCAT only the N-terminal point acts to anchor the PEP domain to the membrane (see **Figure 2.3.5-3c)**. Analysis of specific amino acids contributing to lipid preference found that they clustered at different sites – CLD, transmembrane arm, transmembrane top, and substrate entrance channel. The bulk of the cardiolipin preference interactions occurred in the CLD and substrate entrance channel where the substrate was bound for HlyB, forming an asymmetric pattern of interaction (see **Figure 2.3.6-4**). This was in contrast to PCAT, where no such asymmetry was found (see **Figure 2.3.6-2**). The presence of asymmetry and preferential cardiolipin interaction seemed to suggest that cardiolipin could be forming a particular architecture around HlyB when the substrate was present (see **Figure 2.3.5-4**), a result not found in the PCAT simulation (see **Figure 2.3.5-5**).

Analysis of these results in light of the recently published HlyB/D complex (Zhao, Lee & Chen, 2022) found that two amino acids on the transmembrane arm which interact preferentially with cardiolipin in my simulations (T141 and K369) interact with the HlyD arm. In addition, it was shown in previous studies that this N-terminal arm of HlyD is crucial for secretion, with deletion of the first 60 amino acids abolish HlyA secretion while leaving the HlyB-HlyD and HlyB-HlyA interaction intact (Balakrishnan, Hughes & Koronakis 2001). Analysis of the HlyB/D structure in light of this leads to the conclusion that it is unlikely for HlyD to interact directly with HlyA due to the presence of the HlyB CLD. Another cardiolipin preference amino acid in the transmembrane top region (S402) was also shown in a previous study to be secretion defective when combined with another nearby mutation (S402P/D404K) (Blight *et al*, 1994) (**Figure 2.4.3-1**).

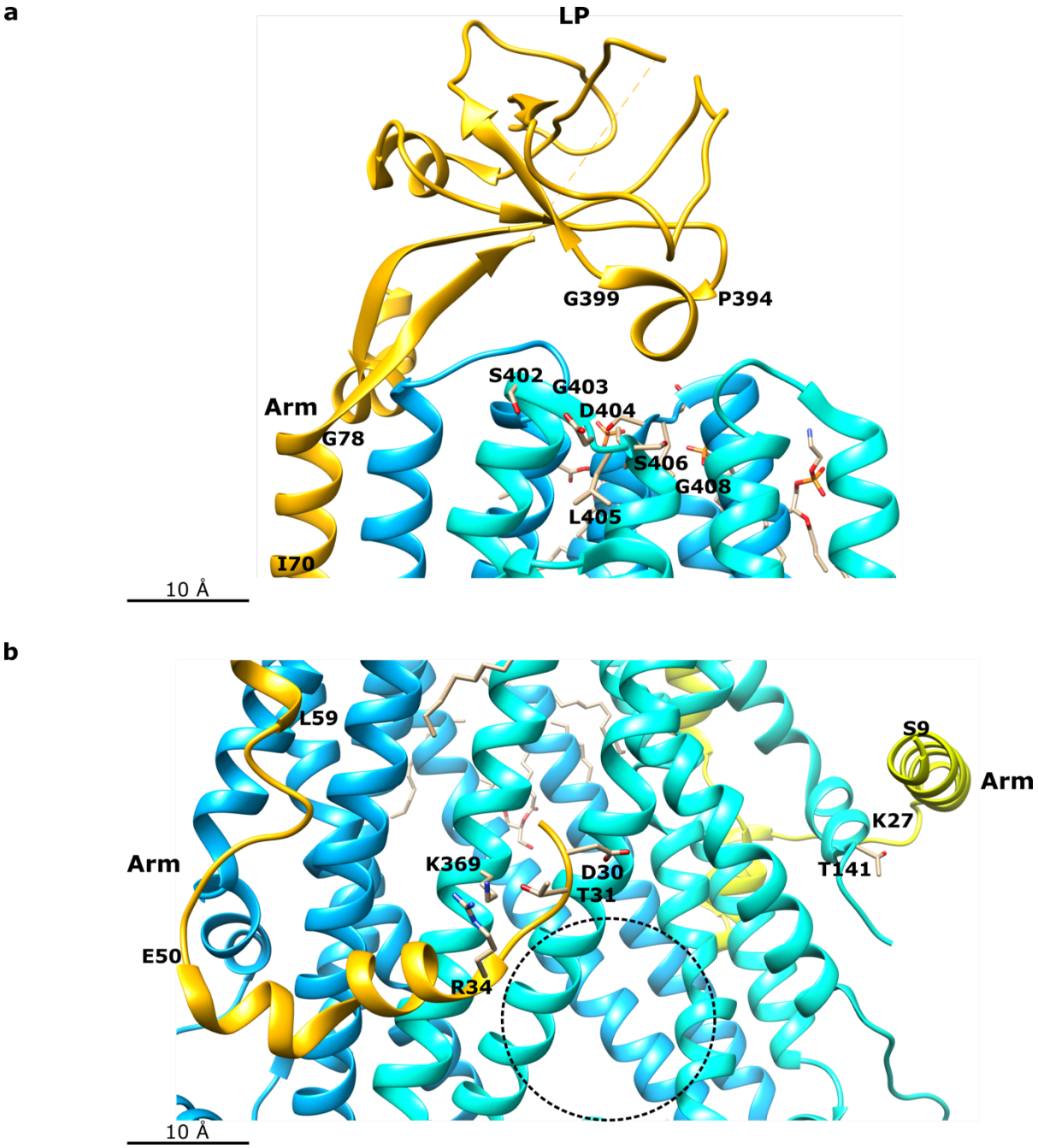


Figure 2.4.3-1. HlyB interaction with HlyD. **a)** Conserved loop S402–G408 (Blight *et al*, 1994) shown with HlyD lipolyl domain (LP) region G399–P394. S402 in this chapter was shown to preferentially interact with cardiolipin. **b)** K369 and T141 on the transmembrane arm region of HlyB (cyan) are shown with respective HlyD arm regions D30–R34 (orange) and S9–K27 (yellow). K369 and T141 in this chapter is shown to preferentially interact with cardiolipin. The first 60 amino acids of HlyD are known to be critical for HlyA translocation (Balakrishnan, Hughes & Koronakis, 2001), although it is not likely to be a direct interaction based on the HlyB/D complex (Zhao, Lee & Chen, 2022). Approximate area where substrate should enter via the CLD is shown (dotted, circle). Structured used is HlyB/D complex (pdb code ID 7sgr) (Zhao, Lee & Chen, 2022).

It was found by Koronakis, Hughes & Koronakis (1991) that HlyA transport may require a proton motive force in addition to ATP hydrolysis, although this result remains contentious (Kanonenberg *et al*, 2018). The presence of a proton motive force in transport has been found for the Type III secretion system (Lee & Rietsch, 2015), as well as the SecYEG system (Corey *et al*, 2018; Ryabichko *et al*, 2020). It was found in the SecYEG system that cardiolipin could act as a proton shuttle to power the proton motive force, due to the distinct pKas of the head group. Moreover, its presence also stimulated ATP hydrolysis (Gold *et al*, 2010). Under simulation, cardiolipin bound transiently to specific sites (Corey *et al*, 2018). Analysis of a cardiolipin deficient bacterial strain found that SecYEG stability was compromised in addition to alkaline phosphatase secretion ability (Ryabichko *et al*, 2020). This presents an intriguing hypothesis, whereby HlyA secretion by HlyB is mediated by the formation of a specific cardiolipin architecture, which could function to stimulate ATP-hydrolysis of the HlyB NBD dimers and as a conduit of the proton motive force.

5. Conclusion

I made a HlyB-HlyA homology model and conducted MD simulations to provide insights into mechanisms of HlyA T1SS transport. I found that HlyB interacts with HlyA via charge-charge and polar contacts, with HlyA tail stabilisation contributing to the favourable contacts. I also found that the lipid environment is impacted by substrate binding; in the presence of substrate, cardiolipin starts to form an architecture around HlyB in the simulation time, driven by specific asymmetric contacts in the CLD and transmembrane regions. I found that the mode of substrate binding is similar between HlyB and related PCAT, with NBD contacts forming a gatekeeping mechanism to allow substrate access to the transmembrane channel. In addition, it is possible that HlyB and PCAT NBDs act in a cooperative manner to facilitate translocation. Interaction with the lipid environment was not as conserved, with PCAT showing a general cardiolipin enrichment but no specific architecture in the simulation time. I propose that the presence of cardiolipin aids HlyA substrate transport by acting as a proton reservoir for the proton motive force, and possibly by stimulating ATP hydrolysis.

Chapter III: Optimising Type I Secretion System Expression & Purification

1. Introduction

1.1 Maintaining the Lipid Environment

Membrane proteins are often challenging to express and purify (Carpenter *et al*, 2008; Hardy *et al*, 2016). Unlike water-soluble proteins, membrane proteins require successful replication of the membrane environment upon extraction. Membranes are not passive environments: they are complex and fluid-like, and interaction with the membrane-embedded protein is often critical for function (Jiang *et al*, 2022). Membrane proteins have historically been divided into three broad classes, depending on their extent of membrane-association: integral membrane proteins, peripheral membrane proteins, and lipid-anchored (Alberts *et al*, 2002). More recent work has attempted to use computational methods for accurate classification (*e.g.,* Guo *et al*, 2019). Membrane proteins in each class interact with the membrane in different ways – for examples, channels and pores have extended hydrophobic cores to interact with the hydrophobic fatty-acid tails, and hydrophilic outer-regions. In contrast, membrane-associated proteins often have a small covalently attached lipid chain for embedding in the membrane (Alberts *et al*, 2002), although there are exceptions (*e.g.* PH domains (Lemmon, 2007) .

Although protein-membrane interactions are complex, their interactions are likely to impact on the structure and activity of the membrane protein (Jiang *et al*, 2022). This means that the membrane-protein interactions must be maintained when extracting membrane proteins from their native environment for structural and functional characterisation.

1.2 Membrane Protein Expression and Purification Workflow

Membrane protein extraction and purification follows a typical workflow (**Figure 3.1.2-1**). Since the yield of membrane proteins is very low, attention must be paid to the amount

produced by the host organism (Grisshammer, 2006). This includes optimising host strains, vector type, and expression conditions. Bacterial strains are popular choices due to their low-cost of growth, especially if high yields are required for downstream processes such as structural studies (Rosano & Ceccarelli, 2014). The correct choice of bacterial strain will depend on balancing the levels of transcription with what the cell can tolerate (Mathieu *et al*, 2019). BL21(DE3) bacterial cells are a popular choice for protein production, however some overexpressed proteins can display a toxic effect on growth in this strain. The C41(DE3) and C43(DE3) cell lines were developed by Miroux & Walker (1996) to tolerate specific integral membrane protein production under the bacteriophage T7 promotor. One mechanism for successful production in C41/C43(DE3) cells is the ability of these cells to limit plasmid instability (Dumon-Seignovert, Cariot & Vuillard, 2004). Later investigations established mutations in the *lac*UV5 promotor controlling T7 RNA polymerase expression were responsible for improved membrane protein expression, potentially by allowing for proper insertion into the membrane by not saturating the Sec translocon (Wagner *et al*, 2008a). However, despite these advances in membrane protein production in bacterial cells, not enough is known about the inter-play of bacterial genetics with the biogenesis of the protein of interest to offer a 'one size fits all' choice of optimal strain (Hattab *et al*, 2015).

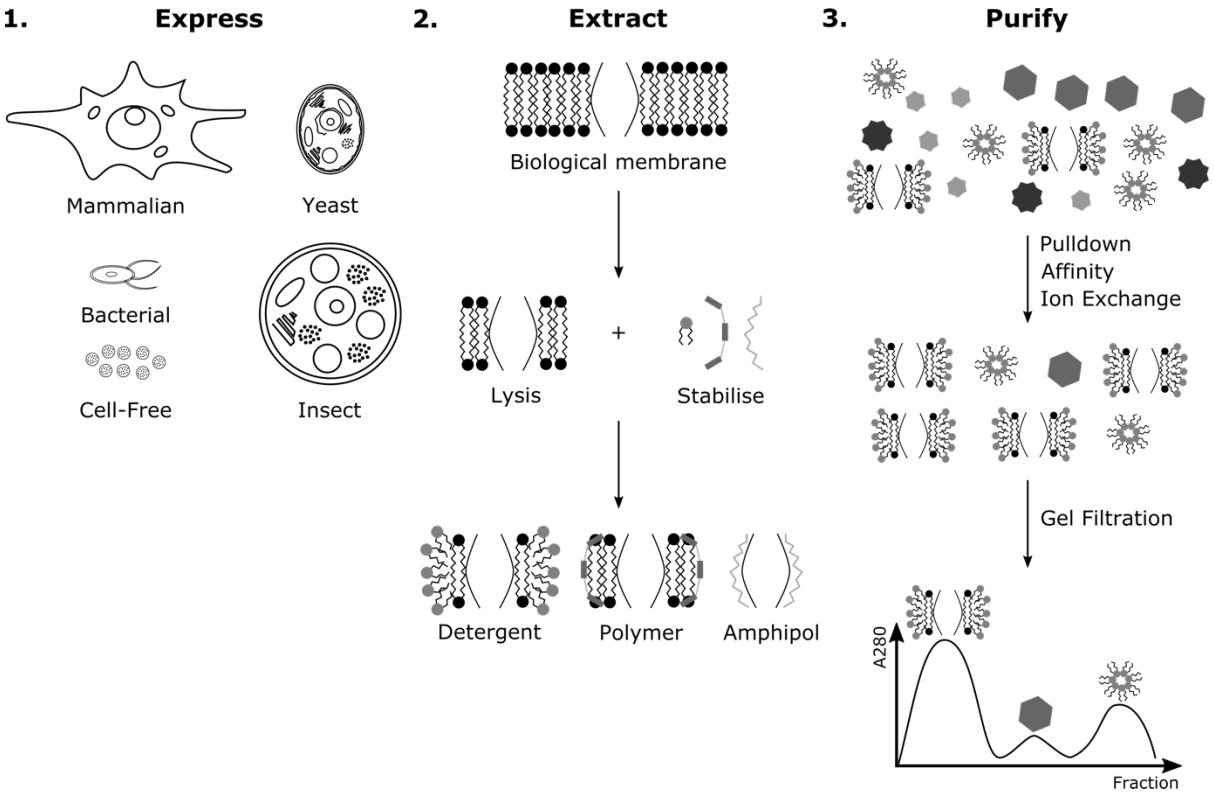
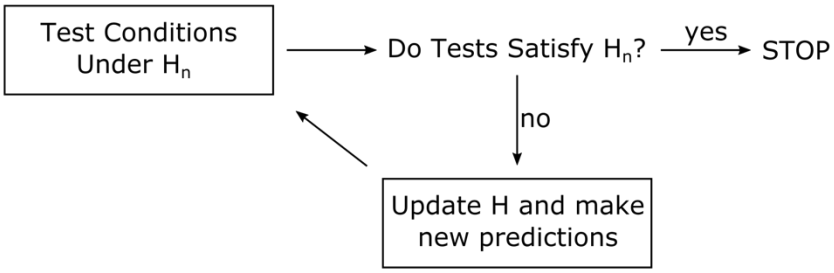


Figure 3.1.2-1. Membrane protein purification workflow. Purifying membrane proteins follows a typical workflow: 1) find a suitable expression host for the target, 2) once the expression host and conditions are found, the target must be extracted from the biological membrane and stabilised, for example through the use of detergents, polymers, or amphipols, and 3) once the target is extracted, it must be separated from other cellular components through purification, for example through an initial pulldown, affinity chromatography, or ion exchange step followed by a gel filtration clean-up step to obtain pure protein without contaminants such as other proteins or detergent micelles.

Once the host has been established, expression conditions must also be determined to maximise yield (Kubicek *et al*, 2014). Due to the multiple combinations of parameters that are possible, screening every condition with replicates is not feasible. However, one can do a sparse screen of initial conditions and narrow down until the successful set is found. One experimental design which is often very successful in multi-parameter optimisation is the sequential design strategy (Lu & Anderson-Cook, 2021). This works by starting with a test set-up, collect results, and then design the next test using the results of the previous set (**Figure 3.1.2-2**). In this set-up replicates are between experiments, saving time and resources while allowing for robust statistics of the output results especially around the region of interest.

Although this method has been applied in some cases (*e.g.,* Rossell & Müller, 2013), it is not widely utilised in protein expression optimisation.

a



b

H_o: Initial conditions will produce T1SS in C43 strain

Test: LB or 2xYT media in C43 strain

Result: not detected above noise

H_1: Initial conditions will produce T1SS under single induction in C43 strain

Test: LB or 2xYT media with expression order in C43 strain

Result: not detected above noise

H_2: LB media at 37°C under single induction conditions will produce T1SS in C43 strain

Test: LB or 2xYT media with BL21(DE3) or C43 strain at 0 mM or 5 mM $CaCl_2$

Result: BL21(DE3) in 2xYT at 37°C produces more T1SS at 5 mM $CaCl_2$ than 0 mM $CaCl_2$

H_3: BL21(DE3) strain will produce T1SS in 2xYT at 37°C under 5 mM $CaCl_2$

Test: Strain (C43, BL21(DE3)), media (2xYT, LB), temperature (20 or 37°C) under 5 mM $CaCl_2$

Result: BL21(DE3) produces most T1SS in LB at 37°C with 5 mM $CaCl_2$

H_4: BL21(DE3) strain will produce T1SS in 2xYT at 37°C under 5 mM $CaCl_2$ using same time of expression

Test: order of induction and induction concentration

Result: BL21(DE3) produces more T1SS under single induction than double induction when HlyB/D is expressed 30 min before HlyA

H_5: BL21(DE3) strain will produce T1SS in 2xYT at 37°C under 5 mM $CaCl_2$ when HlyB/D expressed first under 100 µg/mL carbenicillin with 30 µg/mL kanamycin

Test: antibiotic and $CaCl_2$ concentration

Result: similar T1SS production under tested conditions

.
.
.
.

H_n: BL21(DE3) strain will produce T1SS in 2xYT at 37°C under 5 mM $CaCl_2$ when HlyB/D expressed first under 100 µg/mL carbenicillin with 30 µg/mL kanamycin when total induction time is not more than 2 hours

Test: antibiotic concentration and induction time

Result: tests satisfy hypothesis

Figure 3.1.2-2 (previous page). Sequential design strategy used in T1SS expression optimisation. a) The sequential design strategy follows an iterative procedure of conducting tests of a hypothesis – if tests satisfy the hypothesis then the experiment is stopped, otherwise it proceeds under an updated hypothesis with new predictions. **b)** Example sequential design strategy for optimising T1SS production. Hypotheses and tests are updated until the n^{th} iteration where the hypothesis was satisfied, and the study stopped.

The next consideration after optimising membrane protein expression is an efficacious solubilisation and purification strategy. There are different ways of maintaining protein-membrane interactions when extracting membrane proteins: the use of detergents in the initial stages is common practice, especially for integral membrane proteins. Since detergents interact differently with each membrane protein, they must be screened for compatibility with the membrane protein under investigation (Lin & Guidotti, 2009). In a typical purification workflow, detergent must be present above the critical micelle concentration to maintain protein-lipid interactions (Orwick-Rydmark, Arnold & Linke, 2016). If the detergent concentration is too low, the membrane protein becomes unstable as it is stripped of lipid and hydrophobic regions come into contact with polar water molecules (Lin & Guidotti, 2009). In a typical purification protocol, the membrane portion of the cell is extracted upon cell lysis. The membrane is then initially solubilised in a relatively high detergent concentration (0.1–10% (w/v)) (Lin & Guidotti, 2009). This is to disrupt protein-lipid and lipid-lipid interactions in the native membrane and reconstitute the protein in detergent micelles (le Maire, Champeil & Møller, 2000). In later purification steps, the detergent concentration is dropped significantly (<0.1%) so that it does not interfere with the purification procedure but still maintains the stability of the membrane protein (Lin & Guidotti, 2009). Excess detergent can be removed by a number of methods, including dialysis and chromatographic methods. Although detergents are good at initial solubilisation of membrane proteins, they can be destabilising and deactivate the protein (Seddon, Curnow & Booth, 2004). Detergent can be exchanged for amphipols which are better at mimicking the native lipid environment (Tribet, Audebert & Popot, 1996).

1.3 Aims and Objectives

A trapped form of the T1SS was produced by adding an N-terminal fast-folding eGFP tag onto the substrate HlyA (Lenders *et al*, 2016). However, there is no optimised expression or purification protocol for the T1SS which hampers structural investigation. Hence, optimising for expression and testing different purification strategies would aid in determining the best strategy for producing sufficient T1SS for structural studies. In **Chapter II**, it was found that cardiolipin may play a role in Type I Secretion System (T1SS) substrate secretion. Examining levels of T1SS in a cardiolipin deficient bacterial strain would provide a test for determining if cardiolipin is required for substrate secretion. By comparing T1SS levels in the deficient strain to the parental strain, clues would be provided as to the need for cardiolipin in transport.

The Aims of this Chapter involves answering the following questions:

1) What are the best expression conditions for the T1SS?
2) What is the best purification strategy for obtaining the full T1SS with trapped transport substrate?
3) How can I test the cardiolipin requirement as suggested by Chapter II?

These Aims will be addressed with the following Objectives:

1) Test different detergents to find the best one for solubilising the T1SS
2) Use flow cytometry on antibody labelled T1SS to screen expression conditions and optimise by a sequential design strategy
3) Optimise expression conditions for a cardiolipin deficient strain of *E. coli* and compare the results to the expression optimised parent strain
4) Test use of cryo-correlative light electron microscopy (cryo-CLEM) to see the distribution of T1SS
5) Use these results to help inform a purification strategy for the T1SS

2. Materials & Methods

2.1 Chemically Competent Cells

Bacterial cells were made chemically competent via an in-house method. Starter culture was made by inoculating a single colony in 5 mL LB media and grown overnight at 37°C with shaking at 200 rpm. 250 mL LB with 20 mM $MgSO_4$ was inoculated with 2.5 mL starter culture and grown at 37°C with shaking at 200 rpm until it reached OD_{600nm} 0.4–0.6. Cells were pelleted by centrifugation at 4200 rpm (Beckman JS-4.2, Beckman, California, USA) at 4°C for 15 minutes. Cells were resuspended in cold 100 mL TFB1 (30 mM potassium acetate, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 100 mM RbCl, 15% glycerol, pH 5.8) and left on ice for 5 minutes. Cells were then pelleted by centrifugation at 5000 g for 10 minutes at 4°C. Cells were resuspended in 10 mL TFB2 (10 mM MOPS, 75 mM $CaCl_2$, 10 mM RbCl, 15% glycerol, pH 6.5) and left on ice for 40–60 minutes before being aliquoted and flash frozen in liquid nitrogen.

2.2 Bacterial Cell Transformation

Plasmids containing HlyBD (pK184-HlyBD) and HlyA (pSOI-eGFP-HlyA-FLAGx3-Linker) were obtained from Lutz Schmitt (University of Dusseldorf, Germany). Chemically competent bacterial cells (BL21(DE3), C43, C43ΔAcrAB, MG1655, MG1655Δcls) were transformed with both plasmids (~150 ng each) by heat-shock. Plasmid was incubated with 100 µl cells on ice for 10 minutes before placing in 42°C heating block for 45 seconds followed by 2 minutes on ice. Cells were recovered in 2xYT media at 37°C with shaking 200 rpm for 30–40 minutes before being spread on the appropriate selection plate (50 µg/ml kanamycin + 100 µg/ml carbenicillin). Control cells (heat shock without plasmids) were included to check for bacterial plasmid contamination.

2.3 Bacterial Cell Expression Trials

For expression trials, 5 mL starter culture was grown under antibiotics and used to inoculate larger cultures (1/100 volume). The following conditions were sequentially tested: cell type

(C43, C43ΔAcrAB, or BL21(DE3)), media (LB or 2xYT), temperature (20°C or 37°C), plasmid induction order (HlyA 30 min before HlyBD, same time, or HlyBD 30 min before HlyA), plasmid inducer concentrations (6.8 arabinose + 1 mM IPTG, or 13.6 mM arabinose + 2 mM IPTG), antibiotic concentration (100 µg/mL carbenicillin + 50 µg/mL kanamycin, 50 carbenicillin + 50 µg/mL kanamycin, or 50 carbenicillin + 30 µg/mL kanamycin), and $CaCl_2$ concentration (0 mM, 2 mM, 5 mM, 8 mM, 10 mM). A final expression time course was conducted with samples taken pre-induction and then every 30 minutes until 3 hours post-induction.

2.4 Antibody Labelling

The protocol was adapted from Heinisch *et al* (2018). Cells were normalised to OD_{600nm} 0.5–0.7 in 1 mL with labelling buffer (PBS, pH 7.4, 1 mM $CaCl_2$). Cells were then pelleted (14000 g for 30 seconds) and washed in 1 mL labelling buffer before incubation with primary α-FLAG antibody (Sigma-Aldrich, Missouri, USA) (1/1000 volume) at room temperature for 30 minutes. Cells were washed in 1 mL labelling buffer before incubation with secondary α-mouse Cy5 antibody (Invitrogen, Massachusetts, USA) (1/1000 volume) for 20 minutes on ice. Cells were washed in labelling buffer before the flow cytometry measurement was taken.

2.5 Flow Cytometry

Flow cytometry measurements were taken using the BD LSR II (BD Biosciences, UK) at 530/30 nm bandpass filter (blue channel, GFP fluorescence) and 660/20 nm bandpass filter (red channel, Cy5 fluorescence). The following controls were used for appropriate gating: unlabelled cells (background fluorescence), HlyA-only expressing cells (GFP fluorescence), and antibody labelled HlyBD + HlyA expressing cells (dual GFP and Cy5 fluorescence). Labelling controls were used to check for non-specific antibody binding: primary antibody labelled only cells, and secondary antibody labelled only cells. For each measurement, 10^5 cells were used and first gated to find the relevant population (*E.coli* cells) and then further gated to remove doublets before dividing them into respective quadrants (Q1, Q2, Q3, or Q4).

2.6 Data Analysis

To account for excess signal in the Q1 quadrant, the following standardisation procedure was performed on the Q2 quadrant:

$$Normalised\ T1SS\ (\%) = \frac{T1SS\ (\%) - mean(excess(\%))}{std(excess(\%))/sqrt(N)} \qquad \text{(equation 2.6-1)}$$

where std is the standard deviation of the excess signal and N is the total number of non-control samples. Histograms of the non-normalised and normalised data were checked to ensure the shape of the data remained unchanged after normalisation. For comparisons between samples and plots, only normalised values greater than zero were used. Raw data can be found in **Table C** in **Appendix C**. Information on the data analysis pipeline I developed can be found in **Table B** in **Appendix B**.

2.7 T1SS Expression for Protein Purification

HlyBD (pK184-HlyBD) and HlyA (pSOI-eGFP-HlyA-FLAGx3-Linker) plasmids were transformed in BL21(DE3) by heat shock. Overnight starter culture was made and inoculated (1/1000) in 500 mL 2xYT media. Cells were grown at 37°C with shaking 200 rpm under antibiotic selection (100 µg/mL carbenicillin + 30 µg/mL kanamycin) until OD_{600} 0.5–0.7 was reached. HlyBD was induced using 1 mM IPTG for 30 minutes. 5 mM $CaCl_2$ was added at this step. HlyA was then induced with 6.8 mM arabinose and cells were left at 37°C with shaking 200 rpm for a further 1.5 hours. Cells were pelleted by centrifugation at 4200 rpm (Beckman JS-4.2, Beckman, California, USA) for 30 minutes at 4°C. Pellets were resuspended in 50 mL lysis buffer (50 mM Tris-HCl, 350 mM NaCl, 50 mM KCl, 5 mM $CaCl_2$, pH 7.5) and flash frozen in liquid nitrogen before storing at -80°C.

2.8 Membrane Solubilisation

50 mL bacterial cell suspension was thawed on ice in 1.5% w/v β-DDM with added lysozyme, DNase I, and protease cocktail inhibitor (Sigma-Aldrich, Germany). The thawed pellet was left

at 4°C for 30 minutes under constant agitation before undergoing five passes in the Emulsiflex C5 (Avestin, Ottawa, Canada) <1000 bar to break the cells. Insoluble material was removed by spinning at 40,000 g for 30 minutes. Supernatant was then filtered using a 0.2 µm pore.

2.9 Glycerol Cushioning

To separate free HlyA-eGFP from T1SS complex, a double glycerol cushion used (adapted from Kirykowicz & Woodward (2020)). A 30% glycerol top cushion (lysis buffer + 0.02% w/v β-DDM) and 70% glycerol bottom cushion (lysis buffer + 0.02% w/v β-DDM) was added below solubilised membrane using a fine needle. The layers were allowed to settle for 10 minutes at 4°C before ultracentrifugation at 30,000 rpm (SW 41 Ti, Beckman Coulter, California, USA) in 4°C for 5 hours. The 30–70% sandwich layer was then extracted for FLAG bead pull-down purification.

2.10 FLAG Pull-Down Purification

2 mL FLAG beads (~1 mL bead volume) was three times washed in 15 mL FLAG buffer (lysis buffer with 0.02% β-DDM) by spinning at 500 g for 5 minutes. Washed beads were incubated with 30– 70% sandwich layer (~20 mL) for 45 minutes at 4°C under constant agitation. Beads were pelleted by centrifugation at 500 g for 5 minutes and washed three times in 15 mL FLAG buffer. Complex was eluted with 100 µL FLAG peptide (4 mg/mL) in 3 mL FLAG buffer for 40 minutes at 4°C under constant agitation. Elution mixture was spun at 500 g for 5 minutes and supernatant extracted.

2.11 T1SS Concentration

T1SS complex was concentrated by ultracentrifugation of FLAG elution supernatant at 40,000 rpm for 3 hours at 4°C (SW 60 Ti, Beckman Coulter, California, USA). Pellet was solubilised overnight at 4°C in buffer (20 mM Tris-HCl, 150 mM NaCl, 0.03% w/v β-DDM, pH 8.0).

2.12 Western Blot

Pre-cast mPAGE® (Sigma-Aldrich, Missouri, USA) or NuPAGE (ThermoFischer Scientific, Massachusetts, USA) Bis-Tris gels (4–12%) were run at 170–180 V constant for 0.75–1.5 hours in MOPS buffer (50 mM Tris, 50 mM MOPS, 0.1% w/v SDS, 1 mM EDTA). Gels were transferred to nitrocellulose membrane in transfer buffer (10 mM Tris, 100 mM glycine, 1% v/v methanol) at 100 V for 1 hour at room temperature. Membranes were then incubated in blocking buffer (1xTBS, 3% w/v milk, 0.1% v/v Triton X-100) for 2–3.5 hours. Primary antibody (1:2500 α-TolC, 1:2000 α-HlyB, 1:2000 α-HlyD, and 1:3000 α-HlyA) was added to membranes in 50 mL blocking buffer overnight at 4°C with constant agitation. Membranes were then washed three times for 15 minutes in 1xTBS before incubation with 1:2000 secondary horseradish peroxidase (HRP) linked α-rabbit antibody (ThermoFischer Scientific, Massachusetts, USA) in 50 mL blocking buffer at room temperature for 1 hour with constant agitation. Membranes were washed three times for 15 minutes in 1xTBS before development. The membranes were developed using a chemiluminescence detection kit (ThermoFischer Scientific, Massachusetts, USA) and visualised in film.

2.13 Grid Preparation and Cryo-CLEM

To render grids hydrophilic, negatively charged grids standard glow-discharge in air was performed on Quantifoil® R2/1 holey-carbon grids (Quantifoil Micro Tools, Jena) for 30 seconds on each side. To render grids hydrophilic and positively charged, 5 µl of 1 M of magnesium chloride was applied to each side of air glow-discharged grids for 2 minutes before blotting off. BL21(DE3) with optimised T1SS expression samples were prepared by diluting cultures to OD_{600} 0.5–0.7 using PBS buffer with 0.1% glucose; 30 µL diluted cells were mixed with 1 µL gold fiducials. Cryo-samples were made in a Vitrobot™ (ThermoFischer Scientific, USA); the humidity was set to 100% at a temperature of 4°C. On each side 3 µL of sample was applied, and a blotting time of 3 seconds was used with a blot force of -10 or -15. Grids were stored in liquid nitrogen after plunge-freezing.

For cryo-CLEM, samples were loaded onto a Leica DM6 FS Microscope fitted with a cryo-stage (Leica Biosystems, Germany) available at the electron Bioimaging Centre (eBIC) at the

Diamond Light Source (Harwell, UK). Green and red fluorescence channels were selected; for the green (GFP) channel, 50 ms exposure was used with 100% intensity, and for the red (Cy5) channel, 500 ms exposure was used with 100% intensity. Settings were checked on a signal control sample (pre-induction BL21(DE3) cells) and no fluorescence in the green or red channels was detected. Images were taken and processed using the THUNDER Imager Cryo-CLEM software (Leica Microsystems, Germany) and exported as tif format files.

3. Results

3.1 Membrane Solubilisation Trials

Detergent membrane solubilisation trials were initially conducted in C43Δ*acrAB* cells as these are optimised for membrane protein production (Miroux & Walker, 1996). Addition of 0.3 M NaCl was also trialled to test for positive effect on production of all components (Marlovits *et al*, 2004). The results are summarised in **Table 3.3.1-1**. Triton X-100 was found to be a very poor detergent for all components. There was little difference between α-DMP and β-DDM in terms of their ability to solubilise the components under different expression conditions. The best detergent concentration was found to be 1.5% β-DDM when all components were expressed at the same time in LB media at 20°C without additional salt.

Table 3.3.1-1. Type I Secretion System membrane solubilisation trials. Expression level was quantified visually via immunoblot.

Component	Temp (°C)	0.3 M NaCl	Media	Expression?	Detergent	Culture Vol (mL)	Soluble?
HlyA-eGFP	37	No	LB	Very Weak	1% Triton X-100	150/3 = 50	No
HlyB	37	No	LB	Very Weak	1% Triton X-100	150/3 = 50	Very Poor
HlyD	37	No	LB	Medium	1% Triton X-100	150/3 = 50	Poor
TolC	37	No	LB	Medium	1% Triton X-100	150/3 = 50	Poor
HlyA-eGFP	37	No	LB	Very Weak	1.5% β-DDM	150/3 = 50	Yes
HlyB	37	No	LB	Very Weak	1.5% β-DDM	150/3 = 50	Yes
HlyD	37	No	LB	Medium	1.5% β-DDM	150/3 = 50	Yes

TolC	37	No	LB	Medium	1.5% β-DDM	150/3 = 50	Yes
HlyD	37	No	LB	Medium	1% α-DMP	150/3 = 50	Yes
TolC	37	No	LB	Very Strong	1% α-DMP	150/3 = 50	Yes
HlyA-eGFP	37	Yes	LB	Weak	1.5% β-DDM	150	Yes
HlyB	37	Yes	LB	Strong	1.5% β-DDM	150	Yes – lots insoluble
HlyD	37	Yes	LB	Medium	1.5% β-DDM	150	Yes – tiny amount in pellet
TolC	37	Yes	LB	Very Strong	1.5% β-DDM	150	Yes
HlyA-eGFP	30	No	LB	Weak	1.5% β-DDM	50	Yes
HlyB	30	No	LB	Very Weak	1.5% β-DDM	50	Yes
HlyD	30	No	LB	Very Strong	1.5% β-DDM	50	Yes – small amount in pellet
TolC	30	No	LB	Strong	1.5% β-DDM	50	Yes – lots in pellet
HlyA-eGFP	30	No	2YT	Medium	1.5% β-DDM	50	Yes
HlyB	30	No	2YT	Very Weak	1.5% β-DDM	50	Yes – mostly insoluble
HlyD	30	No	2YT	Very Strong	1.5% β-DDM	50	Yes
TolC	30	No	2YT	Strong	1.5% β-DDM	50	Yes

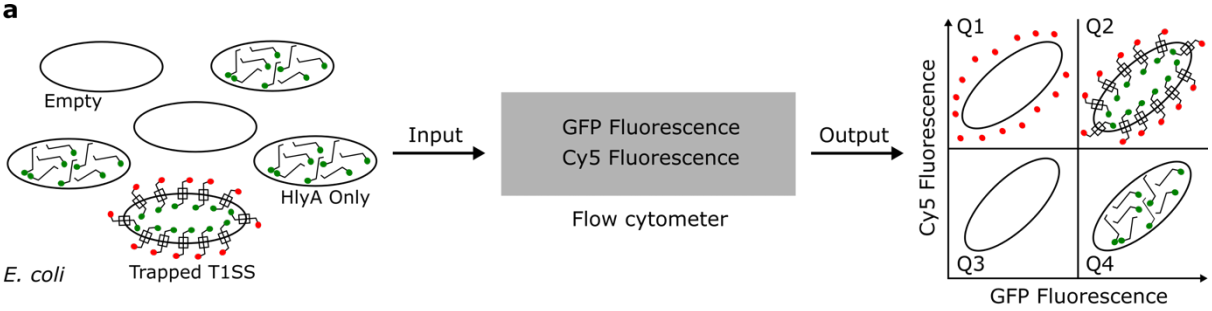
HlyA-eGFP	25	No	LB	Very Strong	1.5% β-DDM	50	Yes
HlyB	25	No	LB	Very Weak	1.5% β-DDM	50	Yes
HlyD	25	No	LB	Strong	1.5% β-DDM	50	Yes
TolC	25	No	LB	Strong	1.5% β-DDM	50	Yes
HlyA-eGFP	25	No	2YT	Very Strong	1.5% β-DDM	50	Yes
HlyB	25	No	2YT	Very Weak	1.5% β-DDM	50	Yes
HlyD	25	No	2YT	Very Strong	1.5% β-DDM	50	Yes
TolC	25	No	2YT	Strong	1.5% β-DDM	50	Yes
HlyA-eGFP	20	No	LB	Very Strong	1.5% β-DDM	150	Yes – not in pellet
HlyB	20	No	LB	Strong	1.5% β-DDM	150	Yes – not in pellet
HlyD	20	No	LB	Strong	1.5% β-DDM	150	Yes – not in pellet
TolC	20	No	LB	Strong	1.5% β-DDM	150	Yes – not in pellet
HlyA-eGFP	20	Yes	LB	Strong	0.5% β-DDM	150/3 = 50	Yes
HlyB	20	Yes	LB	Medium	0.5% β-DDM	150/3 = 50	Yes
HlyD	20	Yes	LB	Extremely weak	0.5% β-DDM	150/3 = 50	Yes
TolC	20	Yes	LB	Strong	0.5% β-DDM	150/3 = 50	Yes
HlyA-eGFP	20	Yes	LB	Strong	1% β-DDM	150/3 = 50	Yes

HlyB	20	Yes	LB	Medium	1% β-DDM	150/3 = 50	Yes
HlyD	20	Yes	LB	Extremely weak	1% β-DDM	150/3 = 50	Yes
TolC	20	Yes	LB	Strong	1% β-DDM	150/3 = 50	Yes
HlyA-eGFP	20	Yes	LB	Strong	2% β-DDM	150/3 = 50	Yes
HlyB	20	Yes	LB	Medium	2% β-DDM	150/3 = 50	Yes
HlyD	20	Yes	LB	Extremely weak	2% β-DDM	150/3 = 50	Yes
TolC	20	Yes	LB	Strong	2% β-DDM	150/3 = 50	Yes

3.2 Expression Optimisation via Flow Cytometry

Although promising initial results were produced from the detergent optimisation trials, the method has the major drawback that expression is confounded with solubilisation. In addition, Western Blotting was used to confirm presence of components, which is very time-consuming and hence fewer trials can be conducted. To fully optimise expression using a more high-throughput method without adding the effect of solubilisation, flow cytometry was used. By tagging the exposed FLAG on exported HlyA with Cy5, I could examine dual GFP and Cy5 fluorescence as an indication of the production of full trapped T1SS (**Figure 3.3.2-1**).

a



b

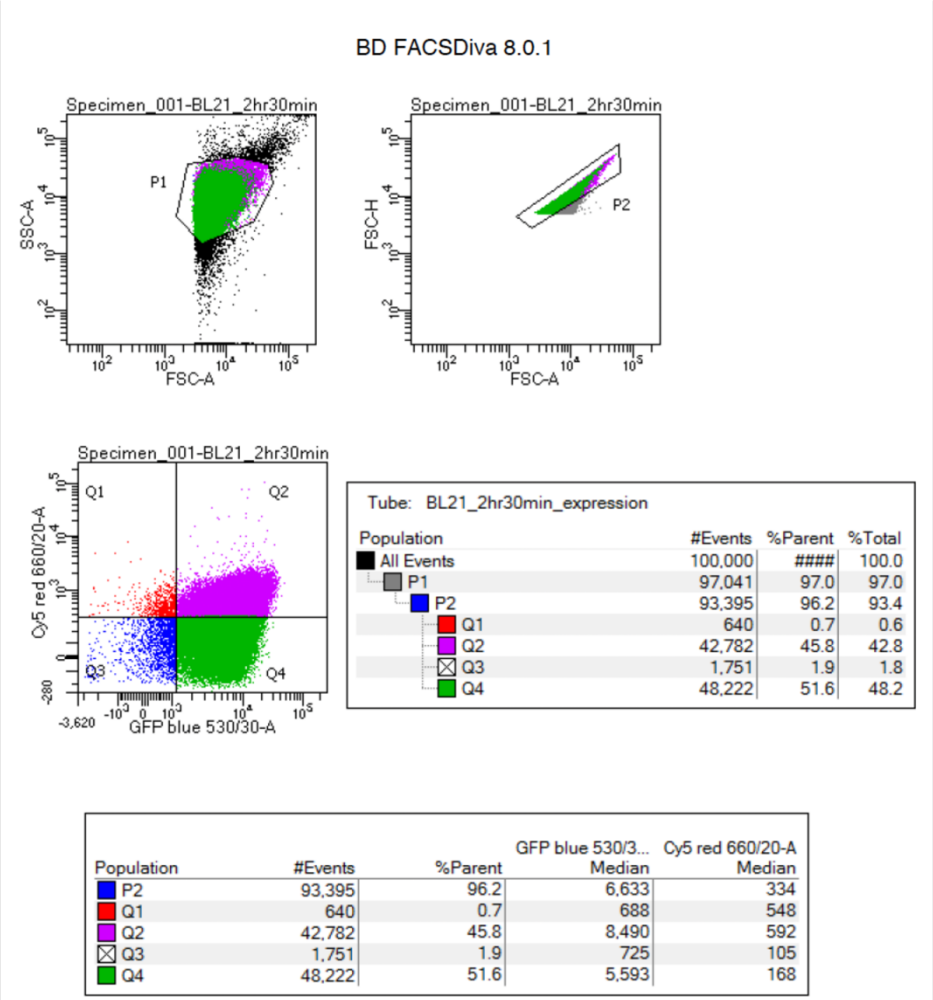


Figure 3.3.2-1. Schematic of expression optimisation via flow cytometry. **a)** *Escherichia coli* containing a mix of non-expressing cells (empty), HlyA-eGFP only expressing cells, and trapped type I secretion system (T1SS) were assessed in the flow cytometer for GFP and Cy5 fluorescence. Cells could be separated as either: low GFP, low Cy5 fluorescence (Q3, empty); low Cy5, high GFP fluorescence (Q4, HlyA only); high GFP, high Cy5 fluorescence (Q2, trapped T1SS); and low GFP, high Cy5 fluorescence (Q4, empty with excess Cy5 antibody). **b)** Example flow cytometry result. The *E. coli* population is first found by plotting side scatter area (SSC-A) vs forward scatter area (FSC-A) and gating appropriately. After finding these events, doublets are removed with another gate on the forward scatter height (FSC-H) vs forward scatter area (FSC-A) plot; in this case, single cells will fall along a linear

path of this plot. Final gates on the GFP vs Cy5 fluorescence are then completed, taking into consideration controls. Percentages are then derived based on number of events from the parent (sorted cells).

The effects of cell type, media, temperature, expression timing, induction concentration, antibiotic concentration, and length of induction were assessed. The noise signal in the Q1 region was large, and a significant optimisation of the antibody labelling protocol was required to minimise the noise. To compare results across trials, a normalisation/standardisation procedure was performed on the Q2 signal using the mean and variance of the Q1 signal, excluding the controls. The results are summarised in **Table C** (see **Appendix C**). Major determinants of T1SS expression levels were investigated. Media and temperature had the biggest effect on T1SS expression, acting as an on/off switch. Expression timing, induction concentration, antibiotic concentration, and length of induction acted as modulators, pushing the cells into either low, medium, or high production levels depending on their combinations.

It was found that BL21(DE3) cells produced the most trapped T1SS in 2xYT media at 37^{o}C under single induction concentrations. The next major effects on T1SS production were found to be expression order, $CaCl_2$ concentration, and total expression time. Expression order had a large effect. Expressing HlyA 30 minutes before HlyB/D resulted in no T1SS production (see **Table C** in **Appendix C**). In contrast, if HlyA and HlyB/D were expressed at the same time, the cells were pushed into a low-expression mode. However, if HlyB/D was expressed 30 minutes before HlyA, the cells fall into medium and high modes of expression (**Figure 3.3.2-3a**). The total expression time followed an interesting pattern: at 1 hour total expression time, cells were at a low-to-medium mode of expression which was rapidly pushed to a high-expression mode after 1.5 hours (**Figure 3.3.2-3b**). This was maintained until 3 hours post-induction, where levels started to fall to medium-to-low modes (**Figure 3.3.2-3b**). There are two explanations for this pattern: either the cells stopped producing T1SS after 1.5–2 hours, and the remaining levels decreased as the cells divided (a “dilution effect”), or the cells continued to produce and turnover T1SS but at a lower level. Previous experimental evidence would suggest that the former explanation is more likely (Lenders *et al*, 2015).

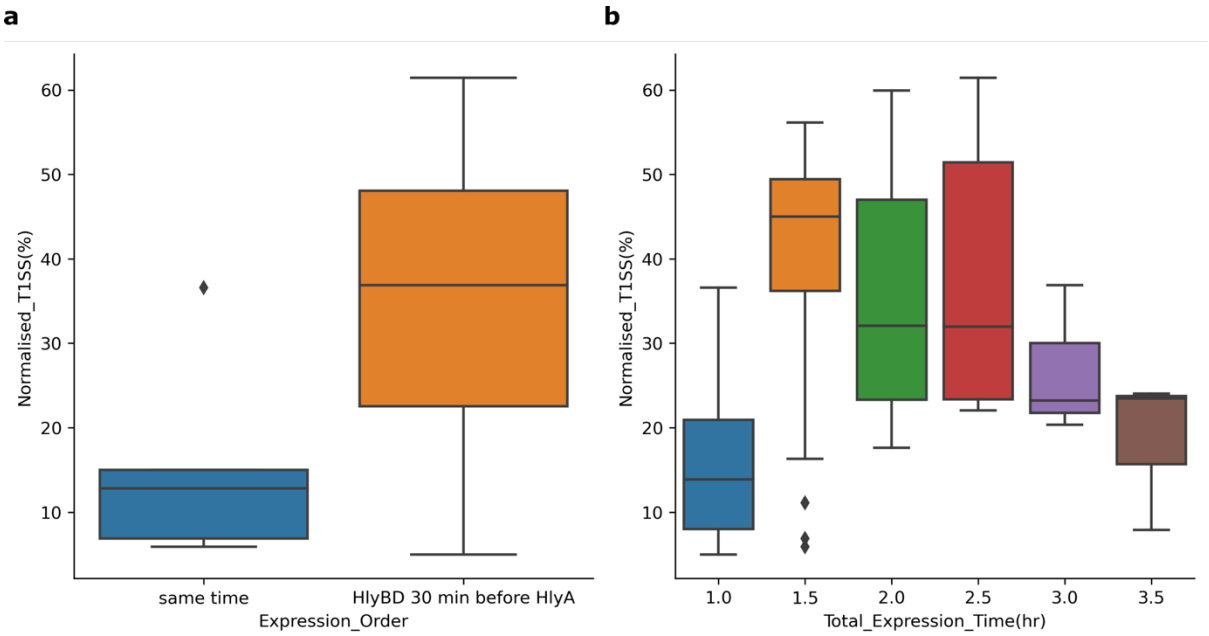


Figure 3.3.2-3. Effect of expression order and time on T1SS production in BL21(DE3) cells. a) Effect of expression order on T1SS production. Note that HlyA 30 min before HlyBD was trialled and the amounts of T1SS produced was found to be less than the noise level (see **Table C** in **Appendix C**). **b)** Effect of total expression time on T1SS production. Production increases after 1 hour and stabilises after 1.5–2.5 hours before decreasing after 3 hours. The colours are for illustrative purposes only. The data was grouped based on measured outcome (*i.e* expression order or total expression time): see the Flow Cytometry Data Analysis Notebook available on GitHub (https://github.com/AMKCam) (see Table B in Appendix B for details and Appendix C for the raw data on the experimental trials). Note that Normalised T1SS (%) refers to a signal standardisation done to take into account excess signal in the Q2 quadrant during flow cytometry measurements (see section 2.6 in Material & Methods).

Antibiotic concentration also had an effect, but this was confounded with $CaCl_2$ concentration and total expression time (**Figure 3.3.2-4**). In general, antibiotic concentration of 100 µg/mL carbenicillin with 30 µg/mL kanamycin performed the best under different $CaCl_2$ concentrations. Interestingly, $CaCl_2$ concentration had a limited effect on T1SS production, with similar levels being produced after 1.5 hours of expression time under the optimal antibiotic concentration (**Figure 3.3.2-4**). Thus, it would seem that antibiotic and $CaCl_2$ concentration had limited ability to modulate T1SS production compared to expression order and total expression time.

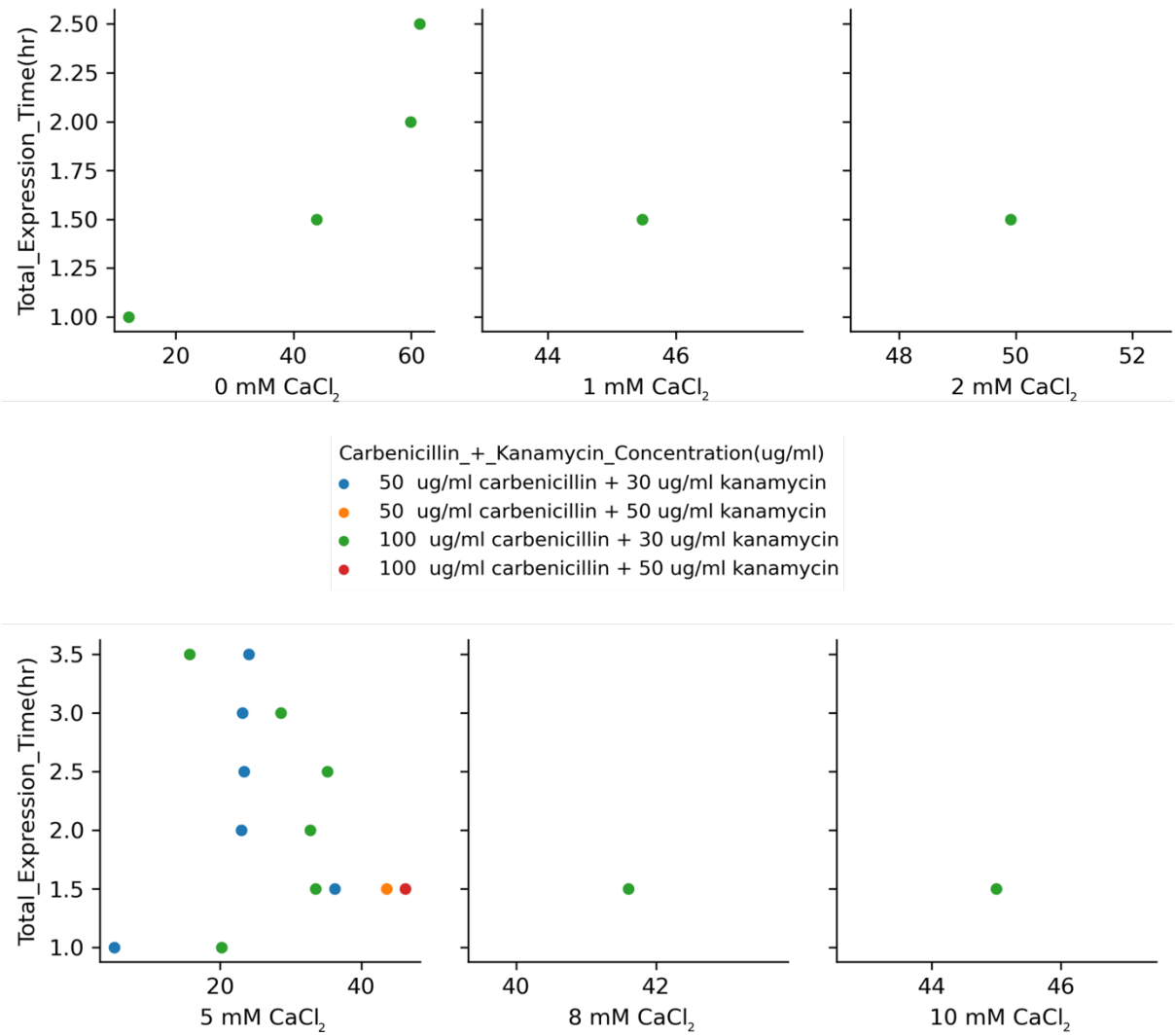


Figure 3.3.2-4. Effect of $CaCl_2$ and antibiotic concentration on T1SS production as a function of total expression time. Mean Normalised T1SS (%) (x-axis) was plotted against the total expression time (hr) (y-axis) at different $CaCl_2$ concentrations, coloured by antibiotic concentration. Note that Normalised T1SS (%) refers to a signal standardisation done to take into account excess signal in the Q2 quadrant during flow cytometry measurements (see section 2.6 in Material & Methods).

3.3 Cardiolipin Deficiency Decreases T1SS Production

After optimising production of the trapped T1SS, I wished to test the hypothesis generated by the molecular dynamics simulation results (see **Chapter II**) that the presence of cardiolipin aids substrate translocation. I utilised a cardiolipin deficient strain (MG1655Δ*cls*) and compared T1SS production levels to its parent strain using the flow cytometry protocol (**Figure 3.3.2-1**). For an initial condition search, I used the results from the BL21(DE3)

optimisation. As found for BL21(DE3) cells, media and temperature had the biggest effect on T1SS production, with cells being switched "on" when using LB media at 37°C. In MG1655, expression order also appeared to act as an on/off switch rather than a modulator, with no T1SS produced expect when all components were expressed at the same time (see **Table C** in **Appendix C**). Comparing strains found that cardiolipin deficient MG1655 is trapped in a low-to-medium expression mode compared to the parent strain (**Figure 3.3.3-1**).

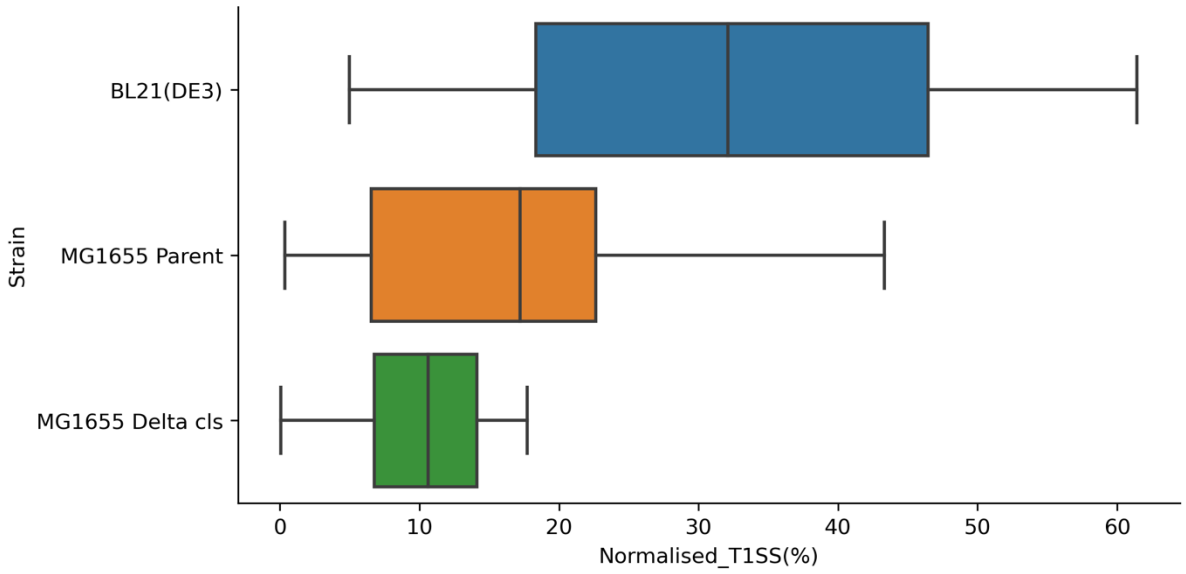


Figure 3.3.3-1. T1SS production by strain. Under different expression conditions, BL21(DE3) cells produce ~32% trapped T1SS on average, with the level dropping to ~17% in the MG1655 parent strain. Cardiolipin deficient MG1655 (MG1655 Delta *cls*) produces ~10% trapped T1SS on average, with cells appearing trapped in a low-to-medium expression mode. Note that Normalised T1SS (%) refers to a signal standardisation done to take into account excess signal in the Q2 quadrant during flow cytometry measurements (see section 2.6 in Material & Methods).

Examining the effect of $CaCl_2$ concentration and total expression time in the MG1655 strain found a similar pattern to BL21(DE3) cells (**Figure 3.3.3-2**). In the parent MG1655 strain, cells stopped producing T1SS after 1.5 hours, while for cardiolipin deficient MG1655, there was no pattern to T1SS levels. Like BL21(DE3) cells, increasing $CaCl_2$ concentration seemed to have a limited effect on the level of T1SS for the MG1655 strain. Across all $CaCl_2$ concentration and expression times, cardiolipin deficient MG1655 remained at around half the level of its parent strain (**Figure 3.3.3-2**). Thus, it would seem that for cardiolipin deficient MG1655, modulation

is limited with significant stochasticity observed, generally fitting with the observation that the cells remain “trapped” in a low-production state irrespective of expression conditions.

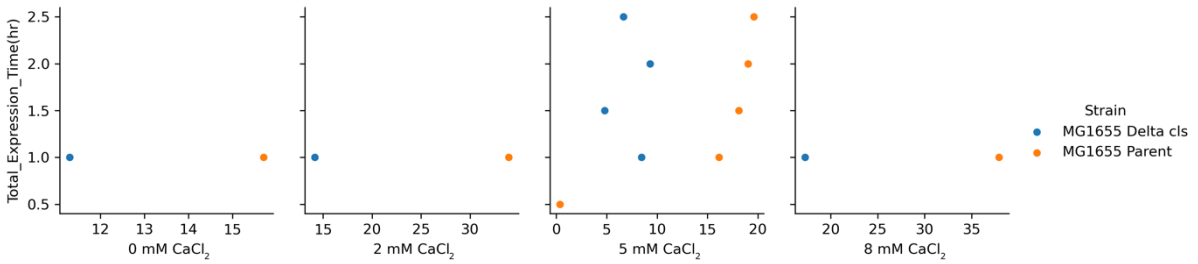


Figure 3.3.3-2. Effect of $CaCl_2$ concentration and expression time on T1SS production. Mean Normalised T1SS (%) (x-axis) was plotted against the total expression time (hr) (y-axis) at different $CaCl_2$ concentrations, coloured by MG1655 strain. Note that Normalised T1SS (%) refers to a signal standardisation done to take into account excess signal in the Q2 quadrant during flow cytometry measurements (see section 2.6 in Material & Methods).

3.4 Cryo-Correlative Light Electron Microscopy

BL21(DE3) cells expressed under the optimum conditions found by flow cytometry were examined under cryo-correlative light electron microscopy (cryo-CLEM). Two different grid conditions were tested: 1) negative charge holey carbon grids, and 2) positive charge holey carbon grids. No noticeable difference was found between the two treatments (**Figure 3.3.4-1**). To test the use of cryo-CLEM for cryo-electron tomography data collection of whole cells, squares with optimal fluorescence in the green and red channels were selected (**Figure 3.3.4-2a**). From these optimal squares, T1SS producing cells which also lie over an open hole were found (**Figure 3.3.4-2b**). In addition, it was observed that the cryo-CLEM agreed with the flow cytometry results in that around 30–50% of the cells in any imaged square were producing trapped T1SS (**Figure 3.3.4-2c**).

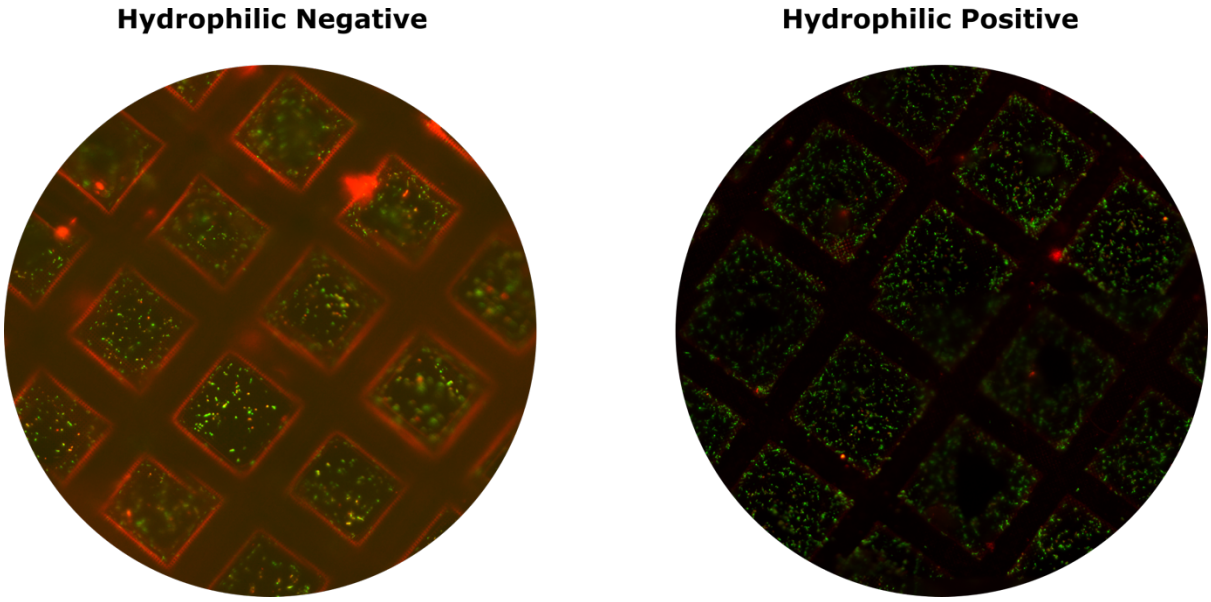


Figure 3.3.4-1. Effect of grid treatments on cryo-CLEM of BL21(DE3) expressing T1SS. Two different grid treatments were trialled: one with normal glow-discharge (hydrophilic negative) and one with magnesium acetate treatment after glow-discharge (hydrophilic positive). No effect of grid treatment was observed on numbers or distributed of BL21(DE3) cells on the grids.

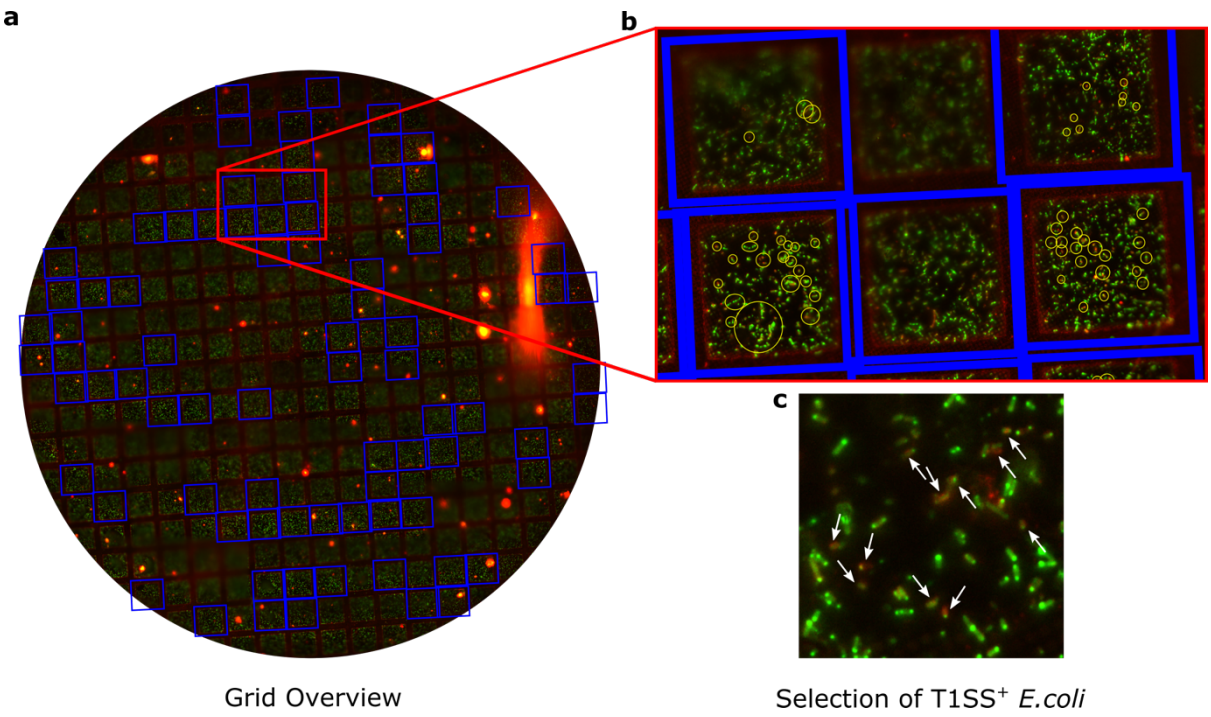
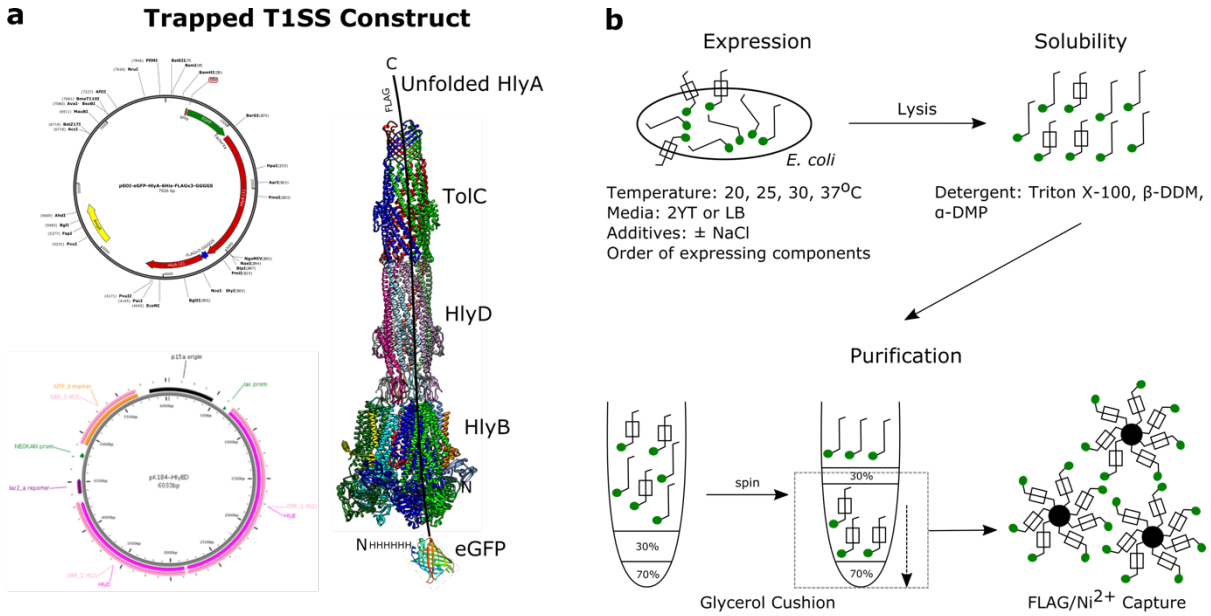


Figure 3.3.4-2. Cryo-CLEM Overview. **a)** Cryo-CLEM was performed on samples of frozen-hydrated BL21(DE3) cells. Squares with a good fluorescence signal are highlighted (blue). **b)** Zoom in view of squares, showing the distribution of HlyA-eGFP producing cells and those producing T1SS (yellow circle) that lie over a hole. **c)** Another

Zoom in view, showing that cell producing T1SS constitute ~30–50% of the cell population, on average (white arrows). An example of EM images taken can be seen in **Chapter IV Figure 4.3.3-1**.

3.5 T1SS Protein Purification

Protein purification tests were conducted to attempt to isolate T1SS for single-particle cryo-EM (**Figure 3.3.5-1**). The tests were done in parallel to expression and detergent optimisation. Purification utilised the FLAG or His-tag present in the trapped complex to isolate the complex from the cell protein mixture (**Figure 3.3.5-1a**). First, I attempted to isolate a membrane fraction using standard protocols. However, in my hands the T1SS complex dissociated in the membrane pellet solubilisation step. Next, I tried chemical cross-linking of the complex using glutaraldehyde on whole *E. coli*; however, that led to failure to solubilise the complex after membrane preparation. To attempt to solubilise the complex without forming a membrane pellet, I directly lysed cells in 1.5% βDDM (**Figure 3.3.5-1b**). This seemed to work, in that all components were present in the soluble fraction on a Western Blot (**Figure 3.3.5-1c**). I also checked for leaky plasmid expression: as expected, TolC is present in all uninduced fractions as it is endogenous; there was a faint band for uninduced HlyA, and an expected band for uninduced HlyB/D; there was also a soluble band for uninduced HlyB and HlyD.


a
Trapped T1SS Construct
C
Unfolded HlyA
FLAG
TolC
HlyD
HlyB
N
N HHHHHH
eGFP
b
Expression
Solubility
Lysis
E. coli
Temperature: 20, 25, 30, 37°C
Media: 2YT or LB
Additives: ± NaCl
Order of expressing components
Detergent: Triton X-100, β-DDM, α-DMP
Purification
spin
30%
70%
Glycerol Cushion
FLAG/Ni2+ Capture

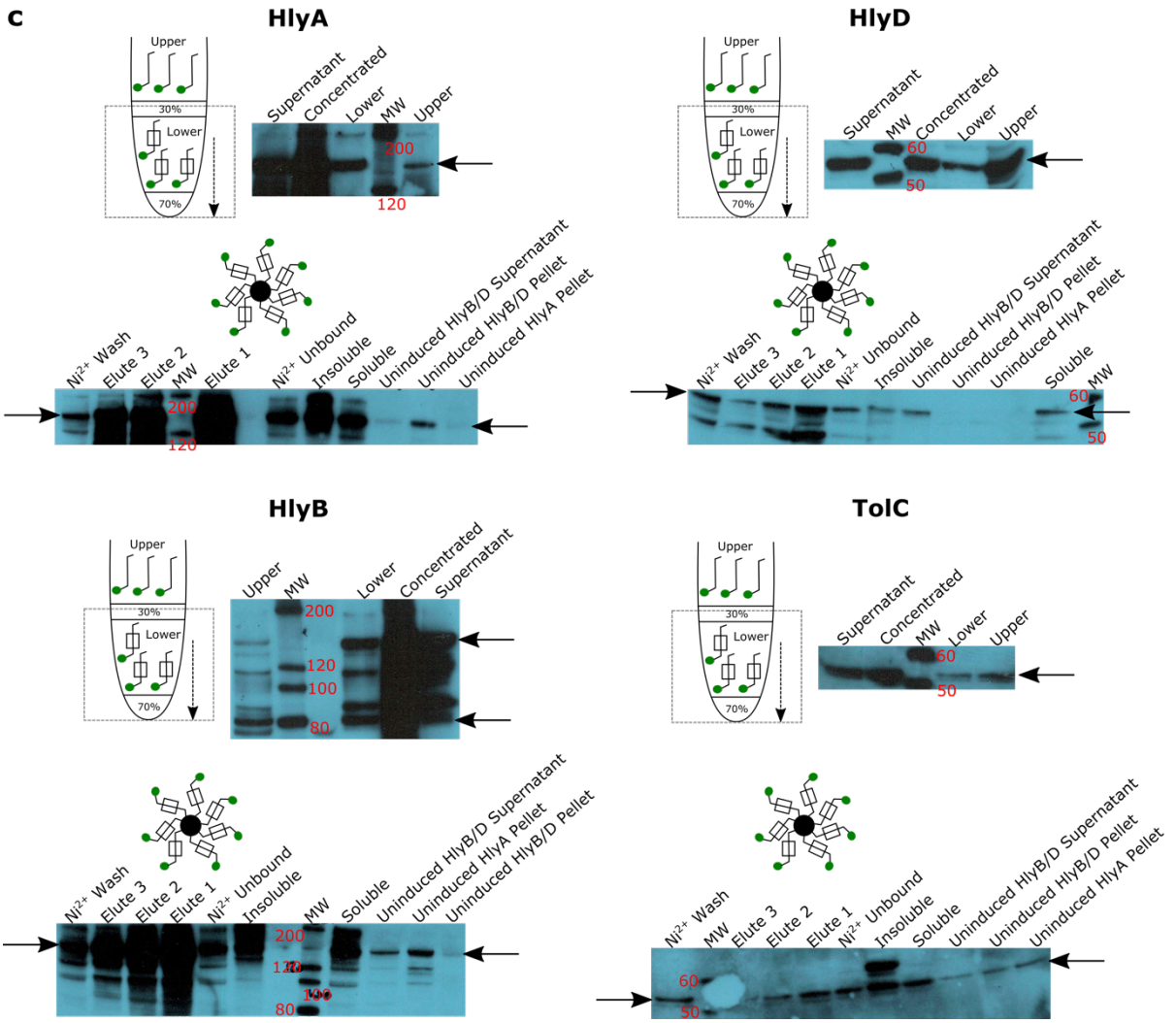

c
HlyA
Upper
30%
Lower
70%
Supernatant
Concentrated
Lower
MW
Upper
200
120
Ni2+ Wash
Elute 3
Elute 2
Elute 1
Ni2+ Unbound
Insoluble
Soluble
Uninduced HlyB/D Supernatant
Uninduced HlyB/D Pellet
Uninduced HlyA Pellet
HlyD
60
50
HlyB
100
80
TolC

Figure 3.3.5-1 (previous page). T1SS Purification Tests. **a)** Schematic of the trapped T1SS used for flow cytometry and purification. N-terminal His-tag and C-terminal FLAG tag can be used for pull-downs. **b)** After optimising for expression conditions and detergent, purifications were trialled, which included FLAG or Ni^{2+} pull-downs and the use of a glycerol cushion to remove excess HlyA-eGFP from trapped T1SS. **c)** Example Western Blots from purification trials. Trials of glycerol cushioning and Ni^{2+} affinity pull-downs were successful in capturing all components. For the glycerol cushion step, all components were found in the soluble supernatant that was loaded onto the column. Components are then separated into lower (between 30 and 70% glycerol) and upper (above 30% glycerol) halves. The lower half was further bound to FLAG beads and concentrated. The Ni^{2+} shows purification without glycerol cushioning. Note the presence of excess HlyA in the soluble fraction. Note that HlyB Western Blot forms two bands: a 79 kDa monomeric band and a ~160 kDa dimeric band.

HlyB/D are on the same plasmid, controlled by a *lac* promoter. Plasmids utilising the *lac* promoter are known to be leaky so this is not unexpected. The *ara* promoter which controls HlyA production is more tightly controlled, but there are sometimes still background levels (Siegele & Hu, 1997). For Ni^{2+} purification after solubilisation, there was a large amount of all components in the unbound fraction; this could be traced to an excess amount of HlyA (**Figure 3.3.5-1c**). Encouragingly, there was substantial amounts of elution for all components. To attempt to remove excess HlyA and improve T1SS yields, a glycerol cushion was trialled before FLAG/Ni^{2+} affinity purification. The results seemed promising, although there were low yields (< 1 mg/mL) for the final Ni^{2+} elution after the glycerol cushioning step (**Figure 3.3.5-2**).

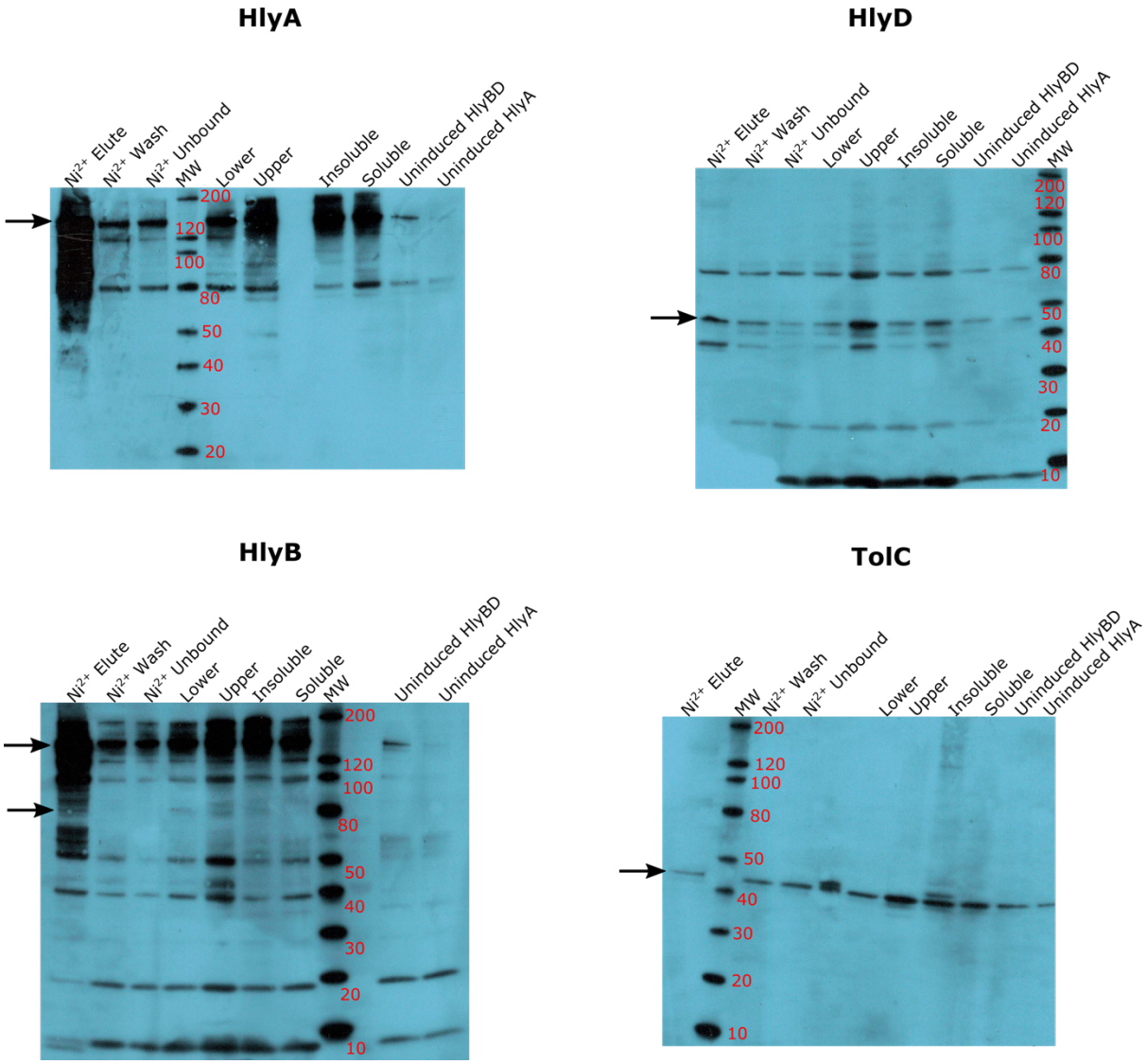


Figure 3.3.5-2. Glycerol cushioning and Ni^{2+} affinity purification Western Blots. A 40–70% glycerol cushion was trialled to attempt to separate excess HlyA-eGFP from T1SS complex. HlyB and HlyD showed leaky expression, while small amounts of HlyA were also detected in the uninduced fraction. All components could be found in the soluble fraction, with varying levels of insolubility. As expected, an excess of HlyA was found in the upper (>40%) glycerol fraction, while all components were found in the lower (between 40 and 70% glycerol) fraction. Small amounts of each component were still detected in the unbound fraction, but this was a big improvement over Ni^{2+} affinity purification without glycerol cushioning (see **Figure 3.3.5-1c**). All components were present in the elution fraction. Note that HlyB Western Blot forms two bands: a 79 kDa monomeric band and a ~160 kDa dimeric band.

Attempts at making cryo-EM grids from this preparation did not yield any encouraging results. There was uncertainty if this was the result of the low yield or due to complex dissociation during plunge-freezing. I attempted the glycerol cushion with a FLAG purification,

concentrating the final elution using ultracentrifugation and resuspension in a smaller buffer volume (**Figure 3.3.5-1c**). This resulted in improved concentration (>1.5 mg/mL); however no complex could be found on cryo-EM grids prepared on this sample. Since concentration is unlikely to be an issue, and Western Blot confirmed the presence of all components (**Figure 3.3.5-1c**), there is a likelihood that the complex dissociates during plunge-freezing. I looked for complex components in the EM grids, but could not find convincing particles even after automated particle picking and classification in Warp using on-the-fly data collection. After these results, purification trials were halted to focus on cryo-ET of whole/ghost cells (see **Chapter IV**). This was to bypass issues with purification and complex dissociation and focus on obtaining an *in situ* structure.

4. Discussion

4.1 Membrane Solubilisation & Protein Purification

Determining the correct solubilisation and purification procedure for membrane proteins relies on a trial-and-error approach, with pipelines adapted for the protein under consideration. Due to time and resource constraints, it is typical for a detergent to be selected which is found to satisfy most requirements for solubilisation and purification (Ratkeviciute, Cooper & Knowles, 2021). I found that the T1SS can be solubilised in β-DDM, with Triton X-100 acting as a very poor detergent for all components. Effects of detergent on solubility were evaluated by Western Blot, which worked better than simple SDS-PAGE as there is less ambiguity for the presence of each component, considering the heterogenous nature of the supernatant and pellet samples. One downside with the Western Blots is that the antibodies used are polyclonal, and hence there is some uncertainty if the off-band targets which appear are degradation products or other proteins that are not the protein of interest. However, the antibodies are more sensitive and can detect proteins at low concentration. Antibody quality can affect the reproducibility of experiments, hence the call for more robust pipelines to produce commercial antibodies (Bradbury & Plückthun, 2015). As pointed out by others (*e.g.,* Freedman, 2015; Polakiewicz, 2015) the key to antibody reproducibility is validation. The antibodies I used are routinely applied in the Lutz group for different experiments (*e.g.,* Lecher *et al*, 2012; Lenders *et al*, 2015, Lenders *et al*, 2016), and have been found to be quite reliable. An additional consideration is that membrane protein oligomers can run anomalously fast or slow due to the binding of SDS; in some cases, the difference can be up to 50% off target molecular weight (MW) (Rath *et al*, 2009). This effect could also produce multiple bands if there are species of oligomers present in the sample. I found a noticeable effect of SDS-PAGE type used on band migration, with significant improvements in generating monomeric HlyB (rather than dimeric HlyB) when mPAGE system was used (see **Figures 3.3.5-1** and **3.3.5-2**). Anomalous SDS-PAGE band migration for HlyB has been noted before (*e.g.,* Lecher *et al*, 2012; Lenders *et al*, 2015, Lenders *et al*, 2016), and may be due to multiple oligomeric species of HlyB present in the sample (Rath *et al*, 2009). Furthermore, multiple bands were found for HlyA-eGFP and HlyD, but not TolC (see **Figures 3.3.5-1** and **3.3.5-2**). This

may be due to multiple oligomers forming as the complex dissociates during denaturation, with TolC forming the weakest interactions in the complex (see **Figure 3.3.5-1a**) and hence full dissociation is more likely. Thus, considering the factors that may cause the presence of multiple bands it is most probable that the observed bands are species of oligomers caused by partial complex dissociation rather than an issue with the polyclonal antibodies.

For the protein purification trials, which were run in parallel with expression and detergent optimisation trials, I found that excess HlyA appeared to interfere with binding of the T1SS to Ni^{2+} or FLAG beads. This was due to its presence in the supernatant, as trials to obtain solubilised T1SS from membrane pellet were unsuccessful. To attempt to remove excess HlyA-eGFP from trapped T1SS, I used glycerol cushioning. This was adapted from a previous sucrose cushioning method I tried for isolating protein complexes (Kirykowicz & Woodward, 2020). One issue with sucrose cushioning is that the isotonicity of the solution changes as the proteins migrate through, leading to precipitation. This means using a low salt starting buffer before cushioning to limit these effects. However, this is not feasible for membrane proteins which require high salt in the starting buffer in order to extract them effectively from the membrane. Switching from sucrose to glycerol can limit isotonicity effects while maintaining the use of high salt starting buffer. The use of a double cushion should separate out excess HlyA-eGFP from T1SS, as the much lower molecular weight (MW) of HlyA-eGFP (~150 kDa) compared to T1SS (>1 MDa) means that it cannot migrate beyond the lower glycerol concentration during ultracentrifugation. Western Blots seemed to indicate that it was generally successful in reducing the levels of HlyA-eGFP and boosting levels of HlyB, HlyD, and TolC (see **Figures 3.3.5-1 and 3.3.5-2**). Cryo-EM samples were made throughout different purification trials, but no T1SS complex was observed. Initially, I thought this was due to low sample concentration, however after concentrating the complex and checking for component presence by Western Blot I still did not observe any T1SS complexes in cryo-EM grid samples. Hence, I concluded that it was possible that the complex is dissociating during plunge freezing. Grid preparation for cryo-EM depends on many factors, including protein concentration, detergent type and concentration, salt concentration, and use of CHAPSO to prevent orientation bias in membrane proteins (Kampjut, Steiner & Sazanov, 2021). In general, protein concentration >1 mg/mL should yield sufficient particles on the grid although numbers will depend on the size of the complex and ice-thickness (Passmore & Russo, 2016). Due to these

factors, I decided to focus on optimising cryo-ET of T1SS samples as this method does not require pure protein. Initial methods for using cryo-CLEM in cryo-ET data collection of whole *E. coli* is discussed in the next section.

4.2 Locating T1SS$^+$ *E. coli* by Cryo-CLEM

I tested cryo-CLEM on BL21(DE3) cells expressing T1SS. As expected from the flow cytometry data, there is a mix population of cells either expressing HlyA-eGFP only or expressing HlyA-eGFP with HlyB/D to form a T1SS channel with endogenous TolC. I tried two grid treatments to test if cell adherence to the grid changed: one with a standard hydrophilic negative charge under glow-discharge conditions, or one with a hydrophilic positive charge following magnesium acetate treatment after glow discharge. I did not find any differences between number or types of cells on the grid after these treatments. Next, I selected grids which had a good distribution of T1SS$^+$ BL21(DE3) and further selected individual cells which were over open holes (see **Figure 3.3.4-2**). This was to image these cells in the TEM for cryo-ET data collection of whole cells. Results of these experiments can be found in **Chapter IV**. Cryo-ET of whole bacteria has been conducted over many years, with the main challenge being the thickness of the cells. Bacterial species with widths ~350 nm (*e.g. Caulobacter crescentus*) allows for visualisation of most internal cellular features, while in species with larger widths ~1000 nm (*e.g. E. coli*) only large cellular components with high-local contrast can be visualised. However, in these larger cells it is possible to visualise membrane protein complexes (*e.g.* chemotaxis arrays, flagellum rotor complex) (Milne & Subramaniam, 2009). Thus, I showed it is possible to isolate T1SS$^+$ cells for tomographic data collection of whole cells.

4.3 Factors Which Influence T1SS Expression

I used the sequential design strategy (Lu & Anderson-Cook, 2021) in order to optimise T1SS production. The strategy works by starting with a prior hypothesis on what to test; this is then updated with each iteration until N observations (Bradt, Johnson & Karlin, 1956). The main benefit is that not all iterations are tested, and thus time and resources are saved. This method was very successful in quickly finding the greatest determinants on T1SS production.

One limitation is that the measured variance for each effect is not equal: by design, more tests are done near the optimum conditions than conditions which are not optimal. Thus, uncertainty is greater for the least optimal conditions. Practically, this may not have appreciable consequence, as the goal is to maximise optimal conditions under a specific set of hypotheses. The solution should converge to the correct set of hypotheses as N is increased (Bradt, Johnson & Karlin, 1956).

I found that the greatest impact on T1SS production was bacterial cell strain, followed by media type and temperature. These seemed to act as primary "on/off" switches. For example, C43 cells did not express T1SS (above detection noise levels) no matter what media type or temperature was used. After these effects, other factors such as order of plasmid induction, length of induction, and antibiotic concentration seemed to act like modulators, tuning levels to what the cell will tolerate. Cells could be divided at the population level into "low", "medium", or "high" expression modes based on these factors. Here, clear differences between strains emerge which can express T1SS above noise levels. BL21(DE3) cells were able to produce higher amounts of T1SS compared to MG1655 cells.

One obvious influencing factor is membrane composition. During cell preparation for flow cytometry, I noticed that the MG1655 cells formed softer pellets which were easier to disturb than the BL21(DE3) cells. MG1655 cells are known to alter their phospholipid composition during their cell cycle; intriguingly it is thought that cardiolipin is primarily localised to the cell poles (Furse *et al*, 2015). This may explain the reduced levels of T1SS in the MG1655 parent strains compared to the BL21(DE3) strain; if cardiolipin is required for transport (see **Chapter II**), then HlyA-eGFP can only be transported from the cell poles rather than the entire cell surface. It is also known that membrane protein production can overwhelm the Sec translocon machinery which is responsible for membrane protein insertion (Schlegel *et al*, 2014). This is why tapering down membrane protein production in overexpressed plasmids may help in boosting overall production (*e.g.,* Zhang *et al*, 2015). Thus, it is possible that the production of HlyB/D creates a cellular stress response, possibly as a result of an overloaded Sec translocon, and hence leading to the triggering of bacterial responses which inhibit plasmid expression. This would explain the "dilution" effect observed after 1.5 hours of total induction time, during which the cells appear to stop T1SS production and any remaining T1SS

is subsequently divided between daughter cells (see **Figure 3.3.2-3b**). This would also explain why the order of plasmid induction seems to have a large influence on T1SS levels. Induction of HlyA 30 minutes before HlyB/D in the BL21(DE3) strain induces a catastrophic failure to produce any T1SS component (not including endogenous TolC), including complete abolition of HlyA-eGFP expressing cells. However, if plasmids are expressed at the same time there is T1SS produced, but the levels remain trapped in low-to-medium expression modes. Indeed, only when HlyB/D is expressed 30 minutes before HlyA-eGFP could cells access the high expression range (see **Figure 3.3.2-3a**). This pattern was not observed for the MG1655 strain, which only tolerated expression of the plasmids at the same time. Thus, MG1655 is likely more sensitive to the presence of HlyB/D in its membrane. The cellular stress response may be caused by ATP-consumption; ATP is required to thread HlyA through the channel, and formation of a trapped complex may induce futile ATP-hydrolysis cycles that places a heavy energy burden on the cell. Indeed, estimates for amounts of ATP hydrolysed per exported HlyA is quite high (see **Chapter VI**).

I found that $CaCl_2$ concentration did not seem to influence T1SS production in either BL21(DE3) cells or MG1655 cells. This is unexpected as $CaCl_2$ is known to aid in the folding of HlyA upon cell exit (*e.g.,* Bumba *et al*, 2016). Indeed, it is also posited to play a role in transport energetics by providing a "Brownian ratchet" and preventing backsliding of HlyA down the channel (Lenders *et al*, 2015; Bumba *et al*, 2016; Lenders *et al*, 2016). However, there is some evidence that $CaCl_2$ may not play a role in T1SS transport energetics (Lenders *et al*, 2016). One factor may be the concentrations of $CaCl_2$ used or the stage of secretion; an effect may only be observed for <1mM $CaCl_2$ concentrations at an early stage of HlyA secretion. I tested 0–10 mM $CaCl_2$ and cells were grown between 1–1.5 hours total induction time. Thus, if $CaCl_2$ is responsible for aiding a "Brownian ratchet" mode of transport, it may only influence an early (<1 hour) stage of secretion.

Lastly, I found that cardiolipin deficiency caused a clear decrease in T1SS levels in the MG1655 strain. As discussed above, decreased levels of T1SS production in MG1655 parent compared to BL21(DE3) may be the result of the phospholipid composition of MG1655 and greater intolerance of HlyB/D in its membrane leading to a cellular stress response which tapers down HlyB/D levels by suppressing plasmid expression. Cardiolipin and PE phospholipid deficiency

can cause severe membrane ultrastructure defects in *E. coli*, with an enlarged periplasmic space observed in cardiolipin deficient cells from exponential phase (Rowlett *et al*, 2017). Thus, the apparent decrease in T1SS levels in the MG1655 cardiolipin deficient strain compared to its parent could be due to these effects. Since my results from **Chapter II** seemed to indicate that cardiolipin could play a role in substrate transport, further experimental tests would need to be designed in order to test this further *in vitro*. This would require careful design, as cardiolipin contributes to overall protein complex stability (*e.g.,* McAuley *et al*, 1999; Lee *et al*, 2020; Xu *et al*, 2021) and also contributes to membrane curvature (Killian *et al*, 1994). To tease out physical effects of lipid (*e.g.* protein stability and membrane curvature) from energetic *(e.g.* charge) effects, multiple *in vitro* tests with different types of lipid would be needed. One option may be to use different nanodiscs with different lipid compositions: a PG/PE composition, a cardiolipin/PG/PE composition, and a phosphatidylserine/PE composition. One challenge would be T1SS assembly; one option is to express HlyB/D in nanodisc and TolC in a proteoliposome. Assembly could be mediated by a substrate which can trigger assembly but not be secreted (*e.g.,* Morgan, Acheson & Zimmer, 2017) until a modified substrate (*e.g*. with luciferase with C-terminal HlyA) is added and secretion measured by light emitted. By comparing secretion efficiency in the different proteoliposome compositions, it would be possible to see if there is an energetic effect on T1SS transport mediated solely by a charge effect or if it is a physical effect or a mix of physical and charge effects. An example is given in **Figure 3.4.3-1**. From this example, it is evident that many different trials could be run: for example, ATP consumption could also be measured with secretion efficiency using an NADH-coupled ATPase assay (*e.g.,* Souabni *et al*, 2021). One evident difficulty is in ensuring control of the experimental setup: this would mean having several controls. This would involve: 1) two background controls consisting of empty nanodiscs/liposomes and a separate control consisting of empty proteoliposome with HlyB/D nanodiscs, and 2) measuring light emission before and after addition of deficient substrate and NanoLuc. Formation of liposomes with nanodisc could be monitored using either a Raman-SEM or cryo-Raman microscope which has particle analysis capabilities.

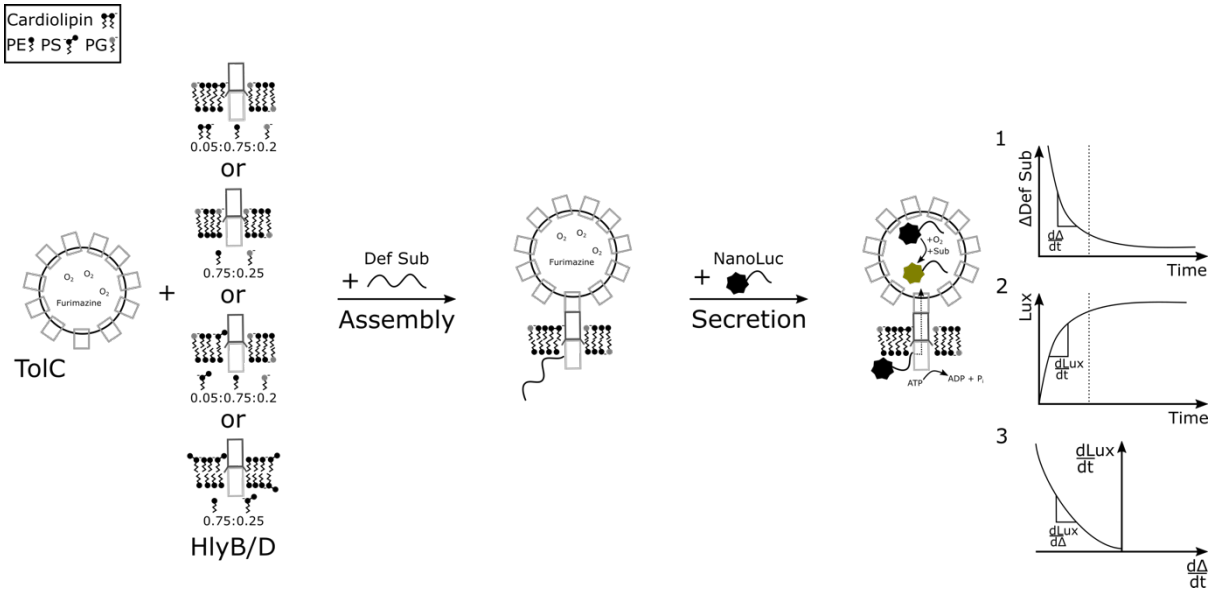


Figure 3.4.3-1. Secretion efficiency test. A small, modified 19kDa luciferase (England, Ehlerding & Cai, 2016) can be used to monitor T1SS efficiency. Proteosome bound TolC containing the NanoLuc substrate, furimazine and cofactor O_2, is mixed with purified HlyB/D containing mixtures of different lipid types, cardiolipin, phosphatidylethanolamine (PE), phosphatidylglycerol (PG), and phosphatidylserine (PS). Assembly is first induced by adding in a secretion deficient substrate which can promote assembly without being secreted (Morgan, Acheson & Zimmer, 2017). Then, NanoLuc with added C-terminal secretion tag is added; secretion of NanoLuc into the proteasome results in a light emitting reaction as furimazine is catalysed to furimamide. Titrating out the secretion deficient substrate as NanoLuc is titrated in will result in two rate changing curves: a decrease in secretion deficient substrate level (1) with an increase in light emission (2) as NanoLuc outcompetes the deficient substrate for secretion. The result will be a measure of the rate change in light emission as deficient substrate is titrated out (3). Note that direct measurements of 1) is not necessary to derive 3) – all that needs to be calculated is light emission rate change and titration levels of deficient substrate and NanoLuc.

5. Conclusion

I optimised the expression conditions for the T1SS using a flow cytometry based sequential design strategy. I found that T1SS production is predominantly driven by a couple of limiting factors, which act as “on” or “off” switches to expression, after which other effects act as modulators which tune levels to what the cell will tolerate. I propose that this observed effect is driven by saturation of the Sec translocon, which is responsible for inserting expressed membrane proteins into the cellular membrane. I find that the poor performance of the C43 strain is likely due to a robust ramping down of the expression of HlyB/D as a result of mutations to its *lac*UV5 promotor controlling T7 RNA polymerase expression which inhibits Sec translocon saturation. I also find that the performance of the MG1655 strain is likely due to the distribution of phospholipids in its membrane, with cardiolipin confined to the poles and hence could result in less routes of exit for the HlyA-eGFP substrate as well as a limited tolerance for HlyB/D in its membrane compared to the BL21(DE3) strain. I used the flow cytometry based sequential design strategy to test the cardiolipin requirement as suggested by MD simulations (see **Chapter II**). Results showed that the MG1655 cardiolipin deficient strain has reduced levels of T1SS production as compared to its parent strain. Since cardiolipin deficiency can cause structural defects in MG1655, I propose *in vitro* tests of secretion efficiency with different lipid compositions to tease out physical effects from charge effects and thus establish a more robust protocol for testing the impact of cardiolipin on substrate secretion. I also optimised detergent for a T1SS purification protocol, and made progress in optimising the purification strategy. I found that use of a glycerol cushion could aid in boosting T1SS levels by separating out competing free HlyA-eGFP from trapped T1SS before affinity pull-downs. I made cryo-EM grids for structural analysis, but found that the complex could be dissociating during plunge-freezing. To circumvent problems with purifying T1SS for structural analysis, I tested the use of cryo-CLEM to identify BL21(DE3) cells which are positively expressing T1SS. I found that these results matched the flow cytometry data and can be used for tomographic data collection. The use of whole cells for tomographic data collection and *in situ* reconstruction is further explored in **Chapter IV**.

Chapter IV: Cryo-Electron Tomography for *In Situ* Structural Elucidation

1. Introduction

1.1 Imaging Biological Complexes

When giving his Nobel Lecture in 1986, Ernst Ruska, one of the founders of the electron microscope, commented:

"*The resolution limit of the light microscope due to the length of the light wave which had been recognized 50 years before by Ernst Abbe and others could, because of lack of light, not be important at such magnifications [*of the electron microscope*]. Knoll and I simply hoped for extremely low dimensions of the electrons. As engineers we did not know yet the thesis of the "material wave" of the French physicist de Broglie that had been put forward several years earlier (1925). Even physicists only reluctantly accepted this new thesis. When I first heard of it in summer 1931, I was very much disappointed that now even at the electron microscope the resolution should be limited again by a wavelength (of the "Materiestrahlung"). I was immediately heartened, though, when with the aid of the de Broglie equation I became satisfied that these waves must be around five orders of magnitude shorter in length than light waves. Thus, there was no reason to abandon the aim of electron microscopy surpassing the resolution of light microscopy*." (Ruska, 1987)

Indeed, the short wavelengths of electrons, their scattering effects from material, and the ability to condense these scattered waves back into an image via electromagnetic lenses has made electron microscopy a powerful tool for atomistic insights. The basic design of a transmission electron microscope has remained relatively unchanged since their invention in the 1930's (**Figure 4.1.1-1**). The power of transmission electron microscopy (TEM) lies in two parts: 1) Image formation of fine details, and 2) Image processing and 3D reconstruction to reveal atomistic insights of a sample. Image formation is the ability of a TEM to produce "undistorted" 2D projection images of a sample. Image processing for 3D reconstruction is

the application of mathematical and computational methods to 2D images to generate insights – this means applying signal-processing methods to convert 2D projection images into a resulting 3D map. The success of 2) will depend on the ability of 1) to produce sufficient structural detail. We will cover both in turn before looking at the limitations in TEM.

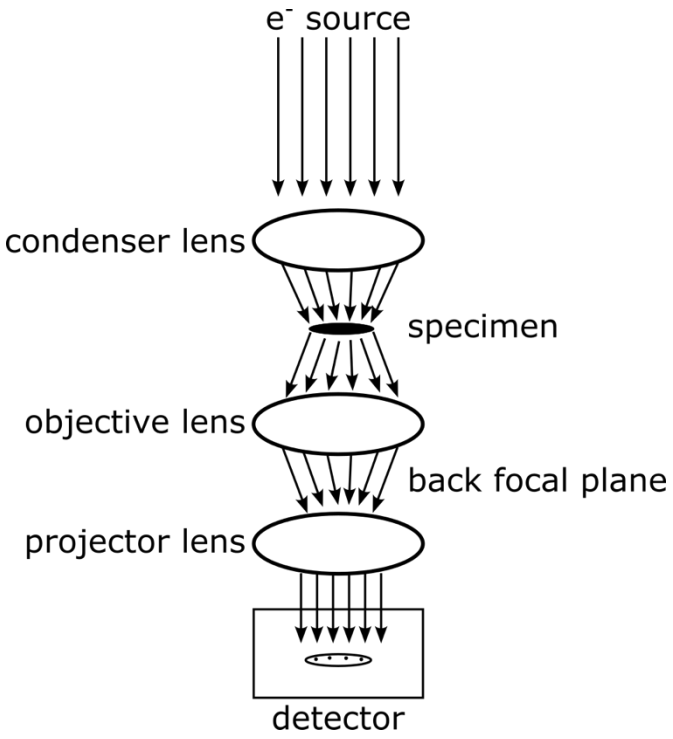


Figure 4.1.1-1. Schematic of an electron microscope. Imaging in an electron microscope works by transmitting electrons from a source towards a specimen. The condenser lens condenses the electron beam and focuses it on the specimen. Electrons passing through the specimen will be diffracted and form the back focal plane after passing through the objective lens. Finally, the real space image is formed by passing the electrons through the projector lens and onto a detector. Note that not all the lenses are shown.

1.2 Part 1: Image Formation

Image formation in a TEM is a complex process, involving sample, the electron beam, electron lensing, and image capture. This process will ultimately affect the output resolution of the 3D structure obtained. Image formation relies on the interaction of the incoming electrons with the atoms in a sample. Upon encountering an atom, the electrons can be scattered either elastically (without loss in energy) or inelastically (with loss in energy). In terms of image formation, elastically scattered electrons will result in an output image whilst inelastically scattered electrons will contribute to noise in the image. Most electrons will not interact with

the material at all and hence contribute to the background (unscattered) wave. It is the intensity of the scattered electrons relative to the background wave from which an image is derived. This scattering can be thought of as creating a change in direction in the paths of the electrons: mathematically, we say that the electronic wave has undergone a *phase shift* relative to the unscattered wave (**Figure 4.1.2-1**). In the case of biological samples, which primarily contain light atoms (*e.g*. C, N, O, P), the observed scattering effect is small and hence small phase shifts are created, leading to low-contrast images (Frank, 2006).

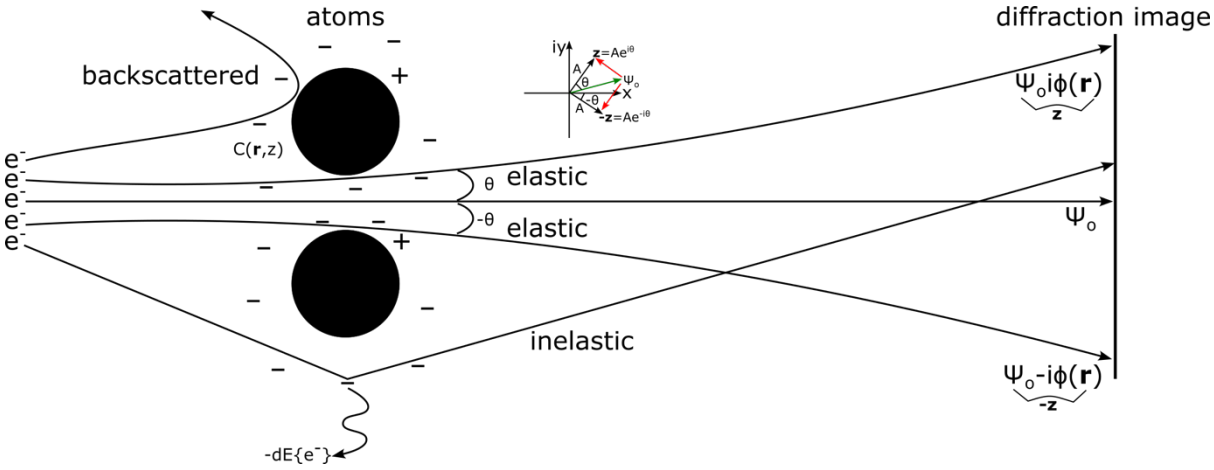


Figure 4.1.2-1. Electron scattering in the sample. Electrons interacting with atoms in a sample can remain unscattered (ψ_o), backscattered, or elastically or inelastically scattered. Inelastically scattered electrons lose energy as a result of collision with the nucleus or electrons of an atom and hence contribute to the noise in an image. By contrast, elastically scattered electrons maintain the same energy after scattering and hence contribute to the diffraction image. Image contrast is derived from the angle of scattering (θ) relative to the unscattered background wave. As shown in the Argand diagram, this scattering (red arrow) produces a resulting phase shift, described by a wave amplitude (A) and phase ($i\theta$) relative to the unscattered wave (green arrow). Interaction of electrons with the atomic Coulombic potential is described by the function C(**r,**z).

The electron microscope essentially takes the scattered electrons and projects them back to a 2D image. Thus, to understand image formation, we must understand the relationship between the 3D object being imaged and its resulting 2D projections.

Since biological samples are weak-phase objects, we start with the weak-phase approximation for biological samples, which is based on Taylor expansion of the spatial

function for the wave outgoing from the illuminated specimen (Frank, 2006) (see **Figure 4.1.2-1**):

$$\psi(\boldsymbol{r}) = \psi_o \left[1 + i\phi(\boldsymbol{r}) - \frac{1}{2}\phi(\boldsymbol{r})^2 + \cdots\right] \quad \text{(equation 4.1.2-1)}$$

where $\psi(\boldsymbol{r})$ describes the outgoing wave and ψ_o the incoming wave. The weak-phase approximation of equation 4.1.2-1 only considers the first two terms due to the nature of biological samples, where only small changes in contrast are observed. The observed phase-shift ($\phi(\boldsymbol{r})$) is described by (Frank, 2006) (see **Figure 4.1.2-1**):

$$\phi(\boldsymbol{r}) = \int C(\boldsymbol{r}, z)\, dz \quad \text{(equation 4.1.2-2)}$$

where $C(\boldsymbol{r}, z)$ is the 3D Coulomb potential of the object and **r** is a 2D vector of coordinates (x,y). Note that in equation 4.1.2-1, the first term describes the unscattered wave, while the subsequent terms describe the scattered wave.
The observed phase contrast in the image is determined by the contrast transfer function (CTF) (Frank, 2006):

$$CTF = \sin\gamma(\boldsymbol{k}) \quad \text{(equation 4.1.2-3)}$$

where **k** is the 2 dimensional spatial frequency (**k**=(kx,ky)) and γ describes the effect of lens aberrations and defocus on the phase shift (see Frank (2006) for more details on equations). Here, frequency is related to scattering: high scattering results in high-frequency, while low scattering results in low frequency.

The CTF describes how phase-contrast is transferred from the lens to the image, dependent on defocus used. This results in regions of negative (black on white), positive (white on black), and no contrast transferred (zero crossings) (**Figure 4.1.2-2**). In practice, the CTF is dampened at high spatial frequencies (**Figure 4.1.2-2**).

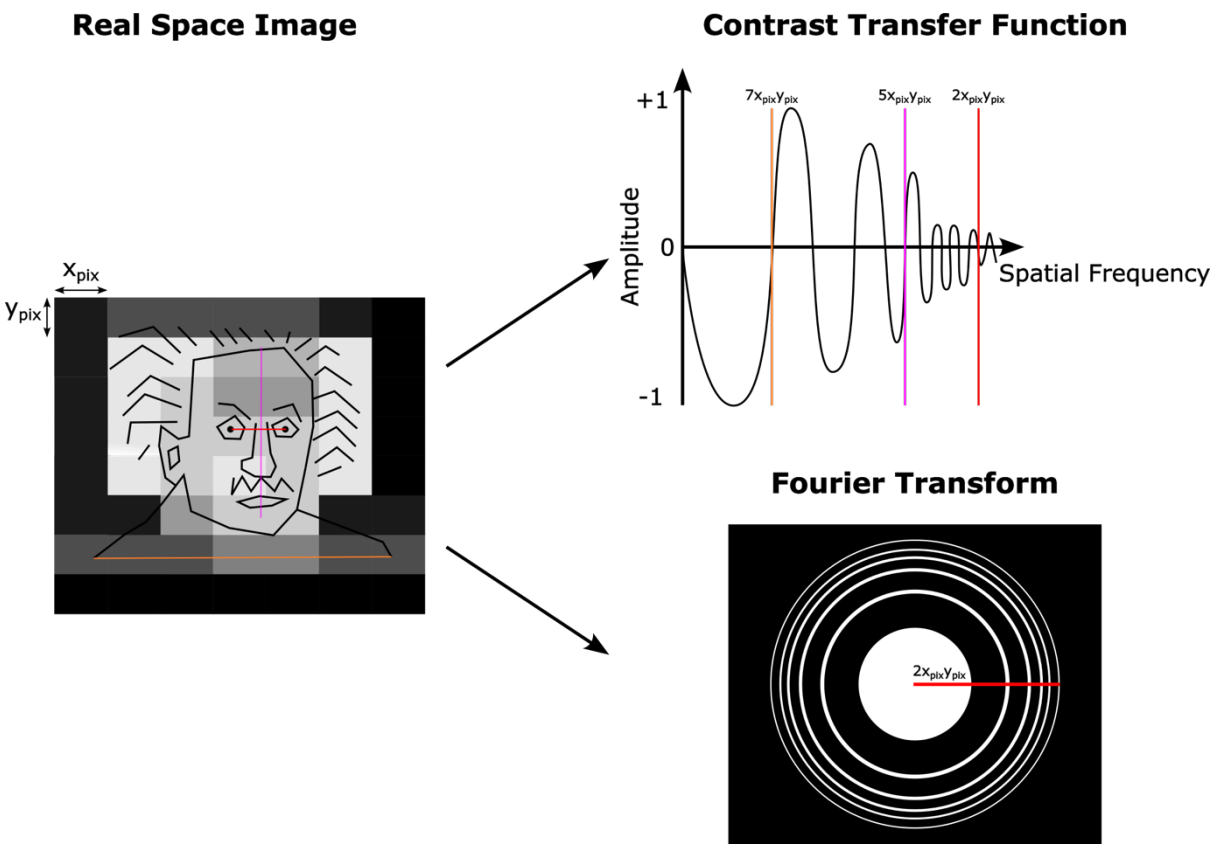


Figure 4.1.2-2. Image contrast and spatial resolution. Resolution information of the real-space image can be viewed as a function of the Contrast Transfer Function (CTF) and Fourier Transform (FT). The CTF describes how the electron microscope transfers contrast (amplitude) from the object to the resulting image as a function of spatial frequency. Contrast transferred can be negative (black on white) or positive (white on black), or no contrast can be transferred (zero crossings). The zero crossings represent a loss of information which must be filled in by applying different defocuses during data collection. The relationship between CTF and image spatial resolution is shown; all spatial resolutions can be described before the first zero crossing (orange line) which shows the largest frequency encapsulating the imaged object. Intermediate frequencies (*e.g.* pink line) then describe different resolvable distances in the object up to the smallest frequency (red line), after which there is no real signal which can be resolved. As the spatial frequency increases the amplitudes are dampened and hence high-resolution information becomes more sensitive to noise compared to the low-resolution information. The FT is another way of showing the same information; in this case, instead of a 1D sine function, it is a 2D contour of image spatial resolution. White "Thon rings" (Thon, 1966) show contrast transferred while black represents no transfer of contrast. The smallest spatial frequency will be visible as the last Thon ring; in this example, it corresponds to the Nyquist limit (Nyquist, 1928) which is twice the pixel size (red line). Downstream image processing will apply CTF-correction, which at a minimum involves converting negative phase contrast to positive phase contrast ("phase-flipping"). Boosting high-resolution information will also involve upweighting high spatial frequency amplitudes and down-weighting low spatial frequency amplitudes (*e.g.* Wiener filter) (Wiener, 1964).

The projected image can be described by (Frank, 2006):

$$I(\boldsymbol{r}) = \phi(\boldsymbol{r})\ o\ h(\boldsymbol{r}) \qquad \text{(equation 4.1.2-4)}$$

where $I(\boldsymbol{r})$ is the intensity of the image, $\phi(\boldsymbol{r})$ is the phase-shift (described in equations 4.1.2-1 and 4.1.2-2), $h(\boldsymbol{r})$ is the point spread function (PSF), and o is a convolution operation between the two functions. This relationship is essentially describing what was mentioned in the first paragraph: over the whole image plane, by what amount (frequency) are the electrons scattered to produce the observed contrast. The point spread function is therefore a 2D representation of spatial resolution in the object.

Finally, we could decompose this spatial resolution information into a series of cosines, describing all frequencies (spatial distances) observed, from lowest to highest frequency (Frank, 2006):

$$FT(\boldsymbol{k}) = O(\boldsymbol{k}).A(\boldsymbol{k}).CTF \qquad \text{(equation 4.1.2-5)}$$

where $O(\boldsymbol{k})$ is the Fourier transform of the outgoing wave, and $A(\boldsymbol{k})$ is the aperture function. The function $A(\boldsymbol{k})$ is step-wise and describes what frequencies are captured based on the objective aperture (see Frank (2006) for details).
We can also write equation 4.1.2-4 as (Frank, 2006):

$$I(\boldsymbol{r}) = |\psi_i(\boldsymbol{r})|^2 \qquad \text{(equation 4.1.2-6)}$$

where $\psi_i(\boldsymbol{r})$ is the inverse Fourier transform of equation 4.1.2-1. Equation 4.1.2-6 is the *power spectrum* of the 2D projection. This is another way of describing equation 4.1.2-3 (**Figure 4.1.2-2**). Here it may be noted that while the Fourier transform decomposes our object plane into a series of observed spatial resolutions, the inverse Fourier transform describes the resulting image intensity. Thus, to recreate the exact 3D slice in which we observe our 2D projection, we only need to apply the inverse Fourier transform.

1.2.1 A note on resolution

One may have noted from equation 4.1.2-4 that although the PSF describes what can be spatially resolved in an object, it is not directly observed in the resulting image (**Figure 4.1.2-2**). In terms of resolution, as noted by Scherzer (1949), the minimum resolvable distance is:

$$d = \frac{0.6\lambda}{sin\alpha} \qquad \text{(equation 4.1.2-7)}$$

where d is the resolvable distance between two object points (*e.g.* atoms), λ is the wavelength in the object plane, and α is the aperture angle of the objective lens. Equation 4.1.2-7 is what is known as the Rayleigh criterion for minimum resolvable distance. As noted above, we cannot directly measure this as the PSF is not directly observed in the resulting image, and hence the FT is used instead (discussed in the next section).

1.3 Part 2: Image Processing and 3D Reconstruction

To convert a series of 2D projection images that include effects of PSF and optical aberrations of the microscope and limitations of the detector into a 3D reconstruction that represents the true object, mathematical methods are applied. DeRosier & Klug (1968) first proposed the method of 3D object reconstruction via Fourier synthesis. Here defocus is estimated by fitting the CTF to the image power spectrum; from the CTF, phase and amplitude are known and 3D reconstruction proceeds by fitting 2D projections to 3D Fourier slices ("central slices"). This is known as the projection theorem: the 2D Fourier transform of a 2D projection gives a central section through the 3D transform of the 3D density. The real-space object is then fully recovered by applying the inverse Fourier transform after all the slices have been determined. This requires adequate sampling of the Fourier space describing the object (**Figure 4.1.3-1**). Here, symmetry is useful as symmetries in real-space will have Fourier counterparts and hence fewer projections will be required to reconstruct the entire object; for example, in the case of n-fold rotational symmetry the number of required projections to represent the entire object will be reduced by a factor of n. Obtaining different projections will then either rely on tilting the microscope stage to obtain different views of the object, or using different images

with different particles. For single-particle 3D reconstruction, the latter method is applied, while for tomographic reconstruction the former is applied. Since these particle projections are in unknown orientations, accurate orientation estimation is paramount in order to correctly determine the 3D Fourier slice the projection represents; DeRosier & Klug (1968) proposed creating an initial model and iteratively refining particle orientations based on this model. All 3D reconstruction algorithms are based on this idea (*e.g.,* Ludtke, Baldwin & Chiu, 1999; Castaño-Díez *et al*, 2012; Scheres, 2012; Galaz-Montoya *et al*, 2016; Punjani *et al,* 2017; Himes & Zhang, 2018).

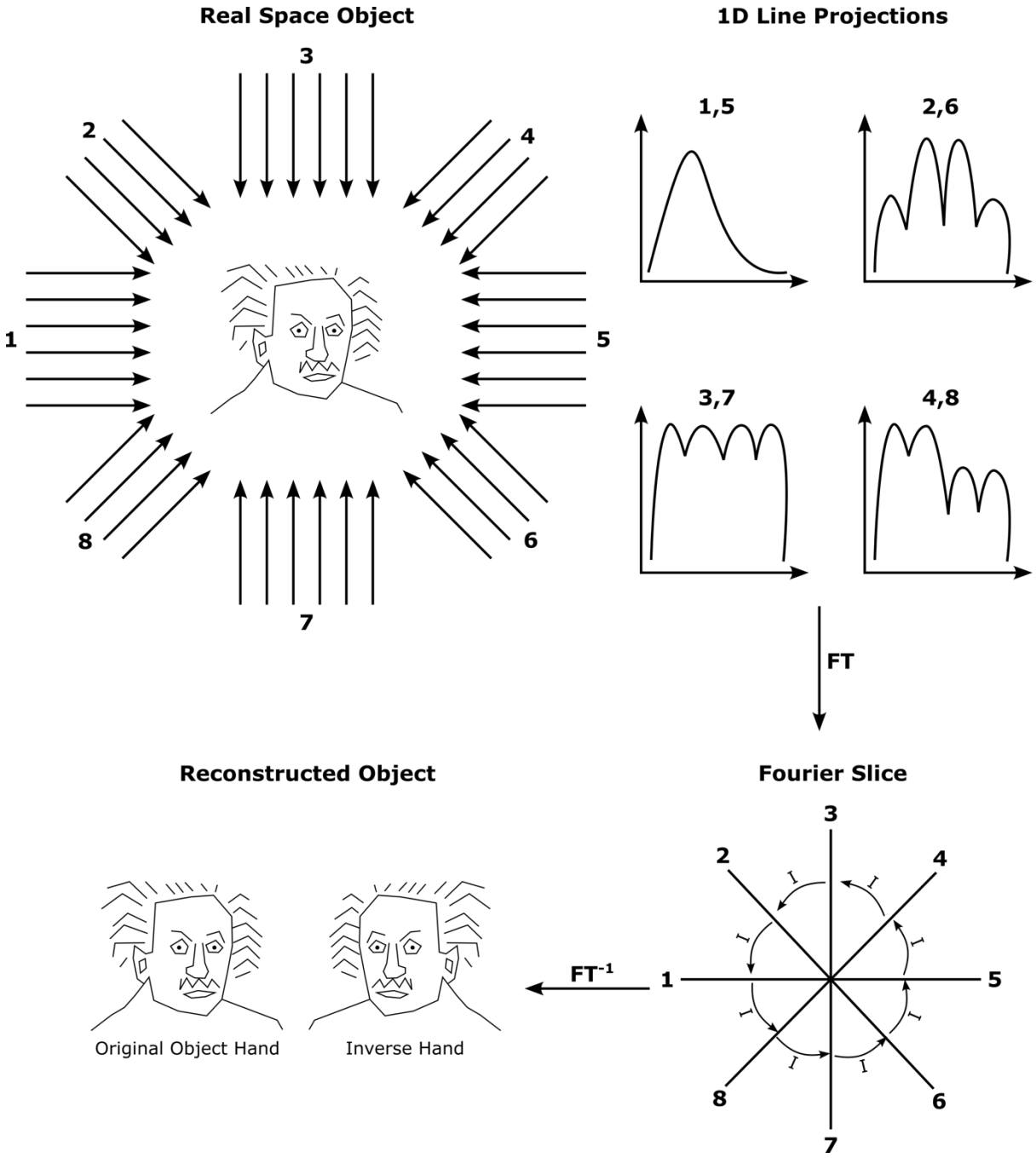


Figure 4.1.3-1. Fourier Slice Theorem and Object Reconstruction. Object reconstruction proceeds by taking projections of the object at different orientations. Some projections will be common depending on the orientation sampled. The Fourier Transform (FT) of the projections gives a central slice through the Fourier transform of the object. Interpolation (I) between Fourier slices is done before the object can be reconstructed by applying the inverse Fourier Transform (FT^{-1}). As shown in this example, faithful reconstruction of the object will depend on adequate sampling of the Fourier space. Note that reconstruction can give two different hands, only one of which will belong to the original object; this is because their projections will be identical. Also note that fine features observed in the reconstructed object will depend on Fourier space sampling.

Klug & Crowther (1972) recognised that the 3D reconstruction procedure is highly sensitive to noise. They introduced the idea of a reliable reconstruction as one in which a certain degree of error can be tolerated. They thus formulated the problem of reconstruction: for a given radial cut-off, R_o, in the spatial frequency spectrum of the object, the number of parameters, P, needed to specify a 3D object is:

$$P = volume\ x\ R_o \qquad \text{(equation 4.1.3-1)}$$

From this, P' is derived: the number of parameters that can be extracted from the given projections according to a criterion of error tolerance. This will depend on the number and relative geometry of various views (*e.g*. symmetry) and also the maximum spatial frequency present in the data.

Klug & Crowther (1972) also recognised that the number of P′ recoverable parameters will not represent the object uniformly and hence interpolation is required between projections to obtain a uniform sampling. Reconstruction is then formulated as:

$$p_r = \sum_P^{P'} \frac{c_p}{\lambda_p} \varphi_p \qquad \text{(equation 4.1.3-2)}$$

where p_r is the number of projections required to reconstruct an object, φ_p and λ_p describe the respective eigenfunctions and eigenvalues of a particular set of projections, and c_p are the unknown coefficients for each φ_p . The summation describes all recoverable P′ components from the data. The function φ is described by:

$$\varphi = F^{-1}ISF \qquad \text{(equation 4.1.3-3)}$$

where F^{-1} is the inverse Fourier Transform, I is an interpolation on the sampling S of the object, and F is the Fourier transform.

The formulation by Klug & Crowther (1972) intuitively describes what can be "restored" unambiguously (with limited error) from a given set of projections, given the presence of

noise. They also introduce the idea that final resolution of a reconstructed 3D object will depend on the size of the object and the number of projections available. Too few projections means that the Fourier space is sampled inadequately, and introduced errors will have a magnified effect due to the limited sampling (i.e., errors ε_p with eigenvectors φ_p will be amplified on the order of $1/\lambda_p$ (Klug & Crowther, 1972). Smaller objects will have fewer pixels (dependent on magnification) to describe their spatial sampling (see equation 4.1.3-3) than larger objects, and hence they will have fewer degrees of freedom. In this case, errors introduced by interpolation is a concern.

In the above description, we introduced the idea of a radial cut-off (R_o) in the spatial frequency spectrum describing an object. Minimum resolvable distance in an object is thus determined by the smallest spatial frequency which can be determined above the noise level. This can be seen by last "Thon" ring (Thon, 1966) in the image power spectrum (**Figure 4.1.2-2**) and is what is known as the Nyquist limit. This comes from the work by Nyquist (1928) in relation to telegraph signalling; to reconstruct a transmitted messaged composed of a series of sinusoidal waves, the minimum number of components is half of the number of signal elements. Similarly, we can derive the sampling rate (v) required to recover all Fourier components in a waveform:

$$f_{Nyquist} = \frac{1}{2}v \qquad \text{(equation 4.1.3-4)}$$

Conversely, in TEM at a magnification with a sampling of S Å/pixel, the resolution will be limited to 2 x S Å. A full description of the sampling theorem and how it relates to 3D reconstruction can be found in Hanszen (2018). It is sufficient to know the following: 1) the lowest spatial frequency contained in an object is:

$$R_e = \frac{1}{2x_e} \qquad \text{(equation 4.1.3-5)}$$

where $2x_e$ is the object field diameter. All higher frequencies are described by:

$$R_k = \frac{k}{2x_e},\ k = 2, 3, 4 \ldots \qquad \text{(equation 4.1.3-6)}$$

We thus have the following sampling points in an object, described by ΔR_s, the change in sampling radial frequency, and Δx_{ps} the change in the sampling point in a Gaussian image:

$$\Delta R_s = \frac{1}{2x_e};\ \Delta x_{ps} = \frac{\lambda f}{2x_e}\frac{(M' - M'_Q)}{M'} \qquad \text{(equation 4.1.3-7)}$$

where λ is the wavelength of the electrons, f is focal length, M' is the magnification, and M'_Q is the magnification of the image of the electron source located in the plane of the exit pupil.

From equation 4.1.3-7, we can see that for each sampling point, the wavefunction (ψ) has only one spatial frequency, i.e:

$$R_e = \frac{1}{100}\text{Å}^{-1}$$
$$R_2 = \frac{2}{100}\text{Å}^{-1}$$
$$R_3 = \frac{3}{100}\text{Å}^{-1}$$
$$\text{etc.} \qquad \text{(equation 4.1.3-8)}$$

The complete wavefunction describing the object field is then determined by the number of countable spectral points (equation 4.1.3-8). At each sampling point, the image of the electron source is broadened by diffraction at the borders of the object; this is further broadened by lens aberrations (Hanszen, 2018).

There are two types of lens aberrations: spherical and chromatic. Spherical aberration is caused by the path distortions during refocusing of electrons in the electromagnetic lens after diffraction; the outer part of the lens is more prone to path distortion than the inner part (Rose, 2008). This causes distortion of the relative values of the spectral points (equation 4.1.3-8) (Hanszen, 2018). Chromatic aberration is caused by small changes in electron velocity as it passes through the lens (Rose, 2008). Both aberrations cause unintended phase shifts (Hanszen, 2018). There are methods to reduce these defects: monochromators ensure the electron energy is more uniform before passing through the specimen which helps reduce the

effects of chromatic aberration; while the use of aberration correctors in the microscope and energy filters to remove inelastically scattered electrons also helps (Rose, 2008).

Finally, we are ready to discuss some algorithms for 3D reconstruction, synthesizing the above points. From the projection theorem, we can see that there must be some relation between different projections; if we project a 2D projection (object), we get a 1D line indicating changes in image intensity through the object. Now we can see that different 2D projections (objects) in different orientations will have intersecting 1D projections, which we call *common lines* (**Figure 4.1.3-1**). Orientation search is then a matter of determining corresponding common lines (Penczek, Zhu & Frank, 1996). For convenience, Euler space is used instead of standard Euclidean space for parameterising a 3D object (**Figure 4.1.3-2a**). Reconstruction is then a matter of determining correct Euler angles for each projection (**Figure 4.1.3-2b**), as well as x- and y-shifts (for 2D single-particle) and an additional z-shift (for 3D tomography). Iteratively optimising these parameters for each projection to achieve a reasonable 3D reconstruction can proceed via back-projection (**Figure 4.1.3-2b**) (Penczek, 2010a). Further developments aimed to use mathematical methods such as maximum likelihood to improve parameter estimation (Scheres, 2012) or a mix of maximum likelihood and stochastic gradient descent for 3D reconstruction (Punjani *et al*, 2017).

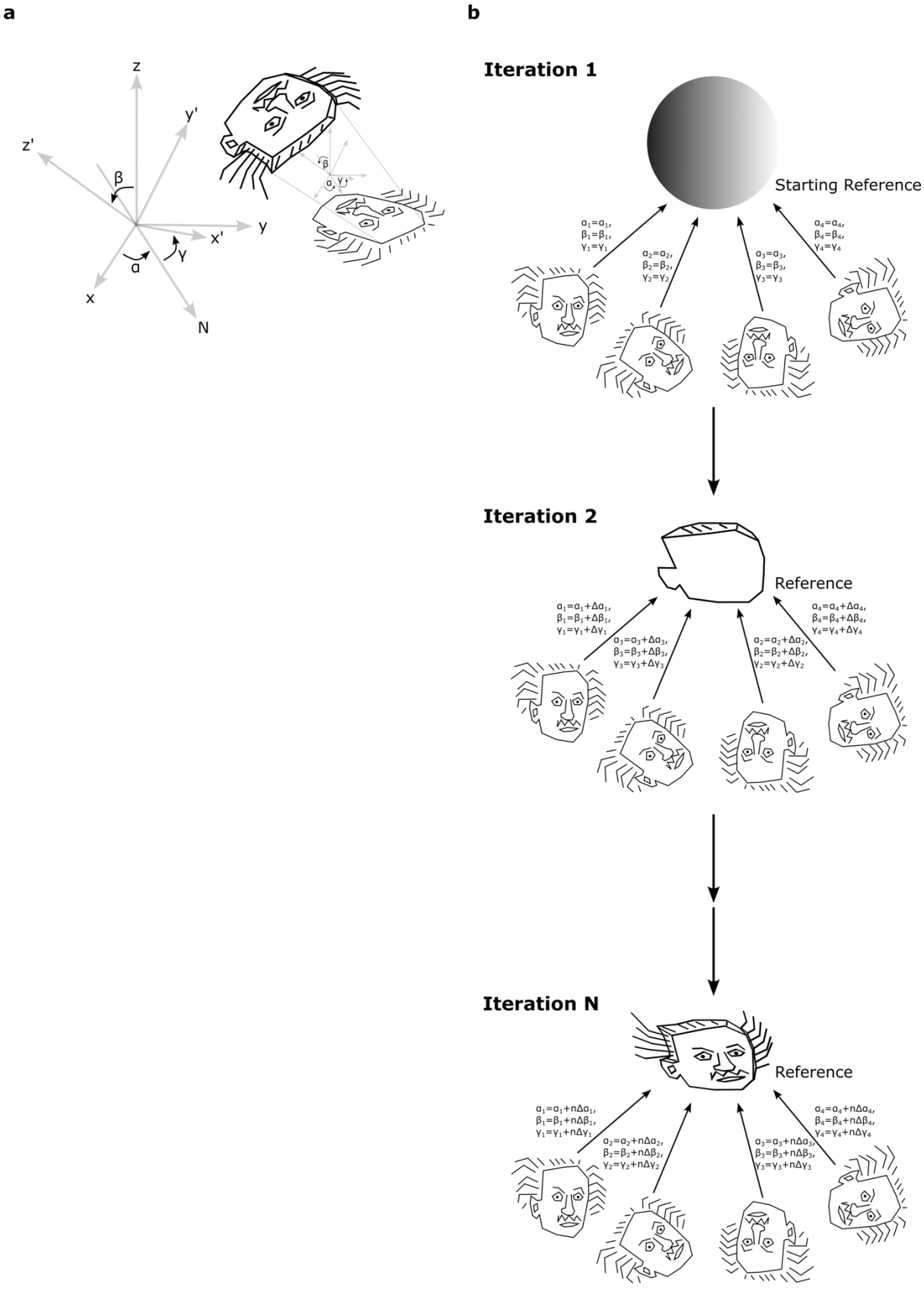


Figure 4.1.3-2. Image orientation and projection matching. a) The orientation of a projection is defined by three Euler angles, α, β, and γ. The angle α represents in-plane rotation of the object while angles β and γ represent

out-of-plane rotations relative to the respective z- and x-axis. **b)** Object reconstruction by projection matching proceeds by starting with a reference and assigning initial Euler angles to all projections. The projections are backprojected along these angles to make a new reference. The angles are updated based on projection matching to this reference and a new reference made by backprojection. This proceeds until the N^{th} iteration, where the model should converge to a solution. Note that particle shifts (in the x- and y-plane or additional z-plane) are not included in this example but they will be taken into account to get optimal alignments.

1.4 Limitations in TEM

Despite the promises of atomistic insights, use of transmission electron microscopes on biological materials was hampered for many decades due to their sensitivity to ionising radiation. To image the biological samples in a vacuum, preservation methods were needed, and first attempts settled on the use of heavy metals such as uranyl acetate in a method called negative staining (Ohi *et al*, 2004). This produced high-contrast images of the shells of the biological samples under investigation, due to the large atomic weight of the metals used that could produce strong amplitude contrast, but also created resolution-limiting effects and structural distortions (Frank, 2006). The next breakthrough was pioneered Jacques Dubochet, who trialled the use of plunge-freezing methods to produce “frozen hydrated” biological samples (Adrian *et al*, 1984). Rapid freezing and preservation at ultra-low temperature allowed these samples to be imaged in a near-native glass-like state. As biological material is highly sensitive to radiation damage, low electron doses are required to preserve sample integrity whilst imaging the frozen specimens (Orlova & Saibil, 2011). This means that fewer electrons are detected per pixel on the camera, creating noise and hampering 3D reconstruction methods. Thus, cryo-TEM was for many decades in a state of “blobology” until the development of direct-electron detectors which vastly improved obtained resolution (Danev, Yanagisawa & Kikkawa, 2019). The main benefit of direct electron detectors is their improved detective quantum efficiency (DQE) over traditional charge-coupled device (CCD) cameras (Faruqi & Henderson, 2007). In CCD cameras, the incoming signal from the electrons is detected by the camera pixels by first being converted to photons which activates an electric current that is detected by the camera sensor. This adds noise to the resulting image, lowering the signal-to-noise ratio (SNR) of the camera detector (Faruqi & Henderson, 2007). In contrast, direct electron detectors remove the need for photon conversion, resulting in

higher SNR of the detector and thus more signal can be extracted from the resulting images to produce high-resolution 3D reconstructions (Kühlbrandt, 2014).

As mentioned, lens aberrations create distortions to 2D images although efforts have been made to reduce these effects. CTF estimation is also an essential requirement to image processing; to account for phase modulations, phase flipping is applied. Additional filters can also be applied to boost the SNR at high frequencies (*e.g*. Wiener (Wiener, 1964)). A range of defocus values must also be used if missing information is to be filled in from the zero crossings (Penczek, 2010b). Although defocus is an essential requirement to obtain contrast in an image (Hanszen, 2018), defocus gradients can arise across an image. One method to account for this is to apply a per-particle CTF-correction (Galaz-Montoya *et al*, 2016). Other problems include Ewald sphere correction for non-flat 2D projections in large objects (Zhu *et al*, 2018), beam tilt, and dynamic electron scattering (Zhang & Zhou, 2011).

Lastly, 3D reconstruction is an optimisation problem and hence one must be careful not to overfit the data; a well-known issue is the "Einstein from noise" problem where a model is fitted to noise (Henderson, 2013). This is particularly an issue if a template is used to find particles ("template matching") (*e.g.*, van Heel, 2013; Subramaniam, 2013). Resolution estimation is also another contentious issue; as mentioned above, direct observation of the PSF to derive a resolution estimate is not possible (as it is convoluted with the microscope FT) and Fourier methods are used. A popular choice is the Fourier Shell Correlation (FSC) which estimates resolution by comparing the Fourier components in two independent reconstructions ("half-maps"); a threshold cut-off is applied upon which the reported resolution is based. Resolution estimates are really two problems: 1) with the data collected, what is the *variation in estimated resolution of my 3D map*?, and 2) with the data collected, what is the overall variance in resolution *had I assigned particles differently between the half-maps*? Although subtle, we can see that problem 1) is not identical to problem 2). Problem 1) is really an issue with particle flexibility; some parts of the map are more resolvable than others as they are more rigid and undergo less thermal motion. Therefore, their atomic B-factors show less "spread". Using B-factor information has been used to "sharpen" resulting 3D maps (*e.g.*, Kaur *et al*, 2021). Other methods have been applied to obtain local-resolution estimates (Vilas *et al*, 2020). Problem 2) appears when it comes to comparisons, for example

trying to determine the effect of a particular collection strategy on the resulting resolution. In this case, a point estimate of resolution is not particularly helpful without the variance information over the entire dataset. This has only been addressed relatively recently (Beckers & Sachse, 2020). In either case, independent methods of resolution validation are recommended, such as examination of observed secondary structure and amino acid density profiles (Orlova & Saibil, 2011).

1.5 Cryo-Electron Tomography

Cryo-electron tomography is a method of creating 3D volumes from a 2D tilt series of a vitrified specimen under investigation (**Figure 4.1.5-1**). ET has been a method practiced since the 1960's (Hart, 1968), where specimens were preserved in plastic resin and thin sections cut-out using a diamond cutter (Olins *et al*, 1983). Due to the preservation methods used, fine structural details were lost (Luther & Crowther, 1984; Skoglund *et al*, 1986; Baumeister, 2005), and cutting often led to artifacts (Al-Amoudi, Studer & Dubochet, 2005). Use of vitrification methods for cryo-ET was optimised in the late 1990's to early 2000's (Grimm *et al*, 1998; Matias *et al*, 2003; Al-Amoudi, Norlen & Dubochet, 2004; Zhang *et al*, 2004). A breakthrough came in 2007 with the publication of cryo-focused ion beam (cryo-FIB) milling (Marko *et al*, 2007). In this case, frozen-hydrated specimens are cut using a beam of gallium ions instead of traditional diamond cutters, avoiding some of the cutting artifacts such as compressions, knife marks, and crevasses (Marko *et al*, 2007). Although successful in making specimens thin enough (<200 nm) for TEM imaging, implementing a cryo-FIB and subsequent lift-out procedure for imaging remains challenging (Parmenter & Nizamudeen, 2020).

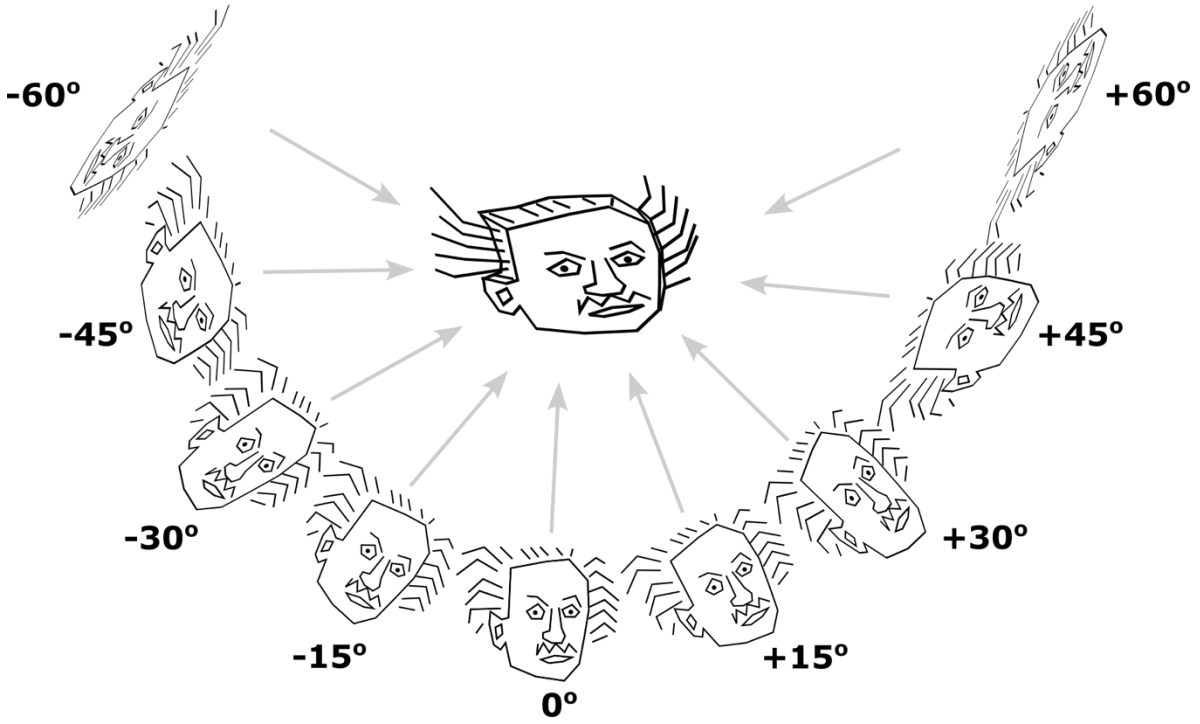


Figure 4.1.5-1. Object Reconstruction by Electron Tomography. Different views of the 3D object are obtained by tilting the stage in increments. These are then projection matched to reconstruct the object in 3D. Note that most stages cannot go beyond 60° and hence a wedge of information will be missing ("missing wedge") (Hagen, Wan & Briggs, 2017).

Data collection of a vitrified specimen proceeds via a tilt-scheme; this is usually dose symmetric if subsequent subtomogram averaging is applied to ensure high-resolution features are not lost due to radiation damage (Hagen, Wan & Briggs, 2017). Data collection follows a procedure of tracking, focusing, and imaging to acquire a tilt series; subsequent processing then follows a typical workflow (**Figure 4.1.5-2**). As a first step, each tilt series must be motion corrected using the acquired movies (*e.g.* MotionCor2 (Zheng *et al*, 2017)), followed by stack generation and CTF estimation (Pyle & Zanetti, 2021). Due to tilting of the stage during acquisition, CTF estimation is not as straightforward as single-particle methods. Higher tilts cause defocus gradients across the tilt axis which must be accounted for if high-resolution information is to be extracted from a subtomogram averaging procedure (Turoňová *et al*, 2017; Himes & Zhang, 2018). Before generating a tomogram, the stacks must be aligned, with procedures accounting for shifts in-plane as the tilt series is acquired (Mastronarde & Held, 2017). Since the microscope stage cannot go beyond a 60-to-70-degree tilt, and the sample becomes too thick at higher tilts for electrons to pass through, reconstructed tomograms will have a "wedge" of missing information. Recent algorithms

have been developed to account for this lost information and reduce resulting Fourier artifacts (*e.g*, Deng *et al*, 2016; Ding *et al*, 2019). Another method is to apply subtomogram averaging to a particle of interest. Here, a good 3D reconstruction of the target particle can be obtained provided that the tomogram contains a sufficient number of particles with different views. For subtomogram averaging, particle coordinates must be determined from the tomograms (usually binned for greater contrast) and used to generate and refine maps. Per particle CTF correction is then usually applied (Pyle & Zanetti, 2021). After model generation, tomogram alignments can be improved (sub-tilt refinement) and particles re-extracted with updated parameters to improve the final resolution of the structure (Pyle & Zanetti, 2021). For cryo-ET data processing, pipelines are typically adapted for each specimen and may involve one or more programmes (*e.g.,* Bharat & Scheres, 2016; Himes & Zhang, 2018; Chen *et al*, 2019; Burt *et al*, 2021).

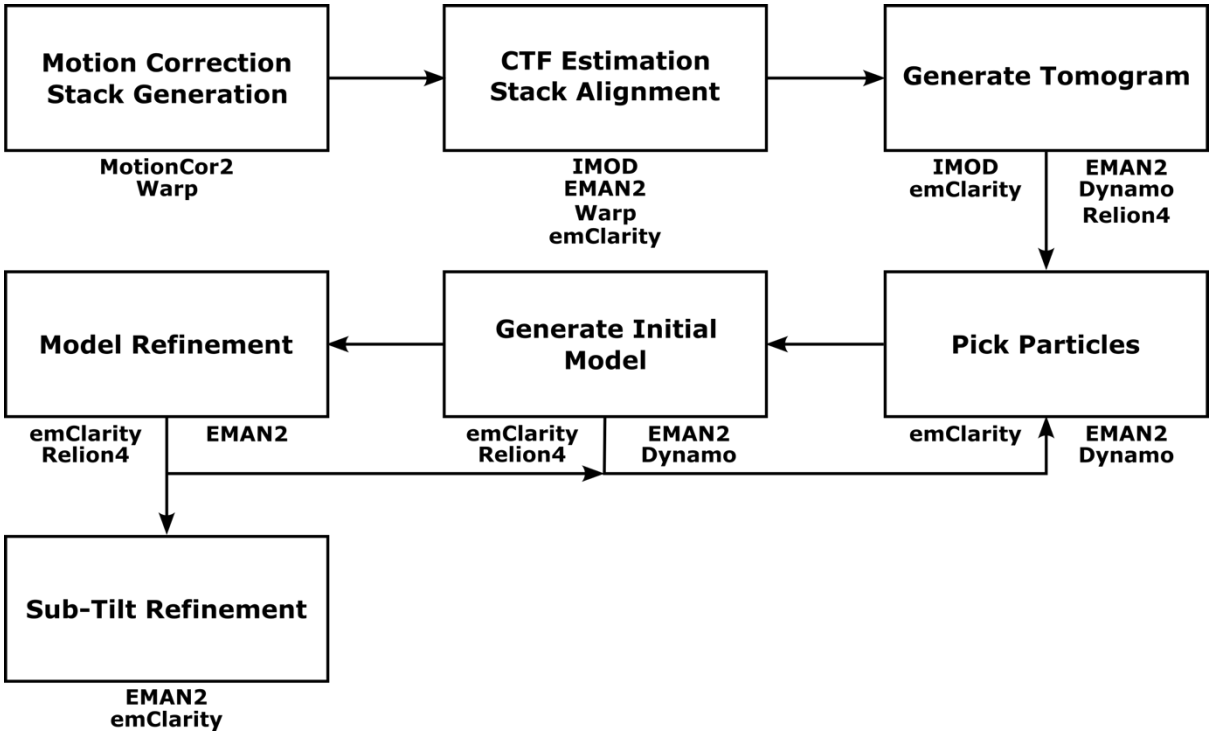


Figure 4.1.5-2. Cryo-Electron Tomography Workflow for Sub-Tomogram Averaging. After data collection, motion correction is applied on the movie frames and a stack generated of the 2D tilt-series. CTF estimation can be applied at this point and the stack aligned before a tomogram is generated. Particles are usually picked on the binned tomogram and positions extracted to make an initial model. This model is then refined. At this stage, more particles can be picked at a lower (or no) binning and the process repeated until the best model is produced. Finally, sub-tilt refinement can be applied by using the particle positions as fiducials to produce a

more accurate tomogram and hence a better model downstream. Examples of common programmes used at each step is given: MotionCor2 (Zheng *et al*, 2017), Warp (Tegunov & Cramer, 2019), IMOD (Mastronarde & Held, 2017), emClarity (Himes & Zhang, 2018), EMAN2 (Chen *et al*, 2019), Dynamo (Castaño-Díez *et al*, 2012), and Relion4 (Bharat & Scheres, 2016; Zivanov *et al*, 2022).

1.6 Aim & Objectives

Cryo-ET with sub-tomogram averaging is a powerful method for determining 3D structures from complex samples. It is particularly useful for *in vivo* structures, especially where purification procedures may strip the sample of key partners, or remove the specimen from its biological context, or such as native membrane in the case of transporters such as MacAB-TolC and the Type I Secretion System (T1SS). The aim of this chapter is to determine a 3D *in vivo* structure for the MacAB-TolC system (see **Chapter V**). This system was chosen as there is a good cryo-EM structure available to act as a reference for subtomogram averaging (Fitzpatrick *et al*, 2017). In addition, its shape and environment are similar to the T1SS so it is a good test specimen for sample preparation and data processing. Finally, there are questions on MacAB-TolC function which an *in vivo* structure can help elucidate (see **Chapter V**). The aim of the chapter will be answered with the following objectives:

1) Test the "ghost cell" protocol for generating thin (<200 nm) bacterial samples expressing MacAB-TolC
2) Determine good cryo-ET data collection parameters for the sample
3) Find a good cryo-ET processing pipeline for the sample
4) Process the collected data and determine a 3D structure using available methods
5) Compare the *in vivo* structure to the cryo-EM structure (pdb ID 5nik) (Fitzpatrick *et al*, 2017) to see if it can answer relevant biological questions (*e.g.* flexibility of the NBDs)

2. Materials & Methods

2.1 Ghost Cell Production

A 1 L expression culture of BL21(DE3) cells was produced for MacAB-TolC (Fitzpatrick *et al*, 2017) or the T1SS (see **Chapter III** for the expression protocol). MacAB-TolC and T1SS cells were resuspended in 50 mL buffer (MacAB-TolC: 20 mM Tris (pH 8.0), 400 mM NaCl; T1SS: 50 mM Tris-HCl, 350 mM NaCl, 50 mM KCl, 5 mM $CaCl_2$, pH 7.5). Ghost cells were made by passing the 50 mL dense suspension twice through an Emulsiflex C5 (Avestin, Ottawa, Canada) at <1000 bar with protease inhibitor cocktail (Sigma-Aldrich, Germany). The suspension was spun at 37,500 g for 20 minutes and resuspended in 45 mL buffer (20 mM Tris-HCl, 100 mM NaCl, 50 mM KCl, 5mM $MgCl_2$, 10% glycerol) with 1 U/µl DNase I. The homogenised mixture was pelleted again at the same speed and time and resuspended in PBS (phosphate buffered saline) (pH 7.4) (with 1 mM $CaCl_2$ for T1SS) normalised to OD_{600} 0.5–0.7. Gold beads (10 nm) were added in a 1:5 ratio of beads to supernatant, 20 µl total volume, before plunge-freezing.

2.2 Plunge Freezing

Plunge freezing was performed using a VitroBot™ (ThermoFischer Scientific, USA) or GP2 (Leica Biosystems, Germany) as done previously in **Chapter III** (see **section 2.13**).

2.3 Dataset Collection

The MacAB-TolC ghost dataset was collected on the Krios III with a Falcon IV detector at the electron Bioimaging Centre (eBIC) at the Diamond Light Source (Harwell, UK). The tilt series was collected at x81,000 magnification which gives a pixel size of 1.5 Å/pixel under a dose symmetric scheme from -50° to +50° in 2° increments. The total dose was 120 $e^-/Å^2$ with a dose rate of 1.426 $e^-/Å^2/s$ and an exposure time of 1.65 seconds. For the T1SS ghost dataset, collection was done on the same microscope at the same magnification under a dose symmetric scheme from -60° to +60° in 3° increments. The total dose was 140 $e^-/Å^2$ with a dose rate of 7.58 $e^-/Å^2/s$ and an exposure time of 1 seconds. For the T1SS whole *E. coli*

dataset, a tilt series was collected at a magnification of x42,000 which gives a pixel size of 2.9 Å/pixel under a dose symmetric scheme from -60° to +60° in 3° increments. The total dose was 120 e^-/ Å^2 with a dose rate of 0.62e^-/ Å^2/s and an exposure time of 4.66 seconds.

2.4 Data Processing & Subtomogram Averaging

Initial motion correction and stack generation was completed using in-house scripts for MotionCor2 (Zheng *et al*, 2017) and etomo (Mastronarde & Held, 2017). For the MacAB-TolC ghost dataset, initial Warp (Tegunov & Cramer, 2019) processing of movie frames and CTF estimation was also trialled. For stack alignment and tomogram generation, IMOD (Mastronarde & Held, 2017) and emClarity 1.5.3.10 (Himes & Zhang, 2018) were trialled as well as Dynamo (Castaño-Díez *et al*, 2012) after Warp processing. For particle picking, emClarity 1.5.3.10 automated reference-based picking was trialled as well as the Dynamo membrane model. For initial model generation, both emClarity 1.5.3.10 and Dynamo were trialled.

For processing in EMAN2, stacks from the five best tomogram stacks identified in previous processing attempts were imported and tomograms reconstructed using their automated workflow. CTF estimation was completed before particles were picked using either manual boxing or neural network segmentation. Initial models were generated using either a template (neural network picking, pdb ID 5nik (Fitzpatrick *et al*, 2017) or a random reference (manual picking). Model refinement was then trialled using the best initial model.

3. Results

3.1 Ghost Cell Protocol for Producing Thin Specimens

Trials of the ghost cell protocol produced thin specimens with intact single and double membranes (**Figure 4.3.1-1**). From zero tilt views, side views of MacAB-TolC were visible. Tomographic dataset collection revealed the presence of top views at higher tilts in the tomograms (**Figure 4.3.1-2**). Due to the tendency of the ghost cells to cluster at the hole edges, the tilt range was restricted to -50°C to +50°C, otherwise the grid bar obstructed the view above this range. Beyond this range the grid bar prevents imaging at higher tilts. The data was further processed for subtomogram averaging.

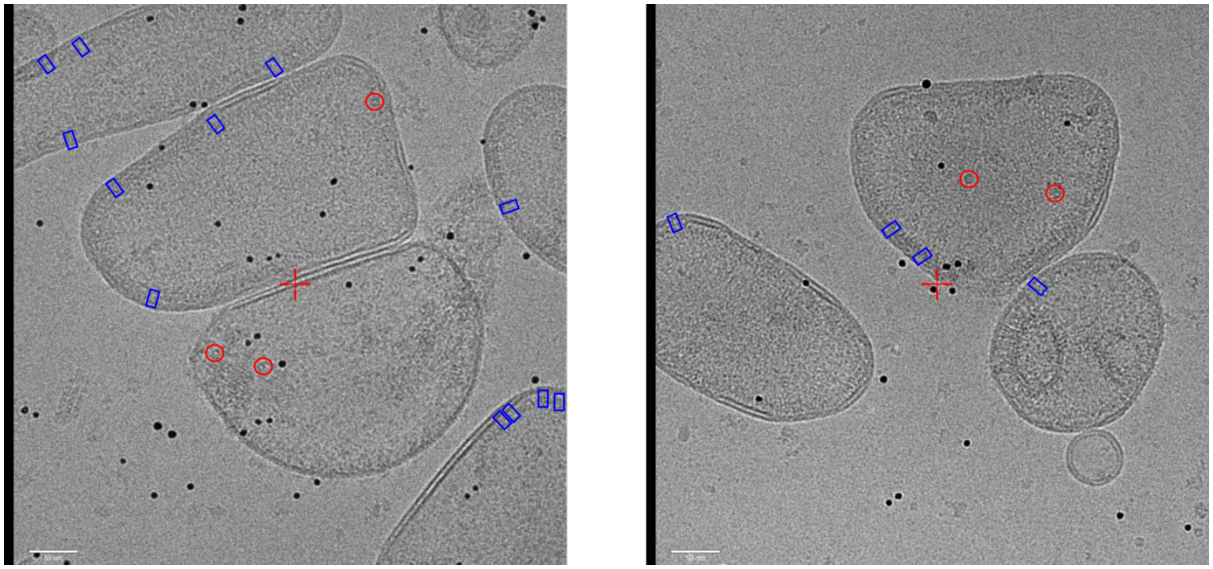


Figure 4.3.1-1. Initial screening of ghost samples. Initial screening results of ghost samples containing MacAB-TolC reveal the presence of single and double-membrane partially lysed cells. Both side (blue box) and top (red circle) views appear to be present. The dense dots are gold fiducials (10 nm).

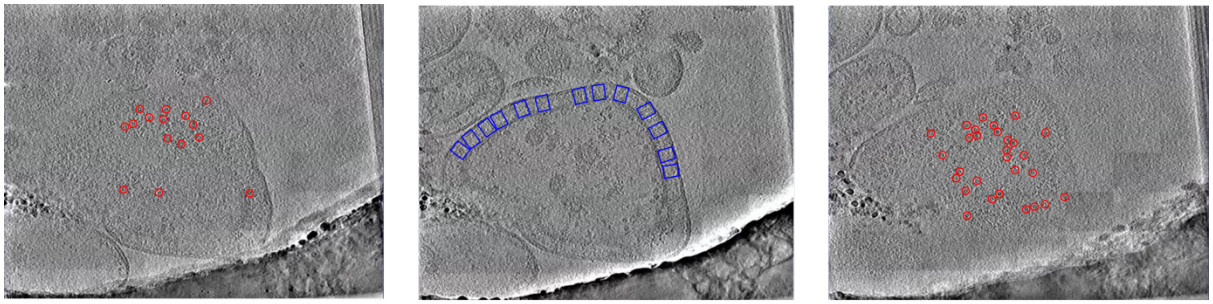


Figure 4.3.1-2. Different views of MacAB-TolC in the tomogram. In the untilted range, side views of MacAB-TolC are present (blue box) with top views becoming apparent at higher tilts (red circle). When tilted, an array-like formation of MacAB-TolC can be seen. Note that not all examples are highlighted.

In the raw data, the appearance of a lattice-like arrangement of MacAB-TolC pumps is apparent (**Figure 4.3.1-3a**). The lattice appears consistent with a diameter of approximately 150 Å and a between pump spacing of around 30 Å. The packing arrangement shows a series of trimeric triangles, consistent with a hexagonal lattice. Side views of the pump shows a between pump spacing that is consistent with a lattice (**Figure 4.3.1-3b**).

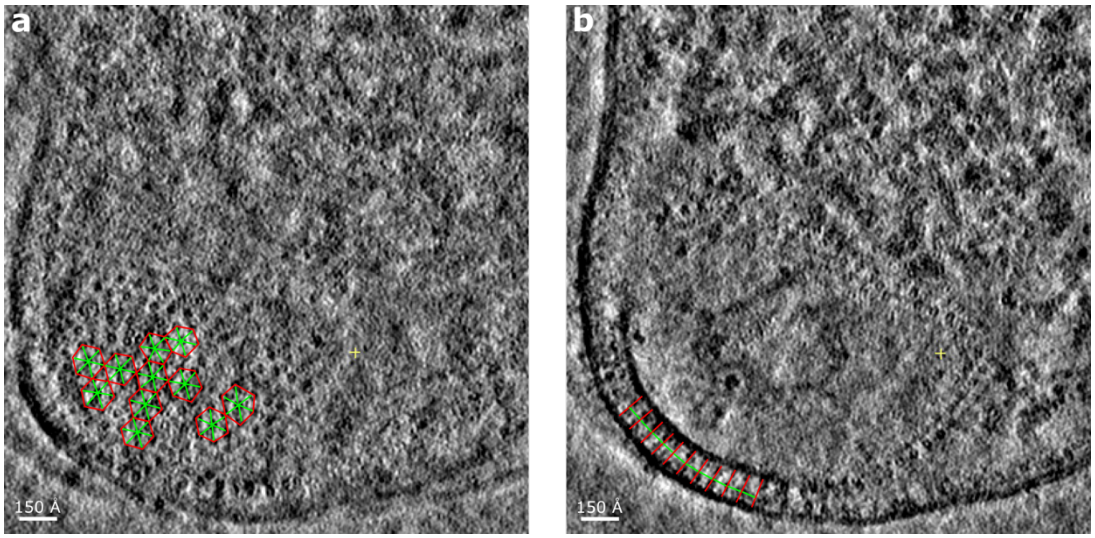


Figure 4.3.1-3. Lattice-like arrangement of MacAB-TolC. a) From the top views of the MacAB-TolC pump, a hexagonal lattice (red) is apparent with a central pump and six on each corner. The lattice is approximately 150 Å in diameter. **b)** Side views of the MacAB-TolC pumps shows a between pump spacing (green) of around 30 Å which corresponds to the lattice spacing seen in the top views.

3.2 Subtomogram Averaging of MacAB-TolC

The 2D images from the tilt series were imported into IMOD (Mastronarde & Held, 2017) for stack production after motion correction by MotionCor2 (Zheng *et al*, 2017). The five best stacks were imported into the EMAN2 tomography workflow (Chen *et al*, 2019). EMAN2 automatically aligns the tilt series for tomogram production. After this, CTF estimation was done. For particle picking, a neural net was trained on each tomogram and a segmentation produced (**Figure 4.3.2-1**). The neural network was checked for performance before being applied to tomogram segmentation (**Figure 4.3.2-1a**). Comparison of this segmentation to the original tomogram showed general isolation of density from the ghost cells, with minimal false positives from background parts of the tomogram (**Figure 4.3.2-1b**). Particles were extracted automatically from these segmentations (**Figure 4.3.2-1b**).

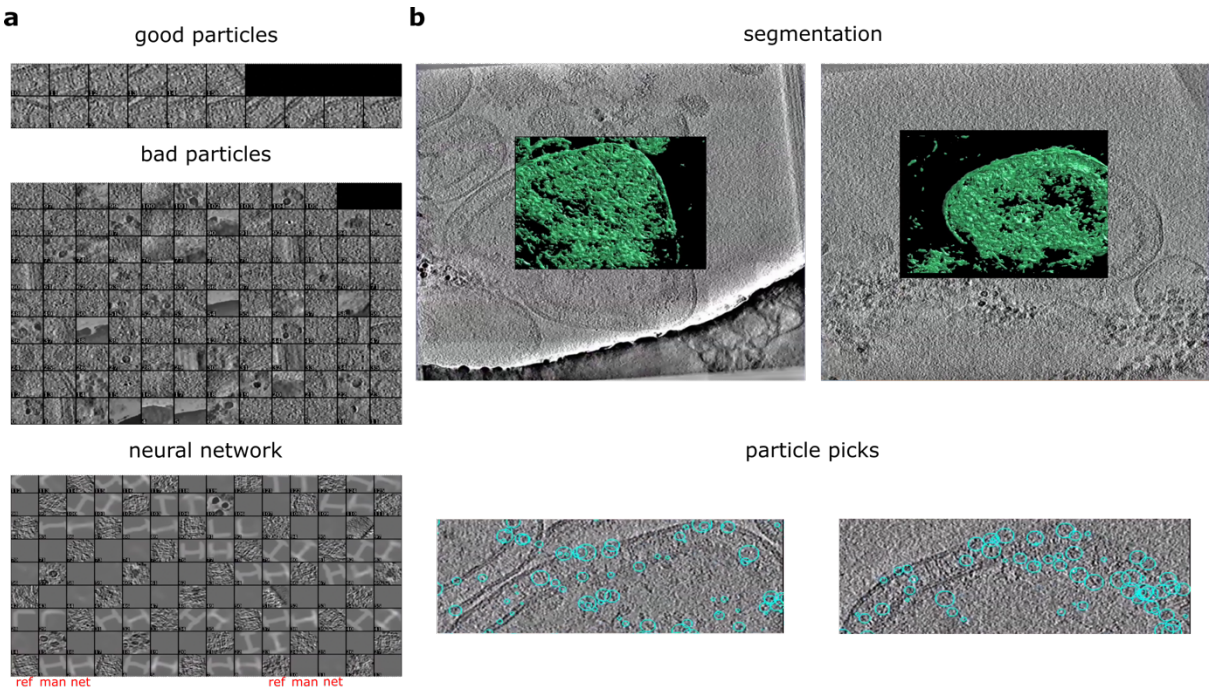


Figure 4.3.2-1. Training of neural network for particle picking. **a)** A neural network was trained to recognise MacAB-TolC in membrane by manually selecting reference sets of good particles (MacAB-TolC) and bad particles (carbon, background, gold beads). Training of the neural network was analysed by comparing the network annotation (net, red) to manual annotation (man, red) of the reference image (ref, red). If a network is well-trained the network should replicate the good particle manual annotation (left example, bottom) and leave bad particles unannotated (right example, bottom). **b)** The neural network was applied to the tomograms (above panel) to find densities corresponding to MacAB-TolC. From this segmentation, particles were extracted (bottom panel). Note that at this stage the programme does not know the orientation of the particles.

Since particle extraction is a very time-consuming and computationally intense task, a test tomogram was used for initial extraction and box size estimation. Approximately 8700 particles were extracted and an initial model produced using the cryo-EM structure as the template (pdb ID 5nik (Fitzpatrick *et al*, 2017) (**Figure 4.3.2-2**). The initial model showed the presence of a MacAB-TolC like density in the centre, with partial densities present which could correspond to the lattice arrangement (**Figure 4.3.2-2**).

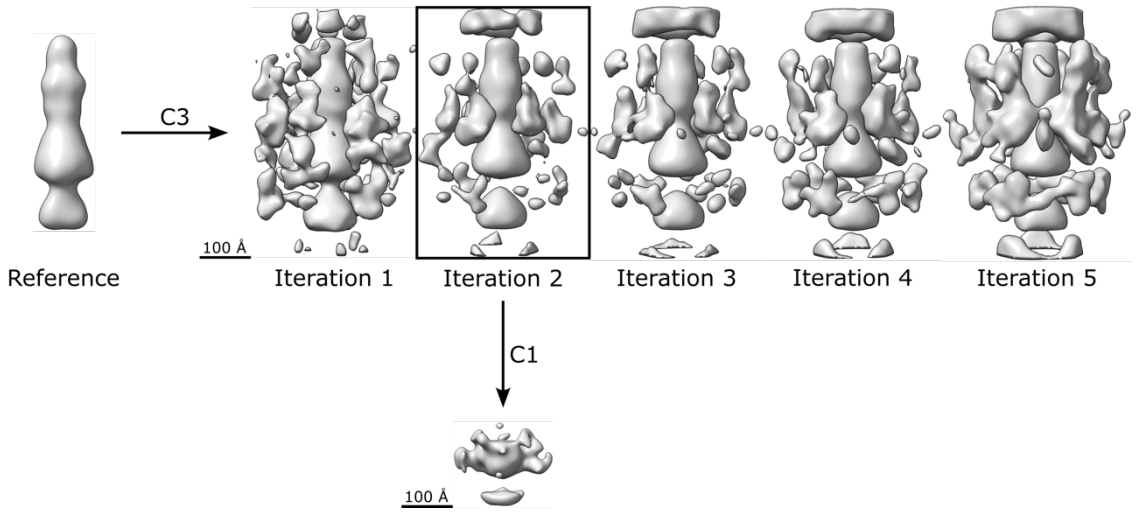


Figure 4.3.2-2. Initial Model Generation and Test Refinement for Neural Network Picked MacAB-TolC. An initial model of MacAB-TolC was generated using the cryo-EM single-particle map as a reference (pdb ID 5nik) (Fitzpatrick *et al*, 2017) with C3 symmetry applied. The second iteration was chosen to go into refinement. However, after one round only a single small density was isolated. This may indicate there are issues with the centering of the particles in the box.

A test refinement on these ~8700 particles was conducted which resulted in isolation of a single small density (**Figure 4.3.2-2**). This could indicate that there may be issue with the automatic picking, in particular the particles may not be centred well-enough to get good alignments for the refinement stage.

To test if automatic picking is a reliable strategy for picking out the pumps, ~4500 particles were manually picked from the five best tomograms and underwent initial model generation and refinement (**Figure 4.3.2-3**). To limit template bias, a random reference model was created and C6 symmetry was applied to try and tease out the MacA part of the structure. This resulted in the extraction of the full pump, with membrane either side and a putative

ring of peptidoglycan (**Figure 4.3.2-3**). The lattice-like arrangement was also apparent, with six partial densities around the central density. Since the hexagonal lattice is not strictly uniform, efforts at producing a refined model from the initial model failed. The use of a tight mask would likely improve refinement efforts.

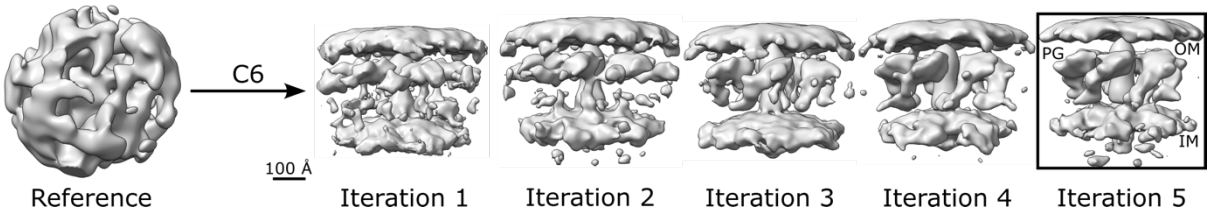


Figure 4.3.2-3. MacAB-TolC Initial Model Generation. An initial model for the MacAB-TolC pump was obtained using an initial random reference from the set of ~4500 manually picked particles. Application of C6 symmetry led to a defined density in the centre of the inner and outer membranes, consisting of TolC and MacA. After five rounds, a density siting below the inner membrane could also be seen, likely consisting of MacB. The structure from iteration 5 was used as the starting model for model refinement. PG: peptidoglycan, OM: outer membrane, IM: inner membrane.

3.3 Tomography Trials Type I Secretion System

Since promising results were obtained from cryo-correlative light microscopy (cryo-CLEM) of whole *E. coli* (see **Chapter III**), dataset collection was attempted after localising T1SS$^+$ cells on the grid; this was done by comparing fluorescence images to bright field and picking cells with dual red and green fluorescence that appeared over a hole (see **Figure 3.3.3.4-2** in **Chapter III**). Screening tests for "ghost" and intact (whole) *E. coli* expressing the T1SS were initially promising (see **Figure 4.3.3-1**). Tomographic datasets were collected for both ghost and whole *E. coli* samples. However, analysis of the whole *E. coli* stack after initial processing showed that the outer membrane was too thick for downstream subtomogram averaging or segmentation. The ghost cell dataset also showed no promising leads, with samples either being thin but containing no visible particles or too thick.

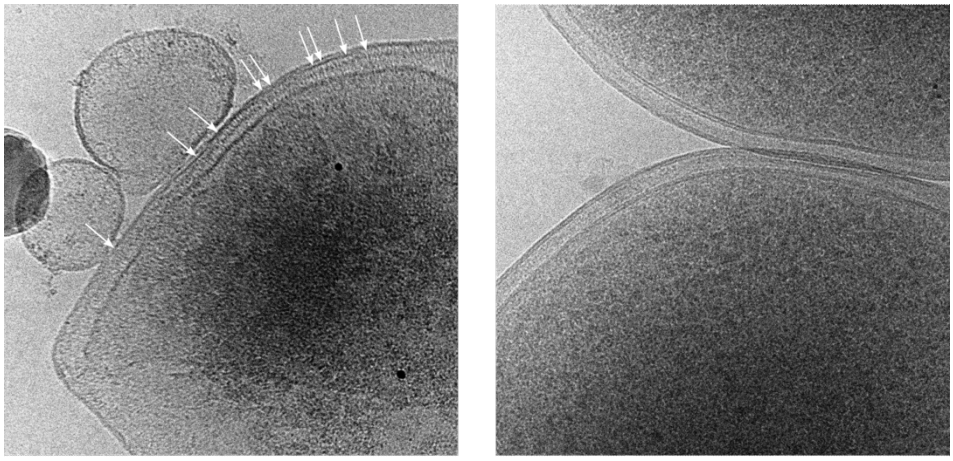


Figure 4.3.3-1. Screening samples for "ghost" and intact (whole) *E. coli* expressing the T1SS. White arrows show striations through the membrane of the ghosts which could be T1SS.

4. Discussion

4.1 Reconstruction Workflow for *In Situ* MacAB-TolC

I successfully reconstructed an initial *in situ* model of MacAB-TolC (see **Figure 4.3.2-3**). The *in situ* structure shows that MacAB-TolC could exist in a lattice-like arrangement, as shown in the raw images (see **Figure 4.3.1-3**); this is not entirely unexpected as previous research has established that outer-membrane proteins (OMPs) cluster in these arrangements (Webby *et al*, 2022). The use of the ghost protocol was very successful in sufficiently thinning the *E. coli* sample (<200 nm) to generate high-contrast images (see **Figures 4.3.1-1 and 4.3.1-2**). In addition, a lattice-like arrangement is clearly apparent and may have functional significance. The initial *in situ* structure also shows a structured density around TolC in addition to the densities corresponding to the membrane bilayer; this is likely a peptidoglycan layer which was previously found to interact with TolC in a tomography reconstruction of the multi-drug efflux pump, AcrAB-TolC (Shi *et al*, 2019). The initial model was produced from a starting random reference, and model convergence started to appear from iteration one (see **Figure 4.3.2-3**). Considering that there is no introduction of template bias in the initial model generation, and that the model has the presence of a ring of PG as seen in the related AcrAB-TolC drug efflux pump, is a good indication of the promise of the overall workflow.

Recent work in the Luisi group has established that the lipoprotein YbjP interacts with TolC. This was based on a single-particle cryo-EM structure of MacAB-TolC, where the TolC-YbjP interaction was maintained after extracting particles in peptidiscs (Kaplan *et al*, in preparation). The exact functional significance of the TolC-YbjP interaction is unknown, although it may aid pump to pump interactions if MacAB-TolC exists in a lattice-like arrangement in cells. Although the raw images (see **Figure 4.3.1-3**) and the initial models hinted at a possibility of this arrangement (see **Figures 4.3.2-2** and **4.3.2-3**), a refined model is needed. One difficulty is the particle alignments, where an irregular lattice-like arrangement would confuse alignments and result in a failed refinement. Another complication is the flexibility of the MacB nucleotide-binding domains (NBDs) (see **Chapter V**). This could be seen in the initial models, where the MacB density is the weakest (see

Figures 4.3.2-2 and **4.3.2-3**). One method for accounting for these complications is to separate the TolC-MacA and MacB parts of the structure with a mask and conduct separate refinements on each – this strategy worked for *in situ* AcrAB-TolC (Shi *et al*, 2019). Refinement of TolC-MacA would determine if the YbjP-TolC interaction occurs *in situ* and thus strengthen the previous finding from the single-particle structure. Furthermore, separate refinement of *in situ* MacB with 3D variability analysis would determine if NBD flexibility occurs in the cell and thus impact translocation. This is covered further in **Chapter V**. Placing the densities back into the tomogram after refinement would then allow us to see if a lattice-like arrangement is possible and if YbjP could be facilitating those pump-to-pump interactions.

Although neural network (NN) picking was able to pinpoint MacAB-TolC density in the ghost cells (see **Figure 4.3.2-1**), an initial model was very close to the input reference and refinement did not lead to an improved structure (see **Figure 4.3.2-2**). This is likely due to off-centering of the picks in the defined box; accurate alignments depend on the particles being relatively centred in the box (*e.g*, Heimowitz, Sharon & Singer, 2021). However, in contrast to attempts at using template matching, where significant false positive picks in the background and carbon were detected, the NN picks had a relatively low background false positive rate. Going forward, improvements in picking for this sample would utilise more than one NN to home in on the region of interest (Chen *et al*, 2017), and attempt to separate the pump from the membrane so that a more "centred" region is defined. The need for manual picking for downstream model generation is not unusual considering the complexity of the sample; this strategy was successfully employed for tomographic reconstruction of AcrAB-TolC (Shi *et al*, 2019). Bias may be introduced when particles are manually picked, however it should be stressed that the risks are the same as for a picking "clean-up" operation that is essential to remove false positives after NN or template-matching picking. Template matching relies on a good signal-to-noise ratio for the programme to cross-correlate template to tomogram positions; the template is rotated and cross-correlations with the tomogram density computed to find matches (Böhm *et al*, 2000). Since membrane has a high signal and a continuous structure that is indistinguishable from the signal generated by a "side view" of a pump-like object, high rates of membrane picks are often observed when template picking is employed, in addition to other false positive signals such as background. This entails a

significant clean-up step to manually remove probable false positives. If this is not done, noise will dominate the downstream reconstructions.

4.2 Biological Relevance of *In Situ* MacAB-TolC

Reconstruction by subtomogram averaging showed that TolC and MacA are relatively static structures *in situ*, with densities matching that of the cryo-EM structure (Fitzpatrick *et al*, 2017). The structure for MacB was less well-defined (see **Figures 4.3.2-2** and **4.3.2-3**). This corresponds to previous single-particle cryo-EM reconstructions for MacB, where attempts resulted in an averaging out of MacB in the full complex model (Fitzpatrick *et al*, 2017). Isolation of MacB density showed a structure in an "open" conformation, and the presence of an unknown density between the MacB protomers. Thus, the orientation of MacB NBDs *in situ* is unknown and its flexibility may have a role in substrate translocation. This is covered further in **Chapter V.** The biological role of the TolC-YbjP interaction is not known, although its absence does not seem to affect the ability of the pump to extrude antibiotics (Kaplan *et al*, in preparation). The interaction may facilitate pump-to-pump interactions in the cell and thus aid in the formation of exporter "islands" in the cellular membrane; lipids have been known to mediate these OMP interactions in the cell and hinder the permeability of antibiotics (Webby *et al*, 2022). Thus the TolC-YbjP interaction could aid in the organisation of the pump in the cell with respect to other membrane transporters. As discussed above, a refined model of TolC-MacA is needed to see if the TolC-YbjP interaction occurs *in vivo*.

Interestingly, a peptidoglycan layer was found hemming TolC, as was the case for the sub-tomogram reconstruction of AcrAB-TolC (Shi *et al*, 2019). This suggests that the peptidoglycan layer may be a general feature of TolC interactions with its myriad periplasmic adapter protein partners (Horiyama, Yamaguchi & Nishino, 2010). The layer may act as an extra stabilising force, facilitating periplasmic adapter protein interactions with TolC. In general, peptidoglycan interactions tend to be lost during the protein purification process required for single-particle EM structural studies. In addition, its unstructured nature means that any density tends to be averaged out during image processing and model generation (Meroueh *et al*, 2006). Thus, my sub-tomogram averaged initial *in situ* structure reveals a rare glimpse of protein-peptidoglycan interactions. As predicted, the peptidoglycan layer forms a ring-like

structure around the pump (see **Figure 4.3.2-3**) (Meroueh *et al*, 2006). However, as also found by sub-tomogram averaging of AcrAB-TolC, my structure shows that the peptidoglycan ring is located in the TolC-MacA interface rather than the TolC equatorial domain (Shi *et al*, 2019; Gumbart *et al*, 2021). In addition, it is known that the peptidoglycan layer is anchored to the outer membrane via Braun's lipoprotein (Lpp); a recent molecular dynamics study has posited that it sits perpendicular to the membrane, providing an anchor between OM and peptidoglycan layer (Gumbart *et al*, 2021). My structure shows a corresponding space (~90 Å) between the OM and peptidoglycan layer that could fit perpendicular Lpp, but no corresponding density, likely as a result of averaging out during reconstruction (see **Figure 4.3.2-3**). Hence, any presence of Lpp is likely to be unordered, in contrast to the peptidoglycan layer which forms an ordered ring. It is known that TolC forms transient tip-to-tip interactions with its partner periplasmic membrane proteins (Horiyama, Yamaguchi & Nishino, 2010; Xu *et al*, 2011; Fitzpatrick *et al*, 2017); the relative weakness of this interface interaction compared to interactions between the periplasmic adapter protein and inner membrane protein is likely a feature that allows TolC to interact with many different periplasmic protein partners. Hence, an additional stabilising force via a structured ring-like peptidoglycan layer may explain the relative static structure of TolC and MacA compared to inner membrane protein MacB.

4.3 Type I Secretion System Tomography Trials

Sample preparation of the T1SS for tomography did not yield a suitable sample for downstream processing. Initial trials of whole *E. coli* expressing T1SS did not yield a thin-enough (<200 nm) sample, while ghost cell preparation resulted in a lack of obvious pump-like striations through the membrane. One difficulty is the lack of T1SS per *E. coli;* estimates of T1SS numbers based on flow cytometry (see **Chapter III**) places a range of 10,000 to 15,000 per cell. This contrasts with a packed membrane, where 40,000 can be accommodated. Thus, T1SS levels could be 3–4x less than that of MacAB-TolC, which forms a packed arrangement in the ghost cells (see **Figure 4.3.1-2**). Another difficulty is the transient arrangement of the T1SS; a stable complex is made *in vivo* via a fast-folding eGFP which prevents full translocation of the HlyA substrate (see **Chapter III**). Disturbance of the complex via purification trials or ghost production could dissociate the complex. Hence, future trials would need to focus on

producing a cross-linked complex which maintains the HlyD-HlyB interaction, as was done to produce a full MacAB-TolC complex (Fitzpatrick *et al*, 2017). Thus, with a cross-link in place to stabilise the complex, ghost production is more likely to be successful. Since the T1SS is very similar in topology and location to MacAB-TolC, downstream processing by EMAN2 is the most likely to be successful.

5. Conclusion

I found that the “ghost” cell protocol produces “thin enough” (<200 nm) bacterial samples for cryo-ET sub-tomogram reconstruction whilst maintaining *in vivo* interactions. This was evidenced by the initial reconstruction of MacAB-TolC, where the inner and outer membranes were present in addition to a ring of peptidoglycan. The initial reconstruction also suggested a lattice-like arrangement of the pumps is possible. Further work to produce a refined structure is needed, although the *in situ* organisation of the pump would entail separate TolC-MacA and MacB refinements. I also tested strategies for cryo-ET data collection of the T1SS; efforts were hampered by the thickness of the whole *E. coli* even in the outer membrane region, whilst efforts to produce “ghost” cells resulted in no visible particles, likely due to the dissociation of the complex during preparation. Future work would focus on chemical cross-linking of the HlyB-HlyD interaction to ensure the complex remains intact during sample preparation. My work has established a successful cryo-ET sub-tomogram workflow, from sample preparation to initial 3D reconstruction of membrane pumps.

Chapter V: Molecular Dynamics to Investigate Transport Mechanisms in the MacAB-TolC Efflux Pump

1. Introduction

1.1 Role of Efflux Pumps in Antibiotic Resistance

Pathogenic bacteria have evolved a range of defences against antibiotics since their introduction in the 1950's (Davies & Davies, 2010). In 2019, 4.95 million deaths were estimated to be associated with antibiotic resistant infections (Antimicrobial Resistance Collaborators, 2022). Indeed, the problem is so severe that the 2016 Review on Antimicrobial Resistance noted that:

"*The magnitude of the problem is now accepted. We estimate that by 2050, 10 million lives a year and a cumulative 100 trillion USD of economic output are at risk due to the rise of drug-resistant infections if we do not find proactive solutions now to slow down the rise of drug resistance*." (O'Neill, 2016)

Thus, it has become imperative to study mechanisms of resistance to antibiotics in order to design effective treatments in the future (Richardson, 2017).

Efflux pumps are one of the molecular mechanisms conferring antibiotic resistance in Gram-negative bacteria (Darby *et al*, 2022). Their function is to form a membrane-spanning channel in order to pump out the antibiotic and protect the cell (Ebbensgaard, Løbner-Olesen & Frimodt-Møller, 2020).

One such efflux pump is the MacA-MacB-TolC (MacAB-TolC) complex, consisting of ATP-binding cassette (ABC)-transporter and inner membrane protein MacB, periplasmic adapter protein MacA, and outer-membrane protein TolC (**Figure 5.1.1-1a**). The complex was first identified based on its ability to export the macrolide antibiotic erythromycin in *Escherichia coli* (Kobayashi, Nishino & Yamaguchi, 2001). Subsequent studies have found that it promotes

virulence in *Salmonella* infections in mice and that its expression is regulated by two-component sensing system PhoP/PhoQ (Nishino, Latifi & Groisman, 2006), which suggests that its main function is to export antimicrobial peptides (Greene *et al*, 2018). Analysis of sub-Saharan African strains of *Salmonella enterica* from clinical isolates have shown that mutations in *macA* and *macB* genes are associated with its evolution as a causative agent of gastrointestinal disease in human hosts to an invasive blood-borne one (Honeycutt *et al*, 2020). This evolution is associated with its ability to export antimicrobial peptides: *macA* and *macB* genes in gastrointestinal *Salmonella* provide resistance to antimicrobial peptides under the regulation of PhoP, while stop-codon mutations in *macB* render it non-functional in invasive blood-borne *Salmonella* strains. This suggests the antimicrobial peptide activity of the MacAB-TolC pump is specifically associated with gut-colonising *Salmonella* strains (Honeycutt *et al*, 2020). The MacAB-TolC pump has also been shown to contribute to biofilm development in *Acinetobacter baumannii*, an opportunistic bacterium associated with hospital-acquired infections (Robin *et al*, 2022).

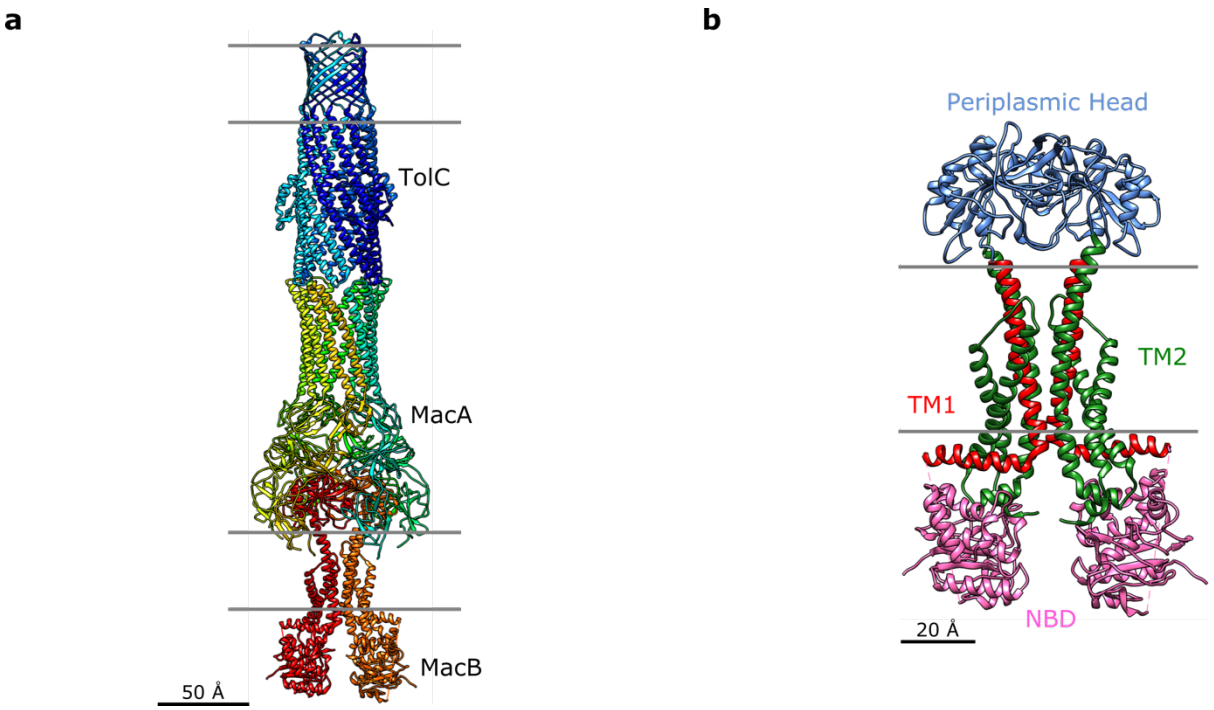


Figure 5.1.1-1. Overview of the MacAB-TolC Complex. **a)** MacAB-TolC complex consists of dimeric MacB coupled to hexameric MacA which in turn interacts with trimeric outer membrane protein TolC. MacA and TolC interact in a tip-to-tip fashion to form a tunnel, with a guard helix above MacB controlling substrate exit (Xu *et al*, 2011; Fitzpatrick *et al*, 2017). **b)** Structural composition of MacB with nucleotide-binding domain (NBD, pink) (residues 1–246) coupled to transmembrane domain 1 (residues 247–305, red). The periplasmic domain (residues 306–506, blue) is then recoupled to the NBD via transmembrane domain 2 (residues 507–648, green). Note the approximate location of the lipid bilayer is shown (grey). Structures shown are from pdb ID 5nik (Fitzpatrick *et al*, 2017).

1.2 Structure and Function of MacAB-TolC

To form a functional pump, the MacAB-TolC complex interacts in a ratio of 6:2:3: MacB interacts with MacA via its periplasmic head region while MacA forms a tip-to-tip interaction with TolC (Xu *et al*, 2011; Fitzpatrick *et al*, 2017) (**Figure 5.1.1-1a**). This tip-to-tip interaction functions as a cogwheel, allowing for substrate transport through the otherwise closed channel (Xu *et al*, 2011), in a similar manner to transport in the resistance to nodulation family (RND) efflux pump AcrA-AcrB-TolC (Wang *et al*, 2017). TolC is a multi-functional outer-membrane protein, associating with many different periplasmic adapter proteins to extrude various substrates (Horiyama, Yamaguchi & Nishino, 2010).

The ABC-transporter inner-membrane component MacB has been classed in the type VII family of transporters (Thomas & Tampé, 2020). Type VII has certain structural features that make it unique from other types of transporters: the presence of a large periplasmic head domain, a thin transmembrane stalk region, and coupling helices to the nucleotide-binding domains (NBDs) (**Figure 5.1.1-1b**). MacB is also unusual from other transporters in that its macrolide antibiotic and polypeptide virulence factor substrates are exported from the periplasm rather than the cytoplasm (Yamanaka *et al*, 2008; Thomas & Tampé, 2020). Like other ABC-transporters, MacB provides the energy for transport via ATP-hydrolysis (Tikhonova *et al*, 2007). ATP binding is stimulated by the presence of MacA via the periplasmic head region of MacB which also aids to recruit TolC during substrate transport (Lu & Zgurskaya, 2012). Furthermore, the presence of substrate does not seem to influence ATP-activity in the MacAB complex (Tikhonova *et al*, 2007; Souabni *et al*, 2021).

1.3 Routes for Substrate Entry and Mechanotransmission

The MacB periplasmic domain contains a port opening to allow for substrate entry, although the exact substrate-binding mechanism remains elusive (Yamanaka *et al*, 2008; Crow *et al*, 2017; Fitzpatrick *et al*, 2017). Comparison of available MacB structures led to the proposal of the "molecular bellows" mechanism of transport by Crow *et al* (2017). In this model nucleotide-free MacB shows an "open" periplasmic port with an undimerised NBD, while ATP-bound MacB shows a "closed" periplasmic port with a dimerised NBD. Substrate export is thought to follow a mechanotransmission mechanism whereby the substrate is squeezed out of the periplasmic cavity after ATP-binding (Crow *et al*, 2017). A crystal structure of MacB from *Acinetobacter baumannii* with bound ADP analogue adenosine-5'-(β-thio)-diphosphate (ADPβS) showed a dimerised NBD with an open periplasmic head domain, suggesting that this is indeed a post-substrate transport state (Okada *et al*, 2017).

Comparison of MacB to more recent MacB-like structures offers some further clues to mechanisms of transport. Haem-detoxification in Gram-positive bacteria is driven by HrtBA; characterisation of this transporter in *Corynebacterium diptheriae*, the causative agent of diptheria, established a mechanism of transport based on comparisons of unliganded, haem-bound, and nucleotide-bound states (Nakamura *et al*, 2022). Unlike MacB, the functional unit

of HrtBA has a separate NBD (comprised of HrtA) from the transmembrane and periplasmic domains (comprised of HrtB). In terms of sequence similarity, *C. diptheriae* HrtA shows ~40% sequence similarity to *E.coli* MacB NBD, while HrtB shows ~19% sequence similarity to *E. coli* MacB transmembrane and periplasmic domains. Such a separation of NBD from transmembrane and periplasmic domains is also found in a MacB-like structure from *Streptococcus pneumoniae* (Yang *et al*, 2018) and also lipoprotein transporter LolCDE (Kaplan *et al*, 2018). HrtB extracts haem from the membrane, squeezing it between its protomers before extruding it by ATP-hydrolysis from HrtA. In the unliganded state, HrtB periplasmic binding site is accessible to haem, however, upon binding of a non-hydrolysable ATP-analogue (adenylyl-imidodiphosphate, AMPPNP), dimerization of HrtA causes this site to become closed. This led to the proposal of a mechanism of transport in which haem binds HrtB in the "open" unliganded state and ATP-binding induced HrtA dimerisation causes haem extrusion from HrtB. ATP hydrolysis then resets the system to the "open" state. Since basal ATP-activity is possible without substrate transport, ATP-binding induced HrtA dimerisation without bound substrate can lead to futile hydrolysis cycles (Nakamura *et al*, 2022).

The structures from an efflux pump from Gram-positive *Streptococcus pneumoniae* was also solved (Yang *et al*, 2018). Like HrtBA, the structure has a separable NBD consisting of Spr0694 which interacts with Spr0695 to from the full inner-membrane component. These then interact with Spr0693 to form a full efflux pump. Although the sequence similarity to *E.coli* MacAB is low (21% sequence similarity Spr0693 to *E. coli* MacA, 37% sequence similarity Spr0694-Spr0695 to *E. coli* MacB), there are striking structural similarities. Like MacB and HrtBA, the functional form of Spr0694-Spr0695 is a dimer, with the periplasmic head domain of Spr0695 containing a guard helix for substrate access. Like MacA, Spr0693 is hexameric, associating with Spr0694-Spr0695 to allow substrate extrusion via its hollow tunnel-like pore. As in MacAB, ATP-activity of Spr0694-Spr0695 is augmented by the presence of Spr0693 and ATP-binding induces Spr0694 dimerisation which is required for hydrolysis (Yang *et al*, 2018).

LolCDE is a complex involved in the transport of lipoproteins from the inner to the outer bacterial membranes of Gram-negative species. Like HrtBA and Spr0694-Spr0695, the NBDs are a separable component from the transmembrane and periplasmic domains: LolC and LolE associate to form the heterodimeric transmembrane/periplasmic domains, while LolD acts as

the dimeric NBD. Together, they capture lipoproteins from the inner membrane to load into carrier protein LolA, which moves through the periplasm to offload the cargo to LolB (Matsuyama, Tajima & Tokuda, 1995; Matsuyama, Yokota & Tokuda, 1997; Yakushi *et al*, 2000; Kaplan *et al*, 2018). Structural analysis of the complex has yielded some insights into the transport mechanism. The periplasmic head of LolC has a similar fold to MacB periplasmic domain and can bind lipoprotein carrier LolA; this interaction is independent of ATP binding or hydrolysis (Kaplan *et al*, 2018). Cryo-EM structures of LolCDE established capture of a lipoprotein sandwiched between the two LolC and LolE transmembrane structures, as was found for HrtBA (Sharma *et al*, 2021). Like HrtBA and Spr0694-Spr0695, LolD dimerisation is induced by ATP-binding, and this leads to a closed LolCDE complex. Hence, as in HrtBA it is postulated that ATP-binding can only occur after lipoprotein binding, after which the dimerisation of LolD induces lipoprotein extrusion to LolA carrier while ATP-hydrolysis then resets the structure (Sharma *et al*, 2021). Further characterisation of LolCDE complex in apo-form, nucleotide-bound, lipoprotein-bound, and LolA-bound showed a diversity of conformations. As found previously, the presence of ATP-analogue AMPPNP induced dimerisation of LolD. However, two states were found: one with lipoprotein bound and one without. This led to the proposal that ATP-hydrolysis provides the power-stroke for lipoprotein export, with the dissociation of post-hydrolysis products ADP+P_i providing the reset mechanism (Tang *et al*, 2021). Intriguingly, an apo-LolCDE form was captured which showed a closed conformation, with a similar structure to lipoprotein-bound LolCDE (Bei *et al*, 2022). This indicates the highly dynamic nature of the transportation process, with available structures providing “snapshots” in time.

1.4 Rationale for This Study

Although available evidence offers clues as to the dynamics of transport in the MacAB-TolC system, there is a lack of functional insight particularly with regard to protein-lipid interactions. Molecular dynamics simulations have shown to be particularly useful in understanding functional dynamism in protein structures and was successfully applied in the Type I Secretion System (T1SS) to yield insights into its transportation process (see **Chapter II**). I applied a similar methodology developed for the T1SS protein-lipid interactions to study these interactions in MacB. Isolation of MacB from the full MacAB-TolC complex for the

simulations was done to make the study more tractable considering the size of the system (Jung *et al*, 2020) and the lack of accurate force-fields for peptidoglycans (Tschampel, Kirschner & Woods, 2006). Nevertheless, considering that functional insights could be obtained from structure snapshots of isolated MacB and MacB-like structures, it is expected that this will also apply in simulation.

As done for HlyB-HlyA coarse grain system (see **Chapter II**), multiple replica trajectories would need to be performed to generate robust statistics (Wan, Sinclair & Coveney, 2021). During equilibration of the first trajectory, POPE lipids were observed to enter the MacB dimer in the upper leaflet, remaining wedged between the protomers during production. At first, I thought this was an artefact of the simulation conditions. However, there is experimental evidence that MacB has internal PE binding sites. It is not clear if PE is a substrate or if it remains bound while other substrates are exported (Barrera *et al*, 2009). Given the amphiphilic nature of phospholipids, studying dynamics of PE transport would offer clues to the transport of antibiotic substrates, considering that they can be either hydrophobic or hydrophilic. This offered an opportunity to simulate MacB under two conditions: 1) with bound POPE, and 2) without bound POPE. Combined with repeat simulations at a sufficient length of simulation time, this would provide an *in silico* experiment for uncovering MacB transport dynamics.

The Aims of this Chapter involves answering the following questions:

1) How does the conformational dynamics of MacB change throughout the simulation? Are the dynamics similar across trajectories?
2) How do potential MacB substrates access the periplasmic port/cavity? What are the dynamics affecting access?
3) Is the lipid environment important for substrate access?
4) Do the observed dynamics fit with the experimental evidence and current model of substrate transport?

These Aims will be addressed with the following Objectives:

1) Prepare the MacB model using available structures
2) Prepare the system(s) by placing the MacB model in simplified membrane
3) Conduct MD simulations with and without internal POPE lipid
4) Develop an analysis pipeline to answer the questions of the Aims

2. Methodology

2.1 Overview of the Simulation Pipeline

The simulation and analysis pipeline were the same as **Chapter II**. Data analysis pipelines are available on GitHub (see **Appendix B**). An overview of the simulations conducted is given in **Table 5.2.1-1**.

Table 5.2.1-1. Overview of simulations conducted for this study

Name	Temperature (K)	# Repeats	Atomistic/Coarse Grain?	Simulation Length (ns)
MacB Internal POPE Lipid (u1–u5)	303.15	4	Coarse Grain	2160
MacB No Internal POPE Lipid (u6–u10)	303.15	4	Coarse Grain	2160

2.2 Structure Preparation

MacB was obtained from the deposited cryo-electron microscopy structure of the MacA-MacB-TolC complex (pdb ID 5nik) (Fitzpatrick *et al*, 2017). The missing loop region G246–N224 in protomers A and B was modelled using MODELLER 9.23 (Sali & Blundell, 1993). The best model was chosen based on DOPE score (Shen & Sali, 2006) and visual inspection of the output structure. Side chain rotamers were then refined using Scwrl 4 (Krivov et al, 2009).

2.3 Molecular Dynamics Simulations

Coarse grain molecular dynamics simulations were conducted as described in **Chapter II** (see **Section 2.2.2**). A membrane of system size 140 $\mathrm{\AA}$ was built using 75% POPE, 25% POPG, and 5% cardiolipin (Raetz & Downhan, 1990) in a 1:1 ratio between top and bottom leaflet.

For repeats, MacB was equilibrated at 310.15 K before being cooled to 303.15 K to reset starting velocities. Repeats 1–2 (trajectories u2–u3/u6–u7) were then run from this starting structure. For repeats 3–4 (trajectories u4–u5/u9–u10), a random starting structure was generated from the last 2 ns of equilibration time.

2.4 Analysis Pipeline

The same analysis pipeline was used as described in **Chapter II** (see **Section 2.2.3**). The code developed for the final analyses is available on GitHub (see **Table B** in **Appendix B**). For cavity estimation, at each time step mean distances were calculated between protomer A and protomer B periplasmic hotspot residues (T349, Y376, F444, and W505) (Crow *et al*, 2017) using the Distances method in MDAnalysis (Michaud-Agrawal *et al*, 2011; Gowers *et al*, 2016). Analysis codes I developed can be found on GitHub (https://github.com/AMKCam) (see **Table B** in **Appendix B**).

3. Results

3.1 MacB Conformational Changes

As with the HlyB MD simulations (see **Chapter II**), energy terms were checked after each equilibration and production runs. Membrane leakage was checked by lipid density analysis and manual inspection. No obvious issues were encountered with the trajectories.

After minimisation, the two halves of MacB adopted a more open conformation, driven by the NBDs. This was maintained in equilibration due to the structural restraint conditions. However, in production when the restraints were removed it was observed that the MacB dimer converted to a closed conformation within 40 ns of simulation time (**Figure 5.3.1-1**). After the NBDs adopted the closed conformation, the periplasmic heads altered conformation such that no substrate access was possible by the end of the simulation. This was an interesting observation considering that the closure of the MacB NBD in the transportation process is supposed to be driven by energy changes from ATP-binding (Crow *et al*, 2017; Fitzpatrick *et al*, 2017). Repeat simulations also showed the same dimeric closure, in both cases where the starting structure was the same (internal lipid trajectories u2 – u3 and no internal lipid trajectories u7 – u8), and the case where the starting structure was randomly selected from the last 2 ns of equilibration (internal lipid trajectories u4 – u5 and no internal lipid trajectories u9 – u10). The system was equilibrated to 37^{o}C and cooled back down to 30^{o}C before the repeat runs, and hence the lipid configuration is also not identical between the repeat trajectories and the initial trajectory.

The simulation results suggest that this closure occurs under normal equilibrium dynamics of the molecule, without any energy from ATP binding or hydrolysis required. In addition, it was observed during initial equilibration of MacB that two POPE residues entered the stalk region separating the two dimeric halves, forming a wedge and opening up the periplasmic head region. These POPE lipids appeared to be drawn up from the membrane to the periplasmic head cavity, possibly indicating a substrate access channel (**Figure 5.3.1-1**). Internal POPE binding sites have been proposed experimentally (Barrera *et al*, 2009). This presented an

opportunity to observe dynamics under two conditions: one with and one without internal POPE residues. Given the diversity of MacB substrates, particularly its ability to excrete a variety of antibiotic substrates, this offered an opportunity to study the dynamics of transport; comparison of the two simulations would provide useful insights into the mechanism of the transportation process.

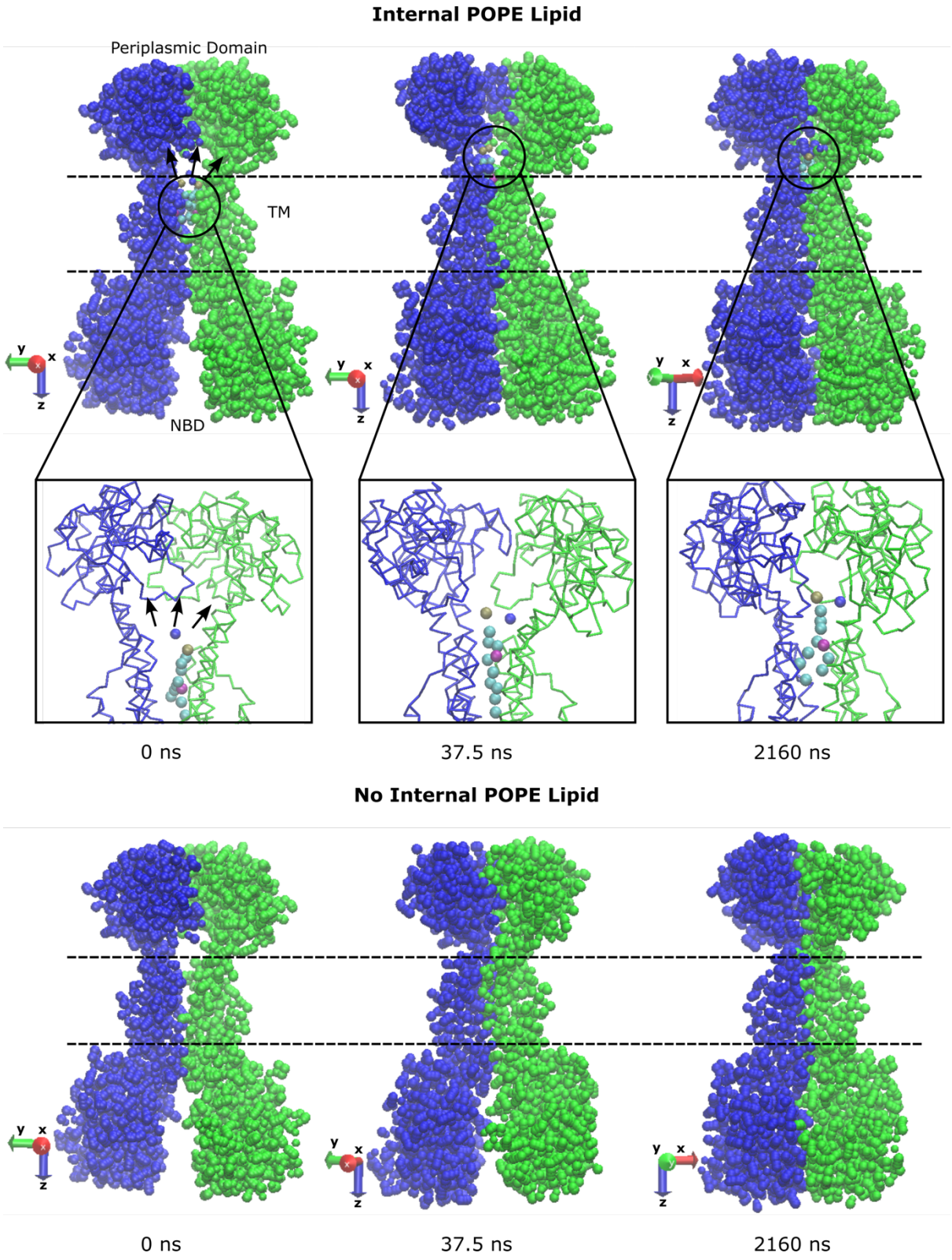


Figure 5.3.1-1. MacB closure of dimeric halves. At the beginning of the initial simulations, dimeric halves (chain A, green; chain B, blue) are in an open state after minimisation and equilibration. After around 40 ns simulation time, the two halves form a closed conformation which remains to the end of the simulation. Note the presence of internal POPE lipids (black, circle) which act as a wedge between the two protomers. Independent closure was observed when the internal POPE residues were deleted. Without the internal POPE residues, the

periplasmic heads appear to close before the NBDs. Approximate location of the lipid bilayer is shown (black line, dotted). Note that the POPE lipid appears to be drawn up to the MacB periplasmic head cavity, which may indicate a substrate access channel. A close-up view of the internal POPE lipid is shown (black box) at each time point. Lipid movement was seen in repeat simulations (u1– u4) in both cases where the starting structure was the same (u1 – u2) and where they were different (u3 – u4).

It should be noted that the MacB protein was simulated using an elastic network model. This model is used to study protein dynamics around a reference structure, with the protein atomic structure represented as point-like objects and their interactions by springs obeying Hooke's Law. This reduces the resultant protein conformational space and hence makes it possible to study motions over a longer timescale. This makes it particularly useful for studying a range of processes, including protein-ligand interactions (Togashi & Flechsig, 2018).

3.2 RMSD, RMSF, PCA, and Ensemble Similarity

As with the HlyB-HlyA MD simulation study (see **Chapter II**), MacB simulations were checked by radius of gyration, root mean square deviation (RMSD), root mean square fluctuation (RMSF), Principle Component Analysis (PCA), and ensemble similarity of the backbone atoms. Two initial trajectories were performed (one for internal POPE lipid and one without internal POPE lipid) and the outputs examined every 1080 ns until the desired length of run-time. Repeat trajectories were then run according to this desired length of time. Example outputs are given in **Figure 5.3.2-1**. There was no difference in the radius of gyration between trajectories – each trajectory showed either a decrease or a stabilisation, indicating that the structure is stable throughout the simulation (**Figure 5.3.2-1a**). For RMSD, the only differences observed was in chain (protomer) A domains (**Figure 5.3.2-1c–d**), while chain (protomer) B showed no differences between domains (**Figure 5.3.2-1e**). RMSF plots were variable between simulations, with no obvious pattern between internal and no internal lipid trajectories (**Figure 5.3.2-1f**). PCA indicated that 2160 ns simulation time was sufficient for the study (**Figure 5.3.2-1g–h**).

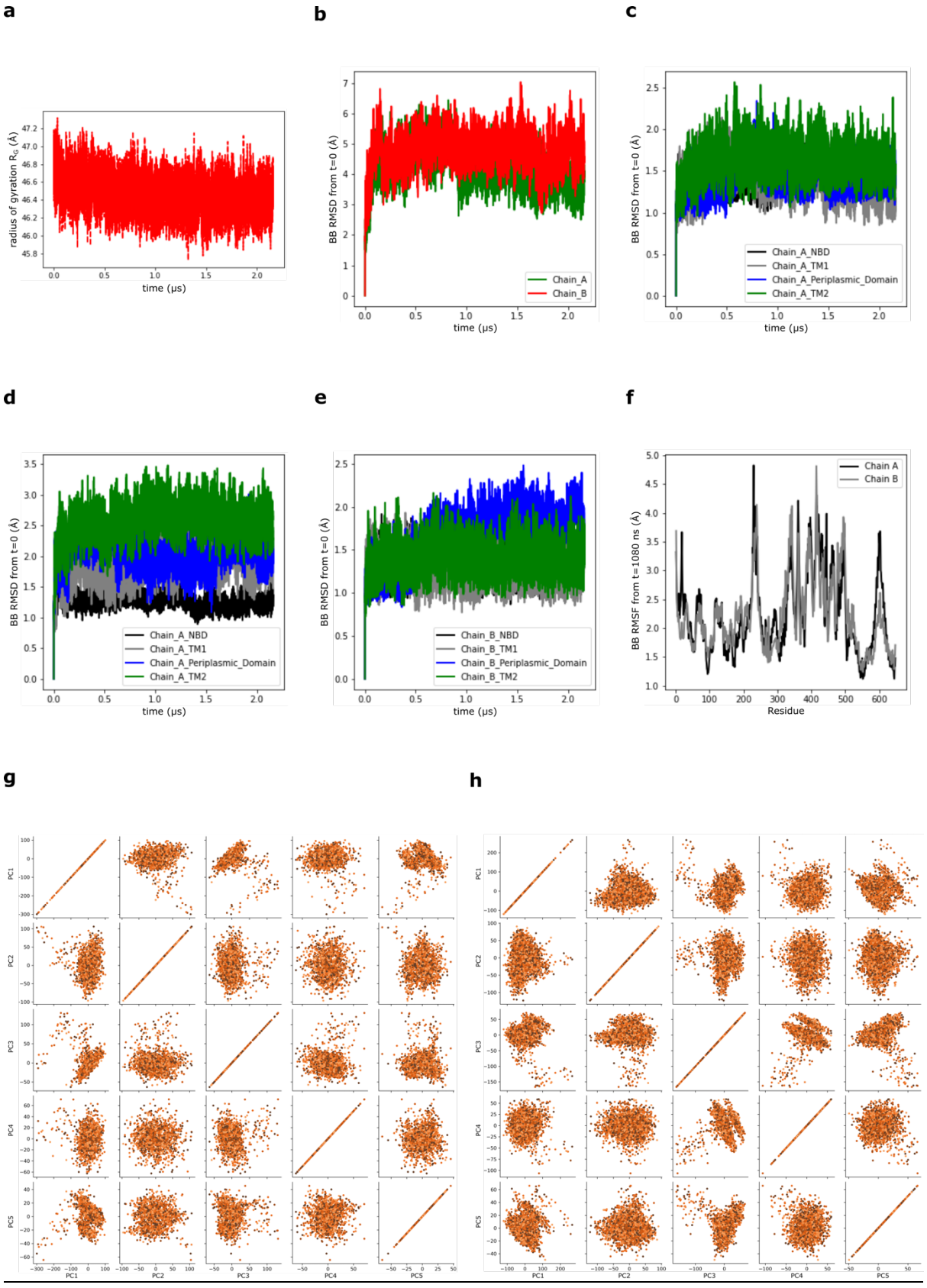


Figure 5.3.2-1. Radius of gyration, RMSD, RMSF, and PCA. a) All simulations showed a decrease or stabilisation in in the radius of gyration. Example taken from trajectory u8. **b)** Root mean square deviation (RMSD) for each chain. Most simulations showed no RMSD difference between chains. Example taken from trajectory u5. **c)** For

no internal lipid trajectories, similar RMSD values were found for all domains (nucleotide-binding domain (NBD) residues 1–126, transmembrane domain 1 (TM1) residues 247–305, periplasmic head domain residues 306–506, and transmembrane domain 2 (TM2) residues 507–648) in chain A. Example taken from trajectory u9. **d)** For internal lipid trajectories, the periplasmic head domain and TM2 showed higher RMSD values than the NBD and TM1 for chain A. Example taken from trajectory u3. **e)** For both internal and no internal lipid trajectories, chain B showed similar RMSD values in all domains. Example taken from trajectory u9. **f)** Example RMSF plot taken from no internal lipid trajectory u6. No difference was found between internal and no internal lipid trajectory RMSF plots. **g) – h)** Example taken from trajectory u6. PCA indicated that conformations were well-sampled after 1080 ns simulation time **(g)**, with increased sampling after 2160 ns simulation time **(h)**. Thus, 2160 ns simulation time was determined to be sufficient for this study. Note each point is a time frame. Also note that **h)** will look denser than **g)** as there are more frames plotted – the clustering of the points gives an indication of the trajectory sampling. More "island" clusters could potentially indicate inadequate sampling of certain conformations, as these were visited less often in the trajectory.

Re-projection of the first five principal components onto the backbone showed similar large-scale movements between trajectories. After taking into account the closure of the protomers, periplasmic head rotations were the next main movements in both internal lipid and no internal lipid trajectories (**Figure 5.3.2-2**). No other large-scale transitions occurred for any of the trajectories.

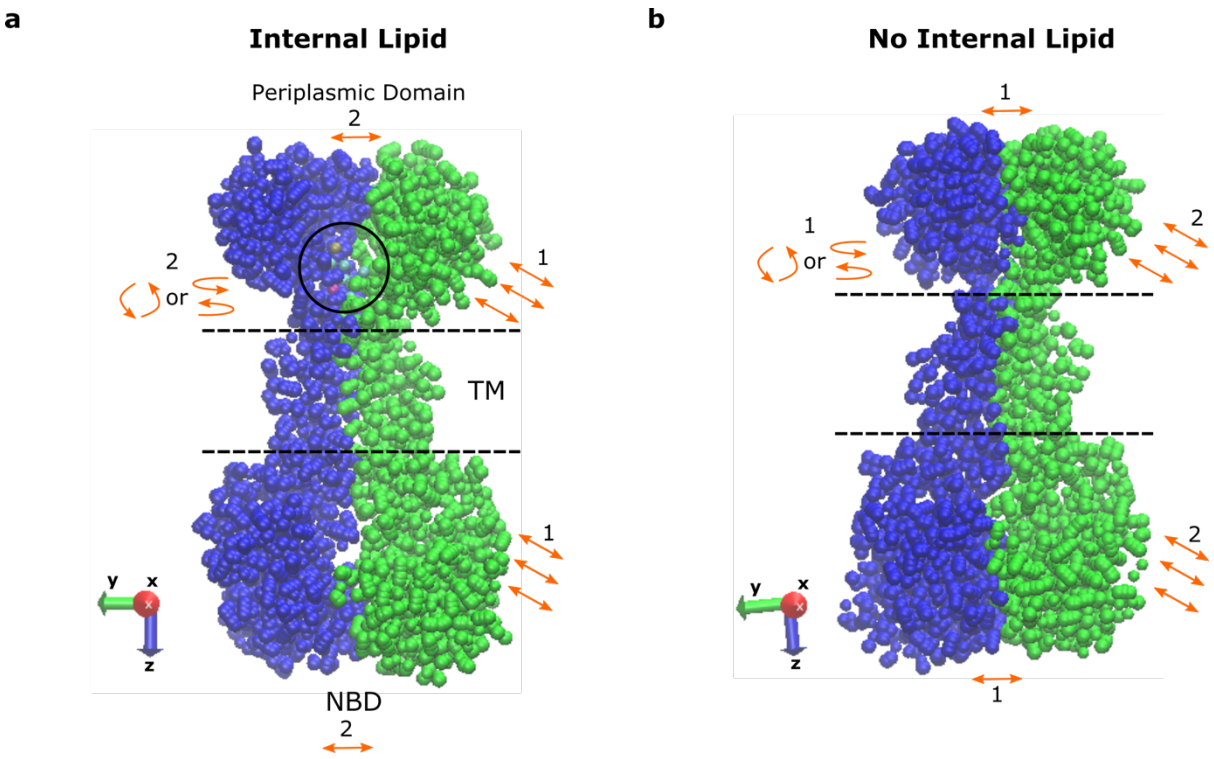


Figure 5.3.2-2. Visualisation of main structure movements. Re-projection of the first five principal components back onto the backbone atoms reveals the main structural movements in the trajectories. After taking into account the dimer closure, the next two main movements involve periplasmic head rotations (x to z rotation, "bobbing" or x to y rotation, "twisting"). Movements of the periplasmic head towards the membrane (black line, dotted) filters down via the transmembrane domain (TM) to the nucleotide-binding domain (NBD) causing a concomitant movement. Expansion and contraction of the dimer is also observed. Note that no difference between the internal lipid (**a**) and no internal lipid (**b**) trajectories was observed. POPE lipid is shown (circled, black). Example structures shown are from trajectory u4 (internal lipid) and trajectory u8 (no internal lipid) sampled at 1080 ns.

Ensemble similarity clustering analysis revealed a distinction in the structures produced by the two conditions (**Figure 5.3.2-3**). Both internal lipid and no internal lipid trajectories showed higher in-group similarity than out-group similarity. Interestingly, there does appear to be a pattern of alternating similarity and dissimilarity for the u7 trajectory, but this is likely to be a result of the K-means clustering method than due to presence of bias as the effect disappears for low values of K.

Figure 5.3.2-3. MacB ensemble similarity. K-mean clustering of samples of protein conformations taken every 1000 frames for all trajectories. Jensen-Shannon divergence provides a measure of conformation similarity, with 0 indicating complete similarity and 1 indicating complete dissimilarity. u1 to u5: trajectories with internal POPE lipid; u6 to u10: trajectories without internal POPE lipid.

3.3 Substrate Cavity Access in MacB

The next question was how MacB interacted with the trapped POPE lipids. This was to provide clues as to its ability to export a variety of substrates, given the amphiphilic nature of phospholipids. It was observed that these POPE lipids underwent a "drawing up" into the MacB cavity after becoming trapped between the dimers (**Figure 5.3.1-1**). Thus, this cavity could possibly show a route of entry during the transportation process. The exact access route for substrates in MacB remains elusive; four periplasmic "hotspot" residues (T349, Y376, F444, and W505) have been proposed as substrate interaction sites based on a mutagenesis study (Crow *et al*, 2017). Thus, to estimate the size of the MacB cavity between internal POPE lipid and no internal POPE lipid trajectories, I calculated the distance between these hotspot residues between protomer A and protomer B. The results reveal a much more open cavity in the internal lipid trajectories (**Figure 5.3.3-1a**). This is likely the result of the presence of the POPE lipid, which acts as a wedge between the protomers. Fascinatingly, each trajectory showed a diversity of cavity openings, reflecting deep dynamics in this region (**Figure 5.3.3-1b**). In contrast, MacB structures without internal POPE lipid showed the same uniform closed structure (**Figure 5.3.3-1c**).

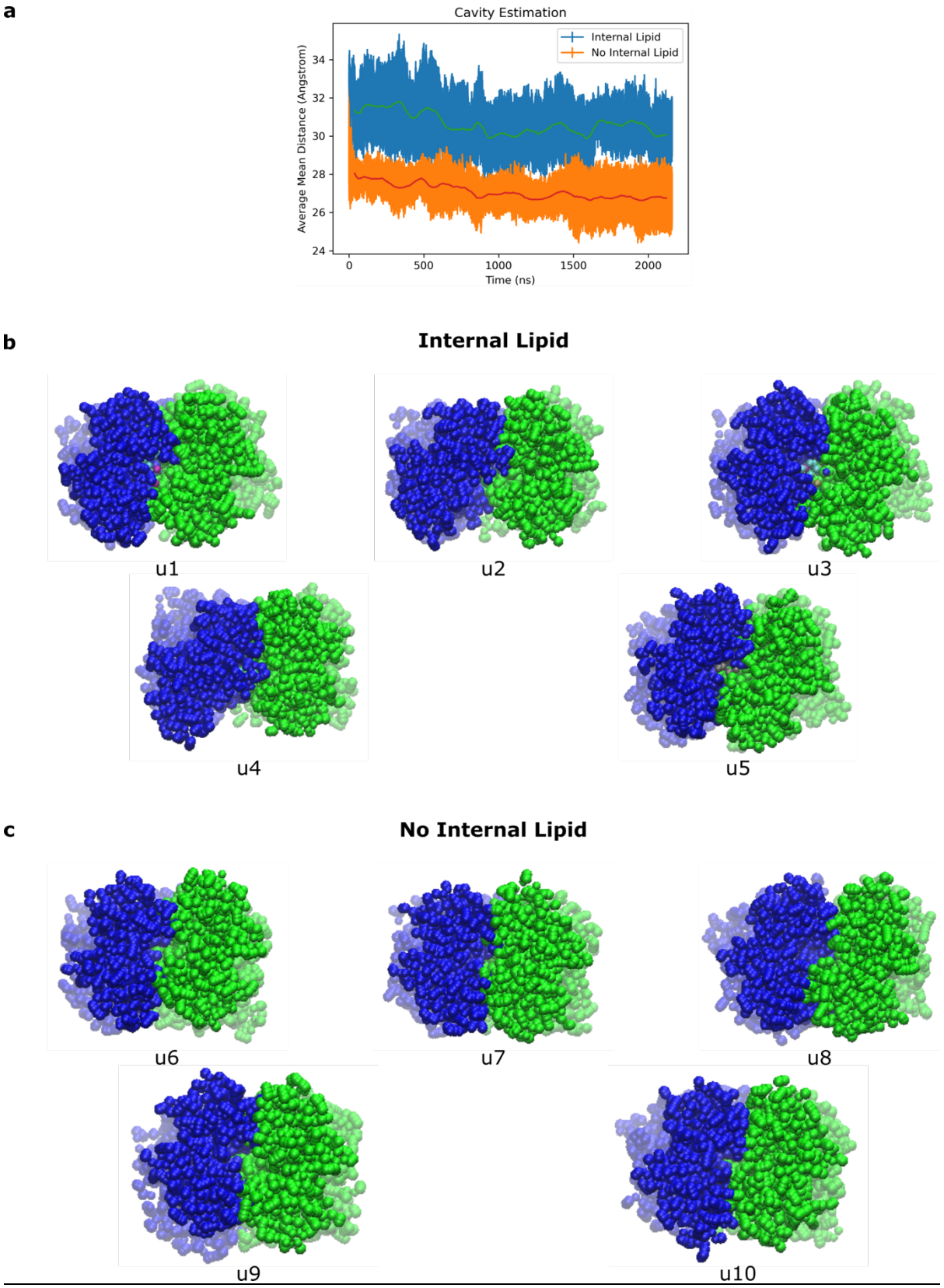


Figure 5.3.3-1. Periplasmic cavity dynamics. a) The MacB periplasmic head cavity maintains an open state when POPE lipid is bound (internal lipid, blue) compared to the closed state without bound POPE lipid (no internal lipid, orange) throughout the simulation time. Moving averages are shown (internal lipid, green, n=5 (u1–u5)

trajectories; no internal lipid, red, n=5 (u6–u10) trajectories). **b)** POPE bound trajectories (u1–u5) show a diversity of structures. Note how protomer A (blue) moves in relation to protomer B (green), indicating more open and closed conformations in the periplasmic cavity. Sample structures taken at time 1080 ns. **c)** No POPE lipid trajectories (u6–u10) show a uniform closed structure. Note how protomer A (blue) does not move in relation to protomer B (green), indicating a uniform closed conformation when no internal POPE is present. Sample structures taken at time 1080 ns.

The binding cavity for the MacB trajectories were examined further. Without POPE lipid, the trajectories showed a similar pocket, with residue F444 pointing away from the pocket and residues T349, Y376, and W505 remaining buried in the central cavity (**Figure 5.3.3-2a**). Unexpectedly, the MacB protomers with internal POPE lipid appeared to adopt either a more open or closed conformation which reflected the degree of binding of POPE (**Figure 5.3.3-2b**). This was not seen in the trajectories without lipid, which all adopted a closed conformation, without rotation of the periplasmic head regions. In this case, binding pockets appeared very similar between trajectories (**Figure 5.3.3-2a**). In contrast, when POPE is present for the most open conformation observed, POPE interacted with residues Y376, F444, and W505. As the conformation became more closed, POPE disengages from F444 and an opening starts to appear at the top of the periplasmic domain (**Figure 5.3.3-2b**). Since T349 was not observed to interact with POPE in any of the trajectories, it is possible that this interaction takes place only during extrusion rather than initial binding. Interestingly, initial trajectory u1 is the only one to contain two POPE lipids; one lipid was subsequently lost during equilibration for the repeat trajectories. This could indicate that substrate binding is reversible. Considering that extrusion is unlikely to occur when the periplasmic port is open, examination of the structural diversity in the trajectories gives a likely path of substrate interaction from most open to most closed conformations (**Figure 5.3.3-2b**). This binding flexibility reflects dynamics of the protein cavity where substrate is likely to pass.

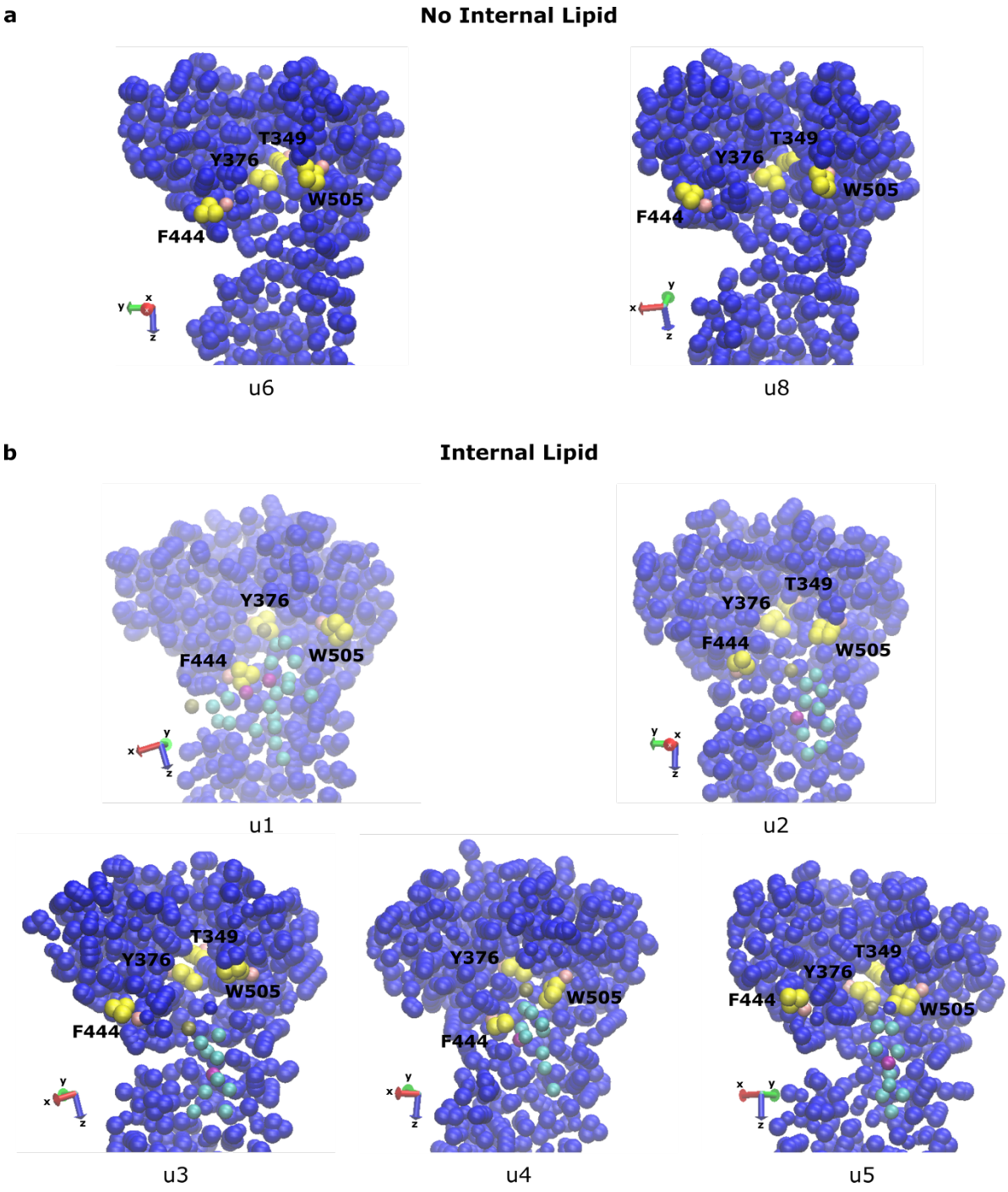


Figure 5.3.3-2. **Changes in MacB Binding Cavity**. **a)** MacB trajectories without an internal POPE lipid show similar binding pockets. Periplasmic hotspot residues (T349, Y376, Y444, W505) are shown. Note how residue F444 points away from the binding pocket in the closed state. Trajectories u6 and u8 shown as examples, taken at time 1080 ns. **b)** Protomer A POPE binding dynamics with the periplasmic hotspot residues (T349, Y376, Y444, W505) for trajectories (u1–u5). In the most open conformations (u4), POPE interacts with Y376, Y444, and W505. As the cavity becomes more closed, POPE disengages from F444 and starts to interact only with Y376 and W505 (u2→u3→u5). Note for u1, there are two engaged POPE lipids. The loss of the second POPE lipid in trajectories u2–u5 during equilibration may indicate this this binding is reversible. Also note that T349 was not observed to

interact with POPE lipid in any of the trajectories, which may indicate that it interacts with substrate during extrusion rather than initial binding. Structures taken at time 1080 ns.

3.4 Protein-Lipid Interactions

Analysis of the MacB cavity with POPE lipid provided a fascinating indication of the dynamics of substrate binding and possible routes for extrusion. Next, I wanted to test if there was any difference in how MacB interacted with the lipid bilayer when internal POPE was present vs when it was absent. As for the HlyB-HlyA MD simulation analysis (see **Chapter II**), I split the interactions into three scales: 1) an overview of interactions across the entire membrane (lipid enrichment), 2) within 15 Å of the protein (lipid clustering), and 3) close protein-lipid contacts (within 3.5 Å). Lipid enrichment analysis found no difference between internal POPE and no internal POPE trajectories, except for POPE lipid which was expected given that it is bound in the cavity (**Figure 5.3.4-1**). It must be noted that lipid enrichment does not account for the length of protein-lipid interactions. Thus, an enriched lipid could occur through long-term interactions with a few lipids or many short-term interactions. The former seems to account for the appearance of cardiolipin enrichment in MacB (**Figure 5.3.4-1**) as shown by the close contact analysis (see below).

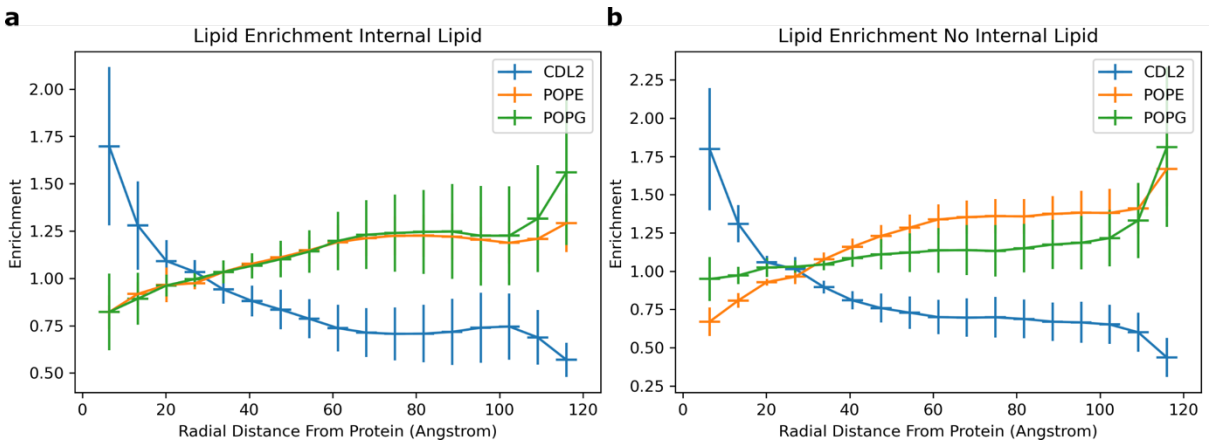


Figure 5.3.4-1. Lipid enrichment between internal and no internal lipid trajectories. No difference was found between lipid enrichments for internal lipid (**a**) and no internal lipid (**b**) trajectories, except for POPE. This was expected as there is POPE bound in the cavity. Note the cardiolipin (CDL2) enrichment does not reflect long-term interactions (see close contact lipid analysis). Bars in the x-axis show each 5 Å band used in the lipid

enrichment calculation, while bars in the y-direction are error bars showing standard deviation (n=5 for internal lipid and n=5 for no internal lipid).

Lipid clustering also shows no difference between internal lipid and no internal lipid trajectories, except for POPE in the protomer A and B periplasmic domain and protomer A transmembrane 1 domain (**Figure 5.3.4-2**). Again, this is expected as this reflects the bound POPE lipid.

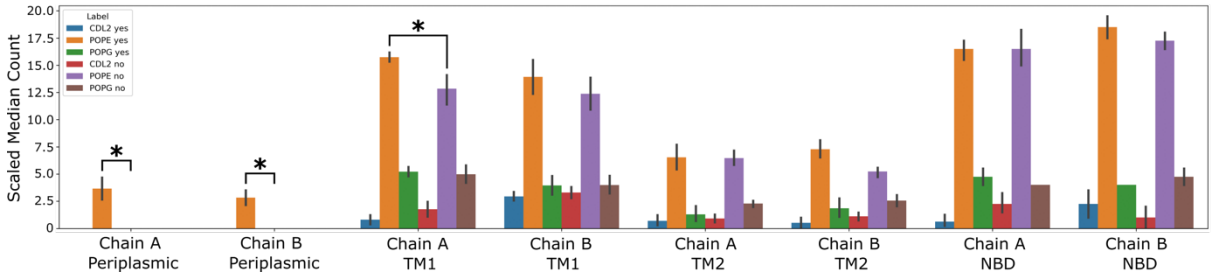


Figure 5.3.4-2. Lipid clustering between internal lipid and no internal lipid trajectories. Lipid clustering analysis for protomer A and protomer B separated by domain. Statistically significant (star, $p < 0.01$) differences between internal lipid ("yes") and no internal lipid ("no") trajectories was found for POPE in the periplasmic domain of protomer A and B, and protomer A TM1 (star). Domains: nucleotide-binding domain (NBD) (residues 1–246), transmembrane domain 1 (TM1) (residues 247–305), periplasmic head domain (Periplasmic) (residues 306–506), and transmembrane domain 2 (TM2) (residues 507–648). Significance was determined using a permutation test with Bonferroni correction for multiple testing (n=5 for internal lipid "yes" and n=5 for no internal lipid "no"). Counts were found by taking the median lipid interaction values over the trajectory and scaling by the number of amino acids queried in each region. Error bars show the 90% confidence interval (CI) centred around the scaled median count.

Next, lipid close contacts ("lipid preference") were examined to find critical interactions between MacB and specific lipids. Close contact (within 3.5 Å) lipid interaction differences between trajectories with and without internal POPE lipid were calculated. **Figure 5.3.4-3** shows the critical lipid interactions (>0.05 fraction frames). Statistically significant (p-value < 0.01) interactions are highlighted.

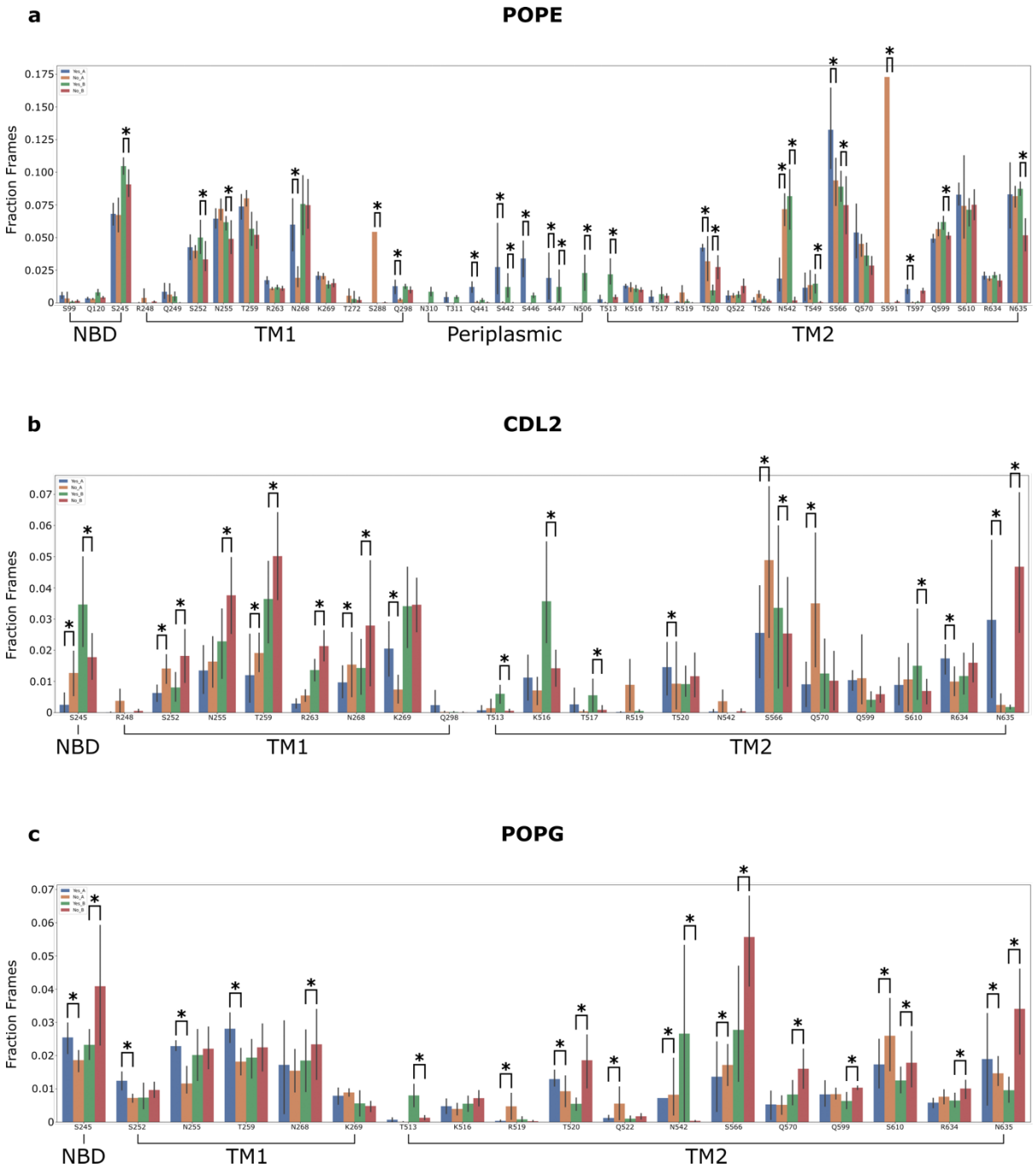
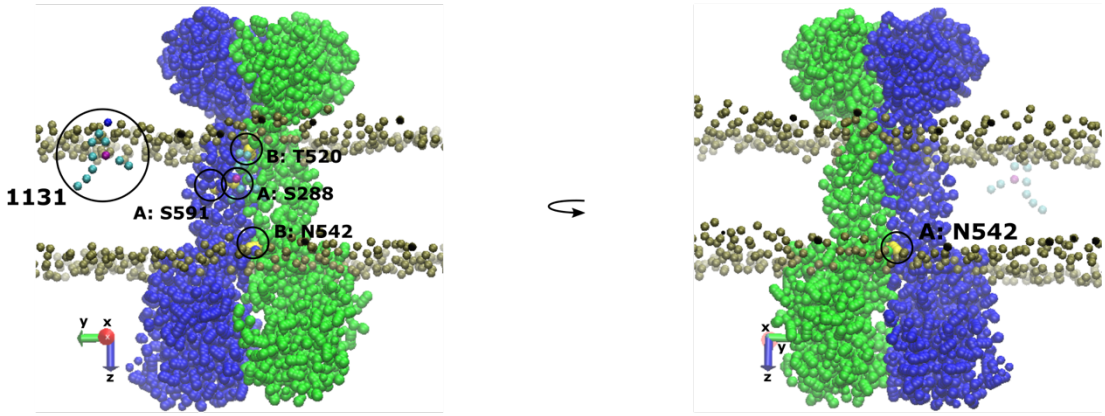


Figure 5.3.4-3. Lipid close contact interactions between internal lipid and no internal lipid trajectories. Differences in lipid close contacts (fraction frames) between internal lipid ("yes") and no internal lipid ("no") trajectories were examined for POPE (**a**), cardiolipin (CDL2) (**b**), and POPG (**c**) lipids separated by interactions in protomer A and B. Statistically significant ($p<0.01$) interactions are highlighted (star), based on the Monte Carlo permutation test with a Bonferroni correction for multiple comparisons (n=5 for internal lipid "yes" and n=5 for no internal lipid "no"). No discernible difference could be found between trajectories, apart from four outliers in POPE interactions at amino acids S288, T520, N542, and S591. Note that POPE interactions in the periplasmic domain are expected due to the presence of the internal POPE lipids. Error bars show the 90% CI centred around the mean fraction frames.

Examination of significant lipid close contact interactions with different states of MacB (see **Figure 5.3.3-1**) yielded no discernible pattern of interaction between states. Lipids tended to interact in certain regions (*e.g.* S245 in the NBD; S252, S255, T259, and N268 in TM1; Q441, S442, S446, S447, and N506 in the periplasmic domain; T513, T520, S566, Q570, Q599, N610, and N635 in TM2), but apart from the periplasmic domain there was no difference in interactions between internal POPE lipid and no internal POPE lipid trajectories (**Figure 5.3.4-3**). Although multiple residues appear as statistically significant, statistical significance on its own does not imply biological significance. For there to be a case for biological significance, I would expect a correspondence of regions between the lipid clustering and lipid preference analysis, as well as differences to appear in the lipid enrichment analysis. The fact that this is not the case, and that the statistically significant regions appear evenly distributed throughout MacB in the lipid preference analysis, seems to suggest that the significance is the result of the simulation rather than reflecting biology.

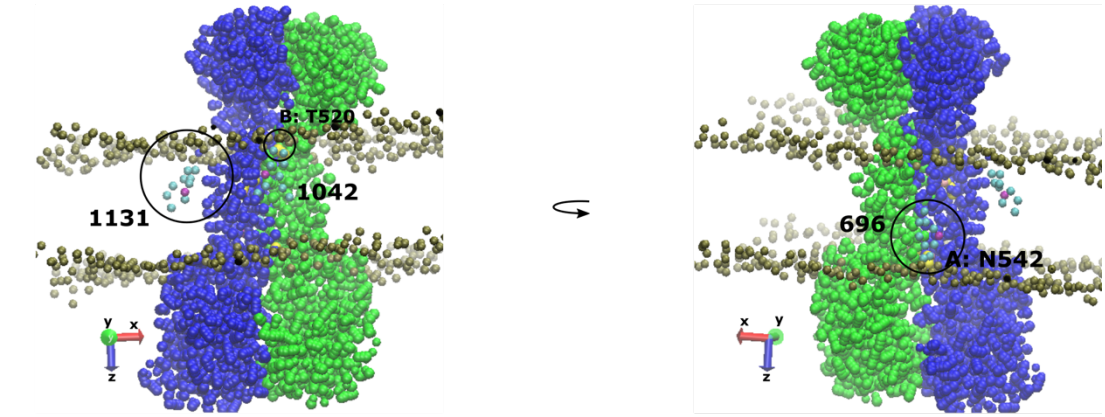
However, there were four noticeable outliers in POPE interactions between internal lipid and no internal lipid trajectories – S288, T520, N542, and S591 showed unusually high levels of interaction when no internal lipid was present. These could be traced to trajectory u10, and examination of this trajectory showed an astonishing interaction of one particular POPE lipid with the transmembrane stalk region of MacB over 1 μs of simulation time (**Figure 5.3.4-4**). It was observed in this trajectory that a POPE residue detached from the lipid bilayer and started to interact with the stalk region after ~1 μs simulation time. This interaction continued to the end of the simulation, with the POPE lipid becoming more buried into the stalk, creating a local distortion of the membrane. It was also found that residue N542 operated an asymmetric gating mechanism when interacting with lipid; in protomer B, N542 remained closed to lipid contact, while in protomer A, N542 started interacting with lipid around 900 ns into the simulation time. Interestingly, a disruption of lipid interaction in residue T520 was observed during the burial process of POPE residue 1131. These interactions could reflect multiple possible sites of entry for MacB substrates, particularly lipids or substrates with a hydrophobic character.

a



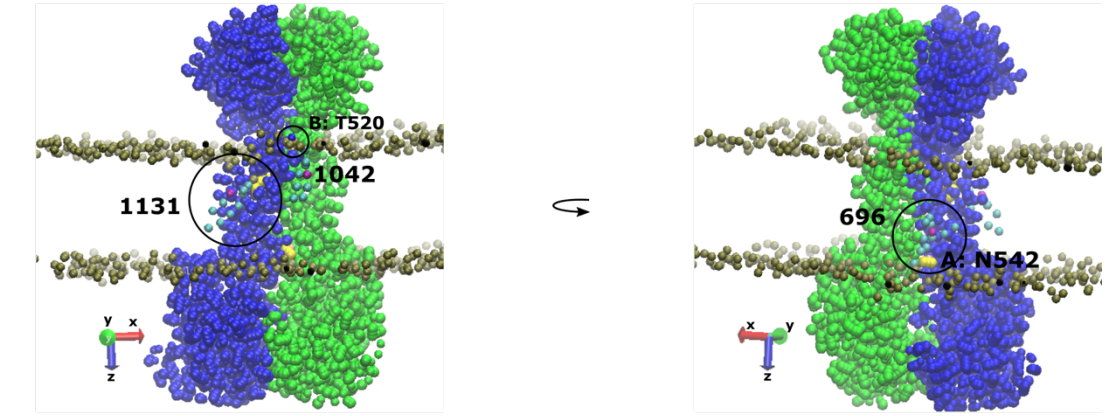
0 ns

b



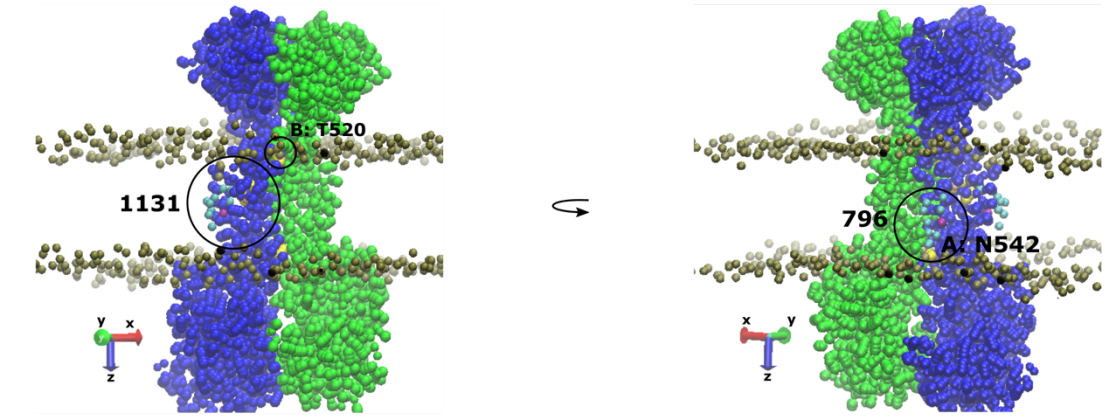
900 ns

c



1050 ns

d



2160 ns

Figure 5.3.4-4 (previous page). u10 trajectory interaction of POPE lipid with MacB. Interaction with POPE lipid 1131 with S288 and S591 in protomer A over 2160 ns simulation time. At the start of the simulation, POPE 1131 is present in the bilayer, while T520 in protomer B interacts with POPE lipids. After 900 ns, protomer A residue N542 protomer starts to interact with POPE lipid while protomer B residue T520 continues interactions with POPE lipids. After 1050 ns, T520 stops interacting with lipids while POPE residue 1131 detaches from the bilayer and interacts with residues S288 and S591. This interaction continues until the end of the simulation, with residue 1131 becoming more buried into protomer A. This process appears to create a local distortion of the bilayer around POPE residue 1131. Note that residue N542 in protomer A maintains interaction with lipid from 1050 ns simulation time, while N542 in protomer B never interacts with lipid in the simulation time.

4. Discussion

4.1 MacB Conformational Changes in Light of "Molecular Bellow" Mechanism of Transport

I found a separation of "closed" and "open" states of MacB which depended on whether there was trapped POPE lipid between the two MacB protomers. The current model of MacAB-TolC transport posits a "molecular bellows" mechanism whereby ATP-binding provides the energy for substrate transport and hydrolysis resets the structure (**Figure 5.4.1-1a**) (Crow *et al*, 2017). In this model, the MacB NBDs remain in an undimerised "open state" until ATP-binding where it then adopts a dimerised "closed state". ATP-binding provides the "power-stroke" to transfer substrate to the MacA channel, while hydrolysis resets the structure from the closed state back to the open state (Crow *et al*, 2017) (**Figure 5.4.1-1a**).

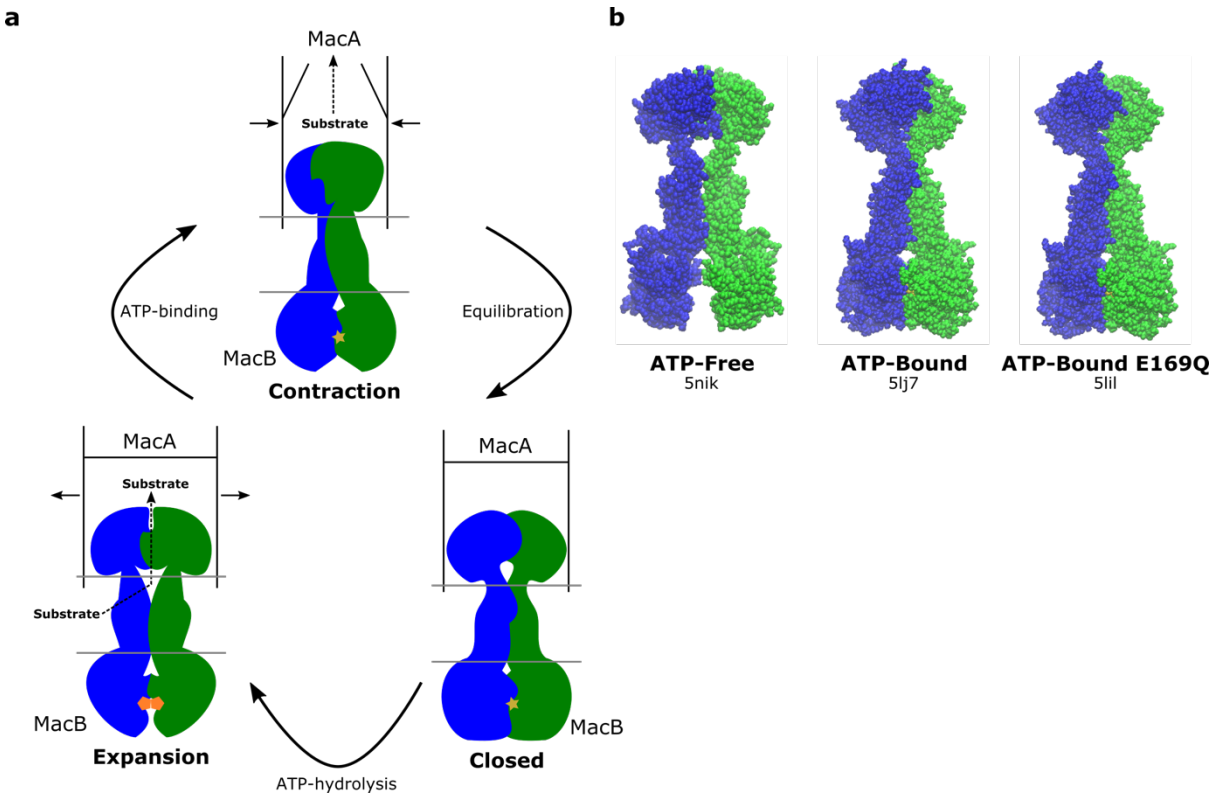


Figure 5.4.1-1. "Molecular Bellows" mechanism of transport by the MacAB-TolC assembly. a) Schematic of the "molecular bellows" mechanism of MacAB-TolC transport proposed by Crow *et al* (2017), where substrate enters MacB via the periplasmic port to the gated MacA channel. Substrate is expelled during ATP-binding which causes a contraction allowing the MacA gate to open. ATP is still bound during MacA gate closure and is finally

hydrolysed as the last step to reset the system for substrate transport by structure "expansion". **b)** Structures used by Crow *et al* (2017) to support their "molecular bellows" model. The ATP-Free state shows undimerised NBDs and an open periplasmic head port (pdb ID 5nik) (Fitzpatrick *et al*, 2017). The unmutated ATP-bound state (pdb ID 5lj7) (Crow *et al*, 2017) shows dimerised NBDs, suggesting that dimerisation of the NBDs is a pre-requisite of hydrolysis. In this structure, the periplasmic port is closed but the apex of the head region is slightly open. Finally, the ATP-bound structure with E169Q mutation (pdb ID 5lil) (Crow *et al*, 2017) shows dimerised NBDs and a slightly open periplasmic port.

The model proposed by Crow *et al* (2017) is based on comparing two structures of MacB which were available at the time: X-ray crystal structures of ATP-bound MacB (pdb IDs 5lil, 5lj6, 5lj7) and a cryo-EM structure without ATP (pdb ID 5nik) (Fitzpatrick *et al*, 2017) (**Figure 5.4.1-1b**). The former was the model I used to conduct these simulations, removing the MacA and TolC components (**Figure 5.4.1-1b**). This model was unusual in that it had density for a feature that appeared trapped in the periplasmic cavity, providing a wedge between the protomers. It is unknown whether these were substrates frozen mid-transport, as the densities could not be identified. The structure also has a disulphide bond linking MacB with MacA in the periplasmic head region of MacB (MacA N271 to MacB G465), which could also potentially limit MacB movement within MacA (Fitzpatrick *et al*, 2017). One downside of the simulations is that the MacB periplasmic head is unconstrained and exposed to water molecules, which is unlikely to occur during active transport as the head region is gripped by MacA. However, the NBDs are located in the cytoplasmic side and are not in contact with MacA. Thus, even if movement of the periplasmic head is constrained by MacA there could still potentially be movement of the NBDs between an open and closed state. To understand dynamics using real examples, we turn our attention to experimental MacB-like structures which have been isolated in various dynamic states.

To date, structures of MacB and MacB-like structures have been solved in various states (**Figure 5.4.1-2a**). Different structures reveal the MacB NBD in undimerised and dimerised states, while the periplasmic head region is also in various states of "openness" (**Figure 5.4.1-2a**). Indeed, MacB appears to be a very dynamic molecule and it is possible that it can adopt different conformations in its energy landscape (Gershenson *et al*, 2014). Conformational diversity can also be seen in MacB-like structures (**Figure 5.4.1-2b**) even when substrate is bound.

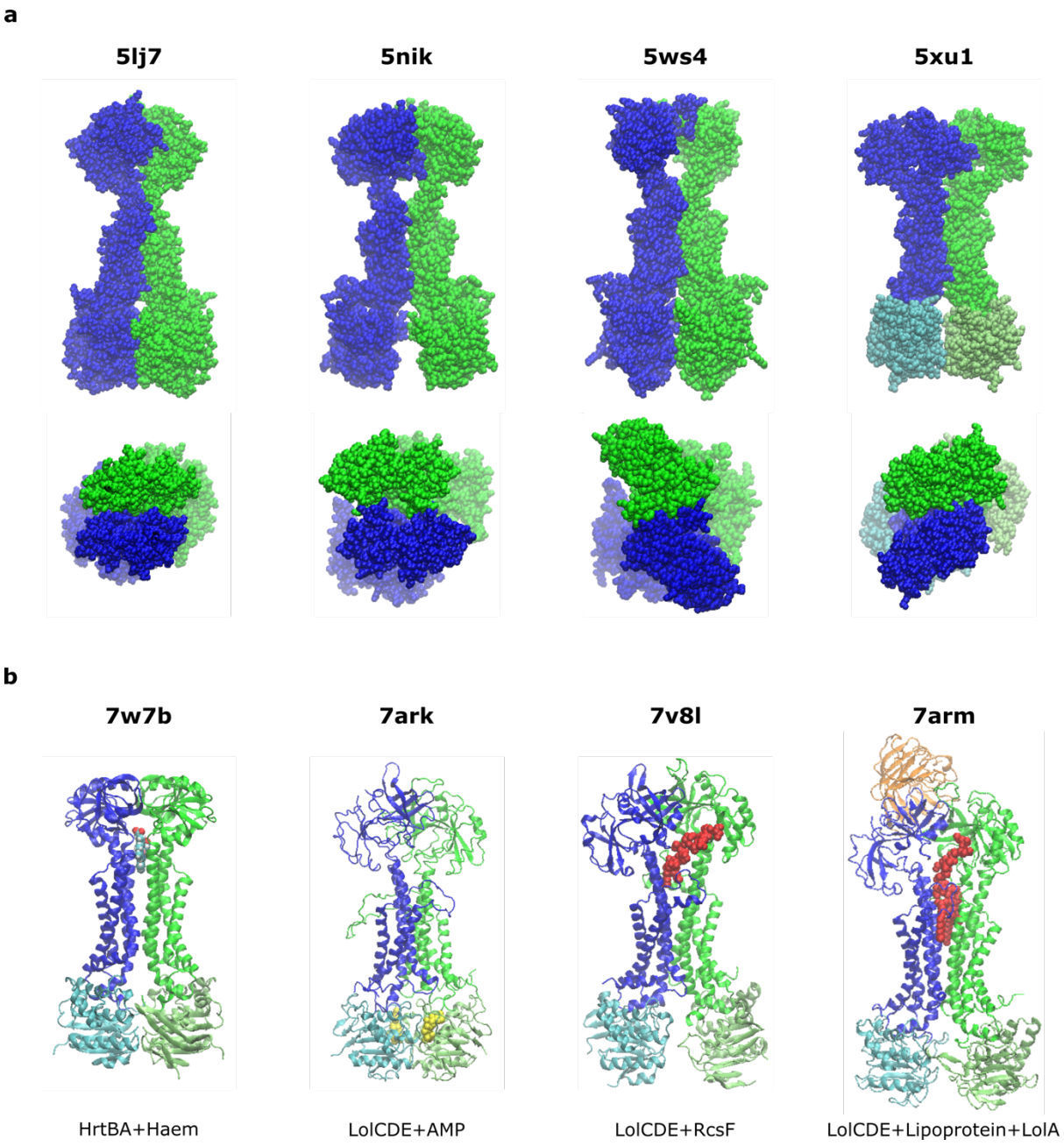


Figure 5.4.1-2. Structural diversity of MacB and MacB-like ABC-transporters. a) MacB and MacB-like structures show a range of conformations, particularly around the placement of the nucleotide-binding domains (NBD) relative to the periplasmic head region. MacB from *Aggregatibacter actinomycetemcomitans* bound to ATP shows a closed conformation of both NBDs and periplasmic head (pdb ID 5lj7) (Crow *et al*, 2017). MacB from *Escherichia coli* without ATP or substrate shows an open conformation in both the NBDs and periplasmic head (pdb ID 5nik) (Fitzpatrick *et al*, 2017). In contrast, MacB with bound ADP-analogue from *Acinetobacter baumannii* shows a closed NBD (like structure 5lj7) but a larger opening in the periplasmic head region compared to structure 5nik (pdb ID 5ws4) (Okada *et al*, 2017). Non-canonical MacB-like structure from *Streptococcus pneumoniae* R6 also shows a closed NBD and an open periplasmic head region (pdb ID 5xu1) (Yang *et al*, 2018). Note that the NBDs (cyan, lime green) do not form a continuous chain with the stalk and periplasmic head (blue,

green). **b)** MacB-like structures with substrates also show structural diversity. Like structure 5xu1, these MacB-like structures have NBDs (cyan, lime green) which bind to the stalk of each protomer (blue, green) rather than form a continuous chain like MacB. Haem-transporter HrtBA from *Corynebacterium diphtheriae* NCTC 13129 binds haem between the protomers (pdb ID 7w7d) (Nakamura *et al*, 2022). Note that the NBDs are dimerised. LolCDE also adopts a closed NBD when AMP (yellow) is bound (pdb ID 7ark), although the periplasmic head is in an open conformation (Tang *et al*, 2021). LolCDE also binds RcsF substrate (red) between the protomers, although in this structure the NBDs are in an undimerised conformation (pdb ID 7v8l) (Bei *et al*, 2022). LolCDE with lipoprotein substrate (red) has recruited carrier protein LolA (orange) with undimerised NBDs (pdb ID 7arm) (Tang *et al*, 2021).

I found that in my simulations MacB could adopt a "closed" NBD dimerised state within 50 ns of simulation time. In the examples of the POPE-trapped "open states", a variety of conformations was observed showing the dynamism of these regions. Given the short simulation time, it is possible that the MacB NBDs and periplasmic head region are in equilibrium between "open" and "closed" state over a micro-to-second timescale. This would conflict with the proposed "molecular bellows" mechanism, which posits that closure can only occur under non-equilibrium ATP-turnover conditions (Crow *et al*, 2017). Recently, there is experimental evidence to suggest that ATP-turnover in MacB is tightly coupled to substrate transport, which is in conflict with the current MacB transport paradigm (Souabni *et al*, 2021). In this study, it was found that 1 ATP molecule was hydrolysed per 3 substrates transported, and that ATPase activity was not linked to the presence of substrate (Souabni *et al*, 2021). Although this would appear puzzling, it is in agreement with a model of substrate turnover where different states are in equilibrium until export which drives MacB to open the periplasmic cavity to the MacA channel. Although this might be considered unusual, recent experimental evidence suggests a range of dynamic conformations are possible under active export for an ABC transporter; in *Thermus thermophilus* multi-drug resistance proteins A and B (TmrAB) it was established that phosphate exit was responsible for a conformational reset (Hofmann *et al*, 2019). Furthermore, it was found through MD simulations that the switch from an inward facing narrow to an inward facing wide conformation were in dynamic equilibrium. ATP-binding caused the NBDs to dimerise, occluding substrate access to the inward-facing wide conformation but allowing for substrate transport through the outward facing open conformation (Hofmann *et al*, 2019). The dynamics of MacB transport is discussed in the next section.

4.2 Substrate Entry and Turnover Mechanisms

The exact route of substrate entry into MacB remains elusive. A recent structure has found haem trapped between the protomers of a MacB-like structure (Nakamura *et al*, 2022), in a similar conformation to my trajectories with trapped POPE lipid. MacB-like structures LolCDE also bind their lipoprotein substrates between protomers (**Figure 5.4.1-2b**). These structures (**Figure 5.4.1-2b**), along with the MacB cryo-EM structure with unknown trapped density (pdb ID 5nik) (Fitzpatrick *et al*, 2017) (mentioned above) suggests that the periplasmic port is an entry point for substrates. It should also be noted that structures of MacB and MacB-like proteins with bound ATP or ATP-analogues show conformations closed to substrate binding (Crow *et al* 2017; Tang *et al*, 2021; Bei *et al*, 2022). Mutagenesis studies of the periplasmic head region also established the existence of "hotspot" residues (T349, Y376, F444, and W505) that could be involved in substrate transport (Crow *et al*, 2017). If the periplasmic port is the route of entry, substrates would need to either reside in the periplasm or be exported there by another transporter. The upper membrane leaflet is also another point of entry.

The diversity of exported substrates would appear to indicate that the MacB port can accommodate substrates in a non-specific manner. The fast turnover of substrates observed by Souabni *et al* (2021) (3 per ATP hydrolysed) would also indicate that the MacB port can accommodate more than one substrate at a time. Indeed, in one of my POPE-trapped simulations (trajectory u1), I observed two POPE lipids in the periplasmic cavity, one of which was subsequently lost during equilibration for the repeat simulations. This would appear to suggest that substrate binding could potentially be an equilibrium process. I observed interactions of the "hotspot" residues Y376, F444, and W505 with the trapped POPE substrate, with these interactions showing a high-level of dynamism between trajectories, which was not observed in the closed, substrate-free state. This led to my proposal that a pathway of "open" to "closed" substrate-bound states could exist during the transportation process. In addition, I did not observe any interaction with T349 which remained buried in the cavity. Thus, it is possible that interaction of substrates with this residue could occur during a conformational change induced either by ATP-binding or hydrolysis.

Mechanisms of transport must account for the energetics of the process (Gershenson *et al*, 2014). Like enzyme catalysis, substrate export is an energetically unfavourable process that requires input energy to overcome the energetic barrier (Sousa *et al*, 2020). In the case of MacAB-TolC export, this energy appears to be derived solely from ATP-hydrolysis. The question then remains not only when the input energy is occurring, but the pre- and post-hydrolysis states. As in enzyme catalysis, we need to consider several states and their energetic barriers: the substrate-free "closed" and "open" states (respectively named M^{closed} and M^{open}), the substrate-bound state (named $M_{substrate}{}^{open}$), the ATP-bound state with and without substrate (named respectively $M_{substrate}{}^{ATP}$ and $M^{ATP}{}_{closed}$), and post-ATP hydrolysis states (named $M^{ADP+Pi}{}_{closed}$). We can presume that $M^{ADP+Pi}{}_{closed}$ is a state that does not bind substrate, since the energy of transport has gone into moving substrate from the cavity into the MacA channel. This irreversible path: $M^{ATP} \rightarrow M^{ADP+Pi}$ must be necessary for one-direction export to be possible. In fact, the release of products ADP+P_i from the MacB NBDs could "reset" the structure to an M^{closed} and M^{open} equilibrium. The exact path would then depend on energy accounting and the evidence we have accumulated so far from the available structures. I will propose that the unobserved "transition" state that places MacB on the irreversible $M^{ATP} \rightarrow M^{ADP+Pi}$ path is in fact the dual ATP- and substrate-bound state. Energetically, binding of ATP first could cause futile hydrolysis cycles as the structures observed indicate that ATP-bound structures are closed to substrate binding (see **Figure 5.4.1-2**). If substrate is bound first while MacB is in the open state, this could lead to the export pathway. I provide a proposed pathway in **Figure 5.4.2-1**.

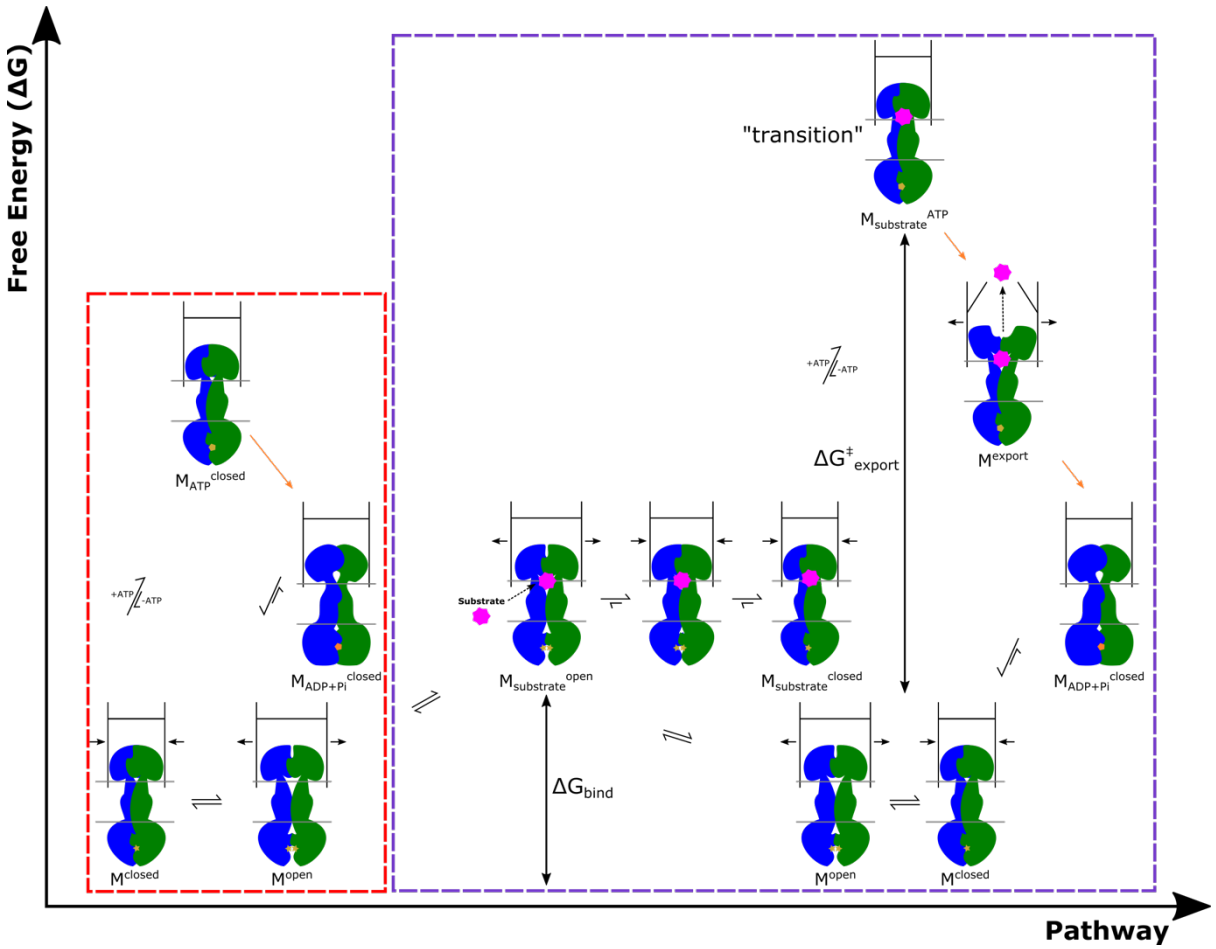


Figure 5.4.2-1. Proposed MacB Transport Free Energy Pathway. Proposed MacAB-TolC transportation process, accounting for current experimental evidence and protein dynamics. At the ground state energy level, MacB can cycle between "closed" (M^{closed}) and "open" (M^{open}) states, causing concurrent expansion and contraction of MacA (box). We assume that like most ABC-transporters, ATP binds when the nucleotide-binding domains (NBDs) are dimerised or favours NBD dimerisation (M_{ATP}^{closed} state). ATP-binding of the M^{closed} state can lead to futile hydrolyses cycles without substrate transport (red box). However, binding of substrate before ATP can lead to non-futile ATP-hydrolyses cycles that actively transport substrate (purple box). Binding of substrate occurs in the M^{open} state ($M_{substrate}^{open}$) leading to a series of equilibrium states towards the NBD dimerised closed state with substrate bound ($M_{substrate}^{closed}$). In this state, binding of ATP leads to irreversible substrate export once hydrolysis occurs. We assume that the dual substrate and ATP-bound state is an unseen "transition" state ($M_{substrate}^{ATP}$) which rapidly leads to hydrolysis and the short-lived M_{export} state. This provides the energy for transport by opening of the periplasmic head region causing rapid expansion of MacA and opening of the periplasmic gate. Release of ADP+P_i resets the structures to the ground energy state.

Available structures have provided "snapshots" of MacB and MacB-like structures. Combined with current experimental evidence, and my MD simulations, the proposed model (**Figure 5.4.2-1**) takes into account the following observations: 1) Dimerisation of NBDs can occur without ATP-binding (structure snapshots), 2) Dimerisation of NBDs can occur under

equilibrium molecular dynamics within a short time-frame (though the reverse may take a longer time) (my study), 3) There appear to be multiple substrate-bound states of MacB (my study and structure snapshots), 4) MacB can undergo futile ATP-hydrolysis cycles were no substrate is transported (Souabni *et al*, 2021), 5) Substrate binding can occur before ATP-binding and needs an open periplasmic cavity (my study and structure snapshots), 6) ATP-binding and hydrolysis requires a dimerised NBD (structure snapshots), 7) Presence of substrate does not prohibit NBD dimerisation (my study and structure snapshots), and 8) Once ATP is bound, substrate cannot be bound (structure snapshots). The proposed model offers a more dynamic free-energy landscape approach to MacAB-TolC substrate transport.

4.3 Effect of Lipid on Transportation

I did not find an effect of the lipid or lipid types between the internal POPE lipid and no internal POPE lipid trajectories, apart from the obvious clustering of the internal POPE lipid to the periplasmic cavity. While cardiolipin did appear to be enriched within 30 Å of MacB, there was no difference between internal lipid and no internal lipid trajectories, suggesting it is a constitutive effect. I did find amino acid sites which appear to favour lipid interaction, but again there was no discernible difference between internal lipid and no internal lipid trajectories which could suggest a biological mechanism of action. The only outliers observed were four sites which favoured POPE interaction when no internal POPE was present – this could be traced to a single trajectory (u10) where it was observed that a POPE lipid detached from the bilayer and interacted with the transmembrane stalk over 1 μs of simulation time. I proposed that this could be a point of entry into the MacB complex, particularly for substrates that are not in the periplasm. Observation of other sites interacting with lipid in this trajectory also suggested multiple routes of entry into the stalk region, where they could be "pulled up" into the periplasmic cavity for export. Of course, it is also possible that the observed dynamics are an artefact of the simulation (since it was only found in one trajectory) and further experimental studies would need to be conducted.

Different lipid types have been shown to influence export by LolCDE, a complex which is in the same family as MacB (Miyamoto & Tokuda, 2007). Both PE and cardiolipin were found to stimulate LolCDE activity, while PG acted as a suppressor. It was suggested that these lipids

exert their effects physically, with PE and cardiolipin exerting a lateral pressure on the LolCDE transmembrane helix at high Mg^{2+} concentration (Miyamoto & Tokuda, 2007). I could not find a correspondence between lipid interactions in my trajectories and different states of MacB. It is possible that any differential effect of lipids on the complex would only be observed at longer simulation times. My results however do correspond with previous MD work establishing that MacB does not appear to cause local membrane distortion (Rao *et al*, 2020). However, LolCDE with open periplasmic heads and dimerised NBDs did appear to cause significant local distortion, which was abolished after the heads closed (Rao *et al*, 2020).

5. Conclusion

I conducted MD simulations of MacB with and without a trapped internal POPE lipid. I found that the MacB NBDs converted to a “closed” state in both simulations. I found that the periplasmic cavity adopts multiple conformations when POPE lipid is present, suggesting dynamic movement in both the MacB NBDs and periplasmic head region. Using these results, previous structures, and experimental evidence I was able to expand the mechanism of the MacB transportation cycle taking into account the free energy landscape of transport. My simulations also suggest there could be multiple routes of entry into the MacB periplasmic cavity. I was not able to find any biologically significant role of lipid in MacB transport, though this could be due to the short simulation times. My study contributes to our understanding of the function of the MacAB-TolC complex to aid future treatments to antibiotic resistant infections.

Chapter VI: Type I Secretion System Substrate Transport in a Free Energy Landscape

1. Introduction

1.1 Hierarchical Free Energy Landscape

Since the 1980's it has been recognised that proteins are highly dynamic biological machines. Working with structures of ligand-bound and unbound haemoglobin, Ansari *et al* (1985) proposed a "hierarchy" of free energy states to explain dynamic behaviour. At the top of the energy hierarchy, the protein is in its folded, functional state. Moving down the hierarchy leads to lowering of energy barriers and the appearance of "substates" of the protein (**Figure 6.1.1-1**). Dynamic function is then explained by moving across these substates; for example, ligand binding causes small changes at the lowest level in the energy hierarchy which then "ripples" up the energy levels leading to a large conformational change. These "proteinquakes" which lead to functional protein states has broad appeal; it elegantly explains how small perturbations in the atoms of the protein lead to large, collective motions which convey function.

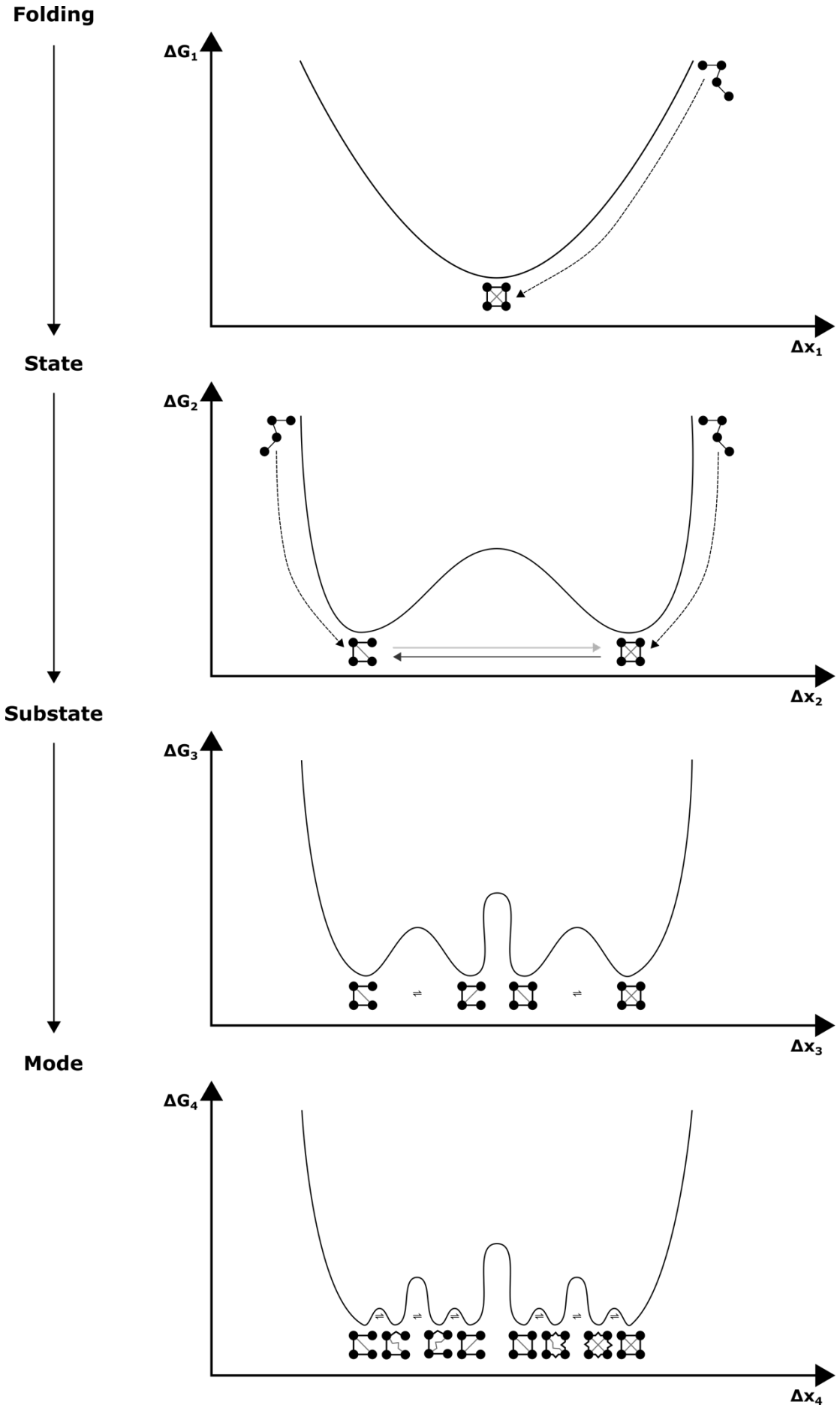

Folding
ΔG1
Δx1
State
ΔG2
Δx2
Substate
ΔG3
Δx3
Mode
ΔG4
Δx4

Figure 6.1.1-1 (previous page). Hierarchical energy landscape. The landscape describes function as produced from a hierarchy of collective motions of the molecule. Folding is at the highest level which confers function via its states in the next energy level. States, in turn, can be described as a collection of substates, and each substate will have a collection of modes. As you go down the hierarchy, the free energy barriers decrease. Function is then a cause of cascading collective motions, where a perturbance in the modes can "ripple up" the energy landscape to cause the functional change in state. Four levels are shown for illustrative purposes.

Itoh & Sasai (2004) applied the "proteinquake" model to understand the photocycling behaviour of photoactive yellow protein. The hierarchical free energy landscape has since been extended and explored (Fraunfelder, Sligar & Wolynes, 1991; Levy & Becker, 2001; Moffat, 2001; Henzler-Wildman & Kern, 2007; Rico & Moy, 2007; Senet *et al*, 2008; Elenewski, Velizhanin & Zwolak, 2019).

1.2 Molecular Motion and the State of Equilibrium

Uncovering dynamic motions in proteins is the goal of molecular dynamics (MD) simulations. Statistical exploration of these dynamics have been investigated using principal component analysis (PCA) (Hess, 2002), root mean square deviation (RMSD) (Kuntz *et al*, 1982; Sargsyan, Grauffel & Lim, 2017), and root mean square fluctuation (RMSF) (Welford, 1962; Keskin, Jernigan & Bahar, 2000; Maguid *et al*, 2006; Fuglebakk, Echave & Reuter, 2012). These methods are descriptive, analogous to using a scalpel to peel back layers of collective motion. At the heart is the idea of protein dynamics existing in a series of microstates, which collectively describe a macroscopic ensemble. In statistical mechanics, a macroscopic state is said to exist in an equilibrium when there is no observed change in the energy between the microstates (Gibbs, 1902). Consider a gas diffusing in a room; as the gas diffuses, it occupies a series of microstates with changing energy until it eventually fills the room. At this point, mixing between the gases is still occurring but there is no observed change in the macroscopic ensemble (**Figure 6.1.2-1**). Overall changes between microstates can be modelled energetically by considering changes in heat (enthalpy) and changes in state order (entropy). At the microscopic level, diffusion dynamics govern the gas particle behaviour; each particle has momentum and velocity, with collisions leading to small heat exchanges. Entropy can be thought of as a measure of the particle ordering: at which position each gas particle occupies in the room at that point in time. The highest entropy level occurs when the first gas is

introduced into the room, and successive microstates are entropically driven to the lowest energy level. From this energetic perspective, we get the idea of sampling the macroscopic ensemble: by observing all possible configurations of microstates. These configurations are "ordered" by a time (entropy) component. In theory, reverse mixing of the two gases could occur, given an infinitely long period of time (Gibbs, 1902).

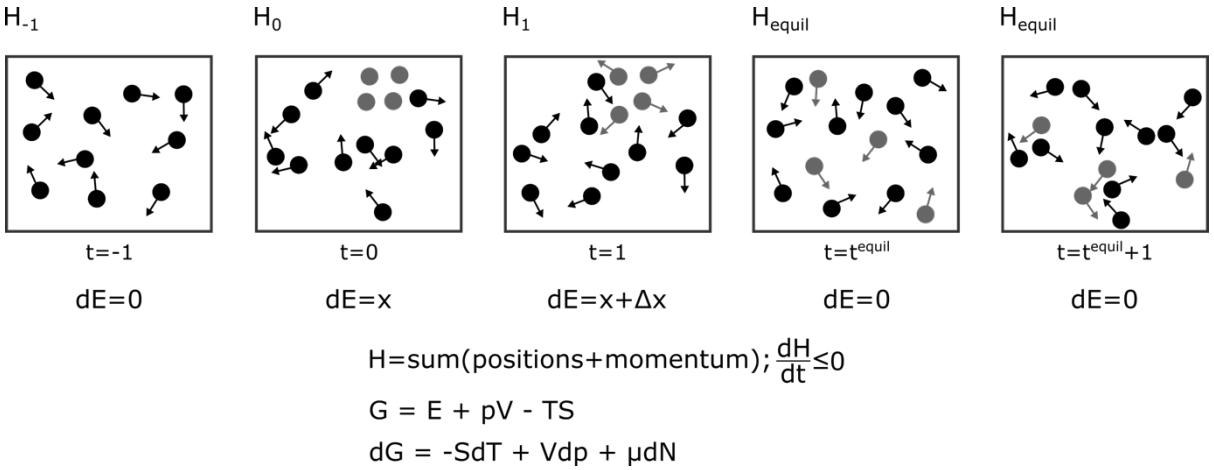


Figure 6.1.2-1. Thermodynamic equilibrium. A gas diffuses in a box – each particle has position and momentum (arrow). At time t=0, another gas is introduced into the box, causing a change in energy of the system. The energy of the box fluctuates until time $t=t_{equil}$, where an equilibrium is achieved, and the energy no longer changes. H ("Hamiltonian") describes all the particle positions and momentums in the box ("the microstate"). Time irreversibility is defined as one microstate following another, based on the previous microstate's observed particle momentums. Particles are assumed to collide elastically with each other and the box. The Gibbs Free Energy (G) is then defined as the sum of the box energy (E), the observed pressure (p) times the box volume (V), and the entropy (S) at a temperature (T). Changes in G is then based on changes to the temperature at entropy (S), changes to pressure on the box from the particle collisions (at constant volume, V), and changes in the number of particles (N) times the particle potential (μ). Explanation adapted from Gibbs (1902).

The idea of proteinquakes leading to changes in protein substates is then a reformulated description of the statistical mechanics described above. A protein state is said to exist in equilibrium when all substates (with submotions) have been observed with no overall changes to the protein macroscopic ensemble. A defining feature of biology, however, is that most of the time things are not in equilibrium; chemical reactions are occurring, substances are being transported, mechanical motion is in place (Astumian, 2012). Although we have a hierarchical description of protein free energy which could explain dynamics, the natural

question is: how does the protein change state? Or: how does changes in protein microstates ripple up the energy landscape to lead to large non-equilibrium changes? This question is non-trivial and several suggestions have been proposed to probe it. The first is to understand the physical first principals: how does perturbations in heat transfer alter dynamics? In proteins, efforts have been made to understand how thermal heat is funnelled through the protein, splitting pathways into short "ballistic" bursts and long-range heat diffusion (*e.g.,* Nadler *et al*, 1987; Li, Magana & Dyer, 2014; Elenewski, Velizhanin & Zwolak, 2019). The second is also physical models, but examining large-scale structure movements, especially in relation to protein allostery: how does changes in one site of the protein affect other sites? On this question, various coupling models have been proposed (*e.g.,* Monod, Wyman & Changeux, 1965; Koshland, Némethy & Filmer, 1966; Hilser, Wrabl & Motlagh, 2012; Weinkam, Pons & Sali, 2012). The third is an examination of protein allostery, but with added evolutionary coupling: what does amino acid conservation at different sites say about potential free energy paths (*e.g.,* Lockless & Ranganathan, 1999)? On the last question, graph-theoretic approaches have been particularly useful (*e.g.,* Böde *et al*, 2007; Amor *et al*, 2014; Amor *et al*, 2016; Mersmann *et al*, 2021). Under this model, atoms in a protein are represented as "nodes" and "edges" define their interactions, split by covalent and non-covalent bonds (Böde *et al*, 2007). Potential allosteric pathways can then be found by considering perturbations in these interactions; subgraphs are made at different scales (to represent different dynamics) and Markov processes modelled to find signal propagation pathways (Amor *et al*, 2014; Amor *et al*, 2016; Mersmann *et al*, 2021).

1.3 Structure Prediction in the Context of the Free Energy Landscape

Utilising biology instead of physical first-principals underlies the success of prediction programmes such as AlphaFold (Jumper *et al*, 2021). Here, physical information is gleamed from multiple-sequence alignments (MSAs) and input experimental structures mined from the Protein Data Bank (PDB). The uncannily accurate predictions (Hegedűs *et al*, 2022; Porta-Pardo *et al*, 2022) is not magic but sophisticated landscape navigation; this is represented as a multi-layer neural network (NN) with nodes and weights ("Evoformer"). Tuning this landscape is achieved by altering the network weights by training it with a large amount of data (Jumper *et al*, 2021). Viewing proteins (and their sequences) as existing in a multi-

dimensional fitness landscape is not new and has been around since the 1970s (Smith, 1970). Navigation of this landscape, however, is a non-trivial problem and has been examined under the guise of various models with differing levels of complexity (*e.g.*, Kauffman & Levin, 1987; Kauffman & Weinberger, 1989; Perelson & Macken, 1995; van Nimwegen, Crutchfield, & Huynen, 1999; Kryazhimskiy, Tkacik & Plotkin, 2009; Nagel *et al*, 2012). Approaching structure-prediction as optimising a graph-based NN can then be seen then as finding the nearest structural minimum for a given input sequence. However, if proteins exist in a series of microstates at equilibrium, sometimes separated by tiny energy barriers, an immediate issue with AlphaFold prediction comes to mind: what is the biological relevance of the output structure? Indeed, for partially unfolded proteins (intrinsically disordered), problems arise with AlphaFold predictions (Perrakis & Sixma, 2021). This is the inevitable consequence of (static) protein prediction on dynamic molecules: the algorithm will output the structure at the nearest minimum it can find, but that does not imply that is the only structure that physically exists. Thus, while AlphaFold is extremely useful for protein structure prediction (at an energy minimum), further work is required to understand the dynamics of the molecule.

Although understanding protein dynamics through physical first-principals remains the goal of many research areas (*e.g.*, Hsueh *et al*, 2022; Moore *et al*, 2022; Chen *et al*, 2023), the success of programmes such as AlphaFold show that it is not necessary to achieve a full physical description to find useful results. Indeed, inputting the biology (through MSAs) can be thought of as a useful heuristic for the underlying physics. Indeed, this has been a noted biological phenomenon for decades; the fact that many proteins share significant structural homology, sometimes with little corresponding sequence conservation hints at the underlying physical driving force (*e.g.,* Shakhnovich, Abkevich & Ptitsyn, 1996; Pearson & Sierk, 2005; He *et al*, 2017).

1.4 Aims & Objectives

Incorporating multiple sources of information (phylogeny, structure, dynamics) is then a powerful method for understanding the biological mechanism of action. In **Chapter II**, I introduced molecular dynamics (MD) simulations of haemolysin B (HlyB), the ABC-transporter component of the Type I Secretion System (T1SS). By examining dynamics with and without

substrate, I was able to tease apart critical interactions, in particular the impact of lipid on protein dynamics. In **Chapter III**, I experimentally tested the hypothesis that cardiolipin aids substrate transport, and found corroboration. In **Chapter V**, I uncovered the dynamics of transport in a related system, the MacAB-TolC efflux pump, and used these dynamics and experimental structures to propose a mechanism of action based on pathways in a free energy landscape. Here, I do the same for the T1SS, incorporating multiple lines of evidence to propose a mechanism of action from a free energy point of view. Furthermore, I propose experiments to test this model by probing free energy states.

The Aims of this Chapter are as follows:

1) What are the dynamics governing T1SS transport (*e.g*. enthalpic and entropic contributions to free energy?)
2) How conserved is the transportation process?
3) How do we test models of transport in relation to free energy?

This will be achieved through the following Objectives:

1) Examining the phylogeny of HlyB and related peptidases/bacteriocins
2) Examining possible allosteric transport pathways through ProteinLens (Mersmann *et al*, 2021)
3) Placing previous MD studies (**Chapter II**) in light of sequence conservation and the results of the allosteric pathway analysis
4) Combining the above points with available experimental structures
5) Combining information (phylogeny, structure, dynamics) with a literature search to propose a mechanism of transport
6) Offer experimental tests of the model as a means to uncover further insights

2. Materials & Methods

2.1 Phylogenetic Analysis

HlyA-like, and related HlyB and peptidase protein sequences were found using the BLAST server (https://blast.ncbi.nlm.nih.gov) (Altschul *et al*, 1990; Altschul *et al*, 1997). *E. coli* HlyA and HlyB, and *Acetivibrio thermocellus ATCC 27405* peptidase amino acid sequences were used as queries on the non-redundant protein sequences database using the blastp (protein-protein BLAST) algorithm. A sample of 25 HlyA-like sequences were aligned in UCSF-Chimera using the default parameters in the "Realign Sequences" option (Petterson *et al*, 2004). For HlyB and peptidase sequence alignment, retrieved sequences were further trimmed manually to have a sample of 150 sequences from a range of organisms. HlyB/peptidase sequences were then aligned using the T-Coffee web-server (https://tcoffee.crg.eu/) (Notredame, Higgins & Heringa, 2000; Tommaso *et al*, 2011). The PSI-TM Coffee option was used for transmembrane proteins using UniRef100 for homology extension. The aligned sequences output was then fed into Phylogeny.fr (http://www.phylogeny.fr/) (Dereeper *et al*, 2008) to make the phylogenetic tree. PhyML 3.1/3.0 aLRT (Guindon *et al*, 2010) was used to compute phylogeny with SH-like Approximate Likelihood Ratio Test to compute branch support, and the Jones-Taylor-Thornton matrix as the substitution model. Other settings were left as default. The output tree was viewed using TreeDyn (Chevenet *et al*, 2006).

2.2 Allosteric Pathway Analysis

Allosteric pathway analysis was computed using ProteinLens (https://www.proteinlens.io) (Mersmann *et al*, 2021). For HlyB, three models from atomistic molecular dynamics trajectories 1–3 sampled at 375 ns were used as input (see **Chapter II**). For peptidase, model 6v9z (Kieuvongngam *et al*, 2020) retrieved from the Protein Data Bank (Berman *et al*, 2000) was used as input. For defining substrate binding site, substrate chains in the respective models were selected (HlyB – chain C; peptidase – chain C/D).

2.3 Sequence Conservation

To visualise sequence conservation, HlyA and HlyB/peptidase aligned sequences were read into Jupyter Notebook (Kluyver *et al*, 2016) using Biopython (Cock *et al*, 2009). Logomaker (Tareen & Kinney, 2019) was then used for visualisation at selected sites. See **Table B** in **Appendix B** for information on the data analysis pipeline I developed.

3. Results

3.1 Phylogeny of HlyB and HlyB-like Structures

To understand the context of HlyB and HlyB-like structures in terms of functional similarity, a phylogenetic tree was calculated for a sample of 150 representative protein sequences. This was further trimmed to 53 sequences for figure and evolutionary clarity (**Figure 6.3.1-1**). HlyB and HlyB-like proteins are known to reside in virulence plasmids in addition to existing in genomes, complicating phylogenetic relationships (Smith & Halls, 1967; Knapp *et al*, 1986; Rawlings & Bateman, 2019). The main issue is that a gene acquired via horizontal gene transfer will show in a phylogeny that the organisms are more closely related than they actually are, in addition to other complicating factors such as the presence of false positives and contaminants (Rawlings & Bateman, 2019). Nevertheless, we can still gain useful insights by how protein sequences cluster, without the need to invoke strict lateral inheritance.

Interestingly, two phylogenetic clusters are formed: a peptidase/bacteriocin branch and a HlyB/HlyB-family branch. Within these branches, the sequences then diverge by Gram-positive and Gram-negative species. This implies that the original split between the two structures occurred before the division of Gram-positive and negative species, which could potentially be billions of years ago (Gupta, 2000; Gupta, 2011; Antunes *et al*, 2016; Hug *et al*, 2016). Intriguingly, the *E. coli* H47 microcin sequence does not fall with the other *E. coli* peptidases and bacteriocin sequence clusters as expected, suggesting that microcins could form a third sequence cluster. This analysis suggests that there was strong evolutionary pressure to maintain multiple copies of these proteins, which then diverged function over time.

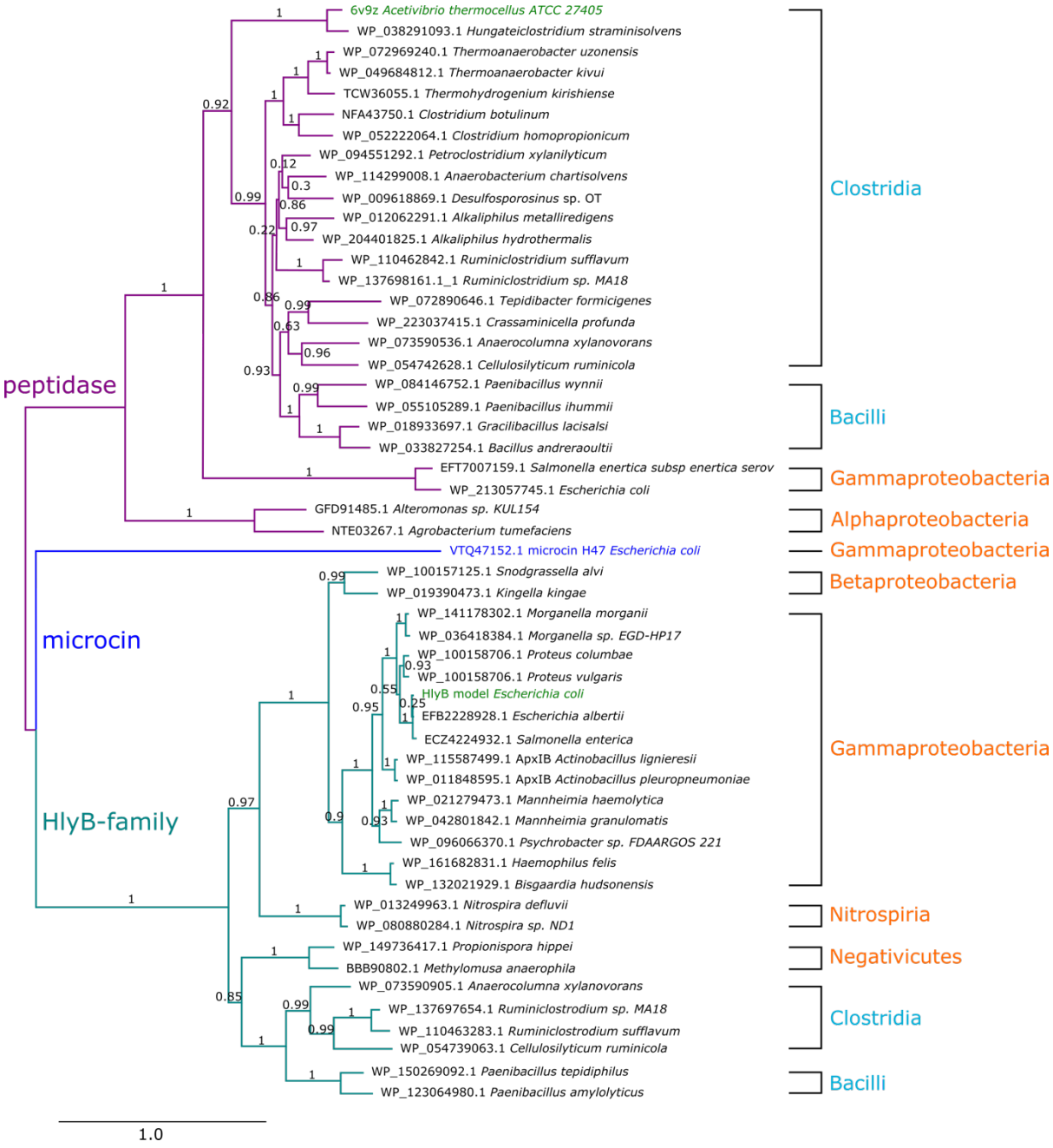


Figure 6.3.1-1. Phylogeny of HlyB-family and peptidases. Representative protein sequences for HlyB-family members (26) and peptidases (26) fall into their own branches, with a further separation by Gram-positive (blue) and Gram-negative (orange) species by class name. Unexpectedly, the *E. coli* H47 microcin did not fall into either the peptidase or the HlyB-family branch, suggesting microcins could form a separate lineage. Sequences for the two structures used in MD simulations (see **Chapter II**) are highlighted in green. Bootstrap values for branches provided by SH-like support via the Phy-ML server (Dereeper *et al*, 2008; Guindon *et al*, 2010). Scale bar shows amino acid substitutions per site.

3.2 Allosteric Pathway Analysis

To examine possible allosteric pathways for substrate binding, both HlyB and peptidase structures were examined by ProteinLens (Mersmann *et al*, 2021). ProteinLens allows for the visualisation of communication away from a queried site. The bond-to-bond propensity is a measure of how likely other residues in a protein will be influenced by energetic changes to the queried site. The strength of connection between the queried site and other residues in the protein is represented as a quantile score, with 0 indicating no coupling to the queried site and 1 indicating complete coupling to the queried site. Transient hotspot analysis is similar to bond-to-bond propensity in that it is also a measure of site communication, except in this case it asks: if I start from the queried site, what is my likely path away from this site in a random walk? In this case, the quantile scoring is a measure of communication speed rather than strength, with 0 indicating no escape and 1 indicating instantaneous communication. Note that in actual proteins, we expect scores to be between 0 and 1 but not to be exactly equal.

The bond-to-bond propensity hotspots clustered around the transmembrane domain for both HlyB and peptidase structures, while the transient hotspot analysis showed clustering of sites at the top of the transmembrane, at the bottom of the NBD, and in the CLD/PEP (**Figure 6.3.2-1**). The transient hotspot analysis seems to match with the main movements observed in the atomistic trajectories; movement from the NBD towards the substrate is simultaneously communicated to the transmembrane domain (see **Chapter II**). This would explain why only one half of HlyB dimer shows transient hotspot residues, as substrate is only bound to one CLD (**Figure 6.3.2-1a**), while in the peptidase structure there is no such discrepancy (**Figure 6.3.2-1b**).

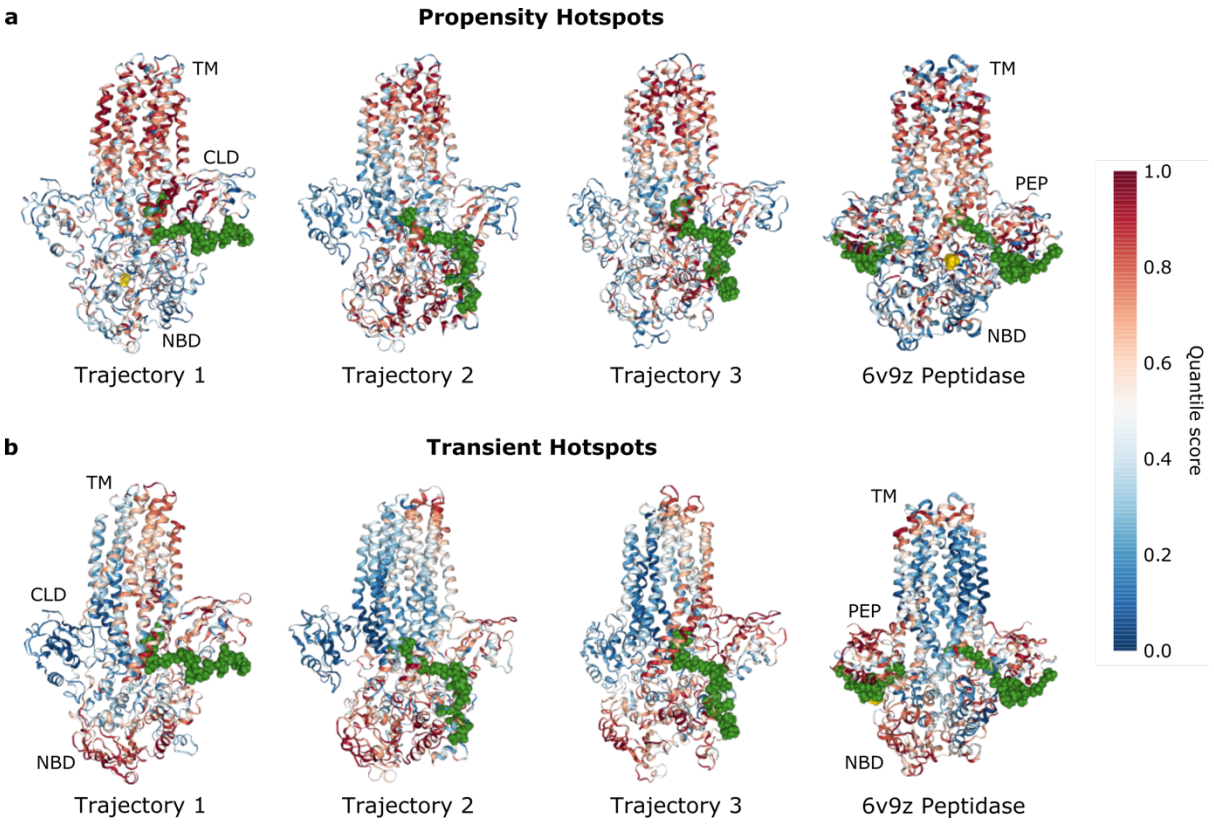


Figure 6.3.2-1. ProteinLens Analysis. ProteinLens (Mersmann *et al*, 2021) analysis results for structures from atomistic trajectories of HlyB and 6v9z peptidase. For each structure, the ligand-binding site was queried (green). Hotspot propensity is an edge-based method for detecting allosteric sites from the queried site, while transient hotspots is a node-based method of site-communication from the queried site. For both HlyB and peptidase structures, the bond-to-bond propensity shows hotspot residues cluster in the transmembrane (TM) domain, while transient hotspots show additional sites in the CLD/PEP and nucleotide-binding domains (NBD). For HlyB, representative structures were taken from the trajectory midpoint (375 ns). Hotspots are coded from a quantile scale of 0 (dark blue) to 1 (dark red), with 1 indicating most significant residues.

3.3 Sequence Conservation at Key Sites

Sequence conservation at key protein-substrate interactions sites found in the MD simulations (see **Chapter II**) was checked across all 150 sequences (**Figure 6.3.3-1**). These residues showed strong sequence conservation, even though overall sequence similarity is low (~30%). In the CLD/PEP domains which bind substrates, very strong sequence conservation was found for a string of lysine residues (**Figure 6.3.3-1a**). In MD simulation, these residues act to stabilise the substrate tail and aid correct positioning in the binding pocket. In peptidases, a catalytic triad is formed between C21, H99, and D115 – these residues

are not conserved in HlyB-family members as they do not need to trim their substrates as required for peptidase function. The residue E79 in HlyB is strongly conserved among HlyB-family members (E/D at this site), but there is no such conservation in peptidases (mutated to G/Q/N/K) (**Figure 6.3.3-1a**). This is probably due to the relative positioning of the loop where it is found (**Figure 6.3.3-1c**). In HlyB, this loop is positioned so that it interacts directly with the substrate, allowing for charge-charge interactions. In peptidases, however, this loop does not directly interact with substrate. In the transmembrane region, striking conservation is found in R247 (K262 in peptidase) and S331 (S346 in peptidase) (**Figure 6.3.3-1b**) which lie at the substrate entrance channel (**Figure 6.3.3-1c**). In simulations these residues form respective critical charge-charge and polar contacts with the substrate. As expected, there is also a pattern of hydrophobic residues in the transmembrane region showing strong conservation. A notable exception is the residue R256 in HlyB which is always R/K in HlyB-family members, but again there is no such pattern in peptidases (mutated to M/S/Q/N). Like R247, this residue could form charge-charge contacts in HlyB-family members which is likely not required for peptidases. In the NBD, very strong conservation is found in the loop region that interacts with the substrate (**Figure 6.3.3-1b/c**). Strikingly, very strong conservation is found in E601 (E618 in peptidase). Polar/charged residues N556 in HlyB, K607 and N617 in peptidase were found in simulation to form polar contacts with the substrate; conservation of these residues thus likely reflects a conserved requirement for polar contacts (**Figure 6.3.3-1b**). In HlyB, R592 can form charge-charge interactions with the substrate in simulation. Interestingly, strong conservation of this residue is only found in other Gammaproteobacteria HlyB-family members, and is otherwise mutated to an inactive proline residue (**Figure 6.3.3-1b**). Thus, this likely reflects a functional requirement for substrates secreted by Gammaproteobacteria.

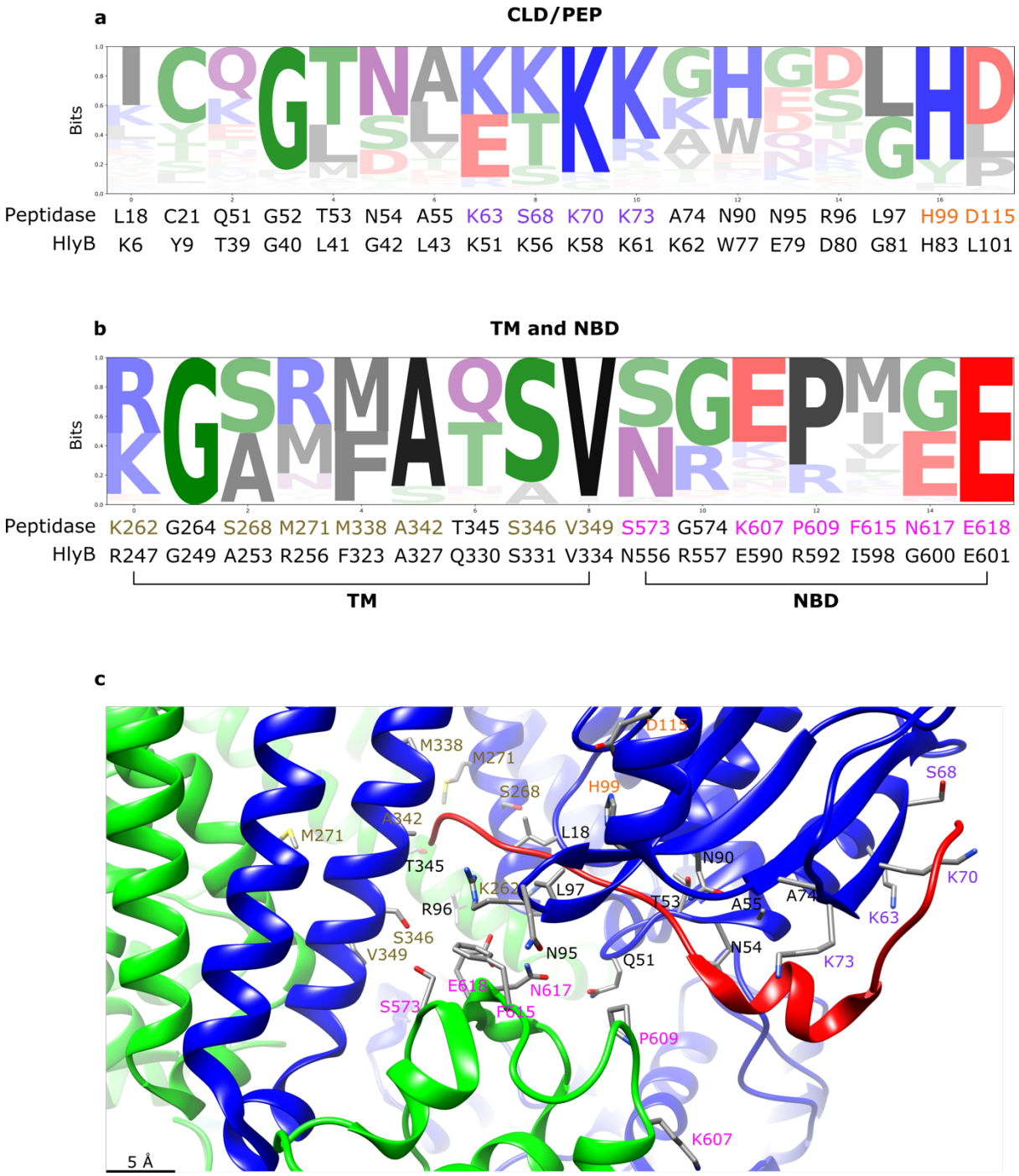
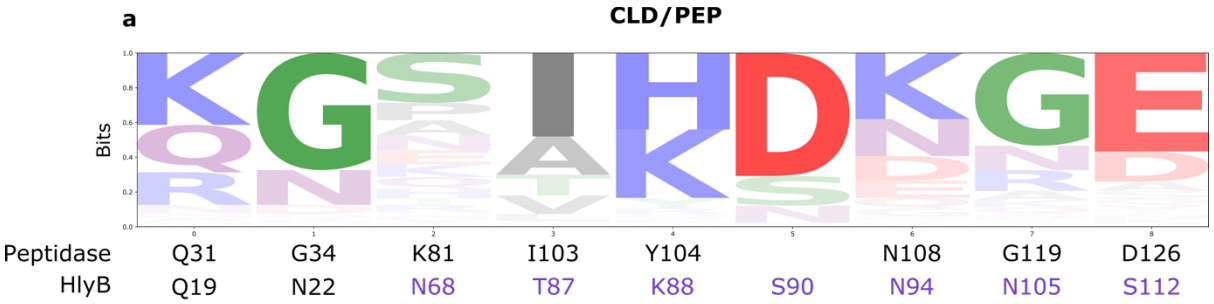


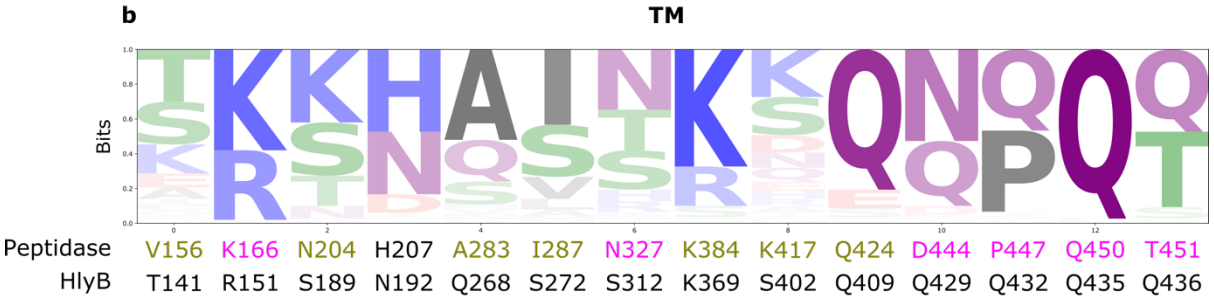
Figure 6.3.3-1. Sequence conservation for key protein-substrate interaction sites. a) Sequence conservation in key residues in the C39-like domain (CLD) and PEP domain. The most conserved amino acids are the K-cluster (purple), which in simulation act to stabilise the tail of the substrate and aid in correct positioning to allow for favourable polar contacts. **b)** Sequence conservation in key residues in the transmembrane domain (TM, gold) and nucleotide-binding domain (NBD, pink). In the TM, conserved residues are the charged R/K (K262) at the transmembrane entrance and a cluster of hydrophobic residues. In the NBD, conserved residues (pink) are found in a critical substrate-interacting loop. **c)** Visualisation of key conserved residues in peptidase structure from *Acetivibrio thermocellum* (pdb ID 6v9z). In the CLD, K tail stabilisation residues are highlighted (purple). Residues

H99 and D115 which form the catalytic triad are shown (orange). Note that C21 (mutated to A21 in 6v9z) is behind H99 (not shown). In the TM, key charged residue K262 and polar residue S346 along with critical hydrophobic residues are highlighted (gold). Note in HlyB, R256 is in place of peptidase M271. In the NBD, key residues lie in a loop region which can interact with the substrate (pink).

Sequence conservation for HlyB key protein-lipid interactions was calculated for all 150 sequences (**Figure 6.3.3-2**). Residues in the CLD/PEP domains show a high-degree of sequence conservation (**Figure 6.3.3-2a**). In the substrate unbound state, it is likely these residues could form charged/polar contacts with residues in the neighbouring protomers NBD and also with HlyD arm regions. However, when substrate is bound there could be a switch to form more contacts with lipid (**Figure 6.3.3-2c**). In simulation, HlyB residues N68, T87, and K88 (PEP residues K81, I103, and Y104) form a preference for cardiolipin when substrate is present. In the HlyB/D complex, these residues form interdomain contacts and are positioned away from the lipid. However, if substrate is present these residues can interact with lipid at the PEP equivalent sites (**Figure 6.3.3-2c**). Certain key residues in the HlyB transmembrane arm and top region could also interact with HlyD (**Figure 6.3.3-2b/c**). In simulation, HlyB residues T141, Q268, K369, and S402 (peptidase residues V156, A283, K384, and K417) were found to interact more with cardiolipin when substrate was present. Mapping these residues onto the HlyB/D complex shows they are in close contact with HlyD arm and lipolyl regions (**Figure 6.3.3-2c**). Interestingly, in simulation R151 and Q435 (peptidase K166 and Q450) showed a decrease in close cardiolipin interactions when substrate was present; these two amino acids are very conserved in both HlyB and peptidases (**Figure 6.3.3-2b**). In the unliganded HlyB/D complex, these amino acids are too far (> 9 $\mathrm{\AA}$) to interact with the HlyD arm (**Figure 6.3.3-2c**). However, in ATP-bound HlyB/D complex, R151 is in close contact (< 5 $\mathrm{\AA}$) with R26 and D30 on the HlyD arm. In simulation HlyB substrate entrance channel residues S312, Q429, Q432, Q435, and Q436 (peptidase residues N327, D444, P447, Q450, and T451) showed either an increase or decrease in POPE/POPG interactions. Of these, Q429, Q432, Q435, and Q436 are the most conserved (**Figure 6.3.3-2b**). In HlyB-family members, Q432 and Q436 is strictly conserved while in peptidases this is mutated respectively to P447 and T451 (**Figure 6.3.3-2b**). The positioning of these residues at the substrate entrance channel in the HlyB/D complex, but too far way to interact directly with HlyD is intriguing (**Figure 6.3.3-2c**). Thus, it is possible that these residues, along with key CLD lipid interacting residues, do allow for a

particular lipid architecture to form around the CLD with substrate entrance channel of the TM as found in simulation.





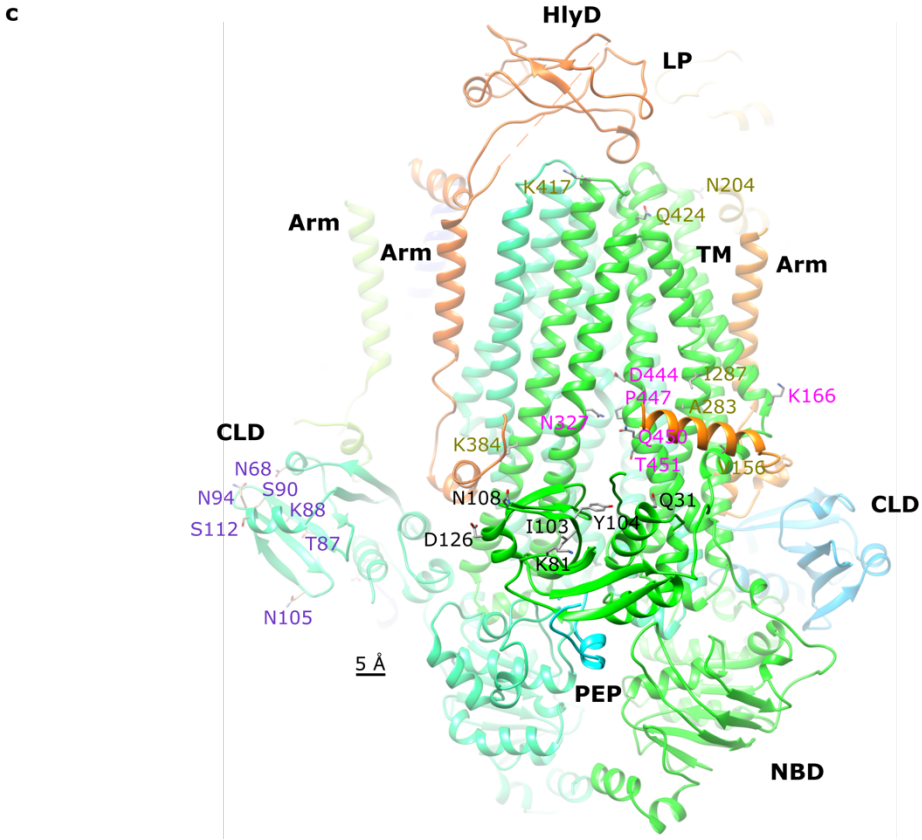


Figure 6.3.3-2 (previous page). Sequence conservation for key protein-lipid interactions. a) Sequence conservation in key residues in the C39-like domain (CLD) and PEP domain. In the unbound state, CLD domain residues (purple) can interact with HlyD arm regions and nucleotide-binding domain (NBD) of the neighbouring protomer. When substrate is bound, the domain is rearranged to accommodate protein-lipid interactions (black). **b)** Sequence conservation in key residues in the transmembrane (TM) domain. Residues in the transmembrane arm and top (gold) can interact with HlyD arm and lipolyl domains. However, residues in the substrate entrance channel and transmembrane (pink) arm are likely to interact with lipid. **c)** Mapping of conserved residues onto the HlyB-HlyD structure (pdb ID 7sgr) (Zhao, Lee & Chen, 2022) with peptidase protomer B (green) with non-translocating substrate (cyan) (pdb ID 6v9z) (Kieuvongngam *et al*, 2020).

4. Discussion

4.1 Relation of Peptidases and HlyB-Family Members in their Evolutionary Context

Phylogenetic analysis placed HlyB-family members as a separate branch from peptidases/bacteriocins and microcins. Within each branch, a separation by Gram-positive and negative species is observed (see **Figure 6.3.1-1**). The current model of gene diversification posits that gene duplication is a prerequisite for diverging function (Ohno, 1971). Once duplicated, the new gene can then undergo selection for novel function (Bershtein & Tawfik, 2008). Peptidases/bacteriocins are remarkably diverse, secreting a range of antimicrobial peptide substrates (Gonzales & Robert-Baudoy, 1996; Paetzel, Dalbey & Strynadka, 1998; Chavan & Riley, 2007; Paetzel, 2014). Peptidase/bacteriocin and microcin diversity seems to reflect ecological niches occupied by bacterial species (Sosunov *et al*, 2007; Nguyen, Myrold & Mueller, 2019; Cole *et al*, 2022). Indeed, my phylogenetic analysis reflects a range of species including ones found in harsh environments (*e.g.* anaerobic *Thermoanaerobacter* and *Thermohydrogenium* species isolated from hot springs (Zacharova *et al*, 1993; Wagner *et al*, 2008b; Hess *et al*, 2014), *Petroclostridium xylanilyticum* isolated from an oilfield in China (Zhang *et al*, 2018), and cold-adapted *Psychrobacter* species (Welter *et al*, 2021). Additionally, I found members of both peptidases/bacteriocins and HlyB-family members in the same species (*e.g. Escherichia coli*, *Salmonella enterica*, *Anaerocolumna xylanovorans*, *Ruminiclostridium sufflavum*, *Ruminiclostridium sp. MA18*, *Cellulosilyticum ruminicola*). Indeed, considering the phylogenetic separation of peptidases/bacteriocins, microcins, and HlyB-family members, it is likely that HlyB-family members evolved from peptidases/microcins and acquired mutations which allowed them to secrete larger substrates without proteolytic processing. It is known that both peptidases (Rawlings & Bateman, 2019) and type I secretion system elements (Smith & Halls, 1967; Knapp *et al*, 1986) can be acquired via horizontal gene transfer which further complicates their phylogeny. Nevertheless, considering the diverse range of organisms which contain peptidases it is likely that these were the functional units in the universal last common ancestor (Rawlings & Bateman, 2019).

4.2 Structure of Peptidases and HlyB-Family Members in Relation to their Function

Previously, experimental structure determination of the *E. coli* HlyB substrate-binding domain found that it was structurally very similar to peptidases – thus it was named "C39-like domain" (CLD) after the C39 proteases (Lecher *et al*, 2012). The main difference between the two lies in the fact that peptidases need to cleave their substrates in order for them to be active after transport (Håvarstein, Diep & Nes, 1995; Ishii *et al*, 2010). This takes place via its PEP ("peptidase") domain, consisting of a conserved catalytic triad (cysteine, histidine, and aspartic acid). For substrates, a conserved double-glycine motif acts to position the sequence for cleavage by the cysteine residue (Schnell *et al*, 1988; Håvarstein, Diep & Nes, 1995; van Belkum, Worobo & Stiles, 1997). Although HlyB-family members and peptidases only share ~30% sequence identity, the structural similarity is very striking. This should not be too surprising considering the phylogenetic relationship between the two (see above). The cryo-EM structure of the peptidase from *Acetivibrio thermocellus* ATCC 27405 (formally *Clostridium thermocellum*) ("PCAT1") with bound substrate was deposited in 2020 (Kieuvongngam *et al*, 2020); this was used as the basis for my HlyB homology model, with the HlyB CLD modelled using the NMR structure solved in 2012 (Lecher *et al*, 2012). The recently solved structure for the HlyB/D complex (Zhao, Lee & Chen, 2022) confirms that the two structures are very conserved in the architecture of their transmembrane domains and NBD; as expected, the main uncertainty lies in the exact positioning of the HlyB CLD. In HlyB/D complex, one CLD in a protomer interacts with the neighbouring protomer, while the other is structurally flexible. Structural flexibility in CLD/PEP domains without substrate has been noted in previous unliganded PCAT1 crystal structure (Lin, Huang & Chen, 2015). In unliganded HlyB/D complex, only one protomer has an open transmembrane helix for possible substrate entry. This becomes closed in all three protomers once ATP is bound (Zhao, Lee & Chen, 2022).

My HlyB homology model with bound substrate offers a prediction for CLD positioning when substrate is present. Since contraction of the transmembrane helices is observed when ATP is bound, caused by the NBD dimerisation, it is unfeasible for ATP-hydrolysis to be driven by the protomer which binds and translocates the substrate. Instead, ATP-hydrolysis must occur in the neighbouring protomers to drive translocation. It has been found that substrate secretion of HlyA is very slow; around 16.0 ± 1.3 amino acids could be translocated per T1SS

per second (Lenders *et al*, 2016). In light of the HlyB/D complex (Zhao, Lee & Chen, 2022), the slow rate could reflect the fact that more than one ATP cycle is required for transport. Indeed, due to the large nature of the substrate being transported (HlyA is 1024 amino acids long), and the fact that the NBDs undergo dimerisation when ATP binds leading to a closure of the transmembrane helices at the channel opening, it is very unlikely that the substrate-binding HlyB protomer undergoes ATP-hydrolysis. In addition, photo-cross linking experiments have demonstrated that substrate is not transported through the central pore of the complex, but must go through via the CLD and transmembrane domain of one protomer (Zhao, Lee & Chen, 2022). This could explain the requirement for HlyB coupling in a trimer which is not found in peptidases. Since peptidases cleave their substrates, NBD dimerisation upon ATP-binding would not preclude transport from the same protomer (Kieuvongngam & Chen, 2022).

4.3 Sequence Conservation and Substrate Transport

In my previous MD simulation study for *E. coli* HlyB and *A. thermocellus* peptidase with and without substrate, I found a cluster of residues responsible for substrate recognition and binding. These seemed to facilitate favourable charge-charge and polar contacts within the CLD and PEP binding pockets (see **Chapter II**). Comparison of these residues in light of their sequence conservation found a cluster of conserved K-residues, which in both HlyB and peptidase simulation acted to stabilise the substrate tail. In HlyB/D complex, lysine residues K56, K58, and K62 interact with the negatively charged arm region of HlyD in the neighbouring protomer (Zhao, Lee & Chen, 2022). In addition, there is a cluster of conserved charged and hydrophobic residues in the transmembrane domain and a highly conserved loop in the NBD which shows some sequence divergence between peptidases and HlyB-family members (see **Figure 6.3.3-1**). Some of these residues have been identified in photo-crosslinking experiments of the HlyB/D complex: R256, F323, and A327 were identified previously (Zhao, Lee & Chen, 2022). In PCAT, residue A55 was also previously identified as involved in substrate recognition (Kieuvongngam *et al*, 2020). However, for both HlyB and PCAT no NBD residues were identified as being critical for substrate transport though this is likely due to the cross-linking experiments used to identify critical residues which may impact ATPase function (Kieuvongngam *et al*, 2020; Zhao, Lee & Chen, 2022). Intriguingly, in simulation it was found that non-translocating PCAT1 peptide substrate loses contact with NBD residue N617, which

is part of the critical NBD loop region (see **Figure 6.3.3-1**). The proximity of the NBD substrate-binding loop to the ATP-binding region in both PCAT peptidase and HlyB/D complex is interesting; conserved residue E590 in HlyB is ~8 Å from E610 which forms part of the catalytic complex in dimerised ATP-bound state (pdb ID 8dck) (Zhao, Lee & Chen, 2022), while equivalent residue K607 in PCAT1 is ~6.5 Å from Q627 (pdb ID 7t54) (Kieuvongngam & Chen, 2022).

Sequence conservation for protein-lipid interacting residues in HlyB may partly be explained by preferable interactions with HlyD in the HlyB/D complex. However, there are lipid interacting residues in the substrate entrance channel which are too far away to interact directly with HlyD. In *E. coli* HlyB, a string of four conserved glutamine residues lies at the substrate entrance channel. In peptidases, only two of the four glutamine residues are conserved. These differences between HlyB-family members and peptidases may reflect formation of a particular lipid architecture as found in simulation for HlyB. It is possible that critical cardiolipin interacting residues on the HlyB CLD also aid in formation of this architecture, as in simulation the CLD with bound substrate was found to promote membrane interaction. As discussed in **Chapter II** and **Chapter III**, cardiolipin does seem to be required for transport. The role of cardiolipin in relation to the free energy landscape of transport will be discussed in the next section.

4.4 Free Energy Landscape of Transport for T1SS

Several structures for PCAT peptidase were solved with bound substrate and ATP (Kieuvongngam & Chen, 2022). Unlike in MacAB-TolC transport, simultaneous substrate and ATP-binding appears to be possible, although the NBDs are not fully dimerised although substrate does appear to enter the cavity. The PCAT1 structures are shown in **Figure 6.4.4-1**.

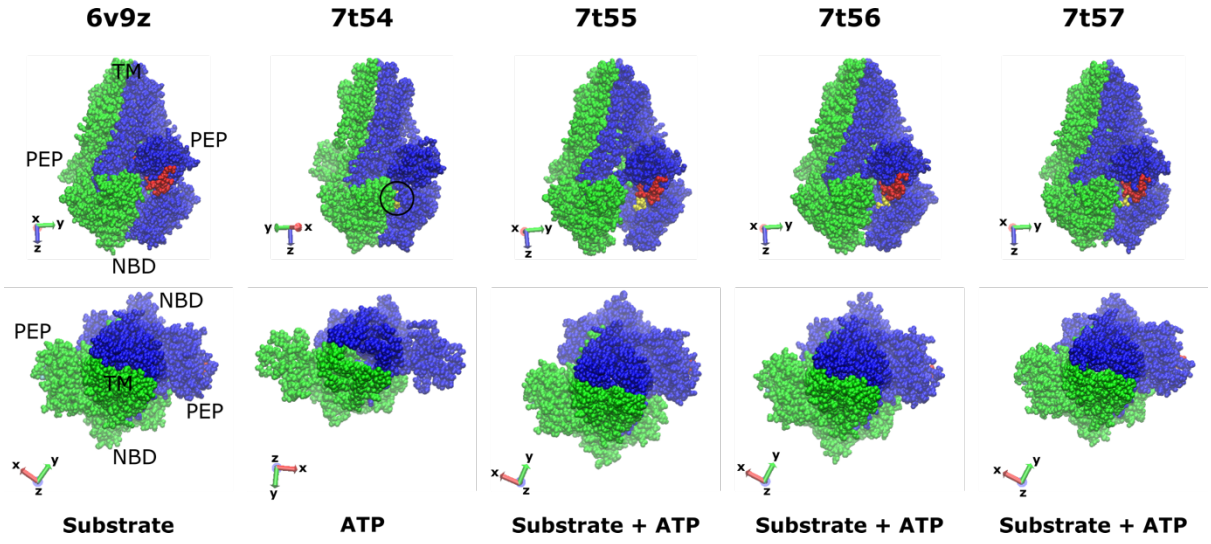


Figure 6.4.4-1. Peptidase PCAT1 Structures. Structures for PCAT1 have been solved in an array of conformations. With bound substrate (red, cyan) but no ATP (yellow), the nucleotide-binding domain (NBD) dimers are closed by not fully dimerised (pdb ID 6v9z) (Kieuvongngam *et al*, 2020). With ATP alone, full dimerisation of the NBDs can take place. This allows the transmembrane domain (TM) to form an opening (pdb ID 7t54). With both bound substrate and ATP, NBDs can exist in a variety of fully open (pdb ID 7t55) to more closed states (pdb IDs 7t56 and 7t57) (Kieuvongngam & Chen, 2022). Note that structure 6v9z was used for MD simulations in **Chapter II**.

Without substrate, ATP-binding causes complete dimerisation of the NBDs which travels up to the transmembrane domain and allows for a pore opening. With substrate, several structures are possible, with and without ATP-binding, although there is no complete NBD dimerisation and the transmembrane pore remains closed (**Figure 6.4.4-1**). Thus, for transport to occur in peptidases with ATP-hydrolysis, cleavage must take place which then allows for complete NBD dimerisation, which transmits to the transmembrane domain to allow pore opening. Hydrolysis may then induce a further conformational change to push the substrate out of the cavity, with release of ADP+P_i products then resetting the structure for a further round of transport. The kinetic control seems to be the presence of Mg^{2+} ions which promote ATP-hydrolysis (Kieuvongngam & Chen, 2022).

Transient hotspot analysis did indicate a strong link between PEP domains, NBD, and the upper transmembrane region (see **Figure 6.3.2-1b**), suggesting there is interdomain communication. In the context of microstates and free energy sampling, it is likely that in the absence of Mg^{2+} to drive ATP-hydrolysis, many structures are sampled in various states of

NBD dimerisation; switching between these states would be possible through low energetic barriers as the structure samples its energy distribution via atomic motions (modes to substates, see **Figure 6.1.1-1**). When Mg^{2+} is introduced, a change of state is possible through the energy release of ATP-hydrolysis, at a particular substate (in this case, fully dimerised NBDs with bound substrate).

In the context of T1SS transport, the following must be accounted for: 1) Substrate transport likely occurs through a single protomer, with the other two protomers providing energy via ATP-hydrolysis (HlyB/D structure with bound ATP analogue (pdb ID 8dck) (Zhao, Lee & Chen, 2022), 2) There may be energy via a proton-motive force (Koronakis, Hughes & Koronakis, 1991), with cardiolipin likely playing a role (see **Chapter II** and **III**), and 3) There is a minimum sequence requirement of 60 amino acids, which roughly corresponds to the length of the Gram-negative cell envelope (Jarchau *et al*, 1994). The main mystery is where the proton-motive force fits into the transportation process (Bumba *et al*, 2016). Recently, a paper was published on the Sec machinery which proposed that a proton-motive force could be generated via lysine and arginine deprotonation of the substrate. Reprotonation of these residues as the substrate exits the machinery would then provide the complete proton-motive force for transport (Allen *et al*, 2022). Intriguingly, I found a string of lysine and arginine residues throughout the *E. coli* HlyA sequence and related Repeat in Toxin (RTX) substrate sequences (**Figure 6.4.4-2**).

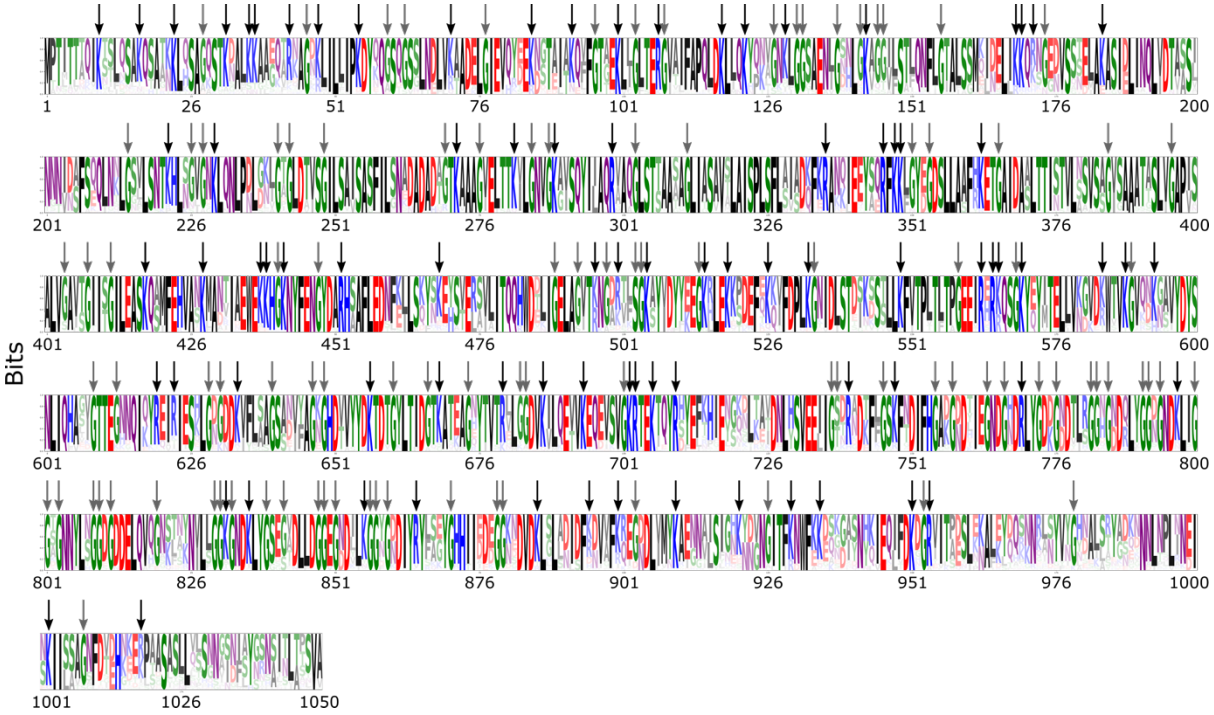


Figure 6.4.4-2. Sequence conservation for HlyA-like substrates. HlyA-like substrates show a pattern of conserved lysine and arginine residues (black arrow) throughout their sequences. Conserved GG-motifs (grey arrow) are also shown. Residues highlighted when conservation is greater or equal to 0.8 bits. See Sequence Conservation Analysis Notebook available on GitHub (https://github.com/AMKCam/Sequence-Conservation-Analysis) [last accessed 15th October 2023] and Table B in Appendix B for a description of the Notebook.

In atomistic simulations, I also found a tendency for the HlyA substrate to be exposed to hydrophobic residues in the pocket leading to tail destabilisation; hydrophobicity promotes lysine and arginine deprotonation of the substrate in the SecYEG mechanism (Allen *et al*, 2022). If this is the case for the T1SS, then this deprotonation and reprotonation mechanism would provide the "Brownian ratchet" of the proton-motive force, and not GG-motif Ca^{2+} binding. The role of Ca^{2+} in the transportation process remains contradictory; although it promotes folding (Holland *et al*, 2016; Spitz *et al*, 2019), tying it to a proposed "Brownian ratchet" motion in transport has remained elusive (Lenders *et al*, 2015; Lenders *et al*, 2016). Indeed, Bumba *et al* (2016) found an electrophoretic effect for CyA secretion, where the presence of negative charges throughout the protein sequence promotes secretion. This may reflect the balancing out of positive K and R residues also present in RTX sequences, as can be seen in sequence conservation analysis (**Figure 6.4.4-2**). Thus the observed electrophoretic effect could be explained by the model proposed by Allen *et al* (2022), where deprotonation

would make an overall local negative charge for the part of the sequence being transported. Considering the relation of HlyB-family to peptidases, whose substrate also contain a GG-motif (for substrate cleavage), it could also be the case that the GG-motif in HlyB-family members underwent neofunctionalization (Rastogi & Liberles, 2005; Deng *et al*, 2010) to promote protein folding upon substrate exit. Thus, the role for cardiolipin could be to act as a proton sink for the deprotonated lysine and arginine residues. A proposed mechanism is provided in **Figure 6.4.4-3**.

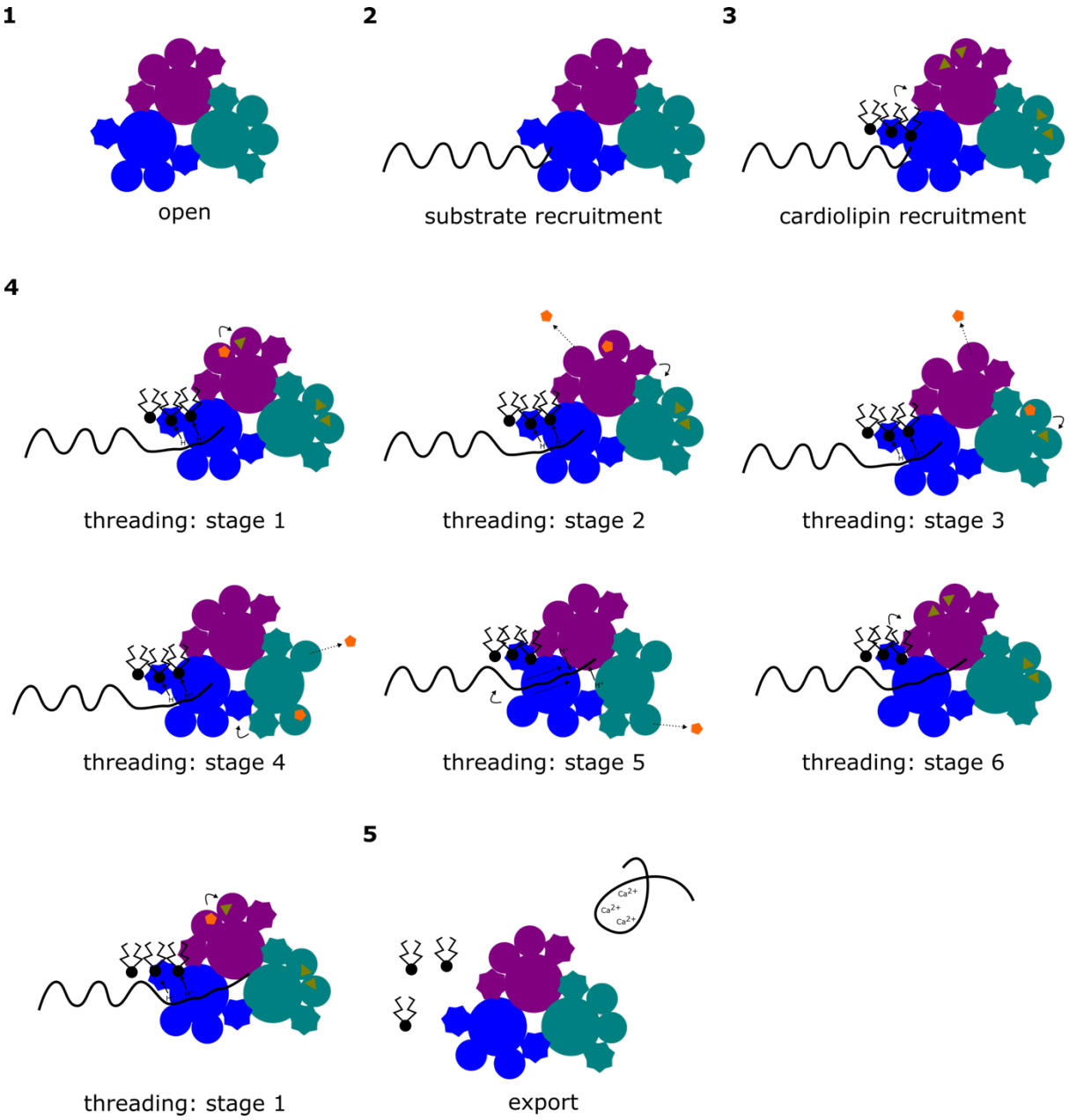


Figure 6.4.4-3 (previous page). Proposed mechanism of transport for the T1SS. Three HlyB dimers assemble with HlyD (not shown). In the open state (1), one dimeric HlyB can recruit substrate (2). Substrate recruitment triggers concomitant recruitment of cardiolipin to the substrate-bound protomer's substrate entrance channel and CLD (hexagon) as well as TolC (not shown). This recruitment is transmitted to the next protomer which starts to stimulate substrate threading (4). In the first stage, the substrate is partially destabilised which promotes deprotonation of lysine and arginine residues in the binding pocket. Protons are shuttled to the recruited cardiolipin head groups. ATP-hydrolysis of the neighbouring HlyB nearby protomer NBD (small circle) is also stimulated, catalysing ATP (triangle, gold) to ADP+P_i products (pentamer, orange). This stimulates the second stage, where atomic movement which is transmitted to its dimeric half, stimulating its ATP-hydrolysis and causing movement to be transmitted to the next protomer. Movements cascade until reaching the protomer with bound substrate, creating work to move the substrate. As the substrate exits, it is reprotonated which creates a proton motive force. The overall work is retransmitted to the next protomer and the cycle starts again until the substrate is fully exported (5).

The proposed mechanism accounts for the following observations are made: 1) substrate-binding can only occur via one HlyB dimeric protomer (HlyB/D complex, pdb ID 7sgr), 2) substrate binding stimulates formation of a cardiolipin architecture around the substrate-bound HlyB (**Chapter II** and **III**), 3) when substrate is bound this creates a polarisation between the dimeric halves which could potentially create a directional signalling-pathway (transient hotspot analysis), 4) for work to be created in order to move the substrate, the neighbouring protomers must catalyse ATP (HlyB/D complex, pdb ID 8dck), 4) ATP-binding and hydrolysis occurs on dimerised NBDs (literature and HlyB/D complex), 5) work to move the substrate also comes from a proton motive force, in this case deprotonation of substrate lysine and arginine residues followed by reprotonation upon substrate exit (*e.g.,* Allen *et al*, 2022; Koronakis, Hughes & Koronakis, 1991), and 6) presence of bound substrate starts an ATP catalysis cascade ("threading") which functions to perform work required to move the substrate. Exactly how the work is transferred from the ATP-hydrolysing dimers to the substrate-exporting dimer is speculative without a structure of a trapped, substrate exporting state. However, based on available structures I would speculate that work could be transferred from the catalytic domains to the CLD and transmembrane via the transmembrane elbow helix which sits adjacent to the Q-loop (see **Chapter I**). Since each CLD acts as a trimer bridge, stimulation of the CLD would be transmitted to the neighbouring NBD to start another cycle.

4.5 Experimental Tests of the Proposed Mechanism

The above proposed mechanism offers some predictions which could be tested by experiment. Overall, I predict around 4 ATP cycles are required to move a portion of substrate (one threading cycle). If we consider the positions of the charged lysine and arginine residues, then a portion moved could equate to 20–30 amino acids. Since the minimum sequence length for transport is around 60 amino acids (Jarchau *et al*, 1994), then at least 2–3 threading cycles are required per proton motive force, which would equate to 8–12 ATP cycles in total. Thus, full-length HlyA would need around 36–54 threading cycles to be exported, equalling 144–216 ATP cycles. We might consider this an overestimate, since it is possible that a full 60 amino acids can be transported per threading cycle, which would impose a lower limit of (17x4=) 72 ATP cycles. A coupling assay could be performed to test the number of ATP cycles per substrate transported, as completed for MacAB-TolC (see Souabni *et al*, 2021). Concomitantly, testing the role of the trimer could be achieved by mixing stalled (NBD H662A mutated) to unstalled (wild-type) HlyB protomers. If only two of the three dimers are catalytically active, then by mixing a 2:4 ratio of inactive to active HlyB should not change the resulting substrate export and ATP-hydrolysis on average (on the assumption that mutating the catalytic unit won't affect inter-domain communication). However, if a ratio of 4:2 or 3:3 is mixed, we would expect to see a substantial decrease in substrate export and concomitant ATP-hydrolysis activity on average. This would also allow us to test the role of the NBD loop (see **Figure 6.3.3-1**) in sensing substrate presence – if we again mix NBD loop mutants with wild type, then all ratios (2:4, 4:2, or 3:3) should show a decrease in substrate export, but not ATP-hydrolysis activity.

Capturing intermediates during transport would also be feasible and allow for comparison to the proposed mechanism. T1SS complex could be purified and incubated with ATP, with and without substrate and Mg^{2+} as performed for PCAT1 peptidase (Kieuvongngam & Chen, 2022). Plunge-freezing and reconstruction by cryo-electron microscopy (cryo-EM) could then be performed to determine the variety of structures present under these conditions. This is likely to be quite difficult considering the size of the complex if protein purification is performed. Another method may be to keep the complex in the cell, or in a lipoproteosome and perform a similar experiment but apply reconstruction by cryo-electron tomography (cryo-ET).

As a very interesting but extremely challenging study would be to perform ancestral sequence reconstruction (Jermann *et al*, 1995) on peptidase and HlyB-family members. This approach relies on inferring ancestral protein sequences through statistical methods at different branching-points in the evolutionary history of the protein family in question. It has been used to address a variety of questions which is difficult to address via standard biochemical and structural studies, for example the emergence of enzyme substrate promiscuity, receptor-binding specificity, and protein fluorescence (Merkl & Sterner, 2016). Finding a plausible ancestral sequence too far back in evolutionary history is likely not feasible, however there are a few branch points which would be of interest to explore: 1) The last common ancestor separating the Gammaproteobacteria, particularly those which later became relevant as disease-causing organisms, and 2) The last common ancestors for Clostridia peptidases and HlyB-family members as these organisms are known to inhabit harsh environments (see **Discussion** section **4.1**). A complicating feature is that substrates for peptidases and HlyB-family members have also undergone co-evolution with their respective transporters. Thus, ancestral sequence reconstruction would also have to be done on the substrates. This would open up exciting possibilities to test the function of these ancestral substrates and could also indicate how these proteins could potentially be modified for therapeutic (*e.g.* natural antibiotic production) or synthetic biology applications (*e.g.* vaccine design).

5. Conclusion

I proposed a mechanism for T1SS transportation based on previous simulation and experimental work completed in **Chapters II** and **III**, along with further bioinformatic analysis (this chapter) and a literature search. This mechanism applies our understanding of protein function via a series of hierarchical energy states: energy from ATP-hydrolysis ripples up the energy hierarchy, creating changes in modes that ultimately lead to a change in state. In the case of the T1SS, presence of substrate leads to a series of such changes that creates a "threading cascade" and allows for substrate transport via ATP-hydrolysis and the proton motive force. A role for cardiolipin is proposed in which it acts as a proton-sink for deprotonated lysine and arginine residues; reprotonation of these residues upon exit from the cell then creates the required proton-motive force. I also offer some experimental tests for the proposed mechanism; this would involve an attempt to capture transport intermediates and solve their structures by single-particle cryo-EM or cryo-ET, and also determining the number of ATP-cycles per substrate transported. I also propose ancestral sequence reconstruction experiments as a possible proof-of-principle method for potentially producing peptidase and HlyB-family proteins for biotechnological applications.

References

Adcock, S.A. & McCammon, J.A. 2006. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. *Chemical Reviews*. 106(5): 1589–615. doi: 10.1021/cr040426m.

Adrian, M., Dubochet, J., Lepault, J., McDowall, A.W. 1984. Cryo-electron microscopy of viruses. *Nature*. 308: 32-6. doi: 10.1038/308032a0.

Akhtar, A.A. & Turner, D.P.J. 2022. The role of bacterial ATP-binding cassette (ABC) transporters in pathogenesis and virulence: Therapeutic and vaccine potential. *Microbial Pathogenesis*. 171: 105734. doi: 10.1016/j.micpath.2022.105734.

Al-Amoudi, A., Norlen, L.P.O. & Dubochet, J. 2004. Cryo-electron microscopy of vitreous sections of native biological cells and tissues. *Journal of Structural Biology*. 148(1): 131-5. doi: /10.1016/j.jsb.2004.03.010.

Al-Amoudi, A., Studer, D. & Dubochet, J. 2005. Cutting artefacts and cutting process in vitreous sections for cryo-electron microscopy. *Journal of Structural Biology*. 150(1): 109-21. doi: 10.1016/j.jsb.2005.01.003.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. 2002. *Molecular Biology of the Cell*. 4th ed. New York: Garland Science.

Alder, B.J. & Wainwright, T.E. 1957. Phase Transition for a Hard Sphere System. *The Journal of Chemical Physics*. 27: 1208-9. doi: 10.1063/1.1743957.

Allen, W.J., Corey, R.A., Watkins, D.W., Oliveira, A.S.F., Hards, K., Cook, G.M., Collinson, I. 2022. Rate-limiting transport of positively charged arginine residues through the Sec-machinery is integral to the mechanism of protein secretion. *eLife*. 11: e77586. doi: 10.7554/eLife.77586.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W Lipman, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology*. 215: 403-10. doi: 10.1016/S0022-2836(05)80360-2.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. 25: 3389-402. doi: 10.1093/nar/25.17.3389.

Amor, B., Yaliraki, S.N., Woscholski, R., Barahona, M. 2014. Uncovering allosteric pathways in caspase-1 using Markov transient analysis and multiscale community detection. *Molecular Biosystems*. 10: 2247-58. doi: 10.1039/C4MB00088A.

Amor, B.R.C, Schaub, M.T., Yaliraki, S.N., Barahona, M. 2016. Prediction of allosteric sites and mediating interactions through bond-to-bond propensities. *Nature Communications*. 7: 12477. doi: 10.1038/ncomms12477.

Andersen, C., Hughes, C., Koronakis, V. 2000. Chunnel vision: Export and efflux through bacterial channel-tunnels. *EMBO Reports*. 1(41): 313-18.

Ansari, A., Berendzen, J., Bowne, S.F., Fraunfelder, H., Iben, I.E.T., Sauke, T.B., Shyamsunder, E., Young, R.D. 1985. Protein states and proteinquakes. *Proceedings of the National Academy of Sciences of the United States of America*. 82(15): 5000-4. doi: 10.1073/pnas.82.15.5000.

Antunes, L.C.S., Poppleton, D., Klingl, A., Criscuolo, A., Dupuy, B., Brochier-Armanet, C., Beloin, C., Gribaldo, S. 2016. Phylogenomic analysis supports the ancestral presence of LPS-outer membranes in the Firmicutes. *eLife*. 5: e14589. doi: 10.7554/eLife.14589.001.

Antimicrobial Resistance Collaborators. 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet*. 399(10325): 629-55. doi: 10.1016/S0140-6736(21)02724-0.

Astumian, R.D. 2012. Microscopic reversibility as the organizing principle of molecular machines. *Nature Nanotechnology*. 7(11): 684-8. doi: 10.1038/nnano.2012.188.

Aqvist, J., van Gunsteren, W.F., Leijonmarck, M., Tapia, O. 1985. A molecular dynamics study of the C-terminal fragment of the L7/L12 ribosomal protein. Secondary structure motion in a 150 picosecond trajectory. *Journal of Molecular Biology*. 183(3): 461-77. doi: 10.1016/0022-2836(85)90014-2.

Balakrishnan, L., Hughes, C. & Koronakis, V. 2001. Substrate-triggered Recruitment of the TolC Channel tunnel during Type I Export of Hemolysin by Escherichia coli. *Journal of Molecular Biology*. 313(3): 501-510. doi: 10.1006/jmbi.2001.5038.

Barrera, N.P., Isaacson, S.C., Zhou, M., Bavro, V.N., Welch, A., Schaedler, T.A., Seeger, M.A., Miguel, R.N., *et al*. 2009. Mass spectrometry of membrane transporters reveals subunit stoichiometry and interactions. *Nature Methods*. 6(8): 585-7. doi: 10.1038/nmeth.1347.

Barrett, M.P., Walmsley, A.R. & Gould, G.W. 1999. Structure and function of facilitative sugar transporters. *Current Opinion in Cell Biology*. 11(4): 496-502. doi: 10.1016/s0955-0674(99)80072-6.

Baumeister, W. 2005. Electron tomography: towards visualizing the molecular organization of the cytoplasm. *Current Opinion in Structural Biology*. 12(5): 679-84. doi: 10.1016/S0959-440X(02)00378-0.

Beckers, M. & Sachse, C. 2020. Permutation testing of Fourier shell correlation for resolution estimation of cryo-EM maps. *Journal of Structural Biology*. 212(1): 107579. doi: 10.1016/j.jsb.2020.107579.

Beckstein, O., Denning, E.J., Perilla, J.R., Woolf, T.B. 2009. Zipping and Unzipping of Adenylate Kinase: Atomistic Insights into the Ensemble of Open↔Closed Transitions. *Journal of Molecular Biology*. 394(1):160–76: 00107. doi:10.1016/j.jmb.2009.09.009.

Bei, W., Luo, Q., Shi, H., Zhou, H., Zhou, M., Zhang, X., Huang, Y. 2022. Cryo-EM structures of LolCDE reveal the molecular mechanism of bacterial lipoprotein sorting in *Escherichia coli*. *PLoS Biology*. 20(10): e3001823. doi: 10.1371/journal.pbio.3001823.

Bekker, H., Dijkstra, E.J., Renardus, M.K.R., Berendsen, H.J.C. 1995. An Efficient, Box Shape Independent Non-Bonded Force and Virial Algorithm for Molecular Dynamics. *Molecular Simulation*. 14(3): 137-51. doi: 10.1080/08927029508022012.

Benabdelhak, H., Kiontke, S., Horn, C., Ernst, R., Blight, M.A., Holland, I.B., Schmitt, L. 2003. A Specific Interaction Between the NBD of the ABC- transporter HlyB and a C-Terminal Fragment of its Transport Substrate Haemolysin A. *Journal of Molecular Biology*. 327(5): 1169-79. doi: 10.1016/s0022-2836(03)00204-3.

Benz, R., Maier, E., Ladant, D., Ullmann, A., Sebo, P. 1994. Adenylate cyclase toxin (CyaA) of *Bordetella pertussis*. Evidence for the formation of small ion-permeable channels and comparison with HlyA of *Escherichia coli. Journal of Biological Chemistry*. 269(44): 27231-9. doi: 10.1016/S0021-9258(18)46973-6.

Berendsen, H.J.C. , Postma, J.P.M. , van Gunsteren, W.F., DiNola, A. , Haak, J.R. 1984. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics*. 81: 3684-90. doi: 10.1063/1.448118.

Berendsen, H.J.C., van der Spoel, D. & Drunen, R. 1995. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*. 91: 43-56. doi: 10.1016/0010-4655(95)00042-E.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Research*. 28: 235-42. doi: 10.1093/nar/28.1.235.

Bershtein, S, Tawfik, DS. 2008. Ohno's Model Revisited: Measuring the frequency of potentially adaptive mutations under various mutational drifts. *Molecular Biology and Evolution*. 25(11): 2311-8. doi: 10.1093/molbev/msn174.

Best, R.B., Zhu, X., Shim, J., Lopes, P.E.M., Mittal, J., Feig, M., MacKerell, A.D. 2012. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone φ, ψ and Side-Chain χ1 and χ2 Dihedral Angles. *Journal of Chemical Theory & Computation*. 8(9): 3257–73. doi: 10.1021/ct300400x.

Bharat, T.A.M. & Scheres, S.H.W. 2016. Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nature Protocols*. 11: 2054-65. doi: 10.1038/nprot.2016.124.

Blight, M.A., Pimenta, A.L., Lazzaroni, J.C., Dando, C., Kotelevets, L., Séror, S.J., Holland, I.B. 1994. Identification and preliminary characterization of temperature-sensitive mutations affecting HlyB, the translocator required for the secretion of haemolysin (HlyA) from *Escherichia coli*. *Molecular & General Genetics*. 245(4): 431-40. doi: 10.1007/BF00302255.

Böde, C., Kovács, I.A., Szalay, M.S., Palotai, R., Korcsmáros, T., Csermely, P. 2007. Network analysis of protein dynamics. *FEBS Letters*. 581(15): 2776-82. doi: 10.1016/j.febslet.2007.05.021.

Böhm, J., Frangakis, A., Hegerl, R., Nickell, S., Typke, D., Baumeister, W. 2000. Toward detecting and identifying macromolecules in a cellular context: Template matching applied to electron tomograms. *Proceedings of the National Academy of Sciences of the United States of America*. 97(26): 14245–50. doi: 10.1073/pnas.230282097.

Bolobova, A.V., Zhukov, A.V. & Klyosov, A.A. 1994. Lipids and fatty acids in cellulosomes of *Clostridium thermocellum*. *Applied Microbiology and Biotechnology*. 42: 128-33. doi: /10.1007/BF00170235.

Bountra, K., Hagelueken, G., Choudhury, H.G., Corradi, V., Omari, K.E., Wagner, A., Mathavan, I., Zirah, S., *et al*. 2017. Structural basis for antibacterial peptide self-immunity by the bacterial ABC transporter McjD. *EMBO Journal*. 36(20): 3062-79. doi: 10.15252/embj.201797278.

Bradbury, A. & Plückthun, A. 2015. Reproducibility: Standardize antibodies used in research. *Nature*. 517: 27-9. doi: 10.1038/518027a.

Bradt, R.N., Johnson, S.M. & Karlin, S. 1956. On Sequential Designs for Maximizing the Sum of n Observations. *Annals of Mathematical Statistics*. 27(4): 1060-74. doi: 10.1214/aoms/1177728073.

Braun, E., Gilmer, J., Mayes, H.B., Mobley, D.L., Monroe, J.I., Prasad, S., Zuckerman, D.M. 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.

Brini, E., Fennell, C.J., Fernandez-Serra, M., Hribar-Lee, B., Lukšič, M., Dill, K.A. 2017. How Water's Properties Are Encoded in Its Molecular Structure and Energies. *Chemical Reviews*. 117(19): 12385–414. doi: 10.1021/acs.chemrev.7b00259.

Brooks, B.R., Brooks III, C.L., MacKerell Jr, A.D., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., *et al*. 2009. CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry*. 30: 1545-614. doi: 10.1002/jcc.21287.

Brown, G.K. 2000. Glucose transporters : Structure, function and consequences of deficiency. *Journal of Inherited Metabolic Disease*. 23: 237-46. doi: 10.1023/A:1005632012591.

Bublitz, M., Morth, J.P. & Nissen, P. 2011. P-type ATPases at a glance. *Journal of Cell Science*. 124(15): 2515–9. doi: 10.1242/jcs.088716.

Bumba, L., Masin, J., Macek, P., Wald, T., Motlova, L., Bibova, I., Klimova, N., Bednarova, L., *et al*. 2016. Calcium-Driven Folding of RTX Domain β-Rolls Ratchets Translocation of RTX Proteins through Type I Secretion Ducts. *Molecular Cell*. 62(1): 47-62. doi: 10.1016/j.molcel.2016.03.018.

Bunduc, C.M., Fahrenkamp, D., Wald, J., Ummels, R., Bitter, W., Houben, E.N.G., Marlovits, T.C. 2021. Structure and dynamics of a mycobacterial type VII secretion system. *Nature*. 593: 445–8. doi: 10.1038/s41586-021-03517-z.

Burt, A., Gaifas, L., Dendooven, T., Gutsche, I. 2021. A flexible framework for multi-particle refinement in cryo-electron tomography. *PLoS Biology*. 19(8): e3001319. doi: 10.1371/journal.pbio.3001319.

Bussi, G., Donadio, D. & Parrinello, M. 2007. Canonical sampling through velocity rescaling. *Journal of Chemical Physics*. 126(1): 014101. doi: 10.1063/1.2408420.

Carpenter, E.P., Beis, K., Cameron, A.D., Iwata, S. 2008. Overcoming the challenges of membrane protein crystallography. *Current Opinion in Structural Biology*. 18(5): 581–6. doi: 10.1016/j.sbi.2008.07.001.

Castaño-Díez, D., Kudryashev, M., Arheit, M., Stahlberg, H. 2012. Dynamo: A flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. *Journal of Structural Biology*. 178(2): 139-51. doi: 10.1016/j.jsb.2011.12.017.

Caves, L.S.D., Evanseck, J.D. & Karplus, M. 1998. Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin. *Protein Science*. 7: 649-66. doi: 10.1002/pro.5560070314.

Chavan, M.A. & Riley, M.A. 2007. Molecular Evolution of Bacteriocins in Gram-Negative Bacteria. In: *Bacteriocins: Ecology and Evolution*. M.A. Riley and M.A Chavan, Eds. Berlin, Heidelberg: Springer-Verlag. 19-43. doi: 10.1007/978-3-540-36604-1_3.

Chen, M., Dai, W., Sun, S.Y., Jonasch, D., He, C.Y., Schmid, M.F., Chiu, W., Ludtke, S.J. 2017. Convolutional Neural Networks for Automated Annotation of Cellular Cryo-Electron Tomograms. *Nature Methods*. 14(10): 983–5. doi: 10.1038/nmeth.4405.

Chen, M., Bell, J.M., Shi, X., Sun, S.Y., Wang, Z., Ludtke, S.J. 2019. A complete data processing workflow for cryo-ET and subtomogram averaging. *Nature Methods*. 16(11): 1161-8. doi: 10.1038/s41592-019-0591-8.

Chen, S.J., Hassan, M., Jernigan, R.L., Jia, K., Kihara, D., Kloczkowski, A., Kotelnikov, S., Kozakov, D., *et al*. 2023. Protein folds vs. protein folding: Differing questions, different challenges. *Proceedings of the National Academy of Sciences of the United States of America*. 120(1): e2214423119. doi: 10.1073/pnas.2214423119.

Chevenet F., Brun C., Banuls A.L., Jacq B. and R. Christen. 2006. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics*. 7: 439. doi: 10.1186/1471-2105-7-439.

Choudhury, H.G., Tong, Z., Mathavan, I., Li, Y., Iwata, S., Zirah, S., Rebuffat, S., van Veen, H.W., Beis, K. 2014. Structure of an antibacterial peptide ATP-binding cassette transporter in a novel outward occluded state. *Proceedings of the National Academy of Sciences of the United States of America*. 111(25): 9145-50. doi: 10.1073/pnas.1320506111.

Christie, P.J. 2019. The Rich Tapestry of Bacterial Protein Translocation Systems. *Protein Journal.* 38(4): 389-408. doi: 10.1007/s10930-019-09862-3.

Claus, S., Jezierska, S. & van Bogaert, I.N.A. 2019. Protein-facilitated transport of hydrophobic molecules across the yeast plasma membrane. *FEBS Letters*. 593(13): 1508-27. doi: 10.1002/1873-3468.13469.

Cock P.A., Antao T., Chang J.T., Chapman B.A., Cox C.J., Dalke A., Friedberg I., Hamelryck T., *et al*. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*. 25: 1422-3. doi: 10.1093/bioinformatics/btp163.

Cole, T.J., Parker, J.K., Feller, A.L., Wilke, C.O., Davies, B.W. 2022. Evidence for Widespread Class II Microcins in Enterobacterales Genomes. *Applied and Environmental Microbiology*. 88(23): 1-19. doi: 10.1128/aem.01486-22.

Corey, R.A., Pyle, E., Allen, W.J., Watkins, D.W., Casiraghi, M., Miroux, B., Arechaga, I., Politis, A., Collinson, I. 2018. Specific cardiolipin–SecY interactions are required for proton-motive force stimulation of protein secretion. *Proceedings of the National Academy of Sciences of the United States of America*. 115(31): 7967–72. doi: 10.1073/pnas.172153611.

Corey, R.A., Song, W., Duncan, A.L., Ansell, T.B., Sansom, M.S.P., Stansfeld, P.J. 2021. Identification and assessment of cardiolipin interactions with E. coli inner membrane proteins. *Science Advances*. 7(34): eabh2217. doi: 10.1126/sciadv.abh2217.

Crow, A., Greene, N.P., Kaplan, E., Koronakis, V. 2017. Structure and mechanotransmission mechanism of the MacB ABC transporter superfamily. *Proceedings of the National Academy of Sciences of the United States of America*. 114(47): 12572–7. doi: 10.1073/pnas.1712153114.

Danev, R., Yanagisawa, H. & Kikkawa, M. 2019. Cryo-Electron Microscopy Methodology: Current Aspects and Future Directions. *Trends in Biochemical Sciences*. 44(10): 837-48. doi: 10.1016/j.tibs.2019.04.008.

Darby, E.M., Trampari, E., Siasat, P., Gaya, M.S., Alav, I., Webber, M.A., Blair, J.M.A. 2022. Molecular mechanisms of antibiotic resistance revisited. *Nature Reviews Microbiology*. (Online). doi: 10.1038/s41579-022-00820-y.

Darden, T., York, D. & Pedersen, L. 1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *Journal of Chemical Physics*. 98: 10089-92. doi: 10.1063/1.464397.

Davies, J. & Davies, D. 2010. Origins and Evolution of Antibiotic Resistance. *Microbiology and Molecular Biology Reviews*. 74(3): 417-33. doi: 10.1128/MMBR.00016-10.

De Vivo, M., Masetti, M., Bottegoni, G., Cavalli, A. 2016. Role of Molecular Dynamics and Related Methods in Drug Discovery. *Journal of Medicinal Chemistry*. 59: 4035-61. doi: 10.1021/acs.jmedchem.5b01684.

Deng, C., Cheng, C.H.C., Ye, H., He, X., Chen, L. 2010. Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. *Proceedings of the National Academy of Sciences of the United States of America*. 107 (50): 21593-8. doi: 10.1073/pnas.1007883107.

Deng, D., Xu, C., Sun, P.C., Wu, J.P., Yan, C.Y., Hu, M.X., Yan, N. 2014. Crystal structure of the human glucose transporter GLUT1. *Nature*. 510(7503): 121-5. doi: 10.1038/nature13306.

Deng, Y., Chen, Y., Zhang, Y., Wang, S., Zhang, F., Sun, F. 2016. ICON: 3D reconstruction with 'missing-information' restoration in biological electron tomography. *Journal of Structural Biology*. 195(1): 100-12. doi: 10.1016/j.jsb.2016.04.004.

Deng, W., Marshall, N.C., Rowland, J.L., McCoy, J.M., Worrall, L.J., Santos, A.S., Strynadka, N.C.J., Finlay, B.B. 2017. Assembly, structure, function and regulation of type III secretion systems. *Nature Reviews Microbiology*. 15: 323–7. doi: 10.1038/nrmicro.2017.20.

Dereeper A., Guignon V., Blanc G., Audic S., Buffet S., Chevenet F., Dufayard J.F., Guindon S., *et al.* 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*. 36(Web Server issue):W465-9. doi: 10.1093/nar/gkn180.

DeRosier, D.J. & Klug, A. 1968. Reconstruction of Three Dimensional Structures from Electron Micrographs. *Nature*. 217: 130-4. doi: 10.1038/217130a0.

Ding, G., Liu, Y., Zhang, R., Xin, H.L. 2019. A joint deep learning model to recover information and reduce artifacts in missing-wedge sinograms for electron tomography and beyond. *Scientific Reports*. 9: 12803. doi: 10.1038/s41598-019-49267-x.

Ditzler, M.A., Otyepka, M., Šponer, J., Walter, N.G. 2010. Molecular Dynamics and Quantum Mechanics of RNA: Conformational and Chemical Change We Can Believe In. *Accounts of Chemical Research*. 43(1): 40–7. doi: 10.1021/ar900093g.

Dolinsky, T.J., Nielsen, J.E., McCammon, J.A., Baker, N.A. 2004. PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research*. 32: W665-7. doi: 10.1093/nar/gkh381.

Doyle, D.A. & Wallace, B.A. 1997. Crystal Structure of the Gramicidin/Potassium Thiocyanate Complex. *Journal of Molecular Biology*. 266(5): 963-77. doi: 10.1006/jmbi.1996.0837.

Drew, D. & Boudker, O. 2016. Shared Molecular Mechanisms of Membrane Transporters. *Annual Review of Biochemistry*. 85: 543-72. doi: 10.1146/annurev-biochem-060815-014520.

Dumon-Seignovert, L., Cariot, G. & Vuillard, L. 2004. The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expression and Purification*. 37(1): 203-6. doi: 10.1016/j.pep.2004.04.025.

Durrant, J.D & McCammon, J.A. 2011. Molecular dynamics simulations and drug discovery. *BMC Biology*. 9: 71. doi: 10.1186/1741-7007-9-71.

Dyla, M., Kjærgaard, M., Poulsen, H., Nissen, P. 2020. Structure and Mechanism of P-Type ATPase Ion Pumps. *Annual Review of Biochemistry*. 89: 583-603. doi: 10.1146/annurev-biochem-010611-112801.

Ebbensgaard, A.E., Løbner-Olesen, A. & Frimodt-Møller, J. 2020. The Role of Efflux Pumps in the Transition from Low-Level to Clinical Antibiotic Resistance. *Antibiotics*. 9: 855. doi: 10.3390/antibiotics9120855.

Elenewski, J.E., Velizhanin, K.A. & Zwolak, M. 2019. Topology, landscapes, and biomolecular energy transport. *Nature Communications*. 10(1): 4662. doi: 10.1038/s41467-019-12700-w.

England, C.G., Ehlerding, E.B., Cai, W. 2016. NanoLuc: A Small Luciferase Is Brightening Up the Field of Bioluminescence. *Bioconjugate Chemistry*. 27(5): 1175-87. doi: 10.1021/acs.bioconjchem.6b00112.

Fabbri, A., Falzano, L., Frank, C., Donelli, G., Matarrese, P., Raimondi, F., Fasano, A., Fiorentini, C. 1999. *Vibrio parahaemolyticus* Thermostable Direct Hemolysin Modulates Cytoskeletal Organization and Calcium Homeostasis in Intestinal Cultured Cells. *Infection and Immunity*. 67(3): 1139–48. doi: 10.1128/iai.67.3.1139-1148.1999.

Famelis, N., Rivera-Calzada, A., Degliesposti, G., Wingender, M., Mietrach, N., Skehel, J.M., Fernandez-Leiro, R., Böttcher, B., *et al*. 2019. Architecture of the mycobacterial type VII secretion system. *Nature*. 576: 321–5. doi: 10.1038/s41586-019-1633-1.

Faruqi, A.R. & Henderson, R. 2007. Electronic detectors for electron microscopy. *Current Opinion in Structural Biology.* 17(5): 549-55. doi: 10.1016/j.sbi.2007.08.014.

Felmlee, T. & Welch, R.A. 1988. Alterations of amino acid repeats in the *Escherichia coli* hemolysin affect cytolytic activity and secretion. *Proceedings of the National Academy of Sciences of the United States of America*. 85(14): 5269–73. doi: 10.1073/pnas.85.14.5269.

Ferguson, A.D. & Deisenhofer, J. 2004. Metal Import through Microbial Membranes. *Cell*. 116(1): 15-24. doi: 10.1016/S0092-8674(03)01030-4.

Fernandez-de-Cossio-Diaz, J. & Vazquez, A. 2018. A physical model of cell metabolism. *Scientific Reports*. 8: 8349. doi: 10.1038/s41598-018-26724-7.

Fitzpatrick, A.W.P., Llabrés, S., Neuberger, A., Blaza, J.N., Bai, X.C., Okada, U., Murakami, S., van Veen, H.W., *et al*. 2017. Structure of the MacAB–TolC ABC-type tripartite multidrug efflux pump. *Nature Microbiology*. 2: 17070. doi: 10.1038/nmicrobiol.2017.70.

Frank, J. 2006. Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State. United States of America: Oxford University Press. doi: 10.1093/acprof:oso/9780195182187.001.0001.

Fraunfelder, H., Sligar, S.G. & Wolynes, P.G. 1991. The Energy Landscapes and Motions of Proteins. *Science*. 254(5038): 1598-603. doi: 10.1126/science.1749933.

Freedman, J.C. 2012. Chapter 4 - Ionophores in Planar Lipid Bilayers. In *Cell Physiology Source Book (Fourth Edition)*. N. Sperelakis, Ed. Academic Press. 61-6. doi: 10.1016/B978-0-12-387738-3.00004-4.

Freedman, L.P. 2015. Antibodies: Validate recombinants too. *Nature*. 518: 483. doi: 10.1038/518483c.

Fuglebakk, E., Echave, J. & Reuter, N. 2012. Measuring and comparing structural fluctuation patterns in large protein datasets. *Bioinformatics*. 28(19): 2431-40. doi: 10.1093/bioinformatics/bts445.

Furse, S., Wienk, H., Boelens, R., de Kroon, A.I.P.M., Killian, J.A. 2015. *E. coli* MG1655 modulates its phospholipid composition through the cell cycle. *FEBS Letters*. 589(19): 2726-30. doi: 10.1016/j.febslet.2015.07.043.

Galaz-Montoya, J.G., Hecksel, C.W., Baldwin, P.R., Wang, E., Weaver, S.C., Schmid, M.F., Ludtke, S.J., Chiu, W. 2016. Alignment Algorithms and Per-Particle CTF Correction for Single Particle Cryo-Electron Tomography. *Journal of Structural Biology*. 194(3): 383–94. doi: 10.1016/j.jsb.2016.03.018.

Gao, X., Lu, F., Zhou, L., Dang, S., Sun, L., Li, X., Wang, J., Shi, Y. 2009. Structure and Mechanism of an Amino Acid Antiporter. *Science*. 324(5934): 1565-8. doi: 10.1126/science.1173654.

Genheden, S. & Ryde, U. 2012. Will molecular dynamics simulations of proteins ever reach equilibrium? *Physical Chemistry Chemical Physics*. 14: 8662-77. doi: 10.1039/C2CP23961B.

Gershenson, A., Gierasch, L.M., Pastore, A., Radford, S.E. 2014. Energy landscapes of functional proteins are inherently risky. *Nature Chemical Biology*. 10(11): 884-91. doi: 10.1038/nchembio.1670.

Gibbs, J.W. 1902. *Elementary Principles in Statistical Mechanics*. Cambridge, USA: John Wilson & Son.

Gillet, J.N. 2022. From molecular dynamics to quantum mechanics of misfolded proteins and amyloid-like macroaggregates applied to neurodegenerative diseases. *Journal of Molecular Graphics and Modelling*. 110: 108046. doi: 10.1016/j.jmgm.2021.108046.

Goebel, W. & Schrempf, H. 1971. Isolation and Characterization of Supercoiled Circular Deoxyribonucleic Acid from Beta-Hemolytic Strains of *Escherichia coli*. *Journal of Bacteriology.* 106(2): 311-17. doi: 10.1128/jb.106.2.311-317.1971.

Gold, V.A.M., Robson, A., Bao, H., Romantsov, T., Duong, F., Collinson, I. 2010. The action of cardiolipin on the bacterial translocon. *Proceedings of the National Academy of Sciences of the United States of America*. 107(22): 10044–9. doi: 10.1073/pnas.0914680107.

Gómez-Márquez, J. 2021. What is life?. *Molecular Biology Reports*. 48(8): 6223-30. doi: 10.1007/s11033-021-06594-5.

Gonzales, T. & Robert-Baudoy, J. 1996. Bacterial aminopeptidases: Properties and functions. *FEMS Microbiology Reviews*. 18(4): 319-44. doi: 10.1111/j.1574-6976.1996.tb00247.x.

Gorovits, N. & Charron, M.J. 2006. What we know about facilitative glucose transporters: Lessons from cultured cells, animal models, and human studies. *Biochemistry and Molecular Biology Education*. 31(3): 163-72. doi: 10.1002/bmb.2003.494031030227.

Gowers, R.J., Linke, M., Barnoud, J., Reddy, T.J.E, Melo, M.N., Seyler, S.L., Domański, J., Dotson, D.L., *et al*. 2016. MDAnalysis: A Python package for the rapid analysis of molecular dynamics simulations. In S. Benthall and S. Rostrup, editors, Proceedings of the 15th Python in Science Conference, pages 98-105, Austin, TX. SciPy. doi:10.25080/Majora-629e541a-00e.

Gray, L ., Mackman, N., Nicaud, J.M., Holland, I.B. 1986. The carboxy-terminal region of haemolysin 2001 is required for secretion of the toxin from *Escherichia coli*. *Molecular and General Genetics*. 205: 127-133. doi: 10.1007/BF02428042.

Gray, L., Baker, K., Kenny, B., Mackman, N., Haigh, R., Holland, I.B. 1989. A novel C-terminal signal sequence targets Escherichia coli haemolysin directly to the medium. *Journal of Cell Science Suppl*. 11: 45-57. doi: 10.1242/jcs.1989.supplement_11.4.

Greene, N.P., Kaplan, E., Crow, A., Koronakis, V. 2018. Antibiotic Resistance Mediated by the MacB ABC Transporter Family: A Structural and Functional Perspective. *Frontiers in Microbiology*. 9: 950. doi: 10.3389/fmicb.2018.00950.

Grimm, R., Singh, H., Rachel, R., Typke, D., Zillig, W., Baumeister, W. 1998. Electron Tomography of Ice-Embedded Prokaryotic Cells. *Biophysical Journal*. 74(2): 1031-42. doi: 10.1016/S0006-3495(98)74028-7.

Grisshammer, R. 2006. Understanding recombinant expression of membrane proteins. *Current Opinion in Biotechnology*. 17(4): 337-40. doi: 10.1016/j.copbio.2006.06.001.

Grossfield, A., Feller, S.E & Pittman, M.C. 2007. Convergence of Molecular Dynamics Simulations of Membrane Proteins. *PROTEINS: Structure, Function, and Bioinformatics*. 67: 31–40. doi: 10.1002/prot.21308.

Guan, L., Mirza, O., Verner, G., Iwata, S., Kaback, H.R. 2007. Structural Determination of Wild-Type Lactose Permease. *Proceedings of the National Academy of Sciences of the United States of America*. 104(39): 15294-8. doi: 10.1073/pnas.0707688104.

Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*. 59(3): 307-21. doi: 10.1093/sysbio/syq010.

Gumbart, J., Aksimentiev, A., Tajkhorshid, E., Wang, Y., Schulten, K. 2005. Molecular dynamics simulations of proteins in lipid bilayers. *Current Opinion in Structural Biology*. 15(4): 423-31. doi: 10.1016/j.sbi.2005.07.007.

Gumbart, J.C., Ferreira, J.L., Hwang, H., Hazel, A.J., Cooper, C.J., Parks, J.M., Smith, J.C., Zgurskaya, H.I., *et al*. 2021. Lpp positions peptidoglycan at the AcrA-TolC interface in the AcrAB-TolC multidrug efflux pump. *Biophysical Journal*. 120(18): 3973-82. doi: 10.1016/j.bpj.2021.08.016.

Guo, L., Wang, S., Li, M., Cao, Z. 2019. Accurate classification of membrane protein types based on sequence and evolutionary information using deep learning. *BMC Bioinformatics*. 20(Suppl 25): 700. doi: 10.1186/s12859-019-3275-6.

Gupta, R.S. 2000. The Natural Evolutionary Relationships among Prokaryotes. *Critical Reviews in Microbiology*. 26(2): 111–31. doi: 10.1080/10408410091154219.

Gupta, R.S. 2011. Origin of diderm (Gram-negative) bacteria: antibiotic selection pressure rather than endosymbiosis likely led to the evolution of bacterial cells with two membranes. *Antonie Van Leeuwenhoek*. 100(2): 171–82. doi: 10.1007/s10482-011-9616-8.

Hagen, W.J.H., Wan, W. & Briggs, J.A.G. 2017. Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *Journal of Structural Biology*. 197(2): 191-8. doi: 10.1016/j.jsb.2016.06.007.

Han, L., Qu, Q., Aydin, D., Panova, O., Robertson, M.J., Xu, Y., Dror, R.O., Skiniotis, G., *et al*. 2022. Structure and mechanism of the SGLT family of glucose transporters. *Nature*. 601: 274-9. doi: 10.1038/s41586-021-04211-w.

Hanszen, K.J. 2018. The Optical Transfer Theory of the Electron Microscope: Fundamental Principles and Applications. In *Advances in Imaging and Electron Physics*. P.W. Hawkes, Ed. Elsevier. 207: 251-341. doi: 10.1016/bs.aiep.2018.04.001.

Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J.Y., Wang, L., Lupyan, D., *et al*. 2016. OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *Journal of Chemical Theory & Computation*. 12: 281–96. doi: 10.1021/acs.jctc.5b00864.

Hardy, D., Bill, R.M., Jawhari, A., Rothnie, A.J. 2016. Overcoming bottlenecks in the membrane protein structural biology pipeline. *Biochemical Society of Transactions*. 44(3): 838–44. doi: 10.1042/BST20160049.

Hart, R.G. 1968. Electron Microscopy of Unstained Biological Material: The Polytropic Montage. *Science*. 159(3822): 1464-7. doi: 10.1126/science.159.3822.1464.

Härtlein, M., Schieβl, S., Wagner, W., Rdest, U., Kreft, J., Goebel, W. 1983. Transport of Hemolysin by *Escherichia coli*. *Journal of Cellular Biochemistry*. 22(2): 87-97. doi: 10.1002/jcb.240220203.

Hattab, G., Warschawski, D.E., Moncoq, K., Miroux, B. 2015. *Escherichia coli* as host for membrane protein structure determination: a global analysis. *Scientific Reports*. 5: 12097. doi: 10.1038/srep12097.

Håvarstein, L.S., Diep, D.B. & Nes, I.F. 1995. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. *Molecular Microbiology*. 16(2): 229-40. doi: 10.1111/j.1365-2958.1995.tb02295.x.

He, Y., Maisuradze, G.G., Yin, Y., Scheraga, H.A. 2017. Sequence-, structure-, and dynamics-based comparisons of structurally homologous CheY-like proteins. *Proceedings of the National Academy of Sciences of the United States of America*. 114(7): 1578-83. doi: 10.1073/pnas.1621344114.

Hearn, E.M., Patel, D.R., Lepore, B.W., Indic, M., van den Berg, B. 2009. Transmembrane passage of hydrophobic compounds through a protein channel wall. *Nature*. 458(7236): 367-70. doi: 10.1038/nature07678.

Hegedűs, T., Geisler, M., Lukács, G.L., Farkas, B. 2022. Ins and outs of AlphaFold2 transmembrane protein structure predictions. *Cellular and Molecular Life Sciences*. 79(1): 73. doi: 10.1007/s00018-021-04112-1.

Heimowitz, A. Sharon, N. & Singer, A. 2021. Centering Noisy Images with Application to Cryo-EM. *SIAM Journal on Imaging Sciences*. 14(2): 689-716. doi: 10.1137/20m1365946.

Heinisch, T., Schwizer, F., Garabedian, B., Csibra, E., Jeschek, M., Vallapurackal, J., Pinheiro, V.B., Marlière, P., *et al*. 2018. *E. coli* surface display of streptavidin for directed evolution of an allylic deallylase. *Chemical Science*. 9: 5383. doi: 10.1039/c8sc00484f.

Henderson, R. 2013. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences of the United States of America.* 110(45): 18037-41. doi: 10.1073/pnas.1314449110.

Henzler-Wildman, K. & Kern, D. 2007. Dynamic personalities of proteins. *Nature*. 450(7172): 964-72. doi: 10.1038/nature06522.

Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M. 1998. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*. 18(12): 1463-72. doi: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

Hess, B. 2002. Convergence of sampling in protein simulations. *Physical Reviews E.* 65: 031910. doi: 10.1103/PhysRevE.65.031910.

Hess, V., Poehlein, A., Weghoff, C.M., Daniel, R., Müller, V. 2014. A genome-guided analysis of energy conservation in the thermophilic, cytochrome-free acetogenic bacterium *Thermoanaerobacter kivui*. *BMC Genomics*. 15: 1139. doi: 10.1186/1471-2164-15-1139.

Higgins, C.F., Hiles, I.D., Salmond, G.P.C., Gill, D.R., Downie, J.A., Evans, I.J., Holland, I.B., Gray, L., *et al*. 1986. A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. *Nature*. 323(6087): 448-50. doi: 10.1038/323448a0.

Hilser, V.J., Wrabl, J.O. & Motlagh, H.N. 2012. Structural and Energetic Basis of Allostery. *Annual Review of Biophysics*. 41: 585-609. doi: 10.1146/annurev-biophys-050511-102319.

Himes, B.A. & Zhang, P. 2018. emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nature Methods*. 15(11): 955-61. doi: 10.1038/s41592-018-0167-z.

Hodges, F.J., Torres, V.V.L., Cunningham, A.F., Henderson, I.R., Icke, C. 2022. Redefining the bacterial Type I protein secretion system. In *Advances in Microbial Physiology*. San Diego, CA, United States: Academic Press. 1-50. doi: 10.1016/bs.ampbs.2022.10.003.

Hofmann, S., Januliene, D., Mehdipour, A.R., Thomas, C., Stefan, E., Brüchert, S., Kuhn, B.T., Geertsma, E.R., *et al*. 2019. Conformation space of a heterodimeric ABC exporter under turnover conditions. *Nature*. 571: 580-3. doi: 10.1038/s41586-019-1391-0.

Holland, I.B., Peherstorfer, S., Kanonenberg, K., Lenders, M., Reimann, S., Schmitt, L. 2016. Type I Protein Secretion— Deceptively Simple yet with a Wide Range of Mechanistic Variability across the Family. *EcoSal Plus*. 7(1). doi: 10.1128/ecosalplus.ESP-0019-2015.

Hollingsworth, S.A. & Dror, R.O. 2018. Molecular dynamics simulation for all. *Neuron*. 99(6): 1129-43. doi: 10.1016/j.neuron.2018.08.011.

Honeycutt, J.D., Wenner, N., Li, Y., Brewer, S.M., Massis, L.M., Brubaker, S.W., Chairatana, P., Owen, S.V, *et al*. 2020. Genetic variation in the MacAB-TolC efflux pump influences pathogenesis of invasive *Salmonella* isolates from Africa. *PLoS Pathogens*. 16(8): e1008763. doi: 10.1371/journal.ppat.1008763.

Horiyama, T., Yamaguchi, A. & Nishino, K. 2010. TolC dependency of multidrug efflux systems in *Salmonella enterica serovar Typhimurium*. *Journal of Antimicrobial Chemotherapy*. 65: 1372–6. doi: 10.1093/jac/dkq160.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., Simmerling, C. 2006. Comparison of multiple AMBER force fields and development of improved protein backbone parameters. *Proteins*. 65(3): 712–25. doi: 10.1002/prot.21123.

Hsu, P.C., Bruininks, B.M.H., Jefferies, D., de Souza, P.C.T., Lee, J., Patel, D.S., Marrink, S.J., Qi, Y., *et al*. 2017. CHARMM-GUI Martini Maker for Modeling and Simulation of Complex Bacterial Membranes with Lipopolysaccharides. *Journal of Computational Chemistry*. 38: 2354-63. doi: 10.1002/jcc.24895.

Hsueh, S.C.C., Nijland, M., Peng, X., Hilton, B., Plotkin, S.S. 2022. First Principles Calculation of Protein–Protein Dimer Affinities of ALS-Associated SOD1 Mutants. *Frontiers in Molecular Biosciences*. 9: 845013. doi: 10.3389/fmolb.2022.845013.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hernsdorf, A.W., *et al*. 2016. A new view of the tree of life. *Nature Microbiology*. 1: 16048. doi: 10.1038/nmicrobiol.2016.48.

Hui, D., Morden, C., Zhang, F., Ling, V. 2000. Combinatorial Analysis of the Structural Requirements of the *Escherichia coli* Hemolysin Signal Sequence. *Journal of Biological Chemistry*. 275(4): 2713–20. doi: 10.1074/jbc.275.4.2713.

Ingólfsson, H.I., Lopez, C.A., Uusitalo, J.J., de Jong, D.H., Gopal, S.M., Periole, X., Marrink, S.J. 2013. The power of coarse graining in biomolecular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science*. 4(3): 225–48. doi: 10.1002/wcms.1169.

Ishii S., Yano T., Ebihara A., Okamoto A., Manzoku M., Hayashi H. 2010. Crystal structure of the peptidase domain of *Streptococcus* ComA, a bifunctional ATP-binding cassette transporter involved in the quorum-sensing pathway. *Journal of Biological Chemistry*. 285: 10777–85. doi: 10.1074/jbc.M109.093781.

Itoh, K. & Sasal, M. 2004. Dynamical transition and proteinquake in photoactive yellow protein. *Proceedings of the National Academy of Sciences of the United States of America*. 101(41): 14736-41. doi: 10.1073/pnas.0402978101.

Jarchau, T., Chakraborty, T., Garcia, F., Goebel, W. 1994. Selection for transport competence of C-terminal polypeptides derived from *Escherichia coli* hemolysin: the shortest peptide capable of autonomous HlyB/HlyD-dependent secretion comprises the C-terminal 62 amino acids of HlyA. *Molecular & General Genetics*. 245: 53-60. doi: 10.1007/BF00279750.

Jardetzky, O. 1966. Simple Allosteric Model for Membrane Pumps. *Nature*. 211: 969-70. doi: 10.1038/211969a0.

Jeffries, D. & Khalid, S. 2021. Atomistic and coarse-grained simulations of membrane proteins: A practical guide. *Methods*. 185: 15-27. doi: 10.1016/j.ymeth.2020.02.007.

Jermann, T.M., Opitz, J.G., Stackhouse, J., Benner, S.A. 1995. Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. *Nature*. 374: 57-9. doi: 10.1038/374057A0.

Jiang, Y., Thienpont, B., Sapuru, V., Hite, R.K., Dittman, J.S., Sturgis, J.N., Scheuring, S. 2022. Membrane-mediated protein interactions drive membrane protein organization. *Nature Communications*. 13: 7373. doi: 10.1038/s41467-022-35202-8.

Jo, S., Kim, T., Im, W. 2007. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS One*. 2(9): e880. doi: 10.1371/journal.pone.0000880.

Jo, S., Kim, T., Iyer, V.G., Im, W. 2008. CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. *Journal of Computational Chemistry*. 29: 1859-65. doi: 10.1002/jcc.20945.

Jones, P.M. & George, A.M. 2009. Opening of the ADP-bound active site in the ABC transporter ATPase dimer: Evidence for a constant contact, alternating sites model for the catalytic cycle. *Proteins*. 75(2): 387-96. doi: 10.1002/prot.22250.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., *et al*. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*. 596: 583-9. doi: 10.1038/s41586-021-03819-2.

Jung, J., Kobayashi, C., Kasahara, K., Tan, C., Kuroda, A., Minami, K., Ishiduki, S., Nishiki, T., *et al.* 2020. New parallel computing algorithm of molecular dynamics for extremely huge scale biological systems. *Journal of Computational Chemistry*. 42(4): 231-41. doi: 10.1002/jcc.26450.

Kaback, H.R. 2005. Structure and mechanism of the lactose permease. *Comptes Rendus Biologies*. 328(6): 557-67. doi: 10.1016/j.crvi.2005.03.008.

Kampjut, D. Steiner, J. & Sazanov, L.A. 2021. Cryo-EM grid optimization for membrane proteins. *iScience*. 24(3): 102139. doi: 10.1016/j.isci.2021.102139.

Kanonenberg, K., Schwarz, C.K.W. & Schmitt, L. 2013. Type I secretion systems – a story of appendices. *Research in Microbiology*. 164(6): 596-604. doi: 10.1016/j.resmic.2013.03.011.

Kanonenberg, K., Spitz, O., Erenburg, I.N., Beer, T., Schmitt, L. 2018. Type I secretion system—it takes three and a substrate. *FEMS Microbiology Letters*. 365(11). doi: 10.1093/femsle/fny094.

Kaplan, E., Greene, N.P., Crow, A., Koronakis, V. 2018. Insights into bacterial lipoprotein trafficking from a structure of LolA bound to the LolC periplasmic domain. *Proceedings of the National Academy of Sciences of the United States of America*. 115(31): E7389–97. doi: 10.1073/pnas.1806822115.

Kaplan, E., Ntsogo, Y., Harris, A., Yu, D.G., Luisi, B.F. A lipoprotein partner for TolC in *Escherichia coli* [working title]. *In preparation*.

Karplus, M. & McCammon, J.A. 2002. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*. 9(9): 646-52. doi: 10.1038/nsb0902-646.

Kauffman, S. & Levin, S. 1987. Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*. 128(1): 11-45. doi: 10.1016/s0022-5193(87)80029-2.

Kauffman, S.A. & Weinberger, E.D. 1989. The NK model of rugged fitness landscapes and its applications to maturation of the immune response. *Journal of Theoretical Biology*. 141(2): 211-45. doi: 10.1016/s0022-5193(89)80019-0.

Kaur, S., Gomez-Blanco, J., Khalifa, A.A.Z., Adinarayanan, S., Sanchez-Garcia, R., Wrapp, D., McLellan, J.S., Bui, K.H., *et al*. 2021. Local computational methods to improve the interpretability and analysis of cryo-EM maps. *Nature Communications*. 12: 1240. doi: 10.1038/s41467-021-21509-5.

Keskin, O., Jernigan, R.L. & Bahar, I. 2000. Proteins with Similar Architecture Exhibit Similar Large-Scale Dynamic Behavior. *Biophysical Journal*. 78(4): 2093-106. doi: 10.1016/S0006-3495(00)76756-7.

Kieuvongngam, V., Olinares, P.D.B., Palillo, A., Oldham, M.L., Chait, B.T., Chen, J. 2020. Structural basis of substrate recognition by a polypeptide processing and secretion transporter. eLife. 9: e51492. doi: 10.7554/eLife.51492.

Kieuvongngam, V. & Chen, J. 2022. Structures of the peptidase-containing ABC transporter PCAT1 under equilibrium and nonequilibrium conditions. *Proceedings of the National Academy of Sciences of the United States of America*. 119(4): e2120534119. doi: 10.1073/pnas.2120534119.

Killian, J.A., Koorengevel, M.C., Bouwstra, J.A., Gooris, G., Dowhan, W., de Kruijff, B. 1994. Effect of divalent cations on lipid organization of cardiolipin isolated from *Escherichia coli* strain AH930. *Biochimica et Biophysica Acta*. 1189(2): 225-32. doi: 10.1016/0005-2736(94)90069-8.

Kim, J.S., Song, S., Lee, M., Lee, S., Lee, K., Ha, N.C. 2016. Crystal Structure of a Soluble Fragment of the Membrane Fusion Protein HlyD in a Type I Secretion System of Gram-Negative Bacteria. *Structure*. 24(3): 477-85. doi: 10.1016/j.str.2015.12.012.

Kirykowicz, A.M. & Woodward, J.D. 2020. Shotgun EM of mycobacterial protein complexes during stationary phase stress. *Current Research in Structural Biology*. 2: 204-12. doi: 10.1016/j.crstbi.2020.09.002.

Klein, J.D., Blount, M.A. & Sands, J.M. 2012. Molecular Mechanisms of Urea Transport in Health and Disease. *Pflügers Archiv: European Journal of Physiology*. 464(6): 561–72. doi: 10.1007/s00424-012-1157-0.

Kleinjung, J., Scott, W.R.P., Allison, J.R., van Gunsteren, W.F., Fraternali, F. 2012. Implicit Solvation Parameters Derived from Explicit Water Forces in Large-Scale Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation*. 8(7): 2391–2403. doi: 10.1021/ct200390j.

Klingenberg, M. 1981. Membrane protein oligomeric structure and transport function. *Nature*. 290(5806): 449-54. doi: 10.1038/290449a0.

Klug, A. & Crowther, R.A. 1972. Three-dimensional Image Reconstruction from the Viewpoint of information Theory. *Nature*. 238: 435-40. doi: 10.1038/238435a0.

Kluyver, T., Ragan-Kelly, B., Fernando, P., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., *et al*. 2016. Jupyter Notebooks – a publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt, eds. Positioning and Power in Academic Publishing: Players, Agents and Agendas. pp. 87–90. doi: 10.3233/978-1-61499-649-1-87.

Knapp, S., Hacker, J., Jarchau, T., Goebel, W. 1986. Large, unstable inserts in the chromosome affect virulence properties of uropathogenic *Escherichia coli* O6 strain 536. *Journal of Bacteriology*. 168(1): 22–30. doi: 10.1128/jb.168.1.22-30.1986.

Knapp, B., Ospina, L. & Deane, C.M. 2018. Avoiding False Positive Conclusions in Molecular Simulation: The Importance of Replicas. *Journal of Chemical Theory & Computation*. 14: 6127−38. doi: 10.1021/acs.jctc.8b00391.

Kobayashi, N., Nishino, K. & Yamaguchi, A. 2001. Novel Macrolide-Specific ABC-Type Efflux Transporter in *Escherichia coli. Journal of Bacteriology*. 183(19): 5639-44. doi: 10.1128/JB.183.19.5639–5644.2001.

Koronakis, V., Koronakis, E. & Hughes, C. 1989. Isolation and analysis of the C-terminal signal directing export of Escherichia coli hemolysin protein across both bacterial membranes. *The EMBO Journal*. 8: 595-605. doi: 10.1002/j.1460-2075.1989.tb03414.x.

Koronakis, V., Hughes, C. & Koronakis, E. 1991. Energetically distinct early and late stages of HlyB/HlyDdependent secretion across both Escherichia coli membranes. *The EMBO Journal*. 10(11): 3263–3272. doi: 10.1002/j.1460-2075.1991.tb04890.x.

Koshland, D.E., Némethy, G. & Filmer, D. 1966. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry*. 5(1): 365-85. doi: 10.1021/bi00865a047.

Kostiuk, B., Santoriello, F.J., Diaz-Satizabal, L., Bisaro, F., Lee, K.J., Dhody, A.N., Provenzano, D., Unterweger, D., *et al*. 2021. Type VI secretion system mutations reduced competitive fitness of classical *Vibrio cholerae* biotype. *Nature Communications*. 12: 6457. doi: 10.1038/s41467-021-26847-y.

Kowalczyk, L., Ratera, M., Paladino, A., Bartoccioni, P., Errasti-Murugarren, E., Valencia, E., Portella, G., Bial, S., *et al*. 2011. Molecular basis of substrate-induced permeation by an amino

acid antiporter. *Proceedings of the National Academy of Sciences of the United States of America*. 108(10): 3935-40. doi: 10.1073/pnas.1018081108.

Krivov, G.G., Shapovalov, M.V., Dunbrack, R.L. Jr. 2009. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins*. 77(4):778-95. doi: 10.1002/prot.22488.

Krishnamurthy, H. & Gouaux, E. 2012. X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. *Nature*. 481: 469-74. doi: 10.1038/nature10737.

Kryazhimskiy, S., Tkacik, G. & Plotkin, J.B. 2009. The dynamics of adaptation on correlated fitness landscapes. *Proceedings of the National Academy of Sciences of the United States of America*. 106(44): 18638-43. doi: 10.1073/pnas.0905497106.

Kubicek, J., Block, H., Maertens, B., Spriestersbach, A., Labahn, J. 2014. Expression and purification of membrane proteins. *Methods in Enzymology*. 541: 117-40. doi: 10.1016/B978-0-12-420119-4.00010-0.

Kühlbrandt, W. 2004. Biology, structure and mechanism of P-type ATPases*. Nature Reviews Molecular Cell Biology*. 5: 282–95. doi: 10.1038/nrm1354.

Kühlbrandt, W. 2014. The Resolution Revolution. *Science*. 343(6178): 1443-4. doi: 10.1126/science.1251652.

Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., Ferrin, T.E. 1982. A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology*. 161(2): 269-88. doi: 10.1016/0022-2836(82)90153-X.

Kutzner, C., Páll, S., Fechner, M., Esztermann, A., de Groot, B.L., Grubmüller. H. 2015. Best bang for your buck: GPU nodes for GROMACS biomolecular simulations. *Journal of Computational Chemistry.* 36(26): 1990–2008. doi: 10.1002/jcc.24030.

Kutzner, C., Páll, S., Fechner, M., Esztermann, A., de Groot, B.L., Grubmüller. H. 2019. More bang for your buck: Improved use of GPU nodes for GROMACS 2018. *Journal of Computational Chemistry*. 40(27): 2418-31. doi: 10.1002/jcc.26011.

Latorraca, N.R., Fastman, N.M., Venkatakrishnan, A.J., Frommer, W.B., Dror, R.O., Feng, L. 2017. Mechanism of Substrate Translocation in an Alternating Access Transporter. *Cell*. 169(1): 96-107. doi: 10.1016/j.cell.2017.03.010.

le Maire, M., Champeil, P. & Møller, J.V. 2000. Interaction of membrane proteins and lipids with solubilizing detergents. *Biochimica et Biophysica Acta*. 1508(1-2): 86-111. doi: 10.1016/s0304-4157(00)00010-1.

Lecher, J., Schwarz, C.K.W., Stoldt, M., Smits, S.H.J., Willbold, D., Schmitt, L. 2012. An RTX Transporter Tethers Its Unfolded Substrate during Secretion via a Unique N-Terminal Domain. *Structure*. 20: 1778-87. doi: 10.1016/j.str.2012.08.005.

Lee, M., Jun, S.J., Yoon, B.Y., Song, S., Lee, K., Ha, N.C. 2012. Membrane Fusion Proteins of Type I Secretion System and Tripartite Efflux Pumps Share a Binding Motif for TolC in Gram-Negative Bacteria. *PLoS One*. 7(7): e40460. doi: 10.1371/journal.pone.0040460.

Lee, P.C. & Rietsch, A. 2015. Fueling type III secretion. *Trends in Microbiology*. 25(3): 296-300. doi: 10.1016/j.tim.2015.01.012.

Lee, J., Cheng, X., Swails, J.M., Yeoms, M.S., Eastman, P.K., Lemkul, J.A., Wei, S., Buckner, J., *et al*. 2016. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation*. 12: 405-13. doi: 10.1021/acs.jctc.5b00935.

Lee, T.S., Cerutti, D.S., Mermelstein, D., Lin, C., LeGrand, S., Giese, T.J., Roitberg, A., Case, D.A., *et al*. 2018. GPU-accelerated molecular dynamics and free energy methods in Amber18: performance enhancements and new features. *Journal of Chemical Information and Modeling*. 58(10): 2043–50. doi: 10.1021/acs.jcim.8b00462.

Lee, R.G., Gao, J., Siira, S.J., Shearwood, A.M., Ermer, J.A., Hofferek, V., Matthews, J.C., Zheng, M., *et al*. 2020. Cardiolipin is required for membrane docking of mitochondrial ribosomes and protein synthesis. *Journal of Cell Science*. 133(14): jcs240374. doi: 10.1242/jcs.240374.

Lemkul, J.A., Huang, J., Roux, B., MacKerell Jr, A.D. 2016. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chemical Reviews*. 116(9): 4983-5013. doi: 10.1021/acs.chemrev.5b00505.

Lemmon, M.A. 2007. Pleckstrin Homology (PH) domains and phosphoinositides. *Biochemical Society Symposia*. (74): 81–93. doi: 10.1042/BSS0740081.

Lenders, M.H.H., Weidtkamp-Peters, S., Kleinschrodt, D., Jaeger, K.E., Smits, S.H.J., Schmitt, L. 2015. Directionality of substrate translocation of the hemolysin A Type I secretion system. *Science Reports*. 5: 12470. doi: 10.1038/srep12470.

Lenders, M.H.H., Beer, T., Smits, S.H.J., Schmitt, L. 2016. *In vivo* quantification of the secretion rates of the hemolysin A Type I secretion system. *Scientific Reports*. 6: 33275. doi: 10.1038/srep33275.

Létoffé, S., Delepelaire, P. & Wandersman, C. 1996. Protein secretion in gram-negative bacteria: assembly of the three components of ABC protein-mediated exporters is ordered and promoted by substrate binding. *EMBO Journal*. 15(21): 5804-11. doi: 10.1002/j.1460-2075.1996.tb00967.x.

Levin, E.J., Quick, M. & Zhou, M. 2009. Crystal structure of a bacterial homologue of the kidney urea transporter. *Nature*. 462: 757-61. doi: 10.1038/nature08558.

Levy, Y. & Becker, O.M. 2001. Energy landscapes of conformationally constrained peptides. *Journal of Chemical Physics*. 114(2): 993. doi: 10.1063/1.1329646.

Li, G., Magana, D. & Dyer, R.B. 2014. Anisotropic energy flow and allosteric ligand binding in albumin. *Nature Communications*. 5: 3100. doi: 10.1038/ncomms4100.

Lin, W., Fullner, K.J., Clayton, R., Sexton, J.A., Rogers, M.B., Calia, K.E., Calderwood, S.B., Fraser, C., *et al*. 1999. Identification of a *Vibrio cholerae* RTX toxin gene cluster that is tightly linked to the cholera toxin prophage. *Proceedings of the National Academy of Sciences of the United States of America*. 96(3): 1071–6. doi: 10.1073/pnas.96.3.1071.

Lin, S.H. & Guidotti, G. 2009. Purification of membrane proteins. *Methods in Enzymology*. 463: 619-29. doi: 10.1016/S0076-6879(09)63035-4.

Lin, D.Y., Huang, S. & Chen, J. 2015. Crystal structures of a polypeptide processing and secretion transporter. *Nature*. 523(7561): 425-30. doi: 10.1038/nature14623.

Lindahl, Abraham, Hess, van der Spoel. 2020. GROMACS Documentation Release 2020.1. Zenodo, URL: https://zenodo.org/record/3685920#.Xun6L_LRbow [Last Accessed 17 June 2020] DOI: https://doi.org/10.5281/zenodo.3685920.

Lindahl, Abraham, Hess, van der Spoel. 2021. GROMACS 2021 Manual. Zenodo, URL: https://zenodo.org/record/4457591#.YvOZvvHMJo4 [last accessed 10 August 2022] DOI: 10.5281/zenodo.4457591.

Linhartová, I., Bumba, L., Mašín, J., Basler, M., Osička, R., Kamanová, J., Procházková, K., Adkins, A., *et al*. 2010. RTX proteins: a highly diverse family secreted by a common mechanism. *FEMS Microbiology Reviews*. 34(6): 1076–1112. doi: 10.1111/j.1574-6976.2010.00231.x.

Liu, P., Agrafiotis, D.K. & Theobald, D.L. 2010. Fast determination of the optimal rotational matrix for macromolecular superpositions. *Journal of Computational Chemistry*. 31: 1561–3. doi: 10.1002/jcc.21439.

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

Lockless, S.W. & Ranganathan, R. 1999. Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. *Science*. 286(5438): 295-9. doi: 10.1126/science.286.5438.295.

Lomize, M.A., Pogozheva, I.D., Joo, H., Mosberg, H.I., Lomize, A.L. 2011. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Research*. 40: D370-6. doi: 10.1093/nar/gkr703.

Lopes, P.E.M., Guvench, O. & MacKerell, A.D. 2015. Current Status of Protein Force Fields for Molecular Dynamics. *Methods in Molecular Biology*. 1215: 47–71. doi: 10.1007/978-1-4939-1465-4_3.

Lopez-Redondo, M.L., Coudray, N., Zhang, Z., Alexopoulos, J., Stokes, D.L. 2018. Structural basis for the alternating access mechanism of the cation diffusion facilitator YiiP. *Proceedings of the National Academy of Sciences of the United States of America*. 115(12): 3042-7. doi: 10.1073/pnas.1715051115.

Lu, S. & Zgurskaya, H.I. 2012. Role of ATP binding and hydrolysis in assembly of MacAB–TolC macrolide transporter. *Molecular Microbiology*. 86(5): 1132–43. doi: 10.1111/mmi.12046.

Lu, L. & Anderson-Cook, C. 2021. Strategies for sequential design of experiments and augmentation. *Quality and Reliability Engineering International*. 37(5): 1740-57. doi: 10.1002/qre.2823.

Ludtke S.J., Baldwin P.R. & Chiu W. 1999. EMAN: semiautomated software for high-resolution single-particle reconstructions. *Journal of Structural Biology*. 128(1): 82-97. doi: 10.1006/jsbi.1999.4174.

Luther, P.K. & Crowther, R.A. 1984. Three-dimensional reconstruction from tilted sections of fish muscle M-band. *Nature*. 307: 566-8. doi: 10.1038/307566a0.

Mackman, N., Nicaud, J.M., Gray, L., Holland, I.B. 1985. Identification of polypeptides required for the export of haemolysin 2001 from *E. coli. Molecular and General Genetics*. 201(3): 529-36. doi: 10.1007/BF00331351.

Mackman, N., Baker, K., Gray, L., Haigh, R., Nicaud, J.M., Holland, I.B. 1987. Release of a chimeric protein into the medium from Escherichia coli using the C-terminal secretion signal of haemolysin. *The EMBO Journal*. 6(9): 2835-41. doi: 10.1002/j.1460-2075.1987.tb02580.x.

Maguid, S., Fernández-Alberti, S., Parisi, G., Echave, J. 2006. Evolutionary Conservation of Protein Backbone Flexibility. *Journal of Molecular Evolution*. 63(4): 448-57. doi: 10.1007/s00239-005-0209-x.

Marko, M., Hsieh, C., Schalek, R., Frank, J., Mannella, C. 2007. Focused-ion-beam thinning of frozen-hydrated biological specimens for cryo-electron microscopy. *Nature Methods*. 4: 215-7. doi: 10.1038/nmeth1014.

Marlovits, T.C., Kubori, T., Sukhan, A., Thomas, D.R., Galán, J.E., Unger, V.M. 2004. Structural insights into the assembly of the type III secretion needle complex. *Science*. 306(5698): 1040-2. doi: 10.1126/science.1102610.

Marrink, S.J., Risselada, H.J., Yefimov, S., Tieleman, D.P., de Vries, A.H. 2007. The MARTINI force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B*. 111(27): 7812-24. doi: 10.1021/jp071097f.

Mastronarde, D.N. & Held, S.R. 2017. Automated tilt series alignment and tomographic reconstruction in IMOD. *Journal of Structural Biology*. 197(2): 102-13. doi: 10.1016/j.jsb.2016.07.011.

Matias, V.R.F., Al-Amoudi, A., Dubochet, J., Beveridge, T.J. 2003. Cryo-transmission electron microscopy of frozen-hydrated sections of *Escherichia coli* and *Pseudomonas aeruginosa*. *Journal of Bacteriology*. 185(20): 6112-8. doi: 10.1128/JB.185.20.6112-6118.2003.

Mathieu, K., Javed, W., Vallet, S., Lesterlin, C., Candusso, M.P., Ding, F., Xu, X.N., Ebel, C., *et al*. 2019. Functionality of membrane proteins overexpressed and purified from *E. coli* is highly dependent upon the strain. *Scientific Reports*. 9: 2654. doi: 10.1038/s41598-019-39382-0.

Matsuyama, S., Tajima,T. & Tokuda, H. 1995. A novel periplasmic carrier protein involved in the sorting and transport of *Escherichia coli* lipoproteins destined for the outer membrane. *The EMBO Journal*. 14(14): 3365–72. doi: 10.1002/j.1460-2075.1995.tb07342.x.

Matsuyama, S., Yokota, N. & Tokuda, H. 1997. A novel outer membrane lipoprotein, LolB (HemM), involved in the LolA (p20)-dependent localization of lipoproteins to the outer membrane of *Escherichia coli*. *The EMBO Journal*. 16(23): 6947–55. doi: 10.1093/emboj/16.23.6947.

McAuley, K.E., Fyfe, P.K., Ridge, J.P., Isaacs, N.W., Cogdell, R.J., Jones, M.R. 1999. Structural details of an interaction between cardiolipin and an integral membrane protein. *Proceedings of the National Academy of Sciences of the United States of America*. 96(26): 14706-11. doi: 10.1073/pnas.96.26.14706.

McCammon, J.A., Gelin, B.R & Karplus, M. 1977. Dynamics of folded proteins. *Nature*. 267: 585-90. doi: 10.1038/267585a0.

McCammon, J.A., Wolynes, P.G. & Karplus, M. 1979. Picosecond dynamics of tyrosine side chains in proteins. *Biochemistry*. 18(6): 927-42. doi: 10.1021/bi00573a001.

Mehmood, S., Corradi, V., Choudhury, H.G., Hussain, R., Becker, P., Axford, D., Zirah, S., Rebuffat, S., *et al*. 2016. Structural and Functional Basis for Lipid Synergy on the Activity of the Antibacterial Peptide ABC Transporter McjD. *Journal of Biological Chemistry*. 291(41): 21656-68. doi: 10.1074/jbc.M116.732107.

Menestrina, G., Moser, C., Pellet, S., Welch, R. 1994. Pore-formation by *Escherichia coli* hemolysin (HlyA) and other members of the RTX toxins family. *Toxicology*. 87(1-3):249-67. doi: 10.1016/0300-483x(94)90254-2.

Merkl, R. & Sterner, R. 2016. Ancestral protein reconstruction: techniques and applications. *Biological Chemistry*. 397(1): 1-21. doi: 10.1515/hsz-2015-0158.

Meroueh, S.O., Bencze, K.Z., Hesek, D., Lee, M., Fisher, J.F., Stemmler, T.L., Mobashery, S. 2006. Three-dimensional structure of the bacterial cell wall peptidoglycan. *Proceedings of the National Academy of Sciences of the United States of America*. 103(12): 4404-9. doi: 10.1073/pnas.0510182103.

Mersmann, S. F., Strömich, L., Song, F. J., Wu, N., Vianello, F., Barahona, M., & Yaliraki, S. N. 2021. ProteinLens: a web-based application for the analysis of allosteric signalling on atomistic graphs of biomolecules. *Nucleic Acids Research*. 49(W1): W551-8 doi: 10.1093/nar/gkab350.

Michaud-Agrawal, N., Denning, E.J., Woolf, T.B., Beckstein, O. 2011. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *Journal of Computational Chemistry*. 32: 2319–27. doi:10.1002/jcc.21787.

Milne, J.L.S & Subramaniam, S. 2009. Cryo-electron tomography of bacteria: progress, challenges and future prospects. *Nature Reviews Microbiology*. 7(9): 666–75. doi: 10.1038/nrmicro2183.

Miroux B. & 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): 289–98. doi: 10.1006/jmbi.1996.0399.

Miyamoto, S. & Tokuda, H. 2007. Diverse effects of phospholipids on lipoprotein sorting and ATP hydrolysis by the ABC transporter LolCDE complex. *Biochimica et Biophysica Acta*. 1768: 1848-54. doi: 10.1016/j.bbamem.2007.04.005.

Moffat, K. 2001. Time-Resolved Biochemical Crystallography: A Mechanistic Perspective. *Chemical Reviews*. 101(6): 1569-81. doi: 10.1021/cr990039q.

Monod, J., Wyman, J. & Changeux, J.P. 1965. On the Nature of Allosteric Transitions: A Plausible Model. *Journal of Molecular Biology*. 12: 88-118. doi: 10.1016/s0022-2836(65)80285-6.

Morgan, J.L.W., Acheson, J.F. & Zimmer, J. 2017. Structure of a Type-1 Secretion System ABC Transporter. *Structure*. 25(3): 522-9. doi: 10.1016/j.str.2017.01.010.

Moore, P.B., Hendrickson, W.A., Henderson, R., Brunger, A.T. 2022. The protein-folding problem: Not yet solved. *Science*. 375(6580): 507. doi: 10.1126/science.abn9422.

Nadler, W., Brünger, A.T., Schulten, K., Karplus, M. 1987. Molecular and stochastic dynamics of proteins. *Proceedings of the National Academy of Sciences of the United States of America*. 84(22): 7933–7. doi: 10.1073/pnas.84.22.7933.

Nagel, A.C., Joyce, P., Wichman, H.A., Miller, C.R. 2012. Stickbreaking: A Novel Fitness Landscape Model That Harbors Epistasis and Is Consistent with Commonly Observed Patterns of Adaptive Evolution. *Genetics*. 190(2): 655-67. doi: 10.1534/genetics.111.132134.

Nakamura, H., Hisano, T., Rahman, M.M., Tosha, T., Shirouzu, M., Shiro, Y. 2022. Structural basis for heme detoxification by an ATP-binding cassette–type efflux pump in gram-positive pathogenic bacteria. *Proceedings of the National Academy of Sciences of the United States of America*. 119(27): e2123385119. doi: 10.1073/pnas.2123385119.

Nakanishi-Matsui, M., Sekiya, M., Nakamoto, R.K., Futai, M. 2010. The mechanism of rotating proton pumping ATPases. *Biochimica et Biophysica Acta*. 1797(8): 1343-52. doi: 10.1016/j.bbabio.2010.02.014.

Navale, A.M. & Paranjape, A.N. 2016. Glucose transporters: physiological and pathological roles. *Biophysical Reviews*. 8(1): 5–9. doi: 10.1007/s12551-015-0186-2.

Newport, T.D., Sansom, M.S.P. & Stansfeld, P.J. 2019. The MemProtMD database: a resource for membrane-embedded protein structures and their lipid interactions. *Nucleic Acids Research*. 47: D390–7. doi: 10.1093/nar/gky1047.

Nguyen, T.T.H., Myrold, D.D. & Mueller, R.S. 2019. Distributions of Extracellular Peptidases Across Prokaryotic Genomes Reflect Phylogeny and Habitat. *Frontiers in Microbiology*. 10: 413. doi: 10.3389/fmicb.2019.00413.

Nicaud, J.M., Mackman, N., Gray, L., Holland, I.B. 1985. Characterisation of HlyC and mechanism of activation and secretion of haemolysin from *E. coli* 2001. *FEBS Letters*. 187(2): 339-44. doi: 10.1016/0014-5793(85)81272-2.

Nicaud, J.M., Mackman, N., Gray, L., Holland, I.B. 1986. The C-terminal, 23 kDa peptide of *E. coli* haemolysin 2001 contains all the information necessary for its secretion by the haemolysin (Hly) export machinery. *FEBS Letters*. 204(2): 331-5. doi: 10.1016/0014-5793(86)80838-9.

Nishino, K., Latifi, T. & Groisman, E.A. 2006. Virulence and drug resistance roles of multidrug efflux systems of *Salmonella enterica serovar Typhimurium*. *Molecular Microbiology*. 59(1): 126-141. doi: 10.1111/j.1365-2958.2005.04940.x.

Nosé, S. 1984. A unified formulation of the constant temperature molecular dynamics methods. *Journal of Chemical Physics*. 81: 511-9. doi: 10.1063/1.447334.

Notredame, C., Higgins, D.G. & Heringa, J. 2000. T-Coffee: A Novel Method for Fast and Accurate Multiple Sequence Alignment. *Journal of Molecular Biology*. 302: 205-17. doi:10.1006/jmbi.2000.4042.

Nyquist, H. 1928. Certain Topics in Telegraph Transmission Theory. *Transactions of the American Institute of Electrical Engineers E*. 47(2): 617-44. doi: 10.1109/T-AIEE.1928.5055024.

Oakes, V. & Domene, C. 2016. Molecular Dynamics Simulations: Principals & Applications for the Study of Membrane Proteins. In: *Computational Biophysics of Membrane Proteins*. pp.19-58. doi: 10.1039/9781782626695-00019.

Ohi, M., Li, Y., Cheng, Y., Walz, T. 2004. Negative Staining and Image Classification – Powerful Tools in Modern Electron Microscopy. *Biological Procedures Online*. 6: 23–34. doi: 10.1251/bpo70.

Ohno, S. 1970. *Evolution by Gene Duplication*. 1st ed. New York, USA: Springer.

Okada, U., Yamashita, E., Neuberger, A., Morimoto, M., van Veen, H.W., Murakami, S. 2017. Crystal structure of tripartite-type ABC transporter MacB from *Acinetobacter baumannii*. *Nature Communications*. 8: 1336. doi: 10.1038/s41467-017-01399-2.

Olins, D.E., Olins, A.L., Levy, H.A., Durfee, R.C., Margle, S.M., Tinnel, E.P., Dover, S.D. 1983. Electron Microscope Tomography: Transcription in Three Dimensions. *Science*. 220(4596): 498-500. doi: 10.1126/science.6836293.

O'Neill J. 2016. Tackling drug-resistant infections globally: final report and recommendations. London: Review on Antimicrobial Resistance.

Oostenbrink C., Villa A., Mark A.E., Van Gunsteren W.F. 2004. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *Journal Of Computational Chemistry*. 25(13): 1656–76. doi: 10.1002/jcc.20090.

Orlova, E.V. & Saibil, H. 2011. Structural Analysis of Macromolecular Assemblies by Electron Microscopy. *Chemical Reviews*. 111(12): 7710–48. doi: 10.1021/cr100353t.

Orwick-Rydmark, M., Arnold, T. & Linke, D. 2016. The Use of Detergents to Purify Membrane Proteins. *Current Protocols in Protein Science*. 84: 4.8.1-4.8.35. doi: 10.1002/0471140864.ps0408s84.

Paetzel, M., Dalbey, R.E. & Strynadka, N.C. 1998. Crystal structure of a bacterial signal peptidase in complex with a beta-lactam inhibitor. *Nature*. 396(6712): 707. doi: 10.1038/24196.

Paetzel, M. 2014. Structure and mechanism of Escherichia coli type I signal peptidase. *Biochimica et Biophysica Acta*. 1843(8): 1497-508. doi: 10.1016/j.bbamcr.2013.12.003.

Parmenter, C.D. & Nizamudeen, Z.A. 2020. Cryo-FIB-lift-out: practically impossible to practical reality. *Journal of Microscopy*. 281(2): 157-74. doi: 10.1111/jmi.12953.

Parrinello M. & Rahman A. 1981. Polymorphic transitions in single crystals: a new molecular dynamics method. *Journal of Applied Physics*. 52: 7182–90. doi: 10.1063/1.328693.

Passmore, L.A. & Russo, C.J. 2016. Specimen preparation for high-resolution cryo-EM. In *Methods in Enzymology.* R.A Crowther, Ed. Academic Press. 579: 51–86. doi: 10.1016/bs.mie.2016.04.011.

Paulsen, M.D., Bass, M.B. & Ornstein, R.L. 1991. Analysis of active site motions from a 175 picosecond molecular dynamics simulation of camphor-bound cytochrome P450cam. *Journal of Biomolecular Structure and Dynamics*. 9(2): 187-203. doi: 10.1080/07391102.1991.10507906.

Pearson, W.R. & Sierk, M.L. 2005. The limits of protein sequence comparison?. *Current Opinion in Structural Biology*. 15(3): 254–60. doi: 10.1016/j.sbi.2005.05.005.

Pechlaner, M., van Gunsteren, W.F., Hansen, N., Smith, L.J. 2022. Molecular dynamics simulation or structure refinement of proteins: are solvent molecules required? A case study using hen lysozyme. *European Biophysics Journal*. 51(3): 265–82. doi: 10.1007/s00249-022-01593-1.

Penczek, P.A., Zhu, J. & Frank, J. 1996. A common-lines based method for determining orientations for N > 3 particle projections simultaneously. *Ultramicroscopy*. 63(3-4): 205-18. doi: 10.1016/0304-3991(96)00037-x.

Penczek, P.A. 2010a. Fundamentals of three-dimensional reconstruction from projections. *Methods in Enzymology*. 482: 1–33. doi: 10.1016/S0076-6879(10)82001-4.

Penczek, P.A. 2010b. Image Restoration in Cryo-electron Microscopy. *Methods in Enzymology*. 482: 35–72. doi: 10.1016/S0076-6879(10)82002-6.

Perelson, A.S. & Macken, C. 1995. Protein evolution on partially correlated landscapes. *Proceedings of the National Academy of Sciences of the United States of America*. 92(21): 9657–61. doi: 10.1073/pnas.92.21.9657.

Periole, X., Cavalli, M., Marrink, S.J., Ceruso, M.A. 2009. Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *Journal of Chemical Theory and Computation*. 5: 2531-43. doi: 10.1021/ct9002114.

Perrakis, A. & Sixma, T.K. 2021. AI revolutions in biology. *EMBO Reports*. 22: e54046. doi: 10.15252/embr.202154046.

Petterson, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*. 25(13): 1605-12. doi: 10.1002/jcc.20084.

Phillips, J.C., Hardy, D.J., Maia, J.D., Stone, J.E., Ribeiro, J.V., Bernardi, R.C., Buch, R., Fiorin, G., *et al*. 2020. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *The Journal of Chemical Physics*. 153(4): 044130. doi: 10.1063/5.0014475.

Pimenta, A.L., Young, J., Holland, I.B., Blight, M.A. 1999. Antibody analysis of the localisation, expression and stability of HlyD, the MFP component of the *E. coli* haemolysin translocator. *Molecular and General Genetics*. 261(1): 122-32. doi: 10.1007/s004380050949.

Polakiewicz, R.D. 2015. Antibodies: The solution is validation. *Nature*. 518: 483. doi: 10.1038/518483b.

Ponder, J.W., Wu, C., Ren, P., Pande, V.S., Chodera, J.D., Schnieders, M.J., Haque, I., Mobley, D.L., *et al*. 2010. Current Status of the AMOEBA Polarizable Force Field. *Journal of Physical Chemistry B*. 114(8): 2549–64. doi: 10.1021/jp910674d.

Porta-Pardo, E., Ruiz-Serra, V., Valentini, S., Valencia, A. 2022. The structural coverage of the human proteome before and after AlphaFold. *PLoS Computational Biology*. 18(1): e1009818. doi: 10.1371/journal.pcbi.1009818.

Poulsen, S.B., Fenton, R.A. & Rieg, T. 2015. Sodium-glucose cotransport. *Current Opinion in Nephrology and Hypertension*. 24(5): 463–9. doi: 10.1097/MNH.0000000000000152.

Prajapati, J.D., Kleinekathöfer, U. & Winterhalter, M. 2021. How to Enter a Bacterium: Bacterial Porins and the Permeation of Antibiotics. *Chemical Reviews*. 121(9): 5158–92. doi: 10.1021/acs.chemrev.0c01213.

Punjani, A., Rubinstein, J.L., Fleet, D.J., Brubaker, M.A. 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*. 14: 290-6. doi: 10.1038/nmeth.4169.

Pyle, E. & Zanetti, G. 2021. Current data processing strategies for cryo-electron tomography and subtomogram averaging. *The Biochemical Journal*. 478(10): 1827-45. doi: 10.1042/BCJ20200715.

Qi, Y., Ingólfsson, H.I., Cheng, X., Lee, J., Marrink, S.J., Im, W. 2015. CHARMM-GUI Martini Maker for Coarse-Grained Simulations with the Martini Force Field. *Journal of Chemical Theory and Computation*. 11: 4486-94. doi: 10.1021/acs.jctc.5b00513.

Raetz, C.R.H. & Dowhan, W. 1990. Biosynthesis and function of phospholipids in *Escherichia coli*. *Journal of Biological Chemistry*. 265(3): 1235-8. doi: 10.1016/S0021-9258(19)40001-X.

Rahman, S. & Mchaourab, H.S. 2020. ATP-dependent interactions of a cargo protein with the transmembrane domain of a polypeptide processing and secretion ABC transporter. *Journal of Biological Chemistry*. 295(43): 14678-85. doi: 10.1074/jbc.RA120.014934.

Rao, S., Bates, G.T., Matthews, C.R., Newport, T.D., Vickery, O.N., Stansfeld, P.J. 2020. Characterizing Membrane Association and Periplasmic Transfer of Bacterial Lipoproteins through Molecular Dynamics Simulations. *Structure*. 28(4): 475-87.

Rastogi, S. & Liberles, D.A. 2005. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evolutionary Biology*. 5: 28. doi: 10.1186/1471-2148-5-28.

Rath, A., Glibowicka, M., Nadeau, V.G., Chen, G., Deber, C.M. 2009. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America.* 106 (6): 1760-5. doi: 10.1073/pnas.0813167106.

Ratheal, I.M., Virgin, G.K., Yu, H., Roux, B., Gatto, C., Artigas, P. 2010. Selectivity of externally facing ion-binding sites in the Na/K pump to alkali metals and organic cations. *Proceedings of the National Academy of Sciences of the United States of America*. 107(43): 18718–23. doi: 10.1073/pnas.1004214107.

Ratkeviciute, G., Cooper, B.F. & Knowles, T.J. 2021. Methods for the solubilisation of membrane proteins: the micelle-aneous world of membrane protein solubilisation. *Biochemical Society Transactions*. 49(4): 1763–77. doi: 10.1042/BST20210181.

Rawlings, N.D. & Bateman, A. 2019. Origins of peptidases. *Biochimie*. 166: 4-18. doi: 10.1016/j.biochi.2019.07.026.

Reback, J., Mendel, J.B., McKinney, W., Van den Bossche, J., Roeschke, M., Augspurger, T., Hawkins, S., Cloud, P., *et al*. 2022. pandas-dev/pandas: Pandas 1.4.3. Zenodo, URL: https://zenodo.org/record/6702671#.Yv0GaPHMJo4 [Last Accessed 17 August 2022] DOI: 10.5281/zenodo.6702671.

Rees, D.C., Johnson, E. & Lewinson, O. 2009. ABC transporters: the power to change. *Nature Reviews Molecular Cell Biology*. 10: 218-27. doi: 10.1038/nrm2646.

Richardson L.A. 2017. Understanding and overcoming antibiotic resistance. *PLoS Biology*. 15(8): e2003775. doi: 10.1371/journal.pbio.2003775.

Rico, F. & Moy, V.T. 2007. Energy landscape roughness of the streptavidin–biotin interaction. *Journal of Molecular Recognition*. 20(6): 495–501. doi: 10.1002/jmr.841.

Robin, B., Nicol, M., Le, H., Tahrioui, A., Schaumann, A., Vuillemenot, J.B., Vergoz, D., Lesouhaitier, O., *et al*. 2022. MacAB-TolC Contributes to the Development of *Acinetobacter baumannii* Biofilm at the Solid–Liquid Interface. *Frontiers in Microbiology*. 12: 785161. doi: 10.3389/fmicb.2021.785161.

Robinson, G.L. 1951. The Haemolysin of *Bacterium coli*. *Journal of General Microbiology*. 5(4): 788-92. doi: 10.1099/00221287-5-4-788.

Rosano, G.L. & Ceccarelli, E.A. 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology*. 5: 172. doi: 10.3389/fmicb.2014.00172.

Rose, H.H. 2008. Optics of high-performance electron microscopes. *Science and Technology of Advanced Materials*. 9(1): 014107. doi: 10.1088/1468-6996/9/1/014107.

Rossell, D. & Müller, P. 2013. Sequential stopping for high-throughput experiments. *Biostatistics*. 14(1): 75–86. doi: 10.1093/biostatistics/kxs026.

Rowlett, V.W., Mallampalli, V.K.P.S., Karlstaedt, A., Dowhan, W., Taegtmeyer, H., Margolin, W., Vitrac, H. 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.

Ruska, E. 1987. Nobel lecture. The development of the electron microscope and of electron microscopy. *Bioscience Reports*. 7(8): 607-29. doi: 10.1007/BF01127674.

Ryabichko, S., de Melo Ferreira, V., Vitrac, H., Kiyamova, R., Dowhan, W., Bogdanov, M. 2020. Cardiolipin is required in vivo for the stability of bacterial translocon and optimal membrane protein translocation and insertion. *Scientific Reports*. 10: 6296. doi: 10.1038/s41598-020-63280-5.

Ryan, R.M. & Vandenberg, R.J. 2016. Elevating the alternating-access model. *Nature Structural & Molecular Biology*. 23: 187-9. doi: 10.1038/nsmb.3179.

Sali, A. & Blundell, T.L. 1993. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*. 234: 779-815. doi: 10.1006/jmbi.1993.1626.

Sánchez-Magraner, L., Cortajarena, A.L., García-Pacios, M., Arrondo, J.L.R., Agirre, J., Guérin, D.M.A., Goñi, F.M., Ostolaza, H. 2010. Interdomain Ca^{2+} effects in *Escherichia coli* α-haemolysin: Ca^{2+} binding to the C-terminal domain stabilizes both C- and N-terminal domains. *Biochimica et Biophysica Acta*. 1798(6): 1225-33. doi: 10.1016/j.bbamem.2010.03.007.

Sargsyan, K., Grauffel, C. & Lim, C. 2017. How Molecular Size Impacts RMSD Applications in Molecular Dynamics Simulations. *Journal of Chemical Theory & Computation.* 13(4): 1518-24. doi: 10.1021/acs.jctc.7b00028.

Sawle, L. & Gosh, K. 2016. Convergence of Molecular Dynamics Simulation of Protein Native States: Feasibility vs Self-Consistency Dilemma. *Journal of Chemical Theory and Computation*. 12(2): 861-9. doi: 10.1021/acs.jctc.5b00999.

Scheraga, H.A., Khalili, M. & Liwo, A. 2007. Protein-folding dynamics: overview of molecular simulation techniques. *Annual Review of Physical Chemistry*. 58: 57-83. doi: 10.1146/annurev.physchem.58.032806.104614.

Scheres, S.H.W. 2012. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*. 180(3): 519–30. doi: 10.1016/j.jsb.2012.09.006.

Scherzer, O. 1949. The Theoretical Resolution Limit of the Electron Microscope. *Journal of Applied Physics*. 20: 20-9. doi: 10.1063/1.1698233.

Schlegel, S., Hjelm, A., Baumgarten, T., Vikström, D., de Gier, J.W. 2014. Bacterial-based membrane protein production. *Biochimica et Biophysica Acta*. 1843(8): 1739-49. doi: 10.1016/j.bbamcr.2013.10.023.

Schnell, N., Entian, K.D., Schneider, U., Götz, F., Zähner, H., Kellner, R., Jung, G. 1988. Prepeptide sequence of epidermin, a ribosomally synthesized antibiotic with four sulphide-rings. *Nature*. 333(6170): 276-8. doi: 10.1038/333276a0.

Seddon, A.M., Curnow, P. & Booth, P.J. 2004. Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica et Biophysica Acta*. 1666(1-2): 105-17. doi: 10.1016/j.bbamem.2004.04.011.

Senet, P., Maisuradze, G.G., Foulie, C., Delarue, P., Scheraga, H.A. 2008. How main-chains of proteins explore the free-energy landscape in native states. *Proceedings of the National*

Academy of Sciences of the United States of America. 105(50): 19708-13. doi: 10.1073/pnas.0810679105.

Shakhnovich, E., Abkevich, V. & Ptitsyn, O. 1996. Conserved residues and the mechanism of protein folding. *Nature*. 379(6560): 96-8. doi: 10.1038/379096a0.

Shamir, M., Bar-On, Y., Phillips, R., Milo, R. 2016. SnapShot: Timescales in Cell Biology. *Cell*. 164: 1302.e1. doi: 10.1016/j.cell.2016.02.058.

Sharma, S., Zhou, R., Wan, L., Feng, S., Song, K., Xu, C., Li, Y., Liao, M. 2021. Mechanism of LolCDE as a molecular extruder of bacterial triacylated lipoproteins. *Nature Communications*. 12: 4687. doi: 10.1038/s41467-021-24965-1.

Shen, M. & Sali, A. 2006. Statistical potential for assessment and prediction of protein structures. *Protein Science*. 15(11): 2507–24. doi: 10.1110/ps.062416606.

Shi, X., Chen, M., Yu, Z., Bell, J.M., Wang, H., Forrester, I., Villarreal, H., Jakana, J., *et al*. 2019. *In situ* structure and assembly of the multidrug efflux pump AcrAB-TolC. *Nature Communications*. 10: 2635. doi: 10.1038/s41467-019-10512-6.

Siegele, D.A. & Hu, J.C. 1997. Gene expression from plasmids containing the *araBAD* promoter at subsaturating inducer concentrations represents mixed populations. *Proceedings of the National Academy of Sciences of the United States of America.* 94(15): 8168-72. doi: 10.1073/pnas.94.15.8168.

Skoglund, U., Andersson, K., Strandberg, B., Daneholt, B. 1986. Three-dimensional structure of a specific pre-messenger RNP particle established by electron microscope tomography. *Nature*. 319: 560-4. doi: 10.1038/319560a0.

Smirnova, I., Kasho, V. & Kaback, H.R. 2018. Engineered occluded apo-intermediate of LacY. *Proceedings of the National Academy of Sciences of the United States of America*. 115(50): 12716-21. doi: 10.1073/pnas.1816267115.

Smit, B., Hilbers, P.A.J., Esselink, K., Rupert, L.A.M., van Os, N.M., Schlijper, A.G. 1990. Computer simulations of a water/oil interface in the presence of micelles. *Nature*. 348: 624-5. doi: 10.1038/348624a0.

Smith, H.W. & Halls, S. 1967. The Transmissible Nature of the Genetic Factor in *Escherichia coli* that Controls Haemolysin Production. *Journal of General Microbiology*. 47(1): 153-61. doi: 10.1099/00221287-47-1-153.

Smith, JM. 1970. Natural selection and the concept of protein sequence space. *Nature*. 225 (5232): 563-4. doi: 10.1038/225563a0.

Sosunov, V., Mischenko, V., Eruslanov, B., Svetoch, E., Shakina, Y.N., Stern, N., Majorov, K., Sorokoumova, G., *et al*. 2007. Antimycobacterial activity of bacteriocins and their complexes with liposomes. *Journal of Antimicrobial Chemotherapy*. 59(5): 919-25. doi: 10.1093/jac/dkm053.

Souabni, H., dos Santos, W.B., Cece, Q., Catoire, L.J., Puvanendran, D., Bavro, V.N., Picard, M. 2021. Quantitative real-time analysis of the efflux by the MacAB-TolC tripartite efflux pump clarifies the role of ATP hydrolysis within mechanotransmission mechanism. *Communications Biology*. 4: 493. doi: 10.1038/s42003-021-01997-3.

Sousa, S.F., Calixto, A.R., Ferreira, P., Ramos, M.J., Lim, C., Fernandes, P.A. 2020. Activation Free Energy, Substrate Binding Free Energy, and Enzyme Efficiency Fall in a Very Narrow Range of Values for Most Enzymes. *ACS Catalysis*. 10: 8444−53. doi: 10.1021/acscatal.0c01947.

Spitz, O., Erenburg, I.N., Beer, T., Kanonenberg, K., Holland, I.B., Schmitt, L. 2019. Type I Secretion Systems—One Mechanism for All? *Microbiology Spectrum*. 7(2). doi: 10.1128/microbiolspec.PSIB-0003-2018.

Stanley, P., Packman, L.C., Koronakis, V., Hughes, C. 1994. Fatty Acylation of Two Internal Lysine Residues Required for the Toxic Activity of *Escherichia coli* Hemolysin. *Science*. 266(5193): 1992-6. doi: 10.1126/science.7801126.

Stansfeld, P.J., Jefferys, E.E. & Sansom, M.S.P. 2013. Multiscale Simulations Reveal Conserved Patterns of Lipid Interactions with Aquaporins. *Structure*. 21(5): 810–9. doi: 10.1016/j.str.2013.03.005.

Stillwell, W. 2016. Membrane Transport. In *An Introduction to Biological Membranes*. Elsevier. 423–51. doi: 10.1016/B978-0-444-63772-7.00019-1.

Stone, J.E., Hardy, D.J., Ufimtsev, I.S., Schulte, K. 2010. GPU-Accelerated Molecular Modeling Coming Of Age. *Journal of Molecular Graphics and Modelling*. 29(2): 116–25. doi: 10.1016/j.jmgm.2010.06.010.

Subramaniam, S. 2013. Structure of trimeric HIV-1 envelope glycoproteins. *Proceedings of the National Academy of Sciences of the United States of America*. 110(45): E4172–4. doi: 10.1073/pnas.1313802110.

Sweet, J.C., Nowling, R.J., Cickovski, T., Sweet, C.R., Pande, V.S., Izaguirre, J.A. 2013. Long Timestep Molecular Dynamics on the Graphical Processing Unit. *Journal of Chemical Theory and Computation*. 9(8): 3267–81. doi:10.1021/ct400331r.

Tajkhorshid, E., Nollert, P., Jensen, M.Ø., Miercke, L.J.W., O'Connell, J., Stroud, R.M., Schulten, K. 2002. Control of the Selectivity of the Aquaporin Water Channel Family by Global Orientational Tuning. *Science*. 296 (5567): 525-30. doi: 10.1126/science.1067778.

Takata, K., Matsuzaki, M. & Tajika, Y. 2004. Aquaporins: water channel proteins of the cell membrane. *Progress in Histochemistry and Cytochemistry*. 39(1): 1-83. doi: 10.1016/j.proghi.2004.03.001.

Tanford, C. 1978. The Hydrophobic Effect and the Organization of Living Matter. *Science*. 200(4345): 1012-8. doi: 10.1126/science.653353.

Tang, X., Chang, S., Zhang, K., Luo, Q., Zhang, Z., Wang, T., Qiao, W., Wang, C., *et al*. 2021. Structural basis for bacterial lipoprotein relocation by the transporter LolCDE. *Nature Structure & Molecular Biology*. 28(4): 347-55. doi: 10.1038/s41594-021-00573-x.

Tareen A. & Kinney J.B. 2020. Logomaker: beautiful sequence logos in Python. *Bioinformatics.* 36(7): 2272-4. doi: 10.1093/bioinformatics/btz921.

Tegunov, D. & Cramer, P. 2019. Real-time cryo-electron microscopy data preprocessing with Warp. *Nature Methods*. 16: 1146–52. doi: 10.1038/s41592-019-0580-y.

Tiberti, M., Papaleo, E., Bengtsen, T., Boomsma, W., Lindorff-Larsen, K. 2015. ENCORE: Software for Quantitative Ensemble Comparison. *PLoS Computational Biology*. 11(10): e1004415, 00031. doi:10.1371/journal.pcbi.1004415.

Tikhonova, E.B., Devroy, V.K., Lau, S.Y., Zgurskaya, H.I. 2007. Reconstitution of the *Escherichia coli* macrolide transporter: the periplasmic membrane fusion protein MacA stimulates the ATPase activity of MacB. *Molecular Microbiology*. 63(3): 895–910. doi: 10.1111/j.1365-2958.2006.05549.x.

Timmons, M.D., Knutson, B.L., Nokes, S.E., Strobel, H.J., Lynn, B.C. 2009. Analysis of composition and structure of *Clostridium thermocellum* membranes from wild-type and ethanol-adapted strains. *Applied Microbiology and Biotechnology*. 82: 929–39. doi: 10.1007/s00253-009-1891-1.

Thanabalu, T., Koronakis, E., Hughes, C. Koronakis, V. 1998. Substrate-induced assembly of a contiguous channel for protein export from *E. coli*: reversible bridging of an inner-membrane translocase to an outer membrane exit pore. *EMBO Journal*. 17: 6487–6496. doi: 10.1093/emboj/17.22.6487.

Theobald, DL. 2005. Rapid calculation of RMSD using a quaternion-based characteristic polynomial. *Acta Crystallographica A*. 61: 478-80. doi: 10.1107/S0108767305015266.

Thomas, C. & Tampé, R. 2020. Structural and Mechanistic Principles of ABC Transporters *Annual Review of Biochemistry*. 89: 605–36. doi: 10.1146/annurev-biochem-011520-105201.

Thon, F. 1966. Imaging properties of the electron microscope near the theoretical limit of resolution. *Proceedings of the Sixth International Conference for Electron Microscopy*. Kyoto, Japan. p. 23.

Togashi, Y. & Flechsig, H. 2018. Coarse-Grained Protein Dynamics Studies Using Elastic Network Models. *International Journal of Molecular Sciences*. 19(12): 3899. doi: 10.3390/ijms19123899.

Tommaso, P.D., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.M., Taly, J.F., Notredame, C. 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Research.* 39(2): W13–7. doi: /10.1093/nar/gkr245.

Tribet, C., Audebert, R. & Popot, J.L. 1996. Amphipols: polymers that keep membrane proteins soluble in aqueous solutions. *Proceedings of the National Academy of Sciences of the United States of America*. 93(26): 15047-50. doi: 10.1073/pnas.93.26.15047.

Tschampel, S.M., Kirschner, K.N. & Woods, R.J. 2006. Incorporation of Carbohydrates into Macromolecular Force Fields: Development and Validation. In *NMR Spectroscopy and Computer Modeling of Carbohydrates*. Vliegenthart, J.F.G & Woods, R.J, Eds. Washington: American Chemical Society: 235-57. doi: 10.1021/bk-2006-0930.ch013.

Tuckerman, M.E. & Martyna, G.J. 2000. Understanding Modern Molecular Dynamics: Techniques and Applications. *The Journal of Physical Chemistry B*. 104(2): 159–78. doi: 10.1021/jp992433y.

Turoňová, B., Schur, F.K.M., Wan, W., Briggs, J.A.G. 2017. Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4 Å. *Journal of Structural Biology*. 199: 187-95. doi: 10.1016/j.jsb.2017.07.007.

Valeva, A., Siegel, I., Wylenzek, M., Wassenaar, T.M., Weis, S., Heinz, N., Schmitt, R., Fischer, C., *et al*. 2008. Putative identification of an amphipathic α-helical sequence in hemolysin of Escherichia coli (HlyA) involved in transmembrane pore formation. *Biological Chemistry*. 389(9): 1201-7. doi: 10.1515/BC.2008.136.

van Belkum, M.J., Worobo, R.W. & Stiles, M.E. 1997. Double-glycine-type leader peptides direct secretion of bacteriocins by ABC transporters: colicin V secretion in Lactococcus lactis. *Molecular Microbiology*. 23(6): 1293-301. doi: 10.1046/j.1365-2958.1997.3111677.x.

van Heel, M. 2013. Finding trimeric HIV-1 envelope glycoproteins in random noise. *Proceedings of the National Academy of Sciences of the United States of America*. 110(45): E4175–7. doi: 10.1073/pnas.1314353110.

van den Berg, B., Black, P.N., Clemons, W.M., Rapoport, T.A. 2004. Crystal Structure of the Long-Chain Fatty Acid Transporter FadL. *Science*. 304(5676): 1506-9. doi: 10.1126/science.1097524.

van Nimwegen, E., Crutchfield, J.P. & Huynen, M. 1999. Neutral evolution of mutational robustness. *Proceedings of the National Academy of Sciences of the United States of America*. 96(17): 9716-20. doi: 10.1073/pnas.96.17.9716.

Vasiliou, V., Vasiliou, K. & Nebert, D.W. 2009. Human ATP-binding cassette (ABC) transporter family. *Human Genomics*. 3(3): 281-90. doi: 10.1186/1479-7364-3-3-281.

Verkman, A.S. 2013. Aquaporins. *Current Biology*. 23(2): R52-5. doi: 10.1016/j.cub.2012.11.025.

Verlet, L. 1967. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Reviews*. 159: 98-103. doi: 10.1103/PHYSREV.159.98.

Vilas, J.L., Heymann, J.B., Tagare, H.D., Ramirez-Aportela, E., Carazo, J.M., Sorzano, C.O.S. 2020. Local resolution estimates of cryoEM reconstructions. *Current Opinion in Structural Biology*. 64: 74–8. doi: 10.1016/j.sbi.2020.06.005.

Villa, A., Fan, H., Wassenaar, T., Mark, A.E. 2007. How Sensitive Are Nanosecond Molecular Dynamics Simulations of Proteins to Changes in the Force Field? *The Journal of Physical Chemistry B*. 111(21): 6015–25. doi: 10.1021/jp068580v.

Wagner, W., Vogel, M. & Goebel, W. 1983. Transport of hemolysin across the outer membrane of *Escherichia coli* requires two functions. *Journal of Bacteriology*. 154(1): 200–10. doi: 10.1128/jb.154.1.200-210.1983.

Wagner, S., Klepsch, M.M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., Högbom, M., can Wijk, K.J., *et al*. 2008a. Tuning *Escherichia coli* for membrane protein overexpression. *Proceedings of the National Academy of Sciences of the United States of America*. 105(38): 14371–6. doi: 10.1073/pnas.0804090105.

Wagner, I.D., Zhao, W., Zhang, C.L., Romanek, C.S., Rohde, M., Wiegel, J. 2008b. *Thermoanaerobacter uzonensis sp. nov*., an anaerobic thermophilic bacterium isolated from a hot spring within the Uzon Caldera, Kamchatka, Far East Russia. *International Journal of Systematic and Evolutionary Microbiology*. 58(Pt 11): 2565-73. doi: 10.1099/ijs.0.65343-0.

Walton, E.B. & VanVliet, K.J. 2006. Equilibration of experimentally determined protein structures for molecular dynamics simulation. *Physical Review E*. 74: 061901. doi: 10.1103/PhysRevE.74.061901.

Wan, S., Sinclair, R.C. & Coveney, P.V. 2021. Uncertainty quantification in classical molecular dynamics. *Philosophical Transactions Royal Society A*. 379: 20200082. doi: 10.1098/rsta.2020.0082.

Wandersman, C. & Delepelaire, P. 1990. TolC, an *Escherichia coli* outer membrane protein required for hemolysin secretion. *Proceedings of the National Academy of Sciences of the United States of America*. 87(12): 4776-80. doi: 10.1073/pnas.87.12.4776.

Wang, K., Sitsel, O., Meloni, G., Autzen, H.E., Andersson, M., Klymchuk, T., Nielsen, A.M., Rees, D.C., *et al*. 2014. Structure and Mechanism of Zn(2+)-Transporting P-Type Atpases. *Nature*. 514(7523): 518-22. doi: 10.1038/nature13618.

Wang, Z., Fan, G., Hryc, C.F., Blaza, J.N., Serysheva, I.I., Schmid, M.F., Chiu, W., Luisi, B.F., *et al*. 2017. An allosteric transport mechanism for the AcrAB-TolC multidrug efflux pump. *eLife*. 6: e24905. doi: 10.7554/eLife.24905.001.

Wassenaar, T.A. & Mark, A.E. 2005. The Effect of Box Shape on the Dynamic Properties of Proteins Simulated under Periodic Boundary Conditions. *Journal of Computational Chemistry*. 27: 316–25. doi: 10.1002/jcc.20341.

Watson, H. 2015. Biological membranes. *Essays in Biochemistry*. 59: 43-69. doi: 10.1042/bse0590043.

Webby, M.N., Oluwole, A.O., Pedebos, C., Inns, P.G., Olerinyova, A., Prakaash, D., Housden, N.G., Benn, G., *et al*. 2022. Lipids mediate supramolecular outer membrane protein assembly in bacteria. *Science Advances*. 8(44): eadc9566. doi: 10.1126/sciadv.adc9566.

Weinkam, P., Pons, J. & Sali, A. 2012. Structure-based model of allostery predicts coupling between distant sites. *Proceedings of the National Academy of Sciences of the United States of America*. 109(13): 4875-80. doi: 10.1073/pnas.1116274109.

Welch, R.A., Dellinger, E.P., Minshew, B., Falkow, S. 1981. Haemolysin contributes to virulence of extra-intestinal *E. coli* infections. *Nature*. 294(5842): 665-7. doi: 10.1038/294665a0.

Welch, R.A. & Falkow, S. 1984. Characterization of *Escherichia coli* Hemolysins Conferring Quantitative Differences in Virulence. *Infection and Immunity*. 43(1): 156-60. doi: 10.1128/iai.43.1.156-160.1984.

Welch, R.A. 1991. Pore-forming cytolysins of Gram-negative bacteria. *Molecular Microbiology*. 5(3): 521-8. doi: 10.1111/j.1365-2958.1991.tb00723.x.

Welford, B.P. 1962. Note on a Method for Calculating Corrected Sums of Squares and Products. *Technometrics*. 4(3): 419-20. doi: 10.1080/00401706.1962.10490022.

Welter, D.K., Ruaud, A., Henseler, Z.M., De Jong, H.N., van Coeverden de Groot, P., Michaux, J., Gormezano, L., Waters, J.L., *et al*. 2021. Free-Living, Psychrotrophic Bacteria of the Genus *Psychrobacter* Are Descendants of Pathobionts. *mSystems*. 6(2): e00258-21. doi: 10.1128/mSystems.00258-21.

Wiener, N. 1964. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Cambridge, Massachusetts, United States of America: MIT Press.

Wong, L.H., Gatta, A.T. & Levine, T.P. 2019. Lipid transfer proteins: the lipid commute via shuttles, bridges and tubes. *Nature Reviews Molecular Cell Biology*. 20: 85–101. doi: 10.1038/s41580-018-0071-5.

Wu, E.L., Cheng, X., Jo, S., Rui, H., Song, K.C., Dávila-Contreras, E.M., Qi, Y., Lee, J., *et al*. 2014. CHARMM-GUI Membrane Builder Toward Realistic Biological Membrane Simulations. *Journal of Computational Chemistry*. 35: 1997-2004. doi: 10.1002/jcc.23702.

Xu, Y., Song, S., Moeller, A., Kim, N., Piao, S., Sim, S.H., Kang, M., Yu, W., *et al*. 2011. Functional Implications of an Intermeshing Cogwheel-like Interaction between TolC and MacA in the Action of Macrolide-specific Efflux Pump MacAB-TolC. *Journal of Biological Chemistry*. 286(15): 13541-9. doi: 10.1074/jbc.M110.202598.

Xu, Y., Erdjument-Bromage, H., Phoon, C.K.L., Neubert, T.A., Ren, M., Schlame, M. 2021. Cardiolipin remodeling enables protein crowding in the inner mitochondrial membrane. *The EMBO Journal*. 40: e108428. doi: 10.15252/embj.2021108428.

Yahyaoui, R. & Pérez-Frías, J. 2020. Amino Acid Transport Defects in Human Inherited Metabolic Disorders. *International Journal of Molecular Sciences*. 21(1): 119. doi: 10.3390/ijms21010119.

Yakushi, T., Masuda, K., Narita, S., Matsuyama, S., Tokuda, H. 2000. A new ABC transporter mediating the detachment of lipid-modified proteins from membranes. *Nature Cell Biology*. 2(4): 212-8. doi: 10.1038/35008635.

Yamanaka, H., Kobayashi, H., Takahashi, E., Okamoto, K. 2008. MacAB Is Involved in the Secretion of *Escherichia coli* Heat-Stable Enterotoxin II. *Journal of Bacteriology*. 190(23): 7693–8. doi: 10.1128/JB.00853-08.

Yamashita, A., Singh, S.K., Kawate, T., Jin, Y., Gouaux, E. 2005. Crystal structure of a bacterial homologue of Na^+/Cl^--dependent neurotransmitter transporters. *Nature*. 437: 215-23. doi: 10.1038/nature03978.

Yang, H.B., Hou, W.T., Cheng, M.T., Jiang, Y.L., Chen, Y., Zhou, C.Z. 2018. Structure of a MacAB-like efflux pump from *Streptococcus pneumoniae*. *Nature Communications*. 9: 196. doi: 10.1038/s41467-017-02741-4.

Zacharova, E.V., Mitrofanova, T.I., Krasilnikova, E.N., Kondratieva, E.N. 1993. Thermohydrogenium kirishiense gen. nov. and sp. nov., a new anaerobic thermophilic bacterium. *Archives of Microbiology*. 160: 492-7. doi: 10.1007/BF00245311.

Zaitseva, J., Oswald, C., Jumpertz, T., Jenewein, S., Wiedenmann, A., Holland, I.B., Schmitt, L. 2006. A structural analysis of asymmetry required for catalytic activity of an ABC-ATPase domain dimer. *EMBO Journal*. 25(14): 3432-43. doi: 10.1038/sj.emboj.7601208.

Zeth, K. & Thein, M. 2010. Porins in prokaryotes and eukaryotes: common themes and variations. *Biochemical Journal*. 431(1): 13–22. doi: 10.1042/BJ20100371.

Zhang, P., Bos, E., Heymann, J., Gnaegi, H., Kessel, M., Peters, P.J., Subramaniam, S. 2004. Direct visualization of receptor arrays in frozen-hydrated sections and plunge-frozen specimens of *E. coli* engineered to overproduce the chemotaxis receptor Tsr. *Journal of Microscopy*. 216(1): 76-83. doi: 10.1111/j.0022-2720.2004.01395.x.

Zhang, X. & Zhou, Z.H. 2011. Limiting factors in atomic resolution cryo electron microscopy: no simple tricks. *Journal of Structural Biology*. 175(3): 253-63. doi: 10.1016/j.jsb.2011.05.004.

Zhang, Z., Kuipers, G., Niemiec, L., Baumgarten, T., Slotboom, D.J., de Gier, J.W., Hjelm, A. 2015. High-level production of membrane proteins in *E. coli* BL21(DE3) by omitting the inducer IPTG. *Microbial Cell Factories*. 14: 142. doi: 10.1186/s12934-015-0328-z.

Zhang, X., Tu, B., Dai, L.R., Lawson, P.A., Zheng, Z.Z., Liu, L.Y., Deng, Y., Zhang, H., *et al*. 2018. *Petroclostridium xylanilyticum gen. nov*., *sp. nov*., a xylan-degrading bacterium isolated from an oilfield, and reclassification of clostridial cluster III members into four novel genera in a new *Hungateiclostridiaceae fam. nov*. *International Journal of Systematic and Evolutionary Microbiology*. 68(10): 3197-211. doi: 10.1099/ijsem.0.002966.

Zhao, H., Lee, J. & Chen, J. 2022. The hemolysin A secretion system is a multi-engine pump containing three ABC transporters. *Cell*. 185: 3329–40. doi: 10.1016/j.cell.2022.07.017.

Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., Agard, D.A. 2017. MotionCor2 - anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods*. 14(4): 331-2. doi: 10.1038/nmeth.4193.

Zhu, D., Wang, X., Fang, Q., Van Etten, J.L., Rossmann, M.G., Rao, Z., Zhang, X. 2018. Pushing the resolution limit by correcting the Ewald sphere effect in single-particle Cryo-EM reconstructions. *Nature Communications*. 9: 1552. doi: 10.1038/s41467-018-04051-9.

Zivanov, J., Otón, J., Ke, Z., von Kügelgen, A., Pyle, E., Qu, K., Morado, D., Castaño-Díez, D., *et al*. 2022. A Bayesian approach to single-particle electron cryo-tomography in RELION-4.0. *eLife*. 11: e83724. doi: 10.7554/eLife.83724.

Appendix A: Statistical Methods

Hypothesis Testing: Permutation Test and Monte Carlo Sampling

For hypothesis testing, a distinction is made between theoretical and empirical distributions to test the null hypothesis against. While theoretical distributions are good when the underlying assumptions are valid for the data, and good estimates can be obtained for the sample mean and variance, it is more difficult to test assumptions when the data is sparse. One difficulty is that with less data there is the risk that the calculated sample variance is very different to the population variance, potentially invalidating the results of the test. One way to overcome these difficulties to use an empirical test on the data. Say you have two samples, A and B, and wish to test if the mean difference between them is 0 (i.e $\mu_A - \mu_B = 0$). A permutation test works by randomly reshuffling our N total observations, calculating the resulting mean difference ($\hat{\mu}_A - \hat{\mu}_B$), and repeating this procedure R times to give a resulting distribution for the estimate ($\hat{\mu}_A - \hat{\mu}_B$) (**Figure A-1**).

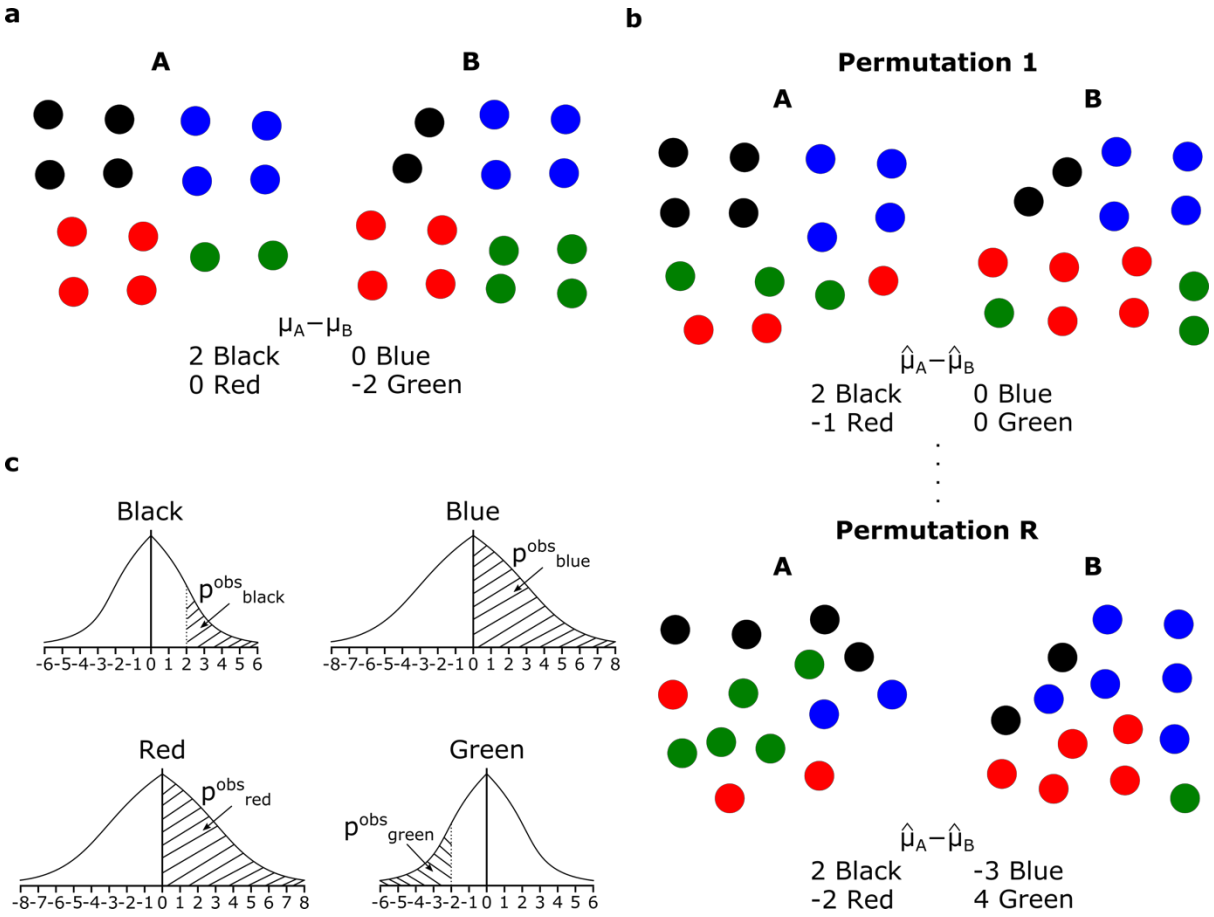


Figure A-1. Permutation test example. a) There are two populations, A and B, each containing black, blue, red, and green balls. The differences between the colours is given by $\mu_A - \mu_B$. **b)** The permutation test works by shuffling balls randomly between the two sample groups and recalculating the differences ($\hat{\mu}_A - \hat{\mu}_B$) for an R number of times. **c)** From the permutation test, expected differences for each colour ball are calculated and an empirical distribution produced. As can be seen, we expect the average difference for black and green to be 1/3 and for red and blue to be 1/2 which converges to the population distribution as the resample rate increases.

The permutation test works well when N is small; however if N is too large, the resulting distribution may not reflect the full variance in the data due to sampling, as not all permutations can be calculated. This becomes an issue as the tail ends of the distribution will show smaller values than expected, and potentially lead to overestimates of "significant" results. One way to correct for this is to apply Monte Carlo sampling; here, instead of reshuffling all of the data we now take a sample, n, from A and B and calculate our mean difference as before. As long as n is sufficiently large to not be skewed by outliers, but not too

large (*e.g.* 0.1–0.3N) the resulting distribution should give a better estimate of the tails of the difference distribution (**Figure A-2**).

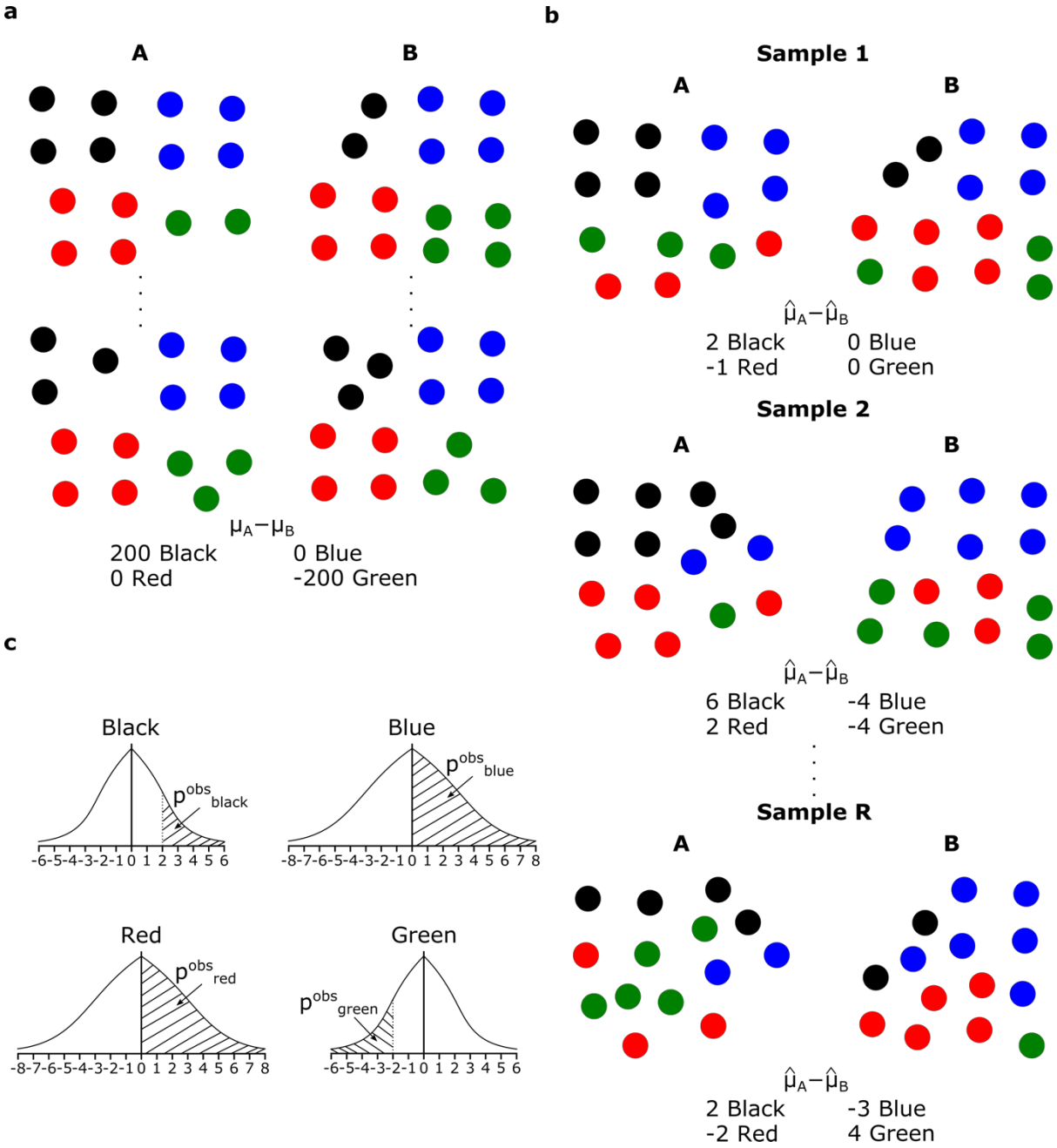


Figure A-2. Monte Carlo Permutation Test. **a)** Instead of 14 balls in each population, A and B, say we have 100 more of each ball. Here, a permutation test is unrealistic as the chances of sampling each permutation decreases as N increases. Instead, we take a Monte Carlo sample. **b)** For a Monte Carlo permutation test, we take a random sample (n=14) for A and B and compute the difference ($\hat{\mu}_A - \hat{\mu}_B$) for R iterations. **c)** The empirical distribution is calculated based on the samples, which should approximate the same probabilities as the permutation test for a sufficient number of iterations.

Appendix B: Data Pipelines

The code for the data pipelines developed for this thesis are available on GitHub (https://github.com/AMKCam) [last accessed 24th July 2023]. An explanation for each analysis Notebook is given in **Table B**.

Table B. Data pipelines used in this thesis.

Analysis NoteBook	Chapters Used In	Data Presentation [Chapter Figures]	Explanation
Coarse-Grain-Analysis-HlyB	II	Figure 2.3.3-3 Figure 2.3.4.1-1 Figure 2.3.4.1-4 Figure 2.3.5-1 Figure 2.3.5-2 Figure 2.3.5-3 Figure 2.3.5-4	This NoteBook analyses coarse grain molecular dynamics trajectory data for HlyB
Coarse-Grain-Analysis-Peptidase	II	Figure 2.3.3-3 Figure 2.3.4.1-2 Figure 2.3.4.1-3 Figure 2.3.4.1-4 Figure 2.3.5-1 Figure 2.3.5-2 Figure 2.3.5-3 Figure 2.3.5-5	This NoteBook analyses coarse grain molecular dynamics trajectory data for PCAT
HlyB-Atomistic-Trajectories-Analysis	II	Figure 2.3.3-2 Figure 2.3.4.2-1	This NoteBook analyses atomistic molecular dynamics trajectory data for HlyB
HlyB-Peptidase-Homologoues-Analysis	II	Figure 2.3.6-1 Figure 2.3.6-3	This NoteBook analyses coarse grain molecular dynamics trajectory data mined

			from MemProtMD (see Table 2.2.2.4-1)
Flow-Cytometry-Data-Analysis	III	Figure 3.3.2-2 Figure 3.3.2-3 Figure 3.3.2-4 Figure 3.3.3-1 Figure 3.3.3-2	This NoteBook analyses flow cytometry data for T1SS production in *E. coli*
Coarse-Grain-Analysis-MacB	V	Figure 5.3.2-3 Figure 5.3.3-1 Figure 5.3.4-1 Figure 5.3.4-2 Figure 5.3.4-3	This NoteBook analyses coarse grain molecular dynamics trajectory data for MacB
Sequence-Conservation-Analysis	VI	Figure 6.3.3-1 Figure 6.3.3-2 Figure 6.4.4-2	This NoteBook analyses aligned sequence data for HlyB, peptidases, and HlyA-like sequences to visualise sequence conservation at key sites

Appendix C: Raw Data

Table C. Flow Cytometry Raw Data

Column Labels (inputs):

1. **Date (DDMMYYY):** date of the experiment
2. **Strain:** bacterial strain used – C43Δ*acrAB*, BL21(DE3), MG1655 parent, MG1655 Δ*cls*, Minicell
3. **Media:** media used to grow the strain (LB, 2xYT)
4. **Expression Order:** order of plasmid induction (HlyA 30 min before HlyB/D, HlyA 1 hr before HlyB/D, HlyB/D 30 min before HlyA, HlyB/D 1 hr before HlyA, same time)
5. **Induction:** inducer concentration (single – 6.8 mM arabinose + 1 mM IPTG, double – 13.6 mM arabinose, 2 mM IPTG)
6. **Carbenicillin concentration (μg/mL):** carbenicillin antibiotic concentration used (0, 100 μg/mL, 50 μg/mL)
7. **Kanamycin concentration (μg/mL):** kanamycin antibiotic concentration used (0 μg/mL, 25 μg/mL, 30 μg/mL, 50 μg/mL)
8. **Chloramphenicol concentration (μg/mL):** chloramphenicol antibiotic concentration used (0 μg/mL, 17.5 μg/mL)
9. **Temperature (°C):** temperature used for induction (20°C, 25 °C, 37 °C)
10. **CaCl (mM):** calcium chloride concentration used for induction (0 mM, 1 mM, 2 mM, 5 mM, 8 mM, 10 mM)
11. **Total expression time (hr):** total time after inducers added (0 hr, 0.5 hr, 1 hr, 1.5 hr, 2 hr, 2.5 hr, 3 hr, 3.5 hr)
12. **Non-expressing (%):** percentage of cells in Q3 quadrant
13. **HlyA (%):** percentage of cells in Q4 quadrant
14. **T1SS (%):** percentage of cells in Q2 quadrant
15. **Excess (%):** percentage of cells in Q1 quadrant
16. **Total (%):** quadrant Q1+Q2+Q3+Q4
17. **T1SS:Excess Ratio:** ratio of cells in Q2 vs Q1 quadrant
18. **Control?:** is this a control sample?

19. **Primary Ig?:** was primary anti-rabbit added?
20. **Secondary Ig?:** was secondary Cy5 anti-mouse added?
21. **Repeat measurement?:** is this a repeat measurement? (i.e measured twice in the flow cytometer?)
22. **Normalised T1SS (%):** quadrant Q2 standardised on quadrant Q1 (see **Chapter IIII** section **2.6**)

Date (DDMMYYYY)	Strain	Media	Expression Order	Induction	Carbenicillin Concentration (µg/ml)	Kanamycin Concentration (µg/ml)	Chloramphenicol Concentration (µg/ml)	Temperature (°C)	CaCl (mM)	Total Expression Time (hr)	Non-expressing (%)	HlyA (%)	T1SS (%)	Excess (%)	Total (%)	T1SS:Excess Ratio	Control?	Primary Ig?	Secondary Ig?	Repeat Measurement?	Normalised_T1SS (%)
17.05.2021	C43deltaAcrAB	LB	HlyA 30 min before HlyBD	double	100	30	0	20	5	1.5	59.1	40.7	0.1	0.1	100	1	No	Yes	Yes	No	-6.2587206
17.05.2021	C43deltaAcrAB	2xYT	HlyA 30 min before HlyBD	double	100	30	0	20	5	1.5	91.7	8	0.1	0.2	100	0.5	No	Yes	Yes	No	-6.2587206
24.05.2021	C43deltaAcrAB	LB	HlyA 30 min before HlyBD	single	100	30	0	20	5	2	44	55.5	0.2	0.2	99.9	1	No	Yes	Yes	No	-6.1643176
24.05.2021	C43deltaAcrAB	LB	same time	single	100	30	0	20	5	2	72.6	27.2	0.1	0.1	100	1	No	Yes	Yes	No	-6.2587206
24.05.2021	C43deltaAcrAB	LB	HlyBD 30 min before HlyA	single	100	30	0	20	5	2	68.1	31.7	0.1	0.1	100	1	No	Yes	Yes	No	-6.2587206
24.05.2021	C43deltaAcrAB	2xYT	HlyA 30 min before HlyBD	single	100	30	0	20	5	2	79.9	19.6	0.2	0.3	100	0.66666667	No	Yes	Yes	No	-6.1643176
24.05.2021	C43deltaAcrAB	2xYT	same time	single	100	30	0	20	5	2	44.1	55.5	0.3	0.1	100	3	No	Yes	Yes	No	-6.0699145
24.05.2021	C43deltaAcrAB	2xYT	HlyBD 30 min before HlyA	single	100	30	0	20	5	2	90	9.6	0.2	0.3	100.1	0.66666667	No	Yes	Yes	No	-6.1643176
09.06.2021	C43deltaAcrAB	LB	same time	single	100	30	0	37	0	1	7.8	91.7	0.5	0	100	inf	No	Yes	Yes	No	-5.8811085
09.06.2021	C43deltaAcrAB	LB	same time	single	100	30	0	37	5	1	9.5	89.4	1	0.1	100	10	No	Yes	Yes	No	-5.4090933
09.06.2021	BL21(DE3)	2xYT	same time	single	100	30	0	37	0	1	16.4	55.4	22.6	5.6	100	4.03571429	No	Yes	Yes	No	14.9819631
09.06.2021	BL21(DE3)	2xYT	same time	single	100	30	0	37	5	1	15.2	20.7	45.5	18.6	100	2.44623656	No	Yes	Yes	No	36.600259
18.06.2021	C43deltaAcrAB	2xYT	same time	single	100	30	0	37	5	1	14.4	83.5	1.8	0.2	99.9	9	No	Yes	Yes	No	-4.653869
18.06.2021	C43deltaAcrAB	2xYT	same time	single	100	30	0	20	5	1	45.7	54	0.2	0.1	100	2	No	Yes	Yes	No	-6.1643176
18.06.2021	C43deltaAcrAB	LB	same time	single	100	30	0	37	5	1	7.4	91.4	1.1	0.1	100	11	No	Yes	Yes	No	-5.3146902
18.06.2021	C43deltaAcrAB	LB	same time	single	100	30	0	20	5	1	36.7	63	0.2	0.1	100	2	No	Yes	Yes	No	-6.1643176
18.06.2021	BL21(DE3)	2xYT	same time	single	100	30	0	37	5	1	4.1	73.6	20.3	2	100	10.15	No	Yes	Yes	No	12.8106932
18.06.2021	BL21(DE3)	2xYT	same time	single	100	30	0	20	5	1	28.9	69.9	0.9	0.3	100	3	No	Yes	Yes	No	-5.5034963
18.06.2021	BL21(DE3)	LB	same time	single	100	30	0	37	5	1	4.8	93.2	1.9	0.1	100	19	No	Yes	Yes	No	-4.5594659
18.06.2021	BL21(DE3)	LB	same time	single	100	30	0	20	5	1	17.9	81.4	0.6	0.2	100.1	3	No	Yes	Yes	No	-5.7867054
21.06.2021	BL21(DE3)	2xYT	same time	single	100	30	0	37	5	1.5	3.7	82.2	13	1.1	100	11.8181818	No	Yes	Yes	No	5.9192714
21.06.2021	BL21(DE3)	2xYT	HlyA 30 min before HlyBD	single	100	30	0	37	5	1.5	4.3	94.9	0.9	0	100.1	inf	No	Yes	Yes	No	-5.5034963
21.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	1.5	3.1	6.3	64.8	25.7	99.9	2.52140078	No	Yes	Yes	No	54.8200455
21.06.2021	BL21(DE3)	2xYT	same time	double	100	30	0	37	5	1.5	3	91.7	5	0.3	100	16.6666667	No	Yes	Yes	No	-1.6329717

21.06.2021	BL21(DE3)	2xYT	HlyA 30 min before HlyBD	double	100	30	0	37	5	1.5	3.1	96.2	0.7	0	100	inf	No	Yes	Yes	No	-5.6923024
21.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	double	100	30	0	37	5	1.5	7.5	16.8	61.7	14.1	100.1	4.37588653	No	Yes	Yes	No	51.8935513
21.06.2021	BL21(DE3)	2xYT	same time	single	100	30	0	37	5	1.5	2.8	82.1	14	1.1	100	12.7272727	No	Yes	Yes	Yes	6.86330178
24.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	1.5	5.1	4	59.5	31.4	100	1.89490446	No	Yes	Yes	No	49.8166845
24.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	50	0	37	5	1.5	5.8	5.7	55.7	32.9	100.1	1.69300912	No	Yes	Yes	No	46.229369
24.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	50	50	0	37	5	1.5	4.8	1.9	52.9	40.4	100	1.30940594	No	Yes	Yes	No	43.5860839
24.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	50	30	0	37	5	1.5	6.1	13.4	66.2	14.3	100	4.62937063	No	Yes	Yes	No	56.1416881
24.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	0	1.5	7.7	6.4	56.6	29.3	100	1.93174061	No	Yes	Yes	No	47.0789964
24.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	1	1.5	6	2.3	54.9	36.8	100	1.49184783	No	Yes	Yes	No	45.4741447
24.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	2	1.5	7	6	59.6	27.4	100	2.17518248	No	Yes	Yes	No	49.9110875
24.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	1.5	7	9.4	58.7	24.9	100	2.35742972	No	Yes	Yes	No	49.0614602
24.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	8	1.5	7	2.2	50.8	40	100	1.27	No	Yes	Yes	No	41.6036201
24.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	10	1.5	7.3	2.8	54.4	35.6	100.1	1.52808989	No	Yes	Yes	No	45.0021295
28.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	0	0	99.7	0.3	0	0.1	100.1	0	No	Yes	Yes	No	-6.3531237
28.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	0	99.7	0.1	0	0.2	100	0	No	Yes	Yes	No	-6.3531237
28.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	0	0.5	99.9	0.1	0	0	100	Undefined	No	Yes	Yes	No	-6.3531237
28.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	0.5	99.8	0.1	0	0.1	100	0	No	Yes	Yes	No	-6.3531237
28.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	0	1	20.4	1.2	16.2	62.3	100.1	0.2600321	No	Yes	Yes	No	8.94016864
28.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	1	20.4	6.1	36.5	37	100	0.98648649	No	Yes	Yes	No	28.1039855
28.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	0	1.5	11.1	7.7	49.9	31.3	100	1.5942492	No	Yes	Yes	No	40.7539927
28.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	1.5	14.7	25.1	49.2	11	100	4.47272727	No	Yes	Yes	No	40.0931715
28.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	0	2	3.3	10.4	70.2	16.1	100	4.36024845	No	Yes	Yes	No	59.9178096
28.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	2	4.2	32	59	4.7	99.9	12.5531915	No	Yes	Yes	No	49.3446693
28.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	0	2.5	2.6	15.8	71.8	9.7	99.9	7.40206186	No	Yes	Yes	No	61.4282583
28.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	2.5	1.1	36.5	61.2	1.2	100	51	No	Yes	Yes	No	51.4215361
30.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	0	99.7	0.3	0	0	100	undefined	No	Yes	Yes	No	-6.3531237
30.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	0.5	99.9	0	0	0	99.9	undefined	No	Yes	Yes	No	-6.3531237

30.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	1	40.1	9.7	26.4	23.8	100	1.1092437	No	Yes	Yes	No	18.5692786
30.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	1.5	16.5	36.3	40.9	6.2	99.9	6.59677419	No	Yes	Yes	No	32.2577192
30.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	2	6.3	41.4	49.1	3.3	100.1	14.8787879	No	Yes	Yes	No	39.9987684
30.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	2.5	4.8	53.1	40.6	1.5	100	27.0666667	No	Yes	Yes	No	31.9745101
30.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	3	1.9	51.6	45.8	0.7	100	65.4285714	No	Yes	Yes	No	36.8834682
30.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	3.5	2.5	82.2	15.1	0.2	100	75.5	No	Yes	Yes	No	7.90173521
13.07.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	0	99.9	0.1	0	0	100	undefined	No	Yes	Yes	No	-6.3531237
13.07.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	50	30	0	37	5	0	99.9	0	0	0	99.9	undefined	No	Yes	Yes	No	-6.3531237
13.07.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	0.5	99.9	0.1	0	0	100	undefined	No	Yes	Yes	No	-6.3531237
13.07.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	50	30	0	37	5	0.5	99.9	0	0	0.1	100	0	No	Yes	Yes	No	-6.3531237
13.07.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	1	66.7	11.9	12	9.4	100	1.27659575	No	Yes	Yes	No	4.97524101
13.07.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	50	30	0	37	5	1	64	17.8	12.1	6	99.9	2.01666667	No	Yes	Yes	No	5.06964405
13.07.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	1.5	42.6	35.3	18.5	3.5	99.9	5.28571429	No	Yes	Yes	No	11.1114385
13.07.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	50	30	0	37	5	1.5	38.3	32.6	24	5.1	100	4.70588235	No	Yes	Yes	No	16.3036057
13.07.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	2	20	44.5	32.3	3.2	100	10.09375	No	Yes	Yes	No	24.1390579
13.07.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	50	30	0	37	5	2	16.6	49.4	31.1	2.9	100	10.7241379	No	Yes	Yes	No	23.0062214
13.07.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	2.5	12.5	55.8	30.1	1.7	100.1	17.7058824	No	Yes	Yes	No	22.062191
13.07.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	50	30	0	37	5	2.5	9.6	57.7	31.5	1.2	100	26.25	No	Yes	Yes	No	23.3838336
13.07.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	3	8.7	62.2	28.3	0.9	100.1	31.4444444	No	Yes	Yes	No	20.3629363
13.07.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	50	30	0	37	5	3	6.3	61.8	31.3	0.6	100	52.1666667	No	Yes	Yes	No	23.1950275
13.07.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	3.5	4.8	63	31.6	0.6	100	52.6666667	No	Yes	Yes	No	23.4782366
13.07.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	50	30	0	37	5	3.5	6.7	60.8	32.2	0.4	100.1	80.5	No	Yes	Yes	No	24.0446549
19.07.2021	Minicell	2xYT	same time	single	50	25	17.5	37	5	1.5	0.5	99.2	0.2	0.1	100	2	No	Yes	Yes	No	-6.1643176
19.07.2021	Minicell	2xYT	same time	single	50	25	17.5	20	5	1.5	5.7	93.8	0.4	0	99.9	inf	No	Yes	Yes	No	-5.9755115
19.07.2021	Minicell	2xYT	same time	single	50	25	17.5	37	5	1.5	0.6	99.1	0.1	0.1	99.9	1	No	Yes	Yes	No	-6.2587206
19.07.2021	Minicell	2xYT	HlyA 30 min before HlyBD	single	50	25	17.5	37	5	1.5	0.5	99.2	0.3	0	100	inf	No	Yes	Yes	No	-6.0699145
19.07.2021	Minicell	2xYT	HlyBD 30 min before HlyA	single	50	25	17.5	37	5	1.5	0.6	99.2	0.2	0	100	inf	No	Yes	Yes	No	-6.1643176

22.07.2021	Minicell	2xYT	same time	single	50	25	17.5	20	5	17	60	39.1	0.8	0.1	100	8	No	Yes	Yes	No	-5.5978993
22.07.2021	Minicell	LB	same time	single	50	25	17.5	20	5	17	66.3	28.2	5	0.5	100	10	No	Yes	Yes	No	-1.6329717
22.07.2021	Minicell	2xYT	same time	single	50	25	17.5	20	5	17	46.3	53.4	0.3	0	100	inf	No	Yes	Yes	No	-6.0699145
22.07.2021	Minicell	2xYT	HlyA 1 hr before HlyBD	single	50	25	17.5	20	5	18	57.8	41.8	0.3	0	99.9	inf	No	Yes	Yes	No	-6.0699145
22.07.2021	Minicell	2xYT	HlyBD 1 hr before HlyA	single	50	25	17.5	20	5	18	61.8	37.8	0.4	0	100	inf	No	Yes	Yes	No	-5.9755115
22.07.2021	Minicell	2xYT	same time	single	50	25	17.5	25	5	17	40.5	59.1	0.4	0	100	inf	No	Yes	Yes	No	-5.9755115
22.07.2021	Minicell	LB	same time	single	50	25	17.5	25	5	17	45.1	54.7	0.2	0	100	inf	No	Yes	Yes	No	-6.1643176
22.07.2021	Minicell	LB	same time	single	50	25	17.5	25	5	17	47.8	51.9	0.3	0	100	inf	No	Yes	Yes	No	-6.0699145
22.07.2021	Minicell	LB	HlyA 1 hr before HlyBD	single	50	25	17.5	25	5	18	49.8	50	0.2	0	100	inf	No	Yes	Yes	No	-6.1643176
22.07.2021	Minicell	LB	HlyBD 1 hr before HlyA	single	50	25	17.5	25	5	18	50.5	49.1	0.4	0	100	inf	No	Yes	Yes	No	-5.9755115
23.07.2021	MG1655 Parent	2xYT	same time	single	100	30	0	37	5	1	62.2	10	6.7	21.1	100	0.31753555	No	Yes	Yes	No	-0.0281201
23.07.2021	MG1655 Parent	LB	same time	single	100	30	0	37	5	1	32.8	49.7	12.6	4.9	100	2.57142857	No	Yes	Yes	No	5.54165924
23.07.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	5	1	69.4	25	3.3	2.3	100	1.43478261	No	Yes	Yes	No	-3.2378234
26.07.2021	MG1655 Parent	LB	same time	single	100	30	0	37	5	2	21.6	71.9	5.9	0.6	100	9.83333333	No	Yes	Yes	No	-0.7833444
26.07.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	5	2	9.3	90.5	0.2	0	100	inf	No	Yes	Yes	No	-6.1643176
26.07.2021	MG1655 Parent	LB	same time	single	100	30	0	20	5	2	41.5	58.3	0.1	0	99.9	inf	No	Yes	Yes	No	-6.2587206
26.07.2021	MG1655 Delta cls	LB	same time	single	100	30	0	20	5	2	56.3	43.7	0	0	100	undefined	No	Yes	Yes	No	-6.3531237
26.07.2021	MG1655 Parent	LB	same time	single	100	30	0	20	5	2	47.4	52.6	0	0	100	undefined	No	Yes	Yes	No	-6.3531237
26.07.2021	MG1655 Delta cls	LB	same time	single	100	30	0	20	5	2	47.1	52.9	0	0	100	undefined	No	Yes	Yes	No	-6.3531237
26.07.2021	MG1655 Parent	LB	HlyA 1 hr before HlyBD	single	100	30	0	20	5	2	44.8	55.1	0.1	0	100	inf	No	Yes	Yes	No	-6.2587206
26.07.2021	MG1655 Delta cls	LB	HlyA 1 hr before HlyBD	single	100	30	0	20	5	2	41.5	58.4	0	0	99.9	undefined	No	Yes	Yes	No	-6.3531237
26.07.2021	MG1655 Parent	LB	HlyBD 1 hr before HlyA	single	100	30	0	20	5	2	98.3	1.7	0	0	100	undefined	No	Yes	Yes	No	-6.3531237
26.07.2021	MG1655 Delta cls	LB	HlyBD 1 hr before HlyA	single	100	30	0	20	5	2	98.5	1.5	0	0	100	undefined	No	Yes	Yes	No	-6.3531237
29.07.2021	MG1655 Parent	LB	same time	single	100	30	0	37	0	1	26	65.1	7.5	1.5	100.1	5	No	Yes	Yes	No	0.72710426
29.07.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	0	1	17.8	76.2	5.7	0.3	100	19	No	Yes	Yes	No	-0.9721504
29.07.2021	MG1655 Parent	LB	HlyA 30 min before HlyBD	single	100	30	0	37	0	1	8.4	89.9	1.7	0	100	inf	No	Yes	Yes	No	-4.748272
29.07.2021	MG1655 Delta cls	LB	HlyA 30 min before HlyBD	single	100	30	0	37	0	1	3.7	96	0.3	0	100	inf	No	Yes	Yes	No	-6.0699145

29.07.2021	MG1655 Parent	LB	HlyBD 30 min before HlyA	single	100	30	0	37	0	1	73.1	0.2	0.1	26.6	100	0.0037594	No	Yes	Yes	No	-6.2587206
29.07.2021	MG1655 Delta cls	LB	HlyBD 30 min before HlyA	single	100	30	0	37	0	1	81.9	0.2	0.3	17.6	100	0.01704546	No	Yes	Yes	No	-6.0699145
29.07.2021	MG1655 Parent	LB	same time	single	100	30	0	37	0	1	8.1	84.8	5.7	1.4	100	4.07142857	No	Yes	Yes	No	-0.9721504
29.07.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	0	1	9.3	84.7	5.9	0.1	100	59	No	Yes	Yes	No	-0.7833444
29.07.2021	MG1655 Parent	LB	same time	single	50	30	0	37	0	1	17.5	66.6	14	1.9	100	7.36842105	No	Yes	Yes	No	6.86330178
29.07.2021	MG1655 Delta cls	LB	same time	single	50	30	0	37	0	1	7.3	88.4	4.2	0.1	100	42	No	Yes	Yes	No	-2.388196
09.08.2021	MG1655 Parent	LB	same time	single	100	30	0	37	5	1	26	57.2	15.8	1	100	15.8	No	Yes	Yes	No	8.56255648
09.08.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	5	1	11.1	73.6	14.5	0.8	100	18.125	No	Yes	Yes	No	7.33531698
09.08.2021	MG1655 Parent	LB	same time	single	100	30	0	37	5	1	26.5	40.8	23.9	8.8	100	2.71590909	No	Yes	Yes	No	16.2092026
09.08.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	5	1	11.9	64.3	22	1.7	99.9	12.9411765	No	Yes	Yes	No	14.4155449
09.08.2021	MG1655 Parent	LB	same time	single	100	30	0	37	0	1	23.9	44.5	24	7.5	99.9	3.2	No	Yes	Yes	No	16.3036057
09.08.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	0	1	8	71.6	19.4	1	100	19.4	No	Yes	Yes	No	11.9610659
09.08.2021	MG1655 Parent	LB	HlyA 30 min before HlyBD	single	100	30	0	37	5	1	19.3	70.4	10	0.3	100	33.3333333	No	Yes	Yes	No	3.08718023
09.08.2021	MG1655 Delta cls	LB	HlyA 30 min before HlyBD	single	100	30	0	37	5	1	9.1	84	6.8	0.1	100	68	No	Yes	Yes	No	0.06628299
09.08.2021	MG1655 Parent	LB	HlyBD 30 min before HlyA	single	100	30	0	37	5	1	41.9	0.1	0.5	57.4	99.9	0.0087108	No	Yes	Yes	No	-5.8811085
09.08.2021	MG1655 Delta cls	LB	HlyBD 30 min before HlyA	single	100	30	0	37	5	1	44.4	0.1	1.6	53.9	100	0.0296846	No	Yes	Yes	No	-4.842675
11.08.2021	MG1655 Parent	LB	same time	single	100	30	0	37	0	1	4.1	56.2	38.6	1.1	100	35.0909091	No	Yes	Yes	No	30.0864494
11.08.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	0	1	5.3	76.3	18	0.3	99.9	60	No	Yes	Yes	No	10.6394233
11.08.2021	MG1655 Parent	LB	same time	single	100	30	0	37	2	1	3.6	52.4	42.7	1.3	100	32.8461539	No	Yes	Yes	No	33.9569739
11.08.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	2	1	7.8	69.4	21.7	1.1	100	19.7272727	No	Yes	Yes	No	14.1323358
11.08.2021	MG1655 Parent	LB	same time	single	100	30	0	37	5	1	3.3	42.3	52.6	1.8	100	29.2222222	No	Yes	Yes	No	43.3028748
11.08.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	5	1	5.6	72.8	21.3	0.4	100.1	53.25	No	Yes	Yes	No	13.7547236
11.08.2021	MG1655 Parent	LB	same time	single	100	30	0	37	8	1	3	49.1	46.9	1	100	46.9	No	Yes	Yes	No	37.9219016
11.08.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	8	1	5.5	68.9	25	0.5	99.9	50	No	Yes	Yes	No	17.2476361
11.08.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	10	1	6.4	67.4	25.5	0.7	100	36.4285714	No	Yes	Yes	No	17.7196513
23.08.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	2	10.8	58	25.4	5.8	100	4.37931035	No	Yes	Yes	No	17.6252482
23.08.2021	MG1655 Parent	LB	same time	single	100	30	0	37	5	0.5	14.8	76.1	7.1	2	100	3.55	No	Yes	Yes	No	0.3494921

23.08.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	5	0.5	10.3	85.9	3	0.8	100	3.75	No	Yes	Yes	No	-3.5210325
23.08.2021	MG1655 Parent	LB	same time	single	100	30	0	37	5	1	8.1	44.5	28.1	19.3	100	1.45595855	No	Yes	Yes	No	20.1741303
23.08.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	5	1	7.5	74.2	13.9	4.3	99.9	3.23255814	No	Yes	Yes	No	6.76889875
23.08.2021	MG1655 Parent	LB	same time	single	100	30	0	37	5	1.5	9.8	47.3	25.9	17	100	1.52352941	No	Yes	Yes	No	18.0972634
23.08.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	5	1.5	5.5	79.8	11.8	3	100.1	3.93333333	No	Yes	Yes	No	4.78643493
23.08.2021	MG1655 Parent	LB	same time	single	100	30	0	37	5	2	10	46.8	26.9	16.3	100	1.65030675	No	Yes	Yes	No	19.0412938
23.08.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	5	2	4.6	73.1	16.6	5.7	100	2.9122807	No	Yes	Yes	No	9.3177808
23.08.2021	MG1655 Parent	LB	same time	single	100	30	0	37	5	2.5	7.4	42.3	27.5	22.8	100	1.20614035	No	Yes	Yes	No	19.607712
23.08.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	5	2.5	5.2	78.2	13.8	2.9	100.1	4.75862069	No	Yes	Yes	No	6.67449571
17.05.2021	C43deltaAcrAB	LB	None	None	100	30	0	20	5	1.5	100	0	0	0	100	N/A	Yes	No	No	No	N/A
17.05.2021	C43deltaAcrAB	LB	None	None	100	30	0	20	5	1.5	99.8	0.1	0	0.1	100	N/A	Yes	No	Yes	No	N/A
17.05.2021	C43deltaAcrAB	LB	HlyA Only	double	100	0	0	20	5	1.5	55.5	44.5	0	0	100	N/A	Yes	No	No	No	N/A
17.05.2021	C43deltaAcrAB	LB	HlyA Only	double	100	0	0	20	5	1.5	55.1	44.8	0	0	99.9	N/A	Yes	No	Yes	No	N/A
17.05.2021	C43deltaAcrAB	2xYT	None	None	0	0	0	20	5	1.5	100	0	0	0	100	N/A	Yes	No	No	No	N/A
17.05.2021	C43deltaAcrAB	2xYT	None	None	0	0	0	20	5	1.5	100	0	0	0	100	N/A	Yes	No	Yes	No	N/A
24.05.2021	C43deltaAcrAB	LB	None	None	0	0	0	20	5	2	99.9	0	0	0	99.9	N/A	Yes	No	No	No	N/A
24.05.2021	C43deltaAcrAB	LB	None	None	0	0	0	20	5	2	99.9	0	0	0.1	100	N/A	Yes	No	Yes	No	N/A
24.05.2021	C43deltaAcrAB	LB	HlyA Only	single	100	0	0	20	5	2	81.7	18.3	0	0	100	N/A	Yes	No	No	No	N/A
24.05.2021	C43deltaAcrAB	LB	HlyA Only	single	100	0	0	20	5	2	73.6	26.1	0.2	0.1	100	N/A	Yes	No	Yes	No	N/A
24.05.2021	C43deltaAcrAB	LB	None	None	100	30	0	20	5	2	99.9	0	0	0.1	100	N/A	Yes	Yes	Yes	No	N/A
24.05.2021	C43deltaAcrAB	2xYT	None	None	100	30	0	20	5	2	99.7	0	0	0.3	100	N/A	Yes	Yes	Yes	No	N/A
09.06.2021	C43deltaAcrAB	2xYT	None	None	0	0	0	37	5	1	99.9	0.1	0	0	100	N/A	Yes	No	No	No	N/A
09.06.2021	C43deltaAcrAB	2xYT	None	None	0	0	0	37	5	1	100	0	0	0	100	N/A	Yes	No	Yes	No	N/A
09.06.2021	C43deltaAcrAB	LB	HlyA Only	single	100	0	0	37	5	1	12.8	87.2	0	0	100	N/A	Yes	No	No	No	N/A
09.06.2021	BL21(DE3)	2xYT	None	None	0	0	0	37	5	1	99.9	0.1	0	0	100	N/A	Yes	No	No	No	N/A
09.06.2021	BL21(DE3)	2xYT	None	None	0	0	0	37	5	1	100	0	0	0	100	N/A	Yes	No	Yes	No	N/A
09.06.2021	BL21(DE3)	2xYT	HlyA Only	single	100	0	0	37	5	1	11.6	88.4	0	0	100	N/A	Yes	No	No	No	N/A

18.06.2021	C43deltaAcrAB	2xYT	None	None	0	0	0	37	5	1	100	0	0	0	100	N/A	Yes	No	No	No	N/A
18.06.2021	C43deltaAcrAB	2xYT	None	None	0	0	0	37	5	1	99.9	0.1	0	0	100	N/A	Yes	No	Yes	No	N/A
18.06.2021	C43deltaAcrAB	2xYT	HlyA Only	single	100	0	0	37	5	1	16.3	83.6	0	0	99.9	N/A	Yes	No	No	No	N/A
18.06.2021	BL21(DE3)	2xYT	None	None	0	0	0	37	5	1	98.9	1	0	0	99.9	N/A	Yes	No	No	No	N/A
18.06.2021	BL21(DE3)	2xYT	None	None	0	0	0	37	5	1	99.9	0.1	0	0	100	N/A	Yes	No	Yes	No	N/A
18.06.2021	BL21(DE3)	2xYT	HlyA Only	single	100	0	0	37	5	1	10.6	89.4	0	0	100	N/A	Yes	No	No	No	N/A
21.06.2021	BL21(DE3)	2xYT	None	None	0	0	0	37	5	1.5	99.9	0.1	0	0	100	N/A	Yes	No	No	No	N/A
21.06.2021	BL21(DE3)	2xYT	None	None	0	0	0	37	5	1.5	99.9	0	0	0	99.9	N/A	Yes	No	Yes	No	N/A
21.06.2021	BL21(DE3)	2xYT	HlyA Only	single	100	0	0	37	5	1.5	6.6	93.4	0	0	100	N/A	Yes	No	No	No	N/A
21.06.2021	BL21(DE3)	2xYT	HlyA Only	single	100	0	0	37	5	1.5	5.9	94.1	0	0	100	N/A	Yes	No	Yes	No	N/A
24.06.2021	BL21(DE3)	2xYT	None	None	0	0	0	37	5	1.5	100	0	0	0	100	N/A	Yes	No	No	No	N/A
24.06.2021	BL21(DE3)	2xYT	None	None	0	0	0	37	5	1.5	99.9	0	0	0	99.9	N/A	Yes	No	Yes	No	N/A
24.06.2021	BL21(DE3)	2xYT	HlyA Only	single	100	0	0	37	5	1.5	6.5	93.5	0	0	100	N/A	Yes	No	No	No	N/A
24.06.2021	BL21(DE3)	2xYT	HlyA Only	single	100	0	0	37	5	1.5	6.8	93.2	0	0	100	N/A	Yes	No	Yes	No	N/A
28.06.2021	BL21(DE3)	2xYT	None	None	0	0	0	37	5	1	100	0	0	0	100	N/A	Yes	No	No	No	N/A
28.06.2021	BL21(DE3)	2xYT	None	None	0	0	0	37	5	1	100	0	0	0	100	N/A	Yes	No	Yes	No	N/A
28.06.2021	BL21(DE3)	2xYT	HlyA Only	single	100	0	0	37	5	1	5.5	94.5	0	0	100	N/A	Yes	No	No	No	N/A
28.06.2021	BL21(DE3)	2xYT	HlyA Only	single	100	0	0	37	5	1	3	97	0	0	100	N/A	Yes	No	Yes	No	N/A
28.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	0	2.5	11.6	87.7	0.6	0.1	100	N/A	Yes	No	Yes	No	N/A
28.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	2.5	6.3	93.5	0.1	0	99.9	N/A	Yes	No	Yes	No	N/A
30.06.2021	BL21(DE3)	2xYT	None	None	0	0	0	37	5	3.5	99.6	0.3	0	0	99.9	N/A	Yes	No	No	No	N/A
30.06.2021	BL21(DE3)	2xYT	None	None	0	0	0	37	5	3.5	99.6	0.1	0	0.3	100	N/A	Yes	No	Yes	No	N/A
30.06.2021	BL21(DE3)	2xYT	HlyA Only	single	100	0	0	37	5	3.5	7.3	92.7	0	0	100	N/A	Yes	No	No	No	N/A
30.06.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	100	30	0	37	5	2	9.9	89.7	0.3	0.1	100	N/A	Yes	No	Yes	No	N/A
13.07.2021	BL21(DE3)	2xYT	None	None	0	0	0	37	5	3.5	99.7	0.3	0	0	100	N/A	Yes	No	No	No	N/A
13.07.2021	BL21(DE3)	2xYT	None	None	0	0	0	37	5	3.5	99.8	0.2	0	0	100	N/A	Yes	No	Yes	No	N/A
13.07.2021	BL21(DE3)	2xYT	HlyA Only	single	100	0	0	37	5	3.5	16.6	83.4	0	0	100	N/A	Yes	No	No	No	N/A

13.07.2021	BL21(DE3)	2xYT	HlyBD 30 min before HlyA	single	50	30	0	37	5	2	16.5	83.5	0	0	100	N/A	Yes	No	Yes	No	N/A
19.07.2021	Minicell	2xYT	None	None	0	50	0	37	5	1.5	92.9	7	0	0	99.9	N/A	Yes	No	No	No	N/A
19.07.2021	Minicell	2xYT	None	None	0	50	0	37	5	1.5	92.9	6.7	0	0.3	99.9	N/A	Yes	No	Yes	No	N/A
19.07.2021	Minicell	2xYT	HlyA Only	single	100	0	0	37	5	1.5	0.6	99.4	0	0	100	N/A	Yes	No	No	No	N/A
19.07.2021	Minicell	2xYT	HlyA Only	single	100	0	0	37	5	1.5	0.5	99.3	0.2	0	100	N/A	Yes	No	Yes	No	N/A
19.07.2021	Minicell	2xYT	same time	single	50	25	17.5	37	5	1.5	0.5	99.5	0.1	0	100.1	N/A	Yes	No	Yes	No	N/A
22.07.2021	Minicell	2xYT	None	None	50	25	17.5	20	5	0	99.3	0.6	0	0	99.9	N/A	Yes	No	Yes	No	N/A
22.07.2021	Minicell	2xYT	HlyA Only	single	100	0	0	25	5	1	1.7	98.3	0	0	100	N/A	Yes	No	No	No	N/A
23.07.2021	MG1655 Delta cls	2xYT	None	None	0	0	0	37	5	1	99.8	0.2	0	0	100	N/A	Yes	No	No	No	N/A
23.07.2021	MG1655 Delta cls	2xYT	None	None	0	0	0	37	5	1	99.9	0.1	0	0	100	N/A	Yes	No	Yes	No	N/A
23.07.2021	MG1655 Parent	2xYT	HlyA Only	single	100	0	0	37	5	1	17.4	82.6	0	0	100	N/A	Yes	No	No	No	N/A
23.07.2021	MG1655 Parent	2xYT	HlyA Only	single	100	0	0	37	5	1	16.5	83.5	0	0	100	N/A	Yes	No	Yes	No	N/A
23.07.2021	MG1655 Delta cls	2xYT	HlyA Only	single	100	0	0	37	5	1	17.4	82.6	0	0	100	N/A	Yes	No	No	No	N/A
23.07.2021	MG1655 Delta cls	2xYT	HlyA Only	single	100	0	0	37	5	1	17.6	82.4	0	0	100	N/A	Yes	No	Yes	No	N/A
23.07.2021	MG1655 Parent	2xYT	same time	single	100	30	0	37	5	1	82.9	17	0	0	99.9	N/A	Yes	No	Yes	No	N/A
26.07.2021	MG1655 Parent	LB	HlyBD 1 hr before HlyA	single	100	30	0	20	5	2	99.3	0.7	0	0	100	N/A	Yes	No	No	No	N/A
26.07.2021	MG1655 Parent	2xYT	HlyA Only	single	100	0	0	37	5	2	15.5	84.5	0	0	100	N/A	Yes	No	No	No	N/A
26.07.2021	MG1655 Delta cls	2xYT	HlyA Only	single	100	0	0	37	5	2	30.8	69.2	0	0	100	N/A	Yes	No	No	No	N/A
26.07.2021	MG1655 Delta cls	LB	HlyA 1 hr before HlyBD	single	100	30	0	20	5	2	31.6	68.4	0	0	100	N/A	Yes	No	Yes	No	N/A
26.07.2021	MG1655 Parent	LB	same time	single	100	30	0	37	5	2	23.1	76.8	0	0	99.9	N/A	Yes	No	Yes	No	N/A
26.07.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	5	2	8.2	91.8	0	0	100	N/A	Yes	No	Yes	No	N/A
29.07.2021	MG1655 Parent	2xYT	HlyA Only	None	100	0	0	37	0	1	100	0	0	0	100	N/A	Yes	No	No	No	N/A
29.07.2021	MG1655 Delta cls	2xYT	HlyA Only	None	100	0	0	37	0	1	100	0	0	0	100	N/A	Yes	No	No	No	N/A
29.07.2021	MG1655 Parent	2xYT	HlyA Only	single	100	0	0	37	0	1	16.7	83.3	0	0	100	N/A	Yes	No	No	No	N/A
29.07.2021	MG1655 Delta cls	2xYT	HlyA Only	single	100	0	0	37	0	1	12.4	87.6	0	0	100	N/A	Yes	No	No	No	N/A
29.07.2021	MG1655 Parent	LB	same time	single	100	30	0	37	0	1	25.1	74.9	0	0	100	N/A	Yes	No	Yes	No	N/A
29.07.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	0	1	16.8	83.2	0	0	100	N/A	Yes	No	Yes	No	N/A

09.08.2021	MG1655 Parent	2xYT	None	None	0	0	0	37	5	1	99.1	0.9	0	0	100	N/A	Yes	No	No	No	N/A
09.08.2021	MG1655 Parent	2xYT	None	None	0	0	0	37	5	1	98.7	0.8	0.1	0.4	100	N/A	Yes	Yes	Yes	No	N/A
09.08.2021	MG1655 Delta cls	2xYT	None	None	0	0	0	37	5	1	99.9	0.1	0	0	100	N/A	Yes	No	No	No	N/A
09.08.2021	MG1655 Delta cls	2xYT	None	None	0	0	0	37	5	1	99.7	0.1	0	0.2	100	N/A	Yes	Yes	Yes	No	N/A
09.08.2021	MG1655 Parent	2xYT	HlyA Only	single	100	0	0	37	5	1	9.6	90.4	0	0	100	N/A	Yes	No	No	No	N/A
09.08.2021	MG1655 Delta cls	2xYT	HlyA Only	single	100	0	0	37	5	1	16.6	83.4	0	0	100	N/A	Yes	No	No	No	N/A
09.08.2021	MG1655 Parent	LB	same time	single	100	30	0	37	5	1	38.7	61.3	0	0	100	N/A	Yes	No	Yes	No	N/A
09.08.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	5	1	11.8	88.1	0	0	99.9	N/A	Yes	No	Yes	No	N/A
11.08.2021	MG1655 Parent	LB	HlyA Only	None	100	0	0	37	5	1	99.7	0.3	0	0	100	N/A	Yes	No	No	No	N/A
11.08.2021	MG1655 Delta cls	LB	HlyA Only	None	100	0	0	37	5	1	100	0	0	0	100	N/A	Yes	No	No	No	N/A
11.08.2021	MG1655 Parent	LB	HlyA Only	None	100	0	0	37	5	1	99.9	0	0	0.1	100	N/A	Yes	Yes	Yes	No	N/A
11.08.2021	MG1655 Delta cls	LB	HlyA Only	None	100	0	0	37	5	1	99.9	0.1	0	0	100	N/A	Yes	Yes	Yes	No	N/A
11.08.2021	MG1655 Parent	LB	HlyA Only	single	100	0	0	37	5	1	9	91	0	0	100	N/A	Yes	No	No	No	N/A
11.08.2021	MG1655 Delta cls	LB	HlyA Only	single	100	0	0	37	5	1	4.9	95.1	0	0	100	N/A	Yes	No	No	No	N/A
11.08.2021	MG1655 Delta cls	LB	same time	single	100	30	0	37	10	1	7.7	92.3	0	0	100	N/A	Yes	No	Yes	No	N/A
23.08.2021	MG1655 Parent	LB	same time	None	100	30	0	37	5	0	99.8	0.2	0	0	100	N/A	Yes	No	No	No	N/A
23.08.2021	MG1655 Delta cls	LB	same time	None	100	30	0	37	5	0	99.5	0.5	0	0	100	N/A	Yes	No	No	No	N/A
23.08.2021	MG1655 Parent	LB	HlyA Only	single	100	0	0	37	5	2.5	2.1	97.9	0	0	100	N/A	Yes	No	No	No	N/A
23.08.2021	MG1655 Parent	LB	HlyA Only	single	100	0	0	37	5	2.5	2.1	97.9	0	0	100	N/A	Yes	No	Yes	No	N/A
23.08.2021	MG1655 Parent	LB	same time	single	100	30	0	37	5	1.5	24.4	75.6	0	0	100	N/A	Yes	No	Yes	No	N/A