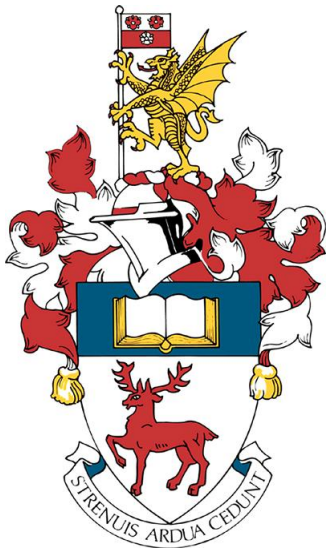
Simulation studies of the bacterial periplasm



Alister Thomas Boags

For the title of Doctor of Philosophy

FEPS, School of Chemistry, Khalid Group
University of Southampton, in collaboration with BII,
A*STAR, Singapore
United Kingdom
30/09/2020

If you're going to walk on thin ice, you might as well dance.

KARIN GILLESPIE

Abstract

Given the current rise in antibiotic resistant bacteria, *E. coli* takes center stage as a bacteria that is used commonly as a model organism. Bacteria possess a complex, crowded cellular environment that appears to be chaotic and disordered at a first glance. Their cellular structure allows them to survive and adapt to harsh conditions and varied environments.

Molecular dynamics is a widely used approach to study biomolecules in biological conditions. Using this the fine, atomic level details of the forces that hold these molecules together and as a sum drive biological function forward can be revealed. This enables the specific interactions of biological building blocks, such as proteins, lipids and other polymers, in the *E. coli* environment, to be studied. Chaperone proteins can transport small molecules in the cellular environment, where this is not limited to bacteria, such as in the case of human apolipoprotein D.

Using a marriage of experimentally sourced data as an anchor to reality and approximations based on theory, the compartment of the bacteria known as the periplasm was studied. In Chapter 3 it was observed that Braun's lipoprotein (BLP) acts as a staple that bends and tilts and can interact with outer membrane protein A (OmpA) and the cell wall. This was extended to a full periplasm, in Chapter 4, where BLP interacts with the cell wall in the presence of OmpA and TolR. In this it was shown that TolR and OmpA can bind with the cell wall simultaneously. Chapter 5 focuses on the lipid transport Mla proteins. These proteins, MlaC, MlaD and MlaA were shown to be a favourable environment for lipid binding, where the docking of the protein components is explored in tandem with modelling. The focus of Chapter 6 is chaperone behaviour. LolA was shown to bind the BLP lipid moiety and that the MAC13243 molecule inhibited interaction, but not binding. Apolipoprotein-D showed preference for arachidonic acid and cholesterol, displaying a similar theme of small hydrophobic ligand binding as LolA.

These studies have provided insight into molecular interactions that occur on a microscopic level within biological simulation.

Acknowledgments

I'd like to thank Syma Khalid and Peter Bond first and foremost for the opportunity to study and complete my PhD, a lot of support has been given to me and I'm very grateful.

A large thanks go to Firdaus Samsudin as a collaborator, who was extremely helpful and assisted me greatly during my time at Southampton.

From Southampton and Singapore I'd like to thank the following colleagues; Graham Saunders, Eilish McBurnie, Taylor Haynes, Jon Shearer, James Graham, Dale Stuchfield, Lorena Zuzic, Jan Marzinek, Alex Krah and many others.

On a personal note I'd like to thank my family who helped me get this far; Thomas Boags, Andrina Boags, Cameron Boags, Sean McGinness, Kerry McGinness and her family, Jim and Jean Boags. I love you all very much and could never have done it without you.

It's hard to leave close friends out, so thanks also go out to Steven Worswick, Michael Moor, Rob Willway and Sean Gilbody for listening to my nonsense over the years.

I'd like to thank Eva Wallis and her family, the support and kindness from them was incredible.

I don't think it's possible to write down all of the names here so I'd like to thank anyone and everyone that gave me a hand over the last four years.

Contents

List of Tables

List of Figures

Declaration of Authorship

Print name: ALISTER THOMAS BOAGS

Title of thesis: Simulations of the *E. coli* bacterial periplasm

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;

2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, that has been clearly stated;

3. Where I have consulted the publish work of others, this is always clearly attributed;

4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;

5. I have acknowledged all main sources of help;

6. Where this thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

7. Parts of this work have been published as:

- A. Boags, P.-C. Hsu, F. Samsudin, P.J. Bond, and S. Khalid. Progress in Molecular Dynamics Simulations of Gram-Negative Bacterial Cell Envelopes. *J. Phys. Chem. Lett.*, 8(11), 2017.
- Firdaus Samsudin, Alister Boags, Thomas J. Piggot, and Syma Khalid. Braun's Lipoprotein Facilitates OmpA Interaction with the *Escherichia coli* Cell Wall. *Biophys. J.*, 113(7):1496-1504,2017.

- Alister T. Boags, Firdaus Samsudin, and Syma Khalid. Binding from Both Sides: TolR and Full-Length OmpA Bind and Maintain the Local Structure of the *E. coli* Cell Wall. *Structure*, 27(4):713-724.e2, 2019.
- Alister Boags, Firdaus Samsudin, and Syma Khalid. Details of hydrophobic entanglement between small molecules and Braun's lipoprotein within the cavity of the bacterial chaperone LolA. *Sci. Rep.*, 9(1):1-8, 2019.
- Conrado Pedebos, Iain P. S. Smith, Alister Boags and Syma Khalid. The Hitchhiker's Guide to the Periplasm: Unexpected Molecular Interactions of Antibiotics Revealed by Considering Crowding Effects in *E. coli*. *BioRxiv*,2020.

Signature:

Date:

List of Abbreviations

BLP	Braun's Lipoprotein
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
MD	Molecular Dynamics
PGN	Peptidoglycan
OM	Outer Membrane
IM	Inner Membrane
QM	Quantum mechanics
QM/MM	Combined quantum mechanics / molecular mechanics
NAM	*N*-Acetylmuramic acid
NAG	*N*-Acetylglucosamine
WHAM	Weighted Histogram Analysis Method
PME	Particle mesh Ewald
LINCS	LINear Constraint Solver
ApoD	Apolipoprotein-D
OMP	Outer membrane protein
OmpA	Outer membrane protein A
LolA	Localisation of lipoprotein (protein) A
Mla	Maintenance of lipid asymmetry
PBP	Penicillin Binding Protein
UDP	Uridine Diphosphate
UTP	Uridine Triphosphate
TLR4	Toll-like receptor 4

Chapter 1

Introduction

1.1 Computational Chemistry

Computational chemistry covers a wide field of study, from simulations involving basic liquids such as water treated as single particles (or single groups of particles) to advanced quantum mechanical calculations to study reactions.[2] It is a theoretical field rooted in the physical sciences area of chemistry, where many mathematical developments have been made over the years in order to approximate chemical behaviour on computational hardware. This chemical behaviour generally revolves around the study of molecules and structures in minute detail, that is not easily observed experimentally.

Computational chemistry can be used to augment experimental studies.[3] It can also be used to model and predict outcomes that are later validated by experiments, such as modelling protein mutations to suggest which are useful for experimental biologists to implement. A large modelling interest has long existed regarding drug and material discovery, where trends between chemical structure and properties are drawn.

Generally the number of components in the simulation causes the computational requirement to increase, as does the level of accuracy that is used. High levels of accuracy include *Ab initio* methods involved with quantum mechanics applications, whereas lower levels of accuracy introduce approximations in order to speed up calculations. In the field of molecular simulation it is generally accepted that the following is true for simulations of the same size:

Equation 1

$$QM_{resource} > QM/MM_{resource} > MM_{resource}$$

where QM is quantum mechanics, QM/MM is the hybrid approach to simulation and MM is molecular (classical) mechanics. In this instance a commonly used method is molecular dynamics (MD), described under section 2.1.1.[4]

Computational chemistry gives the ability to model molecular or atomic systems and produce valuable theoretical data. This in turn leads to thought experiments, where it is possible to focus very specifically on certain aspects of the systems that are being studied. This being said, computational chemistry has certain weaknesses, mainly that the results achieved from the work are only as valid as the models that are used. Due to the approximations involved with some levels of computational chemistry, this introduces potential error due to the estimations that enable longer timescales of simulation. However, this is where theoretical and experimental work is complementary, as the real data that is available improves the models used and allows for simulations to be anchored to reality.

This being said, the parameters, potentials and algorithms in the field are constantly under development and iteratively improved. This is also true of the hardware that is involved, as improving hardware will allow for more detailed systems and faster calculations.

1.1.1 Applications to Biology

The work discussed in this thesis is focused on biological systems[5].[6] Key aspects of biological systems are the building blocks that form cellular components. The modelling of lipids has allowed for the simulation of cellular membranes. Using structural databases that are available today, proteins structures are readily available for simulation, where computational changes, such as mutations or adding lipid moieties, to the structure of the protein and the environment that it will experience can easily be implemented.[7]

It is not uncommon now to see reports of simulations containing millions of atoms to describe entire vesicles that are full of proteins, water and other molecules.[8] Computational calculations are being used to study a range of kinetic and thermodynamics properties present in biology, ranging from protein-protein interactions to improving the water model as a solvent in these simulations[9].[10]

The levels of theory applied to biological studies range from quantum to classic mechanics. Obviously due to the demanding nature of QM calculations the systems that are studied with this method tend to be much smaller and limited to much shorter timescales compared to MM calculations.

1.1.2 Motivation for biological simulations

Research into biological systems has never been more important, specifically in relation to bacteria. As the number of antibiotic strains of bacteria is on the rise, understanding the detailed physical forces involved in their mechanisms of action is crucial. The motivation behind this is that in order to take something apart, it first makes logical sense to see how it works, and from this it is hoped existing anti-bacterial drugs can be improved. This is a solely mechanical point of view, which is much more complicated in biology, particularly inside of humans.

The research is focused on gathering enough basic information about the intermolecular interactions inside of cells in order to provide insight for alternative therapeutic development. With respect to this motivation most of the work contained here is related to the *E. coli* bacterium, whilst a small amount is focused on some human biology aspects. As computational resources improve and the field expands, the amount of information that will be gathered will be enormous, which in time will ideally lead to a full description of biological systems that cannot be visualised in real time using complementary experimental techniques.[11]

1.1.3 Bacterial infections

Bacteria are amongst the oldest forms of life on our planet. The number of bacterial cells present in humans outweighs the number of human cells. A vast number of these foreign cells are vital for human health via symbiosis. This being said, a relatively small proportion of bacterial species are also pathogenic. Commonly reported cases of bacterial infection involve organisms such as *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Neisseria meningitidis* and many others.[12]

Disease caused by bacteria killed countless people before the introduction of antibiotic medicine, initiated by the discovery of penicillin. In the last one hundred years antibiotics have greatly reduced the number of deaths related to bacterial infection. However this is no longer the case. Bacterial cells are capable of reproducing every twenty minutes. This allows them to adapt to new environments through evolution at a much faster rate than the mammals that they can infect. This is the main reason why bacteria are ubiquitous in nature, being found in all environments, even in radioactive waste, thermal vents and salt lakes.[13]

Through the constant use and misuse of antibiotics, bacteria have started to adapt.[14] Resistant strains of bacterial species have begun to appear frequently, due to developing resistance and the ability to easily share this resistance through conjugation. This has rendered common antibiotics ineffective.

Infections caused by resistant bacteria can currently still be treated with some antibiotics, however it is not clear how long this will last until resistance starts to develop to these drugs also.

Never before has there been a greater need for alternative therapeutics, to which bacteria cannot develop a resistance towards. Any developed treatments should be non-toxic to humans and also eliminate the infection entirely.

Drugs can potentially enter the bacterium in a variety of ways. Passive diffusion through porins, active transport through specific proteins that recognise the drug or via self-promoted uptake, such as cationic peptides that permeabilise the bacterial membrane which results in further uptake of more peptide. If there are so many routes to enter a bacterium, where does the resistance come from? Bacteria can inactivate or reduce the effectiveness of drugs by modifying the structure of the incoming molecule.[15] Generally a protective enzyme will functionalise the molecule in a new way, reducing or eliminating ability to function as intended.[16] Bacteria will also alter the structure of the target site, where protective proteins will bind to the target, changing the shape of the binding site, preventing drug function[17].[18]

Further resistance mechanisms include changing the metabolic pathway. If a certain pathway is inhibited by a drug, bacteria can change their metabolic pathway to use a new substrate in order to continue to function.[19] Lipid A modification can also occur, where the barrier to the cell changes chemical properties, resulting in resistance. Finally bacteria are capable of reducing the amount of drugs that are present inside the cell.[20] This is done by either reducing the ability of a drug to permeate the cell initially, or by actively pumping the drugs out using an efflux pump. These pumps have evolved to respond to antibiotics in order to begin to efflux.

Bacteria can be classified into two subcategories: Gram negative or Gram positive. This is as a result of Gram staining, a test where a dye is used to stain bacteria and if the bacteria retains the violet colour of the stain it is referred to as a Gram positive species, otherwise as a Gram negative species[21].[22] These two types of bacteria have varying cellular structure which makes devising specific treatments for bacterial infection an even more complicated task.

Gram-Negative envelope structure

Gram-negative bacteria have a cellular structure that is comprised of an outer membrane, a periplasmic space containing a cell wall and an inner membrane that contains the cytoplasm.[23] The outer membrane is an asymmetric lipid bilayer, where the outer leaflet (surface exposed) is comprised of lipopolysaccharide (LPS) molecules.

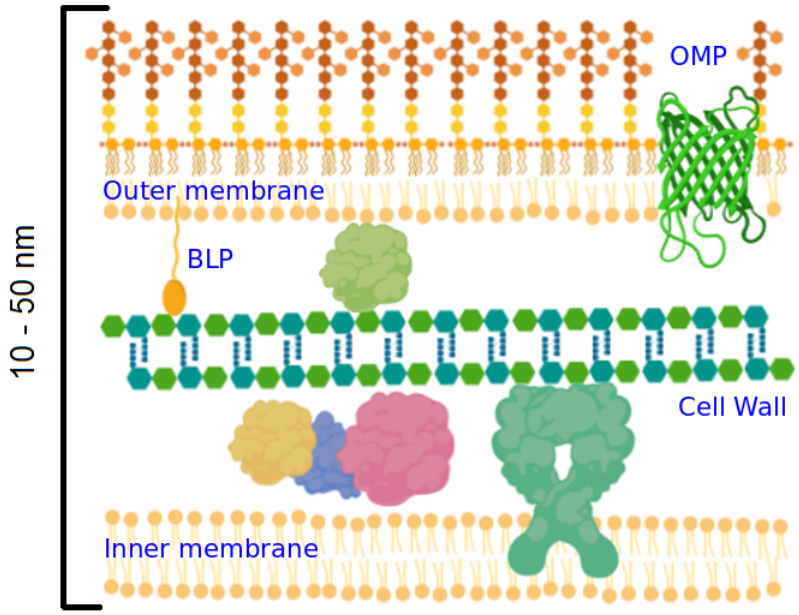


Figure 1.1: A cartoon representation of the Gram-Negative double membrane envelope periplasmic environment. The outer membrane is depicted as an asymmetric environment, with LPS in the outer leaflet and phospholipids in the inner leaflet. An Outer Membrane Porin (OMP) can be seen resting in the outer membrane. The cell wall rests in the periplasm, where the two alternating sugar units of NAM and NAG are coloured in green and turquoise. Braun's lipoportein interacts with both the cell wall and the outer membrane (orange). The inner membrane is comprised of a symmetric phospholipid bilayer, where there are proteins that interact with the cell wall in this membrane (dark green). Periplasmic chaperones and components are packed together in this environment, such as the rough drawing a three protein complex below the cell wall, seen in orange, blue and pink. Figure not to scale.

LPS is a large biomolecule, made up of lipid tails, the Lipid A hydrophobic membrane component, inner core sugars, outer core sugars and O-antigen.[24] O-antigen is a long polysaccharide that has a variable length between bacterial species, see Figure 1.2. It is recognized by antibodies as it is exposed on the cell surface. This polymer is bonded to the inner and outer core sugars which are bonded to lipid A. Whilst the immune system recognises the O-antigen, the lipid A can be responsible for the toxicity of gram-negative bacteria, where when LPS is shed from the surface of the cell into a human extra cellular environment symptoms such as fever and toxic shock can be experienced via TLR4 receptor stimulation.

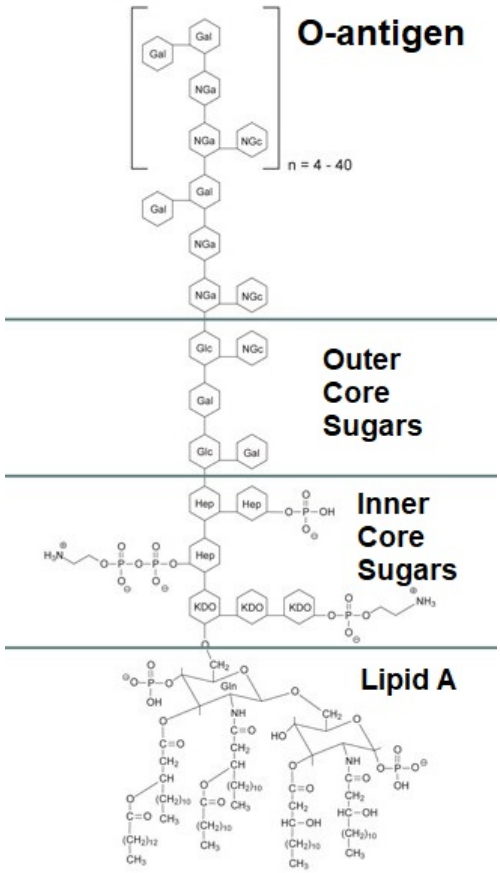


Figure 1.2: The chemical structure of LPS, where the Lipid A is shown, connected to the inner and outer core sugars, with a repeating O-Antigen presenting into the extra cellular environment, ranging between 4 and 40 units as a polymer. This is a general structure, which is observed in *E. coli*.[1]

LPS has been proposed as a main physical barrier for bacterial defense.[25] Due to the interlocking large O-antigen molecules that are less mobile than a normal phospholipid bilayer, it is harder for substances to penetrate the membrane. In addition to this the divalent calcium ions present on the surface of LPS are proposed to lock down the motion of the membrane molecules via phosphate interactions. The interaction between LPS molecules in membrane is an area that has received much interest as a result.

LPS is a heavily studied molecule[26][27].[28] This naturally makes sense as it is the first barrier to periplasmic entry. LPS is highly negatively charged and the layer is stabilised via the presence of divalent cations. Positively charged antimicrobials have been shown to associate with the LPS via simulation and experiments. LPS has been shown to be recognised by TLR4, which initiates the host response that causes the aforementioned toxic shock reaction.[29]

The inner leaflet of the outer membrane (OM) is composed of a phospholipid mixture. This mixture varies between species, including the lipids that are present and the amounts of each. These lipids generally belong to the following families: phosphatidylethanolamine, phosphatidylglycerol and cardiolipin. The two leaflets combined together form the outer membrane.[30]

Other components of the OM are membrane bound proteins, such as OmpA, OmpC or OmpF, known as the Outer Membrane Porins (OMPs). A large number of various OMPs and other proteins exist in the OM, for example OmpA estimated to number 100,000 copies per cell. Proteins localised in the OM generally have a beta barrel structure that rests in the hydrophobic region[31][32].[33] All of these proteins serve various purposes, such as porins, where all typical OMPs are beta barrel proteins that can allow diffusion of molecules in bacteria. The large number of proteins present in the OM make it a variable environment at the microscopic level, with many potential interactions to study.

The OM is wrapped around the periplasm.[30] The periplasm of a gram negative bacterium is the space between the outer and the inner membranes. This space is generally packed with chaperones, substrates and many other molecules. This packed environment also contains a biopolymer known as peptidoglycan[34].[35] Peptidoglycan makes up the bacterial cell wall. The purpose of the cell wall is to provide the bacteria with structure and support. This prevents lysis due to the high internal pressure of the dense periplasm. This is caused as the cytoplasm is a hypotonic region, where water freely flows into the volume, causing it to swell. A large component of the periplasm is Braun's lipoprotein, which is the most abundant source of protein in gram-negative bacteria.[36] This helical protein trimer is covalently bound to the cell wall at the C-terminus via a peptide bond of a terminal lysine and the N-terminus is lipidated, with a cysteine based palmitoyl moiety, where this lipid group rests in the OM, see Figures 1.4 and 1.5. This molecule, and others, essentially connect the cell wall to the OM.[37]

The chemical structure of the cell wall is comprised of *N*-Acetylmuramic acid (NAM) and *N*-Acetylglucosamine (NAG) monomers, see Figure 1.3.[38] These repeating units are glycosidically bonded into strands and the strands are cross linked via three-four peptide linkages between the tetrapeptide side chains on the NAM molecules. This tetrapeptide widely varies between bacterial species, where the chemical structure is flexible, however here the focus is for *E. coli*. This cross linkage is estimated to be between fifty percent of NAM monomers.[39] This bonding allows the cell wall to form a sheet that encloses the inner membrane. The cell wall is an ideal target for therapeutics as it has no human analogue. This provides the potential for a non-cytotoxic drug to be created.[40]

Drugs that interact with the cell wall already exist, e.g. penicillin specifically disrupts the production of the cell wall. This occurs via the inhibition of the penicillin binding proteins (PBPs)[41].[42] These proteins are responsible for the transglycosylation and transpeptidation of the cell wall. Penicillin and many other beta-lactam antibiotics bind to the region responsible for transpeptidation, causing cross linking to fail. This results in a bactericidal effect. However, resistance has developed to beta-lactams, where bacteria produce beta-lactamase, inactivating the drug. As such, now when these drugs are given, beta-lactamse inhibitors may also be administered to allow the desired effect. These enzymes are ancient components of bacteria, where it is assumed their ancestors shared sequence homology with PBPs.

The synthetic route for the cell wall consists first of monomer synthesis in the cytoplasm of the bacteria. It is then proposed that inner membrane proteins such as BacA and MurJ use bactoprenol as a membrane carrier to move the monomer to the periplasm, where the PBPs take action to form the cell wall mesh[43].[44] The cytoplasmic synthesis begins when a glutamine donates an amino group to fructose 6-phosphate, yielding glucosamine-6-phosphate. This molecule is then functionalised with acetyl Coenzyme A (CoA) to give *N*-acetyl-glucosamine-6-phosphate. *N*-acetyl-glucosamine-6-phosphate is then converted to *N*-acetyl-glucosamine-1-phosphate which interacts with UTP, to give UDP-n-actyl-glucosamine (UDP-GlcNAc), the precursor to NAG. A portion of this UDP-GlcNAc is converted to UDP-MurNAc via addition of a lactyl group added to the glucosamine. This is the precursor to the NAM molecule.[45] The precursors are then transported through the membrane for transglycoslyation to form a peptidoglycan monomer.[46]

The cell wall has been characterised using biophysical techniques, where the thickness and elasticity of the cell wall have been assessed using atomic force microscopy. The cell wall is difficult to study, as techniques that can be used to study the periplasm often destroy the periplasmic environment.[47] Decades of study have resulted in two theoretical models of the cell wall, a "layered" model and a "scaffold" model.[35] The layered model is the textbook standard, where a sheet of cell wall that is typically a monolayer surrounds the bacterial

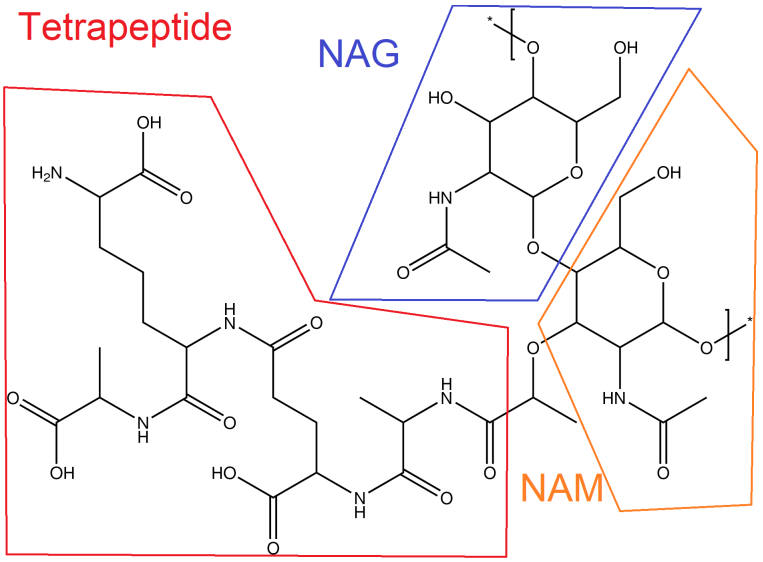


Figure 1.3: The chemical structure of the peptidoglycan monomeric unit, which is comprised of alternating NAG and NAM sugars, where a tetra-peptide is connected to the NAM sugar. This tetra-peptide is comprised of L-alanine, D-glutamic acid, *meso*-diaminopimelic acid and D-alanine. Polymerisation of the NAM and NAG units is shown via square brackets.

inner membrane.[48] The scaffold model contradicts this, where the cell wall runs perpendicular to the membranes of the periplasm. There is evidence that both models exist, however many studies show the layered model to exist and for the purposes of studies in this thesis this model has been selected, as it is the generally accepted theory. It is worth noting that the cell wall is not a static structure, and is subject to damage and requires repair. There is a specific damage response that is involved in this repair, where PBP1b is shown to be active in a repair role, indicating that PBPs are not essentially only for initial synthesis, but also for maintenance and reproduction[49].[50]

The *E. coli* periplasm is a complicated environment. Considering the packed nature and the sheer number of molecules that are present within the space, how does it function? A number of individual processes are occurring simultaneously in this space, all designed to maintain the homeostasis of the bacterial cell.

E. coli contains $\sim$ 4200 unique proteins according to sequencing, where $\sim$ 1000 of these are membrane proteins.[51] This leaves well over 3000 proteins that are located in either the periplasm, cytoplasm or the extracellular environment. Important proteins in the periplasm are chaperones. These proteins are

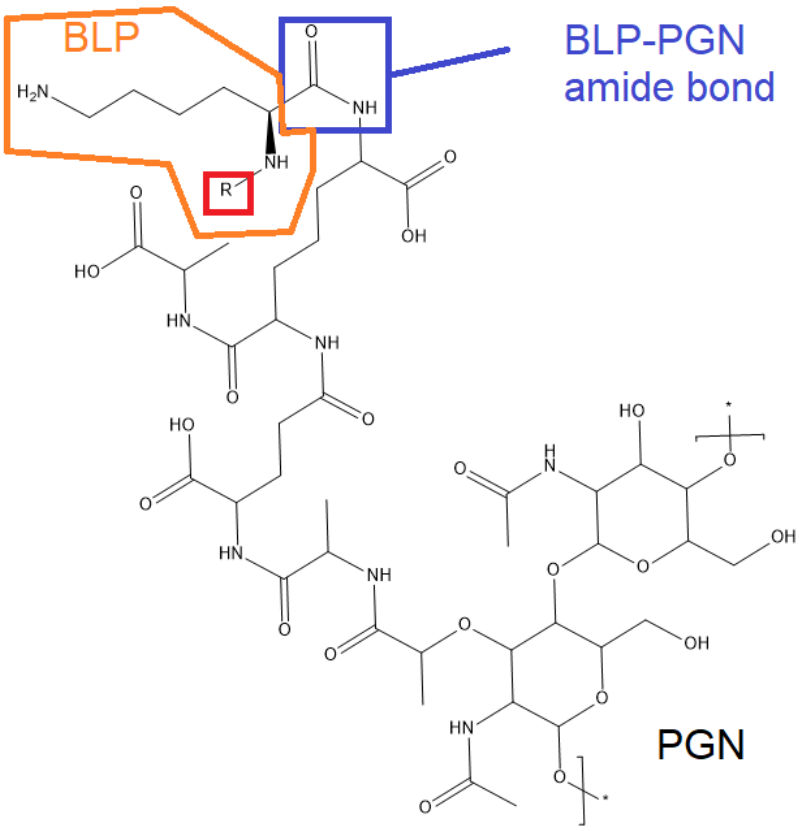


Figure 1.4: The covalent linkage between the C-terminal lysine of Braun's lipoprotein, where the carboxyl group is peptide bonded to the meso-diaminopimelic acid of the peptidogylcan, labelled and indicated via a blue box. The rest of BLP is indicated by the R symbol in a red square for the sake of brevity.

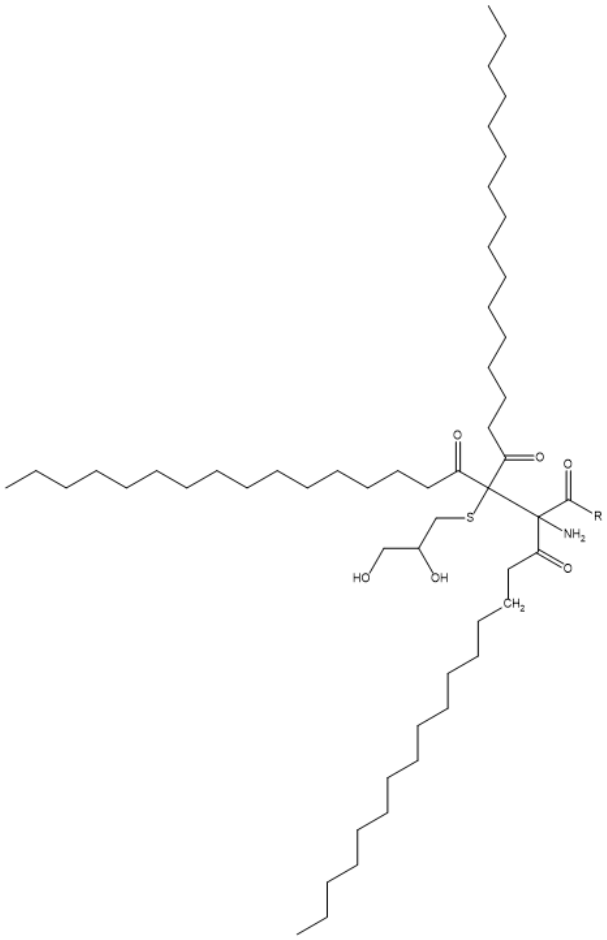


Figure 1.5: Then N terminus of the BLP is lipidated with a cysteine derived pamitoyl lipid. The lipidated cysteine rests in the OM of *E. coli*, non covalently interacting with the membrane. The BLP is indicated by the R symbol.

responsible for shuttling important small molecules between the two membrane compartments. Examples of this function are the Lol proteins, where the chaperone LolA carries lipoprotein between the IM and OM[5223].[53] In the opposite direction it is believed that the MlaC protein removes mislocalised phospholipid from the outer leaflet of the OM to the IM, in order to maintain a high LPS concentration.[54] The cumulative effect of these behaviours through many proteins allows *E. coli* to function normally.

The inner membrane is a symmetric phospholipid bilayer; each leaflet is proposed to have a similar composition to the inner leaflet of the OM[26].[1] The inner membrane surrounds the cytoplasm of the bacteria.[55] The cytoplasm is the area where key life processes such as growth, metabolism and replication take place. This area is a main target for drug delivery.[56] Disruption to key life processes would prevent the bacteria from surviving.

Inner membrane proteins also play a vital role in bacterial architecture. Examples of these proteins are MlaFEDB, TolR and LptF/LptG. These proteins are responsible for many roles, such as substrate transport, structural support and membrane integrity[5758].[59]

1.1.4 Simulations of *E.coli* membranes

Simulations of *E. coli* are varied and numerous[6061].[62] Initially model membranes that were used contained only one type of lipid. It has been shown that multi-component membranes more accurately model the behaviour of proteins with the membranes, such as cardiolipin binding specificity.[63] Electroporation of membranes has additionally been simulated, where it has been shown that the OM of *E. coli* is more resistant to pore formation than the membrane of gram-positive *S. aureus*.[64] This is attributed to the lower transmembrane potential of LPS molecules when compared to symmetric phospholipid bilayers, it is stated that this effect starts at approximately 0.4 V/nm for LPS membranes and 0.3 V/nm for symmetric bilayers. Simulations have advanced to include not just membranes, but also the space between them, where the use of coarse graining has allowed for larger volumes to be explored, such as full length periplasmic transporters[26].[65]

Membrane lipid composition in simulations varies greatly between studies, as do other conditions that are chosen for these simulations. The composition of lipids chosen in the simulation is key, as is changing the ratios present, in order to reproduce biologically relevant systems. Multicomponent *E. coli* inner membranes generally consist of three types of lipid; phosphitadylethanolamines (PE), phosphitadylglycerols (PG) and cardiolipins (CL). It is generally accepted that PE is the dominant lipid (approximately 75 percent) whilst PG and CL make up the rest of the membrane. In simulations a commonly used ratio is 90:5:5 by number of molecules, which has been implemented in many OM simulations for the inner leaflet.[66] Due to the asymmetry of the membrane, the

outer leaflet is generally composed entirely of LPS molecules[26].[67]

Membrane proteins are an integral component of bacterial membranes.[68] In both types of bacteria there are a vast number of membrane proteins that carry out various tasks for vital function. Approximately 25 percent of all proteins in *E. coli* are membrane proteins, leading to heavy study[69][70][71].[72] It is approximated that around 70 percent of all drugs targets are based on membrane proteins.

The action of anti-microbial peptides (AMPs) are a widely studied phenomenon in simulation.[73] This is due to the fact that it is known that some short peptides can harm bacteria, whilst leaving the host unharmed. These short peptides are thought to insert into cellular membranes and cause disruption that leads to the death of the cell. This is also a common mode of action for other medicines, such as daptomycin, a naturally occurring AMP, and other drugs[74].[34]

Peptidoglycan has been simulated to a small extent but this is a relatively new field of study, in contrast to the vast amount of experimental data that has been gathered relating to this biopolymer.[75] Generally, studies of the peptidoglycan before the work in this thesis involved the mechanical effects of stress on the cell wall.[76]

An overview of the published simulations of *E. coli* can be found under Appendix B titled ***Progress in Molecular Dynamics Simulations of Gram-Negative Bacterial Cell Envelopes***.

1.1.5 Aims

The aims of this thesis are as follows:

- To explain the methods involved in general simulation, where each chapter contains specific details for the methods used.
- To explore the behaviour of Braun's lipoprotein (BLP) in the periplasmic environment - Chapter 3.
- To build a simple model of the *E. coli* periplasm, including TolR, OmpA and BLP - Chapter 4.
- To explore the lipid interaction with Mla proteins, specifically MlaD, MlaC and MlaA - Chapter 5.
- Quantification of the binding between the Mla proteins that are proposed to transfer lipid - Chapter 5.
- To assess the chaperone behaviour of the LolA protein as it transports BLP, including possible inhibition - Chapter 6.

- To assess ligand binding to the ApoD monomer - Chapter 6.

Chapter 2

Methods

2.1 Simulation Methods

Improving hardware has allowed the simulation field to expand rapidly, with respect to detail and timescales. In an ideal world, the most accurate method of simulation would be used in every instance, however this is not the case. In molecular simulation, varying levels of detail can be used to describe a system. Generally, the more detailed a system is, the more computationally expensive it is to study.[77]

Simulations are used in many fields. There are a range of applications and software suites that are available in order to study systems on a microscopic scale to the planetary scale. The primary goal of simulation methods is the ability to model and predict behaviour.[78] This predictive ability allows the exploration of many possibilities, providing information to augment experimental studies. In comparison to traditional experiments, simulations can examine properties that impossible to view experimentally e.g. "alchemical" free energy calculations. This is applied in the field of chemistry, once again in a wide range, from studies involving individual atoms to complex intermolecular systems.

In computational chemistry the level of description of techniques varies. Due to the inclusion of electrons, quantum mechanics is highly detailed and allows study of the process of bond formation and dissociation between particles, however this causes the computational requirements to scale heavily with the number of particles. This is due to the complexity of calculating electronic structure during a simulation[79].[80] It also produces results on an extremely short time scale. Molecular mechanics, or classical mechanics, treats particles as spheres, that can carry a charge, that are connected together by bonded potentials that cause spring like behaviour.[81] Using this method is a simplification but allows for longer timescales to be reached.

For biological studies, where solvent is to be included, systems can easily contain hundreds of thousands, if not millions of particles.[82] This eliminates hardcore QM as an option for the study of full protein-protein interactions, drugs, DNA or large membrane systems. It is important however that any simulations capture the behaviour on a biologically relatable time scale.

2.1.1 Molecular Dynamics

Molecular dynamics is a method in computational chemistry that is heavily used in the field of biological simulation. It was introduced to the computational field in the 1950s, with biological applications beginning in 1977.[83] The basis of MD is the use of Newtonian physics, where the system that is studied is an N-body system, where the atoms in the system are treated as spheres that interact with each other through the basis of either potentials that are included in the system, known as the force field.[84] This force field will generally include all of the parameters and potentials to simulate realistic bonding which also includes parameters for bonds, angles and dihedrals present in the system.[85]In addition to this there are non-bonded interactions between the spheres that are also included in the force field to approximate Lennard-Jones and electrostatic interactions.

Historically MD was first conceptualized to study the interactions of spheres by Alder and Wainwright. These initial studies concerned the behaviour of simple liquids, yielding important results.[86] The technique was then further expanded, with the first simulation of liquid argon that used realistic potentials to describe the system, in order to evaluate diffusion coefficients and viscosity.[87] This then allowed for the very first MD simulation of a system to mimic the behaviour of real water in 1974.[88] Following this, simulations of real proteins in solution began to appear only three years later.[83]

In modern times, largely due to the advances in computational hardware, it is commonly seen in academic study for systems to contain a multitude of proteins, or a collection of various molecules such as lipids to form a membrane[89].[90] Additionally, improved algorithms and parallel computing have contributed to this. As systems have iteratively increased in size it has become possible to study increasingly realistic biological systems, such as spaces enclosed by membranes, lipid protein interactions along with a potentially limitless other number of macromolecular complexes.[91]

Advances in MD have moved the field beyond the study of simple hard spheres, and the number of techniques available for use has expanded to include hybrid methods with quantum mechanics, enhanced sampling methods and many other specific use case methodologies for the study of molecular systems[92].[80]

MD relies heavily on experimental techniques to acquire structures for simulation. The experimental data is key due as a realistic anchor is needed to validate analysis of these calculations. X-ray crystallography, solution nuclear magnetic resonance (NMR), cryogenic electron microscopy and other techniques work hand in hand with MD to provide this data.[93] Using computational techniques it is possible to improve on this data via modelling using further non-structural experimental data as a guide.[32]

MD can be performed using many packages such as GROMACS, AMBER and others[9495].[96] Common biomolecular forcefields that are used include iterations of CHARMM, GROMOS and AMBER.

2.1.2 Concepts behind MD

Newton's Equations

As MD is used for the work contained in this thesis the following overview of the mathematical principles shows how MD functions in order to simulate biological systems.

Equation 2:

$$\vec{F} = m\vec{a}$$

Where F is force, m is mass and a is acceleration. The force can also be represented as:

Equation 3:

$$\vec{F} = -\nabla U$$

Where the force, F, is related to the negative of the gradient of potential energy, U. The core principle of MD lies behind Newton's second law of motion (Equation 1). The force acting on a particle can be related to energy via the change in position.

Equation 4:

$$\vec{F} = m\frac{d\vec{v}}{dt}$$

Equation 5:

$$\vec{F} = m\frac{d^2\vec{r}}{dt^2}$$

Acceleration can be represented as the change in velocity over time (Equation 3), where v is the velocity and t is time. This can be further broken down via the second derivative of position over with respect to time, where r is the position.

Combining all of this together results in the following:

Equation 6:

$$m\frac{d^2\vec{r}}{dt^2} = -\nabla U = -\frac{dU}{dt}$$

This essentially shows how Newton's second law allows the change in potential energy to be related to the changes in particle positions during MD

simulation.[97]

As it has been shown that the energy is related to the change in position with respect to time, it is important to clarify how this progression of position, over time during a simulation functions. This is performed gradually as a "timestep", the duration of time in a simulation between a change in position of the particles in the system.

Integrators

Propagation of the simulation over the timestep is applied through the use of integration algorithms to calculate positions, velocities and accelerations, under the assumption that a Taylor series expansion of the following form may be used:[98]

Equation 7:

$$\vec{r}(t+dt) = \vec{r}(t) + \vec{v}(t)dt + \frac{1}{2}\vec{a}(t)dt^2 + ...$$

Equation 8:

$$\vec{v}(t+dt) = \vec{v}(t) + \vec{a}(t)dt + \frac{1}{2}\vec{b}(t)dt^2 + ...$$

Equation 9:

$$\vec{a}(t+dt) = \vec{a}(t) + \vec{b}(t)dt + ...$$

where r is the position, v is velocity, a is acceleration and b is the jerk, the third derivative of position.

This then gives rise to the Verlet algorithm, which can be written as follows, where only up to the second term of the Taylor series is considered, as the third term cancels out, as does the first shown in Equation 11:

Equation 10:

$$\vec{r}(t+dt) = \vec{r}(t) + \vec{v}(t)dt + \frac{1}{2}\vec{a}(t)dt^2$$

Equation 11:

$$\vec{r}(t-dt) = \vec{r}(t) - \vec{v}(t)dt + \frac{1}{2}\vec{a}(t)dt^2$$

The sum of the two (Equations 9 and 10) gives the following form:

Equation 12:

$$\vec{r}(t+dt) = 2\vec{r}(t) - \vec{r}(t-dt) + \vec{a}(t)dt^2$$

From this it can be seen that the Verlet algorithm makes use of positions and accelerations at time *(t)* and the positions from a previous time *(t-dt)* to calculate the new positions at the time *(t+dt)*, where *dt* represents one time step.[99] The benefits of using this form of the Verlet algorithm is that the storage requirements are low, and it is a straightforward equation. The disadvantage is the modest accuracy it provides, considering that velocities are not explicitly calculated. Additionally, the fact that it uses a previous position in time causes issues with the first iteration, since where at $t = 0$, there is no available value for t = -1; however as the acceleration is known during the first step, it is possible to approximate the position, as it has already been shown that position and acceleration relative to time are related.

This gives rise to the Velocity Verlet algorithm, where the positions, accelerations and velocities are all calculated at time *(t)*, solving the issue of the first steps, as follows:

Equation 13:

$$(\vec{r}+dt) = \vec{r}(t) + \vec{v}(t)dt + \frac{1}{2}\vec{a}(t)dt^2$$

Equation 14:

$$(\vec{v}+dt) = \vec{v}(t) + \frac{1}{2}[\vec{a}(t) + \vec{a}(t+dt)]dt$$

Another algorithm commonly used is the leapfrog algorithm; this is another related form of integrator where the velocities are explicitly calculated, but they are not calculated at the same time as the positions.[99] The leapfrog procedure is shown in the following:

Equation 15:

$$\vec{r}(t+dt) = \vec{r}(t) + \vec{v}(t+\frac{1}{2}dt)dt$$

Equation 16:

$$\vec{v}(t + \frac{1}{2}dt) = \vec{v}(t - \frac{1}{2}dt) + \vec{a}(t)dt$$

In this form, the velocities leap over the positions, then the positions leap over the velocities. This form of integrator however means that to get the value of the velocity at a specific timestep, the following approximation must be used:

Equation 17:

$$\vec{v}(t) = \frac{1}{2}[\vec{v}(t - \frac{1}{2}dt) + \vec{v}(t + \frac{1}{2}dt)]$$

where to approximate the value of velocity the average between the two points is taken.

Now that the concept of integrators has been established, the timestep *(dt)* must be evaluated. In MD if the timestep is too low, it will not be possible to achieve meaningful data with current computational resource, but if it is too high then the important motions of bonds in molecules can be missed or instabilities will arise leading to system "crashing". In biological systems this is complicated by the highest frequency motions in the system, such as a chemical bond, which is shown below in the following form as two particles connected by a spring:

Equation 18:

$$\tilde{\nu} = \frac{1}{2\pi c}\sqrt{\frac{K}{\mu}}$$

where μ is the reduced mass of the particles in the bond, $\tilde{\nu}$ is the wavenumber that equates to the vibrational frequency of the bond, K is the force constant of the bond and c is the speed of light. A higher wavenumber indicates a faster bond vibration, if it is assumed that K is constant, lighter atoms connected together will vibrate faster. Given that hydrogen is the lightest atom, this indicates that bonds involving hydrogen will have the highest wavenumbers, vibrating the fastest, which in turn will limit the timestep most.

To select the timestep, it must be less time per step than the fastest vibration in the system which have been established as the vibrations involved with hydrogen atoms. We generally enforce a timestep of 1 fs, unless constraints are imposed upon a system, in which case the timestep can be set to 2 fs.

There are several constraints algorithms that can be used for simulation; LINCS, SETTLE, SHAKE being amongst them[100].[101] A constraint algorithm

is a method that satisfies Newtonian motion of a rigid body, involving mass points. SHAKE is widely used for large molecules, whilst SETTLE is generally used for smaller molecules such as water, to generate rigid water molecules. LINCS is commonly applied to simulations when using the GROMACS software package. LINCS functions by resetting bonds to correct lengths after an unconstrained update of the positions. This is particularly useful for covalent bonds involving hydrogens, removing the previously mentioned high frequency issue. LINCS is more stable and faster than the other algorithms, but does not work with angle constraints, however SHAKE does.

Generally the first sign that a simulation is going to fail, known as "blowing up", is when the constraints fail due to the forces that are applied to a molecule.[102]

2.1.3 Forcefields

Having discussed how an MD simulation progresses through time, the details of what holds particles and molecules together are equally as important[85].[103] The forcefield is made up of parameters and potentials derived from experiments, QM or a combination of both.

This is generally described as a forcefield, where the potential energy of the system is as follows:

Equation 19:

$$U = \Sigma_{non\text{-}bonded} + \Sigma_{bonded}$$

These two terms provide the interactions for non bonded particles in the system and for particles that are connected together to form molecules. This can further be broken down into:

Equation 20:

$$\Sigma\, U_{bonded} = \Sigma\, U_{bonds} + \Sigma\, U_{angles} + \Sigma\, U_{dihedrals}$$

MD forcefields map out the bonding of the particles together before the simulation is started. This means that the particles will always be bonded for the duration of the simulation. Unlike in QM methods, in MD bonds are not broken and are not formed.

Non-bonded interactions such as electrostatics and Lennard-Jones interactions contribute to the overall energy as follows:

Equation 21:

$$\Sigma\, U_{non\text{-}bonded} = \Sigma\, U_{Coulombic} + \Sigma\, U_{Van\ der\ Waals}$$

Coulombic Potentials

Coulombic potentials, also known as electrostatic forces, in MD are dealt with via the application of Coulombs Law:

Equation 22:

$$|\mathbf{F}| = k_e \frac{|q_1 q_2|}{r^2}$$

This law states that the magnitude of the electrostatic force F, either attractive or repulsive, between two point charges, q_1 and q_2, is directly proportional to the product of the magnitudes of the charges and is inversely proportional to the square of the distance, r^2, between the point charges, where k_e is Coloumb's constant.[104] This law allows for a description of the interaction of charges in MD simulation.

Equation 23:

$$U = k_e \frac{q_1 q_2}{r}$$

This can also be expressed in terms of potential energy, where U is the potential energy, and the two point charges, q_1 and q_2, Coloumb's constant, k_e, and the distance, r, between the particles are defined similarly to Equation 22.

Cut-offs

Cut-offs are used to treat the non-bonded interactions, there are multiple methods available for cutoff schemes. The simplest of these is to simply cut off the interaction at a certain distance, referred to as "straight cut-off". Interactions that are beyond this distance are ignored altogether.

Reaction field is also used, for charged interactions only, where each molecule is encompassed by a sphere, where interactions inside the sphere are treated using Coulomb's law, outside of the sphere it is assumed there is a medium that has a standard dielectric constant, where the molecule can induce polarisation in this medium. If an interaction is in the boundary region the charge can be tapered to be equal to zero.

These cut-off schemes are relatively simple and are not very accurate for electrostatics, however they are relatively quick to use to compute the electrostatic potentials. A more complex method is Ewald, not necessarily a true cut-off, where the interaction can be split into two parts, short range and long range, allowing them to converge more quickly.[105]

PME

PME is the Particle Mesh Ewald method.[106] This method is used to treat the charges, long range interactions, in a simulation. This is done by separating the interaction potential into a short range and a long range term:

Equation 24:

$$\phi(r) = \phi_{sr}(r) + \phi_{lr}(r)$$

Where r is distance, ϕ is the total potential, ϕ_{sr} is the short range potential and ϕ_{lr} is the long range potential. The the direct sum of energy between two particles can be shown as:

Equation 25:

$$E_{TOT} = \sum_{ij} \phi(r_j - r_i) = E_{sr} + E_{lr}$$

This is replaced with two sums, one for the sort range and one for the long range:

Equation 26:

$$E_{sr} = \sum_{ij} \phi_{sr}(r_j - r_i)$$

This is a direct sum of the short ranged potential E_{sr} in real space, representing the particle section of PME.

Equation 27:

$$E_{lr} = \sum_{k} \Phi_{lr}(k)|p(k)|^2$$

where Φ_{lr} and *p(k)* represent the Fourier transforms of the potential and charge density, representing the Ewald component. This is done as the short range in real space and long range in Fourier space both converge quickly, as in Equation 22. This allows them to be truncated without much accuracy loss and brings down computational requirements. However to evaluate the Fourier transform of *p(k)* of a charge density field efficiently, a Fast Fourier transform is used, where the density field is computed on a discrete lattice in space, bringing about the mesh component of PME. An important note is that PME should be done using a neutral system and therefore requires a neutralised simulation box. This is rooted in the fact that the Ewald method will only formally converge if the net charge is zero, otherwise due to the use of periodicity in tandem with

PME the sum of the electrostatics will tend towards infinity.

Lennard-Jones Potential

Non-charged interactions are treated using the Lennard-Jones Potential.[107] This model allows for the attractive association of all particles in a system based on Van der Waals interactions. It is the basis of how particles are treated as non-overlapping particles in MD, it has the following mathematical form:

Equation 28:

$$V = 4\epsilon[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^{6}]$$

where V is the potential energy, ϵ is the depth of the potential well, σ is the distance at which the inter-particle potential is equal to zero, r is the distance between the particles. In this model the epsilon term can be referred to as the 'stickiness potential' where the depth of the well indicates the stabilisation between two particles. However as this model treats the particles as non-overlapping spheres, if the distance is reduced to a value less than σ the potential energy in the model quickly tends towards infinity, as hard spheres cannot overlap. Collectively, this is a result of the two main terms in the equation, the r^{-12} term is the short range repulsive term as an approximation of Pauli repulsion, whilst the r^{-6} is the attractive term describing Van der Waals forces. There exists a value of r, r_{eq} which is the equilibrium distance between the two particles, which is shown by the lowest point in the potential energy. If the distances increases past the equilibrium the stabilisation energy rapidly falls off altogether, as the distance increases beyond the cut-off value.

This can also be shown graphically as in Figure 2.1:

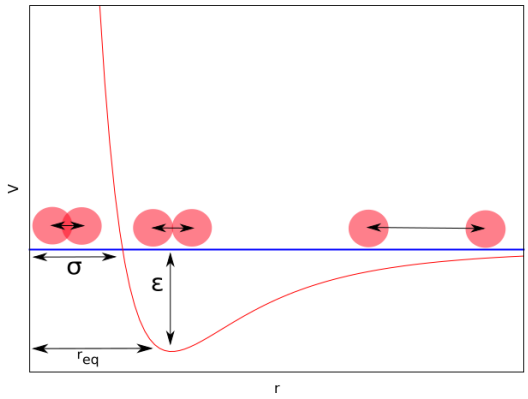


Figure 2.1: A visualisation of the Lennard Jones potential occurring between two interacting particles.

Bonds and Angles

The next idea is to discuss how these particles are connected together in order to form molecules, from simple molecules such as water, to the complex secondary structure of proteins. This is done via the approximation to spring like behaviour that bonds and angles can be treated using harmonic potentials, such that all particles are held together with bonding and angle terms which are described via equations of the following general form:

Equation 29:

$$V = \frac{1}{2}k(x - x_{eq})^2$$

Where V is the potential energy, k is the force constant, x is the angle/bond and x_{eq} is the equilibrium angle/bond value.

This can also be visualised in the following plot, Figure 2.2:

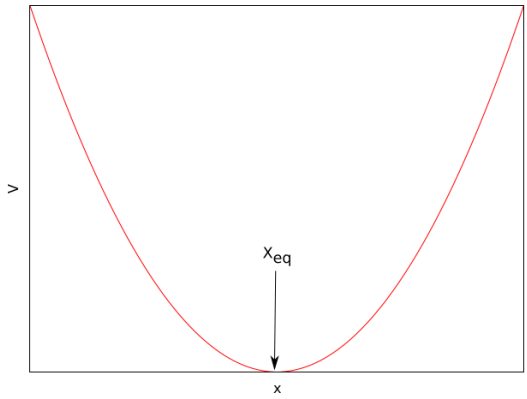


Figure 2.2: A harmonic potential that displays the equilibrium value, used for bonds and angles in simulation.

Where the potential energy is lowest when the angle or bond is at the equilibrium value, and when the value is deviated from there is an energy penalty applied that forces the value to remain at x_{eq}. How heavily this penalty is applied depends on the value of k where higher force constants give the harmonic potential a steeper gradient.

Dihedrals

Dihedrals are treated differently to bonds and angles, as there is more than one value that a dihedral can take, due to the capacity for angles between two planes formed of three atoms to rotate. This takes the form of a cosine potential,as follows:

Equation 30:

$$V = k(1 + cos(n\Theta - \Theta_{eq}))$$

where V is the potential energy, k is the force constant, n is the number of minima, Θ is the angle of the dihedral and Θ_{eq} is the phase angle of the dihedral. This adequately describes the potential energy of the plane of three atoms rotating with respect to another, where the dihedral is composed of four atoms. It can also be visualised in the following figure:[108]

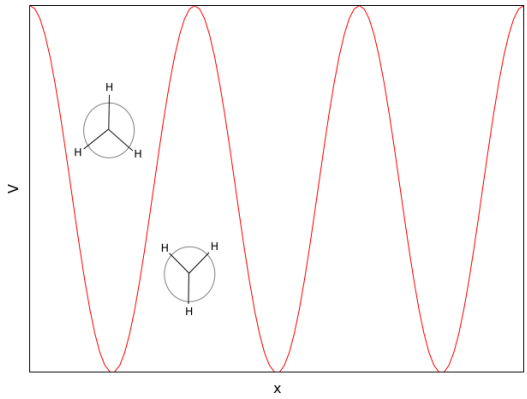


Figure 2.3: A periodic potential used to approximate dihedral potentials in simulation.

Figure 2.3 shows an example of a methyl group rotating in space, where there are three equal minima, which the dihedral can proceed through. This shows the different treatment for dihedral potentials in forcefields in comparison to a single minimum value defined by a harmonic potential.

2.1.4 Energy Minimization

Considering how a forcefield holds together particles using bonded and non-bonded forces, it can be seen that at times the energy associated with a system will be larger than the minimum. The minimum can be a local minimum represented by a stable structure, or the global minimum, which is the lowest energy structure possible. Based on this most systems essentially are never at a global minimum. This can be attributed to overlapping atoms, bonds that are too long or short and other attributes in the system. A good example of this is a protein that has been solvated computationally in order to mimic an aqueous environment. It is highly unlikely that this protein is in a minimum energy conformation after being placed in randomly generated solvent, due to unrealistically large forces or "clashes" being present. This brings about the need for equilibration in systems to achieve natural behaviour, however equilibration of highly unstable systems can still fail if large unfavourable forces are present. Therefore when a system is first created, energy minimisation is needed to remove bad clashes and bring the particles to a lower energy configuration to allow for the equilibration steps to take place.

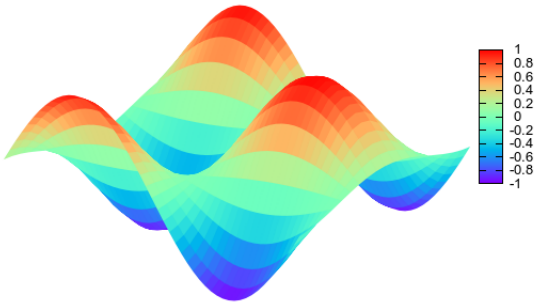


Figure 2.4: An example of a potential energy surface, where red indicates peaks in the landscape and blue indicates troughs.

The potential energy as a function of the particle positions in the system can be visualised as a hypersurface (potential energy surface), such as the one above, where peaks represent maximum energy and valleys represent minimum energy, as in Figure 2.4. There can be many maxima and minima on a hypersurface, but fewer global maxima and global minima. The purpose of energy minimization is to drive the hypersurface towards the minima. However protein hypersurfaces are very rugged, with thousands or millions of potential minima, where minimisation is generally used to reach a configuration that is stable enough to begin simulation.

Steepest Descents (SD)

The first of two commonly used energy minimization algorithms is SD minimisation[109].[110] It is a first order iterative optimization algorithm that searches for the minimum value of the energy. This works by searching in the direction where the gradient is negative or "downhill" on the potential energy surface. The mathematics behind this are as follows:

Equation 31:

$$x_{n+1} = x_n - \gamma \nabla F(x_n)$$

where x is the current point on the potential energy surface, γ is the step size, $\nabla F(x_n)$ is the change in the function which describes the potential energy surface. This equation shows that the energy is always moving towards a smaller value by following the negative change of the local hypersurface, also described as the negative gradient. This ensures that each iterative step produces a smaller energy value, indicating the minimisation of the system.

Advantages of the steepest descents method are that it quickly can minimise a system and is a computationally cheap method. Disadvantages are that it can easily get trapped in local minima and also it struggles to converge around a minimum, where the search for the minimum "zig zags" in direction, leading to many iterative steps that are unnecessary. This can be avoided by setting a reasonable minimum force tolerance that when achieved, ends the minimisation.

Conjugate Gradients

This then gives rise to the conjugate gradients (CG) minimization method.[111] Conjugate gradients is an improvement upon the steepest descent method due to how the searching for the minimum functions. An intuitive explanation for the difference between the two methods, is that whilst regardless of the previous search directions SD will always move in the direction of negative gradient, in the CG method each search direction is based upon the current error and all of the previous search directions, thus allowing for faster minimization when close to a minimum. Therefore instead of zig zagging towards a minimum, the CG method will generally converge in less steps. This is the advantage of the method, however CG is less effective on systems that are far from a minimum, therefore practically steepest descents is best used to approach a minimum, followed by CG to converge the energy when near the minimum.

2.1.5 Simulation Ensembles

An ensemble is defined as a group of systems which have various microscopic states but have a similar macroscopic or thermodynamic state[112113].[114] This is of relevance as MD simulations are performed at the microscopic scale, in order to attempt to replicate the macroscopic behaviour that exists in reality.

As MD generates information at a microscopic level, this information somehow needs to be related to macroscopic properties. Using statistical mechanics makes this possible, where a key component of this is the Ergodic hypothesis:

Equation 32:

$$\langle A \rangle_{ensemble} = \langle A \rangle_{time}$$

This can be equated to much how the X-ray crystallographic structure of a protein is gathered from a crystal that contains many instances of that protein in various conformations, where the resulting structure that is achieved is an average of all of the states present in the crystal. To achieve this in MD simulations, trajectories need to be long enough to give complete sampling of the states present, and often repeat trajectories are also needed to enhance the sampling, or other enhanced sampling methods can be used.

Another way to describe the ensemble is that for a system containing N particles, the system has a space containing $6N$ dimensions. A single point in phase space has the coordinates described by the positions and momentum of the particles in the system. The $6N$ dimensions of the particles in the system represent a single point in phase space, where an ensemble is a collection of points in phase space that meet a certain thermodynamic state. MD generates a series of points in phase space over time which belong to the same ensemble and correspond to various conformations of the system.

NVE

NVE is a thermodynamic state that is composed of a fixed number of particles (N), a fixed volume (V) and a fixed energy (E). This is the equivalent of an isolated system. *NVE* is also referred to as the microcanonical ensemble.[115]

NVT

NVT is a thermodynamic collection of states that is characterized by a fixed number of atoms (N), a fixed volume (V) and a fixed temperature (T). This ensemble is generally used for the initial equilibration of systems after energy minimization, in order to establish a stable temperature. This ensemble is also

useful for crystal simulations. Generally short simulations under the NVT ensemble are suitable for this, the ensemble is also referred to as the canonical ensemble.[115]

NPT

NPT is an ensemble where the number of particles is fixed (N), the volume is fixed (V), a fixed pressure (P) and fixed temperature (T). Simulations under the NPT ensemble are generally used to equilibrate the pressure of the system, before a production run to generate a full trajectory for analysis is run. This ensemble is also called the isothermic-isobaric ensemble and is closest to biological conditions.[116]

2.1.6 Temperature and Pressure

To simulate a realistic system, temperature and pressure must be taken into account. This requires the use of a thermostat to maintain the temperature throughout a simulation and a barostat to maintain the pressure (under NPT simulations only).[115]

Thermostats

In a simulation system the temperature is generally calculated from the total kinetic energy of the system. Generally the goal of the thermostat is not to maintain a constant temperature, as this would require a fixed kinetic energy, but to ensure that average temperature of the system is maintained. As MD is a microscopic method, the fluctuations in temperature are much higher, however the thermostat ensures that the fluctuations are of the correct size and maintain the average value of temperature. This is generally done by coupling the system to a heat bath of constant temperature.

Berendsen thermostat controls temperature via rescaling of the velocities of the particles in a system.[117] The system is weakly coupled to a heat bath with a temperature, T_0. The temperature is then corrected to maintain temperature, T, such that deviations in temperature decay in an exponential fashion, using a time constant τ . This can be shown via the following equation:

Equation 33:

$$\frac{dT}{d\tau} = \frac{T_0 - T}{\tau}$$

The issue with this is that kinetic energy fluctuations are prevented in the system, meaning the results of using this thermostat are not in line with canonical ensemble (NVT), particularly in small systems as the error scales as *1/N* where N is the number of particles in the system. The Berendsen thermostat is mainly used due to the efficiency at which it can set a system to a selected temperature, despite the poor ensemble representation. Generally this causes it to be used for equilibration simulations, however it should not be used in a production trajectory.

Velocity rescale (V-rescale) is a thermostat that is similar to the Berendsen thermostat, however an additional term is added such that the kinetic energy is treated appropriately.[118] This thermostat has the advantage of Berendsen thermostat efficiency and is accurate according to the canonical ensemble. The scaling is similar to Berendsen, tau-t, where an additional stochastic term allows for the correct ensemble behaviour as in Equation 34. In this equation K is kinetic energy, K_0 is the kinetic energy of an external bath, N_f is the number of degrees of freedom, τ_T is the temperature coupling time constant and dW a Weiner process. These alterations make the thermostat produce a correct canonical ensemble and maintain first order decay of temperature deviations.

Equation 34:

$$dK = (K_0 - K)\frac{dt}{\tau_T} + 2\sqrt{\frac{KK_0}{N_f}}\frac{dW}{\sqrt{\tau_T}}$$

Nosé-Hoover is another thermostat that accurately samples the canonical ensemble.[119] It varies from the others as the weak coupling to a heat bath is not ideal for equilibrium simulation. This is done via the introduction of a thermal reservoir and a friction term to the equations of motion. The main practical difference between the two types of algorithm is that with Berendsen/V-rescale there is an exponential relaxation of temperature, with a damped kinetic energy fluctuation, whilst Nosé-Hoover produces oscillatory relaxation of temperature. This generates a fluctuating temperature that gives more realistic dynamics.

Barostats

Pressure is maintained in a similar way to temperature, where the microscopic environment has an average pressure that is maintained for the simulation.[120] In a manner similar to temperature, the system is coupled to a pressure bath.

The Berendsen barostat functions similarly to the thermostat, where the pressure is treated instead of the temperature:[121]

Equation 35:

$$\frac{dP}{d\tau} = \frac{P_0 - P}{\tau}$$

This once again provides first order relaxation towards a defined pressure P_0 from the current pressure P using a time constant τ. This carries the same issue as the thermostat, that the NPT ensemble is not well represented by the barostat, but it is useful for equilibration due to the first order method.

Parrinello-Rahman is a barostat that is used for production dynamics.[122] It simulates the actual NPT ensemble. It functions in a way similar to the Nosé-Hoover thermostat, via providing additional terms to the equations of motion, which allow the box size and the pressure to oscillate. This provides a better description of natural pressure in a system.

Pressure coupling can be applied in a variety of ways. For the systems without interfaces, isotropic coupling can be used, such as a protein in solvent. This causes the scaling of the box size in x,y,z to occur equally and simultaneously. Semiisotropic pressure coupling allows the x,y dimensions to be scaled independently of the z dimension. This is useful for systems such as membranes, where across the periodic boundary there is an interface in the x,y plane.

An important takeaway from this is that equilibration simulations and production dynamics generally do not use the same thermostats and barostats.

2.1.7 Periodicity

In order to simulate a biological system properly the edges and faces of the simulation box must be treated properly, otherwise the system will be suspended in a vacuum giving unrealistic effects, Figure 2.5. This is achieved via the use of periodic boundaries at the edges and faces of the box.[106] During a simulation the simulation box behaves as if it is surrounded by equivalent boxes containing the same molecules. This allows particles to leave one side of the box and move to the other without causing artefacts and also for the realistic simulation of pressure, temperature and diffusion.

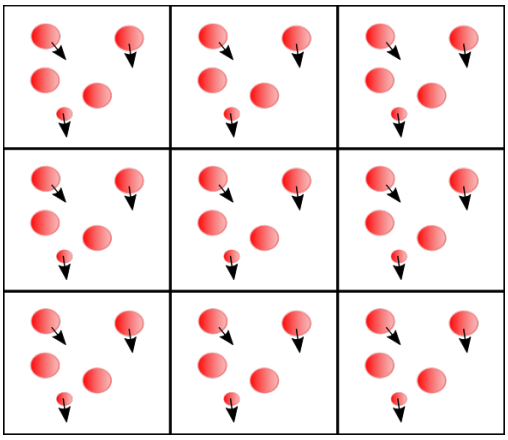


Figure 2.5: An example of periodicity, where the box is replicated around every face of the central box.

This can also be of particular use for some simulations, such as a membrane systems, where by using a small patch of membrane that fills the box in the x and y axes it is possible to simulate a bulk membrane which is uninterrupted. Another notable use for this is periodic molecules, where polymers can be self bounded through the boundaries, such as DNA to produce one continuous chain of monomers that is unbroken during the simulation.[7]

2.1.8 Restraints

Restraints can be imposed upon particles in a system during a simulation. Position restraints cause the particles to be harmonically fixed to their reference positions. This can be necessary in some cases as during equilibration, before a production run, molecules such as proteins can drastically rearrange as a result of unequilibrated solvent. Other uses include using position restraints to create barriers in a system or to pin a molecule in place such as to bias the sampling to a certain conformational space, e.g. in umbrella sampling.

The mathematical expression of a typical position restraint is similar to that of an angle or bond in a forcefield:

Equation 36:

$$V = \frac{1}{2}k|\vec{r} - \vec{R}|^2$$

Where V is the potential energy, k is the force constant of the restraint, r is the position and R is the reference position of the restraint. This essentially forms a harmonic, where an energy penalty is applied in order to prevent the

particles that are restrained from moving during a simulation. This can be expanded upon due to the nature of x,y and z coordinates, see Equation 36, for the expanded form of Equation 35.

Equation 37:

$$V = \frac{1}{2}|k^x(\vec{x} - \vec{R_x})^2\hat{x} + k^y(\vec{y} - \vec{R_y})^2\hat{y} + k^z(\vec{z} - \vec{R_z})^2\hat{z}|$$

This allows for different force constants to be imposed upon the three coordinates for restraints, so for example, particles can be isolated to a certain plane.

Restraints can also be applied in other ways, such as distance restraints or orientational restraints in order to enforce certain behaviours in systems during simulations. Flat bottomed potentials can be used in order to restrict particles to a certain section of a simulation box, where they experience no restriction until they move to a certain boundary where an energy penalty is then applied. The important point of restraints is that they can be a necessary part of building a simulation system in order to produce relevant results.

2.1.9 Simulation Resolution

There are many levels of detail that can be simulated. This generally revolves around the scale of the system that is in question and the aim of the simulation.

All Atom

For extremely specific interactions and possibly the most realistic simulations used in MD, all atom forcefields are used, where polarisable forcefield are available but are much slower. This type of forcefield uses molecular structures that are identical to the skeletal structure of the molecules, where all atoms are included.

CHARMM or AMBER are examples of this. These forcefields are widely used in order to simulate biological molecules, and these forcefields are well validated and trusted for simulation[85].[123] They agree well with experimental data and generally give good results. The downside of all atom simulation is that the number of particles in the system is higher than other resolutions, this means that the computational demand is higher to achieve simulations of a similar length.

United Atom

United atom forcefields are of lower resolution than all atom. This is achieved by the incorporation of non-polar hydrogen atoms into the carbon that they are respectively bonded to. The mass is incorporated into a particle that is the sum of the carbon and hydrogen atoms. This simplifies systems such as membranes and other non-polar environments. This approximation leads to a less accurate forcefield however, but there is a significant speed up as the number of total particles is less.[124]

GROMOS is a forcefield that was developed as a united atom model. This forcefield has also been widely used and validated in a large number of publications.[125]

Coarse Grain

Coarse graining refers to a lower resolution than that of united atom. Similarly to how the hydrogen and carbon atoms are grouped in united atom, heavier atoms can be grouped together into particles known as beads. This grouping is dependent on the forcefield used, for example MARTINI uses an approximate 4:1 heavy atom to bead mapping. An example of this would allow a sixteen carbon lipid tail to be described by four beads[126].[127]

MARTINI is a coarse grain forcefield that is popular and widely used in biomolecular simulations.[65]

Coarse graining comes with two advantages, as the number of particles is significantly reduced the simulations will be much less computationally demanding. The other benefit is that with this method all of the hydrogen atoms in the system have been removed. As mentioned earlier light atoms force timesteps in MD to be short. Considering that all of the atoms are now located inside of beads, which contain their cumulative mass, all of the beads are much heavier. This allows for much longer timesteps, so not only do simulations proceed faster per step, but the timestep used can be much higher, e.g. twenty femtoseconds. Generally in the CG approach potential are also "softer", reducing large forces and contributing to the longer timestep. Coarse grain simulations therefore are extremely popular for large systems of millions of atoms.

With these benefits, come some disadvantages. Considering the level of detail that has been removed from molecules in the system, do they still behave realistically? MARTINI water is also coarse grained, but the increased mass of the beads makes the water less fluid, necessitating the use of anti-freeze particles to stop the water from freezing at room temperature. Elastic networks are sometimes used in order to maintain protein secondary structure in this forcefield, so more approximations are necessary in order to achieve realistic behaviour.[128]

This being said, coarse grain simulation can applied to many situations where it is simply impossible to capture dynamics at atomistic detail due to current limitations in hardware. A prime example of this can be seen in the work of simulating entire viruses, as well as other simulations that are too vast and complex for a higher resolution.

2.1.10 Enhanced Sampling

As mentioned previously, it is possible to enhance the sampling of MD. This is done in order to observe events that require better sampling of phase space than would typically be accessible to a standard MD simulation[129].[130]

Standard MD issues

Standard MD is the simulation of particles with no bias or influence. This allows the sampling of the system as it would naturally evolve according the forcefield and parameters for the simulation, in an unbiased fashion. The issue with this is that only one path through phase space can be revealed. It is known that events occur on a Boltzmann distribution, where there are many conformations available, but if a system gets trapped in a stable state or an energy well then it is not possible to sample all of the potential states in that system. This can be seen with respect to a potential energy surface, where some states are simply inaccessible from a current state due to the large increase in energy that would require the state change.

Umbrella Sampling

This is a method that is widely used to calculate the free energy difference in an evolving system.[131] This is done by generating configurations along the pathway of interest, such as pulling a lipid from a membrane bilayer. By providing these configurations and linking them together with umbrella potentials, which flattens the energy landscape, many states can be sampled in individual simulations. Ideally umbrella sampling then allows for the sampling of all of the states in the system, however it is rare that every possible state is sampled.

The large number of degrees of freedom in even simple molecular systems makes it difficult to sample all of the possible states, as such generally coordinates along a certain axis are used. This practically would involve a simulation to generate the reaction coordinate configurations, followed by many individual simulations of each configuration to explore the phase space. The output of each configuration is then analysed to compute a potential of mean force and the free energy change involved in the selected pathway. One method of analysis used to unbias the data is known as the Weighted Histogram Analysis Method

(WHAM).[92]

WHAM

WHAM is used to analyse the individual windows of the simulation. A key point is data for the variable being studied must have good overlap between the configurations, often represented as a histogram each. It is also important that each configuration has been sampled for long enough, such that a smooth PMF (Potential of Mean Force) profile is produced.

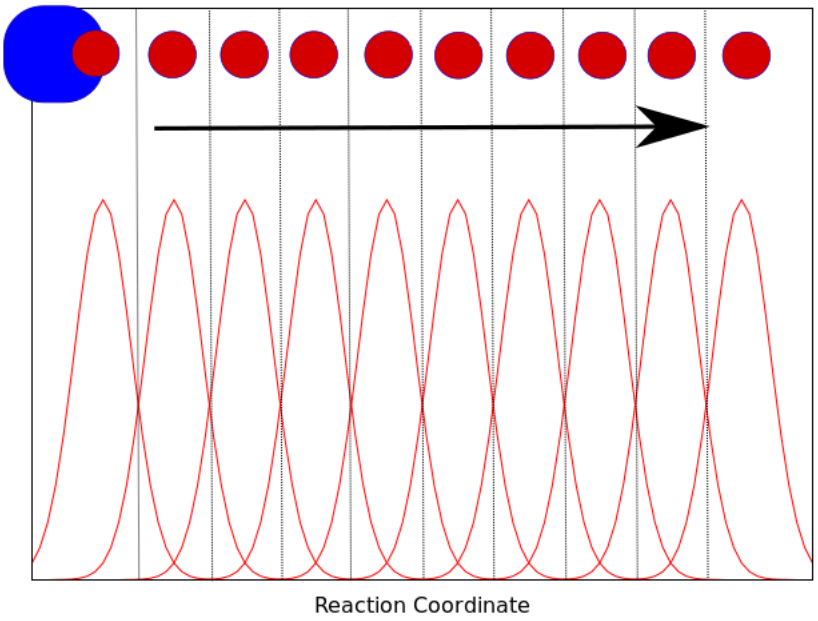


Figure 2.6: A visualisation of how the windows are positioned in umbrella sampling, in blue a binding site and red a substrate being pulled from this binding site.

WHAM allows information from all of the intermediate states to be used. Real data that is analogous to Figure 2.6 is displayed in Figures 2.7 - 2.10, in relation to the enhanced sampling of a protein chaperone, LolA when bound to varying degrees of an inhibitor, present in *E. coli*. This work is discussed in a later chapter in this thesis.

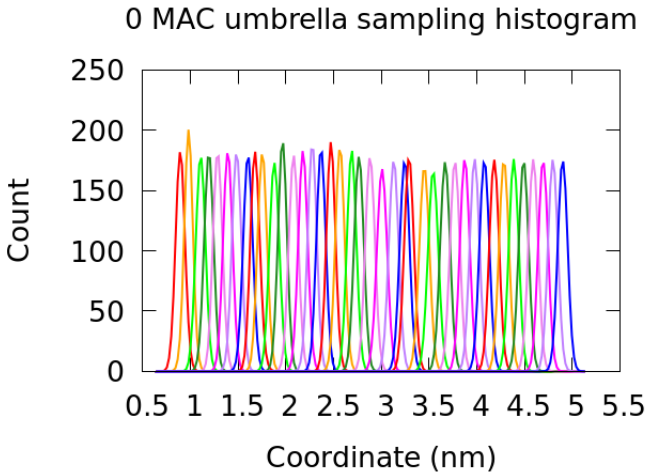


Figure 2.7: Histogram plots to show overlapping regions of sampling along windows for the lipid removal from the apo-LolA umbrella sampling simulations

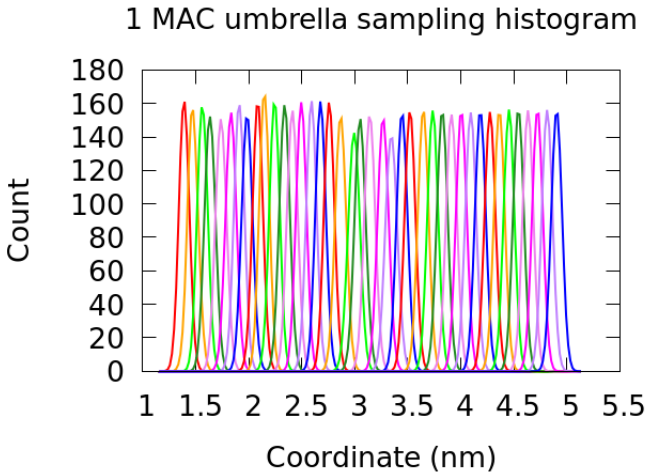


Figure 2.8: Histogram plots to show overlapping regions of sampling along windows for the lipid removal from the 1MAC-LolA umbrella sampling simulations

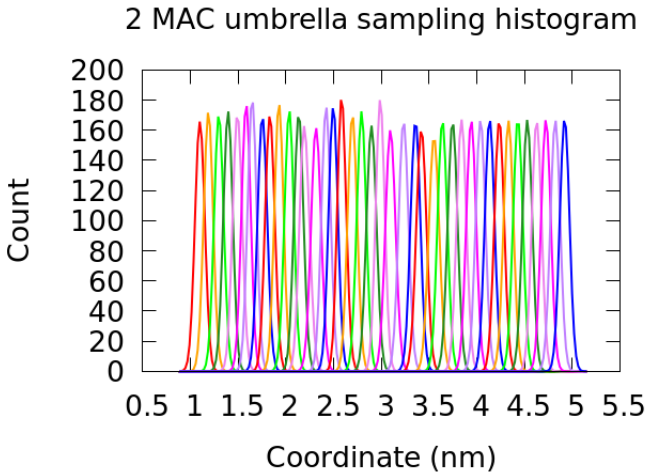


Figure 2.9: Histogram plots to show overlapping regions of sampling along windows for the lipid removal from the 2MAC-LolA umbrella sampling simulations

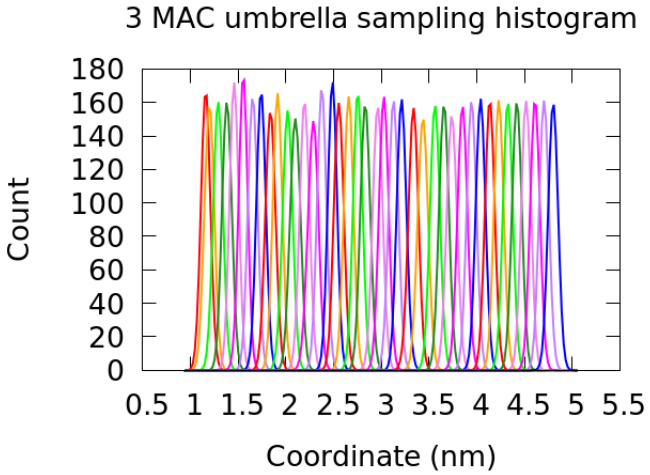


Figure 2.10: Histogram plots to show overlapping regions of sampling along windows for the lipid removal from the 3MAC-LolA umbrella sampling simulations

Visualisation

Visualisation of systems is generally done using various software available such as VMD and Pymol, where these allow brief analysis and visual inspection of static simulation structures, alongside trajectories of structures generated by

simulation[132].[133] Visualisations in this thesis were made using VMD, Pymol or Chimera.

Analysis

Analysis of the data presented in this thesis is done using a variety of programs, chiefly among them self generated python scripts, the MDAnalysis python package and the GROMACS analysis tools[134][135][136].[95]

2.1.11 General Simulation Conditions in this Thesis

Simulation details

For equilibration Berendsen temperature and pressure coupling was used over short timescales (100 ps) before production dynamics were run using the V-rescale thermostat and Parinello-Rahman barostat for GROMOS simulations.

For CHARMM simulations the Nosé-Hoover thermostat was used, with the Parinello-Rahman barostat. Similar values are used for the equilibration and simulation.

Biological Conditions

All simulations presented here are intended to mimic *in vivo* biological conditions. This requires that the conditions for the simulation are equivalent to conditions that would exist in reality. Important aspects of this include temperature, pressure, ionic concentrations and other variables.

For simulations of *E. coli* simulation temperature is that of the human host, 310 K. If any data generated is to be related to *E. coli* cells during a human infection then this makes logical sense.

Pressure was set to 1 bar, which is ~ 0.98 atm. The turgor pressure of bacterial cells, in particular *E .coli* has been estimated to be 0.3 - 3 atm. Therefore this makes this a reasonable value to use for *E. coli* membrane simulations and generic simulations of bacterial environments.[76]

All of the simulations were performed in explicit solvent. The ubiquitous solvent in biology is water, where in the simulations present the water surrounds the heteromolecules in a bulk. Generally visualizations of systems containing water would make it hard to show anything else, so for clarity the water is removed from any images of resulting simulations in this thesis.

The ionic concentration must be within a reasonable range to mimic the biological levels. Ions are included in the bulk solvent to mimic this, additionally in order to produce a system that is of neutral charge to allow for PME. Ionic concentration in bacteria can vary, but a concentration of ~ 0.2 M sodium chloride was typically added to each system. Where LPS is present magnesium cations were inserted on the membrane to follow *in vivo* conditions. This concentration can vary depending on the number of additional counter ions needed to neutralise any charges. However as this is the microscopic scale it follows that in these local environments the ions could be distributed in a ratio that is not one to one.

Procedure for Simulation

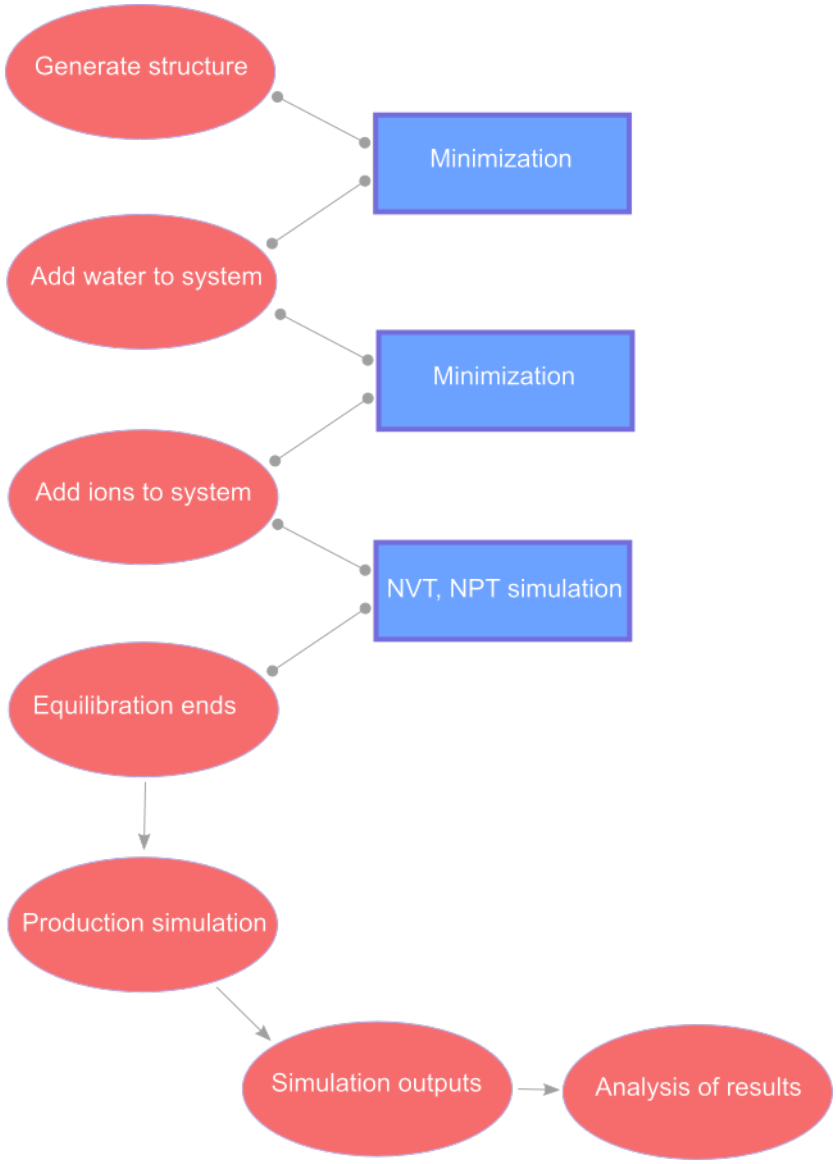


Figure 2.11: A flowchart for the basic set up and simulation of a solvated biological system.

2.1.12 Peptidoyoucan

Modelling polymerization of a monomer is a task that requires scripting. Considering that a small patch of cross linked peptidoglycan can easily contain at least a hundred monomers the following script had to be created. This was necessary due to the need to handle the topologies and structure files that are required to equilibrate the cell wall at each step. This essentially involves the following steps:

- Start with a monomer of peptidogylcan.
- Add an additional monomer, bonding the two together glycosidically and minimize the system.
- Repeat the previous step until the chain of monomers reaches the desired length.
- Take the completed chain and translate copies parallel to the original to form several chains
- Randomly cross link the chains, according to a specified cross linking degree.
- Minimize, solvate and equilibrate the cell wall for a short time.

This has allowed for the quick production of bacterial cell walls of various sizes and degrees of cross linkages. Currently the two models supported are for CHARMM36 and GROMOS54a7. The glycosidic bonds and peptide bonds that are automatically created use existing parameters within the forcefield. By using the standard parameters that are available within the molecule for this type of bond it is possible to replicate the bonding for many monomers and chains. The bonding is additionally periodic, such that the peptidoglycan can behave as one continuous mesh, shown in Figure 2.12.

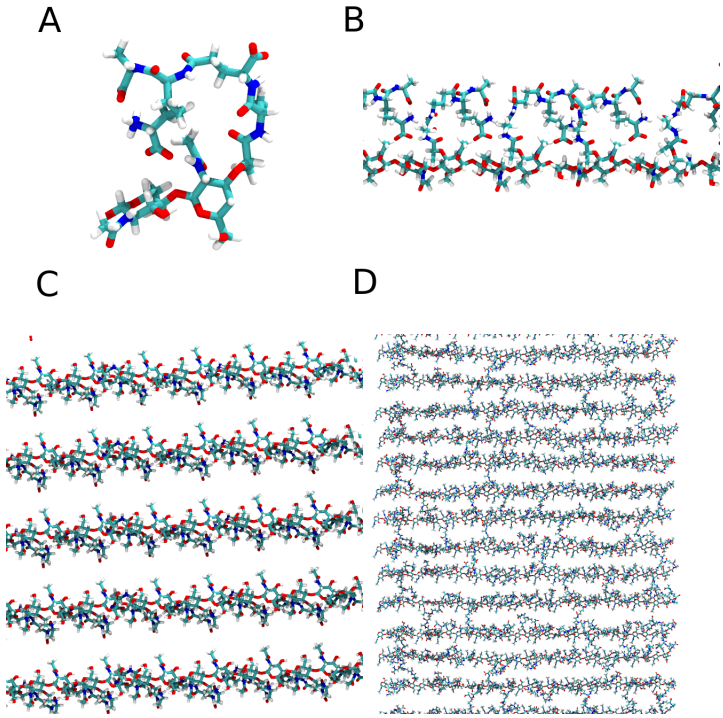
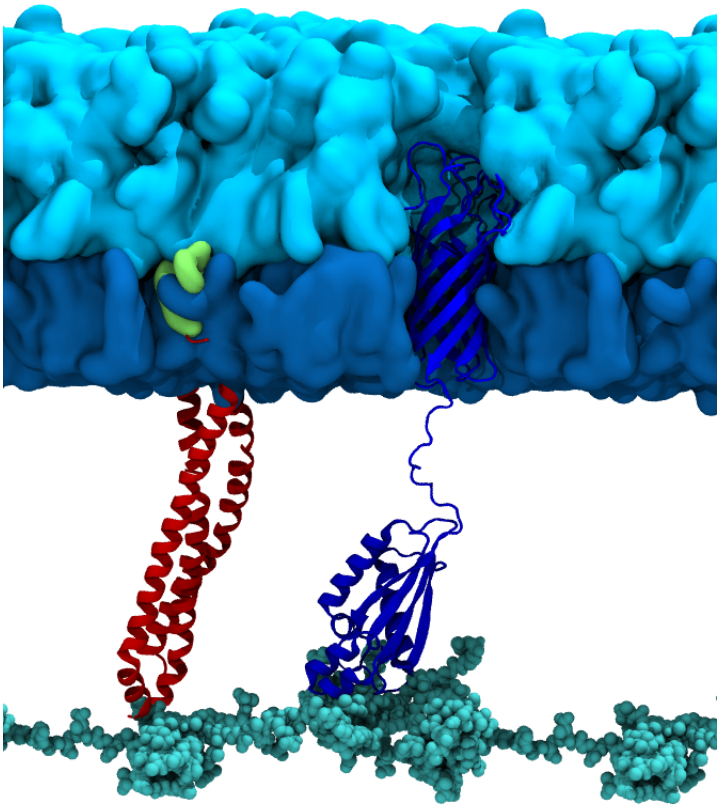


Figure 2.12: The Peptidoyoucan process, where a monomer (A) is grown into a strand (B). Once a periodic strand has been grown, it is copied into a sheet (C). This sheet is then randomly cross linked to give a periodic cell wall in both the X and Y planes (D).

The functionality to allow the system to be flexible with respect to size and the degree of crosslinking allows for the potential study of various properties when these variables are changed. Simulations exploring this are included alongside work studying protein interaction with the cell wall. The working version of this code is available under Appendix A, titled ***Peptidoyoucan.py***. As a long term prospect this code will continue to be developed for use within the group.

Chapter 3

Elucidating OmpA interactions with BLP & the cell wall



3.1 Context in the Literature

BLP is the most abundant source of protein that is present in *E. coli*. This lipoprotein consists of a 58 amino acid sequence followed by a lipid moiety at the N-terminus[36].[137] It was the first discovered lipoprotein.[138] The monomer trimerises into the helical structure that is shown in most textbook representations of BLP, Figure 3.1. One monomer of each trimer is covalently peptide bonded via a C-terminal lysine to the tetrapeptide cross-linking site on the NAM monomer of the cell wall mesh.[137] This linkage to the peptidoglycan is via the diaminopimelic acid molecule present on the NAM monomer of the cell wall. The lipid trimer of this bound form rests in the OM, which makes up a third of BLP present, the rest is free in the periplasm. This interaction spanning the periplasm provides structural integrity to *E. coli*.[38]

Braun's lipoprotein was first discovered by Braun in 1969, where it was named as murein-lipoprotein. This study showed that trypsin can be used to cleave the lysine peptide bond between BLP and the cell wall.[139] In addition to this it was estimated that the ratio of BLP to cell wall is approximately one trimer to ten PGN monomers. In this study it is also proposed that BLP has a stabilising effect on the structure of the cell wall. The lipid component was studied, where after aqueous acid or alkaline hydrolysis the insoluble product could be dissolved in chloroform or methanol. The main component of this lipid component (65 percent) was found to be palmitic acid from the lowest R_f (retention factor) spot on a TLC plate.[36] Further studies showed that the lipid is complex in structure and is most likely bound to either an aspartic acid or serine. This spawned the theory that the entire structure was lipid-protein-murein, where the three moieties are covalently bound. The BLP in this study is proposed to be abundant, on the order of 10^5 molecules present in the periplasm. The inference in this first study is that the lipid component is associated with the OM[140].[139]

Having characterised the lipoprotein, Braun followed with several studies on BLP. It was confirmed that the lysine that remains, post-trypsin, on the peptidogylcan originates on the BLP protein.[38] It was also found that the residues that are split via trypsin digestion were between a lysine and arginine. This paper also suggests that the terminal alanine, residue five, on the NAM molecule is degraded and is not present in the linkage between chains and between the murein and the BLP[141].[139] Along with this structural insight, the model of the "rigid layer" was proposed, suggesting peptidoglycan chains of 10 to 65 disaccharide monomers, with evenly spaced BLP bonded to the wall. Following this, the sequence of BLP was also published by Braun. After establishing the sequence of the structure, it was then found that the lipid component is formed of a triacylated N-terminal cysteine which is anchored in the OM. This completed the structural eludication and the behaviour of the murein-lipoprotein, which gained the name Braun's lipoprotein from Braun after his extensive work. It is thought that *E. coli* expresses approximately 100 lipoproteins that are present

throughout the cell; however, currently BLP is the only known example covalently linked to the cell wall.

Further study of the linkage showed that the bond is added by three transpeptidase proteins; LdtA, LdtB, and LdtC[142].[143] The enzymes are all homologs and contain a conserved cysteine in the active site.[144] Of these enzymes LdtB, is predominant, where it comprises 90 percent of the enzymes of the Ldt group. It has been suggested that the enzymes are required for various conditions, however the BLP linkage to PGN is mostly absent when the gene for LdtB is deleted.[145]

When BLP is not linked to the cell wall, this is known as the free form of BLP. Approximately two thirds of BLP is free.[138] It is still unknown how this free lipoprotein interacts with the cell wall. It is also proposed that this free form of BLP can be surface-exposed on the OM of the cell. This was based on labelling and results have shown that expressing mutants lacking the C-terminal lysine increases the amount of surfaced exposed BLP, as opposed to a mutant that is only PGN bound and is not detected on the surface. It is unclear as to how the BLP is inserted into the OM in such a way that is is surface presenting.[146]

BLP is not unique to *E. coli*, and homologous proteins are present in other species such as *Pseudomonas* and *Salmonella* strains.[138] Generally it is a well established protein across many strains, where the terminal lysine is also well conserved. It has also been shown that mutations in this amino acid results in the loss of BLP-PGN linkages. The reasons behind the large number of BLP molecules present in the periplasm have also been studied. A 2014 study suggests that there are approximately one million copies per cell.[147] The high levels of expression are represented by approximately 8 percent of translation occurrences in the cell involving BLP mRNA. This has been attributed to a high adenine-thymine content, which is thought to aid strand unwinding for transcription. Another factor is that the BLP mRNA is highly stable with a half life of 12 minutes, versus the average 1.3 minutes for other mRNA present in *E. coli*.[147]

E.coli has an included stress response involving BLP.[148] A post transcriptional inhibition of the stable mRNA can occur, resulting in a reduction in BLP production when the cell is experiencing issues such as improper protein folding or mislocalised LPS molecules.

The function of BLP has been studied. It has been reiterated many times that it serves a mainly structural role. This role is referenced in two capacities. The first is that of maintaining the integrity of the OM of the cell. Cells that have the gene for BLP deleted are still viable and can survive, however they do not behave the same as the wild type-containing BLP.[149] These cells experienced increased sensitivity to antibacterial compounds[150].[151] BLP deletion

also causes periplasmic proteins to be released into the surrounding environment, along with OM blebbing.[152] This is evidence for the suggestion that the BLP effectively ties the OM to the PGN layer. Without this linkage, the OM can effectively float away and is easier to disrupt. It has been noted that this phenotype where BLP is entirely deleted is similar to that of the terminal lysine deletion mutant, indicating that it is specifically the linkage that prevents them.[146] The second structural role of the BLP is as a distance maintainer. Studies have shown that BLP insertions in order to lengthen the protein section result in an increased OM to PGN distance[153].[154] The length of BLP is generally conserved with little change in the number of amino acids present in the protein. It has therefore been assumed that BLP determines the size of the periplasm in Gram-negative bacteria.

OmpA is a 325 residue outer membrane protein, consisting a membrane localised barrel region (residues 1 - 171) and a globular periplasmic domain (residues 180 - 325), see Figure 3.1 , where a flexible unstructured region (residues 172 - 179) links the two main domains.[72] The barrel domain is composed of eight antiparralel β strands, with short turns at the periplasmic side and long loop regions at the extra-cellular interface. The globular periplasmic domain is a folded domain that is similar in structure to known peptidoglycan binding domains. OmpA is present in both a monomeric and proposed dimeric form.[155] Much like BLP it interacts with the cell wall, but non-covalently. This is once again proposed to be a stabilising interaction.[18] OmpA can interact with the cell wall as the periplasmic domain contains a binding site for the PGN network.

OmpA has also been extensively studied, due to the surface exposed region, seen as a potential target for therapeutics, as there are hundreds of thousands of copies per cell. It was first purified in 1977, followed by sequencing in 1980, since it has been characterised via many techniques, including crystallography, mass spectrometry and solution NMR[156][157][158].[90] The function of the protein is often stated to maintain the integrity of the OM and to relieve osmotic stress. Mutation of the structure, or removal of the protein gene causes increases sensitivity of the OM to stress. The structure of OmpA was first resolved for the transmembrane domain, followed by inclusion of the periplasmic domain. The periplasmic domain has been studied for PGN binding ability, including temperature dependent structural transitions of the periplasmic domain. It is still a matter of discussion as to whether the protein is present predominantly in a monomeric or dimeric state, where several studies have presented either as possible options. Diffusion through OmpA has been compared to the behaviour of OmpC and OmpF, other outer membrane proteins in the same family, where small solutes were found to diffuse much slower through OmpA.[159] Mutations in the OmpA protein have also been shown to increase pathenogicity of bacteria.[71]

Simulation studies of both of BLP and OmpA currently exist in the literature. Currently there are two simulation studies of BLP, one published based

on the work contained in this thesis and another published by Hwang et al.[160] Whilst our study focuses on the interactions between OmpA, BLP and the cell wall, the other focuses on the stress and mechanical properties that the periplasm undergoes. Here the model used was of BLP connected to the cell wall and non-covalently connected to the OM. This system contains many other proteins. The findings from the simulations were for the values of K_A, the area compressibility of various barriers present in the simulated periplasm. The cell wall was shown to be able to expand when under tension, where the peptide cross-links are the more flexible direction compared to the glycan strands. Comparisons to experimental values for the membranes indicated that the K_A was underpredicted for the IM and overpredicted for the OM.

A key point of the simulations showed that when under higher turgor pressure in the cell, the cell wall will bear much more of the tension, approximately 80 percent, whilst in lower pressure systems the cell wall shares the tension with both membranes equally.[160] This supports the theory that the cell wall exists to protect against the high internal pressure of the cell, via stretching in the glycan and peptide directions. This study is interesting in understanding the mechanics of the periplasmic components, but does not explore the interaction between included membrane proteins.

OmpA simulations are more extensive than those of BLP. The first simulation of this protein, before the periplasmic domain was resolved shows the presence of a pore in the beta barrel membrane domain, along with simulations in a bilayer, where a gating mechanism between the open and closed state is proposed. This exploration of the porin type behaviour showed that perturbing this gate at ARG138-GLU52 allowed for complete permeation of the pore, allowing for agreements with experimental conductance data.[70] This led to the proposal of a mechanism where the ARG138 swaps between pairing with GLU52 and GLU128. The drawback of the simulation study was the timescale, which as the study occurred in 2002, only a total of 19 ns was achieved from the study. An experimental paper by Lucas Tamm and co-workers then confirmed that ARG138 and GLU52 do act as a gate to control the pore.[161]This shows potential meaningful results have been drawn from these short simulations of OmpA and leaves the protein open for further simulation at longer timescales.

A comprehensive overview of OmpA and other OMP simulations followed this, where simulations of homology models were the first to show the proposed periplasmic domain, as it had yet to be characterized for OmpA. The periplasmic domain was first modelled by Khalid et al., where it was first suggested that this domain could interact with PGN.[18] This first model was constructed via fusing a RmpM C-terminal domain (40 percent similarity to *E. coli* CTD OmpA) to the barrel of *P. multocida*. The basis for the interaction with PGN is the determination of a crystal structure of *A. baumannii* in complex with a synthetic PGN molecule. The first proposition of a full length OmpA protein model came from Robinson and co-workers. Using mass spectrometry of sev-

eral OmpA fragments, a low resolution model of full length OmpA was devised, where the proposed model is actually dimeric in form. In this study it was proposed that residues 188-276 are responsible for the dimerization, where cross linking in tandem with analysis of the periplasmic domain truncated in the following patterns: residues 188-325, residues 188-276 and residues 227-325 showed that only the protein at residues 227-325 was monomeric. It was seen during studies of OmpA that it often presented as either a monomer or a dimer.[70] There is evidence in both camps to suggest that it can exist in either form *in vivo*.

Approximately 15 years after the first simulation, Samsudin et al. showed via MD that OmpA can act as a flexible clamp to contact the cell wall such as to promote OM binding to the cell wall. This was done with single strands of PGN. The PGN binding domain was assessed, where ASP241 and ARG251 interacted with the cell wall closely, forming salt bridges with the m-DAP residue on the side chain of the PGN. It was also proposed that the PGN can adopt many conformations inside the binding site. This work included simulations of both monomeric and dimeric OmpA. The monomeric simulations showed linker contraction, which pulled the PGN close to the OM. It was proposed that dimerisation of OmpA would prevent this. However in this simulation, as only single strands were used, it required much less force to pull a strand of PGN to the OM, versus a networked mesh.[18] This work led to the development of a PGN mesh for the study included here.

The motivation for this study comes from the question that if OmpA and BLP both interact at some level with the peptidoglycan, is there some cumulative interaction that allows for this? Considering that the cell wall is present in the periplasm as a sheet this presents the opportunity to build a model for simulation. In particular, the main question is if BLP can regulate the OM to PGN distance, how does OmpA bind to the cell wall easily, as the full length BLP in a perpendicular orientation to the cell wall would prevent contact between the OmpA and PGN? Based on this we have constructed our model of the cell wall and using experimentally derived data for OmpA and BLP, intending to explore the interactions of these proteins and peptidoglycan in a model *E. coli* periplasm.

3.2 Premise

E. coli contains a complex cell environment, which is a target for modern therapeutics.[162] Further exploration of the periplasmic space is required in order to begin to understand the construction of the cell envelope. Using atomistic MD simulations allows for molecular detail interactions to be categorised between OMPs and the peptidoglycan layer that exists in the periplasm.

OmpA has previously been shown to be a flexible clamp for the cell wall that originates in the OM of *E. coli*.[18] BLP also occupies this space, where unlike OmpA there is a covalent attachment to the cell wall.[163] All textbook representations of BLP indicate that the trimeric lipoprotein is perpendicular to the membrane and the cell wall.[164] Based on the previous clamping to the strand of cell wall observed in simulation the following questions are asked in this study:[18]

- How does a mesh of cell wall interact with the OmpA periplasmic domain?
- Does dimerisation of the OmpA protein affect the binding?
- How does BLP behave in the periplasmic environment and does it affect OmpA-PGN binding?
- How does BLP affect the behaviour of the cell wall in the periplasm?

Previously, modelling of the cell wall in a periplasmic environment has been sparse to non existent; therefore this study is one of the first of a kind where both a membrane and a cell wall mesh are used in tandem in a simulation. It is important to ask these questions, as the first step in assessing periplasmic behaviour is to generate as much basic information as possible, by starting to create models of this compartment.

3.3 Methods and Simulation content

All simulations were run for 100 ns with an independent repeat, where resources limited our timescales. The contents of the OM were entirely LPS in the outer leaflet and a ratio of 90:5:5 of PE, PG and cardiolipin lipids in the inner leaflet. These lipids are known as 1-palmitoyl 2-cis-vaccenic phosphatidylethanolamine, 1-palmitoyl 2-cis-vaccenic phosphatidylglycerol, and 1-palmitoyl 2-cis-vaccenic 3-palmitoyl 4-cis-vaccenic diphosphatidylglycerol respectively. The LPS used is referred to as rough Ra-LPS, meaning that the LPS model lacks the long O-antigen; instead only the core sugars are modelled. The negative charge of the LPS was balanced with magnesium counter ions that are known to be present on the membrane surface. The BLP model used consists of 58 amino acids, where the 20 amino acid signal peptide was cleaved; this protein structure was constructed based upon work by Shu et al. (PDB:1EQ7).[165] The N-terminus was functionalised with tripalmytoyl-S-glyceryl-cysteine, using parameters from the GROMOS54A7 forcefield. The linkage between the BLP and the m-DAP residue of the PGN was constructed using standard parameters for a peptide bond from GROMOS54A7. The OmpA model used was the full length model provided by Carol Robinson, which has previously been used in simulation work[155].[166] OmpA was inserted into the membrane via the membed function of GROMACS.

System	OmpA	BLP	Temperature (K)	OmpA contacting PGN?
Control	None	Yes	310	N/A
Control	None	Yes	323	N/A
Sys1	Monomer	Yes	310	Yes
Sys2	Monomer	Yes	323	Yes
Sys3	Monomer	No	310	No
Sys4	Monomer	No	323	No
Sys5	Dimer	Yes	310	Yes
Sys6	Dimer	Yes	323	Yes
Sys7	Dimer	No	310	Yes
Sys8	Dimer	No	323	Yes

Table 3.1: Simulation systems set up for the OmpA, BLP, PGN periplasmic environment.

The peptidoglycan sheet model used was constructed from three strands of 10 repeating monomeric units of NAM-NAG.[18] A single layer of PGN was used, as according to studies approximately 75 percent of the cell wall is monolayered. The strands were then cross linked by peptide bond between residues 3 and 4 on the side chains of opposing NAM monomers. The initial setup of the system comprised of the membrane with the cell wall located approximately 90 Å from the inner leaflet of the membrane. The presence of the BLP and monomeric

and dimeric OmpA was varied in order to compare the ability to bind to the cell wall and how the cell wall location is affected by any binding.

Simulations were run at 310 and 323 K using the V-rescale thermostat at 1 ps time constant. The 323 K simulations were done to study the system at a higher energy. The Parrinello-Rahman barostat maintained pressure semi-isotropically at 1 atm, using a 1 ps time constant. LINCS was applied to allow 2 fs timestep.The short range cutoff for both VdW interactions and charged interactions was set to 1.4 nm.[100] The SPC water model was used to solvate the system and a concentration of 0.2 M NaCl ions was used to neutralise the charge in the system.[167] The simulation used the GROMOS54A7 forcefield.[125] The system sized used was of a membrane approximately 10 nm x 10 nm, with enough space in the box to prevent periodic interactions in the x and y dimensions.

3.4 Results

3.4.1 OmpA dimer 310 & 323 K

Simulation on one side of the *E.coli* periplasm has shown that the proteins OmpA and BLP are key staples that maintain the distance between the OM and the cell wall. OmpA achieves this via the periplasmic domain of the protein, using a flexible binding region that appears to be specifically evolved to contact the cell wall. This interaction can vary over time, as it is non-permanent and governed by intermolecular forces. BLP however is covalently bonded to the cell wall, where the N-terminal region of the protein rests in the OM, whilst the C-terminal region of the BLP timer is located at the cell wall.

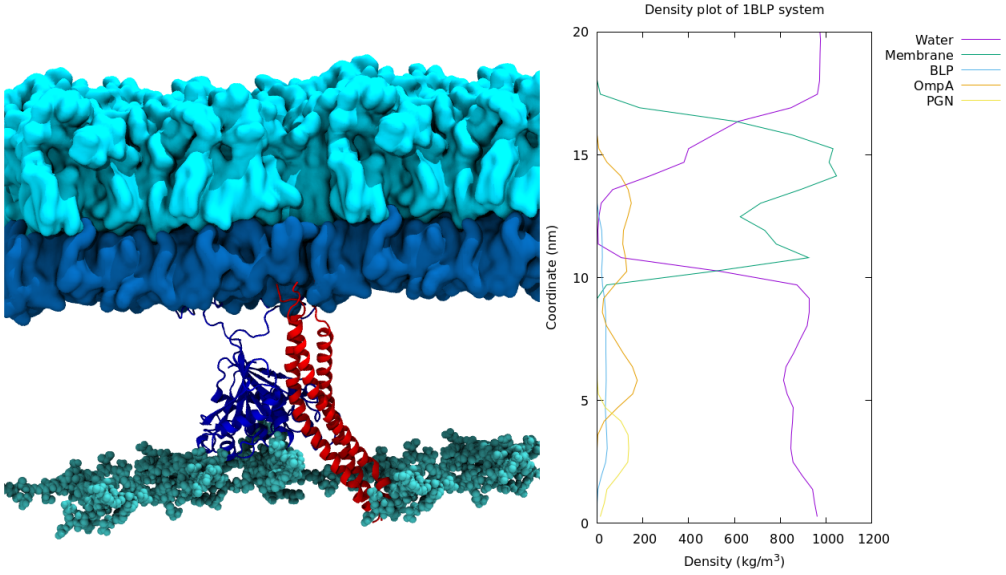


Figure 3.1: A visualisation of the OmpA dimer (blue cartoon), BLP (red cartoon), PGN (cyan spheres)and the membrane (light blue for outer leaflet and dark blue for inner leaflet, both in the surface representation) periplasmic system (left). The corresponding density of the components in the z-axis over 100 ns (right). In this image, the OmpA periplasmic domain, also known as the clamp is bound to the cell wall.

OmpA is proposed to exist in both monomeric and dimeric form, where the interaction between OmpA and BLP must occur at some level in *E.coli*, due to the abundance of BLP. Due to the microscopic scale of these simulation, a small

number (1-2) BLP trimers were included in the system. Figure 3.1 shows that the BLP will spontaneously interact with the OmpA dimer, and that the distance between the cell wall and the OM was defined and limited by the length of the BLP, Figure 3.2. See Appendix Figure 7.1 for further visualisation of OmpA and BLP.

It can be seen in Figure 3.1 that densities of the BLP, OmpA and PGN overlap, indicating that during simulation the association of the proteins occurred and therefore can be further examined, where the physical effects on this rudimentary model of the periplasm can be measured.

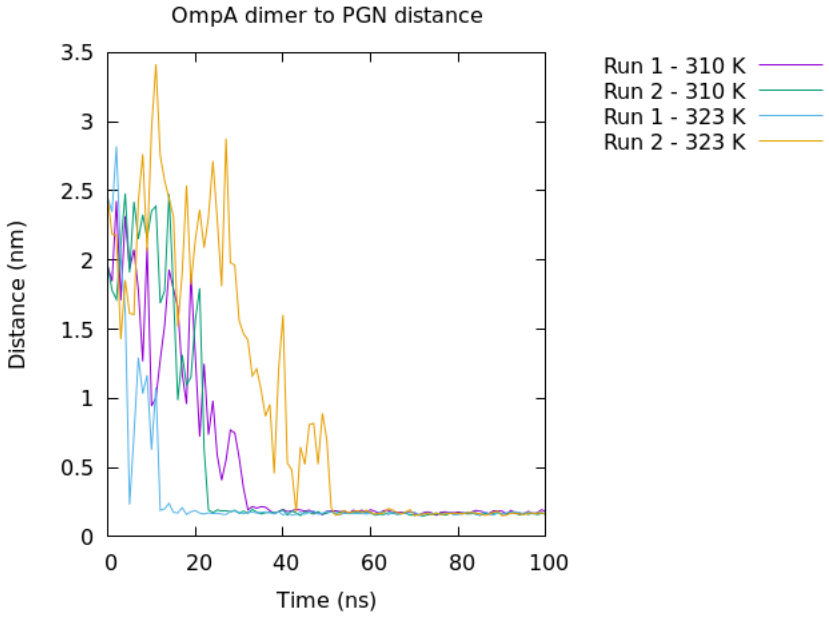


Figure 3.2: The minimum distance between the OmpA dimer and the PGN mesh over 100 ns, at 310 and 323 K.

Firstly, across all simulations at 310 and 323 K, the minimum distance between the dimer and the cell wall was seen to reduce, Figure 3.2, until the OmpA dimer was in physical contact with the cell wall. This can be seen in the reduction of distance of ∼2 nm, to ∼0.2 nm during all four runs. This occurred over varying timescales during 100 ns, however generally the OmpA dimer was bound, below 0.5 nm distance, to the cell wall after 50 ns of simulation.

It is noteworthy that once the binding occurred, there was no dissociation of the cell wall and OmpA. Compared to previous simulations of OmpA dimers and strands of cell wall, where the OmpA dimer associated with both PGN strands,

but the linker region between the barrel and periplasmic clamp contracted.[18] Figure 3.1 shows that the linker regions remained extended whilst the clamp contacted the cell wall.

This PGN-binding occurred either as a result of the movement of the OmpA dimer periplasmic domain, or because of the cell wall moving through the periplasm. The binding is proposed, via crystallographic studies, to involve two key residues, ASP241 and ARG256 that exist in the periplasmic clamp. Considering the negatively charged nature of the PGN, it follows that charged interactions would then govern the binding of the dimer. These two residues are located deep in the periplasmic domain, indicating that initial contact between the protein and PGN is not mediated by this binding region.

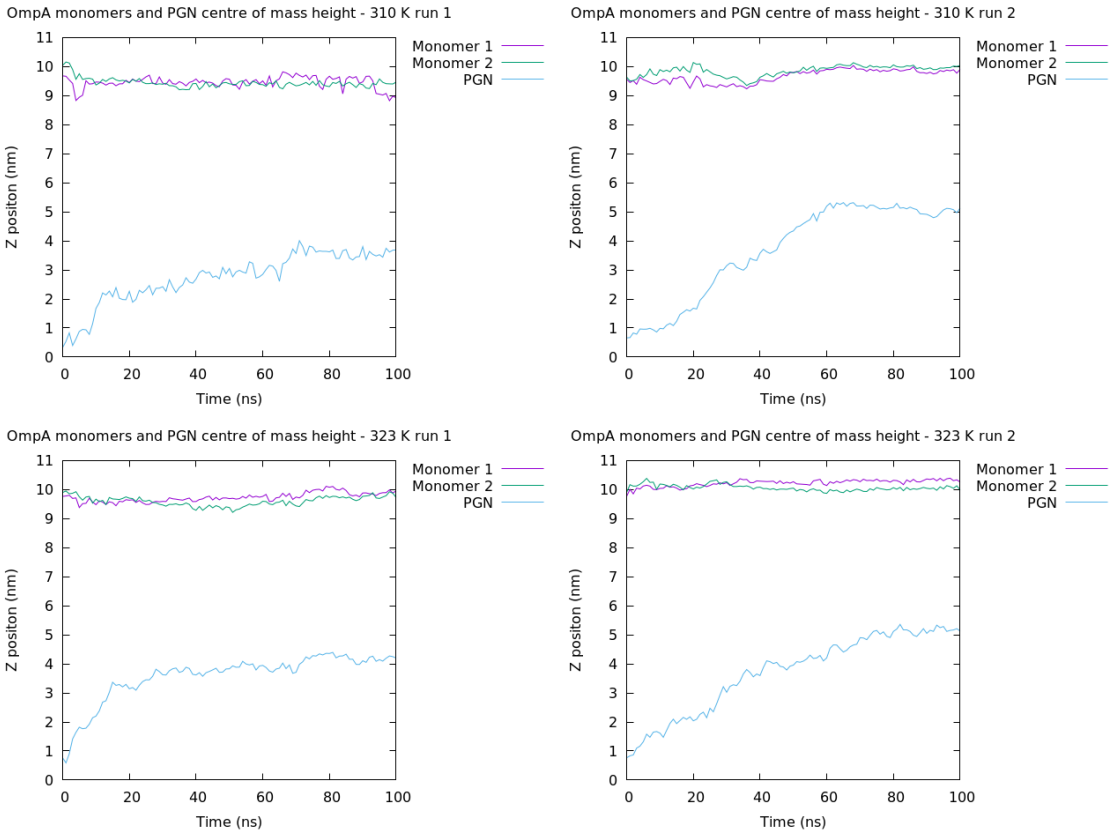


Figure 3.3: Centre of mass measurement in the z-axis of each OmpA protomer and the cell wall.

Measurement of the centres of mass of the OmpA monomers and PGN in

the axis perpendicular to the cell wall (Z), shows the initial positions of the OmpA proteins did not change much in this axis.This is in contrast to the cell wall. Initially OmpA and PGN were separated by a $\sim$ 3 nm gap, Figure 3.3. It can be seen that the cell wall during the simulation moved between 3 nm to 4 nm closer to the monomers in the Z direction.

This movement can be ascribed to either attraction between the cell wall and OmpA, or by another physical effect. BLP can bend and tilt, Figure 3.1 shows this visually whilst Figure 3.10 confirms this, in contrast to most representations showing the BLP at a 90 degree perpendicular angle to the cell wall. It is this natural bending of the protein, which pulls the cell wall upwards, reducing the distance between the OM and the PGN. This is evidence to support that BLP mainly functions as a structural staple in *E. coli*, however this lends evidence to the flexibility and mobility of the protein, rather than a rigid structure.

This reduction in distance caused the initial interaction between OmpA and PGN in these systems.

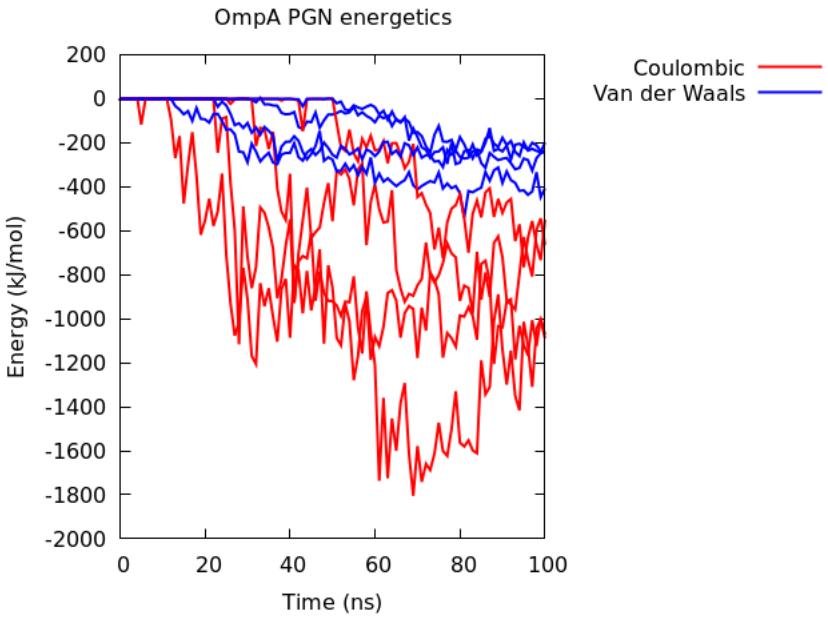


Figure 3.4: The short range electrostatic and Van der Waals interactions between the OmpA dimer and PGN over 100 ns.

The initial interaction as OmpA approached the cell wall can be seen in Figure 3.4, where across all simulations it can be seen that the stabilising interaction between the two structures was due to charged interactions. Van der

Waals interactions played a role in stabilising this binding, however at approximately one order of magnitude less than the coulombic energies.

As both structures exist in the aqueous periplasm, it is to be expected that charged interactions would would be key to cell wall binding. Given the negative charge of the cell wall, it is proposed that there are complementary positively charged residues on OmpA that allowed for strong binding. This is backed up by the proposed binding residues, which can interact with the side chain of PGN. The NAM monomer side chain, containing alanine, glutamic acid and meso-diaminopimelic acid, is similar in composition to amino acids, indicating that it would require a zwitterionic type binding site.

The Van der Waals interaction can generally be assumed to be $\sim$ -200kJ/mol, once the protein was bound, as the values were similar during each run. As the flexible clamp and cell wall are both dynamic regions it is not surprising that the charged interactions fluctuated significantly. This can be seen where in two runs this energy was $\sim$ -600kJ/mol after 100 ns, whilst in the others was $\sim$ -1000kJ/mol, this can be attributed varying area of the OmpA surface bound to PGN. It is possible that the clamp was either tilted, or interacting with other components of the system such as BLP.

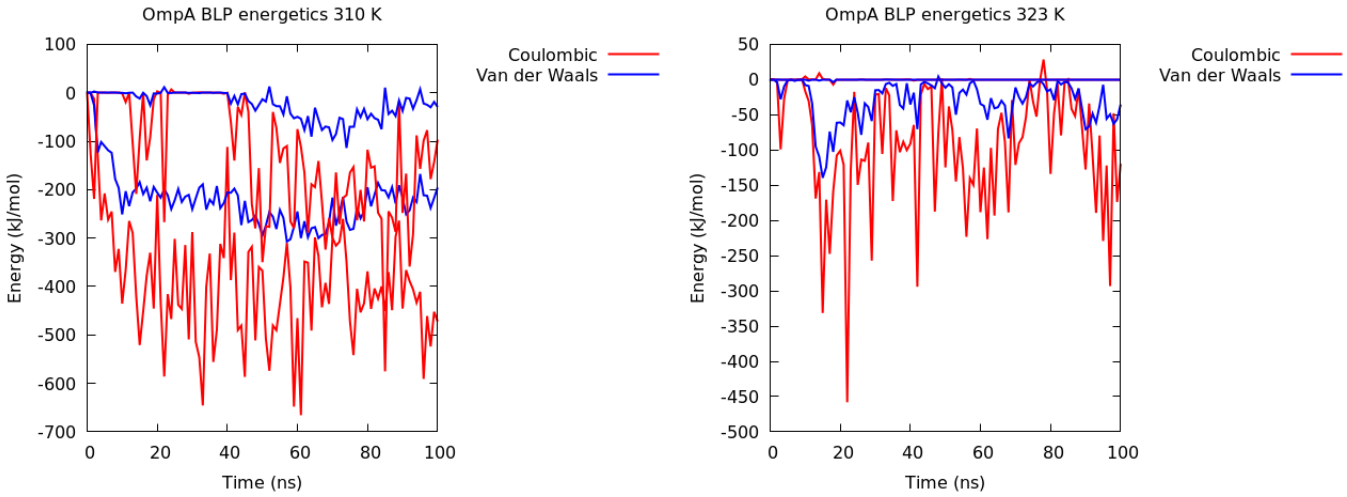


Figure 3.5: The short range electrostatic and Van der Waals interactions between the OmpA dimer and BLP trimer over 100 ns, 310 K (left), 323 K (right).

The dimer interacted with BLP, another charged structure in the simulation, to varying degrees. This was a weaker interaction compared to PGN, however it was still generally attractive, as seen in Figure 3.5.

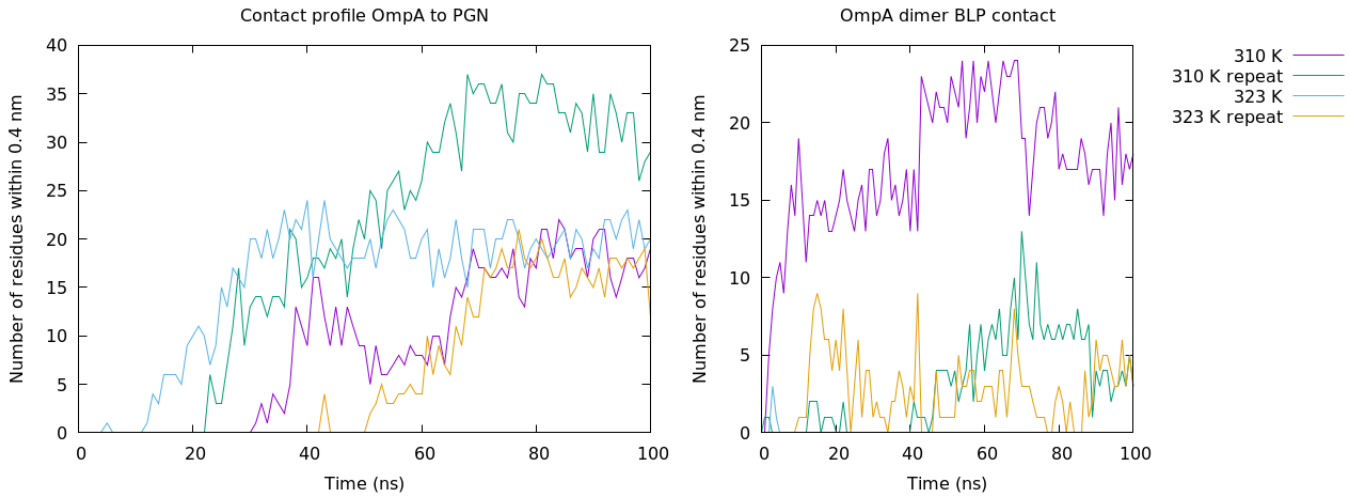


Figure 3.6: Number of OmpA residues contacting PGN over 100 ns (left), number of OmpA residues contacting the BLP trimer over 100 ns (right).

Given the starting position of OmpA in the systems, it is possible that OmpA used BLP as a ladder; where as the cell wall moved, OmpA remained close to BLP until the cell wall was within range for binding. Unlike with the cell wall where physically and energetically association is strongest after 50 ns, Figure 3.6 shows that OmpA and BLP interacted almost immediately. This allows for the hypothesis that BLP is not only a structural staple, but also a point of interaction for OmpA and also many other proteins that exist in the crowded periplasm.

Figure 3.5 indicates for systems at 310 K this interaction was stronger and more consistent, where lower energies (-500 kJ/mol) were reached and maintained by the end of the simulation. In this scenario OmpA was bound to both the BLP and the cell wall, causing lower binding energies to the PGN, as some of the OmpA binding sites were bound to the amino acids in BLP. In systems at 323 K the interaction was weaker, where the energy was approximately 0 for one simulation indicating no interaction, whilst one system showed a lesser interaction (-150kJ/mol) with respect to coulombic energies as seen in Figure 3.5.

This dimer-BLP interaction shows that the separate periplasmic components interacted synergystically in order to maintain local protein binding.

Assessing this amount of protein to PGN contact reveals that the OmpA dimer in three runs had a similar level of contact to the PGN after 100 ns, of approximately 20 unique residues. One simulation shows a higher number at

30 residues, and whilst there was fluctuation there was no significant change in the contact with the cell wall after 80 ns of simulation in all systems.

Analysis of protein to BLP contacts echoes the energetics data, showing that the total number of contacts was generally lower than for PGN, Figure 3.6. This is due to there being a lower surface area for contact, and also the BLP is a less favourable site for OmpA dimer binding sterically. In once instance there was essentially no contact at all over 100 ns, whilst two systems fluctuated between 5 and 10 contacts during a simulation, once again seen in Figure 3.6.

Comparing these two contacts, it can be seen that large increases in the contact to the cell wall were mirrored by decreases in contact to the BLP. With respect to the simulation at 323 K however, it can be seen that contact and interaction between the OmpA dimer and the BLP was not necessary for binding to the cell wall, where in fact this system established the highest amount of contact with the PGN earliest, and then maintained a consistent contact of close to 20 residues. Once this PGN contact was established by 40 ns there was no large change however, which can be attributed to the lack of interference from the BLP trimer, seen in Figure 3.6.

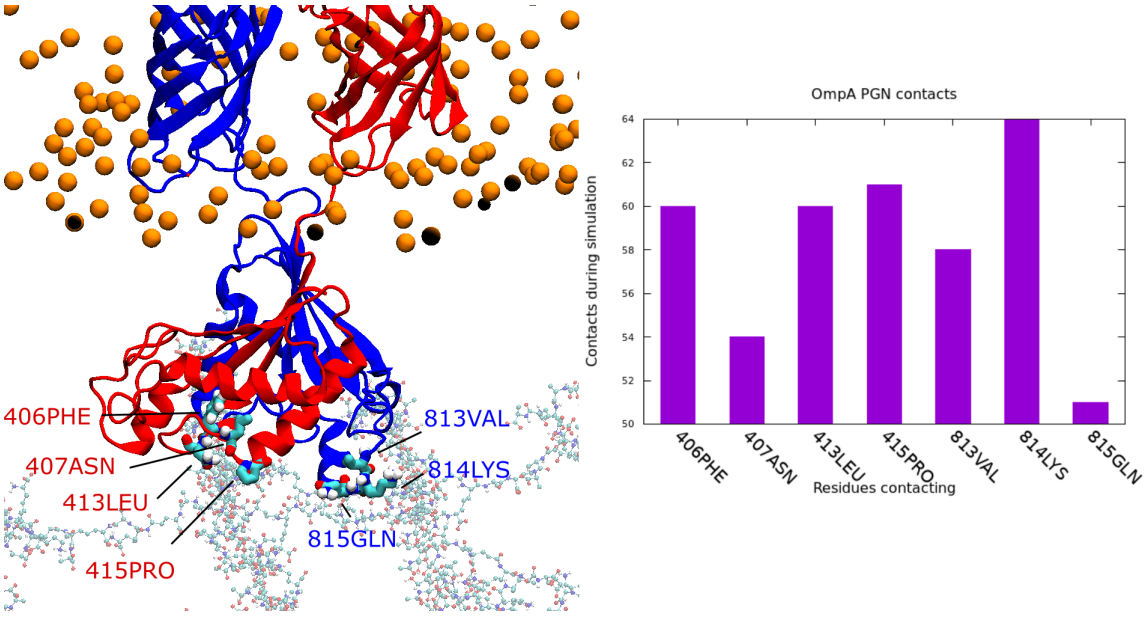


Figure 3.7: A visualisation of the periplasmic clamp of the OmpA dimer to PGN binding in the unstructured region at the edge of the clamp where maximum contact was observed. The orientation the PGN is parallel to the outer membrane, where the periplasmic domains of OmpA contact the flat PGN surface. Separate OmpA protomers are shown in red and blue, with similarly labelled residues. The cell wall is shown transparent, in the CPK representation, whilst the headgroups of the membrane are shown as orange spheres (left). Per residue contact of the PGN binding site of the OmpA over 100 ns, where a LYS residue has the highest contact (right).

After establishing the amount of OmpA-PGN contact, the specificity of the contact comes into question. Fig 3.7 shows that the contact between the residues was highest for a LYS residue. This residue was in contact for 64 percent of the simulation, where it is located at the edge of the periplasmic domain as part of an unstructured regions between two exposed surface helices, as are other residues highlighted in red that had the highest contact. With respect to these there are two separate sites, where one is composed of a VAL-LYS-GLN pocket and the other a PHE-ASN-LEU-PRO site. These highlighted residues form grooves with which the side chain of the PGN monomer can interact, specifically the fourth m-DAP residue in the chain, which carries a charge. This initial interaction between these edge grooves on the clamp and flexible side chains allowed for the further formation of more contacts throughout the simulations.

This charged interaction, along with surrounding hydrophobic residues that prefer to bind the cell wall allowed for the initial stabilisation of binding, and

continued once OmpA is bound to PGN. When dimerised these sites are located at the edge of the periplasmic domains of each monomer, whilst the central region of the OmpA monomers is self associated, indicating that theoretically the dimerisation allowed the formation of this first contact.

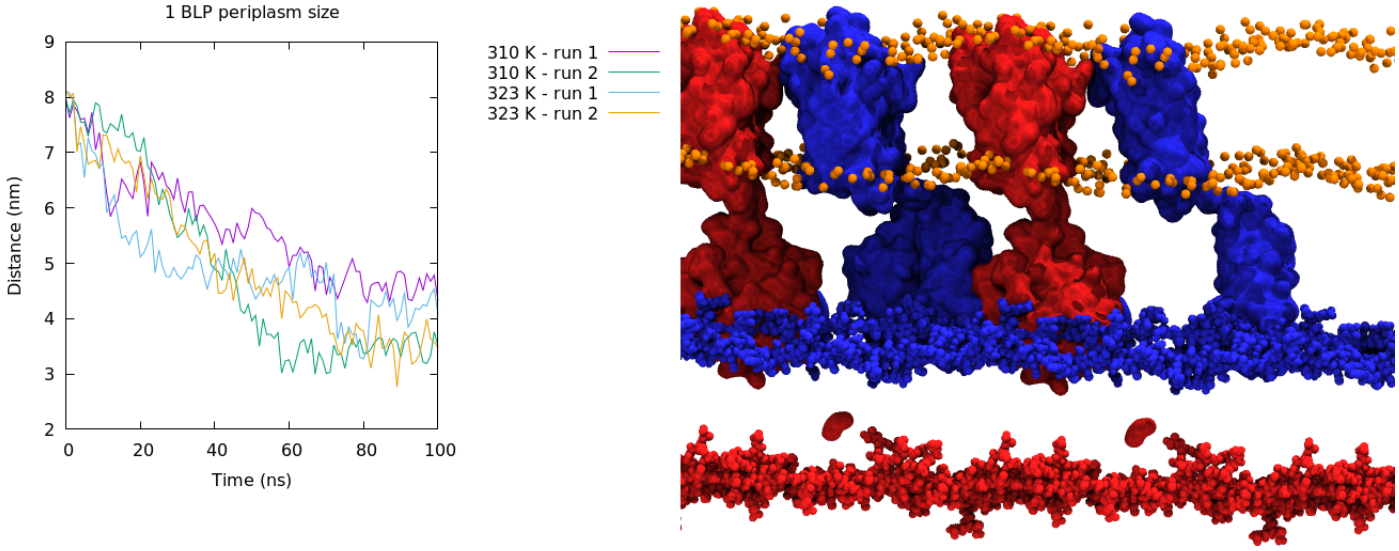


Figure 3.8: The OM to PGN distance, inner leaflet headgroups to PGN, over 100 ns at 310 and 323 K (left). A visualisation of the OmpA dimer and PGN at 0 ns (red) and 100 ns (blue) where the orange spheres represent the membrane headgroups, OmpA and PGN coloured with respect to time. One protein dimer is present in the visualisation, where periodic imaging effects show multiple proteins.

Systems using various amounts of BLP were constructed. Using the main case where one BLP was included it is possible to see the effects on the size of the periplasm. This is achieved by measuring the distance between the inner leaflet of the OM and the PGN mesh. Over four systems it can be seen that this periplasmic distance reduced. Generally there was a 50 % reduction in the size of the periplasm. This can be seen displayed visually in Fig 3.8, where at 0 ns the cell wall was not contacting OmpA and is further from the membrane in the perpendicular axis, whilst at 100 ns it was contacting . It is once again proposed bending of the BLP caused a smaller periplasmic size than the length of straight BLP ($\sim$ 90 Å).

Whilst BLP can pull the cell wall higher, and allow OmpA to interact, it also provides a minimum distance that the cell wall is from the OM. Compared

to previous simulations using free strands in place of PGN meshes with BLP it was been observed that the cell wall can contact the membrane. Including one BLP trimer gives a periplasm that after 100 ns was 3.8 - 4.8 nm thick, however smaller distances were not possible due to the BLP forcing a certain periplasmic volume to exist.

BLP bending can be seen with one BLP trimer, Figure 3.10, but in *E. coli* exists as a macroscopic effect, where it is unlikely that the cell wall to be at the same height throughout the cell, it is probable that there are sites around the bacterium that have higher and lower concentrations in order to facilitate difference in cell wall positions in the periplasm.

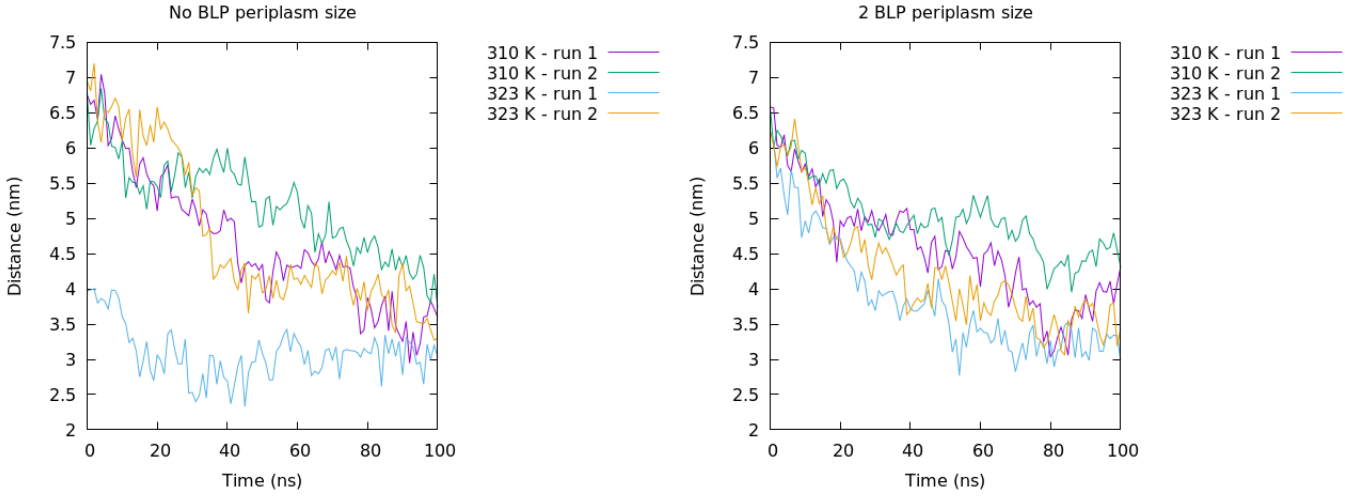


Figure 3.9: The OM to PGN distance over 100 ns when BLP is absent from the system at 310 and 323 K (left), The OM to PGN distance over 100 ns, when 2 BLPs are covalently attached to PGN, at 310 and 323 K (right).

Comparatively systems that contain no BLP reduced in periplasmic size slower than the two BLP system, Figure 3.9, where one system at 310 K displayed a final size of ∼3 nm. Generally the periplasmic size when BLP was absent converges to a lower value, where the largest value was ∼4 nm. This is an indication that the cell wall was still attracted to the OmpA dimer, but there was a significant difference in this distance, as OmpA was now the deciding factor in establishing this distance.

Similar values were seen when comparing the BLP systems, where similar periplasmic sizes were seen. In the 2 BLP systems a range of final sizes, 3.3 - 4.5 nm is seen in Figure 3.9, which was lower than that of 1 BLP but higher than none. Interestingly this could indicate a trend that increasing the num-

ber of BLP proteins bound to the cell wall will in turn reduces the periplasmic size, considering that when BLP is absent the cell wall can be located anywhere in the periplasm. Overall this is an indicator that the cell wall position in the periplasm is very malleable, and that varying the amount of BLP will have an effect on where the cell wall lies and in turn what it was able to interact with.

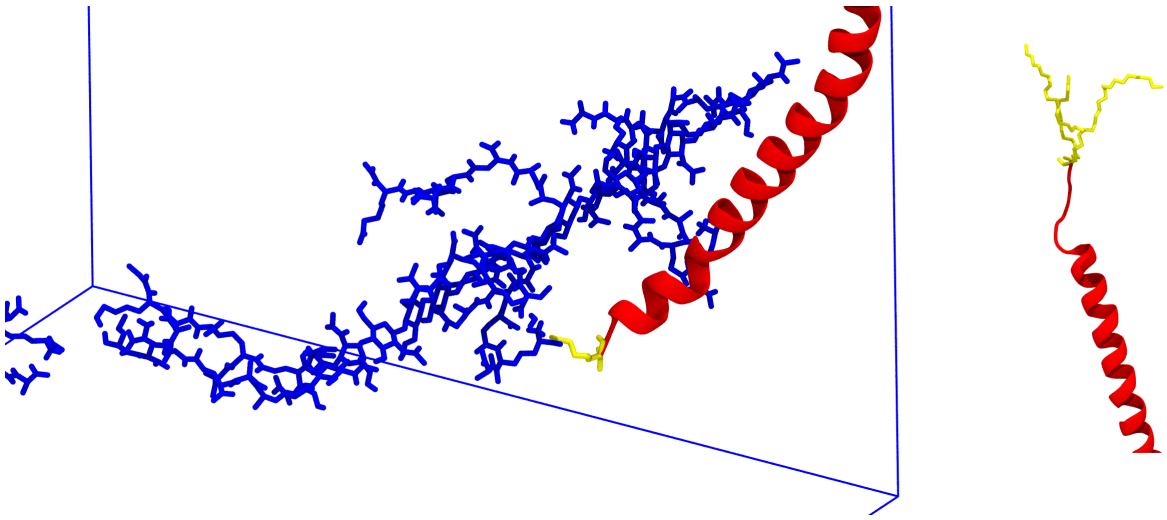


Figure 3.10: A snapshot of the BLP to PGN peptide connection. BLP is shown as a red cartoon, PGN as blue licorice, the terminal lysine bonded to PGN is shown in yellow licorice . The box is displayed to show the bounds of the system, which for simplicity shows only one BLP monomer and a single strand of PGN (left). A snapshot of BLP with the lipid functionalisation, BLP is shown as a red cartoon and lipid as yellow licorice (right). The respective chemical structures can be found in Figures 1.4 and 1.5.

A primary physical attribute of the BLP that controls the size of the periplasm was the tilt angle with respect to the OM bilayer. Tilting and bending seems to be the key aspect of BLP behaviour in the cell, Figure 3.11. It would appear that changing the number of BLPs affects that rate at which the cell wall is pulled closer to the membrane, however that periplasmic size is roughly similar after 100 ns in all three systems.

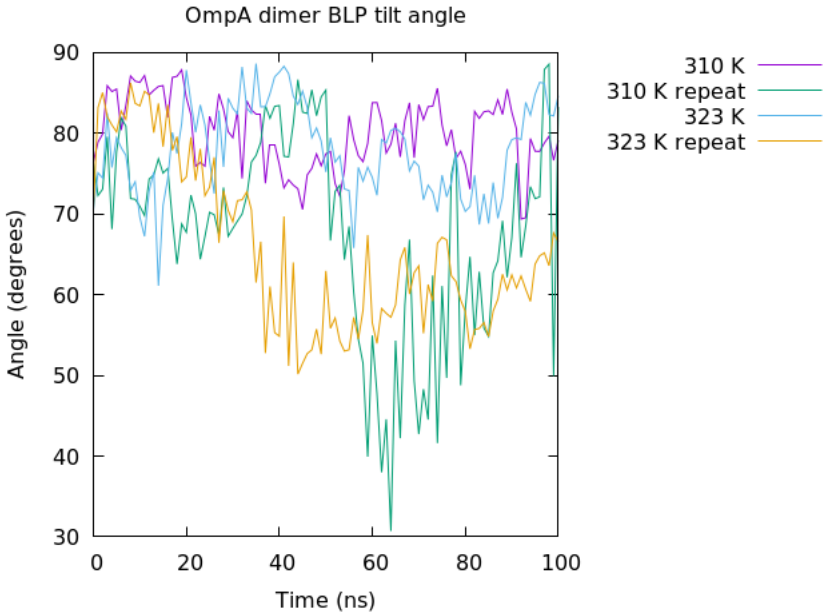


Figure 3.11: The tilt angle of the vector of BLP against the surface of the inner leaflet of the OM over 100 ns at 310 and 323 K.

BLP tilting was characterised across systems with 1 BLP. This tilting is measured via establishing the direction of a vector from the end to end in BLP; the angle of this vector is then measured with respect to the headgroups in the OM inner leaflet.

This data showed that the BLP angle can vary vastly in each simulation. In one instance the angle of the BLP (310 K repeat) was such that it reaches 30 degrees, indicating that the lipoprotein was lying at an angle that allowed the cell wall to be pulled very close to the OM. All systems start at a close to perpendicular angle, where in two systems this angle fluctuated, however reductions were seen in the other two, that last throughout the simulation. Referencing earlier contact data, Figure 3.6, the repeat run at 310 K was the simulation that revealed the highest degree of contact with OmpA. This supports the theory that the bending of BLP facilitated this contact, as the other three systems showed a less severe angle change in the BLP with respect to the OM. It is also interesting to note that the BLP trimer that bent the least had the highest OmpA-BLP contact and the largest periplasmic size. This gives credence to the idea that OmpA interacting with BLP affected the bending and tilting motion, which in turn affected the cell wall positioning.

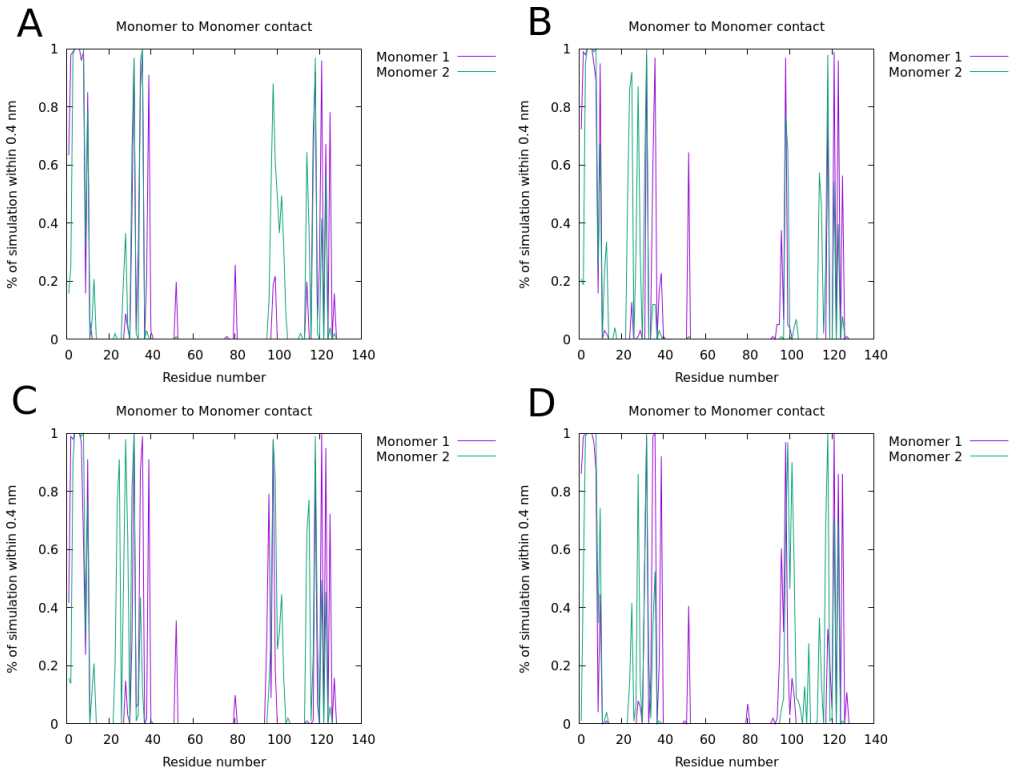


Figure 3.12: The contact between the protomers of the OmpA dimer over each independent repeat, at 310 K (A and B) and 323 K (C and D).

Before considering the monomeric data it is wise to consider the behaviour of the OmpA dimer with respect to the individual monomers. Shown here is the lifetime of monomer to monomer contacts.

It can be seen in the barrel residues (residue number less than 20) that the contact was consistent and that the association membrane allowed these residues to stay firmly in contact over the simulation, in Figure 3.12. In the periplasm however this contact varied, where one would expect to see two symmetrical peaks for the periplasmic domains if the clamp existed as one solid unit.

This was not the case, as these peaks that showed periplasmic domain contacts are not consistent across separate simulations. This indicates that the dimeric clamp is extremely flexible, to the point that, whilst not entirely dissociating, binding to the cell wall weakened the association between the OmpA monomers. This can be seen in the fall in certain peaks in the lifetime contact, such as at residue number 100, and also the reduction in width in others, indicating that during the simulation certain contacts were not maintained for

similar timescales, in Figure 3.12.

3.4.2 OmpA monomer 310 & 323 K

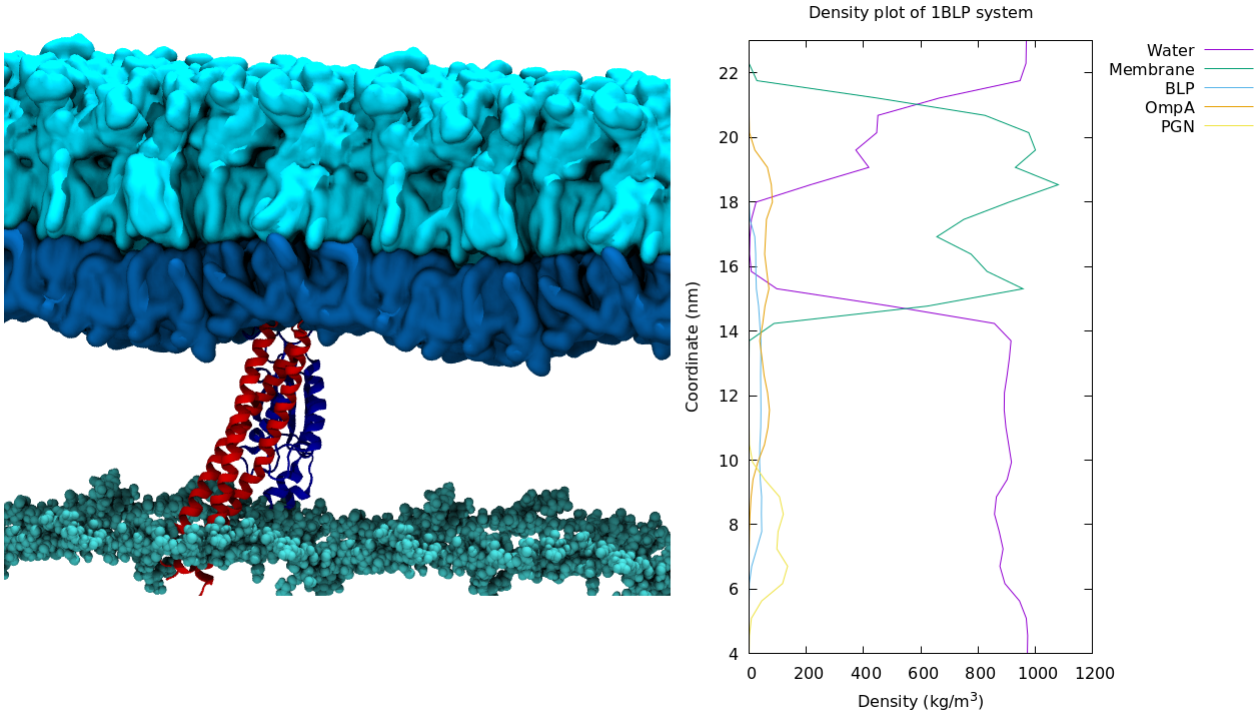


Figure 3.13: A visualisation of the OmpA monomer (blue cartoon), BLP (red cartoon), PGN (cyan spheres, Membrane (inner leaflet dark blue surface, outer leaflet light blue surface) periplasmic system (left). The corresponding density of the components in the z-axis over 100 ns (right).

Analysis of the OmpA monomeric system has been carried out in a similar manner to the dimeric system, in order to compare certain properties, but also to explore individual properties of the monomer. The monomer displayed different characteristics compared to the OmpA dimer. Previous work has shown that there is a difference in contact with PGN strands.[18]

This can be attributed to a number of factors, that include a greater degree of flexibility, greater exposure to solvent and a lower surface area for binding to the cell wall. It was seen in monomeric simulations that it is also possible for OmpA to contact the cell wall. Visualisation of this shows the BLP once again

at a tilted angle, in Figure 3.13. Previous simulations showed the OmpA CTD contracting to contact the inner leaflet of the OM.[18] In the presence of the PGN mesh this did not occur, as the contraction of the linker region in OmpA cannot significantly affect the position of the PGN in the periplasm in the simulation systems, see Figure 3.14.

A density plot of a typical system, for OmpA monomer and one BLP, shows a similar result to the OmpA dimer system, where the densities for OmpA, BLP and PGN overlapped, indicating they occupied the same space in the Z dimension, in Figure 3.13. It can also be seen that the BLP overlapped with the OM, where the lipidated N-terminus rested in the inner leaflet. The membrane forms a barrier where there was little solvent present, where any water was due to that permeating the OmpA barrel.

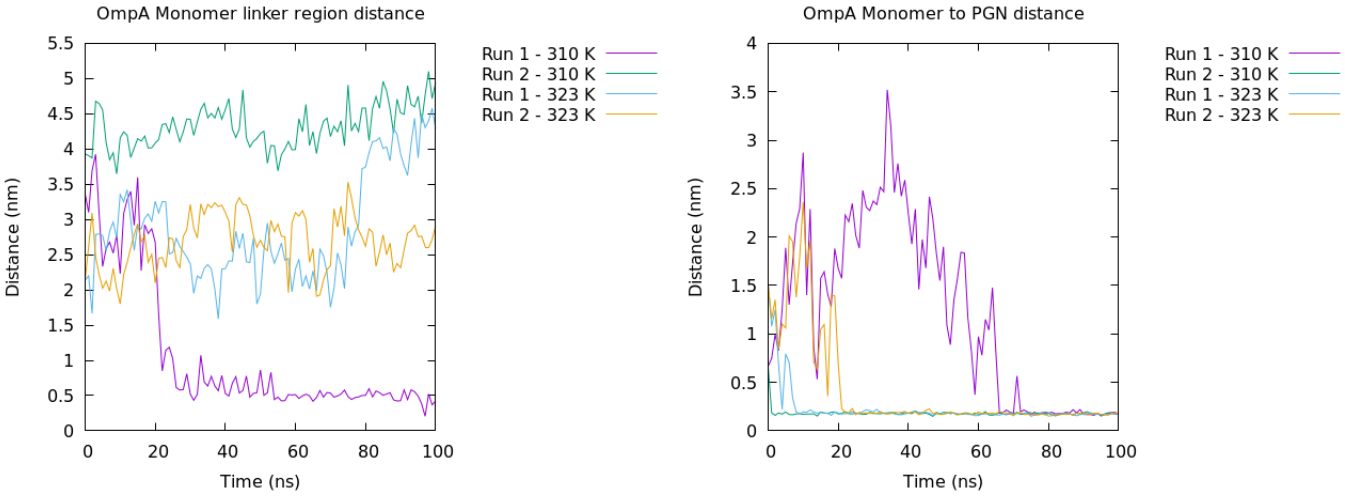


Figure 3.14: The minimum distance between the start and end of the flexible linker region in OmpA over 100 ns, at 310 and 323 K (left). The minimum distance between the OmpA monomer and PGN over 100 ns, at 310 and 323 K (right).

Monitoring the linker region distance, end to end, shows that the OmpA monomer linker did not behave consistently during simulation. After 100 ns distances of ∼5 nm are seen for two simulations, whilst one distance of ∼2.5 nm is seen and a much lower distance of ∼0.5 nm for another. Distances of less than 1 nm would indicate that the linker region has contracted, and that the OmpA monomer was associated to the membrane. This would indicate that in three simulations there were various extended OmpA monomers and one contracted monomer. This contraction can be see in Appendix Figure 7.4

If during simulation these linker regions contract, is PGN contact still established? It can be seen that the minimum distance from OmpA to PGN indicates that there was PGN contact. Comparatively in all but one system OmpA was contacting (distance less than 0.4 nm) PGN after roughly 20 ns, where the outlying system eventually shows PGN contact after 60 ns, Figure 3.14. Considering that the system in which OmpA established PGN contact last was also the contracted linker system, this shows that the monomer did not extend towards the PGN in the same fashion as the dimer.

The variation in the linker distance indicates how quickly the PGN contact was established, where run 2 at 310 K shows that when the linker was extended for the whole simulation, the PGN contact was quickly established and maintained. In all systems when PGN contact was established, the monomer did not unbind from the PGN.

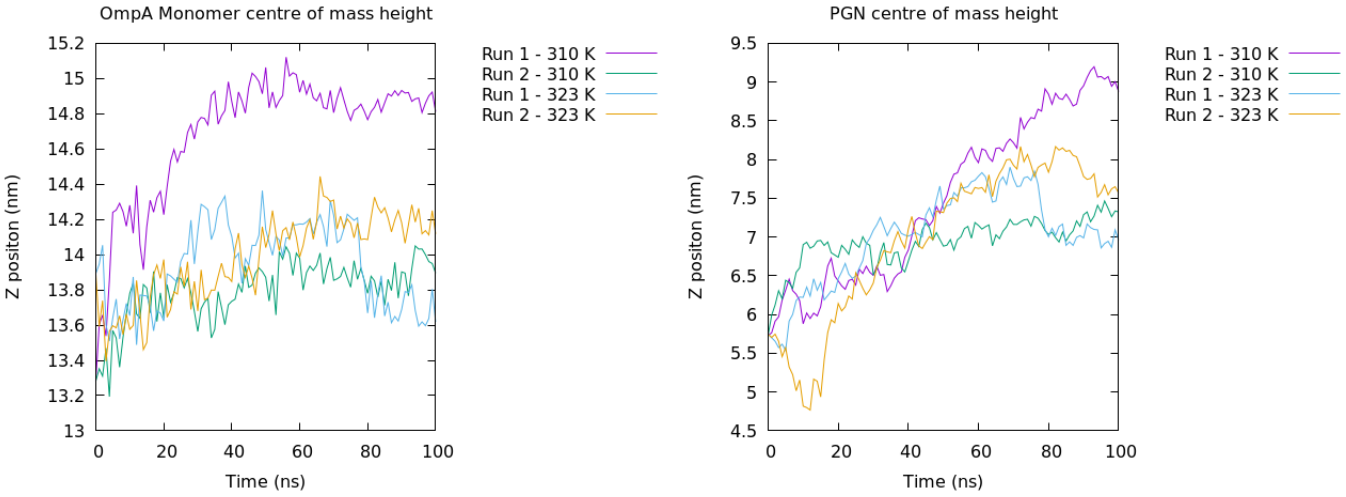


Figure 3.15: The centre of mass measurement of the OmpA monomer in the z-axis over 100 ns, at 310 and 323 K (left). The centre of mass measurement of PGN in the z-axis over 100 ns at 310 and 323 K (right).

The centre of mass in the Z dimension for the OmpA monomer shows that the centre of mass is similar for three of four simulations. The run with the highest centre of mass is consistent with the shortest linker region. The height of the monomer COM for this run was roughly 1 nm higher than the other monomers. The other systems all achieved a similar value within a 0.5 nm range for the centre of mass after 100 ns, see Figure 3.15.

If OmpA contracts then how does it contact PGN? Once again the cell wall moved in the Z dimension during the simulation, as a result of BLP, Figure 3.15.

It can be seen that the PGN COM correlates with the OmpA COM, where the monomeric clamps that rest higher in the periplasm caused a resulting higher PGN layer. As a result of this movement the PGN moved at least 1.5 nm higher in the periplasm, with a maximum increase of close to 3 nm in one instance.

Interestingly this showed that the PGN is free to move large distances over short periods of time in the periplasm, and that it will continue to move even after the monomer binds to the cell wall.

The deviation in centre of mass height was similar for the cell wall when compared to the dimer. However, relative to the individual monomers in the dimer, a higher deviation was seen with respect to the monomers.

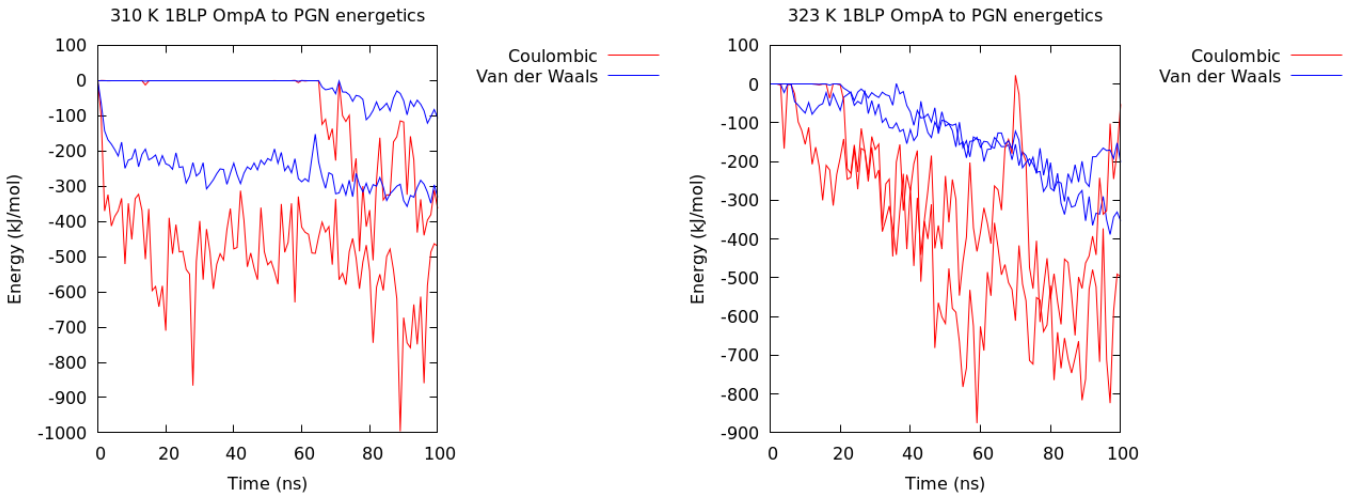


Figure 3.16: The short range electrostatic and Lennard-Jones interactions between OmpA and PGN over 100 ns, at 310 K (left) and 323 K (right).

The energetics of the monomer to PGN contacts showed variation when compared to the dimer. All systems once again showed a stabilising interaction between the two structures. As expected, charged interactions dominated with the monomer, however the largest stabilising coulombic interaction ($\sim$ -1000 kJ/mol) was less than the largest stabilising force in the dimeric systems. The Van der Waals force that occurred between the protein and the PGN generally reached -300 kJ/mol as binding was established. This was similar to the dimeric systems, indicating that the VdW force aided the binding of the protein to PGN but once again only makes up a small proportion of the attractive energy.

At 310 K, once binding began to occur at 80 ns in one system it could be seen to reach a similar coulombic value of -300 kJ/mol compared to the system

that immediately binds, indicating the delayed binding is not weaker but was not fully established after 100 ns. At 323 K, both monomers were bound after 20 ns, where it was seen that the energy is of a similar value as binding at 310 K, Figure 3.16. The energies at both temperatures can be seen to fluctuate largely during simulation, due to the natural movement of the protein during simulation.

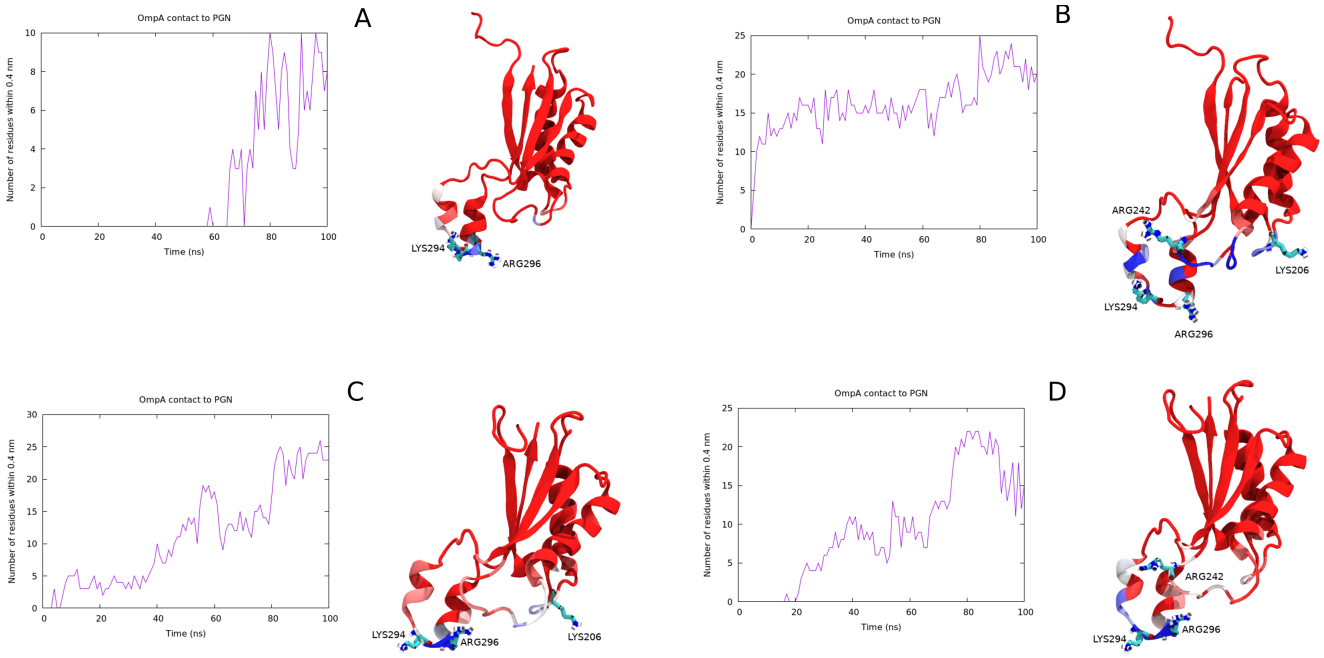


Figure 3.17: The OmpA monomer to PGN unique residue contact, with protein coloured to represent contact using the BWR scheme, where blue is highest contact and red is lowest over 100 ns. Simulations at 310 K (A and B) and 323 K (C and D).

As contact was established in all four simulations, the specificity of this contact was examined for the monomeric systems. It was seen that in all four systems this contact fluctuated but generally trended upwards as the simulation progressed.

With respect to the simulation that did not bind until much later, it was seen that the edge loop of the monomer formed most of the contact, similar to the dimer, in comparison to the other systems where there were two points of contact in a second adjacent loop region that is located deeper in the clamp, Figure 3.17.

It was seen that the specific residues LYS294 and ARG298 were consistently responsible for the contact between the monomer and the PGN. These residues are located in the flexible external loop on the edge of the CTD that can contact the PGN first. This contact was orientation dependent, where if both points of contact are able to touch the cell wall then binding was established more quickly, and according to coulombic energies, more strongly, Figure 3.16.

It was hypothesised that these positively charged residues that are present on the unstructured loop at the edge of OmpA CTD were crucial for establishing initial contact and maintaining this contact between the monomer and the cell wall.

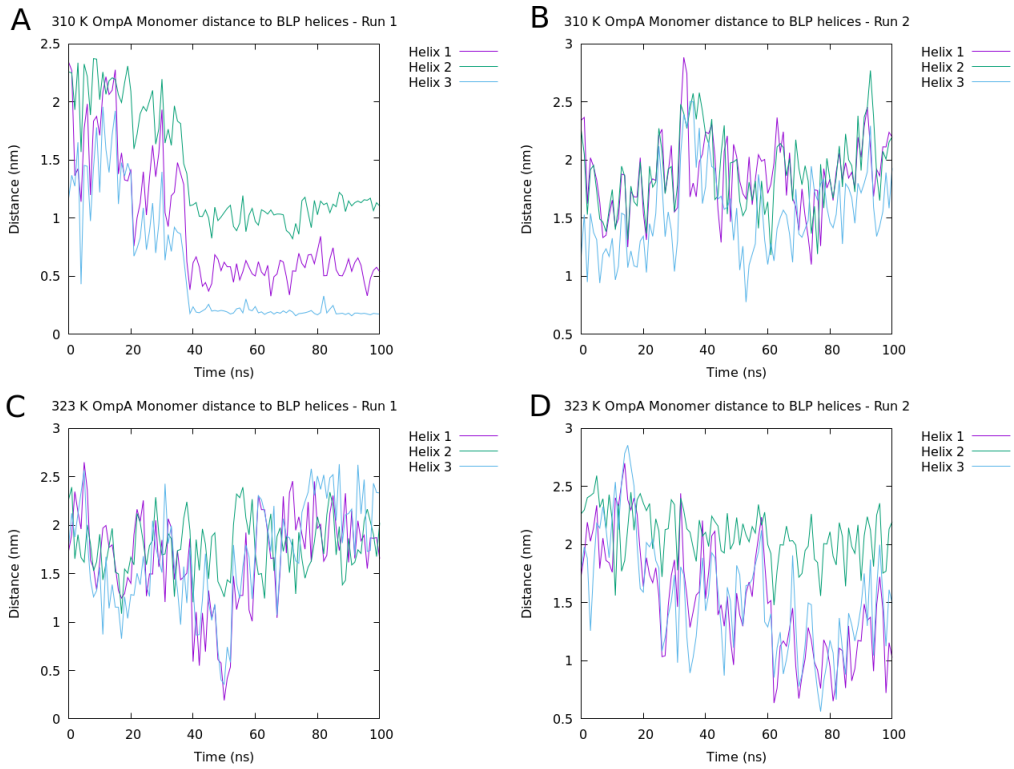


Figure 3.18: The minimum distance between OmpA monomer and BLP individual helices in the each system over 100 ns, 310 K (A and B), 323 K (C and D).

Across all four simulations a varying distance was seen between each OmpA monomer and the helices of the BLP trimer. Generally the monomer was not in contact with the BLP, however in one instance after 40 ns the OmpA monomer (run 1 310 K) was seen to be bound to BLP, specifically to one helix of BLP,

helix 3, Figure 3.18.

In this instance PGN binding to OmpA did not occur until after 80 ns, but the binding to the BLP was established after 40 ns. OmpA CTD in this instance was free to fluctuate in solution, where it did not contact the cell wall, but once it was bound to BLP, it contacted the cell wall using the BLP as a framework to bind until the cell wall had approached close enough for PGN binding.

Considering that in this system OmpA was bound to both BLP and PGN, this explained the slower rate of binding and less favourable energetics. Comparing the data in Figure 3.15 to this indicated that the centre of mass in z of the OmpA monomer stopped increasing after 40 ns, indicating that the BLP served as an anchor for the OmpA. Brief contact occurred in run 1 at 323 K, however it was not consistently maintained.

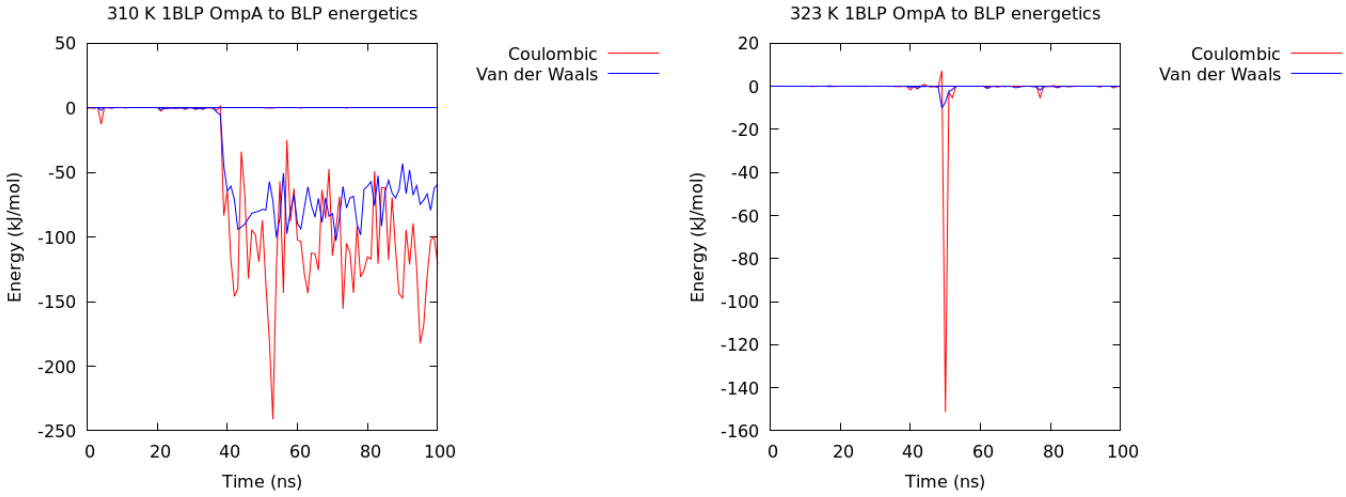


Figure 3.19: The short range electrostatic and Lennard-Jones interactions between the OmpA monomer and BLP over 100 ns, 310 K (left) and 323 K (right).

Energetics of this binding showed that as a majority BLP and OmpA did not interact for significant lengths of time. This however was not the case in the run where OmpA-BLP binding was observed. In this case, once again run 1 310 K, there was a consistent attractive force observed between the OmpA and BLP. This interaction was of a much lower magnitude than OmpA-PGN. Considering that after 100 ns the coulombic energy was close to -100 kJ/mol and the VdW energy was close to -50 kJ/mol it can be assessed that the OmpA-BLP binding was not as strong as OmpA-PGN binding.

However, it is clear that this binding event was strong enough to halt the

free movement of the OmpA monomer in the periplasm, such that eventually the cell wall did bind to it. BLP was serving not only as a structure that allows the OM and cell wall to remain associated, but also as a pillar that surrounding proteins cling to. When the periplasm is crowded, it would effectively allow charged interfaces of proteins to bind to it favourably, facilitating the packing of these proteins around the periplasm.

It is estimated that there are 100,000 OmpA proteins present in the OM of *E. coli*, whilst BLP is the most abundant source of protein in the cell. It is therefore logical to assume that there is interaction between either dimeric or monomeric OmpA consistently occurs in the bacterial cell *in vivo*. It is also key to note that despite the binding of other structures and bending and tilting of BLP during these processes, the trimer remained stable.

3.4.3 Effect of changing BLP content

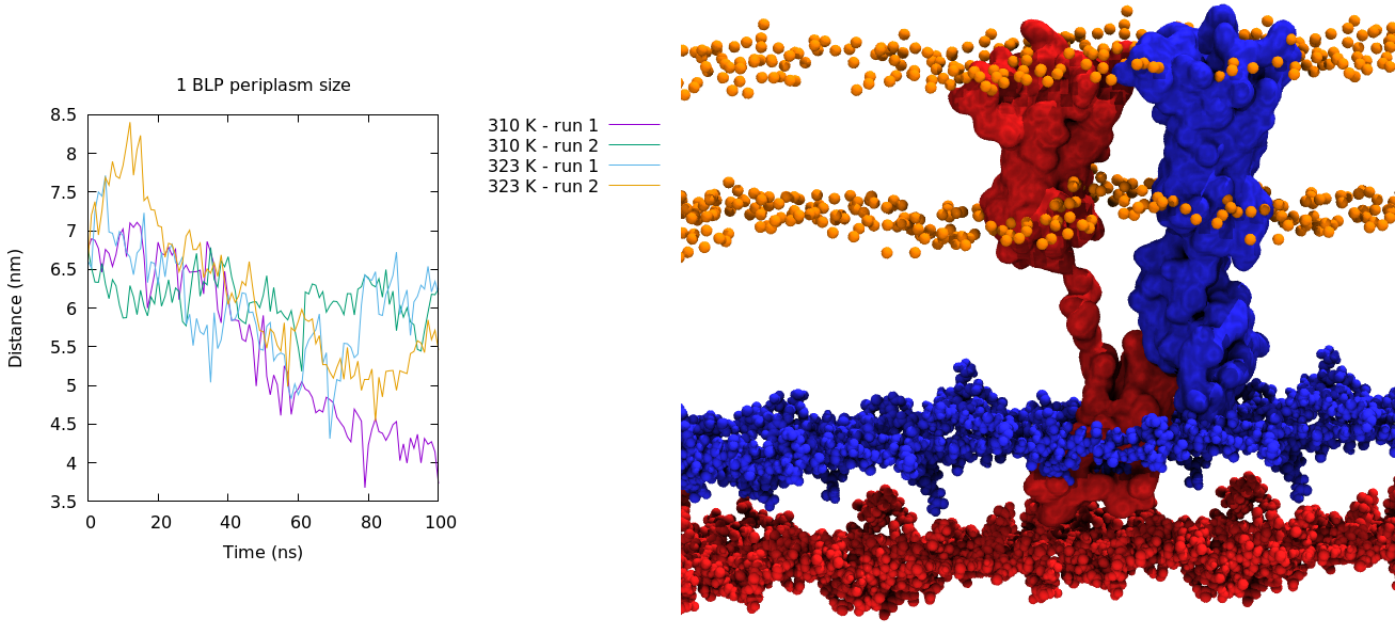


Figure 3.20: The OM to PGN distance over 100 ns at 310 and 323 K (left). A visualisation of the OmpA monomer and PGN at 0 ns (red) and 100 ns (blue) with membrane headgroups as orange spheres and OmpA and PGN coloured according to time.

Relative to the dimer, a lesser amount of protein was present in the periplasm in the monomeric simulations. The periplasm in these systems also shrinks. There was a reduction from 7 nm to $\sim$ 6 nm in three systems, and in the other a reduction to 4 nm. This follows that the BLP tilt and bend has once again produced a smaller periplasm. Generally these periplasmic sizes were larger than the dimeric counterparts. It follows that the extension of the monomer allowed it to contact the cell wall, meaning that the periplasm was larger, however the dimeric systems could extend or contract also.

The dimer compared to the monomer is a larger periplasmic clamp. This clamp bound more strongly to the cell wall and for longer lifetimes. This was a result of a greater surface interface with the PGN. However due to increased size, the dimers were shown to interact more often and more strongly to the BLP, as seen previously.

Comparing the periplasmic size trend with the dimeric 1 BLP system it can clearly be seen that the combination of the OmpA and the BLP controlled the size of the periplasm in both the dimeric and monomeric systems. This can be seen visually where the OmpA was close to the cell wall at both 0 and 100 ns in Figure 3.20, but regardless of this the periplasm still shrank, due to the physical effects of BLP.

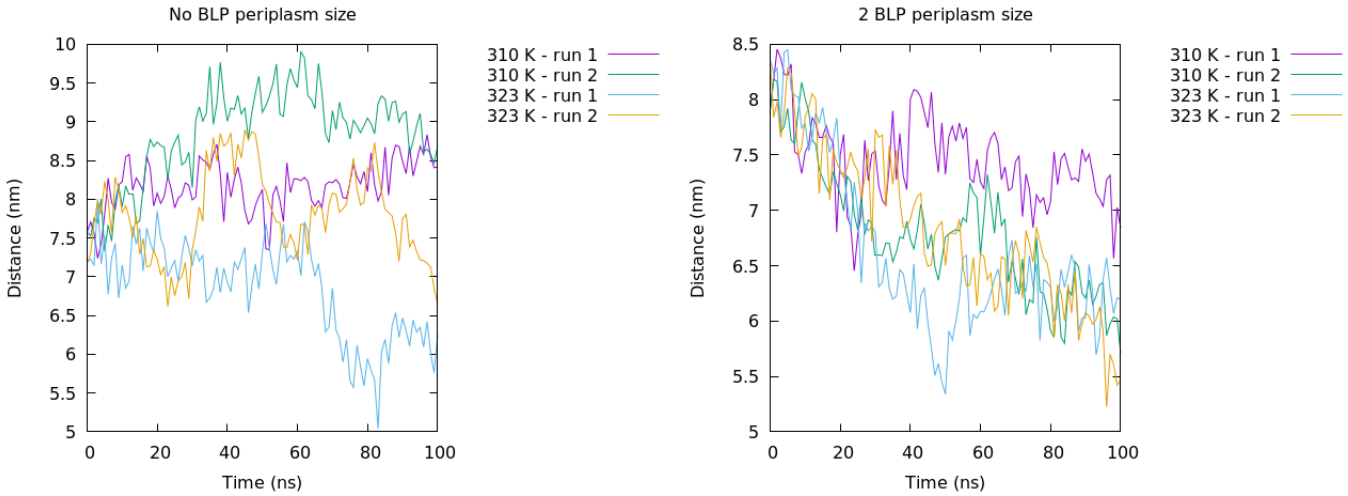


Figure 3.21: The OM to PGN distance over 100 ns when BLP is absent from the system at 310 and 323 K (left), The OM to PGN distance over 100 ns, when 2 BLPs are covalently attached to PGN, at 310 and 323 K (right).

Periplasmic size in the presence of 2 or 0 BLP also varied largely, compared to the dimer, see Appendix Figure 7.2 for an image of the 2 BLP system. In

the monomeric system with no BLP, there were a large range of final periplasmic sizes from 8.5 - 6 nm, so generally these are larger than 1 BLP systems. Also of note was that the periplasm did not consistently decrease in size during the simulation. This furthers the theory that BLP controls the periplasmic size and causes it to shrink, see Appendix Figure 7.3 for further confirmation of this.

The reason that this occurred in the monomeric and not the dimeric systems is that without BLP the monomer will not bind to the PGN as seen in Figure 3.23. This confirms that when the cell wall was free to move in the periplasm that the motion of the cell wall was random, therefore the BLP provides structure and maintains OM to PGN distance, as is proposed experimentally.

Systems with 2 BLP once again displayed the trend that the periplasmic size reduces over the course of the simulation. A range of sizes from 7 - 5.5 nm were observed, indicating that the BLP maintained this distance. This correlates with the 2 BLP dimer system, as the BLP content increased the size of the periplasm increased.

Via comparisons of the OmpA monomeric and dimeric systems it can be seen that dimer more stably bound the cell wall and therefore was capable of maintaining cell wall position but the more flexible and less stable monomer required the BLP as a structural aid to achieve this.

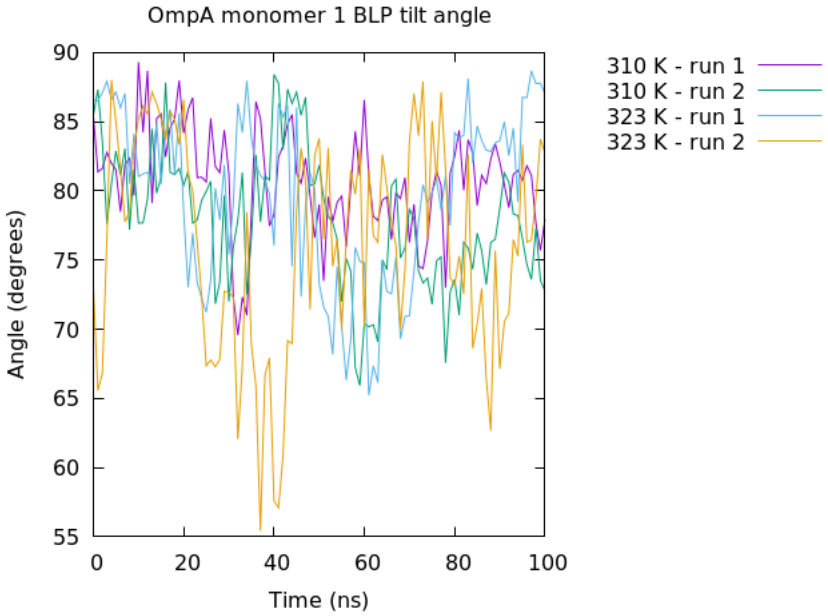


Figure 3.22: The tilt angle of the vector of BLP against the surface of the inner leaflet of the OM over 100 ns, at 310 and 323 K.

Tilting of BLP in the monomeric system has been measured. The largest degree of tilting in these systems was less than the dimer. The minimum angle achieved by the BLP in any system was close to 55 degrees, compared to almost 30 degrees in the dimeric systems. The majority of angles adopted by the BLP in the monomeric simulations rested in the 70-90 degree range. This tilting was still significant as it could allow the cell wall to move upwards and contact OmpA monomers in even the most contracted OmpA proteins, Figure 3.22.

Simulations of BLP show that the tilting fluctuated quickly, however considering the crowded nature of the periplasm it is less likely that this tilting can change so rapidly and more so that the BLP adopted a continuously tilted angle that allows it to accommodate all of the proteins in the periplasm whilst providing the structural strength necessary.

Tilting of BLP has been shown to have a great effect on the cell wall, as the polymer can be easily influenced by the trimeric protein, particularly due to the covalent linkage between the C-terminus and the peptidoglycan side chain. This tilt allowed the cell wall to be pulled upwards, see Figures 3.21 and 3.22.

3.4.4 Removal of BLP in monomeric and dimeric systems

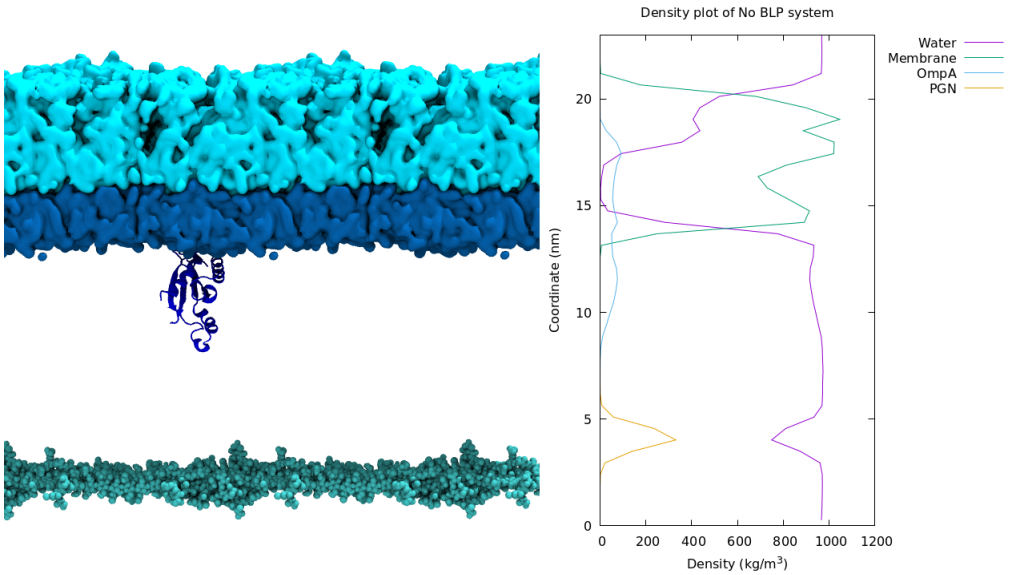


Figure 3.23: A visualisation of the OmpA monomer (blue), PGN (cyan) periplasmic system (left). The corresponding density of the components in the z-axis over 100 ns (right).

Earlier data has been shown for the periplasmic size when BLP is absent. Figure 3.23 shows the effects of BLP removal, where we see that in density there was no bridging protein density, in this case for the OmpA monomer, indicating that there was no contact or structural link between the OM and the PGN.

These lack of interactions form the basis for a simplified version of BLP-deleted *E. coli*, where the OM will bleb and become more permeable to external molecules. In this representation it can be seen that without BLP or cell wall to bind to the monomer will contract and contact the surface of the inner leaflet. As indicated in table 1.2, there was no contact between the OmpA monomer and the cell wall when the BLP is removed from the system.

Conversely the dimer contacted the cell wall even when BLP is removed. This has been shown in earlier periplasmic size plotting, Fig 3.21, This is attributed to the size of the dimer and the stabilisation that occurred from dimerisation. The dimer cannot tilt and bend at the linker region as freely as the monomer can, hence the lack of movement towards the membrane, where the cell wall eventually bound the dimer.

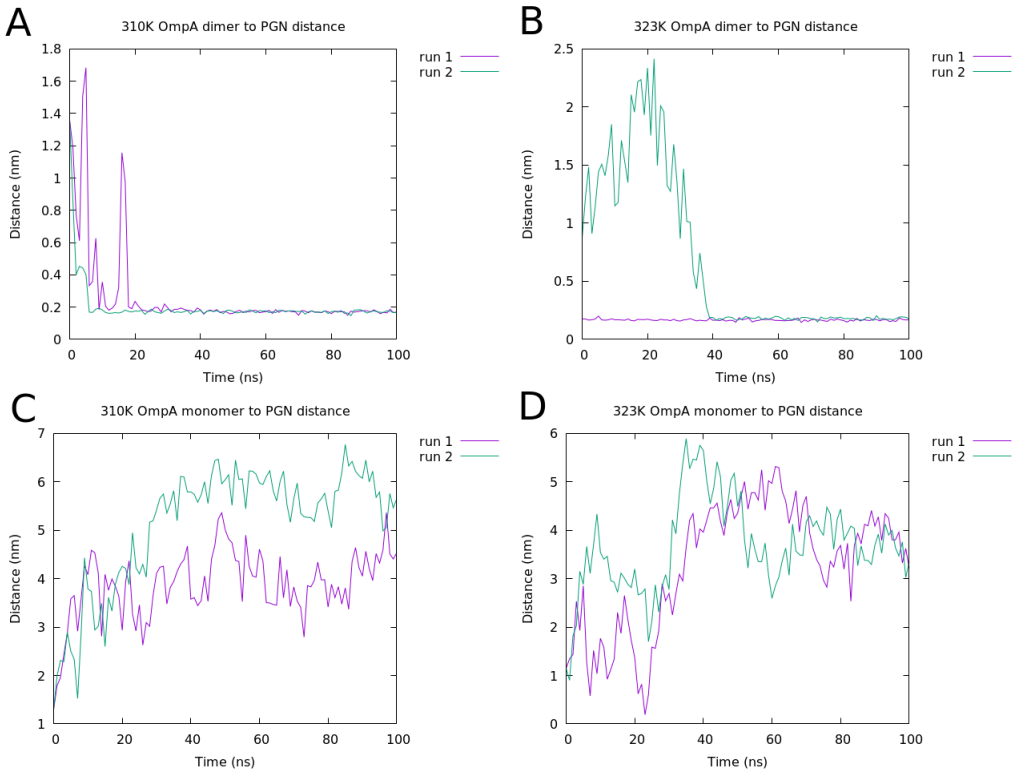


Figure 3.24: The minimum distance between the OmpA dimer to PGN, with BLP absent, over 100 ns. Simulations at 310 K (A and B) and 323 K (B and C).

The minimum distance between OmpA dimer CTDs and the cell wall can be seen in Figure 3.24, at both 310 K and 323 K, where assuming that the cell wall was placed close to the cell wall, there was binding and continuous association in four systems after 40 ns, and that if the protein was placed close to the cell wall the binding occurred very quickly, as was seen the in BLP included simulations, Figure 3.24.

Distance between monomer and PGN shows that the protein did not bind the cell wall. At 310 K it can be seen that even when the monomer was placed 1 nm away from the PGN, the OmpA CTD still moved away and did not bind the cell wall. This eventually led to the distance reaching 4 - 5 nm after 100 ns. At 323 K the distance was similar, as it increased and fluctuated for the whole simulation, indicating no contact.

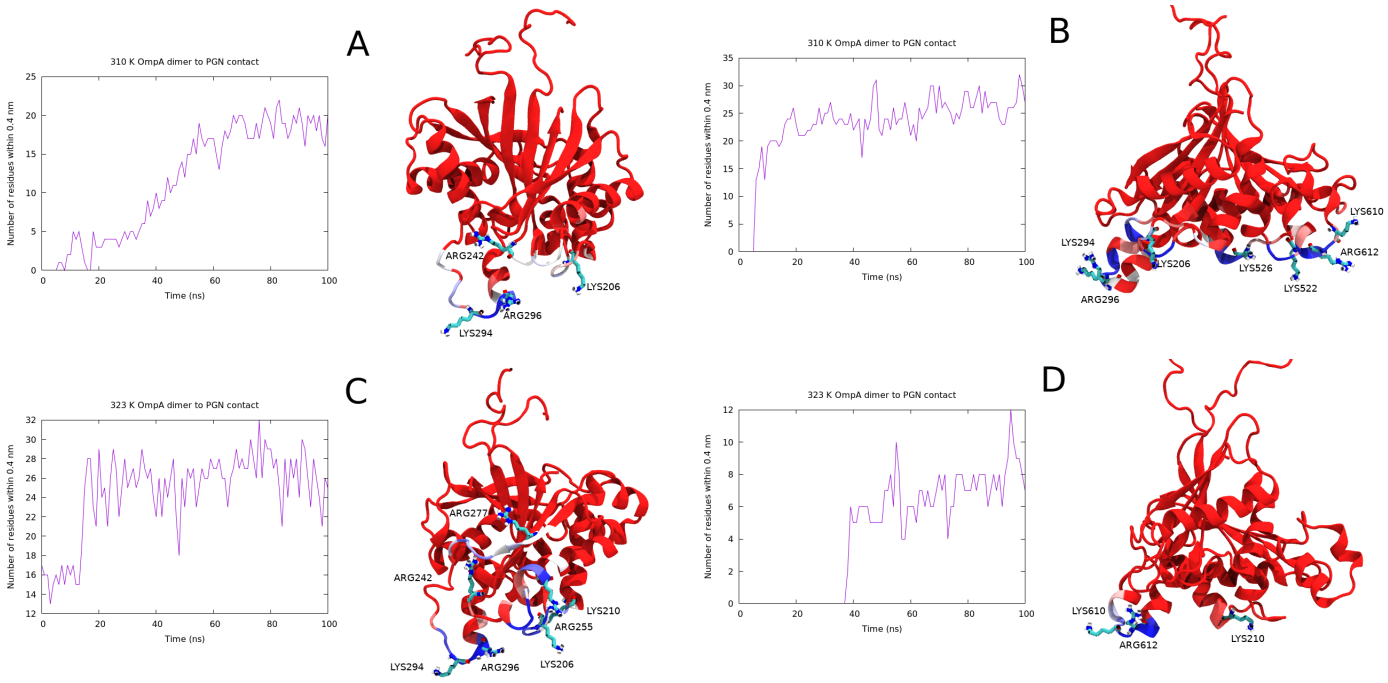


Figure 3.25: The OmpA dimer to PGN contact when BLP is absent, with protein coloured to represent contact using the BWR scheme, where blue is highest contact and red is lowest over 100 ns. Simulations at 310 K (A and B) and 323 K (C and D).

As the dimer contacted PGN in these systems the same initial regions of OmpA contacted the cell wall where the same positively charged residues once again facilitated the binding as seen in previous OmpA to PGN contact and in Figure 3.25.

The highest amount of contact occurred when both monomers of the OmpA interacted with the PGN, at the same unstructured loop region at the edge of the CTD of the protein, involving LYS294 and ARG 296. This is shown where both sides of the protein are highlighted in Figure 3.25, indicating that the positive residues on both sides contacted the PGN.

With respect to the minimum distance data and this, it can be seen that the contact was not immediately established and when comparing No BLP systems to BLP systems in most simulations the same level of contact was established. Once again in run 2 at 323 K only one side of the OmpA dimer made initial contact and therefore did not bind as quickly and with a maximum of 12 residues at 95 ns as opposed to the other systems where the contact was approximately

double this.

3.4.5 Removal of OmpA

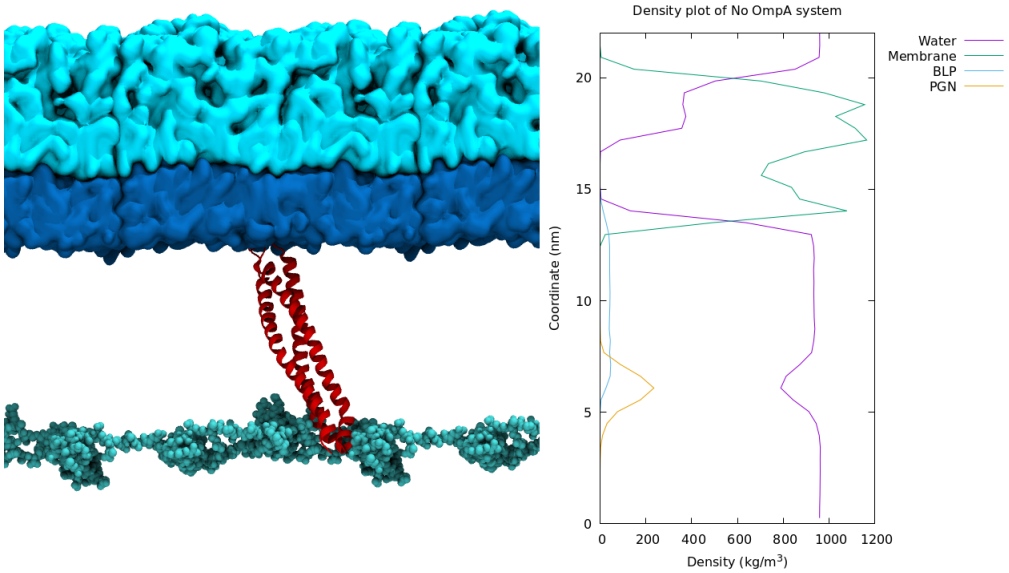


Figure 3.26: A visualisation of the BLP (red), PGN (cyan) periplasmic system (left) where OmpA is removed. The corresponding density of the components in the z-axis over 100 ns (right).

Simulations without OmpA showed that the BLP rested in the membrane and formed a periplasmic bridge between the cell wall and the OM. Naturally there was no binding to assess here without the OmpA present in the system. It is observed that the densities of the BLP and the OM model overlapped, indicating the association between the two over 100 ns of simulation, Figure 3.26.

Without OmpA, only the membrane and PGN can influenced the behaviour of the BLP. It can be seen that the BLP after 100 ns of simulation looked very similar to the "textbook representation" of a straight trimer that maintains a OM-PGN distance similar to the length of the helices of BLP, Figure 3.26.

The density of the PGN here and the no BLP system showed a single sharp peak, indicating a flat layer where the density was highest at 6 nm and $\sim$ 5 nm respectively. There appeared to be a different shape for the PGN density in the OmpA dimer and monomer systems, where sharp peaks were not observed for the dimer and two smaller peaks were observed for monomer, Figures 3.26, 3.13

and 3.1.

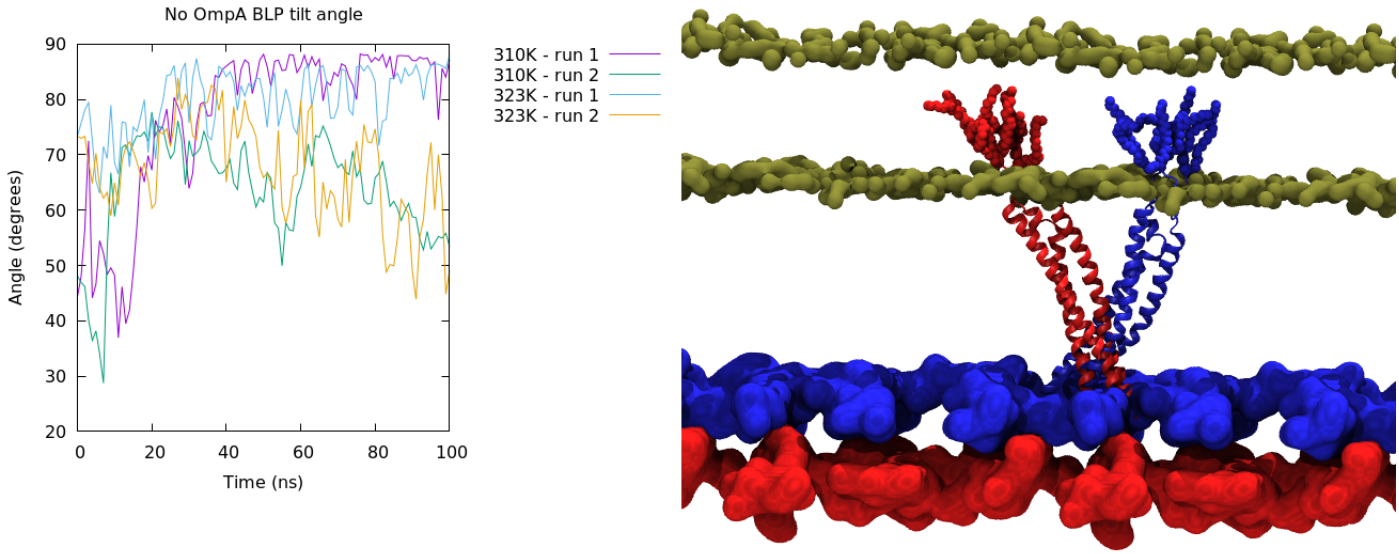


Figure 3.27: The BLP tilt angle over 100 ns at 310 and 323 K (left), A visualisation of the BLP and PGN at 0 ns (red) and 100 ns (blue).

Analysis of tilt angles of BLP without OmpA present once again showed a large range of possible tilting during simulation. The largest tilt angle observed was $\sim$ 90 degrees, whilst the lowest was $\sim$ 30 degrees. Generally the tilting of the BLP was within the 90 - 50 degree range. This free BLP was not influenced by OmpA binding, but the tilting behaviour that is characteristic of this protein was still evident.

The tilt angle for all four systems was observed to vary dramatically, however there was no established trend for the tilting of BLP. In all of the systems studied the bending of the BLP was present, therefore indicating that this is characteristic behaviour of this protein. Generally at the start of the simulation, large deviations were observed in the system without OmpA, but after 40 ns the tilt angle fluctuated around the same value, where this fluctuation was in the range of 20 degrees.

Visualisation of this at 0 and 100 ns showed how the orientation of the BLP can change, where the protein region of the lipoprotein was seen to bend significantly from the original straight orientation. It can be seen that the PGN moved upwards towards the membrane as a result of this, Figure 3.27.

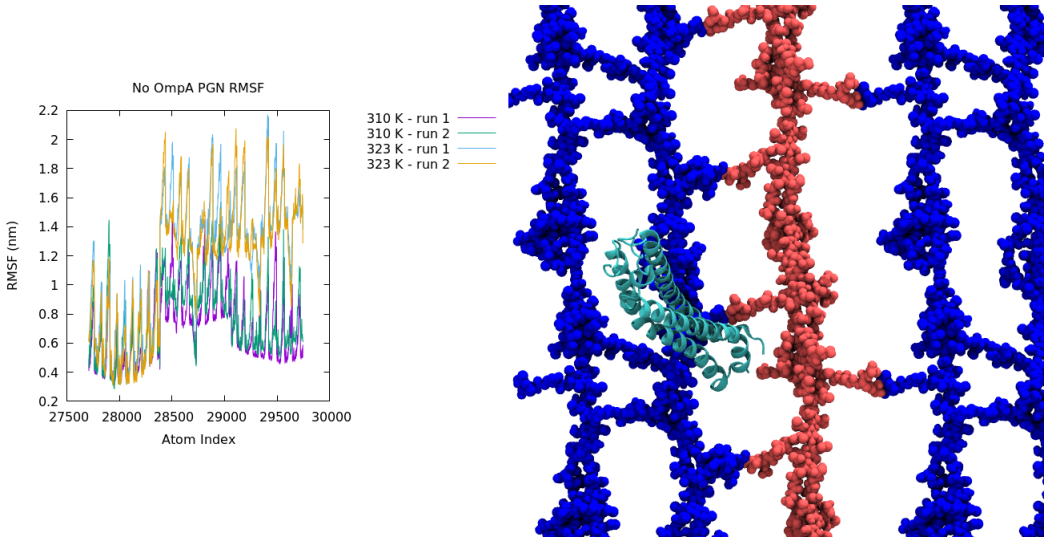


Figure 3.28: The RMSF of the cell wall with BLP bound, no OmpA bound over 100 ns, where the red box highlights the least fluctuating region (left). A top down visualisation of BLP (cyan) bound to PGN (blue) where the red strand of PGN is the bound covalently to the BLP.

The cell wall is a flexible polymer, that can experience expansion and contraction. Regions that are unbound either covalently or non-covalently will fluctuate more. This was seen here, Figure 3.28, where in all four simulations without OmpA, where only BLP was contacting the cell wall, the strand that is connected to the BLP fluctuates the least. This fluctuation across four systems deviated between 0.3 - 1.4 nm for the least mobile strand. This is in contrast to the other strands in the system, where fluctuation was 0.8 - 2 nm and 0.5 - 2.1 nm. As can be seen for the separate temperatures, the fluctuation of the cell wall is higher was the 323 K systems.

This shows that BLP is not only a "structural staple" for the cell wall to OM interaction. BLP also reduced motion of the cell wall, where the fluctuation of the strands was reduced. It follows that other binding events will have the same effect. It is possible this slowing of motion of the PGN aids binding of other molecules, as the structure of the polymer becomes more rigid. BLP and other binding events to PGN cause the layer to become less flexible.

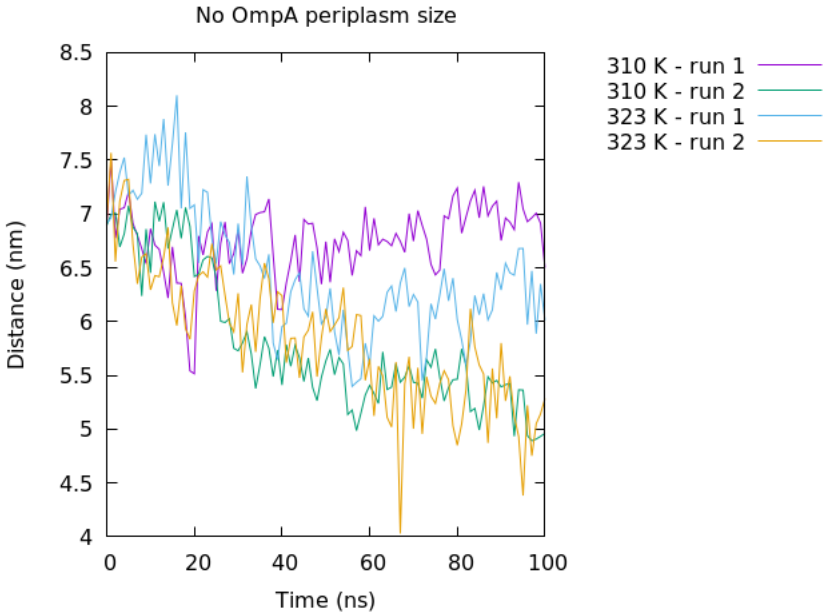


Figure 3.29: The OM to PGN distance over 100 ns, when OmpA is absent from the system, at 310 and 323 K.

The periplasm in these systems varied in size after 100 ns. After simulation, three systems showed a reduction in the size of the periplasm, indicating cell wall movement due to BLP. Two systems showed a reduction from 7 - 5 nm over the simulation. Other simulations showed either a smaller reduction to $\sim$ 6 nm or essentially no change in the final size of the periplasm, Figure 3.29.

These periplasmic sizes correlated well with the tiling angle of the BLP. Large changes in the tilting angle also facilitated large changes in the size of the periplasm. Overall size was proportionate to maximum tilting angle. This can be seen where the largest periplasmic space was run 1 310K, which with respect to tilting displayed a close to perpendicular angle during simulation, indicating the BLP was straight and the cell wall was at the maximum distance caused by BLP. The opposite holds true, where two runs showed a smaller final periplasm, these runs indicated a similar pattern of tilting during simulation and the reduction to 5 nm was caused by a tilt angle of $\sim$ 50 degrees.

This is yet another indication that the BLP does serve as a structural staple, but a dynamic one. Instead of behaving like scaffolding to hold the periplasm together the BLP acted as a hinge, controlling the size of the periplasmic space and fluctuations of the cell wall.

3.5 Discussion

The OM to PGN periplasmic dynamics have been examined via MD simulation, where the OmpA, BLP and PGN are interacting during these simulations. The construction of a cell wall mesh allows for the exploration of covalent and non-covalent linkages between the two barriers present in the periplasm. These simulations reveal molecular insights regarding OmpA to PGN binding, with BLP facilitating the interaction.

BLP can bend and tilt, at the PGN interface, which in turn facilitates a movement of the cell wall toward the OM in the direction perpendicular to the bilayer. This movement allows for the contact between OmpA and PGN, particularly in the case of the OmpA monomer.

Upon contact with the cell wall the OmpA monomer and dimeric contact is stable. This initial contact is due to the positively charged residues present in a flexible loop present at the edge of the OmpA periplasmic domain. These residues are numbered as LYS294 and ARG296. This is not the only binding groove as there appears to be a secondary groove, located at LYS206, which contacts PGN after the initial contact of the first groove with the m-DAP residue of the PGN side chain. Comparing to known literature, binding was not observed for the key residues in the binding pocket located at GLU241 and ARG256; however this is attributed to the simulation timescale, as only initial interaction is explored here.

Considering the timescale of these simulations longer timescales would reveal further data of the OmpA to PGN binding and further repeats would allow for better sampling of the interaction events. Initial interaction shows strong binding OmpA-PGN binding, where dissociation of OmpA from the cell wall is entirely possible in longer simulations.

Comparing the OmpA monomer and dimer shows that dimer binds PGN in every simulation, with or without BLP. The dimer binding without BLP is attributed to the larger surface area of the periplasmic clamp, hence increasing the number of positively charged residues that can interact favourably with the negatively charged PGN sheet. The monomer does not bind the cell wall when BLP is absent, but binds in the presence of the lipoprotein. This is attributed to the BLP reducing the size of the periplasm such that the OmpA monomer can contact the cell wall. Otherwise, OmpA has been shown to contact the membrane instead, via the linker region contraction, which the less mobile dimer cannot do.

With respect to OmpA, it has been proposed that OmpA exists as both monomers and a homodimer *in vivo*. Mass spectrometry has revealed a potential structure for the dimer and as previously stated, simulation studies for the dimer exist, there is no physiological confirmation of the role of the dimer in the cell[33].[166]

OmpA dimers and monomers interact favourably with the BLP, where analysis shows that the monomer contacts the BLP less often. This favourable interaction between OmpA and BLP is an anchoring point for the OmpA, which in turn facilitates cell contact, where the periplasmic clamp can essentially 'walk' down the BLP similarly to a ladder to contact the PGN. This walking effect is controlled by the end to end length of the linker region, where if the region reduces in distance, OmpA contacts the membrane, and the linker region does not expand afterwards.

Binding between OmpA and other components of the periplasm is dominated via coulombic interactions. Given that these interactions are non-covalent and therefore distance dependent, it is likely that the periplasmic clamp detaches and attaches to surfaces such as BLP and PGN in a dynamic equilibrium.

Tilting of BLP is the physical effect that dominates periplasmic size. Generally it was observed that greater degrees of tilting (lower angles) in the lipoprotein with respect to the OM result in a smaller periplasmic volume. This is confirmed as the removal of BLP in the monomeric systems led to a large range of periplasmic sizes, in some instances where the distance between the OM and PGN increased during simulation. This was not observed during dimeric simulation as the binding of the dimer caused the cell wall to retain its position during simulation, indicating that the dimeric binding of OmpA is more favourable and stronger than the monomer.

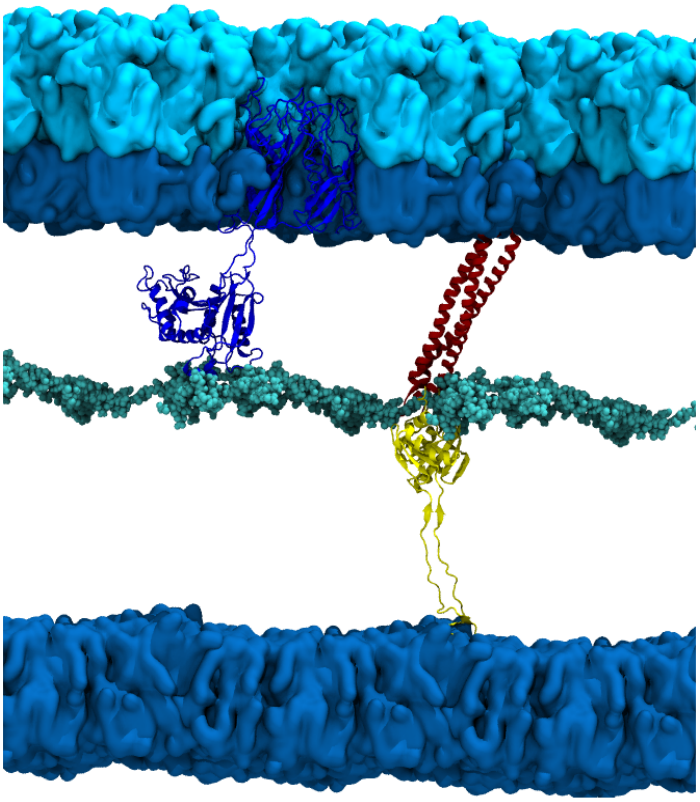
The structural role of BLP was confirmed in simulation, where the connection between the OM and PGN was maintained by the covalent PGN linkage and the N-terminal lipid association to the membrane. Exploration of this role has shown that the BLP is a highly mobile, flexible protein, that functions as a periplasmic hinge. BLP also stabilises the mesh of the cell wall, reducing fluctuations of the atoms in strands that it is bound to.

In the field this represents the first simulation using a membrane and cell wall to model the periplasm. This study advances our knowledge about OmpA interactions, and BLP is a relatively new candidate for simulation. It can be seen that the presence of these two proteins in the periplasmic environment drastically changes the behaviour of the cell wall: rather than only being a barrier to prevent lysis of the cell, it is a dynamic mesh that adjusts to the various interactions surrounding it.

Other discussion and analysis surrounding this topic can be found under Appendix C, published as **Braun's Lipoprotein Facilitates OmpA Interaction with the *Escherichia coli* Cell Wall**, where this work was done in collaboration with Firdaus Samsudin and Syma Khalid.[168]

Chapter 4

Simulations spanning the entire envelope



4.1 Context in the Literature

TolR is an inner membrane protein.[169] The monomer is comprised of 142 residues and in the membrane the protein is a dimer.[59] The protein has three regions, a helical membrane domain, an extendable linker and a periplasmic domain. It is proposed that there are two states of the TolR protein, open and closed.[59] The open state presents the C-termini of the protein to the periplasm, where the protein is extended via the linker domains similar to OmpA. The closed state of the protein involves the coiling of the linker domain, where the flexible termini of the C-terminal domain are buried in the protein. TolR is a protein included in the Tol family, which participates in the Tol/Pal system.[170] Other Tol proteins are TolQ, TolA, and TolB, which form a periplasmic complex that form interactions with Pal (Peptidoglycan Associated Lipoprotein), specifically where TolB interacts with Pal[171].[150] TolR interacts with TolA and TolQ *in vivo* to form an activated TolRAQ complex[59].[172]

The function of the Tol/Pal system is to maintain the integrity of the OM, particularly during OM invagination as part of cell division.[150] TolRAQ are required to complex in order to cause the proton motive force activation of TolA and to allow TolA to form a complex with Pal.[59] It has also been shown that TolR is necessary for the movement of colicins and bacteriophage into the cell.[173] Hundreds of TolR molecules are present in a normal wild type cell.[172] Structural data for the entire Tol/Pal system is not available. The structure of TolR and TolB have been resolved, even in the case with TolB associated to Pal[174175].[176]

Studies show that the Tol/Pal system when disturbed via mutation will result in the release of vesicles that contain periplasmic proteins.[177] This system is shown to be localised at constriction sites in *E. coli*. It has been shown that the invagination of the membrane is due to accumulation at the septum during the division of the cell.[178] In addition the energy state of the Tol complex has been shown to regulate the activity of PBP1B, an important PGN synthase.[177] This would seem to suggest that Tol can regulate PGN production during cell division. It has been shown that during cell division TolR requires TolA or TolQ to be localized correctly, however a key point is that for Pal localization at a division site every single Tol protein is required. This raises the question that if the TolA or TolQ proteins are not present, then where does the TolR reside? Localization only occurs during cell division, where the Tol proteins are actually seen to be dispersed throughout the cell, and then migrate to future division sites, followed by dispersal. An interesting point to note is also that the number of Pal molecules present is at least a magnitude higher then Tol proteins, suggesting that the Tol/Pal system can have multiple Pal binding partners.

Kleanthous et al. have presented a structure of TolR from *E. coli* (residues 36 - 142) relative to the only other structure that originates from *H. influenzae* (residues 62 - 133). In the *H. influenzae* structure there is a deep groove that

is suggested to support PGN binding.[59] Here they show that the architecture of the TolR dimer is very different in *E. coli* where the C and N termini are intertwined with each other compared to the *H. influenzae* structure. The resulting strand swapping from this intertwining causes the monomers of TolR to be rotated 180 degrees to each other. This effectively removes the binding groove that is seen in the 62-133 structure of *H. influenzae* TolR. By removal of this strand swapping, resulting in residues 60-133, a potential PGN binding domain is revealed that is not present in the full length *E. coli* structure. This in turn suggests that TolR can function as a stator protein in the Tol/Pal complex, similar to Mot proteins in the flagellum. To do this it must undergo structural remodelling via termini unfolding that allows the protein to reach and bind the cell wall at a distance approximately 90 Å from the IM. This binding to PGN has not been confirmed with modelling, however it is possible that at some stage during the Tol/Pal mechanism that it influences the cell wall function.

Stator proteins are starting to become better understood in bacterial biology. Generally these proteins complex to others to form a molecular motor. These molecular motor complexes share some sequence and structural properties. Essentially they function via converting energy from proton-motive force (transmembrane potential) to mechanical energy in order to allow molecules to move. There are well characterized motor like complexes such as MotB/MotA which drive flagellar motion.[179] Unlike other similar complexes, the function of the Tol motor is not well understood. However despite this, it is obvious that considerable mechanical energy would be required to assist membrane constriction, so the key to this behaviour could rely on Tol protein interactions with the surrounding environment. TolR has been compared with MotB due to the need for many hydrogen bonds (approximately 30) to be broken in order for the extension and binding to PGN to occur. MotB is known to continuously cycle between an IM and PGN bound state, which is similar to what is proposed for TolR behaviour.[180] TolR is thought to act as a plug for the total complex, where the IM bound state represents the closed pore. The 27 C-terminal residues are suggested to stabilise the intertwined dimer state of TolR. It has been shown that ,using cysteine mutagenesis scanning, TYR117C can only form a disfulide bond when strand swapping occurs. This is supported by a study that indicates that this disulfide bond prevents activation of the complex.[181]

In the context of simulation TolR has been studied before. Wojdyla et al. used MD simulations to describe a model of transmembrane-TolR (36-142) and a model with modelled transmembrane regions (15-142). A dimeric model was constructed using the x-ray structure for TolR. Their TM-TolR model was simulated in a 1:4 PG/PE membrane, first equilibrated in coarse grain resolution for 200 ns and then backmapped to a united atom model in GROMOS96 53ab for 100 ns atomistic simulations. These simulations highlighted the stability of the dimer in a model membrane. Additionally it was noted that ASP23 remained bound to a sodium ion throughout, despite being part of the TM helix. This lends support to the theory that it is the critical site for protonation to allow

transmembrane potential coupling for the Tol/Pal system. They observed that the TM helices remained apart during the run, despite experimental research that a disulfide bond between cysteine mutated L22C exists in the helix. TolQ and TolA, lacking experimental data for structures, were absent from the simulations, therefore it was proposed the TM helices of these proteins keep the TolR dimer helices in contact. This is currently the only other simulation of the TolR protein.

The other component added ,in this work, to the existing OM-PGN system is the inner membrane. IMs have been extensively simulated using MD to study models of *E.coli* previously.[182] It is generally accepted that the OM inner leaflet and IM leaflets are of a similar composition and this is applied in these simulations.

It has been shown that there are two states of TolR in *E. coli*; one which can bind PGN and the other which cannot, and the states are dependent on the structural state of the periplasmic domain and the presence of the potential binding groove. We observed that the previously mentioned open TolR state allows for PGN binding, whilst the closed state of TolR does not. Given that a simulated model for the OM exists including OmpA and PGN, adding TolR in both states allows for this binding to be validated and to examine cell wall behaviour when under the influence of two proteins and BLP simultaneously.

4.2 Premise

Having simulated one side of the periplasm, following this with a study of the whole periplasm is the next logical step. Now that OmpA has been studied in this environment, including an inner membrane and components from the other side of the periplasm allows for construction of an initial model of this compartment in *E. coli*.
TolR is an inner membrane protein, see Figure 4.1. It is proposed to contain a PGN binding domain, to interact with the cell wall. It is part of a complex of proteins that is currently not well understood.

Based on this the following aims are proposed for this study:

- To add TolR to the periplasmic model that was studied in Chapter 3.
- To assess the PGN binding ability of the TolR dimer compared to the OmpA dimer.
- Investigation of the TolR dimer in both the "open" and "closed" states.
- To vary the periplasmic system in order to observe how to the cell wall position changes based on which proteins are present.

This is explored in this study, whereby the two conformations of TolR, both closed and open are monitored when interacting with the cell wall. Open TolR is proposed to reveal the binding site that is favourable for the PGN mesh, whilst closed TolR is proposed to be IM bound and cannot bind PGN. This is done in tandem with monitoring how the presence of OmpA and TolR on each side change the position and shape of the cell wall model giving information regarding cumulative interactions can affect the biopolymer.

4.3 Methods and Simulation content

All simulations were run for 200 ns, using the GROMACS 2018 code, where there were at least 3 independent repeats. The forcefield used was GROMOS54A7.[125] The full length OmpA model used was the same as in previous OmpA dimer simulations in the periplasm.[18] The TolR dimer models were obtained from Phillip Stansfeld. These models represented the open and closed states of TolR, where they are based on the structure from Wojdyla et al. (PDB:5BY4). Modelling to produce the TM helices was done using MODELLER, where ILE38-PRO141 is based on structural data, whilst PRO20-PRO37 were modelled as helical, with all other residues modelled as loops. This was done based on secondary structural prediction, via Jpred and PSIPRED. It is proposed that the TolR can cycle between the open and closed states, which are different due to domain swapping, but require significant rearrangement to do so. The peptidoglycan model used is the same three strands of 10 repeating monomeric PGN units as in the previous chapter. The cell wall was periodic, behaving as a self bonded infinite sheet in the x and y dimensions, as in the previous chapter. Once again this sheet was positioned $\sim$ 9 nm from the OM inside the periplasm. The BLP model is once again built on the same structure from Shu et al (PDB:1EQ7).[165] The N-lipidation was a tripalymitoyl-S-glyceryl-cysteine, constructed from standard GROMOS54A7 parameters, as was the covalent linkage between BLP and PGN.

System	OmpA dimer structure	TolR structure	BLP	TolR contacting PGN?
Sys1	Wild type	open	Yes	Yes
Sys2	Wild type	open	No	Yes
Sys3	CTD Truncated	open	Yes	Yes
Sys4	CTD Truncated	closed	Yes	No

Table 4.1: Simulation systems set up for the full periplasm model environment, containing OmpA, TolR, BLP and PGN.

The OM model consisted of the same rough LPS outer leaflet and a 90:5:5, PE, PG, CL ratio in the inner leaflet of the membrane. These lipids are 1-palmitoyl 2-cis-vaccenic phosphatidylethanolamine, 1-palmitoyl 2-cis-vaccenic phosphatidylglycerol, and 1-palmitoyl 2-cis-vaccenic 3-palmitoyl 4-cis-vaccenic diphosphatidylglycerol respectively. The IM model model uses this 90:5:5 membrane composition with the same lipids, for both leaflets, modelling a symmetric bilayer. Simulations were run at 310 K using the V-rescale thermostat at 1 ps time constant. The Parrinello-Rahman barostat maintained pressure semi-isotropically at 1 atm, using a 1 ps time constant. LINCS was applied to allow a 2 fs timestep.The short range cutoff for both VdW interactions and charged interactions was set to 1.4 nm.[100] The SPC water model was used and all of the systems contained a neutralising concentration of 0.2 M NaCl ions and neutralising Mg^{2+} ions were used to balance the LPS charge.[167]

Simulations were started with the full length OmpA dimer and the extended TolR dimer, both inserted into their respective membranes. These proteins were positioned close the the PGN (less than 0.5 nm) on the respective sides of the cell wall. OmpA truncations consisted of removal of residues 173-316 of the periplasmic domain in each protomer in the dimer. BLP was included in truncated systems to maintain OM to PGN association.

4.4 Results

4.4.1 Including TolR in the periplasm

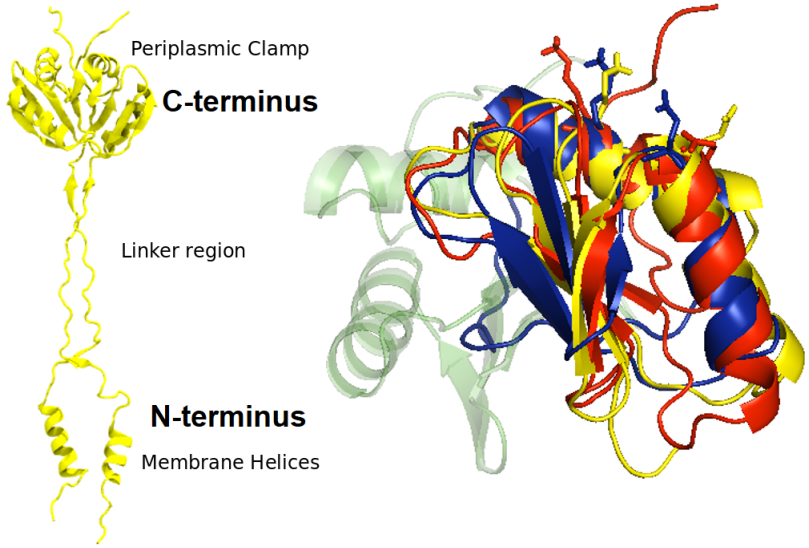


Figure 4.1: A visualisation of the TolR protein in the open state extended state conformation (left). Aligned structures of one TolR protomer C-terminal periplasmic clamp from three separate bacterial species, *E. coli* ExbD (blue), *E. coli* TolR (red) and *H. influenzae* TolR (yellow) where the second protomer is shown in a translucent representation (right). The licorice stick representations in each colour show the charged residues that are conserved between species types, where these residues can contact PGN.

The extended TolR dimer has been simulated in the periplasm. It is proposed that the dimer will extend based on a proton-motive force that allows it to contact the cell wall.[183] For this interaction only cartoon representations exist,

with the cell wall as a flat geometric shape, which we know not to be true.

Shown in Figure 4.1 is the TolR dimer, extended after 200 ns of simulation, note the terminal regions which were extended. The unstructured termini are well conserved between structures that are known for various species. The structural overlap between these species is highly similar.

It is proposed that this exposed terminal region will be presented to the cell wall first. The composition of the 5 terminal amino acids in this region is LEU-MET-THR-GLN-PRO. In contrast to OmpA, these residues are non-charged, but some are polar, indicating that the initial TolR-PGN association is not entirely dominated via charged interactions. The TolR dimeric clamp is a smaller and more flexible unit than the OmpA clamp. In addition, the membrane bound component is helical, as opposed to the porin like barrel structure associated with OMPs.

Insertion of this TolR structure within the membrane yielded simulations that can be used to test protein binding on either side of the cell wall, and the first full periplasmic model.

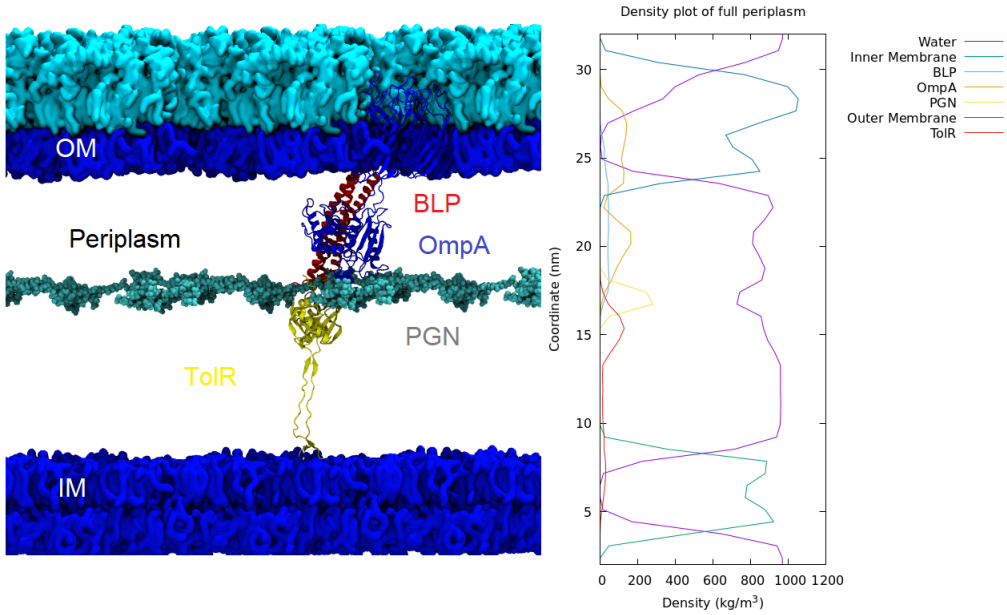


Figure 4.2: A visualisation of the OmpA (blue), BLP (red), PGN (cyan) and TolR (yellow) periplasmic system after 100 ns (left). A density plot of the model periplasmic system over 100 ns (right).

After 200 ns the OmpA and TolR dimers became associated to the PGN. This binding for both proteins is seen in density plotting of Figure 4.2, where the overlap of density of both with PGN is observed.

It can also be seen that the periplasm essentially functions as a closed unit, aside from a small amount of water movement that was allowed via the OmpA barrels. The periplasm was stable and no significant distortion of the cell wall occurred as a result of the enclosed space. The OM to PGN distance was seen to still be a function of BLP and OmpA interaction, as shown previously. TolR remained in the extended conformation over the simulation, indicating that the PGN binding is strong enough such that the flexible regions and linker region between the clamp and membrane helices did not contract.

From the density plot, Figure 4.2, the asymmetry of the membrane present in the system is observed, where in the OM density, the highest peak was greater than 1000 kg/m^3, compared to the lower peak at 800 kg/m^3, where the rough LPS lipids have a greater mass than the PL inner leaflet. IM peaks for the density showed a similar value for both leaflets, where the slight difference in the two can be attributed to the presence of the helices of TolR.

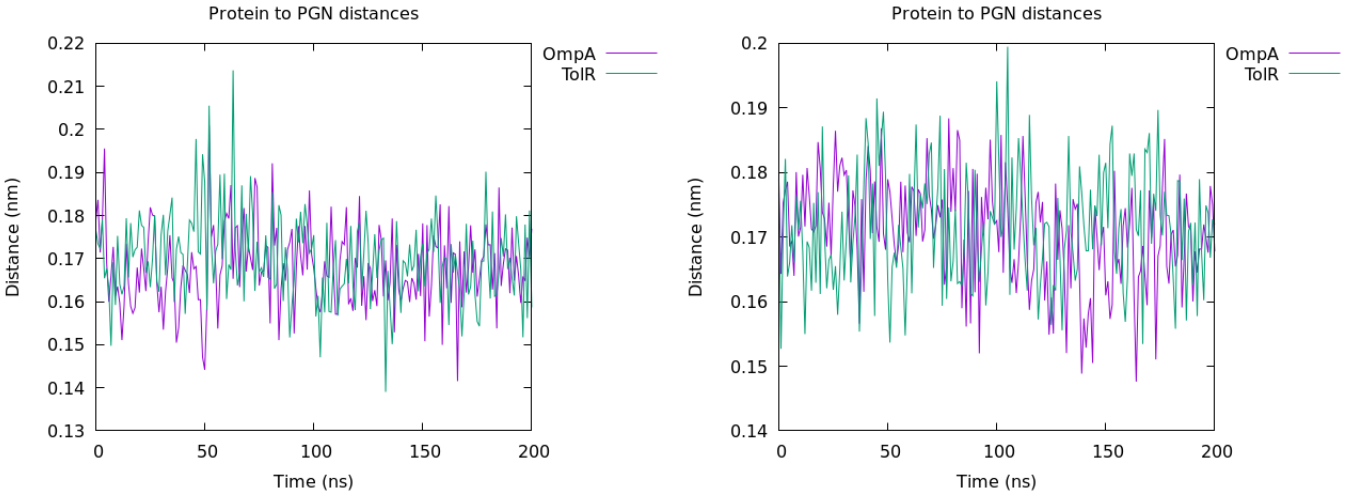


Figure 4.3: The minimum distance between TolR-PGN and OmpA-PGN over 200 ns, two independent repeats shown (left and right).

Visually the proteins were in contact with the PGN. Distance analysis shows that across separate 200 ns trajectories OmpA and TolR were in contact with the cell wall. This can be seen where both proteins were generally within at least 0.2 nm of the PGN mesh, as in Figure 4.3.

This distance fluctuated where the minimum values for TolR are 0.14 and 0.155 nm respectively, whilst the maximum values were 0.21 and 0.2 nm for TolR to PGN distance. OmpA binding adopted similar values, that fluctuated in the same fashion, indicating that both were positioned on the cell wall for the entire 200 ns.

As the TolR was presented to the cell wall in the extended open conformation in these systems, this indicated that the TolR will bind favourably to the cell wall and that the closest extended domain was contacting PGN, as seen in Figure 4.2. As the TolR was fully extended, no change in distance between the PGN and TolR was observed.

TolR, similarly to OmpA, did not unbind from the cell wall during 200 ns of simulation, there was no large change in separation between protein and PGN.

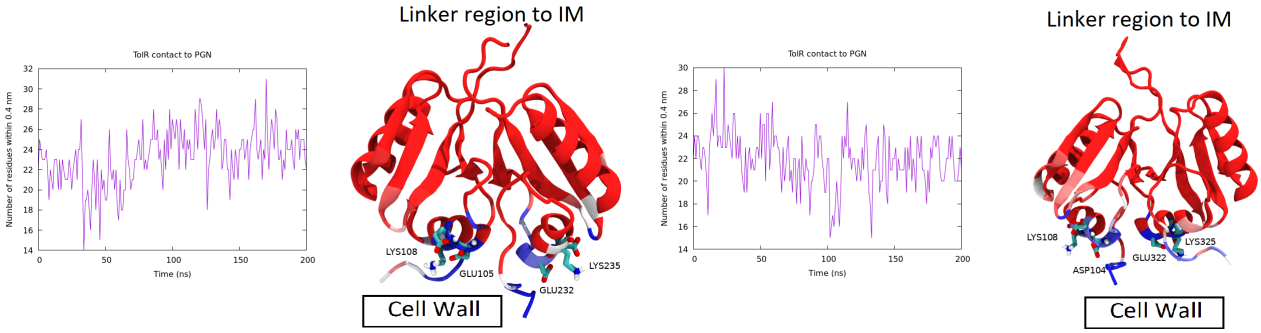


Figure 4.4: The TolR dimer to PGN contact, with protein coloured to represent contact using the BWR scheme, where blue is highest contact and red is lowest over 200 ns, where the direction of the PGN and IM are indicated. Two independent repeats shown (left and right).

The specificity of TolR-PGN binding is shown in Figure 4.4. This binding occurred, as proposed, in the terminal regions of the TolR periplasmic domain. The highest contact was present on the flexible termini of the TolR, indicating that these residues enmeshed with the cell wall and "snorkeled" towards the PGN.

The initial placement of TolR allowed for a contact of $\sim$ 25 amino acid residues that were contacting PGN. Over 200 ns it can be seen that this contact fluctuated, reaching close to half of the initial contact, whilst also increasing to 30 residues. Generally the initial contacts and the contacts that were established after 200 ns are similar.

The residues that were present in the helix adjacent to the unstructured termini contacted the PGN, to a lesser degree. For comparison against OmpA the charged residues in these regions are shown , in Figure 4.4, where it is possible to see that both positively and negatively charged residues interacted with the cell wall. It can be seen that the positively charged lysine side chains were oriented in the same direction as the termini, toward PGN.

Presented on the end of the helix are negatively charged GLU and ASP residues, which will interact unfavourably with the negatively charged cell wall. It is noted that the adjacent beta sheets were also interacting with PGN, but to a lesser degree than the region surrounding the unstructured termini.

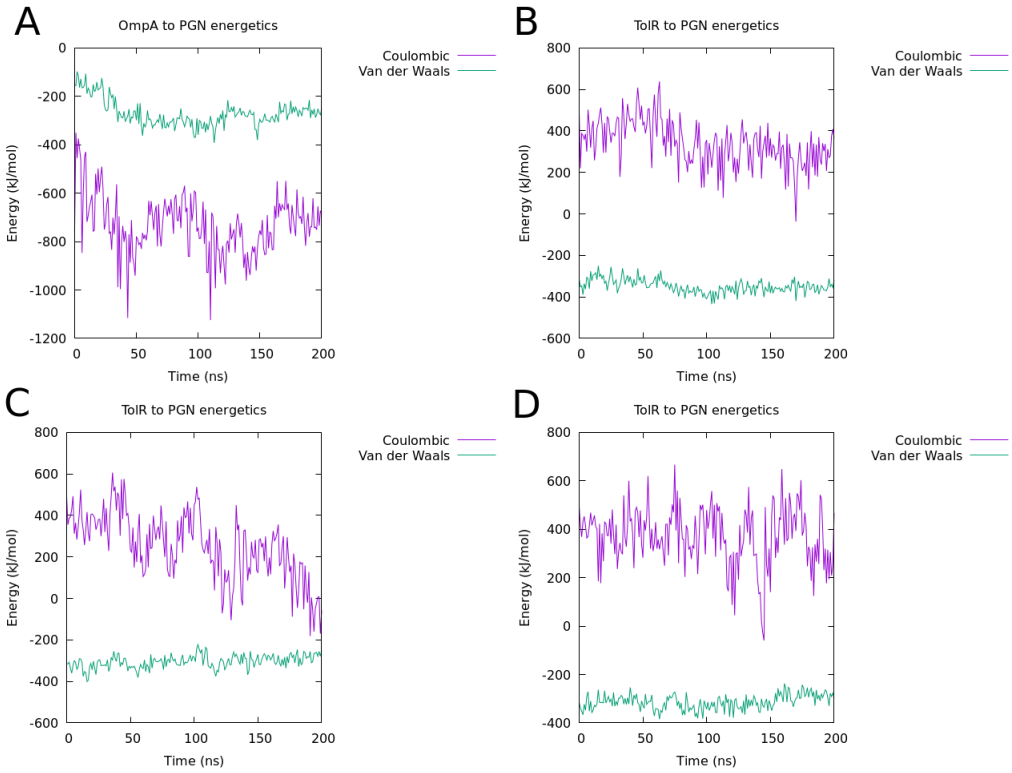


Figure 4.5: The short range electrostatic and Lennard-Jones interactions between OmpA and PGN over 200 ns (A). The short range electrostatic and Lennard-Jones interactions between TolR and PGN over 200 ns (B, C and D).

Given that these negatively charged residues were interacting with the cell wall, across all 200 ns simulations the energetics for wild type open TolR show that there was a generally repulsive charged interaction with the cell wall, Figure 4.5.

This is in contrast to the VdW energy, which is naturally attractive. TolR experienced an overall stabilising interaction as a combination of the two energies. The binding was not as stable as OmpA, but was stable enough to provide lasting interaction with PGN. In one simulation the charged repulsion was seen to trend towards attraction in the last 100 ns, in others there were periods where this repulsive energy dropped to lower values. It is possible that the open structure must reorganise in order to achieve stable PGN binding.

4.4.2 Pulling the OmpA and TolR from the cell wall

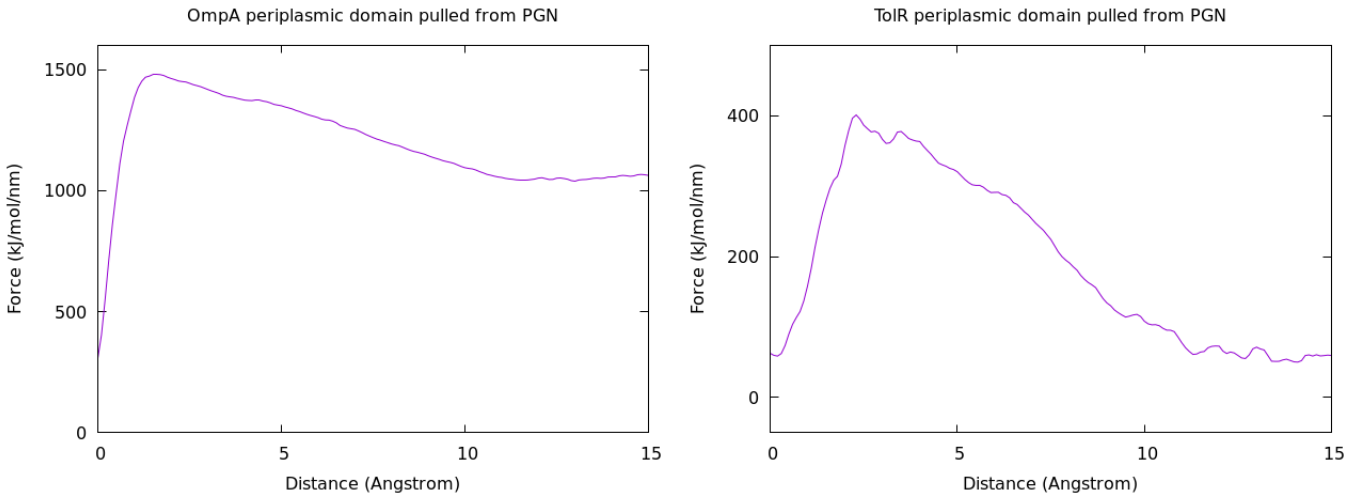


Figure 4.6: Force required to pull the bound OmpA dimer from PGN over a 15 angstrom distance (left). Force required to pull the bound TolR dimer from PGN over a 15 angstrom distance (right).

Pulling simulations of both the OmpA and TolR proteins show that the PGN binding was favourable, where Figure 4.6 shows the force associated with pulling the clamp regions from the cell wall when bound. Pulling of the bound OmpA periplasmic clamp away from PGN shows that within the first angstrom there was a required maximum 1,500 kJ/mol/nm force in order to cause dissociation of the protein from the cell wall. TolR showed a maximum force of 400 kJ/mol/nm to instigate unbinding from PGN. This force also did not peak as quickly as OmpA. The maximum force was reached at approximately 2.5 Å. From this it is confirmed that PGN binding for OmpA was stronger than that of TolR.

After the TolR was pulled 10 Å it was seen that the force required drops off and was stable, indicating that the TolR can easily be pulled through the solution of the periplasm, at a value of $\sim$ 50 kJ/mol/nm. OmpA however continued to experience a force that is magnitudes larger than this at $\sim$ 1000 kJ/mol/nm. This indicated that the smaller TolR periplasmic domain was easier to pull through the periplasm than that of OmpA, indicating that little force was required to cause the linker region of TolR to contract, compared to the OmpA dimer. Visualisations of the system before and after pulling can be seen in Appendix Figures 7.5 and 7.6

4.4.3 Movement of TolR, OmpA, BLP and PGN in the periplasm

Comparing the number of PGN contacting residues, see Figure 4.4, between OmpA and TolR shows that there were a similar number of residues involved in the binding. This infers that a similar area of protein was contacting the cell wall in both cases, however visualisation, pulling force and energies show vastly different binding properties. From these results *in vivo* one would expect a lower proportion of TolR-PGN complexes than OmpA-PGN ones.

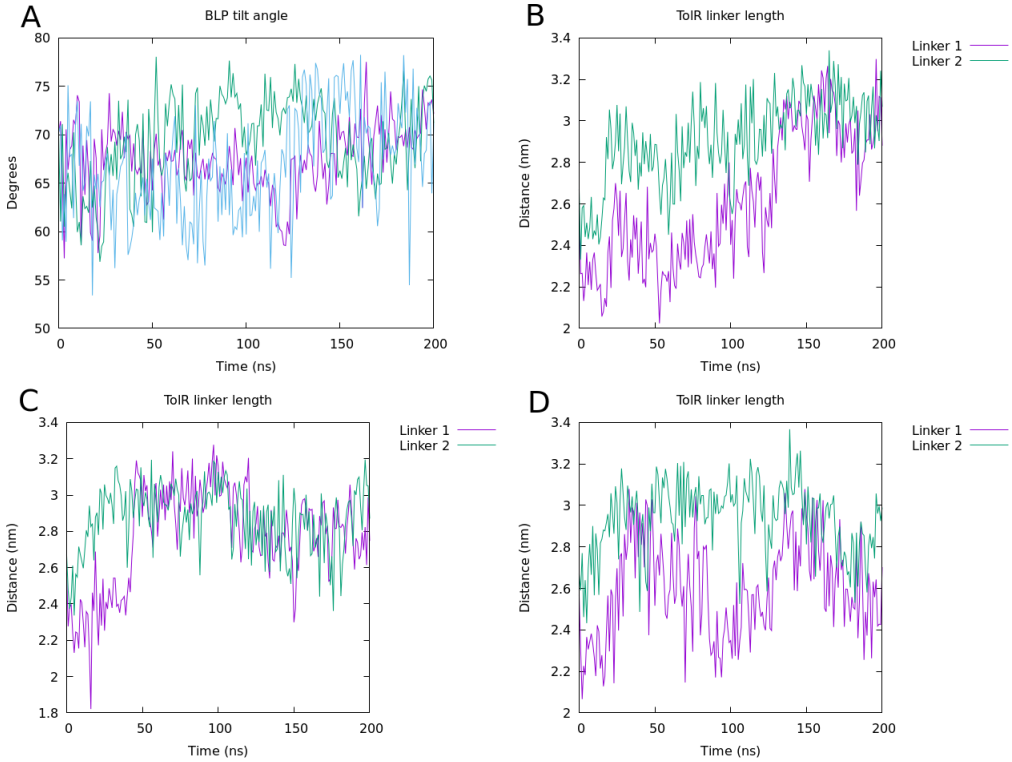


Figure 4.7: BLP tilt angle with respect to the OM over 200 ns, over three independent repeats (A). The linker length of each individual protomer in the TolR dimer over 200 ns, in the wild type system. Independent repeats shown (B, C and D).

The flexible linker in TolR can behave in a similar fashion to OmpA. This linker distance will depend on PGN position, and therefore on BLP tilt angle. The BLP tilt angle in these systems was seen to behave similarly to previously presented simulations, where generally the fluctuations were in the 60 - 75 degree range. BLP tilting to this degree caused the PGN to move close to the OM in the periplasm.

As TolR was bound for the whole simulation, any PGN movement towards the OM should cause extension of the linker in each protomer present. This was the case in each system, where generally an increase was observed. This increase was $\sim$ 30 % starting from 2.4 nm and ending at close to 3 nm, as seen in Figure 4.7. All simulations showed that both linkers adopted the same distance and converged by 150 ns, where one system did not show as much increase.

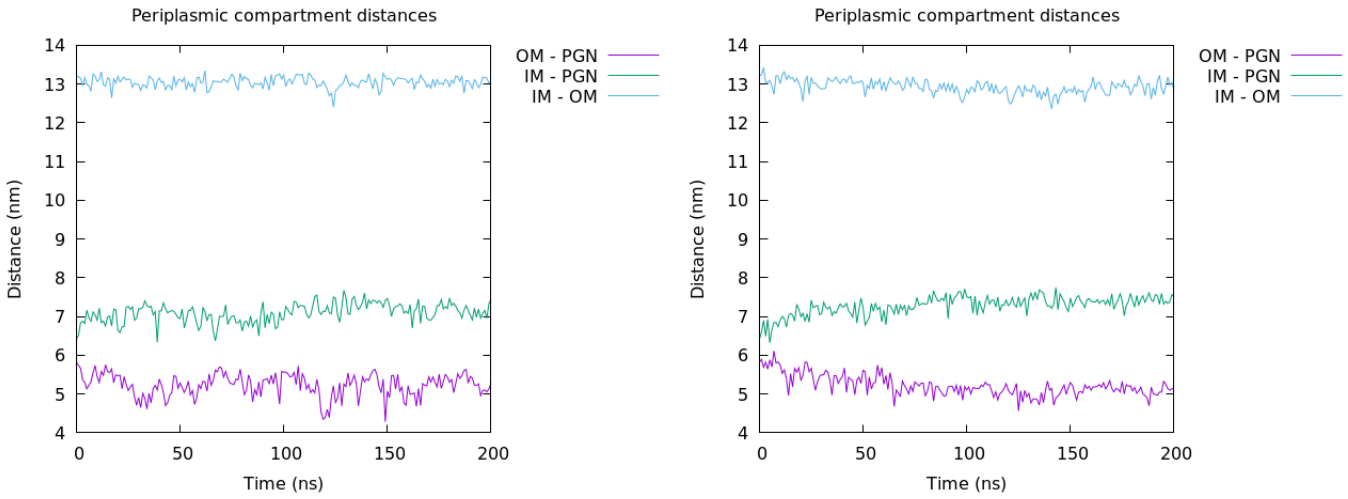


Figure 4.8: Minimum distance between the membranes and PGN in the periplasm over 200 ns, two independent repeats shown (left and right).

With a full periplasm, periplasmic size gained a new definition. As this is a double envelope system, the cell wall divided the periplasm into two separate regions, inner membrane to PGN and outer membrane to PGN. In addition to this the total periplasmic volume was measured in Figure 4.8.

Across 200 ns it was seen that the total periplasmic size does not vary significantly, indicating that the membranes do not drift significantly from each other, shown here where the length perpendicular to both membranes (Z direction) fluctuated close to 13 nm for the duration, seen in Figure 4.8. This total periplasmic size was maintained by the binding of all three protein structures present. This showed that the TolR-PGN periplasmic interaction was strong enough to ensure a stable periplasmic space formed.

Monitoring each segment of the periplasm showed drifting behaviour. This was seen where the IM-PGN distance increased over the first 50 ns and the OM-PGN distance decreased at the same time. Generally this showed that the cell wall during the simulation is moving closer to the OM.

BLP tilting moved the cell wall upwards, and as a result it was seen that the periplasm is not 'cut into halves', based on the simple model here. It indicated that the IM-PGN distance was larger than that of the OM-PGN distance, but that there was no drastic change in either of these distances after 50 ns. Based on this the OmpA, BLP and TolR held the cell wall at an equilibrium position in the *E. coli* periplasmic model.

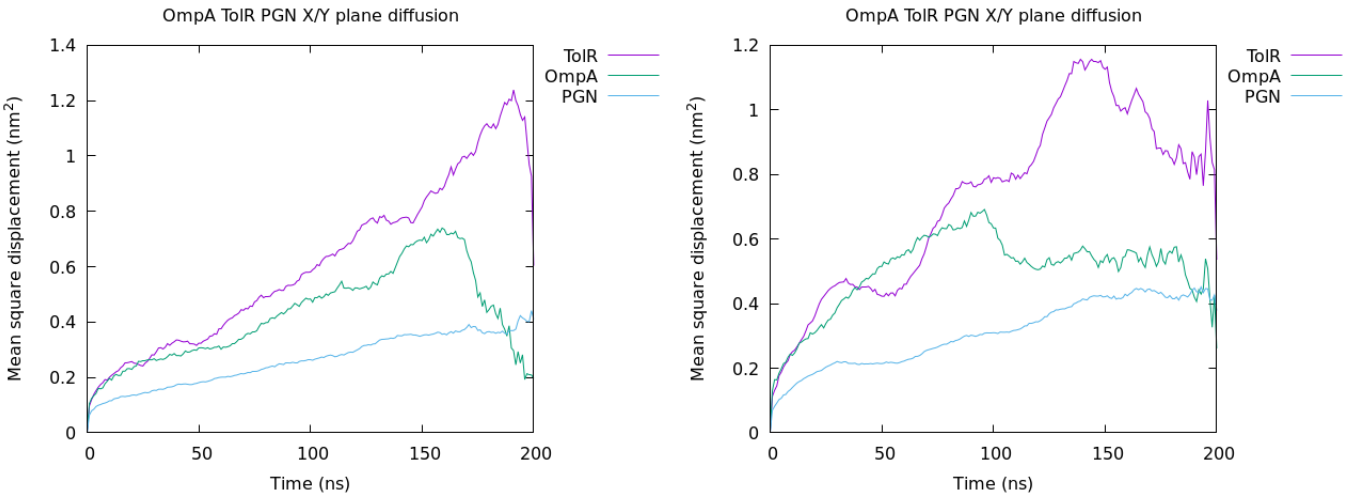


Figure 4.9: Mean square displacement from the initial positions of TolR, OmpA and PGN over 200 ns, across independent repeats shown (left and right).

Displacement perpendicular to the z direction of the OmpA and TolR showed that all three did not generally diffuse across the surface of the PGN. As a result of the binding of OmpA and TolR, the PGN also displayed a low degree of lateral displacement.

In both instances, it was seen that TolR was more mobile than OmpA. This is to be expected due to the smaller size of the unit and a lesser binding affinity for PGN. Comparatively TolR drifted by close to 1 nm^2 in both simulations, whilst the OmpA drifted close to 0.6 nm^2 during simulation. The displacement also showed that back and forth motion of both proteins, where reductions show movement back to the original coordinates.

PGN displacement shows a generally increasing trend, which is to be expected due to the movement of the sheet during simulation, however it was seen that the sheet is much less mobile than either protein present. This indicated that even in the simple model the PGN was relatively static compared to the molecules that are associated with it. In a fully crowded periplasm, it would be expected that the position of the cell wall would deviate even less, but that this microscopic diffusion would still be observed.

Thus, part of the function of PGN binding is to maintain the local geometry of the cell wall, as it is likely that the position of the cell wall in the periplasm is crucial for the function of certain proteins, such as periplasmic transporters.

4.4.4 The closed state of the TolR periplasmic domain

Having studied the open state of the TolR periplasmic domain interacting with the cell wall, a comparison to the closed state was done. These simulations were set up to assess the behaviour of the closed state, alongisde the PGN binding ability of this different form of the TolR protein. In the closed state the flexible uncharged terminal region was buried in the centre of the periplasmic domain. In this closed state, the linkers for the TolR protein were contracted as seen in Figure 4.10, causing the protein to associate with the IM. Simulation of this structure in conjunction with truncated OmpA structures (periplasmic domain removed) were done in order to assess the binding capability of the closed state against the open state.

It can be seen that when placed in the membrane this closed state is associated closely to the membrane, and does not extend as far into the periplasm as the open state. This in turn will affect the binding capability as the closed TolR will not be able to reach across the periplasm to contact PGN. In setup this required that the cell wall was placed closer to the TolR closed model, which requires a smaller IM to PGN distance, hence limiting the size of the periplasm. It logically follows that as many proteins are packed into this space that one would not expect TolR-PGN interactions to occur in such a small volume. Truncation of the OmpA, to weaken OM-PGN association, whilst maintaining a BLP connection shows that the cell wall will move upwards, away from the TolR as the BLP pulls the cell wall higher and out of interaction range.

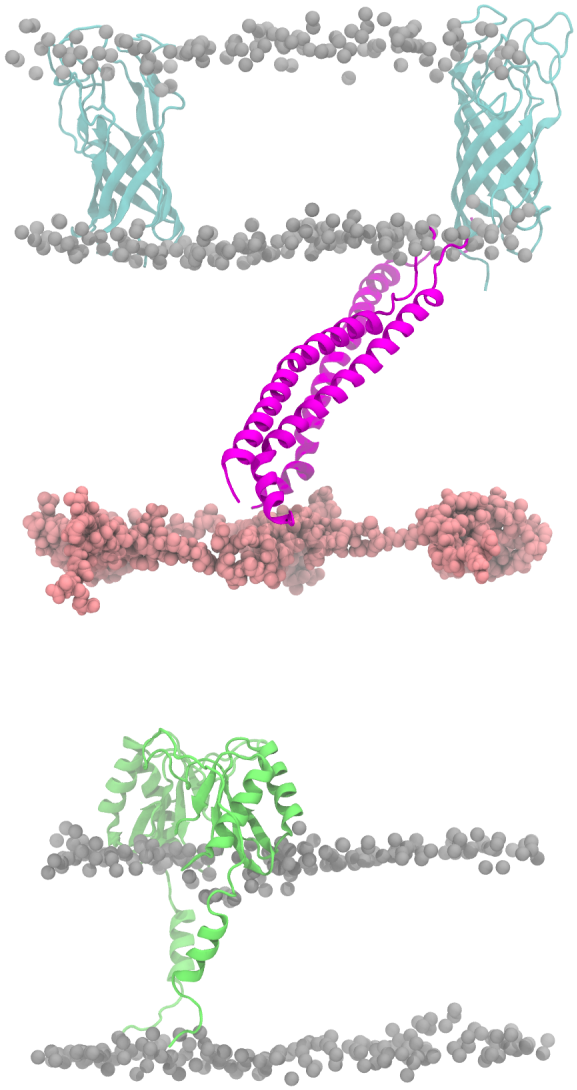


Figure 4.10: Closed TolR (lime green) does not contact the PGN mesh (pink) after 200 ns, in the truncated OmpA (grey) systems, where BLP is seen to bend (magenta).

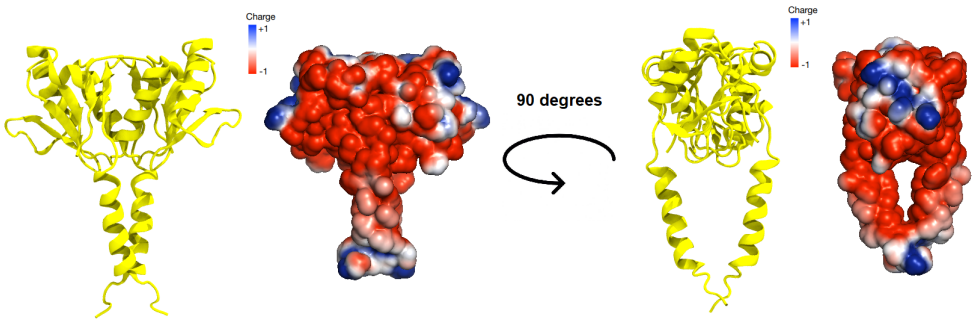


Figure 4.11: Closed TolR (yellow) is shown straight on and at a 90 degree rotation, where the electrostatic profile of the protein is also indicated, where blue is positively charged, red is negatively charged and white is neutral.

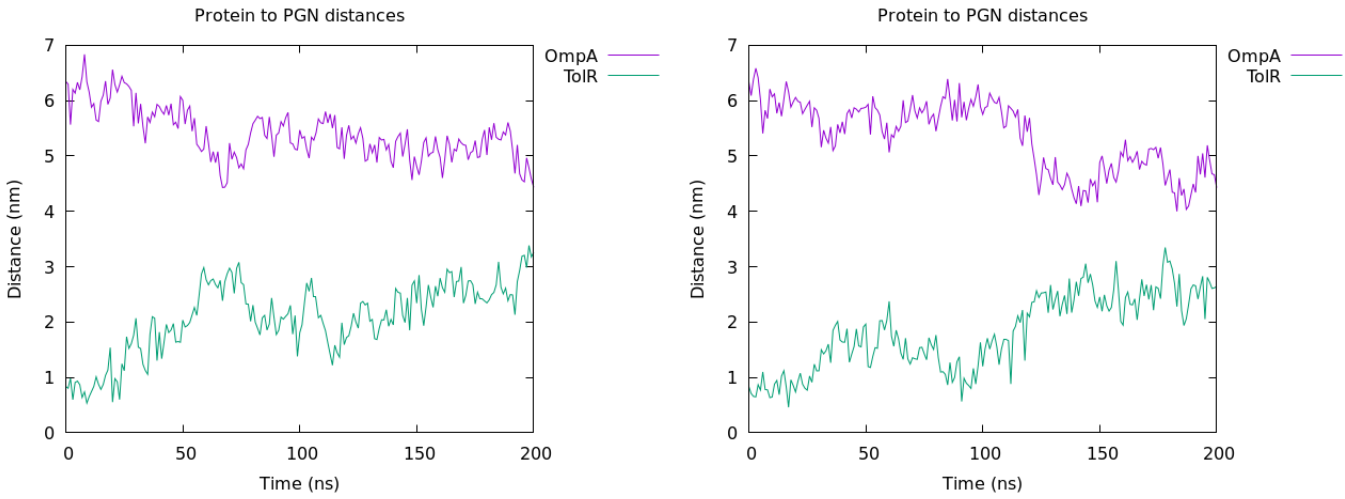


Figure 4.12: Minimum distances between OmpA to PGN and TolR to PGN over 200ns for closed TolR, independent repeats shown (left and right).

It was observed here that when the PGN sheet was placed 1 nm from closed TolR, PGN binding did not occur; this can be seen via the threefold increase in distance between TolR and PGN in Figure 4.12. Distance to the truncated OmpA was measured to track cell wall movement, where this distance was reduced by a similar amount. Without the exposure of the terminal binding region the TolR did not associate with PGN, as the BLP pulls the cell wall out of binding range, where the TolR closed conformation cannot extend to contact PGN.

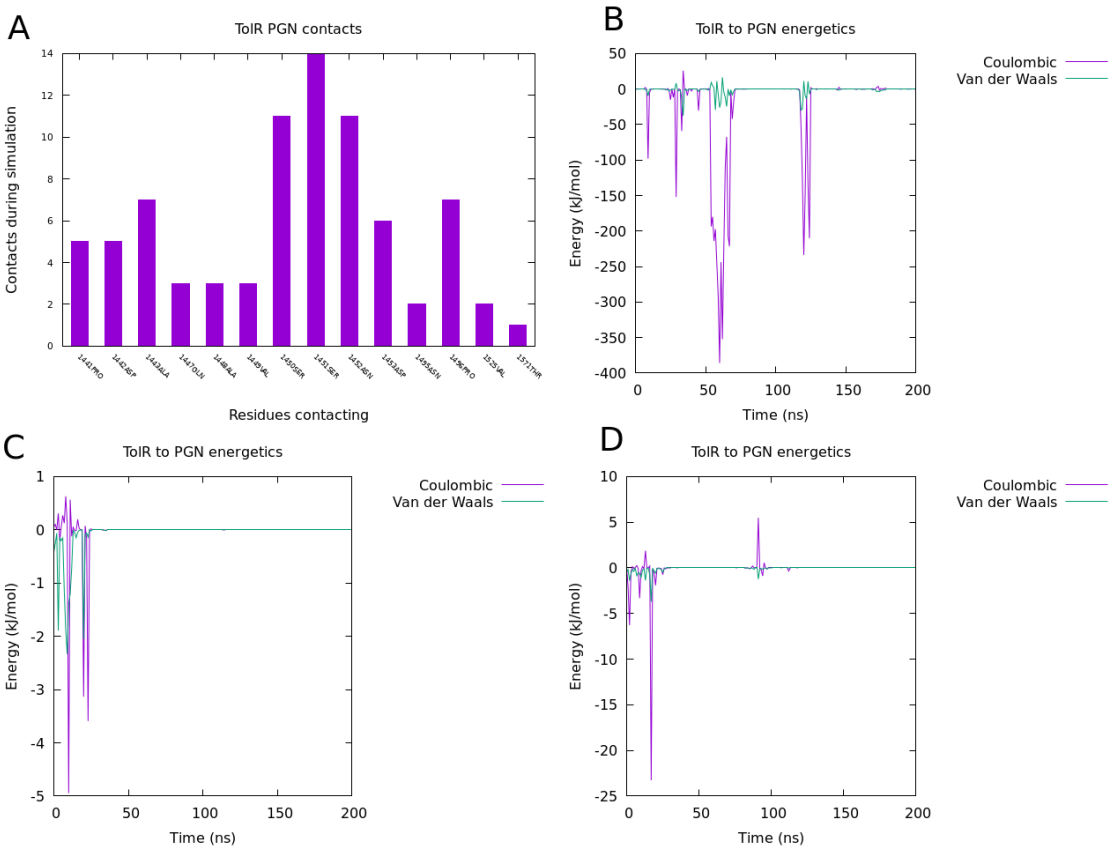


Figure 4.13: TolR to PGN contacts shown for the highest contact system over 200 ns (A). Short range electrostatic and Lennard-Jones interactions for TolR-PGN over 200 ns are shown (B, C and D) indicating little contact.

This lack of contact is shown in Figure 4.13, where the highest amount of contact that was achieved over simulation came from the serine residue exposed to the PGN mesh at the start of simulation, where this residue only contacted PGN in 7% of the trajectory. Energetics confirmed this lack of interaction, where there was generally little interaction, as seen in Figure 4.13. The interaction between closed TolR and PGN was not strong enough to prevent BLP induced cell wall height increase in the periplasm. In the one case where there was significant interaction, it was seen that the coulombic energies were attractive, as opposed to the open TolR where generally the initial charged interaction was repulsive, however these interactions were temporary as the cell wall eventually moved out of significant interaction distance.

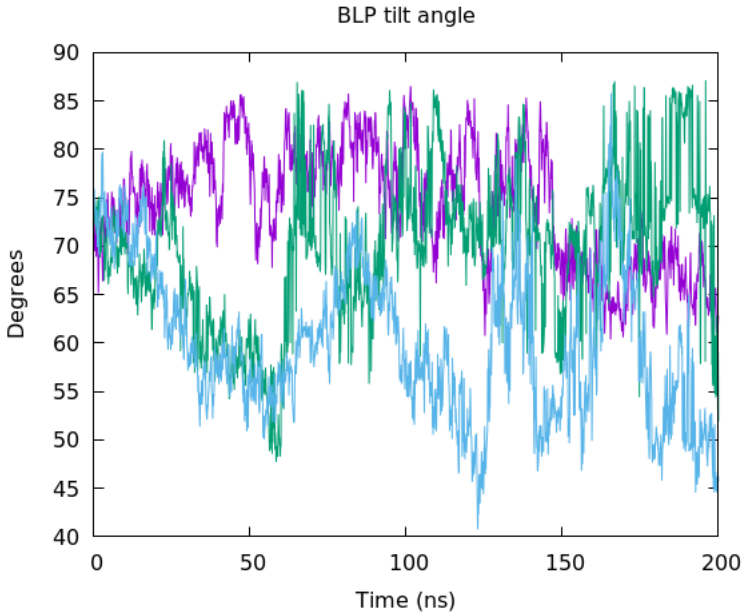


Figure 4.14: BLP tilt angle using the vector of the BLP against the surface of the OM over 200ns, across independent repeats.

Measuring BLP tilt in the TolR closed state system confirmed this; there were large deviations in the tilting angle throughout the simulations, indicating that the BLP was pulling the cell wall towards the OM, see Figure 4.14. After 200 ns all results showed that there has been a decrease in BLP tilt angle. The cell wall during simulation was only interacting with BLP. In two instances it was seen that during the first 50 ns of simulation there was a large decrease in the tilt, from 70 - 50 degrees, over this period was also the largest change in protein-PGN distance, see Figure 4.12, indicating the tilt and cell wall movement are related.

The cell wall was positioned close to the surface of closed TolR at the start of simulation. In two simulations this decrease indicated that the BLP pulled PGN out of range almost immediately, indicated by the two systems that experienced close to 0 kJ/mol charged or VdW interaction. In the other system, BLP did not tilt to a low angle in the first 50 ns. Tilt increased here to 85 degrees indicating a straightening of BLP. This system showed a higher amount of contact and attractive energies in the hundreds of kJ/mol. In the second half of this simulation, tilt eventually decreased below the initial value, such that little to no interaction was seen in the last 50 ns.

From this it is proposed that two factors prevented TolR-PGN association

in the closed state. Firstly TolR can stably associate with the outer leaflet of the IM, preventing any extension of the linker regions, therefore TolR did not have the same capability to contact the cell wall at larger distances from the IM. Secondly the buried flexible termini do not allow the VdW stabilising energy to form, which was seen to be almost non existent in all three systems.

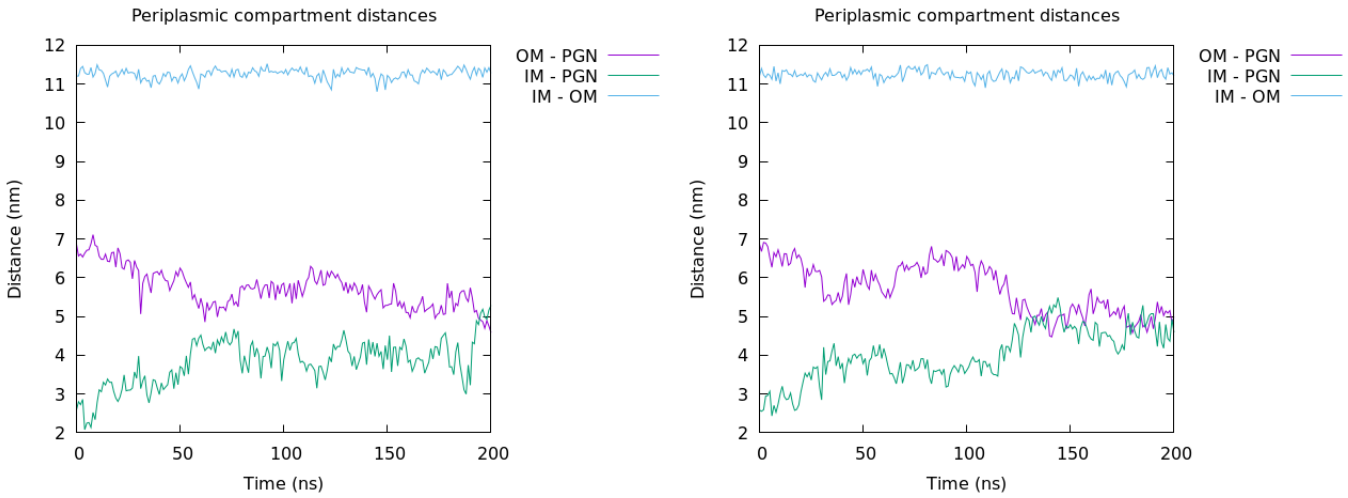


Figure 4.15: Minimum distance measurements between the membranes and PGN in the periplasm over 200 ns with closed TolR, independent repeats shown (left and right).

Similarly to open state TolR, the periplasmic compartment distances were characterised. In the absence of TolR binding it was seen that the cell wall essentially found an equilibrium between the two membranes, where after 200 ns it was equidistant from both, as in Figure 4.15.

It should be noted here that the full periplasmic distance was smaller than in the open TolR simulations, due to the need to locate TolR close to PGN initially, hence the IM-PGN distance was initially smaller than OM-PGN distance, 11.5 nm compared to 13 nm. The BLP forced a minimum starting distance for OM-PGN, generally close to 7 nm, as the BLP was inserted in a close to perpendicular orientation to the membranes and PGN.

Unlike with open state TolR, the periplasmic distances were not maintained, which is an indication that TolR and OmpA binding stabilised the cell wall position, considering that these systems lacked either interaction. Generally the distances for either membrane to PGN eventually were equal, at 5 nm. The periplasm fluctuated around the initial value of 11.5 nm, however there was no significant change in the full periplasm size after 200 ns. This showed that over

these timescales the periplasm held together, regardless of the lack of IM-PGN interaction. However at physiological timescales, it is likely that diffusion of the membrane perpendicular to the surface of the membrane would occur.

4.4.5 Truncation of the OmpA periplasmic domain

Following the comparison of the wild type and truncated OmpA closed TolR, the open state of TolR was studied with OmpA CTD truncation. It can be seen here that open TolR did bind PGN, in the absence of the OmpA periplasmic domain. This is the point of the truncation, where the OmpA periplasmic domain is not present to pull the cell wall towards the OM. In this system the OM-PGN association was maintained via BLP. Once again the open state was positioned close to the cell wall, and contact ensued. It was seen from the density plot of the system that the TolR once again was enmeshed in the cell wall during the simulation, as in Figure 4.16.

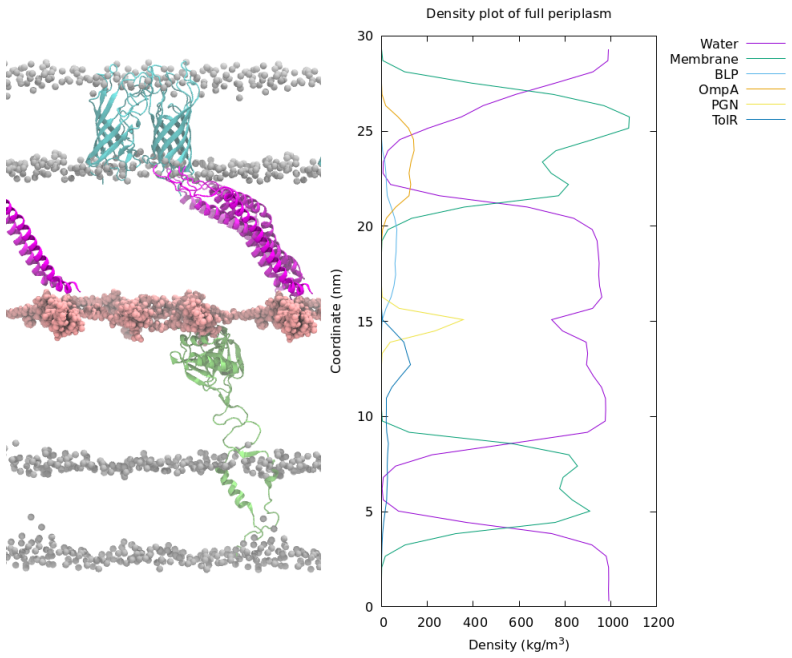


Figure 4.16: A visualisation of the OmpA truncated (grey cartoon), wild type TolR system (lime green cartoon), containing BLP (magenta cartoon) and PGN (pink spheres) with membrane headgroups (grey spheres), shown (left). A density plot of the system in the z-axis over 200 ns (right).

It can be seen that the TolR linker region was in fact not full extended, as

it was in the wild type simulation, when looking at Figures 4.2 and 4.16. Comparing the density of both open TolR states across systems, with the OmpA clamp present it was seen that the TolR density spanned a distance of 12.5 nm, relative to 10 nm in this case. As open state TolR can bind the cell wall, whilst the clamp was bound, this reduction in linker length reduced the distance TolR covers. This is attributed to the lack of OmpA binding, where OmpA and TolR contraction were opposing forces, the two proteins reached an equilibrium distance where TolR was fully extended. Here the OmpA force was absent, allowing TolR contraction, Figure 4.16.

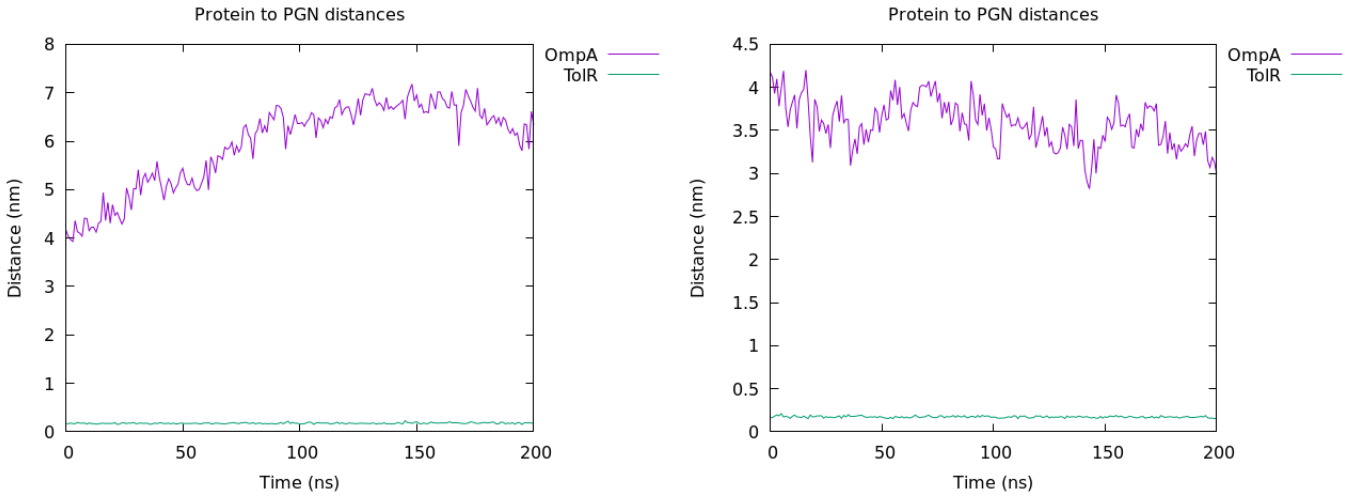


Figure 4.17: Minimum distance between OmpA and TolR proteins to PGN over 200ns, independent repeats shown (left and right).

Analysis of protein-PGN distances shows that the TolR was contacting the PGN, where a distance of less than 0.5 nm was maintained throughout the simulation, indicating that once TolR is placed close to the PGN, it does not unbind over the simulation, as in the wild type simulations, Figure 4.17.

The OmpA barrel to PGN distance increased over 200 ns in one run, whilst in the other there was a decrease in distance between OmpA and PGN, Figure 4.17. This indicated that in certain cases the TolR pulled the cell wall downwards, towards the IM, or it remained in the extended conformation. In the case of contraction it can be seen that the there was roughly 50 % increase in the OmpA-PGN distance.

This contraction was most likely due to the strength of the interaction TolR experiences with PGN. It has been shown that the OmpA dimer strongly bound the cell wall, and that this binding generally resulted in linker contraction of

the protein, as seen in Chapter 3. TolR-PGN binding should produce a similar result, however in the wild type systems it was seen that short range interactions were significantly weaker than OmpA-PGN, essentially rendering the periplasmic domain as a "weaker" clamp to the cell wall.

It is hypothesised that simultaneous binding of both proteins produced a less favourable binding environment for TolR, hence the truncation of OmpA, to study TolR binding alone.

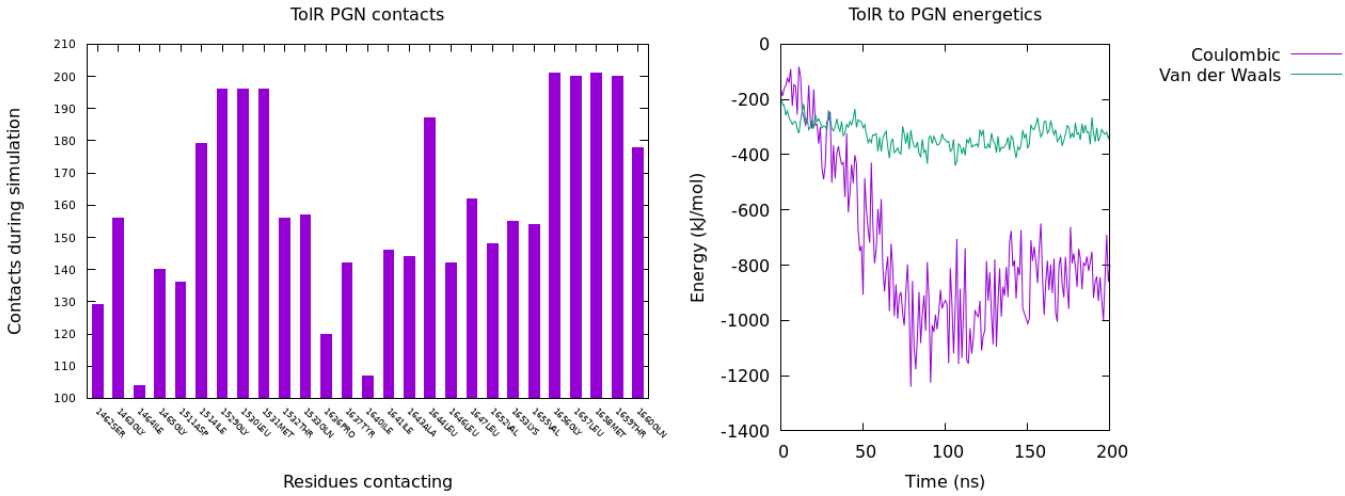


Figure 4.18: TolR to PGN contacts per residue for the first independent trajectory of the system, where TolR is in high contact with the PGN over 200 ns (left). Short range electrostatic and Lennard-Jones potentials for the binding of TolR over 200 ns (right).

The contact can be split into two separate regions, where 1462SER - 1533GLN represents binding in the first protomer and 1636PRO - 1660GLN represents binding in the second protomer, where these regions were present on the TolR termini, Figure 4.18. The highest contact was established in the termini residues GLY-LEU-MET in both protomers (1529-1531 and 1656-1658 respectively) where contact occurred for over 90% of simulation time, indicating that these residues were essentially always in contact with PGN. Residues upstream and downstream from this region also made a high amount of contact. Interestingly out of all contacting amino acids, only two were charged, 1511ASP and 1653LYS, confirming that relative to OmpA, there are less charged residues in the clamp that interacted favourably with PGN.

Comparing the protomers that bind, the second protomer of the dimer interacted more with the PGN. There were a greater number of amino acids that

had over 140 contacts during the simulation, where the residues with the highest contact were all located on this protomer. This was the first observed instance showing that this binding was highly favourable, where after 100 ns a fluctuating stabilising charged energy of ∼-1000 kJ/mol was established, whilst the usual stabilising VdW energy, -300 kJ/mol was present, seen in Figure 4.18. This system displayed the linker contraction, pulling the cell wall closer to the IM.

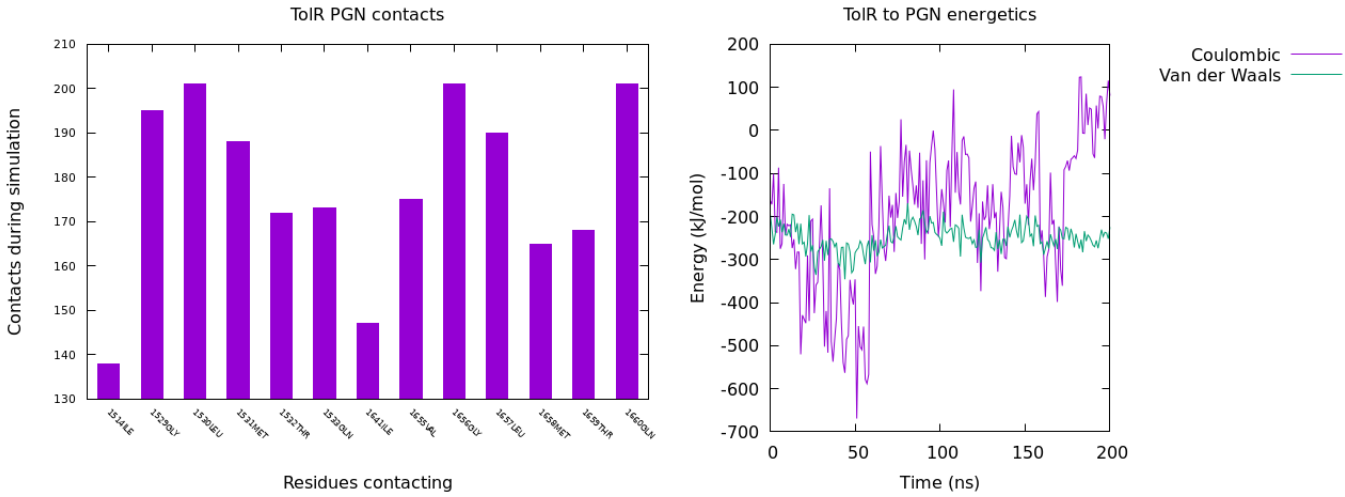


Figure 4.19: TolR to PGN contacts per residue for the second independent trajectory of the system, where TolR is in contact with the PGN over 200 ns (left). Short range electrostatic and Lennard-Jones potentials for the fluctuating binding of TolR over 200 ns (right).

This second independent repeat simulation showed less binding, where both protomers showed only 6/7 residues with over 100 contacts to PGN, Figure 4.19. Once again the previously mentioned sequence in the TolR termini showed the greatest contact to PGN, where the regions of highest contact in each protomer differed. This indicated that the TolR dimer experienced a degree of asymmetry in binding to the PGN mesh model.

Analysis of energetics of this showed that the overall interaction was attractive, however in this case the coulombic energy was attractive for the first 100 ns, after this there was an increase, where this energy trended towards repulsion. Charged interaction were measured, with a minimum of ∼ -700 kJ/mol and maximum of ∼ 100 kJ/mol, the initial and final charged energies were different, where after 200 ns the energy has increased by 300 kJ/mol.

Considering this contrast to the first simulation, but that interaction with

the cell wall was maintained, it is indicated that PGN binding was initially caused by the snorkelling termini in this case, but to achieve strong binding similar to that of the OmpA dimer, closer association to the rest of the dimeric unit was required, where in this instance the rearrangement of the TolR dimer over time weakened the binding.

In this instance there was a contraction in the TolR linker, such that the cell wall moves toward the IM, indicating that even though this binding is less favourable than the first instance, the TolR can still pull the cell wall through the periplasm.

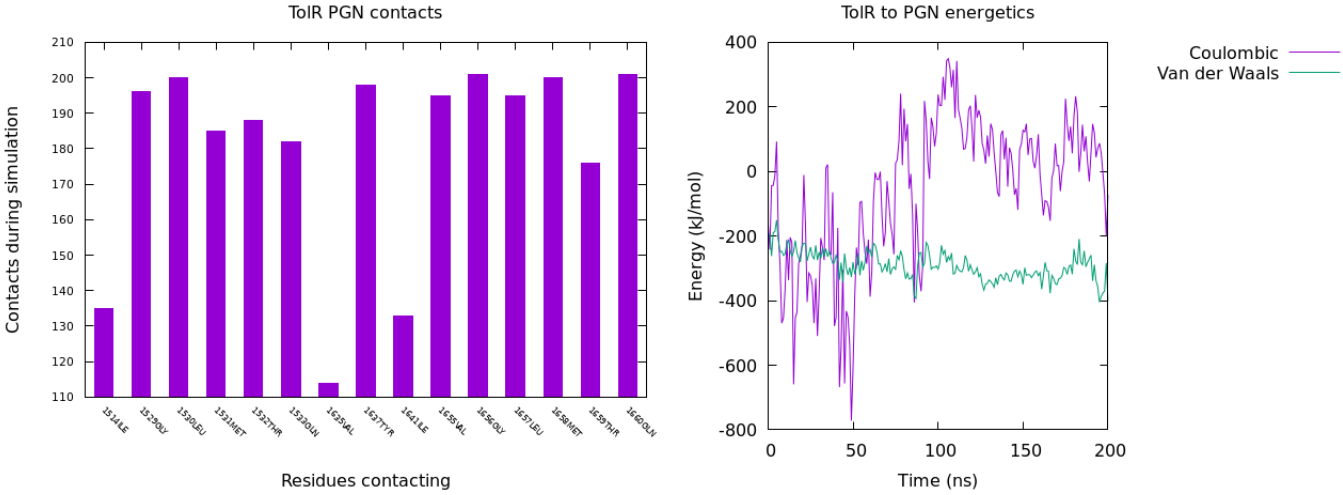


Figure 4.20: TolR to PGN contacts per residue for the third independent trajectory of the system, where the TolR is in contact with the PGN over 200 ns (left). Short range electrostatic and Lennard-Jones potentials for the binding of wild TolR over 200 ns (right).

Thirdly, the binding of TolR here indicated that there were a greater number of residues that cross the contact threshold, six amino acids on protomer one and nine on protomer two, Figure 4.20. The same regions as stated previously showed a large degree of contact, where these terminal residues were interacting frequently, as in the first simulation.

Contact analysis for all three simulations shows that greater association did not necessarily infer stronger binding. In the first case, there were charged residues on TolR interacting with the PGN, which lends greater stability to binding, which is not seen otherwise. It is clear that the terminal region contacting the PGN via hydrophobic residues helped to establish the initial contact, but if the correct association with the rest of the dimer was not achieved then

this interaction will become unfavourable. Based on this it is likely that TolR was in equilibrium between bound and unbound states.

The energy displayed here was the least attractive of all three, where the highest value comparatively for coulombic energy was seen after 100 ns, and that the total interaction became net repulsive at certain times. Here it was seen that the coulombic energy is attractive for the first 100 ns, reaching a minimum of $\sim$ -800 kJ/mol and a maximum of $\sim$ 350 kJ/mol.

As shown in Figure 4.20, regardless of TolR-PGN distance that it was possible for repulsive coulombic energy to interfere with the overall binding of the TolR periplasmic domain to PGN. Contact analysis indicates there is selective binding between TolR, where certain charged residues and a minimum interface must be reached to achieve consistently stable binding.

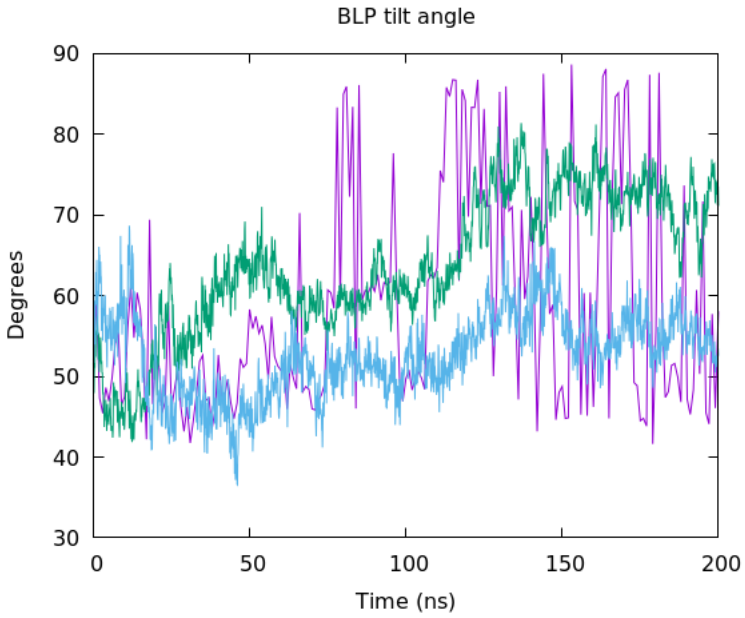


Figure 4.21: BLP tilt angle using the vector of the BLP and the surface of the OM over 200 ns for truncated OmpA and wild type TolR, independent repeats are shown.

BLP tilt is an indication of cell wall movement, where in the case we see that two simulations result in changes in PGN location. Based on three varying binding strengths of TolR, three varying tilt patterns were also seen for BLP in separate simulations, Figure 4.21.

Firstly, when binding was the strongest there were large variations in tilt angle, starting at 50 degrees and generally trending towards higher values at 85 degrees. This is attributed to the contraction of the TolR, indicating that the BLP was experiencing the most disruption, where the fluctuations to an almost 90 degree angle show that the BLP was being pulled into a straight orientation.

Secondly, where there was binding that was attractive, but significantly weaker, it can be seen that there was a consistent increase, where the initial angle of 50 degrees ends at $\sim$ 70 degrees, indicating that the cell wall was being pulled, but less so than the first simulation.

Thirdly, where binding at times was repulsive, there was fluctuation in BLP tilt however with respect to the initial and final values there was a decrease in tilt angle. This is indicative that the BLP was controlling the cell wall movement entirely at points, considering the weak TolR clamp interactions.

4.4.6 Removal of BLP from the envelope

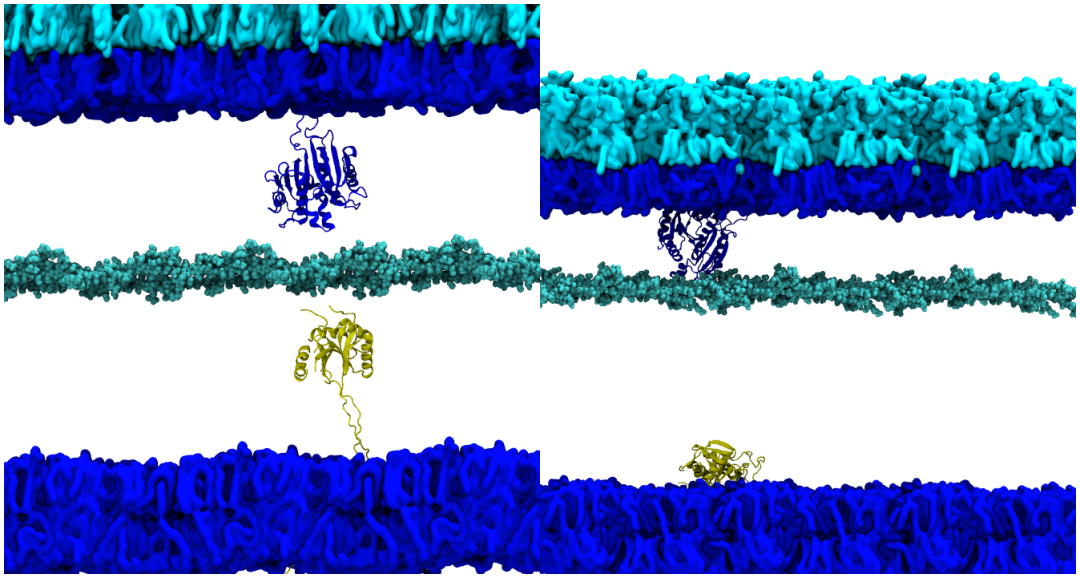


Figure 4.22: Visualisations of the TolR (yellow), OmpA (blue) and PGN (cyan) periplasm without BLP at 0 ns (left) and 200 ns (right).

Finally the removal of BLP from the periplasm once again changed the behaviour of the entire system. There was no longer a structural staple that maintained the cell wall position. Cell wall location was entirely decided by the

short range interactions of TolR and OmpA.

Based on data thus far it would then be expected that OmpA binding would dominate, such that the cell wall would generally be located close to the OM. It is seen that this was possible, where without BLP present in the system there was no size maintenance for the OM-PGN distance, meaning that OmpA binds the cell wall, the linker region contracted and pulled the cell wall to the OM. In this scenario the TolR could not extend far enough to bind PGN, causing linker contraction in the inner membrane protein, where the protein associated with the IM.

Comparing this to previous visualisations of the full and half periplasm models in Chapter 3 and Figures 4.22 and 4.16, it can clearly be seen the large effect BLP had on PGN and where the periplasm was divided into separate components.

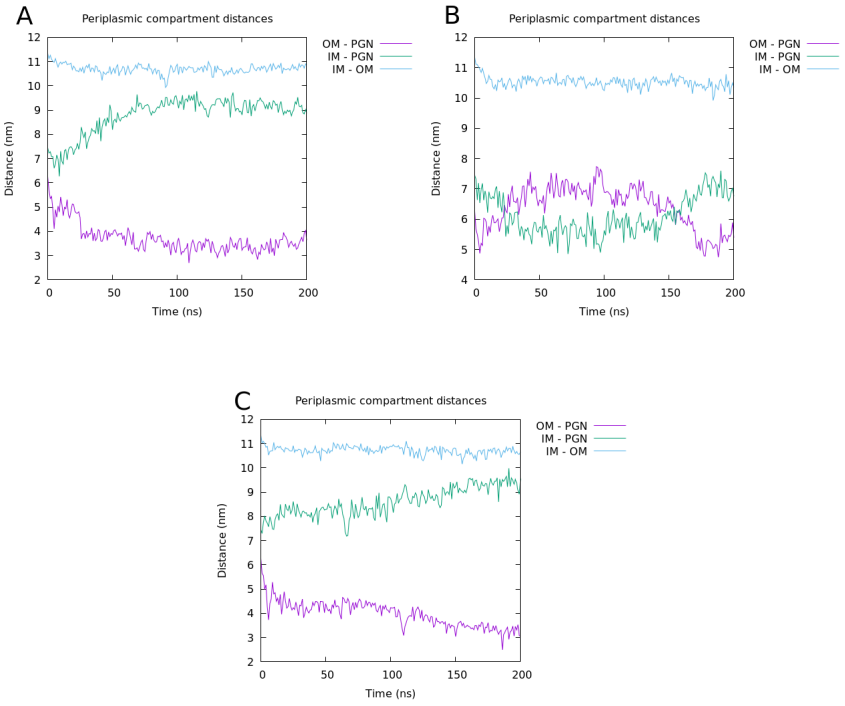


Figure 4.23: Minimum distance measurements between the membranes and PGN in the periplasm over 200 ns with closed TolR, independent repeats shown (A, B and C).

Analysis of periplasmic distances show that without BLP the cell wall moved drastically, as in Figure 4.23. Overall in the periplasm, slight decreases in the

size of the system were seen, but generally periplasmic size was well maintained.

In two cases OmpA could bind PGN and this caused a large change in the periplasm, where the cell wall moved close to the OM, away from the IM. In the other, it can be seen that the distances between the membranes and PGN fluctuated, but generally were the same after 200 ns. This effect was caused by a lack of binding of either protein, where the cell wall was not pulled in either direction, Figure 4.24

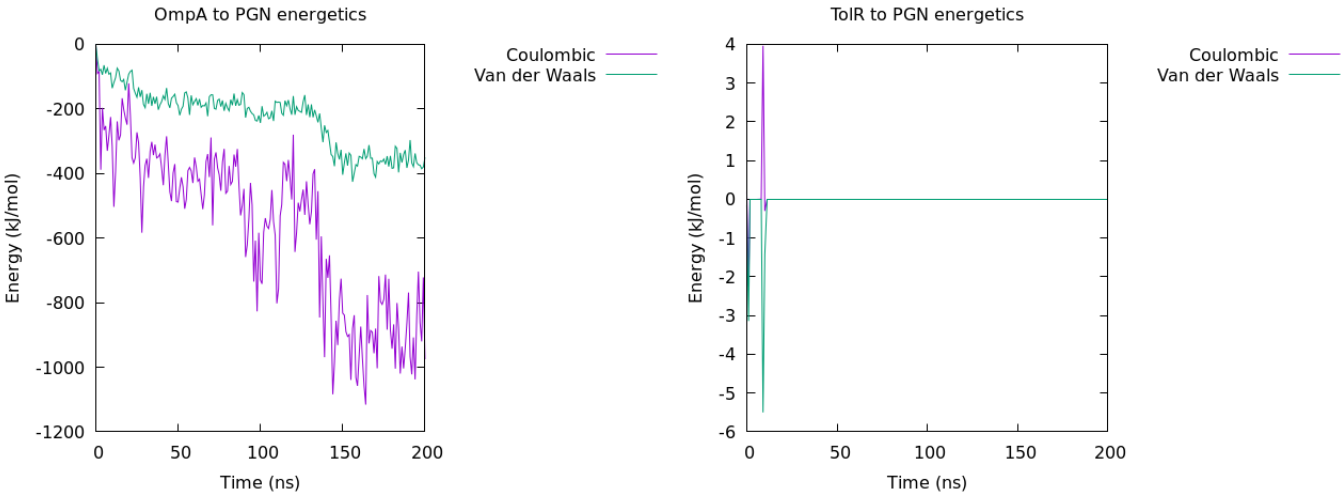


Figure 4.24: Short range electrostatic and Lennard-Jones interactions for OmpA-PGN (left) and TolR-PGN (right) over 200 ns without BLP - simulation 1.

As energetics are now well characterised in multiple instances for both TolR and OmpA, these serve as an indicator for the interaction that occurred each system.

Firstly, where the PGN-OM distance dramatically reduced, the standard binding for the OmpA dimer to the cell wall was observed, where both forces were attractive, indicating that the OmpA was bound to the cell wall and the linker size reduction has moved PGN to the OM. This binding was established quickly and maintained throughout 200 ns, Figure 4.24.

In contrast to this the TolR-PGN interaction was essentially zero, where there was minuscule interaction, which generally was net 0 kJ/mol, and after 10 ns, there was no TolR-PGN association for the remainder. In the related distance plot, a sharp decrease in OM-PGN distance was observed as at this point the cell wall was pulled out of interacting range from the open state TolR

and did not return, Figure 4.23.

Once again, this displays the role of BLP; the cell wall structure is maintained as a result of lipoprotein covalent bonding, slowing down movement and preventing drastic movement perpendicular to the membranes, allowing simultaneous TolR and OmpA association to the cell wall.

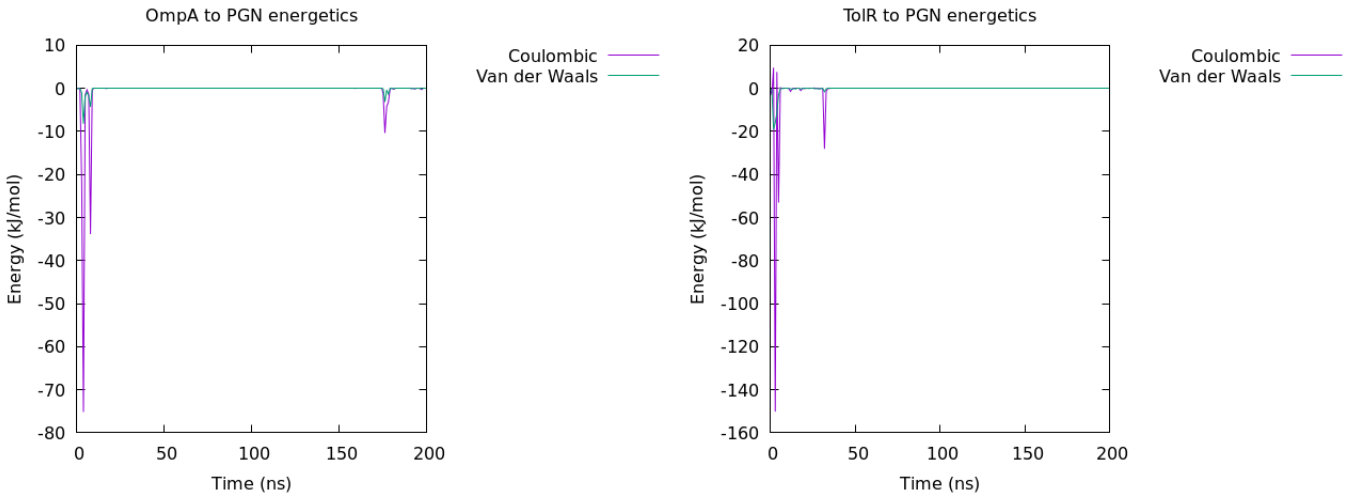


Figure 4.25: Short range electrostatic and Lennard-Jones interactions for OmpA-PGN (left) and TolR-PGN (right) over 200 ns without BLP - simulation 2.

Secondly, in the system where the cell wall fluctuated but did not consistently move in one direction, there was little to no association from either protein. Both proteins have been shown to be able to bind PGN, even when energetics show that the binding was weak or became unfavourable, however in this case the cause is that the cell wall was outside of interaction range.

In this system it is shown that even after 150 ns OmpA could potentially interact with PGN, as a very small attractive trough was seen, at -10 kJ/mol, but this was not enough to allow the protein to bind. In this case both proteins appeared to adopt an orientation that prevents binding, where the OmpA did not have BLP to attract the cell wall or use as a scaffold to bridge the OM-PGN gap.

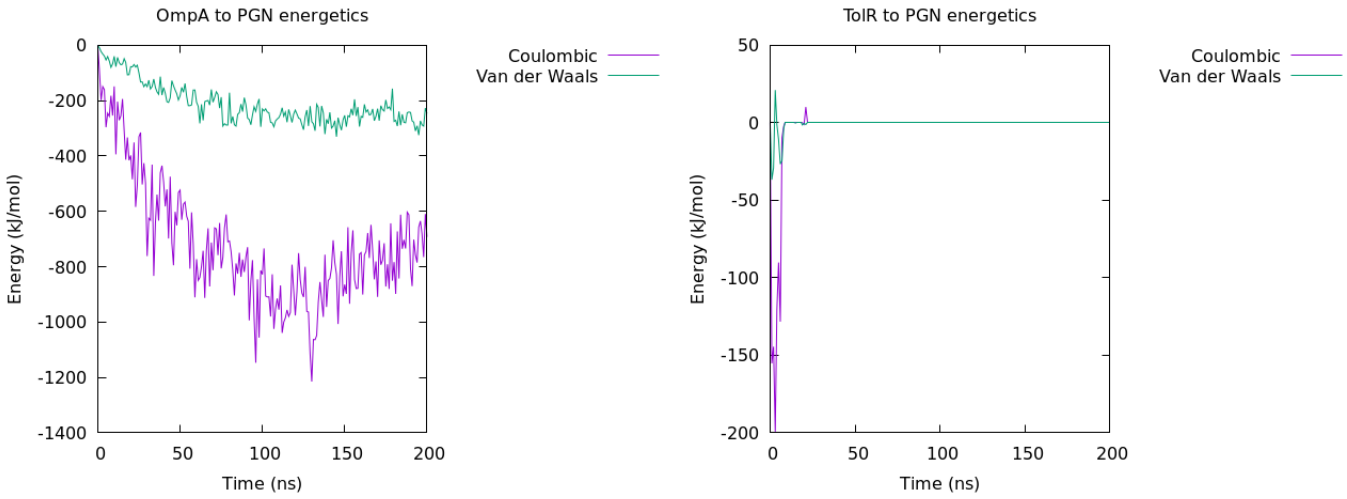


Figure 4.26: Short range electrostatic and Lennard-Jones interactions for OmpA-PGN (left) and TolR-PGN (right) over 200 ns without BLP - simulation 3.

Thirdly, it can be seen that the OmpA bound strongly to PGN. Figure 4.26 indicates that the bound OmpA reached what appears to be the equilibrium binding more quickly. Once this energy was reached there was an increase, to above -800 kJ/mol. For TolR it was seen that there was an initial interaction but that this fell off, for the rest of the simulation after 20 ns there was no interaction with PGN. This was an indication that PGN was outside of interaction range, preventing any association with TolR.

Overall, it has been shown that without the presence of BLP, there was little to no contact between open state TolR and PGN. The larger slower moving OmpA dimer showed a similar degree of interaction when compared to previous simulations in Chapter 3, where dimeric binding was possible and generally was favourable. In the instance where there was no binding and significant PGN displacement, BLP removal caused the cell wall to fluctuate in the periplasm, but without structural support the cell wall was not a good binding target for the periplasmic domain of TolR.

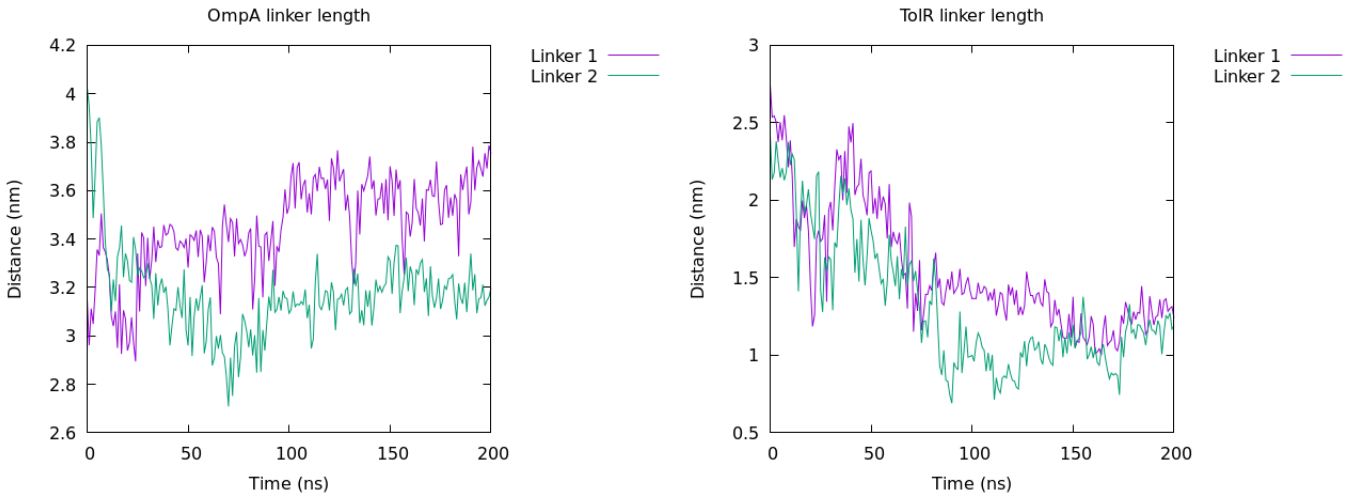


Figure 4.27: Minimum distance between the start and end of the linker for each protomer in OmpA (left) and in TolR (right) over 200 ns - simulation 1.

Linker region behaviour has been broken down in a similar manner. In the first system it can be seen that the OmpA linkers changed length differently, there was a swap between the length of both linkers, where linker 1 extended and linker 2 contracted. This expansion and contraction was generally expressed in a 1 nm range, in Figure 4.27.

TolR experienced a contraction of both linkers, indicating that the periplasmic domain relocated to the surface of the inner membrane. This contraction was a reduction of 1.5 nm in end to end distance of the linkers. This measurement shows that without a driving force, such as binding, to keep TolR extended, the open conformation contracted.

As OmpA is shown to bind, considering the linker distances this would indicate that the dimer did not bind with the entire face of the clamp flat to the cell wall. This has been observed previously and in this case this behaviour can be seen, where as linker 2 was the shortest after 200 ns, the second protomer will be in more contact with the cell wall than the first.

It is observed that the closest protomer will bind first, and that after the binding is established the linker on the opposite protomer will behave in an opposite fashion. Using this instance as an example, linker 2 was furthest from the OM, so closest to PGN, where after binding occurred a contraction in this linker occurred, pulling PGN up, whilst the other, linker 1, expanded to facilitate more PGN contact in the first protomer.

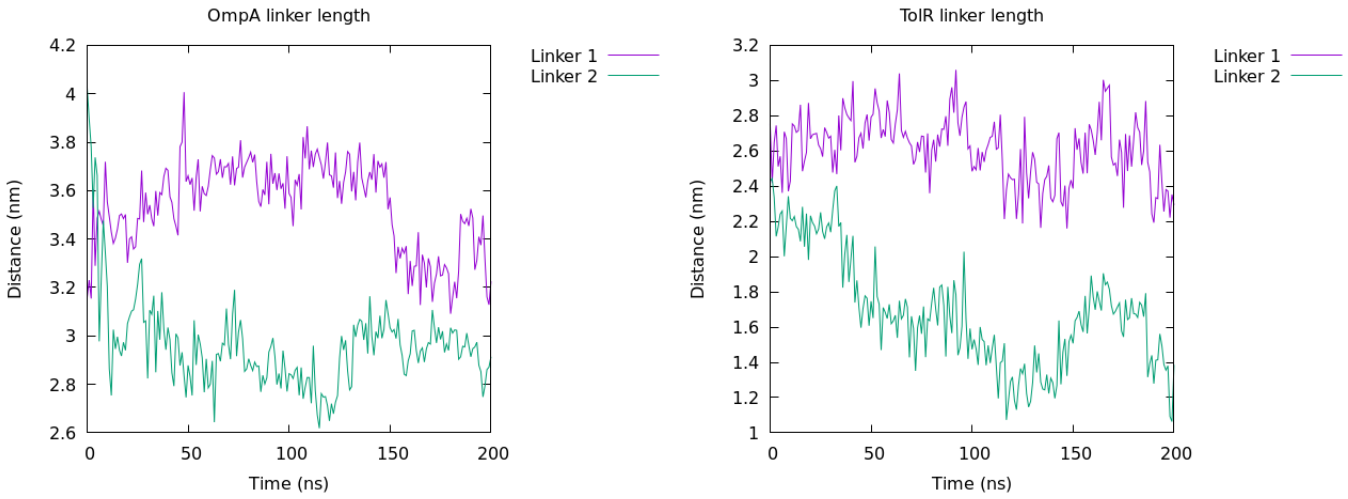


Figure 4.28: Minimum distance between the start and end of the linker for each protomer in OmpA (left) and in TolR (right) over 200 ns - simulation 2.

With respect to the system where there was almost zero binding of either protein, OmpA showed a pattern similar for linker 2, where the same contraction was seen, however linker 1 did not display consistent expansion. After 100 ns there continued to be significant change in linker length, indicating the dimer was not bound and that the linker regions were not influenced by PGN interaction.

TolR in this system displayed different behaviour. Linker 1 remained extended for the simulation, unlike in the first, where both linkers contracted. The second linker contacted, shrinking by 1 nm, however this was not a consistent reduction as there was an increase back to 1.8 nm from 1.2 nm within a 50 ns window. This fluctuation indicated free movement, where the TolR remained extended, compared to the previous simulation, where both linkers behaved in an almost identical fashion. This indicates that PGN binding was not obligatory for open TolR to remain extended.

The different TolR linker lengths indicate that the TolR was not in the extended straight form, but was bent, as one linker pulled the periplasmic clamp to the side as a result of contraction. Orientation of the clamp as result can affect favourable binding to PGN, Figure 4.28. This effect was also observed in OmpA.

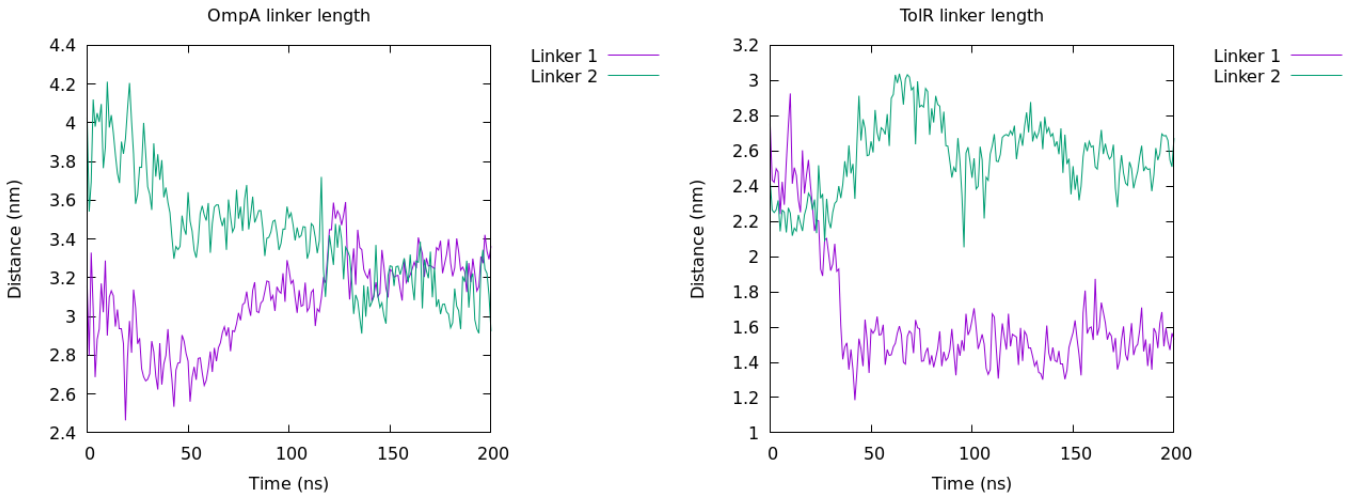


Figure 4.29: Minimum distance between the start and end of the linker for each protomer in OmpA (left) and in TolR (right) over 200 ns - simulation 3.

Where the quickest OmpA-PGN association was observed both linkers in OmpA converged to the a similar distance. In this instance this indicated a similar amount of binding per protomer present. This correlated well with the energetics, indicating that OmpA binding was either achieved by approach of one protomer, then binding of another, or that simultaneous binding did occur.

TolR displayed linker contraction in linker 1 and expansion in linker 2. This profile once again suggests a tilted conformation, where one protomer was much further from the membrane helices compared to the other.

Comparing the two proteins in general it can be confirmed that the TolR linker was shorter than OmpA in the periplasmic solution. Generally the longest linker region in TolR was close to 2.5 nm, whereas in OmpA the linker was frequently extended to a distance of 4 nm. TolR contracted to much smaller distances than OmpA, where the linker regions were shown to be able to contract to less than 1 nm, whilst the OmpA never contracted to a distance of less than 2.5 nm, Figure 4.29.

Based upon this it is proposed that OmpA extended further into the periplasm and that the longer linker region was key in PGN binding, where without BLP the binding will easily occur, compared to TolR which never bound in these simulations.

4.5 Discussion

A simple *E. coli* periplasmic model has been constructed and was stable for 200 ns of simulation. This is the first atomistic representation in a recent wave of periplasmic simulations where due to technological improvements this whole compartment can now be studied. The simple periplasmic model allows protein to cell wall modelling and exploration of these interactions.

The open TolR dimer structure and OmpA dimer can bind PGN at the same time, when BLP is present. In many cases it was seen that the open TolR can approach the cell wall and enmesh with PGN interacting attractively according to analysis of energetics and minimum distances from protein to PGN.

TolR C-termini are key to facilitating the initial binding interaction. The five terminal residues in the TolR open conformation are capable of threading into the cell wall and causing the initial binding event between PGN and TolR. These flexible regions are not charged and can easily change shape in order to attach to PGN. With respect to TolR the structure of the termini is similar between bacterial species, indicating that this region is a necessary component, where it is proposed to be involved in PGN binding.[59]

TolR exists in two conformations, open and closed. Open TolR is generally a weaker binder to PGN than OmpA. Analysis of energetics of the TolR-PGN association indicate that in the majority of simulations the overall interaction that is observed is always attractive, but that the coulombic energy involved is highly variable, generally fluctuating between attractive and repulsive. It is proposed that the exposed surface of the TolR clamp must achieve a specific orientation to achieve favourable coulombic interaction.

Extensive TolR interaction requires interplay of charged residues and other residues surrounding the termini interface. This is seen when during simulation, other segments of the periplasmic clamp, particularly charged residues come into contact with PGN. When only the termini interact throughout the simulation, a weaker association is observed. The VdW energy involved in this binding interaction is consistent, however the only case where charged energies are similar to the attractive OmpA dimeric profile are when aforementioned charged residues that are downstream from the termini are involved.

Closed TolR did not bind to PGN in truncated OmpA systems. The closed conformation is organised such that the flexible termini are buried. This then removes the initial interaction that causes TolR-PGN binding. Without this there is not an attractive enough force between protein and cell wall. In the case of truncated OmpA this is shown where the closed conformation remains bound to the membrane, whilst BLP pulls PGN away from the IM.

BLP continues to act as a structural staple that can facilitate OmpA and

TolR PGN binding. Without BLP present in the periplasm the cell wall is subject to random motion, where it has shown a propensity to bind OmpA, but without BLP there is no TolR-PGN association during simulation. This structural role indicates that the BLP forces the cell wall to exist at a certain height, where proteins can bind to it in both directions, so long as this local PGN position is maintained.

BLP tilt angle in the full periplasm is affected by TolR binding. When OmpA cannot interact, TolR open state will bind PGN. This TolR open state binding has a variety of strengths, dependent on specific residues. Stronger interactions result in a cell wall movement that causes the BLP tilt angle to trend towards 90 degrees, compared to other simulations, where a more acute BLP-OM angle is observed.

When bound, the periplasmic clamps of the proteins do not move much in the x/y plane, where TolR is shown to be more mobile. The cell wall provides a platform for various proteins to clamp onto, where when both are bound it can be seen that the cell wall is less mobile in the parallel perpendicular to the membrane. This effect is expected to amplified *in vivo*, to the extent that the cell wall is almost fixed in position relatively to the molecules interacting around it.

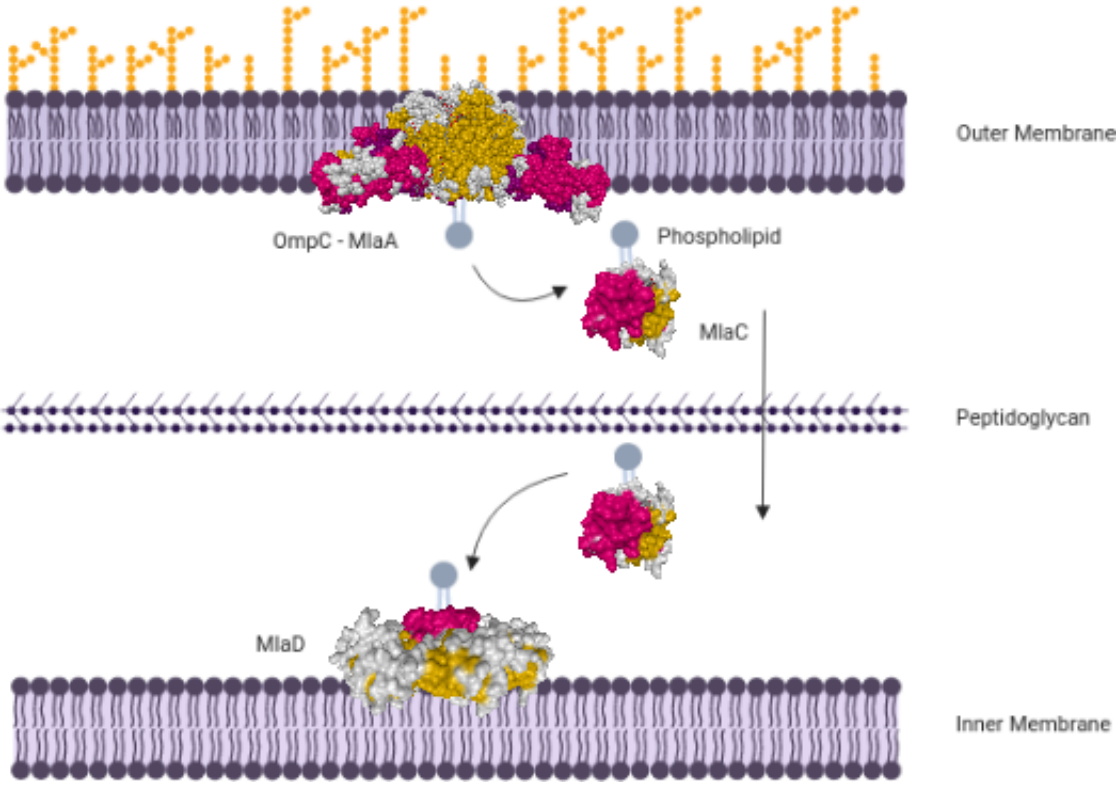
Linker length of the OmpA/TolR complex can suggest binding orientation and also controls the extent of the clamp-PGN interaction. Symmetric linker length profiles correlate to stronger binding, where if one linker contracts, the clamp will obtain a tilted geometry, which affects PGN association. Further to this, linkers in TolR can full contract to pull the periplasmic clamp to the IM, prevent PGN binding.

The full *E. coli* periplasm *in vivo* is a much more complex and crowded environment than this simple model. Based on this simple model a new model has been derived containing more proteins and structures. Much research has begun to focus on the periplasm and it is expected there will be many iterations of this system in years to come.

Other discussion and analysis surrounding this topic can be found under Appendix D, published as **Binding from Both Sides: TolR and Full-Length OmpA Bind and Maintain the Local Structure of the *E. coli* Cell Wall**. For a draft of the publication of a more complex periplasm see Appendix E, titled **Hitchhiker's guide to the periplasm**.

Chapter 5

The Mla Protein Family



5.1 Context in the Literature

Controlling the location of lipids in the cell is of critical importance.[184] All cells, including bacteria are known to be experts at this, preventing toxic ligand entry and even going so far as to export small molecules that are harmful[185].[186] This is not only true for antibiotics and other bactericidal but also for naturally occurring small molecules in the cell. Lipid transport is one of these pathways, where lipid transport between the IM and OM of gram-negative bacteria is of interest[187].[188]

Proteins that control these pathways are from the MCE (mammalian cell entry) superfamily.[189] These proteins can be identified as containing at least one conserved domain in sequence but without similarity to other proteins that are known. They are common in all bacteria containing a periplasm, whilst they do not occur in single membrane cells. This generally indicates that MCE proteins are exclusive to gram-negative bacteria, such as *E. coli*.[190] These proteins are key to virulence in bacteria, where mutations have been shown to reduce bacterial virulence.[191] Multiple studies have indicated that these proteins are also responsible for the movement of small hydrophobic molecules, such as lipids and steroids.[192] Initially the plant based proteins followed by bacterial proteins have been confirmed to bind PLs.[189]

The Mla (Maintenance of Lipid Asymmetry) proteins in *E. coli* are present throughout the periplasm. This family of proteins consist of MlaA, MlaC, MlaDEFB. Periplasmic proteins from this group are MlaA, MlaC and MlaD. MlaA is an outer membrane protein, MlaC is a periplasmic chaperone and MlaD is an inner membrane protein that is complexed to the rest of the Mla inner membrane proteins. The purpose of the MlaC, MlaA and MlaD proteins is to transport mislocalised phospholipid molecules from the outer membrane to the inner membrane. However it has been proposed that this pathway can also function in reverse, even to export molecules out of the cell.[54] Functionally this protects the organism as having a high LPS content in the OM increases the survivability of the bacterial cell.

The structure of MlaD has been resolved for *E. coli* for both X-ray and EM structures, where it has been shown that the MlaD forms a hexameric ring.[190] This hexameric ring is present at the IM outer leaflet and is complexed with other Mla proteins that occur in the transmembrane region.[193] This complex is thought to interact with MlaC, a chaperone, in order to transport hydrophobic molecules through the periplasm, (see Chapter 5 title figure).[194] MlaD is not the only MCE protein in *E. coli*, where other proteins (YebT and PqiB) form PL-binding hexameric rings that stack, creating tubes that span the periplasm[195].[192] MlaD is a 183 residue protein, which organises into a hexameric ring. The monomer consists of a seven stranded beta barrel, with a periplasm-facing helical structure and an N-terminal helix that rests in the IM.[190] The protein may be split into three domains; a short cytoplasmic region

(7 residues), a transmembrane region (residues 8 -28) and a periplasmic region (residues 29-183.[190] The X-ray structure consists of residues 38-152, with some missing residues in the loops at the edge of the hexamer. In the X-ray structure the C-terminal residues are missing. These are predicted to be unstructured and therefore are harder to characterize.

MlaC is a 221 residue protein, with a signal peptide at residues 1-21. It consists of 198 residues in the periplasm.[194] The secondary structure consists of nine helices and five beta strands, where the beta strands form a cavity that can potentially contain small ligands.[54] It is proposed to non covalently interact with the MlaD hexamer in order to transfer lipid between the two proteins.[193] The PL binding cavity is controlled via a β sheet lid.[196] MlaC has been shown to bind PL. Short MD simulations of potential sites for the hydrophobes to occupy inside the protein have already been reported.[194] A recent study has further shown that the MlaFEDB complex exports PL to MlaC, via an ATP independent pathway.[54] This study suggested that the direction of transport is from IM to OM. Additionally it has been shown that MlaC can exist in two states, closed and open[194193].[196] The open state occurs when a PL is bound to the hydrophobic site, whilst the flexible cavity will close when no PL is bound, in order to stabilise this hydrophobic region in the aqueous environment. This is controlled via a β sheet lid, that behaves almost like a hinge. The PL that is bound to the protein has previously been assessed to be either PE/PG without the presence of cardiolipin, but this newest study indicated that the protein had been purified with cardiolipin present[54].[197]

MlaA is lipoprotein that contains 234 residues, where the lipidation site occurs at residue 18, via an N-palmitoyl cysteine and an S-diacylglycerol cysteine.[198] The protein is membrane bound and associated with the trimeric OmpC unit.[198] The protein is largely helical with two β strands. This lipoprotein rests in the outer membrane and has been shown to exist in complex with OmpC/F[199].[32] The lipoprotein is found in close proximity to OMPs.[32] A joint computational and biochemical study showed the presence of a potential water pore and pathway that a lipid could be transferred through.[32] A complex with OmpC has been proposed to be 1:1 stoichiometry where a trimeric OmpC barrel complex interacts with an MlaA at each dimeric interface.[190] It has been proposed that this complex somehow facilitates the removal of PL from the OM; however the mechanistic details have yet to be revealed. The interaction with OmpC was confirmed however via UV irradiation and proteolytic degradation.[190]

Structural data exists for these Mla proteins, however the full structure for the MlaFEDB complex has yet to be resolved, at the time of writing several publications are in preprint with further structural data available[197200].[201] The available Cryo-EM data shows a strong correlation with the MlaD ring structure and suggests the rough dimensions of the total complex. This data does not have a high enough resolution to elucidate the fine details of the total complex

structure.

5.2 Premise

Lipid asymmetry is one of the hallmarks of the OM of *E. coli* and other Gram-negative species. The LPS molecules in the outer leaflet are presented as a vital barrier to entry to the cell. This layer must be maintained in order to protect the cell, where Mla proteins remove mislocalised PL to keep the concentration of LPS high. The proposed mechanism is that the MlaA molecule interacts with OmpC/OmpF proteins located in the OM and receives PL for transport to the periplasm. The PL is then picked up by MlaC, which contains a proven hydrophobic binding pocket, transported to the IM, where after docking with the MlaD the PL is transferred and eventually localised to the IM. This study explores the PL movement inside the MlaD and MlaC proteins, including how the two proteins can potentially dock, and seeks to model a terminal region for which there is no structural data, providing potential insight into the modes of interaction between the two periplasmic components. This attempts to glean more insight into the PL trafficking pathway.

Given this the following aims are proposed;

- To study the lipid interaction with the MlaD hexamer associated with a membrane.
- To study the same lipid interaction with the MlaC chaperone.
- Docking of MlaC to the MlaD hexamer
- Modelling of the MlaD terminal regions missing from available structures.
- To observe atomistic resolution lipid abstraction propagated by the MlaA/OmpC membrane complex.

5.3 Methods and Simulation content

All simulations were carried out for the shown lengths in tables 5.1 - 5.4 with the CHARMM36 forcefield, using the GROMACS2018 code[85].[202] Temperature was maintained at 310 K using the Nosé-Hoover thermostat (1 ps time constant). Pressure was maintained using the Parinello-Rahman barostat, at 1 atm (5 ps time constant), using semiisotropic pressure coupling for membrane systems and isotropic pressure coupling for aqueous systems. The TIP3P water model was used to solvate the systems, alongside a neutralising concentration of 0.2 M NaCl.[10] The short range electrostatic and Van der Waals cut-off distance was set to 1.2 nm. Simulations were run with a 2 fs timestep, using the LINCS constraint algorithm.[100] Three independent repeats for the MlaD, MlaC-lipid,OmpC-MlaA and MlaD-MlaC docking for unfolded termini were carried out. Multiple starting positions were tested in simulation (greater than 5 configurations) for the remainder of the docking simulations. Single shorter OmpC-MlaA runs for the mutagenesis were performed to investigate any change in lipid behaviour.

MlaD

Structure	Environment	MlaC docked?	Simulation length
X-ray repaired	Membrane	No	100 ns
Unfolded termini	Membrane	No	100 ns
Folded termini	Membrane	No	100 ns
Unfolded termini	Aqueous	Yes	100 ns
Folded termini	Aqueous	Yes	100 ns

Table 5.1: Varying MlaD termini model systems, including docking of the MlaC chaperone.

MlaC

Structure	Environment	Lipid bound	Simulation length
X-ray	Aqueous	POPE	100 ns
X-ray	Aqueous	POPG	100 ns
X-ray	Aqueous	PVCL2	100 ns

Table 5.2: MlaC systems, where different bound lipids are studied.

OmpC/MlaA

System	MlaA mutation	No. Lipid/MlaA complexes	Simulation length
Sys1	No	1	500 ns
Sys2	No	2	500 ns
Sys3	No	0	500 ns
Sys4	137ALA	1	100 ns
Sys5	137-138ALA	1	100 ns
Sys6	137-139ALA	1	100 ns
Sys7	137-140ALA	1	100 ns

Table 5.3: The wild-type and OmpC/MlaA systems that were set up to observe membrane protein interactions, using a symmetric bilayer of PL.

MlaA/MlaC

Structure	Environment	Cross-linking	Simulation length
X-ray	Aqueous	20%	50 ns
X-ray	Aqueous	40%	50 ns
X-ray	Aqueous	1 per group	50 ns

Table 5.4: Docking simulations of the MlaA to MlaC chaperone.

The same membrane composition was used for all membrane systems, 90:5:5 POPE:POPG:PVCL2, where POPE is short for 1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine, POPG is an abbreviation of 1-palmitoyl-2-oleoyl-sn-glycero-3-(phospho-rac-(1-glycerol)), PVCL2 is 1,1-palmitoyl-2,2-vacenoyl cardiolipin. The CHARMMGUI membrane builder was used to create a 10 x 10 nm membrane. Parameters for the lipids, were taken from the CHARMMGUI server for CHARMM36.[203]

MlaD was simulated using PDB: 5UW2. This crystal structure is lacking residues at the both the C and N terminus which were built using MODELLER along with missing loops. In membrane simulations, MlaD was mounted onto the surface of the membrane by positioning the N-terminus close (less than 0.5nm) to the surface of the membrane. An additional extended run of the X-ray repaired and unfolded termini systems of 500 ns was also done. Lipid pulling simulations were done using the GROMACS pull code. This was done by pulling the lipid centre of mass in a single direction (Z) away from the protein, using a rate of 0.001 nm/ps, with a force constant of 3000 kJ/mol/nm^2.

MlaC was simulated using PDB: 5UWA.[190] This structure is missing the first 23 residues, which were built using the MODELLER software.[204] Lipid binding systems were set up via close positioning (less than 0.5 nm) of the lipid to the cavity of MlaC, allowing the lipid to bind the cavity stochastically during

simulation. Enhanced sampling for MlaC-lipid systems was done using umbrella sampling, where the GROMACS pull code was used to generate the configurations for each window and the analysis was done with WHAM to generate the PMF data. Windows every 0.1 nm were used and the simulations were all run for 100 ns.

The OmpC-MlaA trimeric complex was simulated using PDB: 5NUP.[198] Mutations of this complex were done using the PYMOL mutagenesis tool.[133] Protein-protein docking was done via the ClusPro server.[205] Restraints were applied during the docking via a minimum distance of 0.6 nm between proposed docking residues on either protein. Cross-linking data and other helpful insights were provided by the Shu Sin lab, NUS. This data is as follows: The following residues on MlaD are proposed to interact with MlaC: F119, M141, Q149, Y152 and V179. The following residues on MlaC are proposed interact with MlaD: P124, E169, V171 and S172. Cryo-EM density was taken from the work of Ekiert et al.[190]

5.4 Results

5.4.1 Modelling MlaD membrane association

The starting model for MlaD was the loop repaired crystal structure (see Section 5.3). In the MlaFEDB complex, MlaD is proposed to rest on the surface of the membrane. Thus, initial simulations of each of the MlaD models were done via placement of MlaD onto a membrane surface. In this section the aim is to predict stable membrane bound states of MlaD and to attempt to assess binding modes to the termini regions and loops, as they could be important regarding membrane association. Modelling of the termini is essential due to the presence of residues that are proposed to interact during MlaD-MlaC association that are not present in the crystal structure.

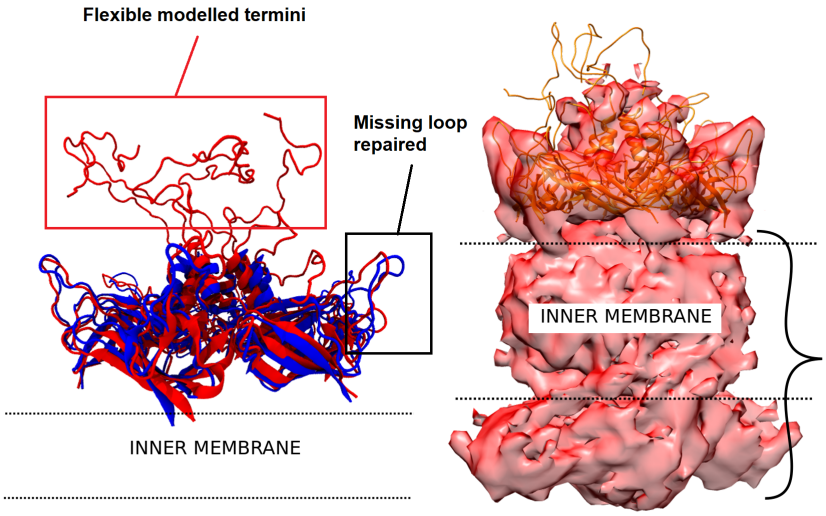


Figure 5.1: Aligned structures of the MlaD loop-repaired protein (blue) and the flexible-termini modelled MlaD hexamer (red) (left). The Cryo-EM density of the MlaFEDB protein complex (pink), with a post 100 ns simulation flexible termini MlaD hexamer fitted to the density (orange), where the bracketed region indicates the MlaFEB proteins,(right). In both, the position of the membrane is shown with dotted lines, where the flexible termini point into the periplasm.

Shown in Figure 5.1 are the separate structures of the MlaD repaired crystal structure and the modelled termini region. It can be seen that after 100 ns of simulation and alignment, the core crystal structure regions of the proteins had

a similar structure, where some variation was seen at the repaired loop. The modelled termini region was free to move and it can be seen that the termini extended into the periplasm, where each flexible region can adopt a variety of positions and self associate.

Alignment of this data with Cryo-EM density showed that the core regions agree with the density, but that the termini fitted less well due to the flexibility of these regions and/or the resolution of the Cryo-EM data.

The MlaD loop repaired model can be seen associated with the membrane, (Figure 5.2) where it remained associated for the full length of the simulation, as is expected. The RMSD of this protein indicates that the protein is stable over 100 ns, fluctuating at a value of $\sim$ 0.3 nm over 100 ns (Figure 5.3). The repeating pattern of the RMSF indicates similar behaviour by each monomer, where the flexible regions are at the repaired loops and termini (Figure 5.3).

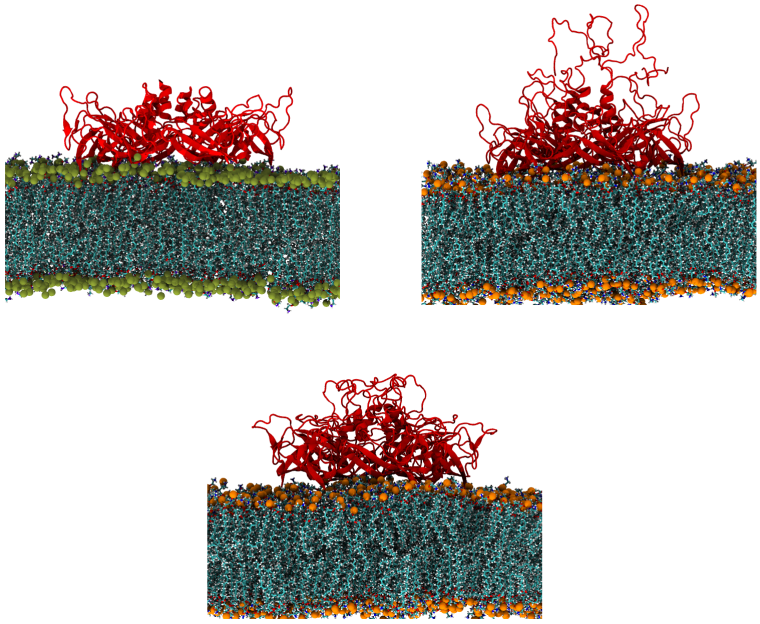


Figure 5.2: The loop repaired x-ray MlaD structure (red) associated to the membrane (grey licorice) with phospholipid headgroups shown as spheres (brown) (top left). The unstructured termini modelled MlaD structure (red) associated to the membrane (grey licorice) with phospholipid headgroups shown as spheres (orange) (top right). The folded termini modelled MlaD structure (red) associated to the membrane (grey licorice) with phospholipid headgroups shown as spheres (orange) (bottom).

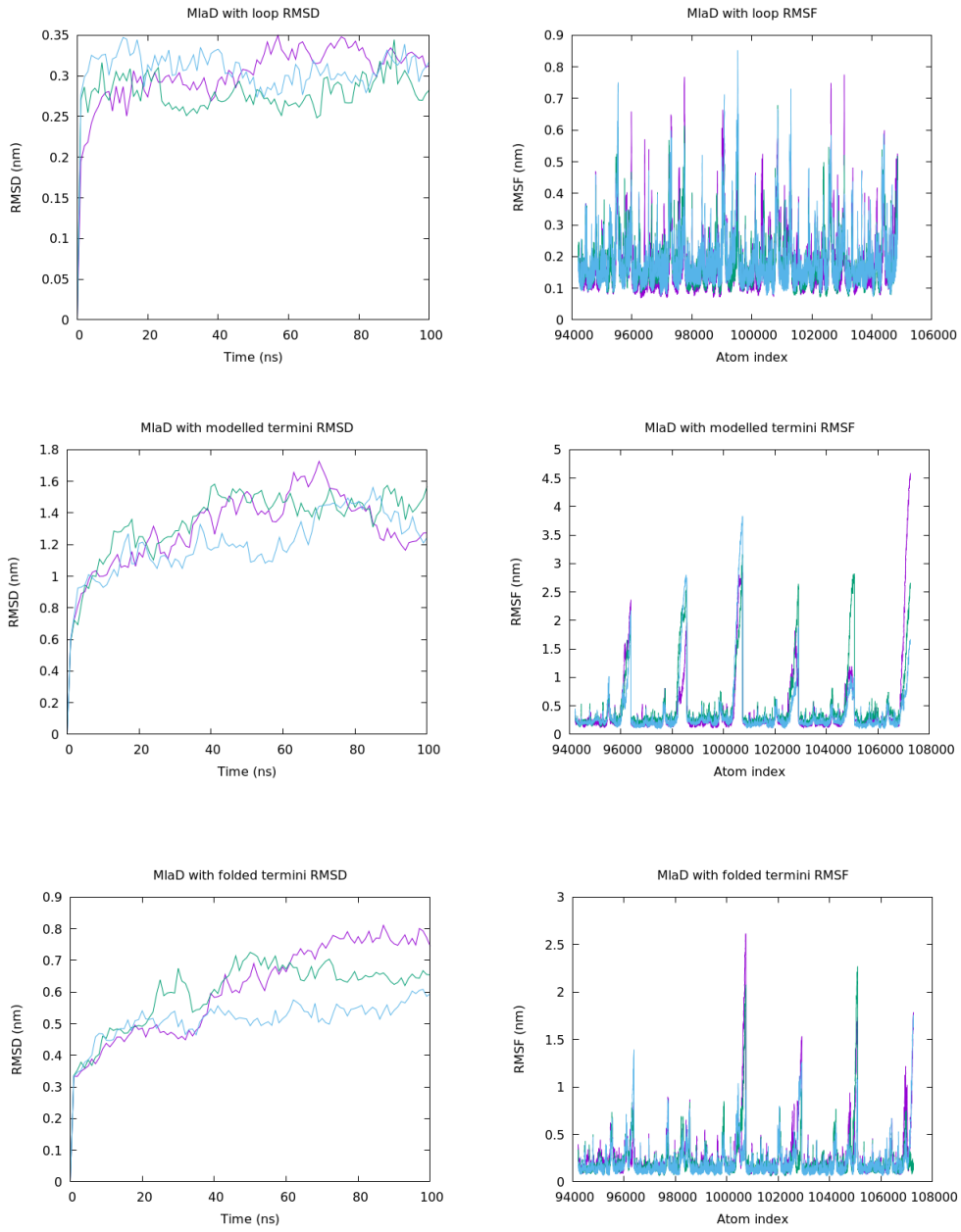


Figure 5.3: The RMSD and RMSF the entire MlaD loop repaired hexameric structure across independent repeats (top). The RMSD and RMSF of the entire MlaD modelled termini hexameric structure across independent repeats (middle). The RMSD and RMSF of the MlaD folded termini structure, across independent repeats (bottom).

Adding the modelled termini to MlaD introduced a significant amount of conformational space that can be explored. Figure 5.2 shows that this structure associated with the membrane but the termini were flexible and could associate with lipid or extend into solvent. RMSD was significantly increased for the protein as a whole upon addition of the termini, plateauing around 1.4 nm for the protein. Comparatively the core barrel regions behaved similarly between the modelled termini and loop repaired structures, where similar values of less than 0.5 nm RMSF were seen. These unstructured terminal regions could possibly act like a "fishhook" that could potentially attract the MlaC protein chaperone for docking. This poses the question, if MlaD has a specific interaction to MlaC for PL transport then how does this interact occur considering the highly flexible termini?

During simulation it was observed an unstructured terminus is interacting with an adjacent monomer in each monomer in the ring. This structure bears a high degree of similarity to the Cryo-EM density that is available. Therefore a homohexameric model of this state was next built. Simulation of this showed that the RMSD of the protein was largely reduced when the flexible termini started in this folded conformation. Comparison between the "free" and "folded" termini showed RMSD values in the range of 0.6 - 0.8 nm for the folded termini, approximately half of the value for the free structure (Figure 5.3). RMSF once again indicated that the termini were the most flexible regions present in the protein, however six distinct peaks were no longer observed, where fluctuation in one terminus was relatively similar to the rest of the protein and in a similar range to the RMSF of the original repaired x-ray structure. This shows that self association and interaction with other monomers in MlaD has a stabilising effect on these unstructured modelled termini, indicating a theoretical membrane associated structure that is stable and could also potentially interact with MlaC.

5.4.2 Lipid binding to membrane bound MlaD

The loop repaired structure of MlaD, without any modelled termini, was next simulated on a membrane surface of the same inner membrane model as previously described (90:5:5 PE:PG:Cardiolipin).

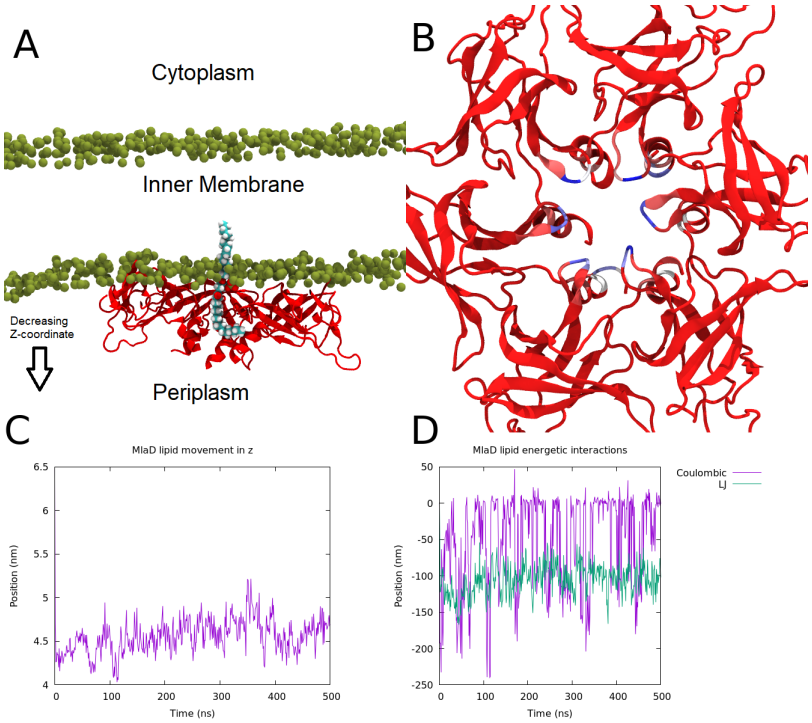


Figure 5.4: A visualisation of the MlaD hexamer (red) mounted on the membrane, only headgroups are shown for clarity (brown). The POPE lipid (cyan) can be seen to unbiasedly insert one tail into the MlaD hexameric pore (A). A top down view of the MlaD hexamer, where contact probability is displayed using the Blue-White-Red scheme, where blue is highest contact probability and red is no contact (B). The POPE lipid movement in the z-axis, perpendicular to the membrane shows the movement of the lipid into the MlaD hexamer (C). The short range electrostatic and Lennard-Jones interactions between the inserted lipid and the MlaD (D).

The MlaD hexameric pore is comrpised of six helices at the N terminus, pointing into the periplasm. During simulation of the membrane bound X-ray repaired structure, a POPE lipid bound the hexameric pore at the center of the MlaD was observed in multiple simulations at the 100-500 ns timescale. The lipid entered into the MlaD hexameric pore in a “one up one down” configuration (Figure 5.4). Thus, one lipid tail remained associated to the membrane whilst another extended down into the periplasmic helical region formed by the hexamer. Over the simulation the lipid remained bound to this region, where the highest contact was observed at the loop regions between the monomeric interfaces and on the helices in the pore (Figure 5.4).

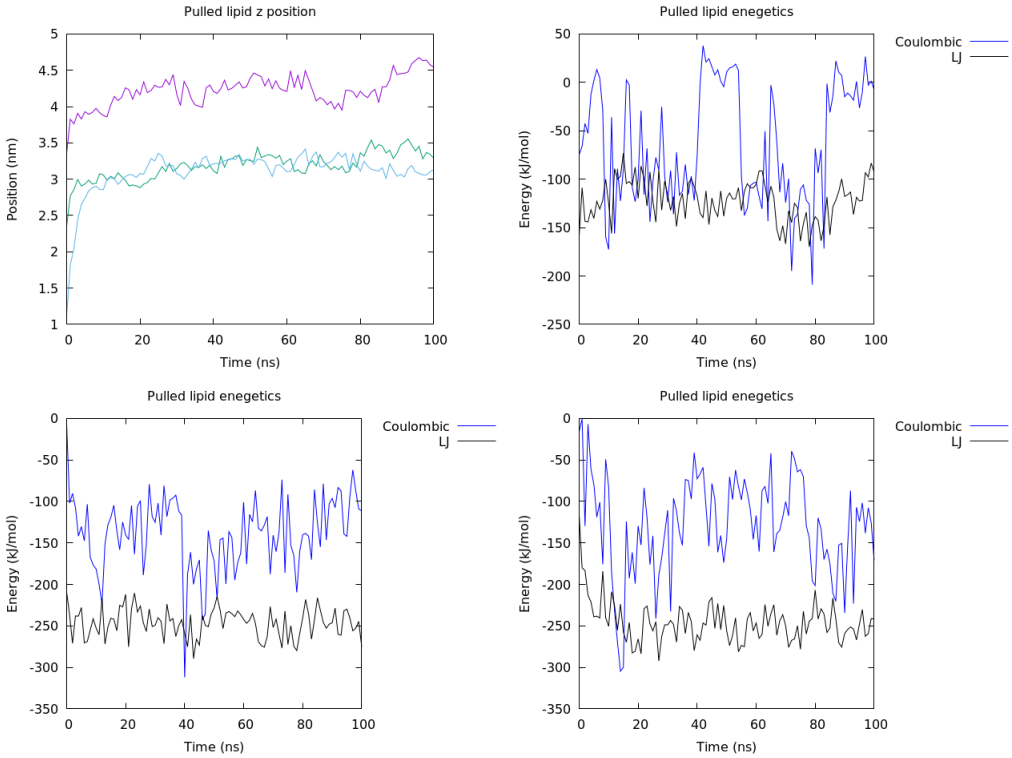


Figure 5.5: The movement of the lipid centre of mass at three different starting heights, where increases in Z show movement towards the membrane, whilst decreases show movement towards in the periplasm, in the MlaD pore over 100 ns, in the z-axis, perpendicular to the membrane (top left). The short range electrostatic and Lennard-Jones interactions between each lipid and the MlaD over 100 ns (top right, bottom left and bottom right).

Lipid configurations were generated by pulling the lipid through a short region of the pore. Simulation of the lipid in three starting poses after pulling in this pore showed an upwards movement in the lipid, indicating that the lipid relaxed back into the pore, towards the membrane, indicated by an increase the Z-coordinate of the lipid. This was done to access more of the potential lipid pathway, where various binding sites in the MlaD pore could be explored. Analysis of energetics showed that the first relaxation was less favourable, where both charged and non charged interactions were less stabilising, however in two cases both lipids relaxed to the same position, which was maintained by a larger coulombic and Van der Waals (LJ) energy (Figure 5.5). This serves as further confirmation that lipids could stably associate with the pore.

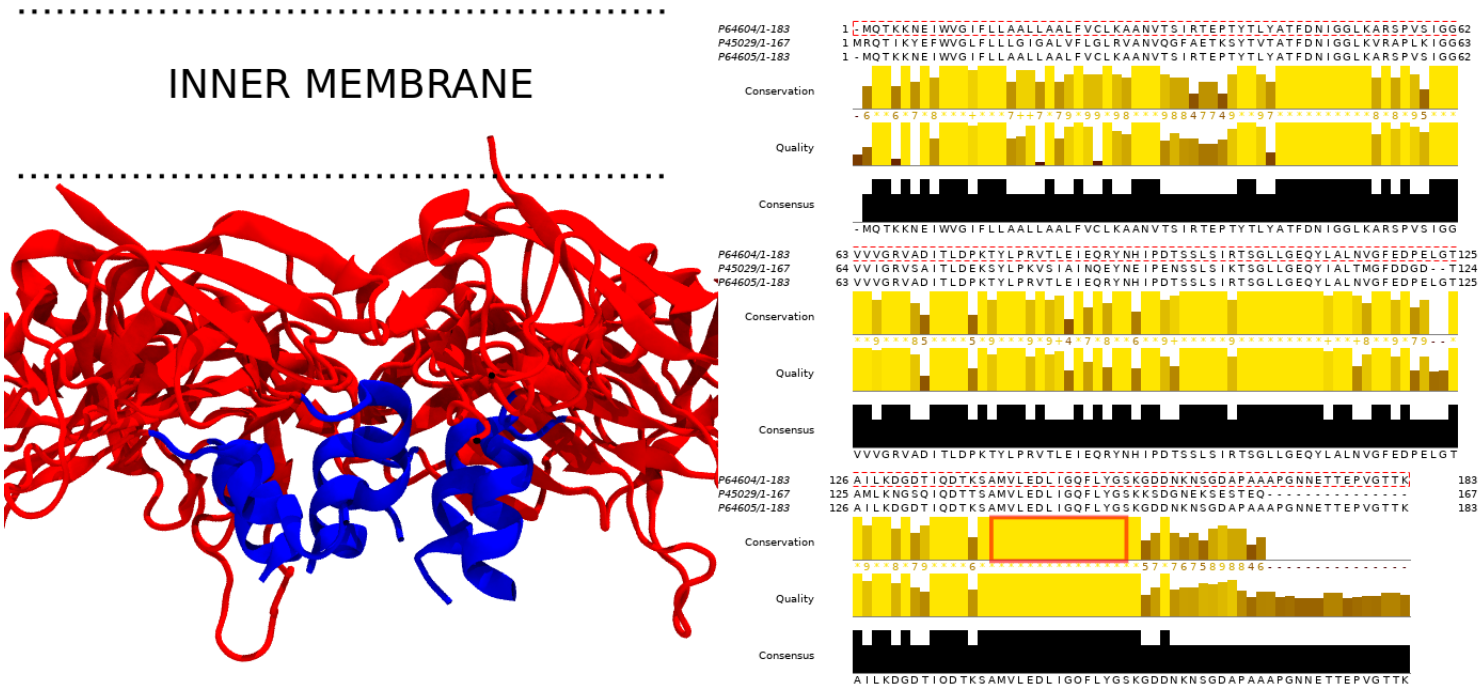


Figure 5.6: A close up visualisation of the MlaD hexamer (red), where the highlighted region (blue) forms the periplasmic helical bundle, creating a pore (left). The sequence alignment of three separate MlaD structures, where the highlighted region (red rectangle) represents the helical region in MlaD, where it is 100% conserved (right).

Sequence alignment of the MlaD protein across three species of gram negative bacteria; *E. coli (strain K12)*, *E. coli (O157:H7)*, *H. influenzae (strain ATCC 51907 / DSM 11121 / KW20/ Rd)* indicated that certain regions of the MlaD hexamer are well conserved. One key region, close to the N-terminus, that indicated 100% conservation was the highlighted $\sim$ 17 helical pore region. It is possible the tyrosine and phenylalanine residues present in this region create a Pi stacking interaction, to maintain the pore stability and/or interact favourably with the lipids that are transported between the MlaFEDB complex and MlaC. Regardless the data from various species shows that the helical pore is a conserved region.

5.4.3 MlaC ligand binding

As previously stated, the first 23 residues of MlaC were missing.[190] The terminal region in MlaC exists at the edge of the protein (Figure 5.7). None of the cross linking residues measured experimentally by the Shu Sin lab for the MlaD-MlaC interaction exist in this region.

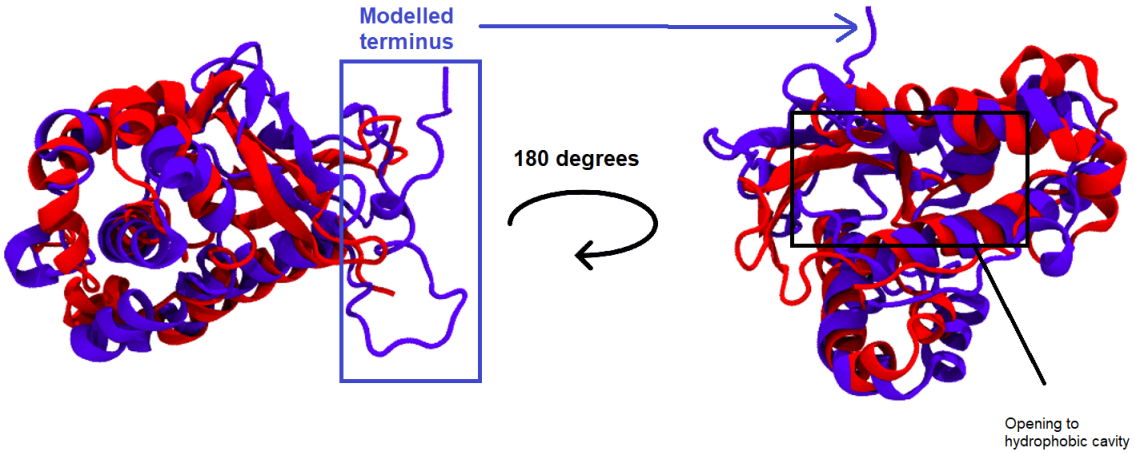


Figure 5.7: The MlaC chaperone, where the crystal structure (red) and the structure containing the modelled termini region (blue) are aligned.

The terminal modelled region is on the wrong side of the protein to interact with the MlaC cavity. It is proposed that it may not significantly contribute to MlaC-lipid binding or MlaC-MlaD binding. With this information available after modelling both structures in aqueous simulation, the original x-ray structure was selected for simulation i.e. without modelled termini, reducing the level of uncertainty in simulations involving the MlaC protein.

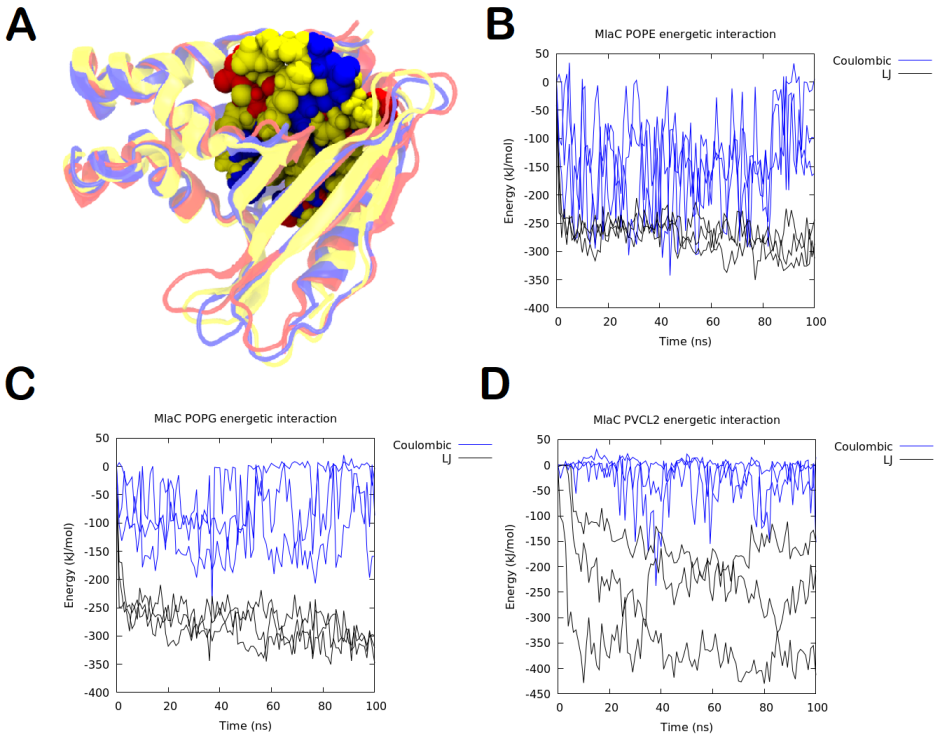


Figure 5.8: A snapshot of aligned MlaC chaperone bound to different lipids MlaC-POPE(blue), MlaC-POPG (red) and MlaC-PVCL2 (yellow) with protein shown as a carton and lipid as spheres (A). The short range electrostatic and Lennard-Jones interactions between POPE and MlaC, across independent repeats (B). The short range electrostatic and Lennard-Jones interactions between POPG and MlaC, across independent repeats (C). The short range electrostatic and Lennard-Jones interactions between PVCL2 and MlaC, across independent repeats (D).

The start condition for this was to place the lipid tails within 1.0 nm of the cavity. In Figure 5.8 it can be seen that MlaC will non-specifically bind differently lipid types. In these simulations (see Table 5.2) we saw spontaneous association between protein and POPG, POPE and PVCL2. Within 10 ns, in each system the lipid inserted into the MlaC cavity, where the tail groups were bound to the beta barrel cavity core region, whilst the head groups were oriented to the surface of the protein. The start condition for this was to place the lipid tails within 1.0 nm of the cavity.

Considering that the MlaC crystal structure was bound to POPE, these simulations once again confirm stable binding (Figure 5.8). Interestingly, cardiolipin bound to the pocket, where all four tails fit inside the cavity, indicating that the cavity could flex and change size in order to admit different lipid types, occupying approximately twice the volume of PE or PG.

As would be expected the Van der Waals interaction (LJ) was the dominant force involved in the lipid binding. Driven by the hydrophobic effect, POPE and POPG displayed similar profiles, where after 100 ns the LJ interaction was $\sim$ -300 kJ/mol for both lipids. PE displayed larger but fluctuating charged interactions with MlaC. PG showed a charged interaction that generally varied between 0 - -150 kJ/mol. Cardiolipin showed different behaviour, where the LJ forces were the most stabilising by far and generally the charged interaction was less than -100 kJ/mol during simulation, which makes sense in light of its greater number of acyl tails.

5.4.4 MlaC-lipid umbrella sampling

Umbrella sampling of the lipid binding to MlaC was done to assess specificity. Is there a specific lipid that MlaC has a preference for, or is it non-selective and promiscuous? To assess the binding, three separate protein conditions were used in order to study the removal of the POPE, POPG and PVCL2 lipids. The "free state" of the protein simply involved pulling the lipid out of the protein, with no restrictions on the movement of the protein. The "elastic state" of the protein differed via the introduction of an elastic network between the carbon backbone atoms of the protein, essentially preventing secondary structure change. The "restrained state" of the protein took this one step further, where the position of the backbone of the protein is restrained in x, y and z. This was done to assess if structural changes of MlaC influence lipid binding.

There was an energy penalty associated with removing any of the lipid types from the MlaC cavity. Using umbrella sampling, three different states of the protein were examined (Figure 5.9). In the "free" state, the order of binding affinity was as follows: PVCL2>POPE>POPG. Considering the charged attraction shown by the POPE shown in the previous section, it is expected to be harder to remove, however the additional tails in cardiolipin caused a hydrophobic interaction that is stronger than for POPE. The "restrained" state showed poorer sampling, as can be seen in the profiles. Testing the effect of position restraints leads to poor convergence and unreliable data, but was done to explore if malleability of the MlaC affected the binding affinity. The "elastic" state of MlaC generally showed a similar value to the "free" state, indicating that malleability did not significantly influence the binding energetics, seen in Figure 5.9.

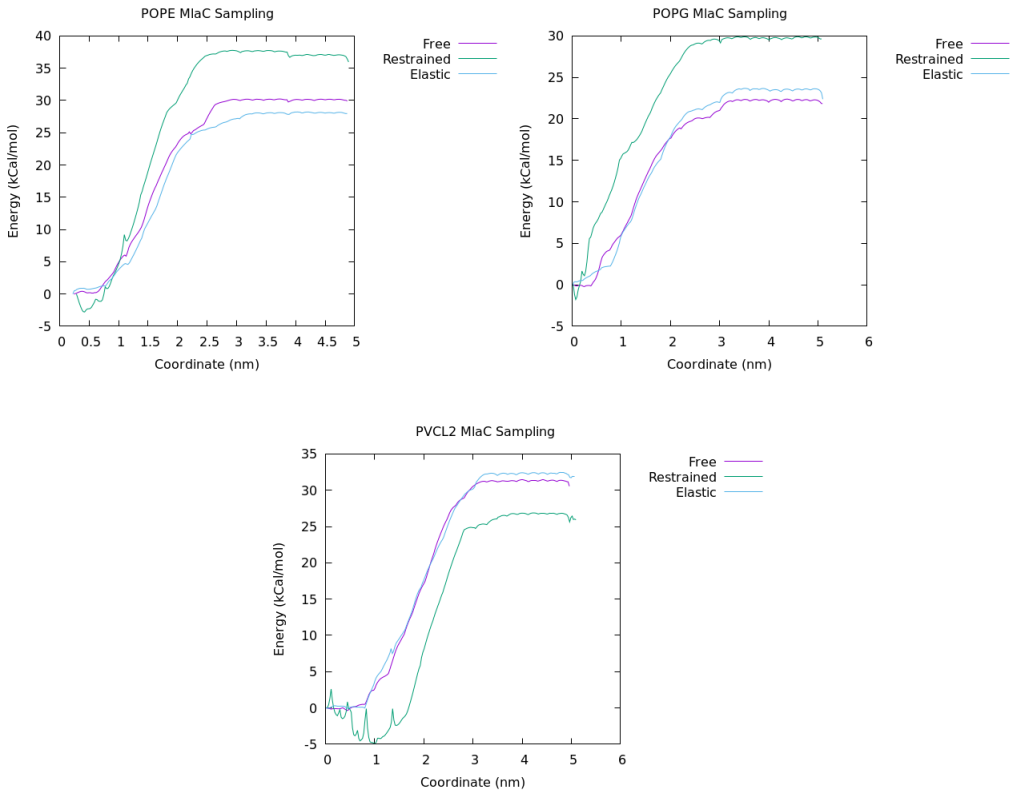


Figure 5.9: The enhanced sampling energy profile for the removal of POPE from MlaC, where the free protein, elastic network and positions restraints are compared (top left). The enhanced sampling energy profile for the removal of POPG from MlaC, where the free protein, elastic network and positions restraints are compared (top right).The enhanced sampling energy profile for the removal of PVCL2 from MlaC, where the free protein, elastic network and positions restraints are compared (bottom).

Histograms for this umbrella sampling can be found in Figure 7.9 in the Appendix.

5.4.5 MlaD to MlaC docking

Considering that MlaC bound lipid favourably and the hexameric MlaD pore also did, protein-protein docking may allow investigation of possible lipid transfer pathways between the two. Using the provided crosslinking data from the Shu Sin lab, this was integrated with docking to model potential binding poses

between MlaC and MlaD with C-termini present.

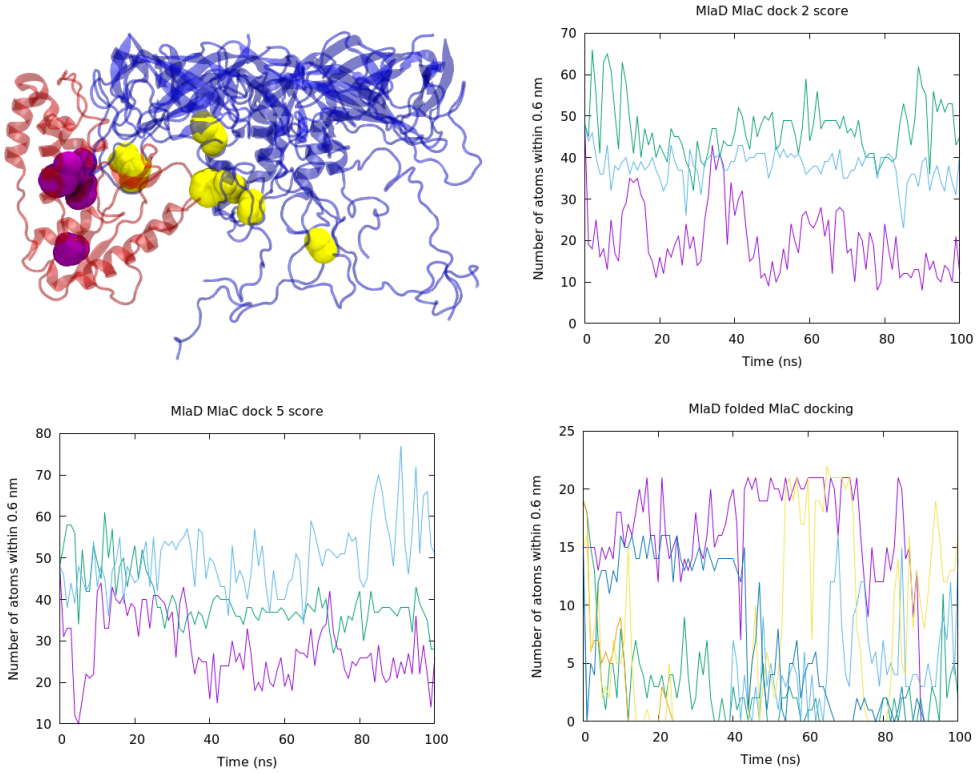


Figure 5.10: A visualisation of the highest contact structure can be seen (top left) where MlaC (red) is docked to MlaD (blue), with the hydrophobic pocket pointing towards the helical pore, MlaC crosslinking sites are shown as purple spheres, MlaD cross linking sites are shown for the closest monomer as yellow spheres. The docking poses of the MlaC with highest contact between MlaC-MlaD (flexible termini) crosslinking residues are shown (top right and bottom left), across independent repeats. The docking poses of the MlaC with highest contact between MlaC-MlaD (folded termini) crosslinking residues are shown (bottom right), across independent repeats.

An example of one such docking structure, where high contact was established can be seen (Figure 5.10). This indicates that the docking was successful when the open cavity of MlaC was directed towards the helical pore of MlaD. Interestingly in this scenario it can also be seen that the unstructured termini were pushed to one side, indicating that despite the need for termini-MlaC interaction, high binding contact occurred when the other termini were not blocking

the binding site.

It can be seen that this docking was subject to a large deviations, where separate repeats showed very different data in many cases. This was attributed to the variability that was introduced in the flexible termini and the need for one residue on this region to contact the MlaC in precisely the correct position. This data shows that there are multiple potential binding poses that produced significant contact between cross linking residues between the modelled MlaD and MlaC proteins, shown in Figure 5.10.

A similar docking protocol was performed for the "folded termini" MlaD structure with MlaC. In comparison to the "free termini" docking data, it can be seen in Figure 5.10 the docking was arguably worse in the folded system, compared to the docking in Figure 5.10 of the free termini structure. Across 6 simulation systems in many cases contact fell off entirely, whilst the highest degree of contact that occurred was ~ 20 atoms, suggesting that in the "folded" conformation the binding site was not as accessible as the "unfolded" one. The highest contact seen in the unfolded docking was ~ 75 atoms, more than three times that amount seen in the highest folded MlaD hexamer.

The folded termini did not move as much as the unfolded, shown previously. This indicates that there must be some form of termini movement in order to facilitate the MlaC binding. From this docking it is proposed that *in vivo* there is an equilibrium that is established between the unfolded and folded states. In this equilibrium the folded termini can move to a more flexible position to admit MlaC, and return to the folded state once lipid transfer is complete.

5.4.6 The OmpC/MlaA trimer in a symmetric PL bilayer

The OM component of the Mla family, the OmpC/MlaA trimeric complex was next simulated. This is the starting point for the maintenance of lipid asymmetry. The role of MlaA in the OmpC/MlaA complex is to abstract PL from the outer leaflet. In a more realistic model containing LPS and patches of PE it would require vast amounts of sampling in order to possibly observe PL movement into MlaA. Thus, a simplified symmetric PL bilayer was used.

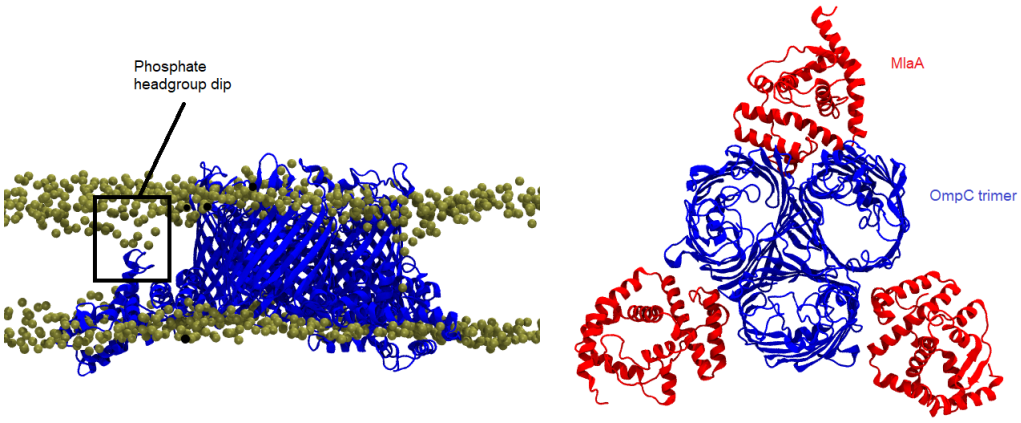


Figure 5.11: The MlaA-OmpC trimeric complex (blue) embedded in a symmetric bilayer (90:5:5 PE:PG:CL), only headgroups shown for clarity (brown), where the head groups are seen to dip toward the MlaA (left). The MlaA-OmpC trimer shown, MlaA (red) at the edges of the OmpC dimeric interfaces, the OmpC trimer (blue) is associated with three MlaA proteins (right).

After 500 ns it can be clearly observed that the PL headgroups moved from the outer leaflet and into the MlaA protein (Figure 5.11), where a disruption in the normally flat bilayer was seen. Across three independent repeats it was observed that the MlaA components in the first repeat did not abstract PL, in the second that one MlaA protein abstracted PL and in the third that two MlaA proteins instigated lipid movement. In the case where one lipid was seen to enter the MlaA cavity, it can be seen that within 200 ns the lipid centre of mass decreased by 2 nm and reached a steady state (Figure 5.12). In the second case, the movement of both lipids into separate MlaA proteins was tracked. The profiles for both lipids are similar and mimic that of the first case of lipid transfer into MlaA (Figure 5.12). A reduction from $\sim$ 6.5 nm to 3.5 - 4 nm was once again seen in 200 ns. Notably this was an irreversible movement; there was no sign of the lipid returning to the outer leaflet once this process occurs and one lipid was bound to the water pore cavity of MlaA at each time. The PL headgroups surrounding the abstracted PL were also lower in the Z dimension than the bulk bilayer, indicating that this effect was not local to single lipids.

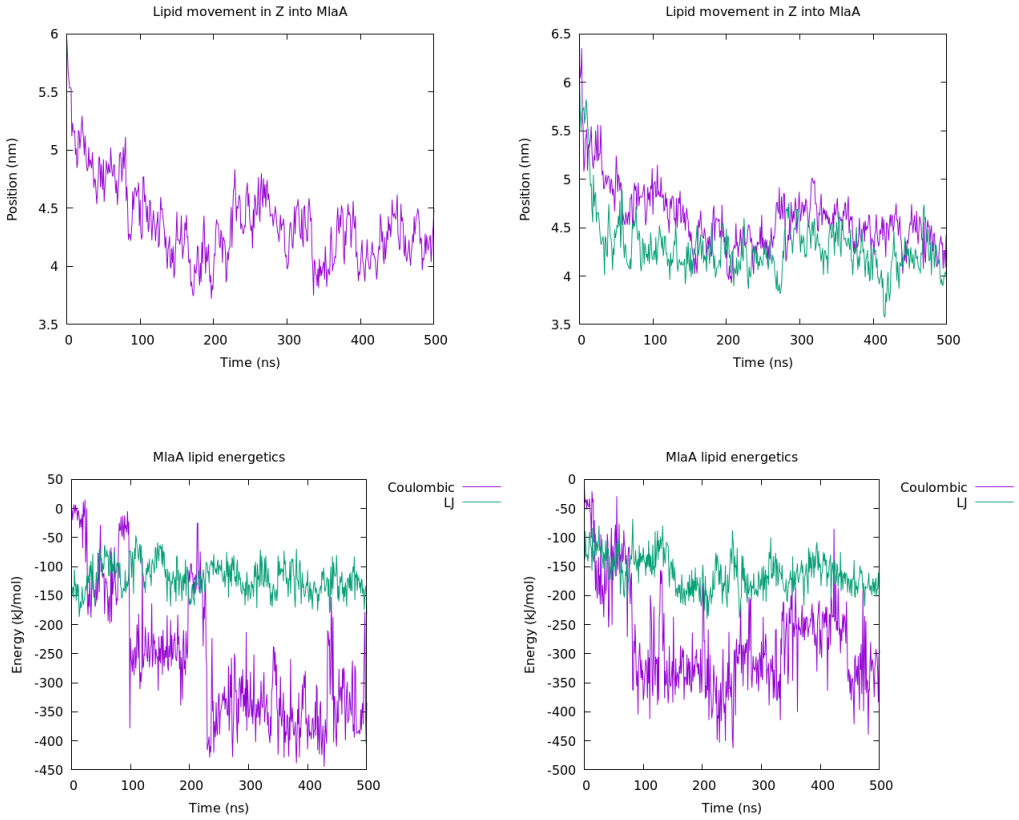


Figure 5.12: The movement of one lipid into the MlaA protein, over 500 ns in one replica, measured via position in the z-axis perpendicular to the membrane (top left). The movement of two lipids into separate MlaA proteins, over 500 ns in another replica, measured via position in the z-axis perpendicular to the membrane (top right). The short range electrostatic and Lennard-Jones interactions between the lipids and MlaA during movement into the MlaA cavity, shown over 500 ns in independent replicas (bottom left and bottom right).

The lipid interaction with the MlaA cavity was favourable. Highlighting two instances of this, it was seen that there was a consistent LJ interaction between the lipid and MlaA, from the outset of the simulation. In the case of the charged energy there appeared to be several stages involved with the MlaA lipid abstraction. Both lipids appeared to be associated with an initial stabilising charged interaction that initiated PL movement ~ 50 ns; during this window the lipid was abstracted 1 nm from the outer leaflet. Following this, after a brief instance where the charged energy destabilised, the coulombic energy sharply

stabilised to roughly -300 kJ/mol, followed by a further stabilisation to roughly -400 kJ/mol that generally occurred close to 200 ns, shown in Figure 5.12.

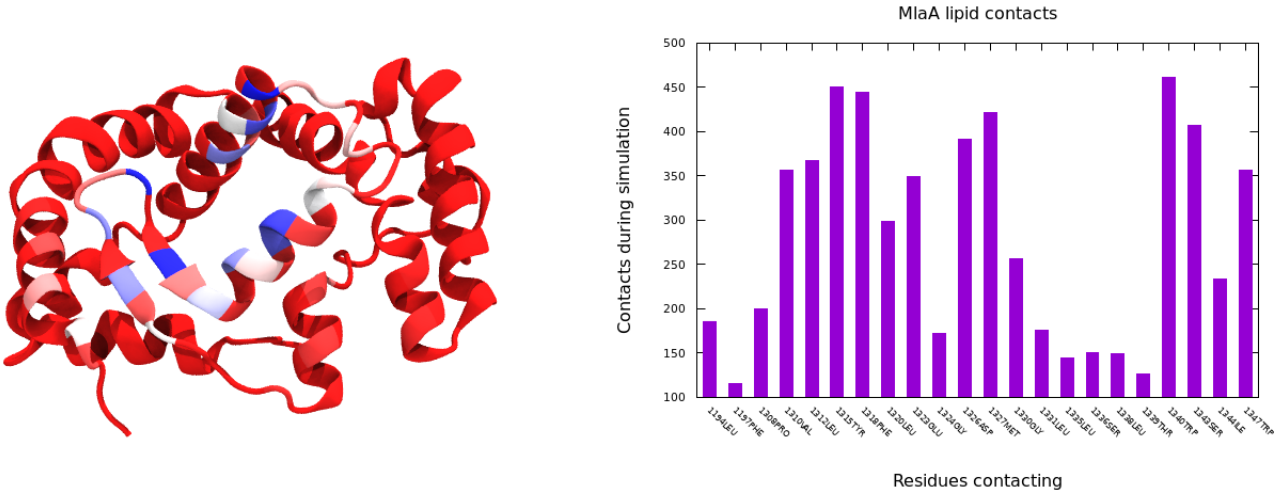


Figure 5.13: The contact probability between lipid and MlaA during the abstraction process, using the BWR colour scheme, where blue is highest contact and red is lowest (left). The specificity of the lipid-MlaA contact over 500 ns (right).

Contact analysis shows that the lipid "slides down" the helical pore region where water flows, and comes to rest where the lipid tails were contacting the beta sheet of MlaA (Figure 5.13). In this orientation, the head group was pointing into the water pore and the tails were perpendicular to the bilayer. Based on the high level of contact with the beta sheet it is proposed that this may act as a "gating region" for abstraction of PL, see Appendix Figure 7.7. The beta sheet gating loop presented a barrier through the MlaA cavity for PL. Mutations of this loop showed that successive mutations of the group generally led to decreased stability of the lipid when interacting with the loop see Appendix Figure 7.8 and the associated text.

The highest contact occurred between residues 1315TYR, 1318PHE and 1340TRP, over 90% of the simulation. The lipid also had relatively high contact with the residues surrounding these. 1315TYR and 1318PHE are located on the proposed beta sheet gate, whilst the 1340TRP is located on the helical region upstream from the beta sheet. These three residues played a key role in the stabilisation, due to their aromatic nature, aiding the movement of the

lipid through MlaA. Polar and charged residues such as 1343SER and 1326ASP interacted significantly with the lipid. These, in conjunction with the solvent present are proposed to stabilise the charged headgroup of the lipid as it is transferred.

5.5 Discussion

Separate components of the Mla protein family have been studied in order to elucidate phospholipid interactions with each protein, including MlaD, MlaC and OmpC/MlaA. At the outset of the project accurate data for the full MlaFEDB complex was not reported, but new structures are now being released[19754].[190]

Modelling of the MlaD hexamer has allowed for the atomistic characterisation of potential conformations of the unstructured termini. In the unfolded conformation these termini fluctuate more and diffuse freely in the periplasm; it is proposed that this behaviour could help to capture MlaC for docking, but based on to experimental data it seems unlikely that this random motion would successfully drive lipid transfer. The folded state of the termini are more stable, fluctuate less and add a degree of symmetry to MlaD, where the full complex is observed to be highly symmetric via Cryo-EM. It is possible the termini exist between these states, in a folded, but more dynamic conformation, or that they can alternate between the two.

Docking of MlaC to MlaD indicated that the unfolded termini allow the proposed interaction sites, based on experimental crosslinking data, the highest amount of contact. Docking of the folded state conversely was less successful: where unfolded termini can move to admit MlaC, the folded version cannot, indicating that the specific folded conformation used inhibits binding by blocking the active site. Docking of the MlaA to MlaC was not descriptive enough without the addition of the full membrane bound complex. Short simulations of this complex indicated that contacts between experimental crosslinking residues were short lived and the further addition of data from the Shu Sin lab could be used in future to further explore this docking in a membrane environment.

The x-ray structure of MlaD contains a hydrophobic helical pore, and this pore is well conserved over multiple bacteria. This short region extends into the periplasm , where experimentally confirmed interaction sites are present, indicating that lipid transfer can occur at this region. In fact, the MlaD pore can abstract lipid tails from the IM model bilayer, where hydrophobic interactions dominate. It was seen that lipids will freely associate with the MlaD pore. Without the interactions between MlaD and MlaE the specifics of how the lipid enters this pore in the full complex remain unresolved, but initial models of this were proposed.

Simulations of the lipid in the helical pore indicate that lipid relaxes toward the bilayer, with the lipid binding in the interface between the MlaD monomeric units. This indicates once lipid enters MlaD, there are key binding regions for the headgroup, whilst the tails enter the inside of the pore.

MlaC has been shown to favourably bind three different lipids. With variations in head group and the number of tails present. The chaperone can thus be

seen to contain a promiscuous cavity, capable of transporting a variety of lipid across the periplasm.

Enhanced sampling has shown that the flexible movement of the MlaC chaperone did not significantly change the energy required to remove a lipid from the binding site. Without any restraints, cardiolipin was shown to be most difficult to remove from the MlaC cavity. POPE and POPG displayed a similar affinity for the cavity of the protein. When an elastic network was applied it can be seen that these energies differ slightly, but the values were similar to the "free" state of MlaC. Application of position restraints caused sampling issues and varied energy values that can't be relied upon.

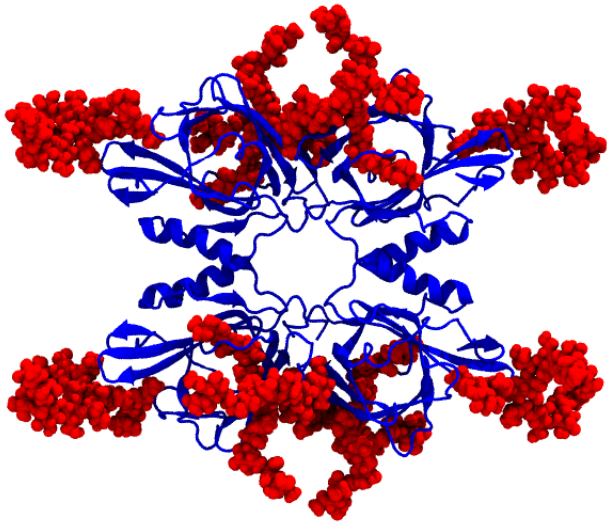
Atomistic detail of lipid abstraction from a symmetric bilayer into the MlaA protein of the MlaA/OmpC complex has been observed. MlaA lipid abstraction is controlled by charged interactions, where the movement of the lipid is eventually blocked by a gating loop. This gating loop is a short region that can serve as a landing pad for the abstracted lipid, however once this loop binds lipid it is unclear how the remainder of the lipid abstraction to allow MlaA-MlaC lipid transfer occurs.

Mutation of the MlaA gating loop weakened the interaction between abstracted lipid and this region of the protein. Removing the specificity of this loop, both by reducing bulk and removing charges indicated that the loop is specifically tailored to bind the lipid. With this loop in mind, it is hypothesised that an unknown mechanism such as MlaC binding triggers the "opening" of the gating region necessary for the lipid to completely transfer through MlaA.

It has been observed that all three periplasmic components of the Mla protein family will readily bind lipid. The molecular level movement of these lipids provides key insights to the components of the Mla system that are necessary for lipid transport. As more experimental data becomes available, as is the case at the time of writing where several new structures are on the verge of release, knowledge of this system will become more detailed. *In silico* a full lipid transport pathway is not apparent, but this data will provide insight to our experimental collaborators at the Shu Sin lab.

Chapter 6

Lol and ApoD Chaperones



6.1 Context in the Literature

The Lol (localisation of lipoproteins) system was discovered to transport lipoprotein by the Tokuda research group in the 1990s.[206] It was first proposed that a novel protein that allowed the release of BLP from the IM existed. The results indicated that this protein, termed p20, formed a soluble complex with OM lipoproteins. Further analysis showed that BLP released from the IM in the presence of p20 was localised into the OM. This was the first study of the protein known as LolA. A following study showed that the LolCDE complex exists in the IM and that the release of lipoprotein from the IM is an ATP dependent process.[207] It was stated that lipoproteins destined for the OM are selectively sorted by this complex, where the amino acid at position 2 on the lipoprotein defines this specificity. It is generally found that ASP at this site causes the lipoprotein to be ignored by the complex, suggesting that these lipoproteins are retained at the IM. The crystal structure for the LolCDE complex has not been resolved, however the same research group characterised LolA and LolB in 2003.[206]

LolA has the structure of an 11 stranded unclosed antiparralel beta-barrel where one side of the barrel is covered by three alpha helices, Figure 6.1. The inner component of the barrel and helices are hydrophobic, allowing for lipid transport. LolA is proposed, like most chaperones, to exist in an open and closed state. This has been shown to be facilitated by these helices, where essentially they can "pop" up to allow for hydrophobes to enter[208].[209] This mechanism of motion is referred to as a helical lid, which can open and close to change the protein state. This has been quantified where the closed state is stabilised via hydrogen bonding between ARG43 and alpha helix one and two. As this hydrogen bonding must be broken to open the lid it is suspected that the LolCDE complex catalyzes the lid movement. Research into the lipoprotein transfer shows that LolA favourably transfers these ligands to LolB in the OM.[210] Analysis of this transfer in real time indicates that the transfer is irreversible.[211]

LolB has a similar structure to LolA, despite a surprising level of sequence similarity of 8% between the two proteins[212].[206] Once again the structure is composed of an 11 stranded antiparallel beta sheet that is capped by 3 alpha helices. Similarly a hydrophobic cavity is present in LolB.[213] A key difference is that LolB is a lipoprotein, where the N-terminus is lipidated with N-palmitoyl cysteine and S-diacylglycerol cysteine, common lipoprotein moieties.[214] The similarity in structure between the two proteins leads to a theory that they can dock and somehow transfer the lipid from one hydrophobic cavity to another, yet precise details of this mechanism are still not known[211].[210]

A known inhibitor of LolA is MAC13243. When LolA is bound to this molecule, the *E. coli* OM becomes more permeable. This is attributed to the lack of essential molecules such as BLP and Pal to provide OM structural integrity. Permeation has been proven via the application of 1-N-phenylnapthylamine

(NPN), which can only enter OM compromised cells and is also a fluorophore. Cells treated with MAC13243 present fluorescence 15 times higher than DMSO controls, and the fluorescence changes based on concentration used.[215] This is further confirmed when MAC13243 allows the entry of antibiotic compounds such as vancomycin that normally cannot pass the OM. MAC13243 is not unique in this example, as it is known to degrade easily in solution, where these degradation products (S-(4-chlorobenzyl)isothiourea and S-(4-dichlorobenzyl)isothiourea) also can bind to the LolA to cause inhibition.[216]

The complete structure of the LolCDE complex that releases lipoproteins has yet to be solved. Given the data available the following questions can be asked: How does LolA bind to LolB? How does BLP bind to LolA for transport? What are the molecular details of this MAC13243 inhibition that reduces LolA function? Using MD and enhanced sampling, these interactions can be analysed.

ApoD is another example of a small protein (169 residues) that is a key transporter for cell processes, but is eukaryotic in nature instead of prokaryotic. It is a human glycoprotein that is present in the plasma.[217] Structurally the protein contains an eight stranded beta barrel, a mainly hydrophobic region that is close to an alpha helix. It is involved in lipid metabolism, where its binding activity for progesterone and arachidonic acid affects cancerous and neurological disease development[218].[219] Not only has the protein been shown to transport these molecules but also can directly affect their oxidation state.[220]

This protein has been shown to exist as a monomer or a dimer, but recently a tetrameric structure has been proposed[221].[222] The structure of the protein is highly conserved between various eukaryotes. The close association of ApoD to HDL particles has been confirmed, in particular via a disulfide bond with the apolipoprotein A-II.[223] This protein has been expressed in *E. coli* organisms, therefore it can exist in a similar environment to other periplasmic chaperones, such as LolA or MlaC.[224] One study shows that ApoD can be selective in its binding to similar compounds, where progesterone and arachidonic acid will bind but pregnenolone and bilirubin will not[224].[225] Previous ApoD MD studies give evidence to show lipid binding to the hydrophobic pocket, where lipids closely associate with MET93, proposed to allow for the conversion of the MET93 to MET93 sulfoxide, where glycosylation of the protein shields a second site at MET49.[220]

6.2 Premise

Lol (localisation of lipoproteins) proteins are present in the *E. coli* periplasm. The Lol system comprises of an IM (LolCDE), OM (LolB) and periplasmic component (LolA). LolA is proposed to accept lipoprotein ligands from the IM complex and transport them to LolB in the OM for membrane localisation. BLP could be transported in this system, and given the relationship between BLP and OM integrity, some studies have been done on the Lol system in order to shed light on lipoprotein trafficking. The mechanism of this binding and transport has yet to be explored, however there are known compounds that inhibit this transport behaviour.

ApoD is a well known human glycosylated protein that has a known interaction with high-density lipoprotein (HDL). It is generally found in the brain and testes and is found in elevated levels for patients suffering neurological disorders such as Alzheimers. ApoD can also carry small molecules in a hydrophobic pocket.

The following aims for simulations of LolA have been devised:

- To assess BLP binding to the LolA chaperone.
- Similarly to assess the MAC13243 inhibitor binding to the LolA chaperone.
- Assess if the MAC inhibitor prevents BLP binding to LolA.
- Implementing umbrella sampling to sample the free energy of the lipid removal from the LolA pocket.

Similarly the aims for parallel ApoD simulation were to study the association of cholesterol, arachidonic acid, pregnenolone and progesterone with the ApoD cavity.

Therefore this simulation work explores chaperone binding for LolA with BLP and various inhibitor compounds, to answer the question of how this inhibition occurs at a molecular level. Alongside this is an exploration of how ApoD can stably bind various small molecules for transport in aqueous systems. This is a continuation of a similar type of work involved with MlaC and highlights how atomic level understanding of ligand binding could reveal new details for future use.

6.3 Methods and Simulation content

6.3.1 LolA

All simulations of the LolA protein were done using the GROMOS54A7 forcefield, using the GROMACS 2018 code. At least three independent repeats of each simulation were performed. Temperature was maintained at 310 K, using the velocity rescale thermostat, 1 ps time constant. Pressure was maintained using the Parrinello-Rahman barostat, at 1 atm, time constant 1 ps. The systems were solvated using the SPC water model, alongside a 0.2 M neutralising concentration of NaCl. LINCS constraints allowed for a 2 fs integration timestep.[100] The charged and Van der Waals cut-offs were set to 1.4 nm.

BLP bound	Number of inhibitors	Simulation length
Yes	0	100 ns
Yes	1	100 ns
Yes	2	100 ns
Yes	3	100 ns

Table 6.1: Simulation systems set up for the inhibition of BLP-LolA binding, using the MAC inhibitor.

The LolA model used was obtained from the PDB (PDB: 1IWL), and the parameters used were the standard GROMOS54a7 potentials.[206] The inhibitor used was MAC13243 (MAC), the parameters for this small molecule were obtained from the Automated Topology Builder (ATB), using the GROMOS54a7 forcefield[226125].[215] The MAC is in an uncharged state, where the behaviour of the drug protein binding pockets could potentially carry a different charge compared to the solvated drug, considering the imine group on the ring, which would generally be charged at neutral pH. Further to this the parameters provided by the ATB can be questioned, where the accuracy using automated services is always in doubt, leading to the need for further paramterisation. The BLP model was the structure deposited by Shu *et al.* (PDB: 1EQ7), where the N-terminus is palmitoylated as previously described.[165]

PMF calculations were performed as previously, where using a pulling simulations the windows for simulation were generated, followed by 100 ns simulation of each window and analysis using WHAM in order to compute the free energy of lipid removal from LolA, with respect to increasing amount of MAC inhibitor present in the protein cavity. The pulling simulations were done via pulling the lipid centre of mass from the protein centre of mass. A force constant of 100 kJ/mol/nm was used and a rate of 0.0001 nm/ps was used. To assist in the convergence of this sampling, the BLP protein segment was removed, leaving only the lipid to be pulled from the cavity.

Starting coordinates for these simulations were generated via manual placement of MAC in the LolA cavity, followed by stochastic association of the BLP lipid.

6.3.2 ApoD

All simulations were run for the lengths shown in Table 6.2. The CHARMM36 forcefield was used to simulate the systems.[85] Temperature was maintained at 310 K using the Nose-Hoover thermostat, 1 ps time constant. Pressure was maintained at 1 atm, using the Parinello-Rahman barostat, 5 ps time constant. The charged and Van der Waals interaction cut-offs were set to 1.4 nm. The system is solvated using the TIP3P water model, in conjunction with a neutralising concentration of 0.2 M NaCl ions.[10] LINCS constraints were applied to covalent hydrogen bonds to allow for a 2 fs integration time step during simulation.[100] Three independent repeats of each of these simulations were performed.

Glycosylated	Number of Units	Substrate	Simulation length
No	monomer	-	100 ns
Yes	monomer	-	100 ns
Yes	monomer	Cholesterol	100 ns
Yes	monomer	Arachidonic acid	100 ns
Yes	monomer	Progesterone	100 ns
Yes	monomer	Pregnenolone	100 ns

Table 6.2: Simulation systems set up for the glycosylated vs non-glycosylated ApoD structures and systems for the ligand binding to the ApoD monomer.

The ApoD models were derived from the tetrameric unit present on the SASBDB database, (SASDD83)[227].[228] Glycosylation was added to two sites: N45 and N78, according to the literature. N45 is attached to a trisialo trianternnary glycan, whilst N78 carries a fucose disialo biantennary glycan.[229] This glycosylation was done using the CHARMMGUI glycosylation feature, in order to modify the protein.[203] Ligands bound to ApoD were generated using CGENFF server, and the standard CHARMM36 parameters, where no large penalties involving the parameters were observed.[230] Ligands were placed close to the ApoD cavity and allowed to bind freely during simulation.

6.4 Results

6.4.1 Apo LolA binds the BLP lipid component

Initial LolA simulations showed that the lipid moiety of BLP can bind to the hydrophobic cavity inside of the chaperone, Figure 6.1. When this occurred the protein component was bound to the outside of the LolA, where the majority rested in the solvent bulk.

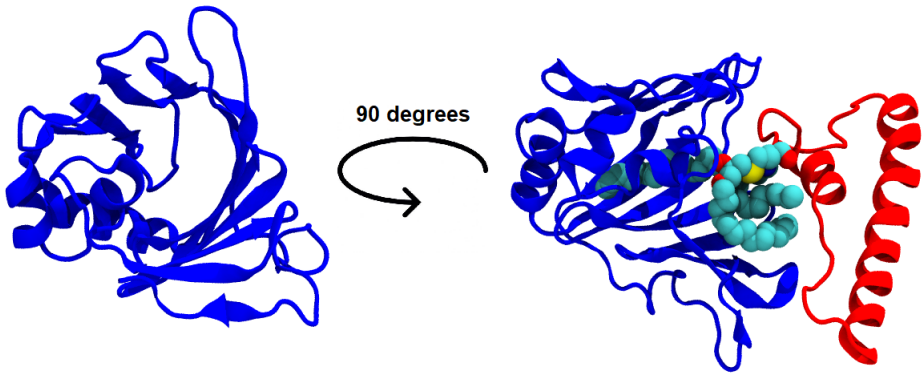


Figure 6.1: A visualisation of the LolA chaperone (blue) where the protein is oriented showing the binding cavity (left). A visualisation of the LolA apochaperone (blue) bound to the monomer of BLP (red), where the lipid moiety (cyan) associates with the hydrophobic cavity (right).

It was observed that the LolA cavity was open in this configuration, where there was significant volume available for the the lipid tails to bind. With respect to the stoichiometry of the BLP binding, it was assumed that a single monomer of the BLP trimer can bind the LolA cavity. This was due to the folding and change in structure of the protein, in contrast to straight helical structures seen previously, in Chapters 3 and 4, when bound to LolA. It was noted that this helical structure of the BLP was not eliminated entirely but was compressed in order to facilitate LolA binding, Figure 6.1.

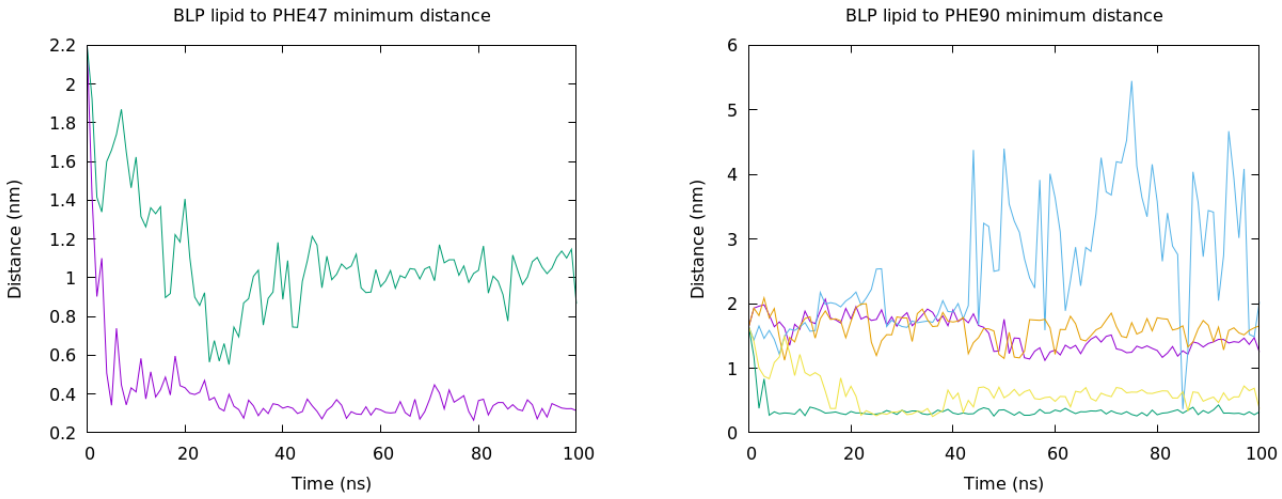


Figure 6.2: The lipid to LolA minimum distance to the PHE47 surface residue (left) and the PHE90 buried cavity residue (right).

This binding was numerically confirmed via the minimum distance of two residues, PHE47 and PHE90, where both both residues are located at the base of the LolA cavity, where full lipid binding in the pocket would allow for association to the lipid. In two instances close association to these buried phenylalanine residues was seen, where in one case a minimum distance to PHE47 reached a value of 0.2 - 0.4 nm by 40 ns and was maintained. In another, a reduction to 0.6 nm was seen after 20 ns, however there was an increase increase to a 1 nm distance. Comparison of this contact showed a similar distance to PHE90, where all independent runs are shown. In this instance it was seen there there are BLP lipids closely associated to LolA, however the tails of the lipids were not fully extended into the hydrophobic pocket. With respect to PHE90 in LolA, two lipids only contacted the mouth of the cavity, whilst one lipid was not bound at all, while the two remaining lipids were fully bound.

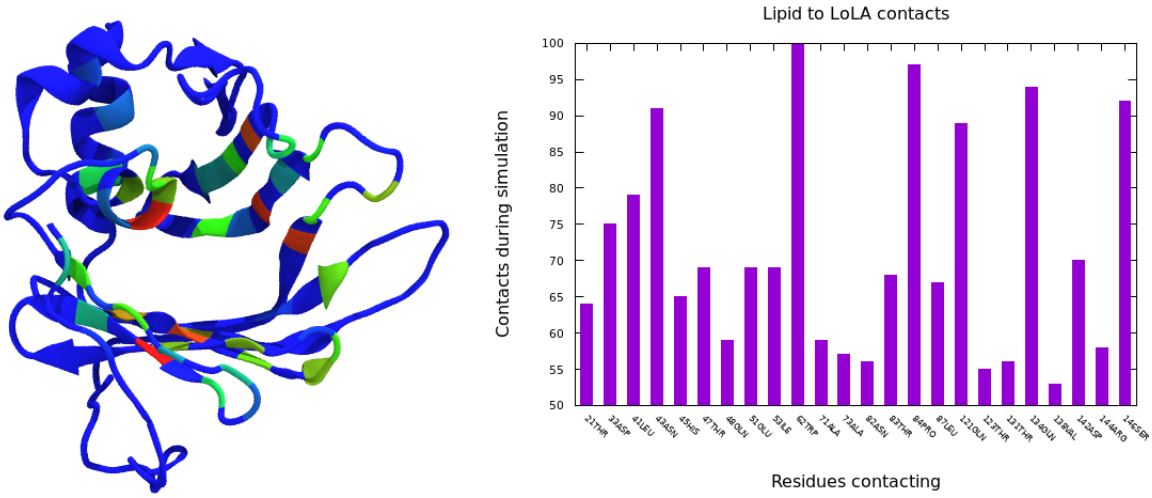


Figure 6.3: The binding contact probability of the lipid to the LolA cavity in the first repeat, shown in the BGR (Blue-Green-Red) colour scheme, where blue is lowest contact (none) and red is highest (left). Lipid contact to specific residues on the protein in the first repeat (right).

In the first repeat where the most contact between the lipid and hydrophobic core was seen there was a high amount of contact between LolA and the residues buried deep in the pocket. It was seen that the lipid contacts a large number of residues in this region, where significant contact was seen around the mouth of the cavity. Residues that were in contact with the lipid for over 90% of simulation time were: ASN43, TRP62, PRO84, GLN134 and SER146 where the protein has been renumbered to start at residue 1, Figure 6.3. These residues are highlighted where it can be seen that there is a pattern of high contact along adjacent beta sheets in LolA, indicating that the lipid adhered to the hydrophobic wall of the half barrel structure and the tail extended deep into the pocket.

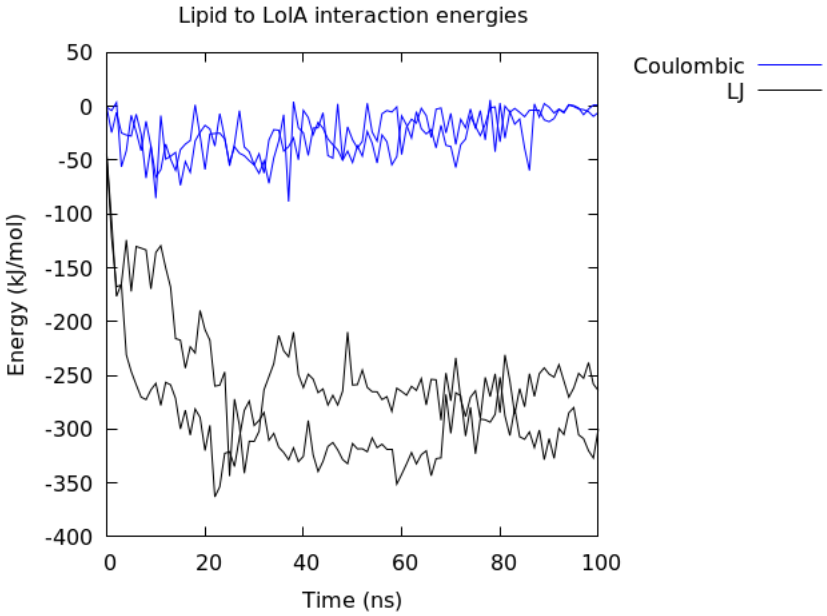


Figure 6.4: The short range electrostatic and Lennard-Jones interactions between the lipid and LolA over 100 ns, for two independent repeats.

Energetically comparing the two best binding events shows that in the case of both the charged energy did not play a significant role, a small contribution of only $\sim$ -50 kJ/mol was attributed to coulombic effects, in Figure 6.4. Non-charged interactions formed the bulk of the attractive energy between the lipid moiety and the LolA cavity. Within 20 ns there was an increase in attractive energy to $\sim$ -350 kJ/mol, indicating that the lipid was stably bound to the pocket. It was seen that a similar degree of stabilisation was reached and that fluctuation occurred in the same range, however in the case of deeper lipid binding an increased stabilisation of $\sim$ -50 kJ/mol was seen. This is attributed to the more frequent contact with hydrophobic residues in the pocket of LolA, confirming that this was the most stable region for the lipid to bind.

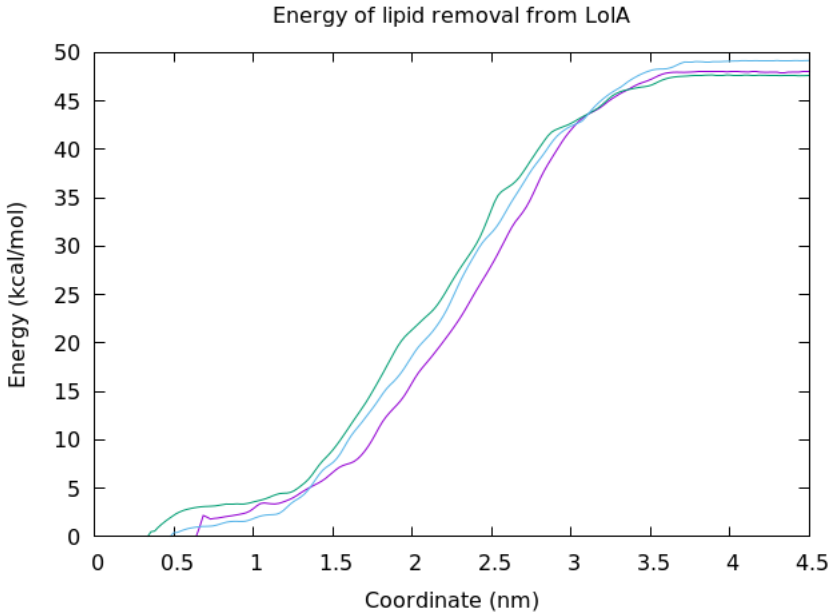


Figure 6.5: The PMF profile for the removal of BLP lipid alone from the apo LolA cavity, where each line represents an independent repeat.

Based on the best bound state from the simulations above, calculations for the lipid binding affinity estimation were performed. Given the favourable environment for lipid binding, there was an energy penalty involved in the removal of the lipid group in the cavity. Umbrella sampling calculations showed an energy of 47 - 49 kcal/mol in order to fully remove the lipid from the protein across three independent repeats (Figure 6.5). As the lipid was extracted there was an increase to $\sim$ 5 kcal/mol over the first 1.5 nm of the sampling region. From 1.5 - 2.5 nm there was a large increase in energy to $\sim$ 30 kcal/mol. In this region the lipid was removed entirely from the LolA cavity, into bulk solvent.

6.4.2 LolA binds MAC inhibitor and BLP lipid simultaneously

1 MAC Molecule

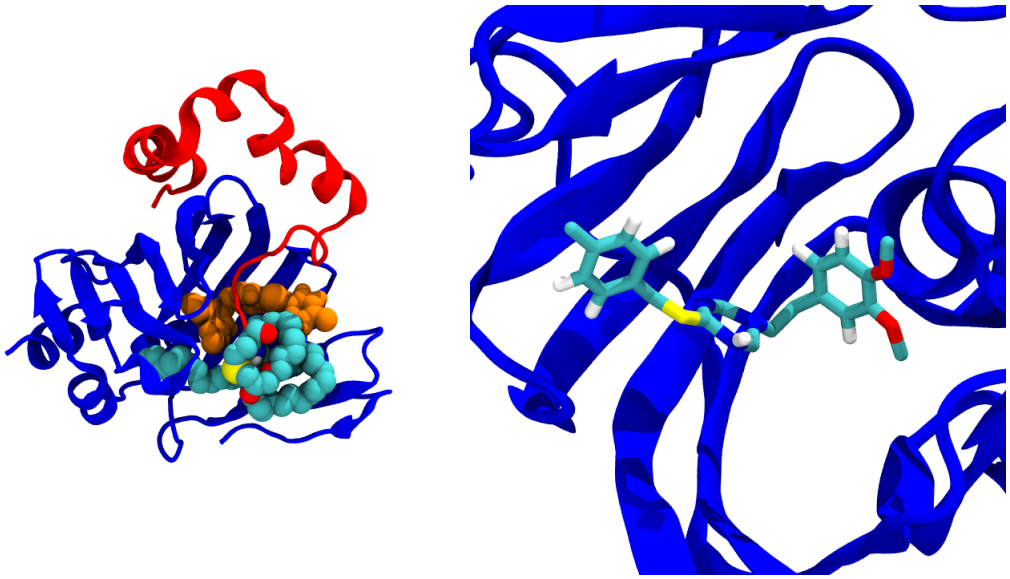


Figure 6.6: An illustration of LolA (blue) to BLP (red) lipid (cyan) binding in the presence of the MAC inhibitor (orange), (left). A visualisation of the MAC inhibitor bound (coloured according to atom type) to the LolA protein cavity (blue), (right).

The MAC inhibitor has been shown to effectively inhibit lipoprotein binding to LolA. Manual placement of this inhibitor in the hydrophobic pocket, followed by simulation, showed that this inhibitor can be stably associated with the chaperone (Figure 6.6). When an inhibitor molecule was placed into the cavity of LolA, it can be seen that BLP lipid will additionally bind to the cavity, spontaneously during simulation.

The inhibitor can span the cavity, where the aromatic rings of this ligand were in contact with the lipid binding regions shown previously (Figure 6.6). Whilst it may be expected that BLP binding be inhibited via introduction of MAC, the opposite was seen, where the lipid tails were interacting with the MAC, and lipid actually intertwined with the ligand structure. This behaviour contradicts the competitive binding model for inhibition.

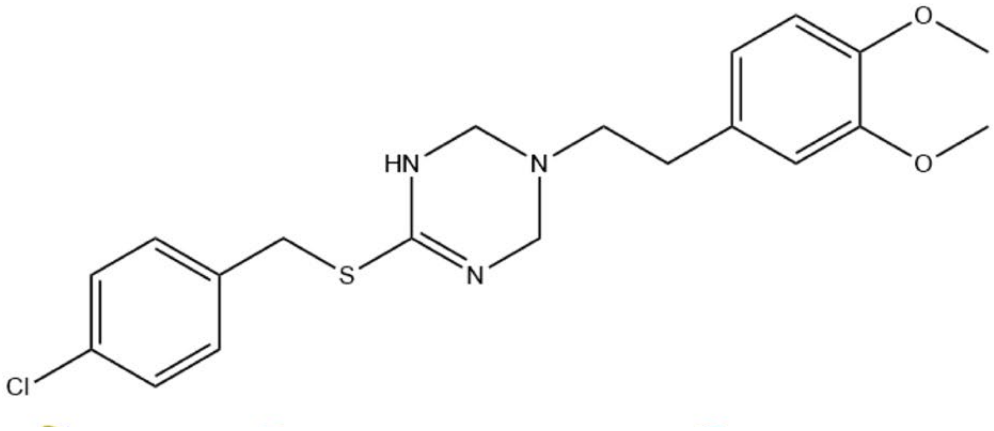


Figure 6.7: The skeletal chemical structure of the MAC inhibitor.

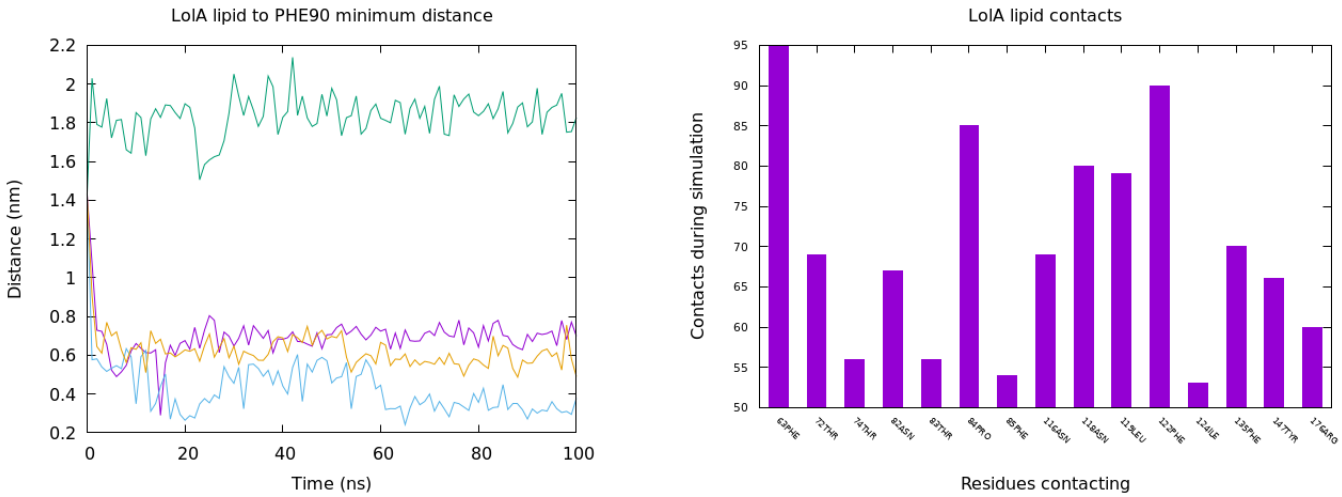


Figure 6.8: The lipid minimum distance to the buried PHE90 residue in LolA when one MAC inhibitor is bound (left). The contact between the LolA protein and lipid in a case where lipid binds when MAC is present (right).

The structure of MAC shows both hydrophobic and hydrophilic groups (Figure 6.7). This is similar to the lipid ligand that bound LolA, where hydrophobic tails and polar groups are present. Considering this, it is reasonable that both lipid and inhibitor can interact with the pocket simultaneously as was seen in Figure 6.6.

From four independent repeats it was seen that three simulations led to

lipid binding deep into the LolA cavity, where these lipids occupy approximate distances of 0.7, 0.6 and 0.3 nm from PHE 90 respectively after 100 ns (Figure 6.8). With respect to contact in the case of lipid and MAC binding it was seen that lipid once again bound to a separate region of the pocket, where highest contact was with residues buried deeper in the pocket, generally interacting with PHE or THR residues.

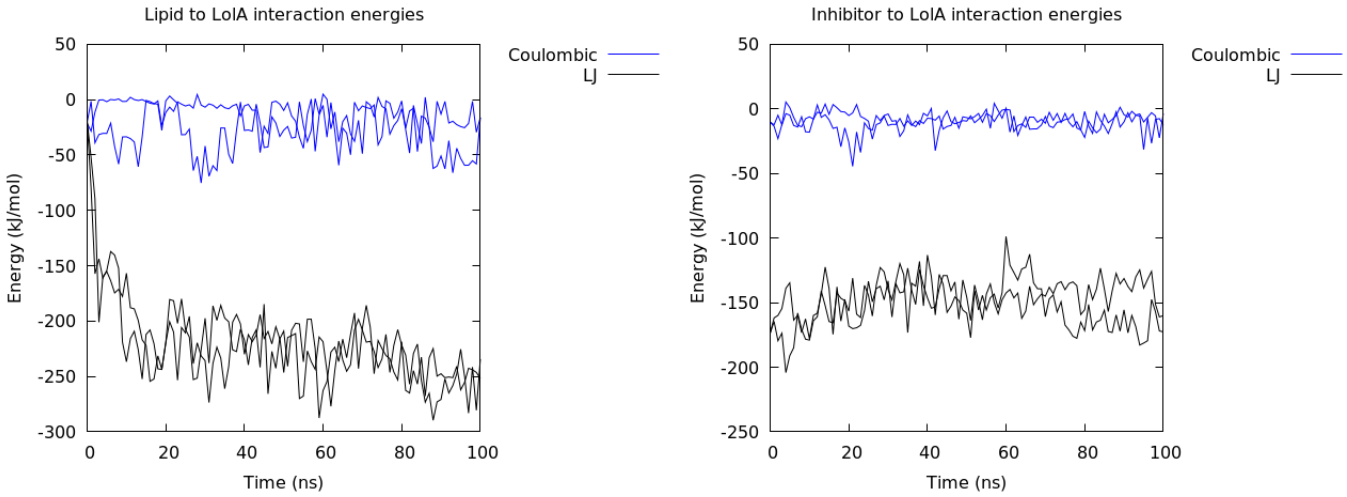


Figure 6.9: The short range electrostatic and Lennard-Jones interactions between the Lipid and the LolA, when one MAC is bound (left). The short range electrostatic and Lennard-Jones interactions between one MAC molecule and LolA (right).

In theory, inhibitors should result in a weakened interactive energy between lipid and the LolA pocket. The interaction between inhibitor and LolA indicates that the inhibitor molecules behaved in a similar manner to lipids when bound to the cavity (Figure 6.9). LJ interaction formed the bulk of the stabilising energy in these systems, where there was a fluctuation between -100 to -200 kJ/mol over 100 ns, as in Figure 6.9. From this it can be seen that the inhibitor bound LolA favourably, but less so than the lipid group on BLP. Lipid to LolA interactions are seen to behave similarly to the previous data indicated in the apo state (Figure 6.4) with LJ energy quickly stabilised within the first 20 ns, $\sim$ -250 kJ/mol. Based on the placement of the inhibitor inside of the LolA cavity the minimum distance shows that the MAC molecule was always contacting the LolA, where the minimum distance in any simulation was between 0.18 and 0.32 nm. The RMSD of the inhibitor molecules shows that there was limited variation, of between 0.1 - 0.3 nm across independent repeats.

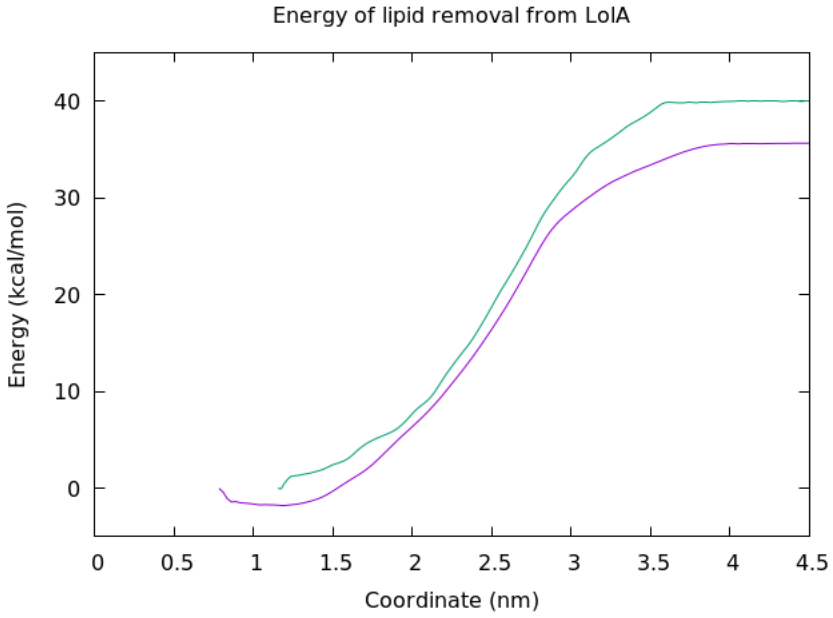


Figure 6.10: The umbrella sampling energy profiles for the removal of BLP lipid from the one MAC bound LolA cavity in two independent repeats.

Umbrella sampling across two independent repeats for lipid pulling from the LolA cavity in the presence of MAC shows two profiles of a similar shape, that converged to slightly different values. These values are $\sim$ 35 kcal/mol and 40 kcal/mol, but do not have the same starting value due to scrambled structures between repeats (Figure 6.10).

This once again shows that the LolA was a favourable environment for the lipid to be bound to, where a large energy penalty resulted in the lipid removal from the pocket. Compared to the apo sampling, it was seen that the energy involved to remove the lipid in the presence of the MAC inhibitor was lower (Figure 6.5). Based on two repeats of this sampling there was a reduction in the range of 10 - 15 kcal/mol.

It is worth noting that when one MAC inhibitor was present this was the equivalent of a concentration of around 375 μg/ml, considering the system size. The MIC of MAC is 16 μg/ml, indicating that at much lower concentrations than this system the macroscopic effect of this is a reduction in LolA function.[215] Despite this, binding was still possible at the microscopic level.

2 MAC Molecules

Despite having already simulated the effect of MAC binding to LolA at a concentration higher than the MIC, it was still possible for more than one inhibitor to bind to LolA (Figure 6.11). Simulations of two inhibitor molecules, via manual cavity placement, showed a similar effect, where there was binding, despite the concentration effectively equalling greater than 40 times the minimum amount required. BLP continued to bind to LolA, with the inhibitors rested at the mouth of the LolA cavity in this example where BLP could entangle with them, further adding weight to the "hydrophobic entanglement" effect that the inhibitors may cause.

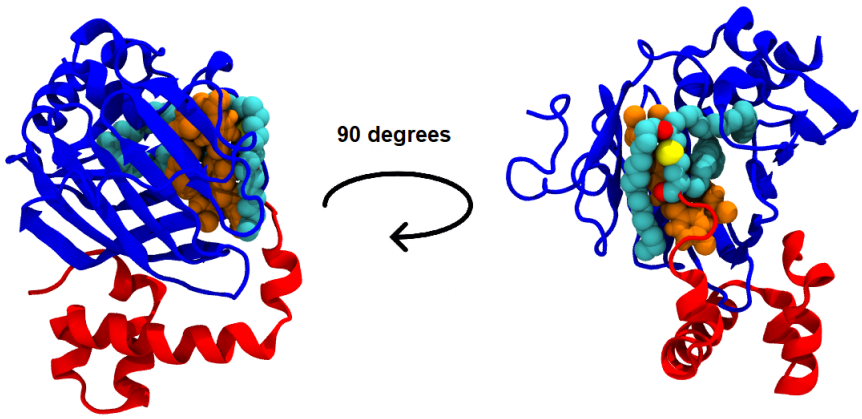


Figure 6.11: A visualisation that demonstrates LolA (blue) to BLP (red) lipid (cyan) binding in the presence of two MAC inhibitors (orange), where the lipid tails are obstructed by MAC, at a 90 degree rotation angle between the images (left and right).

From increasing amounts of MAC it was observed that the MAC will self associate, here seen bundled into one cluster (Figure 6.11). This clustering effect then clearly can block the lipid tails from entering the pocket. One lipid tail was shown to be parallel to the mouth of the pocket, where the cluster of inhibitors prevented the extension of this tail into the pocket. Despite this, other tails extended into the pocket and maintained BLP binding.

It has been shown that the energy required to remove lipid was lower in MAC presence; it can be proposed that increasing MAC concentration caused

increased BLP-MAC binding, which whilst favourable, was weaker than the BLP-LolA binding. Based on this the molecular mechanism of inhibition was not binding site obstruction, but rather of a scaffold of MAC molecules which was a poorer binding site than the LolA pocket.

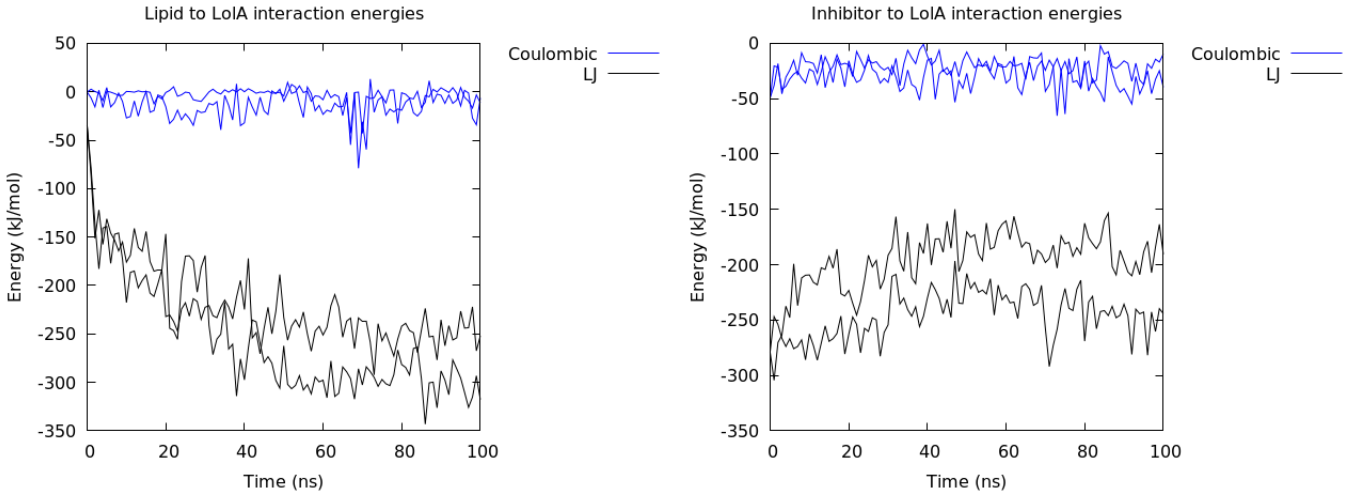


Figure 6.12: The short range electrostatic and Lennard-Jones interactions between the Lipid and the LolA, when two MAC are bound (left). The short range electrostatic and Lennard-Jones interactions between the two MAC molecules and LolA (right).

Energetically similar magnitudes were once again observed with respect to the lipid and MAC interaction. Charged energies in both systems were generally fluctuating around the same value of up to -50 kJ/mol, regardless of MAC presence. The hydrophobic interactions between MAC and LolA vary compared to the 1 MAC simulations. The initial LJ energy was more stable than in the one inhibitor systems. The simulation with higher LolA contact converged to $\sim$ -250 kJ/mol, where more contact with hydrophobic residues caused a greater LJ attraction. Where MAC binding was less this energy converges to $\sim$ -175 kJ/mol.

The lipid behaved in a similar fashion as before; there was a quick association from lipid to LolA, followed by a continuous increase in the attractive energy between the two. The two profiles reached separate values of around -250 kJ/mol and -300 kJ/mol respectively. Generally the largest increase in the attractive energy occurred in the first 40 ns, which correlates with the decrease of the MAC-LolA interactions. This shows that as MAC concentration increased the lipid and MAC were competing for binding sites in the cavity, where the lipid became more stabilised and the inhibitor less so.

A similar minimum distance between MAC and LolA was observed of ~ 0.2 nm, this in conjunction with the energetics indicates constant association of the MAC molecules and the protein. There is once again no ejection of the MAC observed; expectations are that as MAC concentration increases the total bulk of the cavity complex, lipid and MAC, increases, therefore essentially meaning that there is increased volume in the cavity. The MAC distance to the core PHE90 was also a near constant of ~ 0.3 nm across all replicas. When lipid maintained this distance it was deeply bound to LolA; this indicates that the 2 MAC complex was deep in the LolA cavity. Increasing the concentration of MAC led to an intercompeting effect between lipid and inhibitor, indicating that the stability of lipid binding was dependent upon the orientation and position of the MAC.

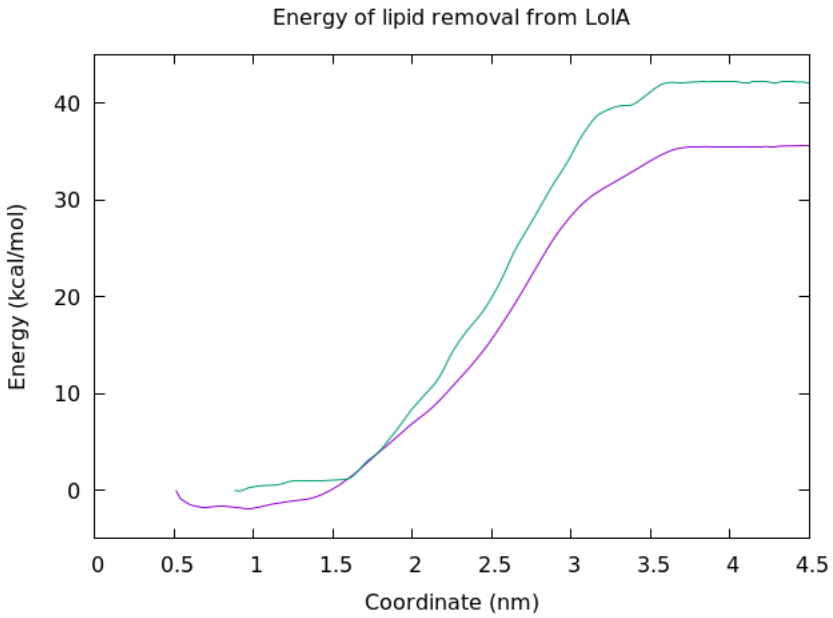


Figure 6.13: The umbrella sampling energy profiles for the removal of BLP lipid alone from the two MAC inhibitors bound LolA cavity.

Pulling the lipid from the protein in the presence of 2 MAC inhibitors once again showed that the removal of the lipid was disfavoured. The total energy for separate repeats gave approximate values of 42 and 35 kcal/mol starting from separate values due to scrambled starting coordinates (Figure 6.13). These energies were less than the apo state enhanced sampling, confirming that 2 MAC complexation with lipid in the LolA cavity causes less favourable binding than the presence of no inhibitor. According to this lipid binding in the presence of

the 2 MAC molecules is roughly as favourable as 1 MAC, this could be because the protein is already saturated at the point of 1 MAC being bound, alongside sampling issues.

3 MAC Molecules

Finally, insertion of 3 MAC molecules into the LolA protein prior to ligand binding was performed, using manual placement (Figure 6.14). The number of atoms of MAC in the pocket now outnumbered the lipid. In this example one tail of the lipid was wrapped around the complex, whilst the other two were blocked from entry by the MAC molecules. It is proposed that when the site is occupied by 3 MAC there is not enough space for full lipid binding to LolA, where the lipid mainly associates with the MAC molecules.

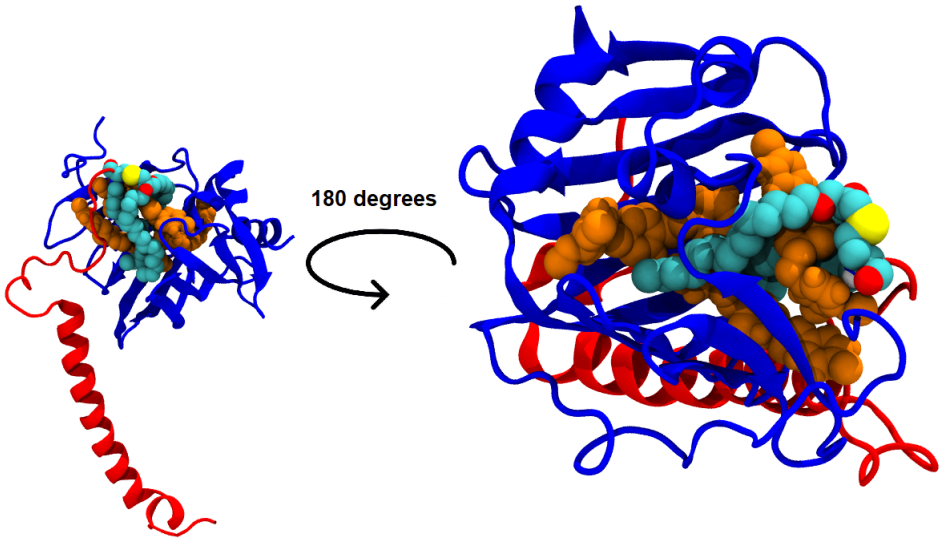


Figure 6.14: A visualisation of LolA (blue) to BLP (red) lipid (cyan) binding in the presence of three MAC inhibitors (orange), where closer inspection (right) shows that the three MAC inhibitors bind deeply to the pocket of LolA.

Even though the lipid-LolA binding was inhibited, there was still close association of the BLP to the LolA chaperone. It is clear that increasing MAC content affects lipid binding; however in our simulations, there was no driving force for BLP to unbind from LolA. As the lipid has been shown to compete for binding sites with the inhibitor, it is proposed that the in high concentrations of MAC the inhibitors are pushed further into the cavity on lipid binding, blocking the deepest binding residues within the pocket.

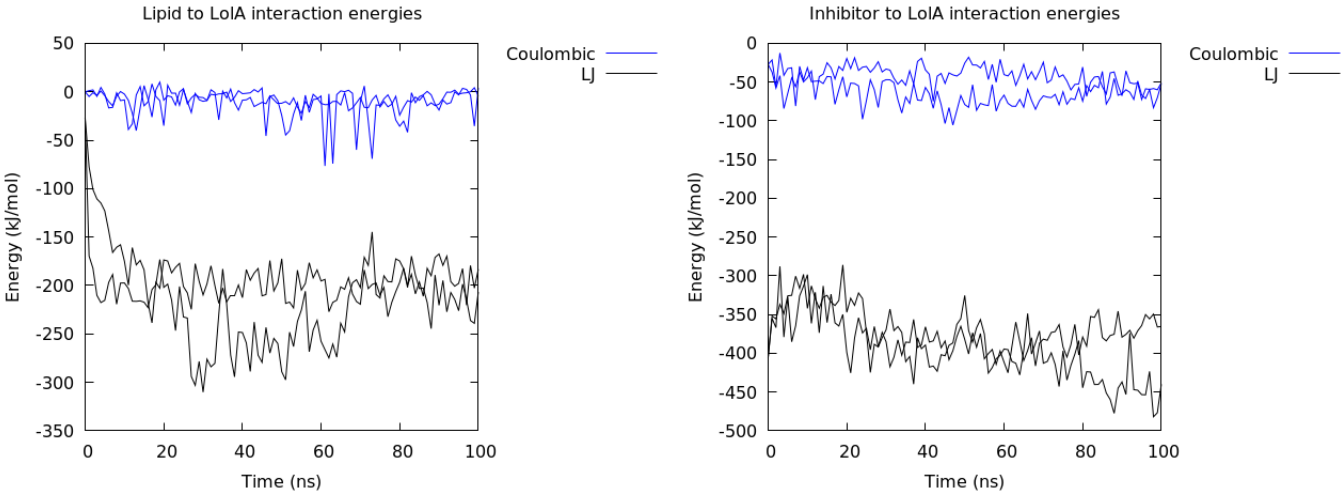


Figure 6.15: The short range electrostatic and Lennard-Jones interactions between the Lipid and the LolA, when three MAC are bound (left). The short range electrostatic and Lennard-Jones interactions between the three MAC molecules and LolA (right).

Charged interactions were not important for lipid to LolA association. Otherwise a similar rapid association between lipid and LolA was observed, and the hydrophobic energy stabilised in a range of -200 to -250 kJ/mol after 100 ns. Considering all systems from 0, 1 and 2 MAC bound, this was the least stable binding of the lipid to the pocket. Directly compared to the apo state where deepest binding occurs, lipid association was approximately 100 kJ/mol weaker. Obstruction of hydrophobic sites in the pocket led to this energy change.

Inhibitor attraction to the pocket was observed, however, to be the strongest out of any of the MAC concentrations. This indicates that increasing the amount of MAC in the cavity increased the stability of MAC binding. An increase in attractive energy is expected as there were an increasing number of atoms of inhibitor. The 3 MAC complex experienced a stabilising LJ energy at around -350 kJ/mol and -450 kJ/mol in two separate simulations, after 100 ns (Figure 6.24). This may be compared to 1 MAC, where a value of approximately -150 kJ/mol was reached in the same timeframe (Figure 6.9). Here there is more evidence of competition between the lipid and inhibitor, as the further stabilisation to -450 kJ/mol occurred simultaneously with lipid destabilisation from -300 to -200 kJ/mol.

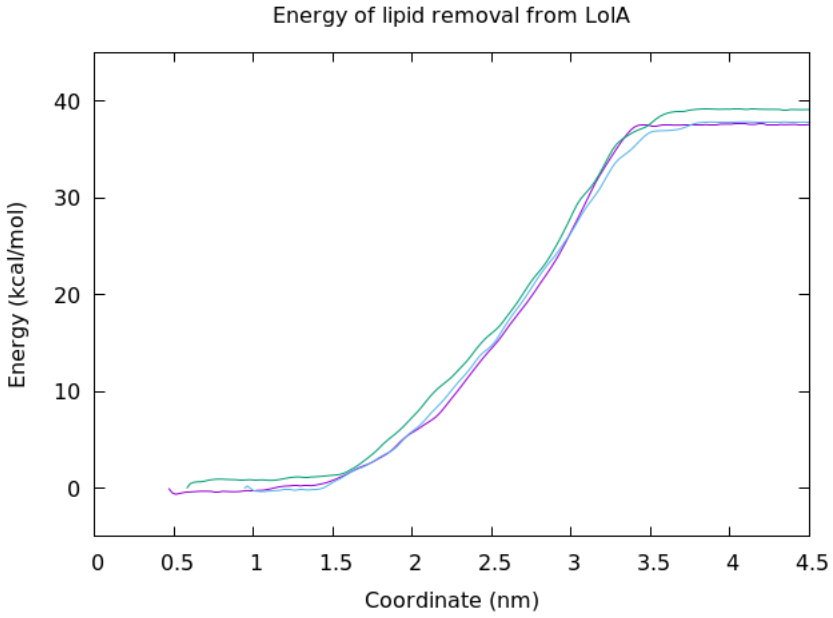


Figure 6.16: The enhanced sampling energy profiles for the removal of BLP lipid alone from the three MAC inhibitors bound LolA cavity in three independent repeats.

Based on data shown so far, the presence of 3 MAC should mean that the lipid was more easily removed from the binding site than in the apo state. The energy associated with this removal was in the range of 38 to 40 kcal/mol, Figure 6.25. Monitoring the PMFs across each concentration of MAC indicates that in the presence of inhibitor the lipid removal was always more favourable than the apo state. A range of values are seen for the 1, 2 and 3 MAC sampling, where there was not significant evidence to suggest which one of these was most favourable for lipid removal. From this it can be seen that MAC binding at all levels allows the lipid to be removed more easily. However there is not evidence to confidently say that there is a marked difference when comparing the MAC concentrations between the simulation systems. For verification that overlapping histogram data was achieved to indicate good sampling please see Figures 2.7 - 2.10 regarding umbrella sampling.

6.4.3 Apolipoprotein D

Having assessed LolA ligand binding, the ApoD chaperone discussed in the premise of Chapter 6 has been studied. The structure of ApoD is similar to LolA, where there is a hydrophobic pocket that is open to the solution in one face, where an unstructured loop blocks the end of the cavity. As previously mentioned this is a human lipocalin, that has been expressed in *E. coli*.

Non-glycosylated ApoD

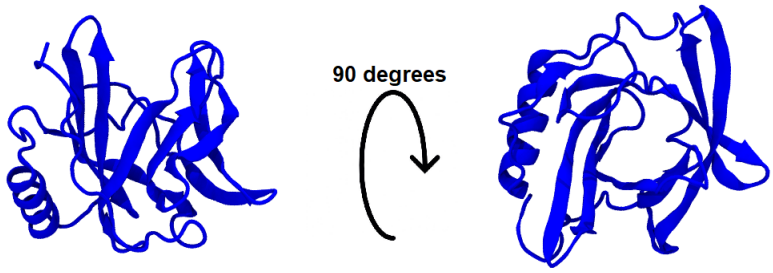


Figure 6.17: Visualisation of the non-glycosylated ApoD monomer (blue), where the chaperone has been rotated around 90 degrees, (left and right)

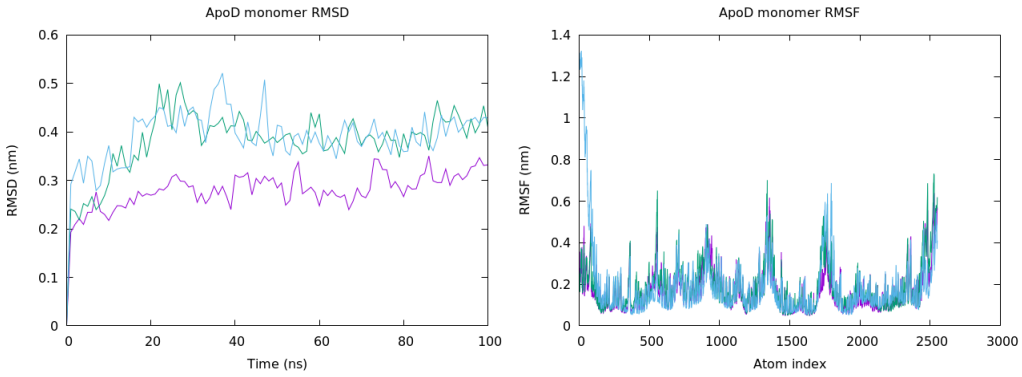


Figure 6.18: The root mean square deviation of the ApoD monomer, across independent repeats (left). The root mean square fluctuation of the ApoD monomer across independent repeats (right).

Simulation of the ApoD monomer, without glycosylation was performed. This showed that the protein was a stable unit in an aqueous environment. RMSD

of the protein plateaued to 0.2 - 0.4 nm. The RMSF was generally less than 0.6 nm, aside from the N terminus, which was more flexible than the rest of the protein in Figure 6.18.

Glycosylated ApoD

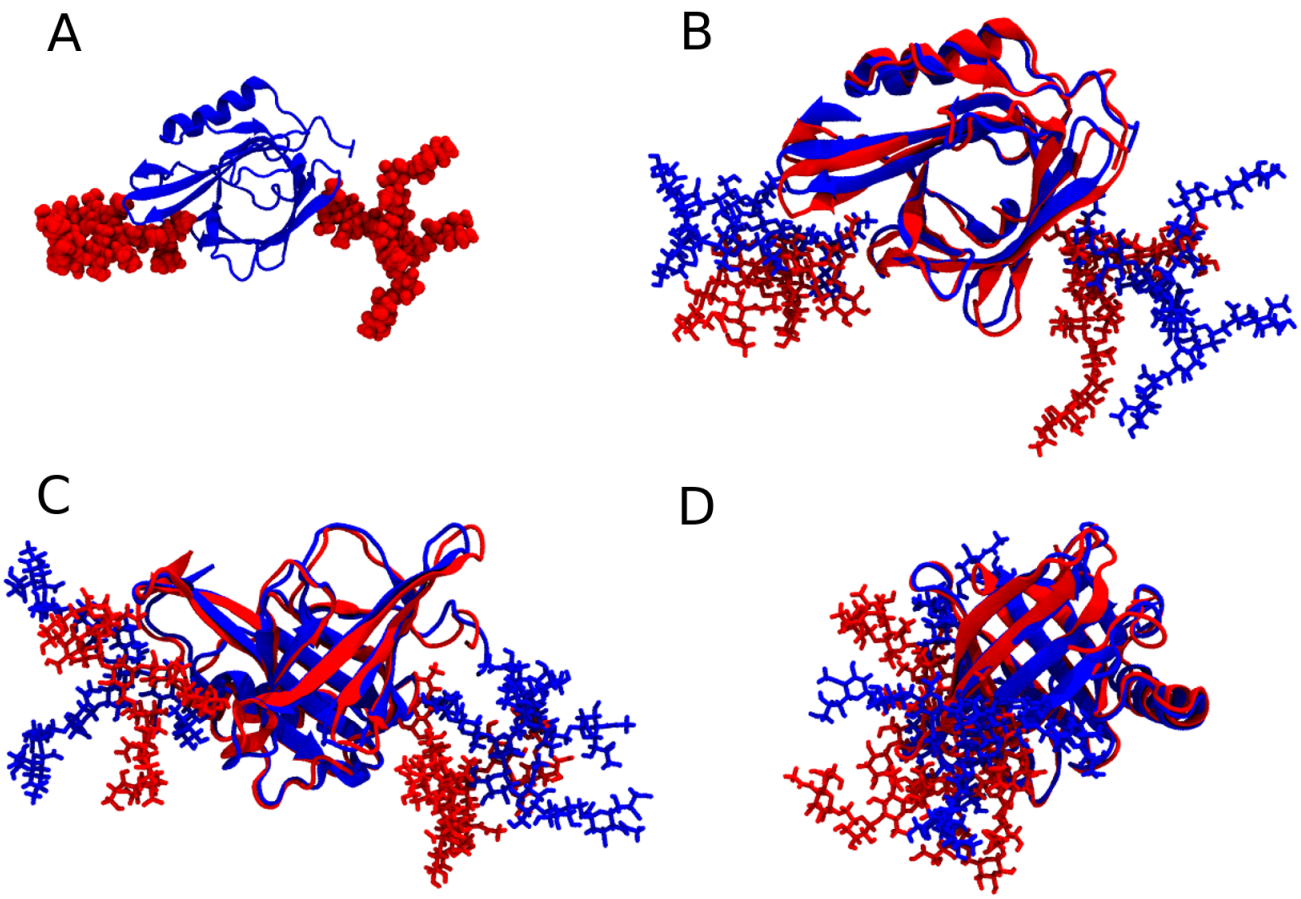


Figure 6.19: Visualisation of the ApoD monomer (blue) after glycosylation (red), (A). Aligned structures of the ApoD glycosylated monomer after 100 ns simulation (red), 0 ns (blue) from three angles, (B, C and D).

Construction of the ApoD glycosylation used experimental information proposing two separate sites that are functionalised with different glycans. These glycans extended into the space either side of the hydrophobic pocket of the lipocalin (Figure 6.19). Given their location it is extremely likely that the glycans on the lipocalin are actively involved in ApoD *in vivo* oligomerisation.

The glycosylation added a highly flexible, hydrophilic region to ApoD. Alignment of this structure before and after 100 ns of simulation shows that the glycan occupied a volume around the barrel of the protein, however that these glycan units were facing the direction opposite to the helical region flanking the hydrophobic barrel. These glycan regions did not associate with the cavity, where the pocket was still open and presented to the solution, indicating that they will not have a significant effect on the role of the monomer as a lipocalin.

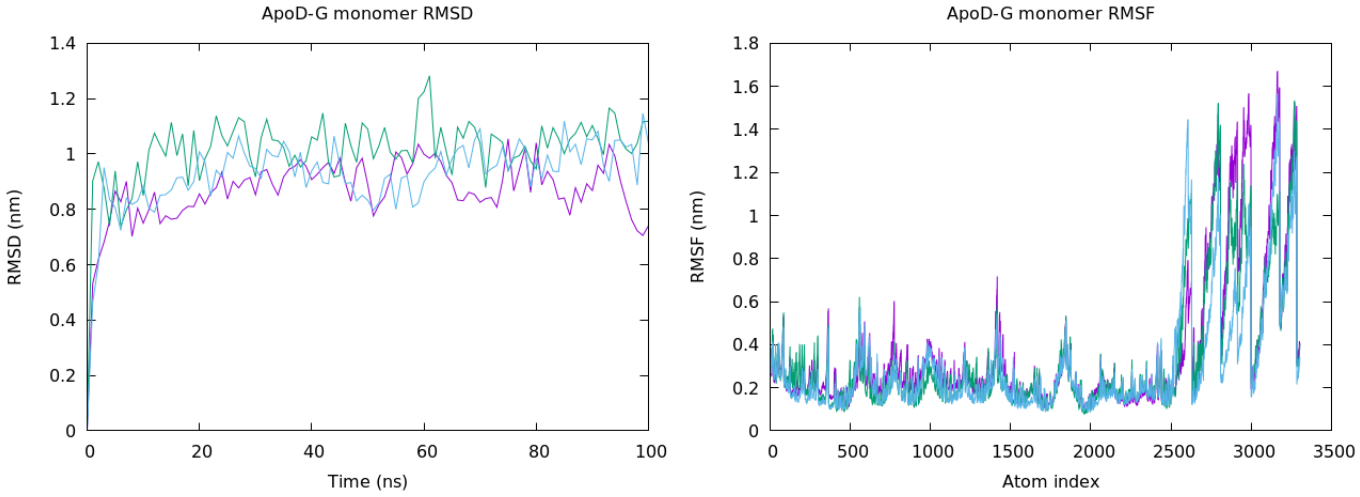


Figure 6.20: The root mean square deviation of the ApoD glyco-monomer, across independent repeats (left). The root mean square fluctuation of the ApoD glyco-monomer across independent repeats (right).

It can be seen that the presence of the glycans increased the RMSD of the ApoD monomer (Figure 6.20). RMSD values of the monomer are in the range of 0.8 - 1.2 nm over 100 ns. This increase is attributed to the glycan units, as the protein remained structurally similar to the non-glycosylated monomer. RMSF of the protein was similar in the protein region, where large fluctuations were observed in the glycan regions that are added to the protein. In comparison to core protein regions, the glycans were five times more flexible. From this it is observed that the fluctuation and structure of the protein region was highly similar after simulation with glycan regions. Thus, the biologically relevant glycosylated form of ApiD was used in subsequent ligand binding studies.

6.4.4 ApoD ligand binding

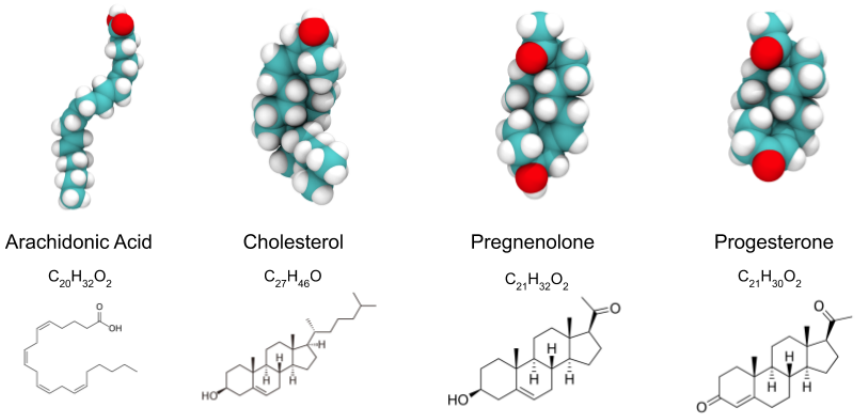


Figure 6.21: Illustrations of the ApoD binding substrates and their skeletal structures.

Multiple binding studies exist to indicate that the following ligands are possible binding partners for ApoD: arachidonic acid, progesterone, pregnenolone and cholesterol (Figure 6.21). It has been proposed that ApoD can carry such a wide variety of ligands that the protein could be multi-functional, unlike LolA which is proposed to be dedicated to lipoprotein transport.

Arachidonic Acid

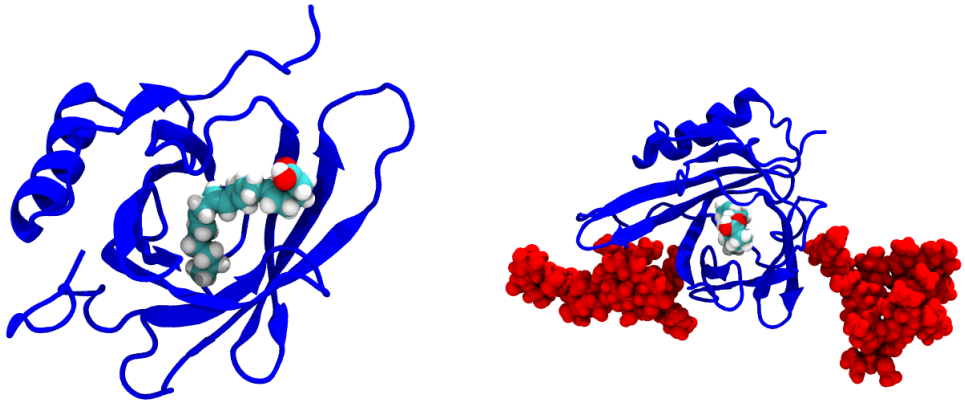


Figure 6.22: Arachidonic acid binds to the cavity of the glycosylated ApoD monomer (blue), glycans (red) and ligand (cyan, white, red).

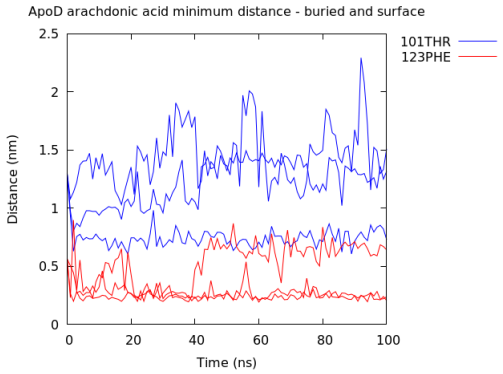


Figure 6.23: The minimum distance between arachidonic acid and the buried 101THR residue and the surface 123PHE residue.

Unbiased substrate binding to ApoD is possible, where experimentally the specificity of this binding with small organic ligands has been proposed. The glycans self interact with the external regions of the protein and do not block the binding of the substrate to the monomer. The interaction of arachidonic acid, cholesterol, progesterone and pregnenolone with the glycosylated monomer has been assessed. All four of these small molecules are ubiquitous in human biology, but

exist to serve difference purposes. It can be seen that the hydrophobic pocket of ApoD admitted these lipids, where when bound to the pocket there is limited contact between ligand and glycans (Figure 6.22).

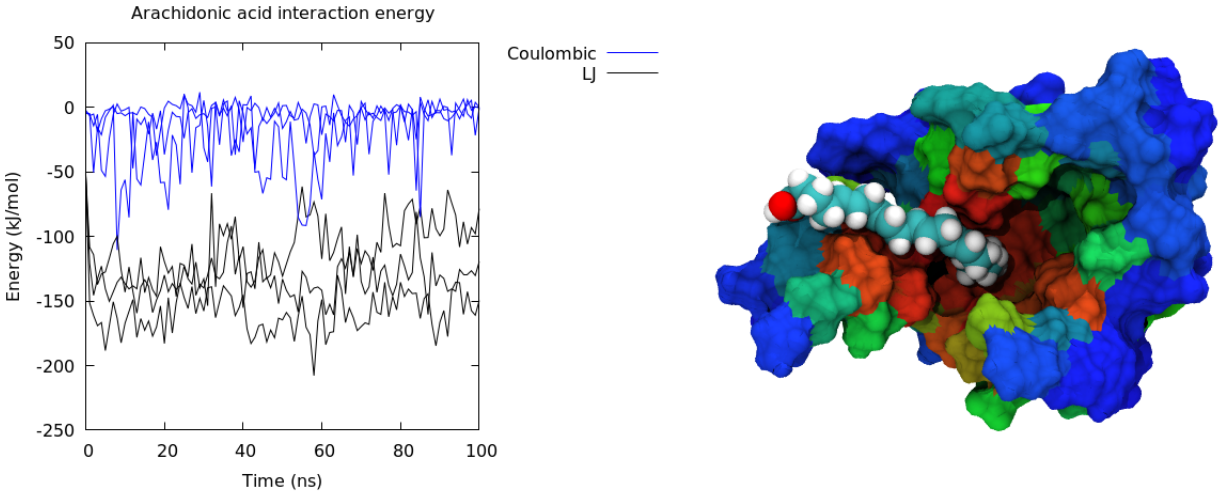


Figure 6.24: The short range electrostatic and Lennard-Jones interactions between arachidonic acid and the ApoD cavity (left). A visualisation of the arachidonic acid contact with ApoD, using the BGR colour scheme where red is high contact and blue is low contact. The arachidonic acid is coloured according to name (right).

Arachidonic acid is proposed to bind the ApoD cavity strongly in Figures 6.23 and 6.24. Placing the ligand near (within 1.0 nm distance) the hydrophobic pocket and allowing free entry of the acid led to spontaenous binding and the ligand then remained constantly associated with the protein. Further measurement of this, using a phenylalanine residue near the mouth of the pocket, and a threonine residue buried in the core of the pocket showed this association further. From this it can be seen that there was a close association to the PHE residue, whilst there was a varying distance to the core THR residue. Distance to 123 PHE was normally less than 0.5 nm throughout two repeats whilst in the third there was a movement from this residue. The substrate whilst bound varied in distance from the buried THR between 0.7 - 2.3 nm.

The unsaturated carbon chain was closely associated with the deepest regions. This interaction was generally maintained by a mainly hydrophobic association. LJ energy was attractive, in the range of -100 - 200 kJ/mol. The charged energy involved was minimal, generally fluctuating between 0 and -50

kJ/mol. This energy was a result of the head group interacting with the mouth of the cavity, where the acidic carboxylate group can be seen pointing toward the surface of the protein. It is surprising that the arachidonic acid is experimentally proposed to have a higher affinity than the other ligands, where there is no ring like structure present in the hydrophobic region. Given this, there is only one charged region which may stabilise binding.

Cholesterol

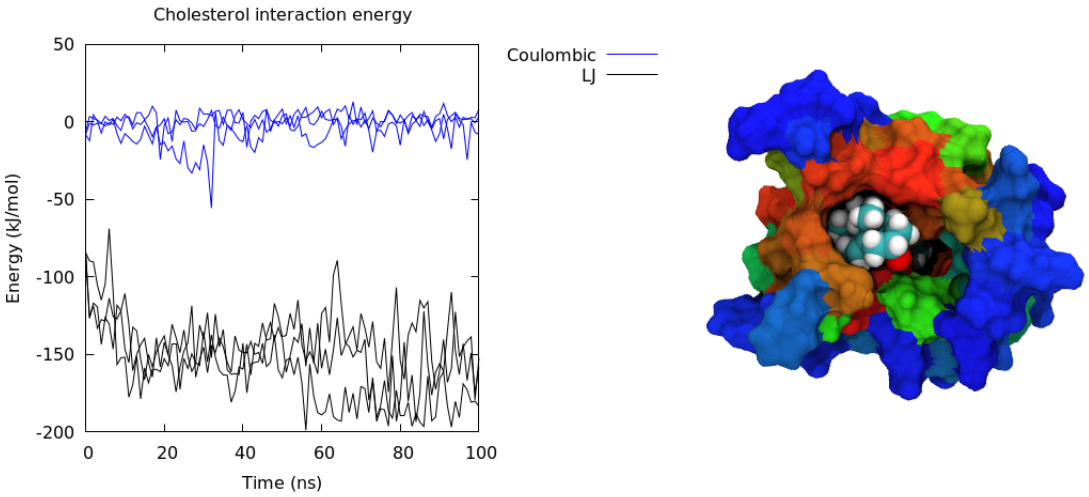


Figure 6.25: The short range electrostatic and Lennard-Jones interactions between cholesterol and the ApoD cavity (left). A visualisation of the cholesterol contact with ApoD, using the BGR colour scheme where red is high contact and blue is low contact. The cholesterol is coloured according to name (right).

Cholesterol binds to the pocket of the protein, where the ligand was also deeply associated with the cavity. Experimentally cholesterol is proposed to have a weak binding affinity to ApoD. Energetically the association was favourable. Whilst there were varying degrees of LJ energy fluctuation in the arachidonic acid simulations, here cholesterol was shown to stabilise quickly, where after 100 ns, the LJ energy reached the value of $\sim$ -175 kJ/mol (Figure 6.25). Relatively the coulombic activity contributed less to the binding, generally less than -25 kJ/mol. In the cavity it was seen that the cholesterol was coiled up, where the majority of the ligand was bunched deep in the pocket. In contrast, the arachidonic acid was extended, laying along the hydrophobic cavity to the extent that it was surface exposed to the solvent. The ring like versus straight chain ligand

structure could hold the key to the binding stability.

Examining the LJ energy alone, it appears that the cholesterol was more stable in the pocket than arachidonic acid, indicating a disagreement with experimental data. However it is worth noting that the arachidonic acid ligand used was the neutral form of the acid, where the charged form could experience a different interaction. However, based on the hydrophobic pocket, the charged headgroup would be less stable in this environment. Another relevant factor not taken into account is potential ligand aggregation. In addition, the experimentally known propensity for ApoD oligomers to form could be important. Regardless of the arachidonic acid comparison, this data indicates that the cholesterol has a similar affinity to the ApoD pocket.

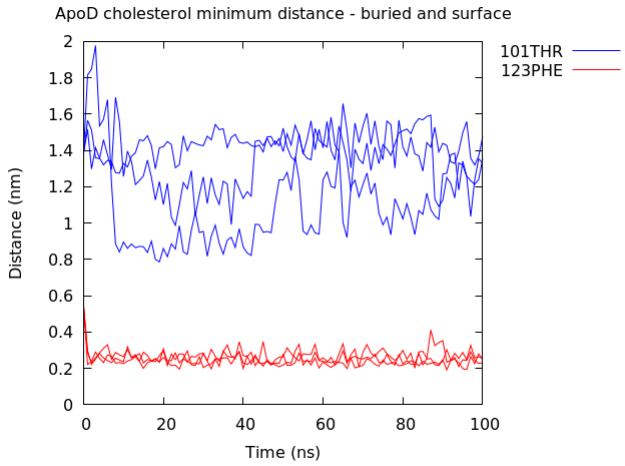


Figure 6.26: The minimum distance between cholesterol and the buried 101THR residue and the surface 123PHE residue.

Based on the residues in the cavity, it can be seen that the PHE residue was consistently in contact with cholesterol throughout the simulation once the initial binding occurs. With respect to the buried THR residue, generally contact did not occur, where the ligand was at least 1 nm away from this buried region. This was similar to the binding of arachidonic acid, where there was closer association to PHE123 here, which makes sense as the hydrophobic interactions between the cholesterol will be stronger with ring like residues.

Pregnenolone

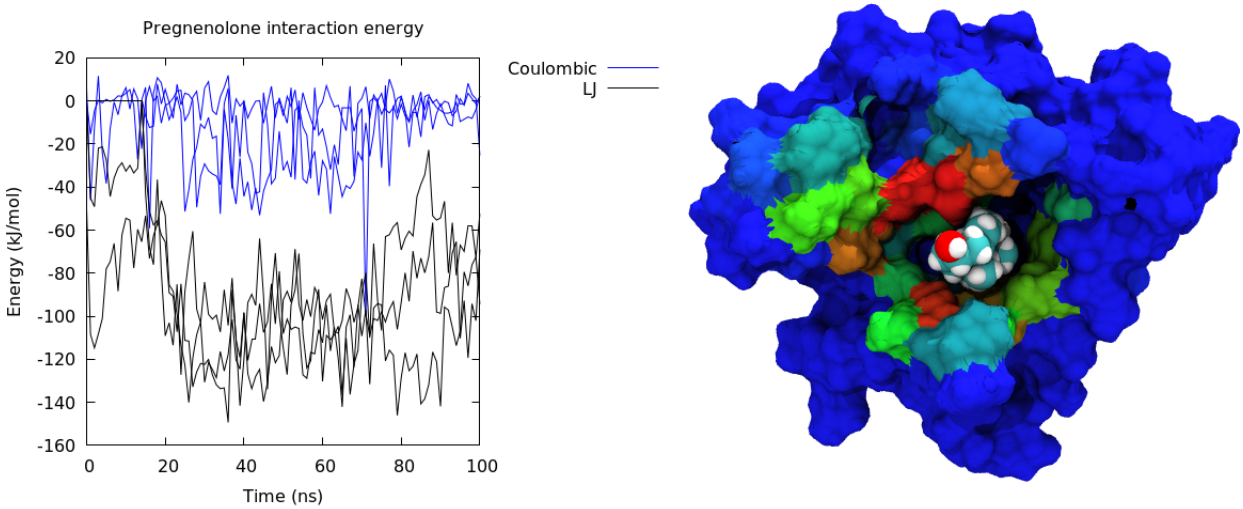


Figure 6.27: The short range electrostatic and Lennard-Jones interactions between pregnenolone and the ApoD cavity (left). A visualisation of the pregnenolone contact with ApoD, using the BGR colour scheme where red is high contact and blue is low contact. The pregnenolone is coloured according to name (right).

ApoD has been proposed to bind hormone type ligands. Pregnenolone is a small ring like molecule, where in contrast to arachidonic acid and cholesterol, it lacks a straight chain carbon moiety. Relative to the previous ligands, pregnenolone bound to the hydrophobic pocket of ApoD less favourably, Figure 6.44. When bound, the smaller molecule was associated deep into the cavity of the lipocalin. However despite the fact that little contact was made with the mouth of the pocket, the attractive energy was less between the protein and ligand.

Firstly in one repeat, there was no association for most of the first 20 ns, indicating that the ligand moved away from the mouth of the pocket. When binding did not immediately occur, followed by association to the pocket, an attractive coulombic energy that was generally less than - 40 kJ/mol was formed. This energy fluctuated between this value and 0, indicating once again that charged interactions were not key for association. The LJ energy was less after 100 ns for the binding of pregnenolone. It can be seen that at this point the stabilising energy was less than -100 kJ/mol. However this interaction was not consistent, and varies in stability. In comparison to the previous ligands, not only was the overall association weaker, it was not maintained for the entire 100

ns in all independent repeats.

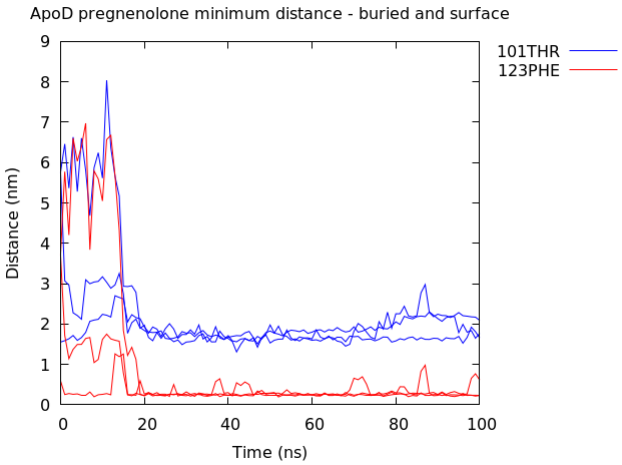


Figure 6.28: The minimum distance between pregnenolone and the buried 101THR residue and the surface 123PHE residue.

Using the distance to the surface and buried areas, it can be seen that there was a close association to the PHE123 residue in the ApoD, where generally a distance less than 1 nm to this residue was maintained over 80 ns (Figure 6.28). In this time period this distance was mostly smaller than this. In one simulation, even as binding occurred, close association to the PHE123 residue was not established until 18 ns. With respect to the buried THR101, once association occurred, a distance of approximately 2 - 2.5 nm from this residue was established for 80 ns. Pregnenolone therefore upon binding did not reach as far into the pocket as the other ligands, explaining the poorer binding affinity.

Pregnenolone has been shown experimentally to have a poorer binding affinity than arachidonic acid, as supported by these simulations.

Progesterone

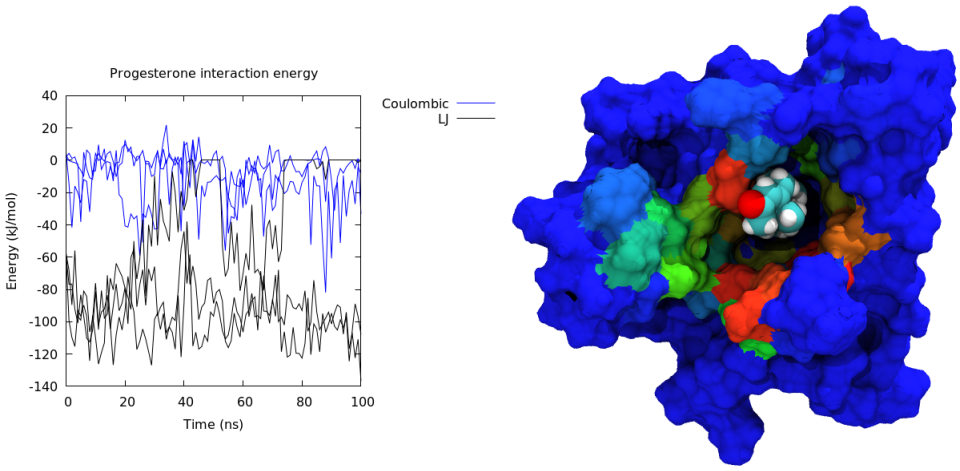


Figure 6.29: The short range electrostatic and Lennard-Jones interactions between progesterone and the ApoD cavity (left). A visualisation of the progesterone contact with ApoD, using the BGR colour scheme where red is high contact and blue is low contact. The progesterone is coloured according to name (right).

The last ligand assessed was the hormone progesterone (Figure 6.29). Progesterone and pregnenolone are highly similar in structure, where the variation between the two is the placement of one double bond in a ring and a ketone versus an alcohol functional group at one end of the molecule. Based on this the binding properties of progesterone would be expected to be highly similar to pregnenolone, as they occupy a similar shape in space and chemical structure. With respect to the binding it can be seen that this was true. The small hormone rested deep in the cavity of the ApoD, with little to no relative contact with the surface residues. Similar to pregnenolone, this ligand is proposed to have a poorer affinity to the ApoD than arachidonic acid.

Once again, poorer binding affinity than either cholesterol and arachidonic acid binding was seen. During simulation the charged energy varied, where the interaction was generally less than -40 kJ/mol. Dissociation of the progesterone was observed; the interaction energies were 0 at several points during one simulation. Prior to this the hormone was bound and interacting favourably, indicating that the ApoD monomer ejected this molecule after binding.

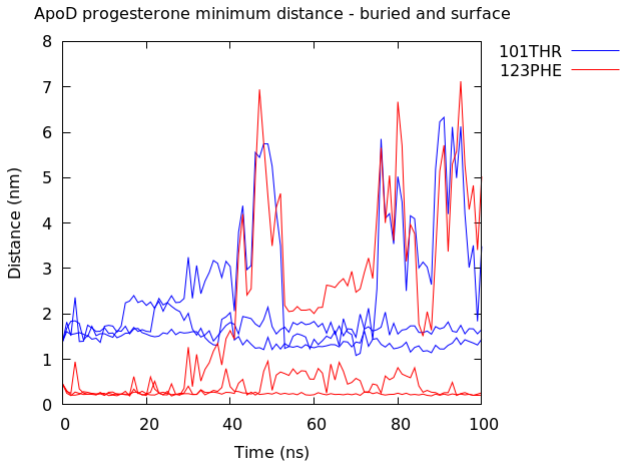


Figure 6.30: The minimum distance between progesterone and the buried 101THR residue and the surface 123PHE residue.

The attractive LJ energy for progesterone binding generally ranged from -60 to -120 kJ/mol in two simulations, whilst in the other at 20 ns there was a destabilisation until dissociation occurs (Figure 6.29). This follows as the distance dependent interactions were 0 kJ/mol for most of the latter half of the simulation.

Distance to the surface and buried regions shows that the ligand maintained close contact, 0.2 - 1 nm with the surface PHE123 residue when bound. A distance of less than 2 nm from the buried THR101 was maintained, where in this binding the hormone was deeper in the pocket than pregnenolone. In the unbinding simulation, the movement away from these residues began at 25 ns, where by 40 ns contact was removed. Importantly, when the re-association of the ligand does occur, it does not interact with either region.

6.5 Discussion

Atomistic simulations of both the bacterial LolA chaperone and the human ApoD chaperone have been done. From these simulations, the binding of hydrophobic ligands to small protein chaperones, a vital biological interaction was shown to have common themes in mechanisms. With respect to the LolA chaperone, the binding of BLP has been observed in the hydrophobic pocket. This binding is observed over 100 ns, stabilising the BLP monomer for lipoprotein transport. The inhibitor MAC13243, known as MAC, binds stably to the pocket of the LolA during simulations. This inhibitor occupies a smaller space than the BLP lipid, indicating that multiple inhibitors could access this site.

When MAC is bound, the binding of the lipid moiety could still be observed in the LolA chaperone. The lipids can interact favourably with, and wrap around the inhibitor molecules. From this an inhibitor-lipid complex is formed that is stable inside of the LolA cavity. This cavity is hydrophobic in nature, where the MAC occupies the binding sites inside the chaperone. These sites generally contain amino acids with ring-like R groups, where the proposed interaction occurs due to the pi-stacking of the inhibitor molecules and hydrophobic attraction of the lipid molecules.

Inhibitor concentration was observed to result in a reduction of the lipid contact to the protein cavity. The number of unique amino acids inside of the cavity contacting the lipid was inversely proportional to the number of inhibitors bound. As a result of this the two different structures are competing to access hydrophobic sites in the protein structure.

Energetically, competition between the lipid and inhibitor was observed, where as inhibitor concentration increases the hydrophobic interaction between lipid and protein is reduced. This intercompetition results in significantly reduced Lennard-Jones interactions between the lipid and LolA cavity in the highest concentration of inhibitor, relative to the lowest concentration of inhibitor present. At higher concentrations the inhibitor is pushed into the cavity, essentially blocking the lipid from binding to deeper regions of the LolA. However PMFs do not indicate a significant difference between MAC concentrations.

Free energy calculations indicated that lipid binding is less favourable when inhibitor molecules are present in the cavity, when compared to uninhibited states of the protein. The presence of the lower concentrations of MAC indicated that introduction of any number of inhibitors allow for easier removal of the lipid from the cavity. *In vivo* this indicates that the transport ability of the LolA will be reduced by this inhibitor, where the propensity for the bound lipoprotein to become dislodged during transport across the periplasm is increased.

At this point a discussion surrounding the validity of the umbrella sampling is required. Considering that the state of the MAC molecule could be charged or uncharged inside the protein cavity, alongside the use of the GROMOS forcefield could potentially affect these results. Umbrella sampling is notoriously hard to achieve flawless convergence, where local convergence is generally the accepted level. Looking at this system, a lipid is pulled from a complex protein environment, which has had inhibitor, sometimes multiple inhibitors added. The complexity of this and the degrees of freedom that are sampled in the umbrella sampling windows, i.e. one direction only, mean that there are a large number of variables that are not being considered. Having stated all of this the following conclusion can be drawn: in simulation the MAC inhibitor will lower the energy required to extract the lipid from the cavity, however the varying concentration

of MAC is a complex problem, where it is not possible to draw conclusions between different MAC molecules. In addition, unrealistic states could be forced by packing in a large number of MAC inhibitors, which once again brings the PMF results into doubt. However, with respect to the sampling we can see that we have at least sampled part of the pathway well.

From this it has been seen that the inhibition of the LolA chaperone is not a simple case of binding site obstruction. Experimentally, MAC has been shown to increase the permeability of the *E. coli* cell. As BLP is the structural staple holding the cell together, a reduction in BLP localisation will inherently destabilise the double membrane envelope architecture. Based on previous work including BLP, it is obvious the landscape of the periplasm is very different without the OM to PGN association caused by the lipoprotein. In this context, LolA is an important chaperone that is inhibited by the MAC molecule, without the binding of the lipoprotein being excluded, according to simulation.

In contrast to all other simulations present in this thesis, Apolipoprotein-D (ApoD) is a human glycoprotein. It is known to be a transporter and chaperone, with many proposed ligands and interesting associations with human biology.

ApoD, similarly to LolA, will consistently bind to hydrophobic ligands. The specificity of this has been studied, where due to the number of potential ligands it is proposed to have many functions as a transporter. Generally it appears that ApoD will associate with a vast range of small hydrophobic molecules. In these simulations highest affinity observed was for the cholesterol molecule, followed by arachidonic acid, where progesterone and pregnenolone are relatively unfavourable in their binding to the ApoD pocket. This agrees and disagrees with experimental conclusions, where the binding for arachidonic acid is described as favourable whilst cholesterol is not.

ApoD is known to function as a link between HDL and LDL particles. Considering this it could be that the transport of cholesterol between these two particles functions in this way, which would allow for the strong cholesterol association and provide a reasoning for high binding affinity. It is agreed with experiments that the progesterone and pregnenolone can bind, but poorly to ApoD. From this it is generally proposed that small hormone molecules bind poorly to the lipocalin, but that the longer, hydrocarbon chain containing, molecules such as arachidonic acid and cholesterol are better ligands for association.[224]

It is possible that the ApoD-ligand binding occurs in the dimeric, or even the tetrameric states which are known to occur biologically, where the quaternary structure may affect ligand binding. It has been proposed that the ApoD is a multi-ligand multi functional protein, with many biological purposes. It can be proposed that the quaternary structures could present different binding to ligands, where multiple ligands are bound at one time, however this remains to be studied in simulation.

For additional information on the topic LolA, BLP and inhibitor simulations, see Appendix F, for a publication titled **Details of hydrophobic entanglement between small molecules and Braun's lipoprotein within the cavity of the bacterial chaperone LolA**.

Chapter 7

Conclusions

Using molecular dynamics allows for the visualisation and analysis of biological systems. As has been seen these can range from a single protein in solution, to a complex compartment bounded by multiple membranes and other structures. Whilst simulations are an approximation to reality, potentially limited by timescales sampled, molecular level insights have been revealed in this thesis about many interactions occurring in *E. coli*, particularly in the periplasm, in conjunction with a brief view of a human chaperone, apolipoprotein-D.

Constructing the OM to PGN region of the periplasm in *E. coli* highlighted the importance of Braun's lipoprotein, where the proposed role as a structural staple was confirmed. Alongside this the dynamics of the cell wall and membrane as a result of the OM-PGN association provided by BLP have been revealed. Generally BLP is not a static staple, but a flexible hinge, where the movement of the lipoprotein can change the physical size of the periplasmic space. This change in the size of the periplasm allows contact with a proposed cell wall clamp, OmpA. From this it is confirmed that interaction via both covalent and non-covalent linkages to the cell wall maintain periplasmic integrity.

Expanding upon this with a novel simulation protocol for the full periplasm showed interactions on both side of the periplasm maintain the position of the cell wall. TolR interacts non-covalently with the peptidoglycan, where an open state conformation allows binding, but a closed state conformation will not bind. Once again BLP is shown to be a major player in these systems, where the consistent interaction between OmpA, TolR and PGN breaks down without the presence of the lipoprotein. Removing the non-covalent protein interaction from either side of the periplasm causes a shift in PGN height in the periplasm. From this it is inferred that the macroscopic interactions of many proteins throughout the entire periplasm will determine cell wall landscape.

The Mla protein system is key for maintaining the LPS barrier that *E. coli* is well known for. Simulations showed that the MlaD hexamer, MlaC chaperone

and MlaA membrane protein can all bind lipid. In the case of the membrane associated MlaD and MlaA, lipid removal from the associated bilayer was seen. Favourable interactions between MlaC and various lipids were shown, where the transporter can behave non-selectively and bind different lipid types. Modelling of the MlaD termini indicated that they have an important biological function, where various conformations of this region are possible. With modelled residues present, the MlaD can bind to the MlaC favourably, agreeing with experimental data. From this it can be seen that the various stages of the lipid transport pathway are possible. The overall mechanism of this lipid transport remains elusive, however further study and new data of the MlaFEDB complex, recently reported, could shed further light on this lipid transport pathway.

Chaperones are key for life processes in any organism, where the shuttling of substrates to varying cell locations is crucial. BLP can be transported by the LolA protein, where the binding of the lipid moiety of BLP to LolA has been shown. Inhibition of this binding is possible, shown both experimentally and in our simulations. Inhibition takes on an interesting definition in this context, where experimentally *E. coli* becomes more permeable, whilst simulation reveals a reduced tendency for BLP binding to LolA, but still a favourable environment for lipid association due to hydrophobic entanglement of the acyl tails to the MAC inhibitor molecules. This in turn brings into question how inhibition is understood in chaperones, as this inhibition could be a red herring, where increasing the inhibitor concentration in the LolA binding site has no significant effect on inhibition.

Parallels in the binding of small hydrophobic ligands to ApoD glycoprotein were also found. ApoD will bind small ligands, where disagreement occurs on cholesterol binding affinity to the chaperone, but that binding of arachidonic acid and hormones are within expectations.

Based on the data shown, it can be seen that simulation is a powerful tool, where as technology continues to improve, in the future it will be possible to examine biomolecular systems of increasing complexity at atomistic resolution. Contributing towards such efforts, there is no doubt that generating the basic information of protein contacts, polymer interactions and membrane behaviour will continue to be of use for many years to come.

Bibliography

[1] Thomas Silhavy, Daniel Kahne, and Suzanne Walker. The bacterial cell envelope., 2010.

[2] Martin Karplus and J. Andrew McCammon. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, 9(9):646–652, 2002.

[3] Teresa Paramo, Thomas J. Piggot, Clare E. Bryant, and Peter J. Bond. The structural basis for endotoxin-induced allosteric regulation of the toll-like receptor 4 (tlr4) innate immune receptor. *Journal of Biological Chemistry*, 288(51):36215–36225, 2013.

[4] Jakob P. Ulmschneider and Martin B. Ulmschneider. Molecular Dynamics Simulations Are Redefining Our View of Peptides Interacting with Biological Membranes. *Accounts of Chemical Research*, 51(5):1106–1116, 2018.

[5] Georg E. Schulz. The structure of bacterial outer membrane proteins. *Biochimica et Biophysica Acta - Biomembranes*, 1565(2):308–317, 2002.

[6] D. Jefferies and S. Khalid. *Molecular Simulations of Complex Membrane Models*. Elsevier, 2017.

[7] A. Boags, P.-C. Hsu, F. Samsudin, P.J. Bond, and S. Khalid. Progress in Molecular Dynamics Simulations of Gram-Negative Bacterial Cell Envelopes. *Journal of Physical Chemistry Letters*, 8(11), 2017.

[8] Gongpu Zhao, Juan R. Perilla, Ernest L. Yufenyuy, Xin Meng, Bo Chen, Jiying Ning, Jinwoo Ahn, Angela M. Gronenborn, Klaus Schulten, Christopher Aiken, and Peijun Zhang. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature*, 497(7451):643–646, 2013.

[9] Jianzhuo Zhu, Erkuang Zhu, Jing Gao, Xingyuan Li, and Jiguo Su. Structural and dynamic properties of water molecules in a uniformly charged nanopore. *The Journal of Chemical Physics*, 149(7):074703, 2018.

[10] Pekka Mark and Lennart Nilsson. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *Journal of Physical Chemistry A*, 2001.

[11] Adam Hospital, Josep Ramon Goñi, Modesto Orozco, and Josep L. Gelpí. Molecular dynamics simulations: Advances and applications, 2015.

[12] James R. Johnson, Brian Johnston, Connie Clabots, Michael A. Kuskowski, and Mariana Castanheira. Escherichia coli Sequence Type ST131 as the Major Cause of Serious Multidrug-Resistant E. coli Infections in the United States . *Clinical Infectious Diseases*, 51(3):286–294, 2010.

[13] Gregory J. Dick. The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally, 2019.

[14] Saad Nseir, Christophe Di Pompeo, Stéphane Soubrier, Pierre Delour, Hélène Lenci, Micheline Roussel-Delvallez, Thierry Onimus, Fabienne Saulnier, Daniel Mathieu, and Alain Durocher. First-generation fluoroquinolone use and subsequent emergence of multiple drug-resistant bacteria in the intensive care unit*. *Critical Care Medicine*, 33(2):283–289, 2005.

[15] Kok Fai Kong, Lisa Schneper, and Kalai Mathee. Beta-lactam antibiotics: From antibiosis to resistance and bacteriology, 2010.

[16] E W Goodell, R Lopez, A Tomasz, Zemer Gitai, and Ned S. Wingreen. Suppression of lytic effect of beta lactams on Escherichia coli and other bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 73(9):3293–7, 1976.

[17] A. M. Egorov, M. M. Ulyashova, and M. Yu Rubtsova. Bacterial enzymes and antibiotic resistance, 2018.

[18] Firdaus Samsudin, Maite L. Ortiz-Suarez, Thomas J. Piggot, Peter J. Bond, and Syma Khalid. OmpA: A Flexible Clamp for Bacterial Cell Wall Attachment. *Structure*, 24(12):2227–2235, 2016.

[19] Sam Ogwang, Hoa T. Nguyen, Marissa Sherman, Saralee Bajaksouzian, Michael R. Jacobs, W. Henry Boom, Guo Fang Zhang, and Liem Nguyen. Bacterial conversion of folinic acid is required for antifolate resistance. *Journal of Biological Chemistry*, 286(17):15377–15390, 2011.

[20] Amanda J. McBroom and Meta J. Kuehn. Release of outer membrane vesicles by Gram-negative bacteria is a novel envelope stress response. *Molecular Microbiology*, 63(2):545–558, 2007.

[21] TJ Beveridge. Use of the Gram stain in microbiology. *Biotechnic & Histochemistry*, 76(3):111–118, 2001.

[22] Michael J. Wilhelm, Joel B. Sheffield, Mohammad Sharifian Gh., Yajing Wu, Christian Spahr, Grazia Gonella, Bolei Xu, and Hai Lung Dai. Gram's Stain Does Not Cross the Bacterial Cytoplasmic Membrane. *ACS Chemical Biology*, 10(7):1711–1717, 2015.

[23] Anna Pavlova, Hyea Hwang, Karl Lundquist, Curtis Balusek, and James C. Gumbart. Living on the edge: Simulations of bacterial outer-membrane proteins. *Biochimica et Biophysica Acta - Biomembranes*, 1858(7):1753–1759, 2016.

[24] David J Sherman, Ran Xie, Rebecca J Taylor, Alexander H George, Suguru Okuda, Peter J Foster, Daniel J Needleman, and Daniel Kahne. Lipopolysaccharide is transported to the cell surface by a membrane-to-membrane protein bridge. *Science (New York, N.Y.)*, 359(6377):798–801, 2018.

[25] Haohao Dong, Quanju Xiang, Yinghong Gu, Zhongshan Wang, Neil G. Paterson, Phillip J. Stansfeld, Chuan He, Yizheng Zhang, Wenjian Wang, and Changjiang Dong. Structural basis for outer membrane lipopolysaccharide insertion. *Nature*, 511(7507), 2014.

[26] Pin Chia Hsu, Firdaus Samsudin, Jonathan Shearer, and Syma Khalid. It Is Complicated: Curvature, Diffusion, and Lipid Sorting within the Two Membranes of Escherichia coli. *Journal of Physical Chemistry Letters*, 8(22):5513–5518, 2017.

[27] Christian R.H. Raetz, C. Michael Reynolds, M. Stephen Trent, and Russell E. Bishop. Lipid A Modification Systems in Gram-Negative Bacteria. *Annual Review of Biochemistry*, 76(1):295–329, 2007.

[28] Hoangdung Ho, Anh Miu, Mary Kate Alexander, Natalie K. Garcia, Angela Oh, Inna Zilberleyb, Mike Reichelt, Cary D. Austin, Christine Tam, Stephanie Shriver, Huiyong Hu, Sharada S. Labadie, Jun Liang, Lan Wang, Jian Wang, Yan Lu, Hans E. Purkey, John Quinn, Yvonne Franke, Kevin Clark, Maureen H. Beresini, Man Wah Tan, Benjamin D. Sellers, Till Maurer, Michael F.T. Koehler, Aaron T. Wecksler, James R. Kiefer, Vishal Verma, Yiming Xu, Mireille Nishiyama, Jian Payandeh, and Christopher M. Koth. Structural basis for dual-mode inhibition of the ABC transporter MsbA. *Nature*, 557(7704):196–201, 2018.

[29] Eva M. Palsson-McDermott and Luke A. J. O'Neill. Signal transduction by the lipopolysaccharide receptor, Toll-like receptor-4. *Immunology*, 113(2):153–162, 2004.

[30] Thomas J. Silhavy, Daniel Kahne, and Suzanne Walker. The bacterial cell envelope., 2010.

[31] D.P. Tieleman and H.J.C. Berendsen. A Molecular Dynamics Study of the Pores Formed by Escherichia coli OmpF Porin in a Fully Hydrated Palmitoyloleoylphosphatidylcholine Bilayer. *Biophysical Journal*, 74(6):2786–2801, 1998.

[32] Jiang Yeow, Kang Wei Tan, Daniel A. Holdbrook, Zhi Soon Chong, Jan K. Marzinek, Peter J. Bond, and Shu Sin Chng. The architecture of the OmpC–MlaA complex sheds light on the maintenance of outer membrane

lipid asymmetry in Escherichia coli. *Journal of Biological Chemistry*, 293(29):11325–11340, 2018.

[33] William C. Wimley. The versatile β-barrel membrane protein. *Current Opinion in Structural Biology*, 13(4):404–411, 2003.

[34] Hon-Yeung Cheung, Matthew Man-Kin Wong, Sau-Ha Cheung, Longman Yimin Liang, Yun-Wah Lam, and Sung-Kay Chiu. Differential Actions of Chlorhexidine on the Cell Wall of Bacillus subtilis and Escherichia coli. *PLoS ONE*, 7(5):e36659, 2012.

[35] Emma J Hayhurst, Lekshmi Kailas, Jamie K Hobbs, and Simon J Foster. Cell wall peptidoglycan architecture in Bacillus subtilis. *Proceedings of the National Academy of Sciences of the United States of America*, 105(38):14603–8, 2008.

[36] Volkmar Braun and Helga Wolff. The Murein-Lipoprotein Linkage in the Cell Wall of Escherichia coli. *European Journal of Biochemistry*, 14(2):387–391, 1970.

[37] Masatori Inouye, Jii˜ Shaw, and Cynthia Shen. The Assembly of a Structural Lipoprotein in the Envelope of IkAerichia COW. *Journal of Biological Chemistry*, 247(24):8154–8159, 1972.

[38] Volkmar Braun and Valerie Bosch. Sequence of the Murein . Lipoprotein and the Attachment Site of the Lipid. *European Journal of Biochemistry*, 28(1):51–69, 1972.

[39] Waldemar Vollmer, Didier Blanot, and Miguel A. De Pedro. Peptidoglycan structure and architecture. *FEMS Microbiology Reviews*, 32(2):149–167, 2008.

[40] Dustin T. King, Gregory A. Wasney, Michael Nosella, Anita Fong, Natalie C.J. Strynadka, and F. Peter Guengerich. Structural insights into inhibition of Escherichia coli penicillin-binding protein 1B. *Journal of Biological Chemistry*, 292(3), 2017.

[41] Alexander J. F. Egan, Roberto Maya-Martinez, Isabel Ayala, Catherine M. Bougault, Manuel Banzhaf, Eefjan Breukink, Waldemar Vollmer, and Jean-Pierre Simorre. Induced conformational changes activate the peptidoglycan synthase PBP1B. *Molecular Microbiology*, 110(3):335–356, 2018.

[42] Dirk Jan Scheffers and Jeffery Errington. PBP1 is a component of the Bacillus subtilis cell division machinery. *Journal of Bacteriology*, 186(15):5153–5156, 2004.

[43] Ahmed Bouhss, Amy E. Trunkfield, Timothy D.H. Bugg, and Dominique Mengin-Lecreulx. The biosynthesis of peptidoglycan lipid-linked intermediates. *FEMS Microbiology Reviews*, 32(2):208–233, 2008.

[44] Gizem Özbaykal, Eva Wollrab, Francois Simon, Antoine Vigouroux, Baptiste Cordier, Andrey Aristov, Thibault Chaze, Mariette Matondo, and Sven van Teeffelen. The transpeptidase PBP2 governs initial localization and activity of the major cell-wall synthesis machinery in E. coli. *eLife*, 9, 2020.

[45] Heng Zhao, Vaidehi Patel, John D. Helmann, and Tobias Dörr. Don't let sleeping dogmas lie: new views of peptidoglycan synthesis and its regulation, 2017.

[46] Meriem El Ghachi, Nicole Howe, Chia Ying Huang, Vincent Olieric, Rangana Warshamanage, Thierry Touzé, Dietmar Weichert, Phillip J. Stansfeld, Meitian Wang, Fred Kerff, and Martin Caffrey. Crystal structure of undecaprenyl-pyrophosphate phosphatase and its role in peptidoglycan biosynthesis. *Nature Communications*, 9(1), 2018.

[47] Robert D. Turner, Jamie K. Hobbs, and Simon J. Foster. Atomic force microscopy analysis of bacterial cell wall peptidoglycan architecture. In *Methods in Molecular Biology*, volume 1440, pages 3–9. Humana Press Inc., 2016.

[48] K. C. Huang, R. Mukhopadhyay, B. Wen, Z. Gitai, and N. S. Wingreen. Cell shape and cell-wall organization in Gram-negative bacteria. *Proceedings of the National Academy of Sciences*, 105(49):19282–19287, 2008.

[49] Antoine Vigouroux, Baptiste Cordier, Andrey Aristov, Laura Alvarez, Gizem Özbaykal, Thibault Chaze, Enno Rainer Oldewurtel, Mariette Matondo, Felipe Cava, David Bikard, and Sven van Teeffelen. Class-A penicillin binding proteins do not contribute to cell shape but repair cellwall defects. *eLife*, 9, 2020.

[50] Elizabeth A. Mueller and Petra Anne Levin. Bacterial cell wall quality control during environmental stress. *mBio*, 11(5):1–15, 2020.

[51] Boumediene Soufi, Karsten Krug, Andreas Harst, and Boris Macek. Characterization of the E. coli proteome and its modifications during growth and ethanol stres. *Frontiers in Microbiology*, 6(FEB), 2015.

[52] Shin ichiro Narita and Hajime Tokuda. Bacterial lipoproteins; biogenesis, sorting and quality control, 2017.

[53] Alister Boags, Firdaus Samsudin, and Syma Khalid. Details of hydrophobic entanglement between small molecules and Braun's lipoprotein within the cavity of the bacterial chaperone LolA. *Scientific Reports*, 9(1):1–8, 2019.

[54] Gareth W. Hughes, Stephen C. L. Hall, Claire S. Laxton, Pooja Sridhar, Amirul H. Mahadi, Caitlin Hatton, Thomas J. Piggot, Peter J. Wotherspoon, Aneika C. Leney, Douglas G. Ward, Mohammed Jamshad, Vaclav Spana, Ian T. Cadby, Christopher Harding, Georgia L. Isom, Jack A. Bryant, Rebecca J. Parr, Yasin Yakub, Mark Jeeves, Damon Huber, Ian R. Henderson, Luke A. Clifton, Andrew L. Lovering, and Timothy J. Knowles. Evidence

for phospholipid export from the bacterial inner membrane by the Mla ABC transport system. *Nature Microbiology*, pages 1–14, 2019.

[55] Henrik Strahl and Jeff Errington. Bacterial Membranes: Structure, Domains, and Function. *Annual Review of Microbiology*, 71(1), 2017.

[56] Jeffrey K. Griffiths, Ramaswamy Balakrishnan, Giovanni Widmer, and Saul Tzipori. Paromomycin and geneticin inhibit intracellular Cryptosporidium parvum without trafficking through the host cell cytoplasm: Implications for drug delivery. *Infection and Immunity*, 66(8):3874–3883, 1998.

[57] Alister T. Boags, Firdaus Samsudin, and Syma Khalid. Binding from Both Sides: TolR and Full-Length OmpA Bind and Maintain the Local Structure of the E. coli Cell Wall. *Structure*, 27(4):713–724.e2, 2019.

[58] P. Germon, M. C. Ray, A. Vianney, and J. C. Lazzaroni. Energy-dependent conformational change in the TolA protein of Escherichia coli involves its N-terminal domain, TolQ, and TolR. *Journal of Bacteriology*, 183(14):4110–4114, 2001.

[59] Justyna A. Wojdyla, Erin Cutts, Renata Kaminska, Grigorios Papadakos, Jonathan T S Hopper, Phillip J. Stansfeld, David Staunton, Carol V. Robinson, and Colin Kleanthous. Structure and function of the Escherichia coli Tol-Pal stator protein TolR. *Journal of Biological Chemistry*, 290(44):26675–26687, 2015.

[60] Walter L. Ash, Marian R. Zlomislic, Eliud O. Oloo, and D. Peter Tieleman. Computer simulations of membrane proteins. *Biochimica et Biophysica Acta - Biomembranes*, 1666(1-2):158–189, 2004.

[61] David Y. Haubertin, Hocine Madaoui, Alain Sanson, Raphaël Guérois, and Stéphane Orlowski. Molecular dynamics simulations of E. coli MsbA transmembrane domain: Formation of a semipore structure. *Biophysical Journal*, 91(7):2517–2531, 2006.

[62] Justin Gullingsrud, Dorina Kosztin, and Klaus Schulten. Structural determinants of MscL gating studied by molecular dynamics simulations. *Biophysical Journal*, 80(5):2074–2081, 2001.

[63] Antreas C. Kalli, Mark S. P. Sansom, and Reinhart A. F. Reithmeier. Molecular Dynamics Simulations of the Bacterial UraA H+-Uracil Symporter in Lipid Bilayers Reveal a Closed State and a Selective Interaction with Cardiolipin. *PLOS Computational Biology*, 11(3):e1004123, 2015.

[64] Thomas J. Piggot, Daniel A. Holdbrook, and Syma Khalid. Electroporation of the E. coli and S. Aureus Membranes: Molecular Dynamics Simulations of Complex Bacterial Membranes. *The Journal of Physical Chemistry B*, 115(45):13381–13388, 2011.

[65] Jan K. Marzinek, Daniel A. Holdbrook, Roland G. Huber, Chandra Verma, and Peter J. Bond. Pushing the Envelope: Dengue Viral Membrane Coaxed into Shape by Molecular Simulations. *Structure*, 24(8):1410–1420, 2016.

[66] E. J.J. Lugtenberg and R. Peters. Distribution of lipids in cytoplasmic and outer membranes of Escherichia coli K12. *Biochimica et Biophysica Acta (BBA)/Lipids and Lipid Metabolism*, 441(1):38–47, 1976.

[67] Karl N. Kirschner, Roberto D. Lins, Astrid Maass, and Thereza A. Soares. A glycam-based force field for simulations of lipopolysaccharide membranes: Parametrization and validation. *Journal of Chemical Theory and Computation*, 8(11):4719–4731, 2012.

[68] Mark P. Molloy, Ben R. Herbert, Martin B. Slade, Thierry Rabilloud, Amanda S. Nouwens, Keith L. Williams, and Andrew A. Gooley. Proteomic analysis of the Escherichia coli outer membrane. *European Journal of Biochemistry*, 267(10):2871–2881, 2000.

[69] Alex Pautsch and Georg E. Schulz. High-resolution structure of the OmpA membrane domain. *Journal of Molecular Biology*, 298(2):273–282, 2000.

[70] Syma Khalid, Peter J. Bond, Timothy Carpenter, and Mark S.P. Sansom. OmpA: Gating and dynamics via molecular dynamics simulations. *Biochimica et Biophysica Acta - Biomembranes*, 1778(9):1871–1880, 2008.

[71] J N Weiser and E C Gotschlich. Outer membrane protein A (OmpA) contributes to serum resistance and pathogenicity of Escherichia coli K-1. *Infection and immunity*, 59(7):2252–8, 1991.

[72] Peter J. Bond, José D. Faraldo-Gómez, and Mark S.P. Sansom. OmpA: A Pore or Not a Pore? Simulation and Modeling Studies. *Biophysical Journal*, 83(2):763–775, 2002.

[73] Peter J. Bond, Daniel L. Parton, James F. Clark, and Mark S.P. Sansom. Coarse-grained simulations of the membrane-active antimicrobial peptide maculatin 1.1. *Biophysical Journal*, 95(8):3802–3815, 2008.

[74] T. Schneider, K. Gries, M. Josten, I. Wiedemann, S. Pelzer, H. Labischinski, and H. G. Sahl. The lipopeptide antibiotic friulimicin B inhibits cell wall biosynthesis through complex formation with bactoprenol phosphate. *Antimicrobial Agents and Chemotherapy*, 53(4):1610–1618, 2009.

[75] Peter Schumann. Peptidoglycan Structure. In *Methods in Microbiology*, volume 38, pages 101–129. Academic Press, 2011.

[76] James C. Gumbart, Morgan Beeby, Grant J. Jensen, and Benoît Roux. Escherichia coli Peptidoglycan Structure and Mechanics as Predicted by Atomic-Scale Simulations, 2014.

[77] Jaewoon Jung, Wataru Nishima, Marcus Daniels, Gavin Bascom, Chigusa Kobayashi, Adetokunbo Adedoyin, Michael Wall, Anna Lappala, Dominic Phillips, William Fischer, Chang-Shung Tung, Tamar Schlick, Yuji Sugita, and Karissa Y. Sanbonmatsu. Scaling molecular dynamics beyond 100,000 processor cores for large-scale biophysical simulations. *Journal of Computational Chemistry*, 40(21):1919–1930, 2019.

[78] Edmund Bertschinger. SIMULATIONS OF STRUCTURE FORMATION IN THE UNIVERSE. *Annual Review of Astronomy and Astrophysics*, 36(1):599–654, 1998.

[79] Kieron Burke and Lucas O. Wagner. DFT in a nutshell, 2013.

[80] Thom Vreven and Keiji Morokuma. Chapter 3 Hybrid Methods: ONIOM(QM:MM) and QM/MM, 2006.

[81] Martin Karplus and J. Andrew McCammon. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, 9(9):646–652, 2002.

[82] Juan R. Perilla and Klaus Schulten. Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations. *Nature Communications*, 8(1):1–10, 2017.

[83] J. Andrew McCammon, Bruce R. Gelin, and Martin Karplus. Dynamics of folded proteins. *Nature*, 267(5612):585–590, 1977.

[84] Ju Li. Basic Molecular Dynamics. In *Handbook of Materials Modeling*, pages 565–588. Springer Netherlands, 2005.

[85] Jing Huang and Alexander D. MacKerell. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of Computational Chemistry*, 34(25):2135–2145, 2013.

[86] B. J. Alder and T. E. Wainwright. Studies in Molecular Dynamics. I. General Method. *The Journal of Chemical Physics*, 31(2):459–466, 1959.

[87] B. J. Alder, D. M. Gass, and T. E. Wainwright. Studies in molecular dynamics. VIII. The transport coefficients for a hard-sphere fluid. *The Journal of Chemical Physics*, 53(10):3813–3826, 1970.

[88] B. J. Alder, W. E. Alley, and J. H. Dymond. Studies in molecular dynamics. XIV. Mass and size dependence of the binary diffusion coefficient. *The Journal of Chemical Physics*, 1974.

[89] Jan Domański, Mark S. P. Sansom, Phillip J. Stansfeld, and Robert B. Best. Atomistic mechanism of transmembrane helix association. *PLOS Computational Biology*, 16(6):e1007919, 2020.

[90] Peter J Bond and Mark S.P Sansom. Membrane Protein Dynamics versus Environment: Simulations of OmpA in a Micelle and in a Bilayer. *Journal of Molecular Biology*, 329(5):1035–1053, 2003.

[91] Camille Goemans, Katleen Denoncin, and Jean-François Collet. Folding mechanisms of periplasmic proteins. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1843(8):1517–1528, 2014.

[92] Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992.

[93] E. Phillip Bourne. Protein Data Bank — Nucleic Acids Research — Oxford Academic. *Nucleic Acids Research*, 2000.

[94] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindah. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, 2015.

[95] Mark Abraham, Berk Hess, David van der Spoel, and Erik Lindahl. GROMACS User Manual version 5.0.7. *Www.Gromacs.Org*, 2015.

[96] David A. Pearlman, David A. Case, James W. Caldwell, Wilson S. Ross, Thomas E. Cheatham, Steve DeBolt, David Ferguson, George Seibel, and Peter Kollman. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications*, 91(1-3):1–41, 1995.

[97] R. Car and M. Parrinello. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Physical Review Letters*, 55(22):2471–2474, 1985.

[98] R. W. Hockney, S. P. Goel, and J. W. Eastwood. Quiet high-resolution computer models of a plasma. *Journal of Computational Physics*, 14(2):148–158, 1974.

[99] W. F. Van Gunsteren and H. J.C. Berendsen. A Leap-Frog Algorithm for Stochastic Dynamics. *Molecular Simulation*, 1(3):173–185, 1988.

[100] Berk Hess, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.

[101] Shuichi Miyamoto and Peter A. Kollman. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*, 13(8):952–962, 1992.

[102] Elmar Krieger and Gert Vriend. New ways to boost molecular dynamics simulations. *Journal of Computational Chemistry*, 36(13):996–1007, 2015.

[103] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general Amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 2004.

[104] Histoire de l'Academie royale des sciences - Google Books.

[105] Szilárd Páll and Berk Hess. A flexible algorithm for calculating pair interactions on SIMD architectures. *Computer Physics Communications*, 184(12):2641–2650, 2013.

[106] T. E. III Cheatham, J. L. Miller, T. Fox, T. A. Darden, and P. A. Kollman. Molecular Dynamics Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins. *Journal of the American Chemical Society*, 117(14):4193–4194, 1995.

[107] J. Jones E. On the determination of molecular fields. —II. From the equation of state of a gas. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 106(738):463–477, 1924.

[108] Alexander D. Mackerell. Empirical force fields for biological macromolecules: Overview and issues, 2004.

[109] Jun-Qiang Sun and Klaus Ruedenberg. Quadratic steepest descent on potential energy surfaces. I. Basic formalism and quantitative assessment. *The Journal of Chemical Physics*, 99(7):5257, 1993.

[110] Svetlana S. Petrova and Alexander D. Solov'Ev. The Origin of the Method of Steepest Descent. *Historia Mathematica*, 24(4):361–375, 1997.

[111] R. Fletcher. Function minimization by conjugate gradients. *The Computer Journal*, 7(2):149–154, 1964.

[112] Kurt Binder, Jürgen Horbach, Walter Kob, Paul Wolfgang, and Varnik Fathollah. Molecular dynamics simulations. *Journal of Physics Condensed Matter*, 2004.

[113] Sheng Liu and Yong Liu. Classical Molecular Dynamics. In *Modeling and Simulation for Microelectronic Packaging Assembly*. John Wiley & Sons (Asia) Pte Ltd, 2011.

[114] M. S. Shell. Advanced molecular dynamics techniques. *Advanced molecular dynamics techniques*, 2009.

[115] Alister J. Page, Tetsushi Isomoto, Jan M. Knaup, Stephan Irle, and Keiji Morokuma. Effects of molecular dynamics thermostats on descriptions of chemical nonequilibrium. *Journal of Chemical Theory and Computation*, 2012.

[116] Simone Melchionna, Giovanni Ciccotti, and Brad Lee Holian. Hoover npt dynamics for systems varying in shape and size. *Molecular Physics*, 1993.

[117] A. S. Lemak and N. K. Balabaev. On The Berendsen Thermostat. *Molecular Simulation*, 13(3):177–187, 1994.

[118] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *Journal of Chemical Physics*, 126(1):014101, 2007.

[119] Denis J. Evans and Brad Lee Holian. The Nose-Hoover thermostat. *The Journal of Chemical Physics*, 1985.

[120] Scott E. Feller, Yuhong Zhang, Richard W. Pastor, and Bernard R. Brooks. Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics*, 1995.

[121] H. J.C. Berendsen, J. P.M. Postma, W. F. Van Gunsteren, A. Dinola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 1984.

[122] Shuichi Nosé and M. L. Klein. Constant pressure molecular dynamics for molecular systems. *Molecular Physics*, 1983.

[123] Jing Huang, Sarah Rauscher, Grzegorz Nawrocki, Ting Ran, Michael Feig, Bert L de Groot, Helmut Grubmüller, Alexander D MacKerell, and Jr. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods*, 14(1):71–73, 2017.

[124] Sarah Lee, Alan Tran, Matthew Allsopp, Joseph B. Lim, Jéroîme Hénin, and Jeffery B. Klauda. CHARMM36 united atom chain model for lipids and surfactants. *Journal of Physical Chemistry B*, 2014.

[125] Nathan Schmid, Andreas P. Eichenberger, Alexandra Choutko, Sereina Riniker, Moritz Winger, Alan E. Mark, and Wilfred F. Van Gunsteren. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European Biophysics Journal*, 2011.

[126] Siewert J. Marrink, H. Jelger Risselada, Serge Yefimov, D. Peter Tieleman, and Alex H. De Vries. The MARTINI force field: Coarse grained model for biomolecular simulations. *Journal of Physical Chemistry B*, 2007.

[127] Luca Monticelli, Senthil K. Kandasamy, Xavier Periole, Ronald G. Larson, D. Peter Tieleman, and Siewert Jan Marrink. The MARTINI coarse-grained force field: Extension to proteins. *Journal of Chemical Theory and Computation*, 4(5):819–834, 2008.

[128] Xavier Periole and Siewert Jan Marrink. The martini coarse-grained force field. *Methods in Molecular Biology*, 2013.

[129] Yi Isaac Yang, Qiang Shao, Jun Zhang, Lijiang Yang, and Yi Qin Gao. Enhanced sampling in molecular dynamics, 2019.

[130] Vojtech Spiwok, Zoran Sucur, and Petr Hosek. Enhanced sampling techniques in biomolecular simulations, 2015.

[131] Johannes Kästner. Umbrella sampling, 2011.

[132] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 1996.

[133] W.L. DeLano. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography*, 2002.

[134] Travis E. Oliphant. Python for scientific computing. *Computing in Science and Engineering*, 9(3):10–20, 2007.

[135] Richard Gowers, Max Linke, Jonathan Barnoud, Tyler Reddy, Manuel Melo, Sean Seyler, Jan Domański, David Dotson, Sébastien Buchoux, Ian Kenney, and Oliver Beckstein. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In *Proceedings of the 15th Python in Science Conference*, 2016.

[136] Naveen Michaud-Agrawal, Elizabeth J. Denning, Thomas B. Woolf, and Oliver Beckstein. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, 2011.

[137] Volkmar Braun and Uta Sieglin. The Covalent Murein-Lipoprotin Structure of the Escherichia coli Cell Wall. The Attachment Site of the Lipoprotein on the Murein. *European Journal of Biochemistry*, 13(2):336–346, 1970.

[138] Abir T. Asmar and Jean François Collet. Lpp, the Braun lipoprotein, turns 50—major achievements and remaining issues. *FEMS Microbiology Letters*, 2018.

[139] V. Braun and K. Rehn. Chemical Characterization, Spatial Distribution and Function of a Lipoprotein (Murein-Lipoprotein) of the E. coli Cell Wall: The Specific Effect of Trypsin on the Membrane Structure. *European Journal of Biochemistry*, 10(3):426–438, 1969.

[140] Charles E. Cowles, Yongfeng Li, Martin F. Semmelhack, Ileana M. Cristea, and Thomas J. Silhavy. The free and bound forms of Lpp occupy distinct subcellular locations in Escherichia coli. *Molecular Microbiology*, 79(5):1168–1181, 2011.

[141] Jessica Humann and Laurel L. Lenz. Bacterial peptidoglycan-degrading enzymes and their impact on host muropeptide detection, 2009.

[142] Felipe Cava, Miguel A. De Pedro, Hubert Lam, Brigid M. Davis, and Matthew K. Waldor. Distinct pathways for modification of the bacterial cell wall by non-canonical D-amino acids. *EMBO Journal*, 30(16):3442–3453, 2011.

[143] Akeisha N. Sanders and Martin S. Pavelka. Phenotypic analysis of Eschericia coli mutants lacking L,D-transpeptidases. *Microbiology (United Kingdom)*, 159(PART 9):1842–1852, 2013.

[144] Matthieu Depuydt, Stephen E. Leonard, Didier Vertommen, Katleen Denoncin, Pierre Morsomme, Khadija Wahni, Joris Messens, Kate S. Carroll, and Jean François Collet. A periplasms reducing system protects single cysteine residues from oxidation. *Science*, 326(5956):1109–1111, 2009.

[145] Sophie Magnet, Samuel Bellais, Lionel Dubost, Martine Fourgeaud, Jean Luc Mainardi, Sébastien Petit-Frère, Arul Marie, Dominique Mengin-Lecreulx, Michel Arthur, and Laurent Gutmann. Identification of the L,D-transpeptidases responsible for attachment of the Braun lipoprotein to Escherichia coli peptidoglycan. *Journal of Bacteriology*, 189(10):3927–3931, 2007.

[146] Toshiharu Yakushi, Terutaka Tajima, Shin Ichi Matsuyama, and Hajime Tokuda. Lethality of the covalent linkage between mislocalized major outer membrane lipoprotein and the peptidoglycan of Escherichia coli. *Journal of Bacteriology*, 179(9):2857–2862, 1997.

[147] Gene Wei Li, David Burkhardt, Carol Gross, and Jonathan S. Weissman. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, 157(3), 2014.

[148] Monica S. Guo, Taylor B. Updegrove, Emily B. Gogol, Svetlana A. Shabalina, Carol A. Gross, and Gisela Storz. MicL, a new σE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes and Development*, 28(14):1620–1634, 2014.

[149] Ye Ni, John Reye, and Rachel R. Chen. Lpp deletion as a permeabilization method. *Biotechnology and Bioengineering*, 97(6):1347–1356, 2007.

[150] Eric Cascales, Alain Bernadac, Marthe Gavioli, Jean Claude Lazzaroni, and Roland Lloubes. Pal lipoprotein of Escherichia coli plays a major role in outer membrane integrity. *Journal of Bacteriology*, 2002.

[151] Robert J. Nichols, Saunak Sen, Yoe Jin Choo, Pedro Beltrao, Matylda Zietek, Rachna Chaba, Sueyoung Lee, Krystyna M. Kazmierczak, Karis J. Lee, Angela Wong, Michael Shales, Susan Lovett, Malcolm E. Winkler, Nevan J. Krogan, Athanasios Typas, and Carol A. Gross. Phenotypic landscape of a bacterial cell. *Cell*, 144(1):143–156, 2011.

[152] Y. Hirota, H. Suzuki, Y. Nishimura, and S. Yasuda. On the process of cellular division in Escherichia coli: a mutant of E. coli lacking a murein lipoprotein. *Proceedings of the National Academy of Sciences of the United States of America*, 74(4):1417–1420, 1977.

[153] Abir T. Asmar, Josie L. Ferreira, Eli J. Cohen, Seung Hyun Cho, Morgan Beeby, Kelly T. Hughes, and Jean François Collet. Communication across the bacterial cell envelope depends on the size of the periplasm. *PLoS Biology*, 15(12), 2017.

[154] Eli J. Cohen, Josie L. Ferreira, Mark S. Ladinsky, Morgan Beeby, and Kelly T. Hughes. Nanoscale-length control of the flagellar driveshaft requires hitting the tethered outer membrane. *Science*, 356(6334):197–200, 2017.

[155] Julien Marcoux, Argyris Politis, Dennis Rinehart, David P. Marshall, Mark I. Wallace, Lukas K. Tamm, and Carol V. Robinson. Mass spectrometry defines the C-terminal dimerization domain and enables modeling of the structure of full-length OmpA. *Structure*, 22(5):781–790, 2014.

[156] Etsuko Sugawara and Hiroshi Nikaido. Pore-forming activity of OmpA protein of Escherichia coli. *Journal of Biological Chemistry*, 267(4):2507–2511, 1992.

[157] E Sugawara and H Nikaido. OmpA protein of Escherichia coli outer membrane occurs in open and closed channel forms. *The Journal of biological chemistry*, 269(27):17981–7, 1994.

[158] Anthony W. Confer and Sahlu Ayalew. The OmpA family of proteins: Roles in bacterial pathogenesis and immunity, 2013.

[159] Jonathan Shearer, Damien Jefferies, and Syma Khalid. Outer Membrane Proteins OmpA, FhuA, OmpF, EstA, BtuB, and OmpX Have Unique Lipopolysaccharide Fingerprints. *Journal of Chemical Theory and Computation*, 15(4), 2019.

[160] Hyea Hwang, Nicolò Paracini, Jerry M. Parks, Jeremy H. Lakey, and James C. Gumbart. Distribution of mechanical stress in the Escherichia coli cell envelope. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1860(12):2566–2575, 2018.

[161] Heedeok Hong, Gabor Szabo, and Lukas K. Tamm. Electrostatic couplings in OmpA ion-channel gating suggest a mechanism for pore opening. *Nature Chemical Biology*, 2(11):627–635, 2006.

[162] Samuel I. Miller and Nina R. Salama. The gram-negative bacterial periplasm: Size matters, 2018.

[163] Shigeru Hayashi and Henry C. Wu. Lipoproteins in bacteria. *Journal of Bioenergetics and Biomembranes*, 1990.

[164] Ye Ni and Rachel R. Chen. Accelerating whole-cell biocatalysis by reducing outer membrane permeability barrier. *Biotechnology and Bioengineering*, 87(6):804–811, 2004.

[165] Wei Shu, Jie Liu, Hong Ji, and Min Lu. Core structure of the outer membrane lipoprotein from Escherichia coli at 1.9 Å resolution. *Journal of Molecular Biology*, 299(4):1101–1112, 2000.

[166] Maite L. Ortiz-Suarez, Firdaus Samsudin, Thomas J. Piggot, Peter J. Bond, and Syma Khalid. Full-Length OmpA: Structure, Function, and Membrane Interactions Predicted by Molecular Dynamics Simulations. *Biophysical Journal*, 111(8), 2016.

[167] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*, pages 331–342. Springer, Dordrecht, 1981.

[168] Firdaus Samsudin, Alister Boags, Thomas J. Piggot, and Syma Khalid. Braun's Lipoprotein Facilitates OmpA Interaction with the Escherichia coli Cell Wall. *Biophysical Journal*, 113(7):1496–1504, 2017.

[169] M. M. Muller, A. Vianney, J. C. Lazzaroni, R. E. Webster, and R. Portalier. Membrane topology of the Escherichia coli TolR protein required for cell envelope integrity, 1993.

[170] Jean Claude Lazzaroni, Pierre Germon, Marie Céline Ray, and Anne Vianney. The Tol proteins of Escherichia coli and their involvement in the uptake of biomolecules and outer membrane stability. *FEMS Microbiology Letters*, 1999.

[171] E. Cascales, M. Gavioli, J. N. Sturgis, and R. Lloubes. Proton motive force drives the interaction of the inner membrane TolA and outer membrane Pal proteins in Escherichia coli. *Molecular Microbiology*, 2000.

[172] Rahmona Derouichei, Hélène Bénédetti, Jean Claude Lazzaroni, Claude Lazdunski, and Roland Lloubès. Protein complex within Escherichia coli inner membrane: TolA N-terminal domain interacts with TolQ and TolR proteins. *Journal of Biological Chemistry*, 270(19):11078–11084, 1995.

[173] Laure Journet, Emmanuelle Bouveret, Alain Rigal, Roland Lloubes, Claude Lazdunski, and Hélène Bénédetti. Import of colicins across the outer membrane of Escherichia coli involves multiple protein interactions in the periplasm. *Molecular Microbiology*, 42(2):331–344, 2001.

[174] Thierry Clavel, Pierre Germon, Anne Vianney, Raymond Portalier, and Jean Claude Lazzaroni. TolB protein of Escherichia coli K-12 interacts with the outer membrane peptidoglycan-associated proteins Pal, Lpp and OmpA. *Molecular Microbiology*, 1998.

[175] Mélissa Petiti, Bastien Serrano, Laura Faure, Roland Lloubes, Tâm Mignot, and Denis Duché. Tol Energy-Driven Localization of Pal and Anchoring to the Peptidoglycan Promote Outer-Membrane Constriction. *Journal of Molecular Biology*, 431(17):3275–3288, 2019.

[176] Anne Walburger, Claude Lazdunski, and Yves Corda. The Tol/Pal system function requires an interaction between the C-terminal domain of TolA and the N-terminal domain of TolB. *Molecular Microbiology*, 2002.

[177] Alain Bernadac, Marthe Gavioli, Jean Claude Lazzaroni, Satish Raina, and Roland Lloubès. Escherichia coli tol-pal mutants form outer membrane vesicles. *Journal of Bacteriology*, 180(18):4872–4878, 1998.

[178] Matthew A. Gerding, Yasuyuki Ogata, Nicole D. Pecora, Hironori Niki, and Piet A.J. De Boer. The trans-envelope Tol-Pal complex is part of the cell division machinery and required for proper outer-membrane invagination during cell constriction in E. coli. *Molecular Microbiology*, 63(4):1008–1025, 2007.

[179] Seiji Kojima and David F. Blair. Solubilization and Purification of the MotA/MotB Complex of Escherichia coli. *Biochemistry*, 2004.

[180] Mark C. Leake, Jennifer H. Chandler, George H. Wadhams, Fan Bai, Richard M. Berry, and Judith P. Armitage. Stoichiometry and turnover in single, functioning membrane protein complexes. *Nature*, 2006.

[181] Emilie L. Goemaere, Anthony Devert, Roland Lloub?s, and Eric Cascales. Movements of the TolR C-terminal domain depend on TolQR ionizable key residues and regulate activity of the Tol complex. *Journal of Biological Chemistry*, 282(24):17749–17757, 2007.

[182] Nils A. Berglund, Thomas J. Piggot, Damien Jefferies, Richard B. Sessions, Peter J. Bond, and Syma Khalid. Interaction of the Antimicrobial Peptide Polymyxin B1 with Both Membranes of E. coli: A Molecular Dynamics Study. *PLOS Computational Biology*, 11(4):e1004180, 2015.

[183] Justyna A Wojdyla, Erin Cutts, Renata Kaminska, Grigorios Papadakos, Jonathan T S Hopper, Phillip J Stansfeld, David Staunton, Carol V Robinson, and Colin Kleanthous. Structure and function of the Escherichia coli Tol-Pal stator protein TolR. *The Journal of biological chemistry*, 290(44):26675–87, 2015.

[184] Gerda In't Veld, Arnold J.M. Driessen, and Wil N. Konings. Bacterial solute transport proteins in their lipid environment. *FEMS Microbiology Reviews*, 12(4):293–314, 1993.

[185] Martijn Zwama, Seiji Yamasaki, Ryosuke Nakashima, Keisuke Sakurai, Kunihiko Nishino, and Akihito Yamaguchi. Multiple entry pathways within the efflux transporter AcrB contribute to multidrug recognition. *Nature Communications*, 9(1), 2018.

[186] Catherine M. Oliphant and Kathryn Eroschenko. Antibiotic Resistance, Part 2: Gram-negative Pathogens. *Journal for Nurse Practitioners*, 11(1):79–86, 2015.

[187] Alvin C.Y. Kuk, Aili Hao, Ziqiang Guan, and Seok Yong Lee. Visualizing conformation transitions of the Lipid II flippase MurJ. *Nature Communications*, 10(1), 2019.

[188] Rahul Shrivastava and Shu Sin Chng. Lipid trafficking across the Gram-negative cell envelope. *Journal of Biological Chemistry*, 294(39):14175–14184, 2019.

[189] Damian C. Ekiert, Gira Bhabha, Georgia L. Isom, Garrett Greenan, Sergey Ovchinnikov, Ian R. Henderson, Jeffery S. Cox, Ronald D. Vale, Charles E. Cowles, Yongfeng Li, Martin F. Semmelhack, Ileana M. Cristea, Thomas J. Silhavy, Thomas Steiner, Jonathan G Swoboda, Jennifer Campbell, Timothy C Meredith, Suzanne Walker, Julien Marcoux, Argyris Politis, Dennis Rinehart, David P. Marshall, Mark I. Wallace, Lukas K. Tamm, Carol V. Robinson, Nienke Buddelmeijer, Julien Marcoux, Argyris Politis, Dennis Rinehart, David P. Marshall, Mark I. Wallace, Lukas K. Tamm, Carol V. Robinson, A Bernadac, M Gavioli, J C Lazzaroni, S Raina, R Lloubès, Hiroaki Ishida, Alicia Garcia-Herrero, Hans J. Vogel, T Schneider, K Gries, M Josten, I Wiedemann, S Pelzer, H Labischinski, H-G Sahl, A Bernadac, M Gavioli, J C Lazzaroni, S Raina, R Lloubès, Anke Seydel, Pierre Gounon, Anthony P. Pugsley, Olaf Schneewind, Dominique Missiakas, Marlena M. Wilson, Harris D. Bernstein, Carmen Schwechheimer, Claretta J. Sullivan, Meta J. Kuehn, B. Brooks, M. Karplus, Georg E. Schulz, William C Wimley, Naoko Yokota, Toshiro Kuroda, Shin Ichi Matsuyama, Hajime Tokuda, Walter L. Ash, Marian R. Zlomislic, Eliud O. Oloo, D. Peter Tieleman, Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, Erik Lindah, Damian C. Ekiert, Gira Bhabha, Georgia L. Isom, Garrett Greenan, Sergey Ovchinnikov, Ian R. Henderson, Jeffery S. Cox, Ronald D. Vale, Benjamin I. Baarda, Ryszard A. Zielke, Adriana Le Van, Ann E. Jerse, Aleksandra E. Sikora, Meriem El Ghachi, Nicole Howe, Chia Ying Huang, Vincent Olieric, Rangana Warshamanage, Thierry Touzé, Dietmar Weichert, Phillip J. Stansfeld, Meitian Wang, Fred Kerff, Martin Caffrey, Teresa Paramo, Alexandra East, Diana Garzón, Martin B. Ulmschneider, Peter J. Bond, Marcin Grabowicz, Bilge Ercan, Wen Yi Low, Xuejun Liu, Shu Sin Chng, Stefan U. Vetterli, Katja Zerbe, Maik Müller, Matthias Urfer, Milon Mondal, Shuang Yan Wang, Kerstin Moehle, Oliver Zerbe, Alessandra Vitale, Gabriella Pessi, Leo Eberl, Bernd Wollscheid, John A. Robinson, Wei Shu, Jie Liu, Hong Ji, Min Lu, Rahul Shrivastava, Shu Sin Chng, Jiang Yeow, Kang Wei Tan, Daniel A. Holdbrook, Zhi Soon Chong, Jan K. Marzinek, Peter J. Bond, Shu Sin Chng, Rahul Shrivastava, Shu Sin Chng, Zhi Soon Chong, Wei Fen Woo, Shu Sin Chng, Anatol Luther, Matthias Urfer, Michael Zahn, Maik Müller, Shuang Yan Wang, Milon Mondal, Alessandra Vitale, Jean Baptiste Hartmann, Timothy Sharpe, Fabio Lo Monte, Harsha Kocherla, Elizabeth Cline, Gabriella Pessi, Parthasarathi Rath, Seyed Majed Modaresi, Petra Chiquet, Sarah Stiegeler, Carolin Verbree, Tobias Remus, Michel Schmitt, Caroline Kolopp, Marie Anne Westwood, Nicolas Desjonquères, Emile Brabet, Sophie Hell, Karen LePoupon, Annie Vermeulen, Régis Jaisson, Virginie Rithié, Grégory Upert, Alexander Lederer, Peter Zbinden, Achim Wach, Kerstin Moehle, Katja Zerbe, Hans H. Locher, Francesca Bernardini, Glenn E. Dale, Leo Eberl, Bernd Wollscheid,

Sebastian Hiller, John A. Robinson, Daniel Obrecht, Jiang Yeow, Kang Wei Tan, Daniel A. Holdbrook, Zhi Soon Chong, Jan K. Marzinek, Peter J. Bond, Shu Sin Chng, Bilge Ercan, Wen Yi Low, Xuejun Liu, and Shu Sin Chng. Characterization of Interactions and Phospholipid Transfer between Substrate Binding Proteins of the OmpC-Mla System. *Journal of Biological Chemistry*, 169(2):114–119, 2019.

[190] Damian C. Ekiert, Gira Bhabha, Georgia L. Isom, Garrett Greenan, Sergey Ovchinnikov, Ian R. Henderson, Jeffery S. Cox, and Ronald D. Vale. Architectures of Lipid Transport Systems for the Bacterial Outer Membrane. *Cell*, 169(2):273–285.e17, 2017.

[191] Andrea Gioffré, Eduardo Infante, Diana Aguilar, María De La Paz Santangelo, Laura Klepp, Ariel Amadio, Virginia Meikle, Ignacio Etchechoury, María Isabel Romano, Angel Cataldi, Rogelio Pando Hernández, and Fabiana Bigi. Mutation in mce operons attenuates Mycobacterium tuberculosis virulence. *Microbes and Infection*, 7(3):325–334, 2005.

[192] Georgia L. Isom, Nathaniel J. Davies, Zhi Soon Chong, Jack A. Bryant, Mohammed Jamshad, Maria Sharif, Adam F. Cunningham, Timothy J. Knowles, Shu Sin Chng, Jeffrey A. Cole, and Ian R. Henderson. MCE domain proteins: Conserved inner membrane lipid-binding proteins required for outer membrane homeostasis. *Scientific Reports*, 2017.

[193] Bilge Ercan, Wen Yi Low, Xuejun Liu, and Shu Sin Chng. Characterization of Interactions and Phospholipid Transfer between Substrate Binding Proteins of the OmpC-Mla System. *Biochemistry*, 58(2):114–119, 2019.

[194] Yu ming M. Huang, Yinglong Miao, Jason Munguia, Leo Lin, Victor Nizet, and J. Andrew McCammon. Molecular dynamic study of MlaC protein in Gram-negative bacteria: conformational flexibility, solvent effect and protein-phospholipid binding. *Protein Science*, 25(8):1430–1437, 2016.

[195] Chuan Liu, Jinying Ma, Jia Wang, Hongwei Wang, and Li Zhang. Cryo-EM Structure of a Bacterial Lipid Transporter YebT. *Journal of Molecular Biology*, 2020.

[196] Gareth W Hughes, Stephen C.L. Hall, Claire S Laxton, Pooja Sridhar, Amirul H Mahadi, Caitlin Hatton, Thomas J Piggot, Mohammed Jamshad, Vaclav Spana, Ian T Cadby, Christopher Harding, Georgia L Isom, Jack A Bryant, Rebecca J Parr, Yasin Yakub, Mark Jeeves, Damon Huber, Ian R Henderson, Luke A Clifton, Andrew L Lovering, and Timothy Knowles. Evidence for phospholipid export from the gram-negative inner membrane: time to rethink the Mla pathway? *bioRxiv*, 2018.

[197] Gareth W Hughes, Pooja Sridhar, Stephanie A Nestorow, Peter J Wotherspoon, Benjamin F Cooper, and Timothy J Knowles. MlaFEDB displays flippase activity to promote phospholipid transport towards the outer membrane of Gram-negative bacteria. *BioRxiv*, 2020.

[198] Javier Abellón-Ruiz, Shreyas S. Kaptan, Arnaud Baslé, Beatrice Claudi, Dirk Bumann, Ulrich Kleinekathöfer, and Bert Van Den Berg. Structural basis for maintenance of bacterial outer membrane lipid asymmetry. *Nature Microbiology*, 2(12):1616–1623, 2017.

[199] Zhi Soon Chong, Wei Fen Woo, and Shu Sin Chng. Osmoporin OmpC forms a complex with MlaA to maintain outer membrane lipid asymmetry in Escherichia coli. *Molecular Microbiology*, 98(6):1133–1146, 2015.

[200] Nicolas Coudray, Georgia L. Isom, Mark R. MacRae, Mariyah N. Saiduddin, Gira Bhabha, and Damian C. Ekiert. Structure of MlaFEDB lipid transporter reveals an ABC exporter fold and two bound phospholipids. *bioRxiv*, 2020.

[201] Xiaodi Tang, Shenghai Chang, Wen Qiao, Qinghua Luo, Yuejia Chen, Zhiying Jia, James Coleman, Ke Zhang, Ting Wang, Zhibo Zhang, Changbin Zhang, Xiaofeng Zhu, Xiawei Wei, Changjiang Dong, Xing Zhang, and Haohao Dong. Structural insight into outer membrane asymmetry maintenance of Gram-negative bacteria by the phospholipid transporter MlaFEDB. *BioRxiv*, 2020.

[202] Carsten Kutzner, Szilárd Páll, Martin Fechner, Ansgar Esztermann, Bert L. de Groot, and Helmut Grubmüller. More bang for your buck: Improved use of GPU nodes for GROMACS 2018. *Journal of Computational Chemistry*, 40(27):2418–2431, 2019.

[203] Sunhwan Jo, Taehoon Kim, Vidyashankara G. Iyer, and Wonpil Im. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, 2008.

[204] Narayanan Eswar, Ben Webb, Marc A. Marti-Renom, M.S. Madhusudhan, David Eramian, Min-yi Shen, Ursula Pieper, and Andrej Sali. Comparative Protein Structure Modeling Using Modeller. *Current Protocols in Bioinformatics*, 15(1):5.6.1–5.6.30, 2006.

[205] Dima Kozakov, David R. Hall, Bing Xia, Kathryn A. Porter, Dzmitry Padhorny, Christine Yueh, Dmitri Beglov, and Sandor Vajda. The ClusPro web server for protein-protein docking. *Nature Protocols*, 2017.

[206] Kazuki Takeda, Hideyuki Miyatake, Naoko Yokota, Shin Ichi Matsuyama, Hajime Tokuda, and Kunio Miki. Crystal structures of bacterial lipoprotein localization factors, LolA and LolB. *EMBO Journal*, 22(13):3199–3209, 2003.

[207] Toshiharu Yakushi, Kazuhiro Masuda, Shin-ichiro Narita, Shin-ichi Matsuyama, and Hajime Tokuda. A new ABC transporter mediating the detachment of lipid-modified proteins from membranes. *Nature Cell Biology*, 2(4):212–218, 2000.

[208] Priyadarshini Murahari, Sharmila Anishetty, and Gautam Pennathur. Understanding the lid movements of LolA in Escherichia coli using molecular dynamics simulation and in silico point mutation. *Computational Biology and Chemistry*, 47:71–80, 2013.

[209] Yuki Oguchi, Kazuki Takeda, Shoji Watanabe, Naoko Yokota, Kunio Miki, and Hajime Tokuda. Opening and closing of the hydrophobic cavity of LolA coupled to lipoprotein binding and release. *The Journal of biological chemistry*, 283(37):25414–20, 2008.

[210] Suguru Okuda and Hajime Tokuda. Model of mouth-to-mouth transfer of bacterial lipoproteins through inner membrane LolC, periplasmic LolA, and outer membrane LolB. *Proceedings of the National Academy of Sciences of the United States of America*, 106(14):5877–5882, 2009.

[211] Jun Tsukahara, Shin ichiro Narita, and Hajime Tokuda. Real time analysis of lipoprotein transfer from LolA to LolB by means of surface plasmon resonance. *FEBS Letters*, 583(18):2987–2990, 2009.

[212] A Kovacs-Simon, R W Titball, and S L Michell. Lipoproteins of bacterial pathogens. *Infection and immunity*, 79(2):548–61, 2011.

[213] Shin-ichiro Narita and Hajime Tokuda. Sorting of Bacterial Lipoproteins to the Outer Membrane by the Lol System. In *Protein Secretion*, pages 117–129. Humana Press, 2010.

[214] Anna Konovalova and Thomas J. Silhavy. Outer membrane lipoprotein biogenesis: Lol is not the end. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1679), 2015.

[215] Claudio Muheim, Hansjörg Götzke, Anna U. Eriksson, Stina Lindberg, Ida Lauritsen, Morten H. H. Nørholm, and Daniel O. Daley. Increasing the permeability of Escherichia coli using MAC13243. *Scientific Reports*, 7(1):17629, 2017.

[216] Courtney A. Barker, Sarah E. Allison, Soumaya Zlitni, Nick Duc Nguyen, Rahul Das, Giuseppe Melacini, Alfredo A. Capretta, and Eric D. Brown. Degradation of MAC13243 and studies of the interaction of resulting thiourea compounds with the lipoprotein targeting chaperone LolA. *Bioorganic & Medicinal Chemistry Letters*, 23(8):2426–2431, 2013.

[217] Eric Rassart, Arda Bedirian, Sonia Do Carmo, Olivier Guinard, Jacinthe Sirois, Laurence Terrisse, and Ross Milne. Apolipoprotein D, 2000.

[218] Mario Ruiz, Diego Sanchez, Colin Correnti, Roland K. Strong, and Maria D. Ganfornina. Lipid-binding properties of human ApoD and Lazarillo-related lipocalins: Functional implications for cell differentiation. *FEBS Journal*, 2013.

[219] Maria D. Ganfornina, Sonia Do Carmo, Jose M. Lora, Sonia Torres-Schumann, Marci Vogel, Maria Allhorn, Constancio González, Michael J. Bastiani, Eric Rassart, and Diego Sanchez. Apolipoprotein D is involved in the mechanisms regulating protection from oxidative stress. *Aging Cell*, 2008.

[220] Aaron J Oakley, Surabhi Bhatia, Heath Ecroyd, and Brett Garner. Molecular dynamics analysis of apolipoprotein-D-lipid hydroperoxide interactions: mechanism for selective oxidation of Met-93. *PloS one*, 7(3):e34057, 2012.

[221] Francisco Blanco-Vaca and Henry J. Pownall. Disulfide linked dimers of apolipoprotein D in urine. *ELECTROPHORESIS*, 14(1):1086–1087, 1993.

[222] C. Abergel, A. Walburger, S. Chenivesse, and C. Lazdunski. Crystallization and preliminary crystallographic study of the peptidoglycan-associated lipoprotein from Escherichia coli. *Acta Crystallographica Section D: Biological Crystallography*, 57(2):317–319, 2001.

[223] Sarah Dassati, Andreas Waldner, and Rüdiger Schweigreiter. Apolipoprotein D takes center stage in the stress response of the aging and degenerative brain, 2014.

[224] Martin Vogt and Arne Skerra. Bacterially produced apolipoprotein D binds progesterone and arachidonic acid, but not bilirubin orE-3M2H. *Journal of Molecular Recognition*, 14(1):79–86, 2001.

[225] João H. Morais Cabral, Gordon L. Atkins, Luis M. Sánchez, Yolanda S. López-Boado, Carlos López-Otin, and Lindsay Sawyer. Arachidonic acid binds to apolipoprotein D: implications for the protein's function. *FEBS Letters*, 1995.

[226] Alpeshkumar K. Malde, Le Zuo, Matthew Breeze, Martin Stroet, David Poger, Pramod C. Nair, Chris Oostenbrink, and Alan E. Mark. An Automated force field Topology Builder (ATB) and repository: Version 1.0. *Journal of Chemical Theory and Computation*, 2011.

[227] Erica Valentini, Alexey G. Kikhney, Gianpietro Previtali, Cy M. Jeffries, and Dmitri I. Svergun. SASBDB, a repository for biological small-angle scattering data. *Nucleic Acids Research*, 2015.

[228] Claudia S. Kielkopf, Jason K.K. Low, Yee Foong Mok, Surabhi Bhatia, Tony Palasovski, Aaron J. Oakley, Andrew E. Whitten, Brett Garner, and Simon H.J. Brown. Identification of a novel tetrameric structure for human apolipoprotein-D. *Journal of Structural Biology*, 203(3):205–218, 2018.

[229] Markus Aebi. N-linked protein glycosylation in the ER, 2013.

[230] K. Vanommeslaeghe and A. D. MacKerell. Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing. *Journal of Chemical Information and Modeling*, 2012.

Appendix

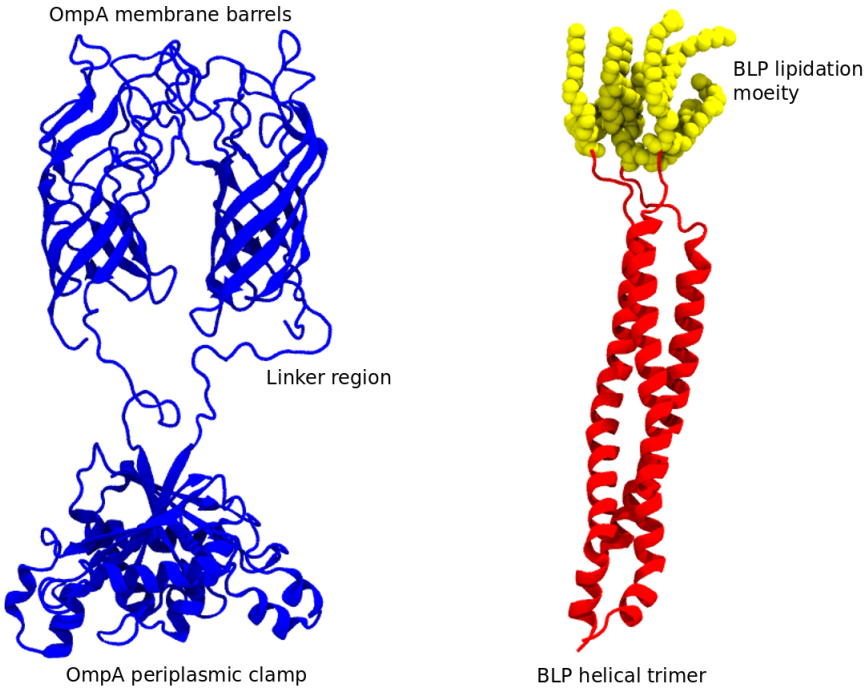


Figure 7.1: A visualisation of the OmpA dimer, OmpA shown in blue (left). A visualisation of the BLP trimer, BLP shown in red, where the lipid component is shown as yellow spheres (right).

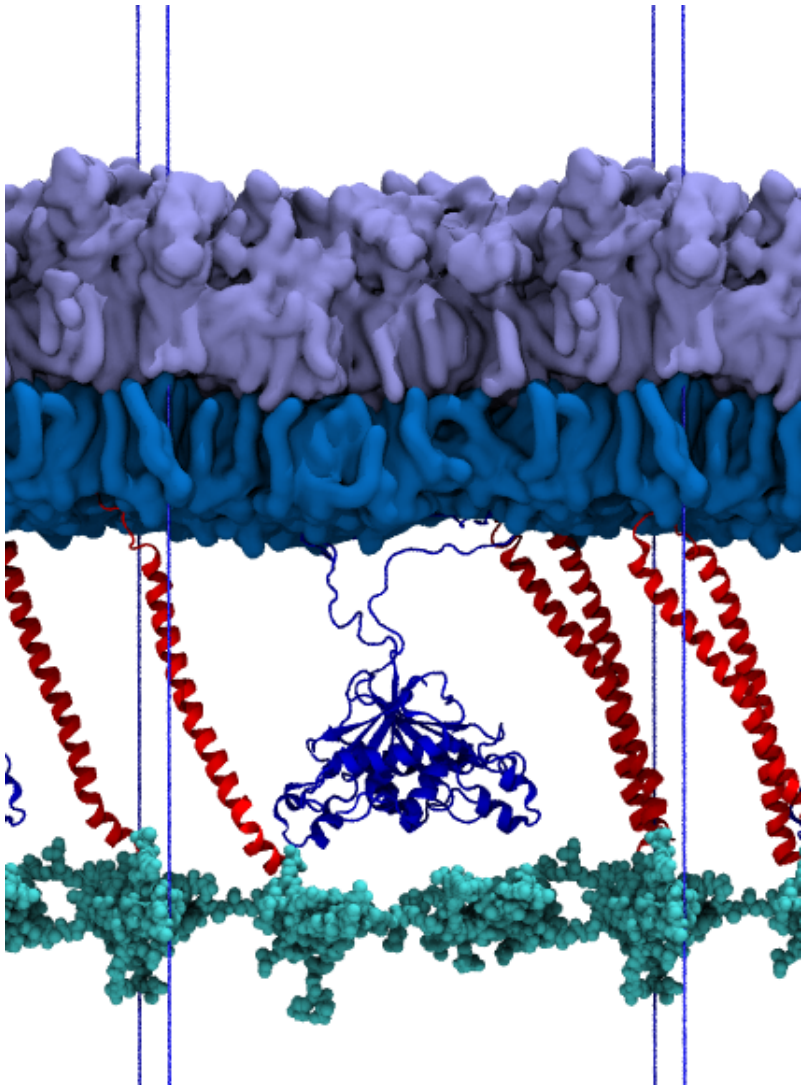


Figure 7.2: A visualisation of the 2 BLP OmpA dimer system. OM (purple and dark blue), BLP (red), OmpA dimer (blue) and PGN (cyan) where the system box is indicated by vertical blue lines.

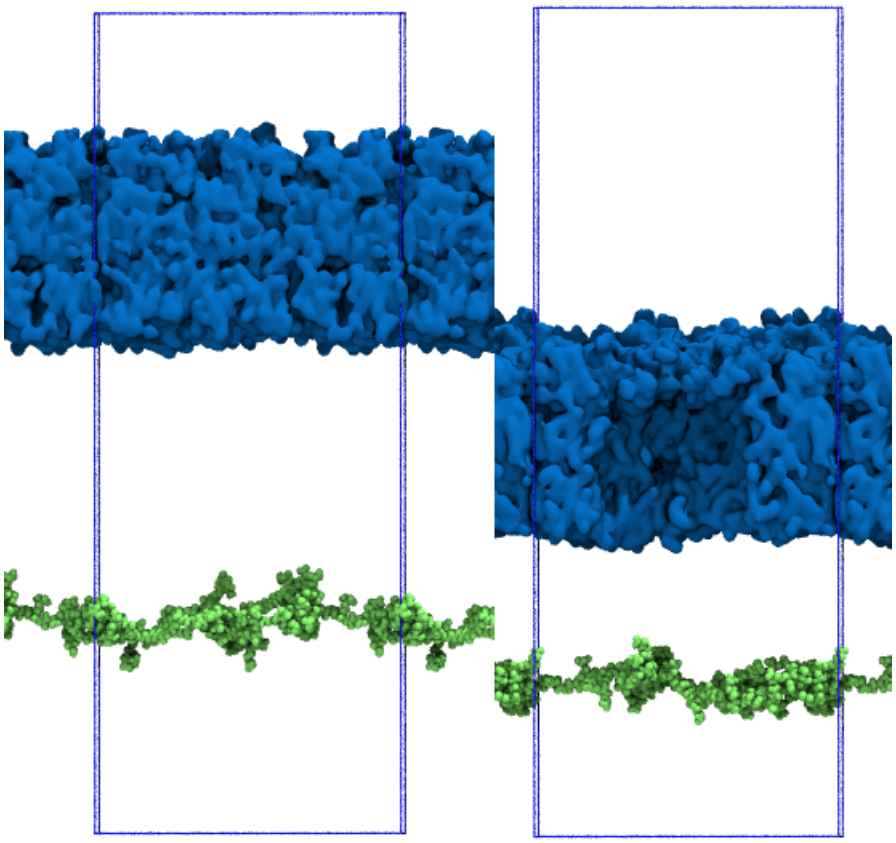


Figure 7.3: A visualisation of the change in periplasmic size in the OmpA dimer and BLP system. OM (blue) and PGN (green) shown alone for clarity, where the system box as been indicated.

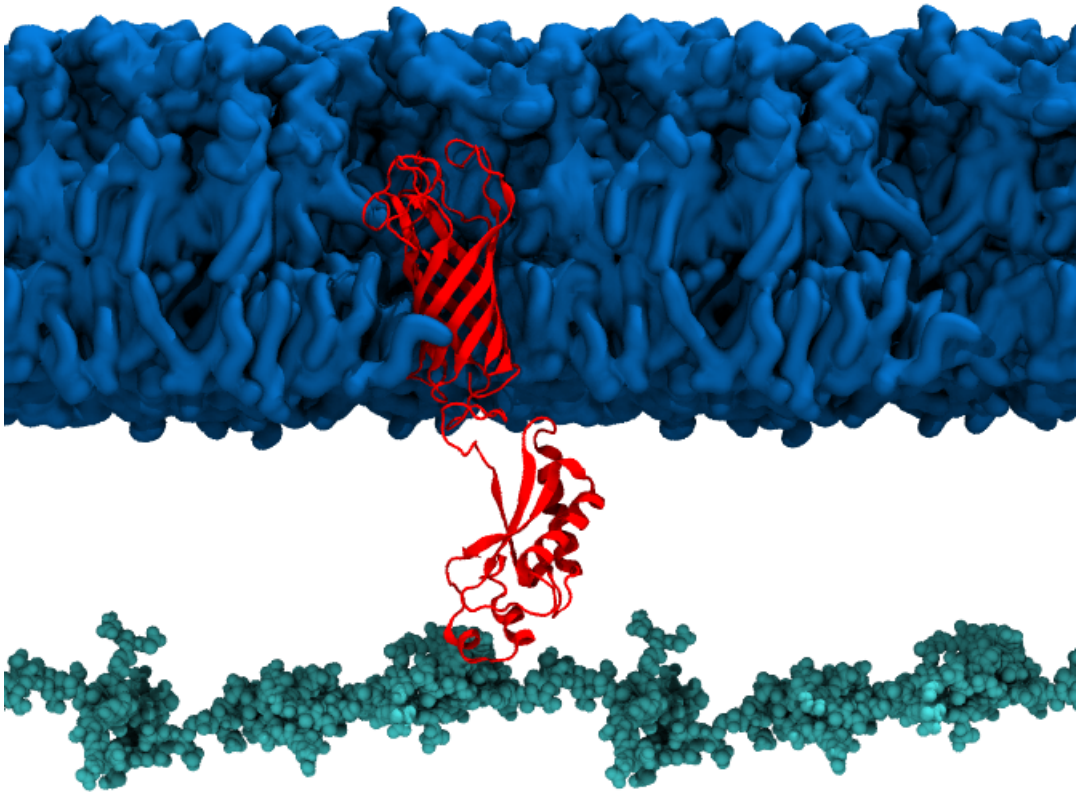


Figure 7.4: The OmpA monomer linker region contracts after 100 ns, OmpA (red), PGN (cyan) and OM (blue).

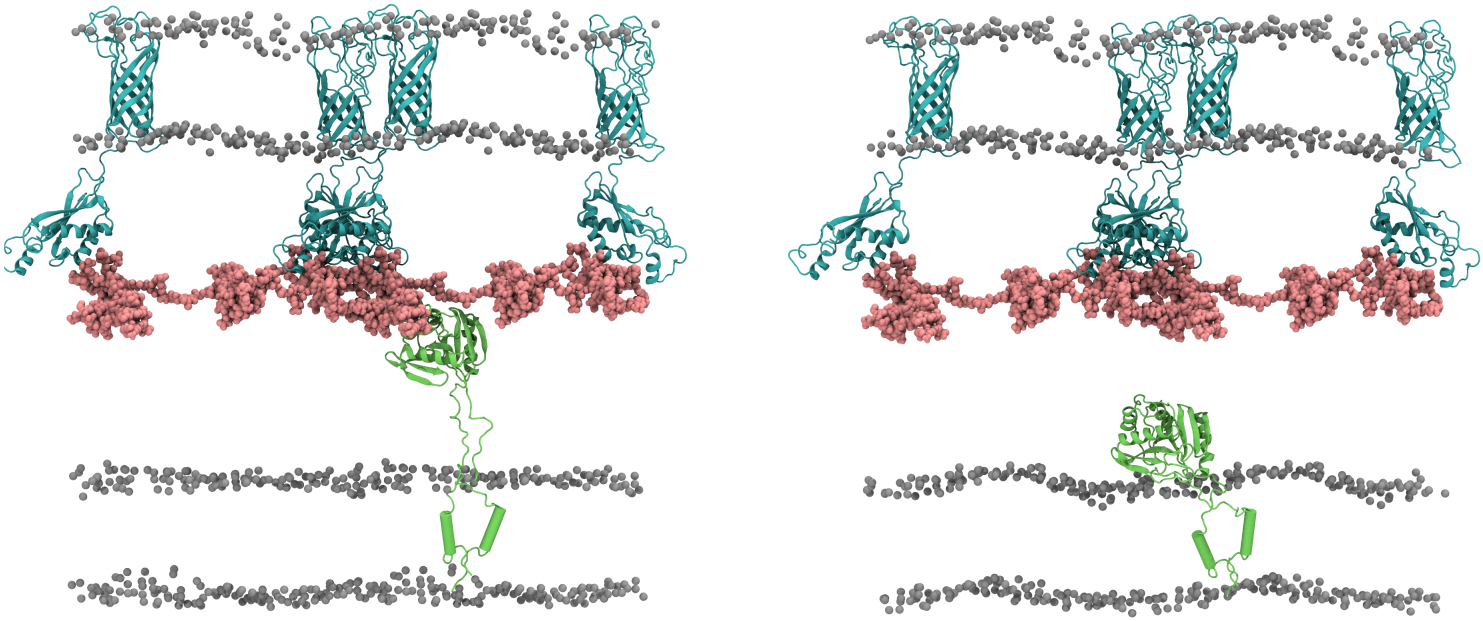


Figure 7.5: TolR being pulled from the cell wall, where initial and final positions are shown (left and right) respectively. TolR (lime green), PGN (pink), OmpA (cyan) and membrane headgroups (grey) are shown.

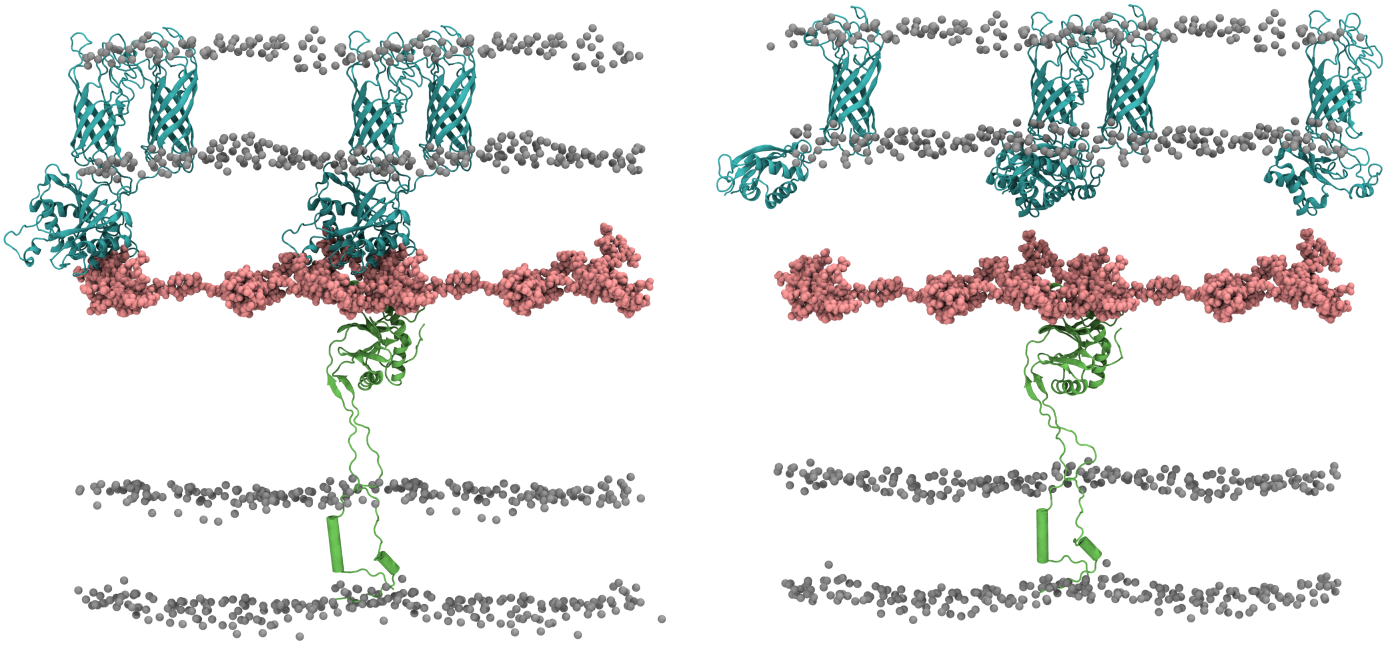


Figure 7.6: OmpA being pulled from the cell wall, where initial and final positions are shown (left and right) respectively. TolR (lime green), PGN (pink), OmpA (cyan) and membrane headgroups (grey) are shown.

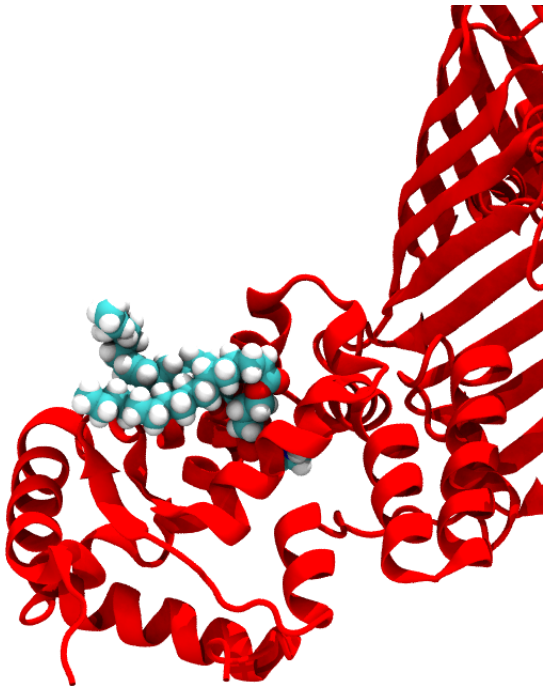


Figure 7.7: A POPE lipid contacts the MlaA beta gate and ”slides” down the water channel in the centre of the protein. MlaA (red) and POPE (cyan and white) are shown.

Mutation of the MlaA residues was done, where the mutation occured in the beta sheet gating region with high contact during lipid abstraction, shown preivously in Figure 5.24. This was done to assess if the interaction becomes less favourable as a result of mutating a small number of residues to alanine in the region.

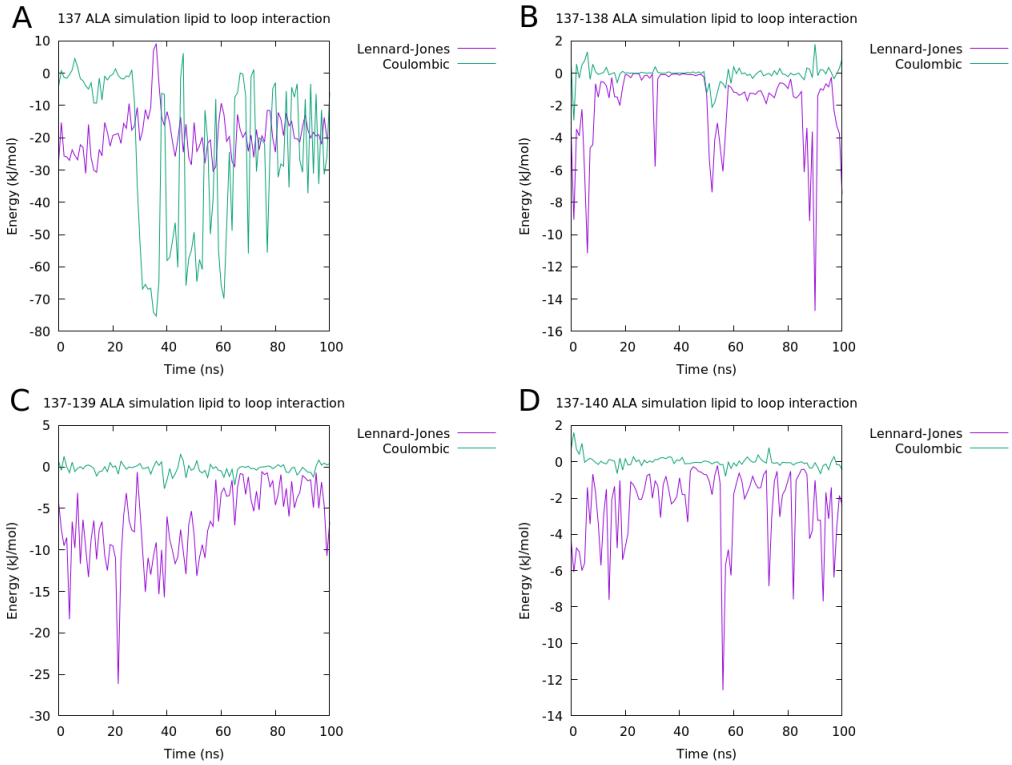


Figure 7.8: The short range electrostatic and Lennard-Jones interactions between the gating loop in MlaA and the abstracted lipid, where the loop is mutated starting at residue 137 to 137-140 all to alanine, proceeding mutations (A-D) through this range are shown.

The loop region 137-140 consists of the following residues: 137PHE, 138TYR, 139GLY and 140SER. Considering the stable interaction with these residues, as previously indicated in Figure 5.25, mutation to alanine disturbed this stabilising energy. Progressive mutation caused the charged energy to essentially drop to 0 or become repulsive, where it was generally attractive in the first mutant. This was also observed in the Lennard-Jones interactions where comparing first to last mutation, a tenfold reduction was seen in Figure 5.26.

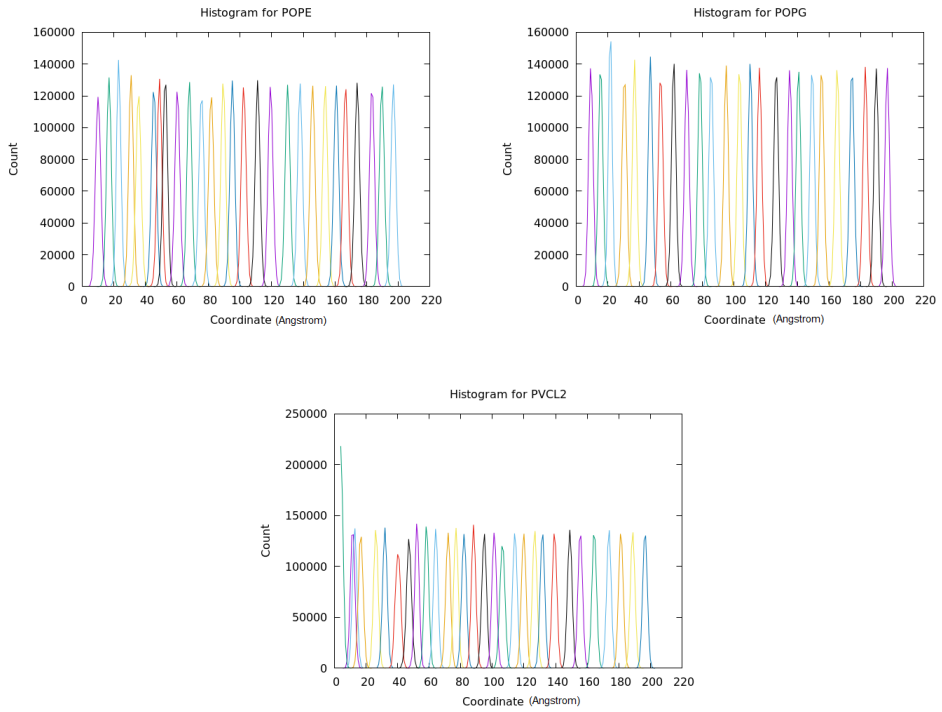


Figure 7.9: The histogram overlap for the umbrella sampling of POPE (top left), POPG (top right) and PVCL2 (bottom) when pulled from the MlaC protein cavity

7.1 Appendix A

```
from __future__ import print_function
from collections import OrderedDict
import fileinput
import sys
import gromacs
import random
import os
import subprocess
sys.dont_write_bytecode = True

os.environ["GMX_MAXBACKUP"]="-1"

#GLUE CODE BEGINS
section = None
atom_count = 0
def Glue(inpt):
        def moltype(fin, fout, sec_name):
                i = 0
                j = 0
                k = 0
                for line in fin:
                        if line.startswith("["):
                                if line.startswith(sec_name):
                                        i = i + 1
                                        if line.startswith(sec_name) and i == 1:
                                                fout.write(line)
                                section = line
                        elif line.startswith("; N") and section.startswith(sec_name)
:
                                        j = j + 1
                                        if j < 2:
                                                fout.write(line)
                        elif line.startswith("PGN"):
                                k = k + 1
                                if k == 1:
                                        fout.write(line)
                        elif section.startswith(sec_name) and not line.startswith("P
GN"):
                                fout.write(line)
                fin.close()
        def atoms(fin, fout, sec_name):
                i = 0

                j = 0
                for line in fin:
                        if line.startswith("["):
                                if line.startswith(sec_name):
                                        i = i + 1
                                        if line.startswith(sec_name) and i == 1:
                                                fout.write(line)
                                section = line
                        elif line.startswith(";") and section.startswith(sec_name):
                                if not line.startswith(";        nr"):
                                        fout.write(line)
                        elif line == "":
                                fout.write(line)
                        elif section.startswith(sec_name):
                                fout.write(line)
                fin.close()
        def dihedrals(fin, fout, sec_name, h_init):
                """New function necessary for Propers and Impropers sections to rema
in distinct"""
                h = h_init

                t = 0
                for line in fin:
                        if line.startswith("["):
                                if line.startswith(sec_name):
                                        h = h + 1
                                        if line.startswith(sec_name) and h % 2 and h
 < 3 + h_init:
                                                fout.write(line)
                                section = line
```

```
                        elif line.startswith(";") and section.startswith(sec_name) a
nd h % 2:
                                t = t + 1
                                if line.startswith(";") and t <= 1 or t >= 3:
                                        fout.write(line)
                        elif line == "":
                                fout.write(line)
                        elif section.startswith(sec_name) and h % 2:
                                fout.write(line)
                fin.close()

        with open('topologies/glued.itp', 'wa') as fout:
                fin = fileinput.input(inpt)
                moltype(fin, fout, "[ moleculetype ]")
                fin = fileinput.input(inpt)
                atoms(fin, fout, "[ atoms ]")
                fin = fileinput.input(inpt)
                atoms(fin, fout, "[ bonds ]")
                fin = fileinput.input(inpt)
                atoms(fin, fout, "[ pairs ]")
                fin = fileinput.input(inpt)
                atoms(fin, fout, "[ angles ]")
                fin = fileinput.input(inpt)
                dihedrals(fin, fout, "[ dihedrals ]", 0)
                fin = fileinput.input(inpt)
                dihedrals(fin, fout, "[ dihedrals ]", 1)
#GLUE CODE ENDS
#Renumber code begins


# The following script will renumber a concatonated .itp file, where the sections ha
ve been joined together appropriately, this is intended for use in building a single
 large .itp file, primarily for a sheet of peptidoglycan due to the bonded existing
between each strand

        #Warning this program only works if your .itp file is in the following order
 :
        #               [ moleculetype ]"
        #               [ atoms ]
        #               [ bonds ]
        #               [ pairs ] (optional)
        #               [ angles ]
        #               [ dihedrals (proper)]
        #               [ dihedrals (improper)]
def renumber(filename):
        lst = []
        def start(filename, lst):
                print("Sorting the start of your file.")
                section = None
                with open(filename, "r") as file:
                        for line in file:
                                line = line.strip()
                                if line.startswith("["):
                                        section = line
                                if section == "[ moleculetype ]":
                                        lst.append(line)
                                if section == None:
                                        lst.append(line)

                return lst


        def atoms(filename, lst):
                section = None
                line_counter = 0
                repeats = 0
                mapping = OrderedDict()
                with open(filename, "r") as file:
                        for line in file:
                                line = line.strip()
                                if line.startswith("["):
                                        if line == "[ atoms ]":
                                                lst.append(line)
```

```
                                                print("Found your atoms, renumbering
 them for you now.")
                                        section = line
                                elif line.startswith(";") and section == "[ atoms ]"
:
                                        lst.append(line)
                                elif line == "":
                                        pass
                                        #lst.append(line)
                                elif line.startswith("PGN"):
                                        pass
                                        #lst.append(line)
                                else:
                                        if section == "[ atoms ]":
                                                line_counter = line_counter + 1
                                                columns = line.split()
                                                if int(columns[0]) == 1:
                                                        repeats = repeats + 1
                                                mapping[columns[0] + "-" +str(repeat
s)] = line_counter
                                                columns[0] = line_counter

                                                new_line = "{0:6d} {1:10s} {2:6s} {3
:6s} {4:6s} {5:6s} {6:10s} {7:10s}".format(columns[0], columns[1], columns[2], colum
ns[3], columns[4], columns[5], columns[6], columns[7])
                                                lst.append(new_line)
                return mapping, lst


        def bonds(filename, mapped, lst):
                section = None
                repeats = 0
                previous_section = None
                saved_line = None
                with open(filename, "r") as file:
                        for line in file:
                                line = line.strip()
                                if line.startswith("["):
                                        if line == "[ bonds ]":
                                                lst.append(line)
                                                print("Found your bonds, renumbering
 them for you now.")
                                        section = line
                                elif line.startswith(";") and section == "[ bonds ]"
:
                                        lst.append(line)
                                elif line == "":
                                        pass
                                        #lst.append(line)
                                elif line.startswith("PGN"):
                                        #lst.append(line)
                                        previous_section = section
                                else:
                                        if section == "[ bonds ]":
                                                if previous_section != section:
                                                        saved_line = line
                                                if saved_line == line:
                                                        repeats = repeats + 1
                                                columns = line.split()
                                                columns[0] = mapped[columns[0] + "-"
 + str(repeats)]
                                                columns[1] = mapped[columns[1] + "-"
 + str(repeats)]
                                                new_line = "{0:5d} {1:5d} {2:5s}".fo
rmat(columns[0], columns[1], columns[2])
                                                lst.append(new_line)
                                        previous_section = section
                return lst


        def pairs(filename, mapped, lst):
                section = None
                repeats = 0
```

```
                previous_section = None
                saved_line = None
                with open(filename, "r") as file:
                        for line in file:
                                line = line.strip()
                                if line.startswith("["):
                                        if line == "[ pairs ]":
                                                lst.append(line)
                                                print("Found your pairs, renumbering
 them for you now.")
                                        section = line
                                elif line.startswith(";") and section == "[ pairs ]"
:
                                        lst.append(line)
                                elif line == "":
                                        pass
                                        #lst.append(line)
                                elif line.startswith("PGN"):
                                        #lst.append(line)
                                        previous_section = section
                                else:
                                        if section == "[ pairs ]":
                                                if previous_section != section:
                                                        saved_line = line
                                                if saved_line == line:
                                                        repeats = repeats + 1
                                                columns = line.split()
                                                columns[0] = mapped[columns[0] + "-"
 + str(repeats)]
                                                columns[1] = mapped[columns[1] + "-"
 + str(repeats)]
                                                new_line = "{0:5d} {1:5d} {2:5s}".fo
rmat(columns[0], columns[1], columns[2])
                                                lst.append(new_line)
                                        previous_section = section
                return lst

        def angles(filename, mapped, lst):
                section = None
                repeats = 0
                previous_section = None
                saved_line = None
                with open(filename, "r") as file:
                        for line in file:
                                line = line.strip()
                                if line.startswith("["):
                                        if line == "[ angles ]":
                                                lst.append(line)
                                                print("Found your angles, renumberin
g them for you now.")
                                        section = line
                                elif line.startswith(";") and section == "[ angles ]
":
                                        lst.append(line)
                                elif line == "":
                                        pass
                                        #lst.append(line)
                                elif line.startswith("PGN"):
                                        #lst.append(line)
                                        previous_section = section
                                else:
                                        if section == "[ angles ]":
                                                if previous_section != section:
                                                        saved_line = line
                                                if saved_line == line:
                                                        repeats = repeats + 1
                                                columns = line.split()
                                                columns[0] = mapped[columns[0] + "-"
 + str(repeats)]
                                                columns[1] = mapped[columns[1] + "-"
 + str(repeats)]
                                                columns[2] = mapped[columns[2] + "-"
 + str(repeats)]
```

```
                                                new_line = "{0:5d} {1:5d} {2:5d} {3:
5s}".format(columns[0], columns[1], columns[2], columns[3])
                                                lst.append(new_line)
                                        previous_section = section
                return lst

        def dihedrals(filename, mapped, lst):
                gi = 0
                i = 0
                section = None
                repeats = 0
                previous_section = None
                saved_line = None
                with open(filename, "r") as file:
                        for line in file:
                                line = line.strip()
                                if line.startswith("["):
                                        if line == "[ dihedrals ]":
                                                lst.append(line)
                                                print("Found your dihedrals, renumbe
ring them for you now.")
                                        section = line
                                elif line.startswith(";") and section == "[ dihedral
s ]":
                                        lst.append(line)
                                elif line == "":
                                        pass
                                        #lst.append(line)
                                elif line.startswith("PGN"):
                                        #lst.append(line)
                                        previous_section = section
                                else:
                                        if section == "[ dihedrals ]":
                                                columns = line.split()
                                                if previous_section != section:
                                                        saved_line = line
                                                if columns[4] == "2" and gi == 0:
                                                        repeats = 0
                                                        saved_line = line
                                                        gi = gi + 1
                                                if saved_line == line:
                                                        repeats = repeats + 1
                                                columns[0] = mapped[columns[0] + "-"
 + str(repeats)]
                                                columns[1] = mapped[columns[1] + "-"
 + str(repeats)]
                                                columns[2] = mapped[columns[2] + "-"
 + str(repeats)]
                                                columns[3] = mapped[columns[3] + "-"
 + str(repeats)]
                                                new_line = "{0:5d} {1:5d} {2:5d} {3:
5d} {4:5s}".format(columns[0], columns[1], columns[2], columns[3], columns[4])
                                                lst.append(new_line)
                                        previous_section = section
                with open("topologies/bonded_clean.itp", "w") as new_file:
                        for line in lst:
                                print(line, file=new_file)
        print("Renumbering Finished.")

        def main(filename):
                list1 = start(filename, [])
                x, y = atoms(filename, list1)
                c = bonds(filename, x, y)
                f = pairs(filename, x, c)
                i = angles(filename, x, f)
                dihedrals(filename, x, i)
        main(filename)


#Renumber code ends
#BONDER CODE BEGINS

# The following script contains the parameters for establising peptide bonds and pep
```

```
tidoglycan sugar linkages, prior to using this script ensure that you have deleted t
he correct atoms from the molecules that you are bonding together, if you're reading
 this there's a good chance you're here because your bond parameters are not include
d, a later comment shows how to add your own bonding parameters for a specific syste
m
def Bonder(x,y):
        lst = []
        def start(filename, lst):
                section = None
                with open(filename, "r") as file:
                        for line in file:
                                line = line.strip()
                                if line.startswith("["):
                                        section = line
                                        if line == "[ moleculetype ]":
                                                lst.append(line)
                                elif line.startswith(";") and section == "[ molecule
type ]":
                                        lst.append(line)
                                elif line.startswith("PGN"):
                                        lst.append(line)
                return lst
        list1=start("topologies/bonded_clean.itp", [])
        def atoms(filename, lst):
                section = None
                with open(filename, "r") as file:
                        for line in file:
                                line = line.strip()
                                if line.startswith("["):
                                        section = line
                                        if line == "[ atoms ]":
                                                lst.append(line)
                                elif line.startswith(";") and section == "[ atoms ]"
:
                                        lst.append(line)
                                elif line == "":
                                        pass
                                elif line.startswith("PGN"):
                                        pass
                                elif section== "[ atoms ]":
                                        columns = line.split()
                                        new_line = "{0:6s} {1:10s} {2:6s} {3:6s} {4:
6s} {5:6s} {6:10s} {7:10s}".format(columns[0], columns[1], columns[2], columns[3], c
olumns[4], columns[5], columns[6], columns[7])
                                        lst.append(new_line)
                return lst

        a = atoms("topologies/bonded_clean.itp", list1)

        def bonds_sec(filename, lst):
                section = None
                with open(filename, "r") as file:
                        for line in file:
                                line = line.strip()
                                if line.startswith("["):
                                        if line == "[ bonds ]":
                                                lst.append(line)
                                        section = line
                                elif line.startswith(";") and section == "[ bonds ]"
:
                                        lst.append(line)
                                elif line == "":
                                        pass
                                        #lst.append(line)
                                elif line.startswith("PGN"):
                                        #lst.append(line)
                                        previous_section = section
                                else:
                                        if section == "[ bonds ]":
                                                columns = line.split()
                                                new_line = "{0:5s} {1:5s} {2:5s}".fo
rmat(columns[0], columns[1], columns[2])
                                                lst.append(new_line)
```

```
                return lst

        c = bonds_sec("topologies/bonded_clean.itp", a)

        def bonding(lst, x, y):
                # y = carbon atom, x = oxygen atom
                bond = (str(x), str(y), "1")
                new_bond = "{0:5s} {1:5s} {2:5s}".format(bond[0], bond[1], bond[2],)
                lst.append(new_bond)
                return lst

        d = bonding(c, x, y)


        def pairs(filename, lst):
                section = None
                with open(filename, "r") as file:
                        for line in file:
                                line = line.strip()
                                if line.startswith("["):
                                        if line == "[ pairs ]":
                                                lst.append(line)
                                        section = line
                                elif line.startswith(";") and section == "[ pairs ]"
:
                                        lst.append(line)
                                elif line == "":
                                        pass
                                        #lst.append(line)
                                elif line.startswith("PGN"):
                                        pass
                                        #lst.append(line)
                                else:
                                        if section == "[ pairs ]":
                                                columns = line.split()
                                                new_line = "{0:5s} {1:5s} {2:5s}".fo
rmat(columns[0], columns[1], columns[2])
                                                lst.append(new_line)
                return lst

        e = pairs("topologies/bonded_clean.itp", d)

        def pairing(lst, x, y):
                # y = carbon atom, x = oxygen atom
                pair1 = (str(x), str(y - 18), "1")
                new_pair1 = "{0:5s} {1:5s} {2:5s}".format(pair1[0], pair1[1], pair1[
2])
                lst.append(new_pair1)
                pair2 = (str(x), str(y + 21), "1")
                new_pair2 = "{0:5s} {1:5s} {2:5s}".format(pair2[0], pair2[1], pair2[
2])
                lst.append(new_pair2)
                pair3 = (str(x), str(y + 36), "1")
                new_pair3 = "{0:5s} {1:5s} {2:5s}".format(pair3[0], pair3[1], pair3[
2])
                lst.append(new_pair3)
                pair4 = (str(x), str(y - 46), "1")
                new_pair4 = "{0:5s} {1:5s} {2:5s}".format(pair4[0], pair4[1], pair4[
2])
                lst.append(new_pair4)
                pair5 = (str(x), str(y + 25), "1")
                new_pair5 = "{0:5s} {1:5s} {2:5s}".format(pair5[0], pair5[1], pair5[
2])
                lst.append(new_pair5)
                pair6 = (str(x), str(y + 1), "1")
                new_pair6 = "{0:5s} {1:5s} {2:5s}".format(pair6[0], pair6[1], pair6[
2])
                lst.append(new_pair6)
                pair7 = (str(x - 4), str(y + 2), "1")
                new_pair7 = "{0:5s} {1:5s} {2:5s}".format(pair7[0], pair7[1], pair7[
2])
                lst.append(new_pair7)
                pair8 = (str(x - 4), str(y - 51), "1")
```

```
                new_pair8 = "{0:5s} {1:5s} {2:5s}".format(pair8[0], pair8[1], pair8[
2])
                lst.append(new_pair8)
                pair9 = (str(x - 4), str(y + 22), "1")
                new_pair9 = "{0:5s} {1:5s} {2:5s}".format(pair9[0], pair9[1], pair9[
2])
                lst.append(new_pair9)
                pair10 = (str(x - 61), str(y), "1")
                new_pair10 = "{0:5s} {1:5s} {2:5s}".format(pair10[0], pair10[1], pai
r10[2])
                lst.append(new_pair10)
                pair11 = (str(x - 1), str(y), "1")
                new_pair11 = "{0:5s} {1:5s} {2:5s}".format(pair11[0], pair11[1], pai
r11[2])
                lst.append(new_pair11)
                pair12 = (str(x - 8), str(y), "1")
                new_pair12 = "{0:5s} {1:5s} {2:5s}".format(pair12[0], pair12[1], pai
r12[2])
                lst.append(new_pair12)
                return lst

        t = pairing(e, x, y)

        def angles(filename, lst):
                section = None
                with open(filename, "r") as file:
                        for line in file:
                                line = line.strip()
                                if line.startswith("["):
                                        if line == "[ angles ]":
                                                lst.append(line)
                                        section = line
                                elif line.startswith(";") and section == "[ angles ]
":
                                        lst.append(line)
                                elif line == "":
                                        pass
                                        #lst.append(line)
                                elif line.startswith("PGN"):
                                        pass
                                else:
                                        if section == "[ angles ]":
                                                columns = line.split()
                                                new_line = "{0:5s} {1:5s} {2:5s} {3:
5s}".format(columns[0], columns[1], columns[2], columns[3])
                                                lst.append(new_line)
                return lst

        f = angles("topologies/bonded_clean.itp", t)

        def angling(lst, x, y):
                pass
                # y = carbon atom, x = oxygen atom
                angle = (str(x - 4), str(x), str( y), "5")
                new_angle = "{0:5s} {1:5s} {2:5s} {3:5s}".format(angle[0], angle[1],
 angle[2], angle[3])
                lst.append(new_angle)
                angle2 = (str(x), str(y), str(y - 51), "5")
                new_angle2 = "{0:5s} {1:5s} {2:5s} {3:5s}".format(angle2[0], angle2[
1], angle2[2], angle2[3])
                lst.append(new_angle2)
                angle3 = (str(x), str(y), str(y+2), "5")
                new_angle3 = "{0:5s} {1:5s} {2:5s} {3:5s}".format(angle3[0], angle3[
1], angle3[2], angle3[3])
                lst.append(new_angle3)
                angle4 = (str(x), str(y), str(y+22), "5")
                new_angle4 = "{0:5s} {1:5s} {2:5s} {3:5s}".format(angle4[0], angle4[
1], angle4[2], angle4[3])
                lst.append(new_angle4)
                return lst

        g = angling(f, x, y)
```

```
        def dihedrals_proper(filename, lst):
                gi = 0
                dihedral_counter = 0
                section = None
                repeats = 0
                previous_section = None
                saved_line = None
                with open(filename, "r") as file:
                        for line in file:
                                line = line.strip()
                                if line.startswith("["):
                                        if line == "[ dihedrals ]":
                                                dihedral_counter = dihedral_counter
+ 1
                                                if line == "[ dihedrals ]" and dihed
ral_counter == 1:
                                                        lst.append(line)
                                        section = line
                                elif line.startswith(";") and section == "[ dihedral
s ]" and dihedral_counter < 2:
                                        lst.append(line)
                                elif line == "":
                                        pass
                                        #lst.append(line)
                                elif line.startswith("PGN"):
                                        #lst.append(line)
                                        previous_section = section
                                else:
                                        if section == "[ dihedrals ]" and dihedral_c
ounter < 2:
                                                columns = line.split()
                                                new_line = "{0:5s} {1:5s} {2:5s} {3:
5s} {4:5s}".format(columns[0], columns[1], columns[2], columns[3], columns[4])
                                                lst.append(new_line)
                return lst

        h = dihedrals_proper("topologies/bonded_clean.itp", g)

        def dihedraling_proper(lst, x, y):
                pass
                # y = carbon atom, x = oxygen atom
                dihedral = (str(x-4), str(x), str(y), str(y-51), "9")
                new_dihedral = "{0:5s} {1:5s} {2:5s} {3:5s} {4:5s}".format(dihedral[
0], dihedral[1], dihedral[2], dihedral[3], dihedral[4])
                lst.append(new_dihedral)
                dihedral2 = (str(x-4), str(x), str(y), str(y+2), "9")
                new_dihedral2 = "{0:5s} {1:5s} {2:5s} {3:5s} {4:5s}".format(dihedral
2[0], dihedral2[1], dihedral2[2], dihedral2[3], dihedral2[4])
                lst.append(new_dihedral2)
                dihedral3 = (str(x-4), str(x), str(y), str(y+22), "9")
                new_dihedral3 = "{0:5s} {1:5s} {2:5s} {3:5s} {4:5s}".format(dihedral
3[0], dihedral3[1], dihedral3[2], dihedral3[3], dihedral3[4])
                lst.append(new_dihedral3)
                dihedral4 = (str(x), str(y), str(y+2), str(y+1), "9")
                new_dihedral4 = "{0:5s} {1:5s} {2:5s} {3:5s} {4:5s}".format(dihedral
4[0], dihedral4[1], dihedral4[2], dihedral4[3], dihedral4[4])
                lst.append(new_dihedral4)
                dihedral5 = (str(x), str(y), str(y+2), str(y+25), "9")
                new_dihedral5 = "{0:5s} {1:5s} {2:5s} {3:5s} {4:5s}".format(dihedral
5[0], dihedral5[1], dihedral5[2], dihedral5[3], dihedral5[4])
                lst.append(new_dihedral5)
                dihedral6 = (str(x), str(y), str(y+2), str(y-46), "9")
                new_dihedral6 = "{0:5s} {1:5s} {2:5s} {3:5s} {4:5s}".format(dihedral
6[0], dihedral6[1], dihedral6[2], dihedral6[3], dihedral6[4])
                lst.append(new_dihedral6)
                dihedral7 = (str(x), str(y), str(y+22), str(y-18), "9")
                new_dihedral7 = "{0:5s} {1:5s} {2:5s} {3:5s} {4:5s}".format(dihedral
7[0], dihedral7[1], dihedral7[2], dihedral7[3], dihedral7[4])
                lst.append(new_dihedral7)
                dihedral8 = (str(x), str(y), str(y+22), str(y+36), "9")
                new_dihedral8 = "{0:5s} {1:5s} {2:5s} {3:5s} {4:5s}".format(dihedral
8[0], dihedral8[1], dihedral8[2], dihedral8[3], dihedral8[4])
                lst.append(new_dihedral8)
```

```
                dihedral9 = (str(x), str(y), str(y+22), str(y+21), "9")
                new_dihedral9 = "{0:5s} {1:5s} {2:5s} {3:5s} {4:5s}".format(dihedral
9[0], dihedral9[1], dihedral9[2], dihedral9[3], dihedral9[4])
                lst.append(new_dihedral9)
                return lst
        i = dihedraling_proper(h, x, y)

        def dihedrals_improper(filename, lst):
                dihedral_counter = 0
                section = None
                with open(filename, "r") as file:
                        for line in file:
                                line = line.strip()
                                if line.startswith("["):
                                        if line == "[ dihedrals ]":
                                                dihedral_counter = dihedral_counter
+ 1
                                                if line == "[ dihedrals ]" and dihed
ral_counter == 2:
                                                        lst.append(line)
                                        section = line
                                elif line.startswith(";") and section == "[ dihedral
s ]" and dihedral_counter == 2:
                                        lst.append(line)
                                elif line == "":
                                        pass
                                        #lst.append(line)
                                else:
                                        if section == "[ dihedrals ]" and dihedral_c
ounter == 2:
                                                columns = line.split()
                                                new_line = "{0:5s} {1:5s} {2:5s} {3:
5s} {4:5s}".format(columns[0], columns[1], columns[2], columns[3], columns[4])
                                                lst.append(new_line)
                return lst

        j = dihedrals_improper("topologies/bonded_clean.itp", i)

        def dihedraling_improper(lst, x, y):
                #pass
                #dihedral = (str(y), str(y+7), str(x), str(y+56), "2", "gi_2")
                #new_dihedral = "{0:5s} {1:5s} {2:5s} {3:5s} {4:5s} {5:8s}".format(d
ihedral[0], dihedral[1], dihedral[2], dihedral[3], dihedral[4], dihedral[5])
                #lst.append(new_dihedral)
                with open("topologies/bonded_clean.itp", "w") as new_file:
                        for line in lst:
                                print(line, file=new_file)

        dihedraling_improper(j, x, y)


#BONDER CODE ENDS

#Structure CODE begins


def Structure(it):
        lst = []
        with open("structure_files/PGN_init.gro", "r") as infile:
                for line in infile:
                        line = line.strip()
                        if line.startswith("Peptidowhocan"):
                                lst.append(line)
                        try:
                                int(line) % 121 == 0
                                numline = str(121 * (it + 2))
                                lst.append(numline)
                        except ValueError:
                                pass
                        if line.startswith("1PGN"):
                                columns = line.split()
                                columns[0] = columns[0]
```

```
                                columns[1] = columns[1]
                                columns[2] = int(columns[2]) % 9999
                                columns[3] = columns[3]
                                columns[4] = columns[4]
                                columns[5] = columns[5]
                                new_line = "{0:>9s} {1:>5s} {2:>4d} {3:>7s} {4:>7s}
{5:>7s}".format(columns[0], columns[1], columns[2], columns[3], columns[4], columns[
5])
                                lst.append(new_line)
        with open("structure_files/PGN_mon.gro", "r") as infile:
                for line in infile:
                        line = line.strip()
                        if line.startswith("Peptidowhocan"):
                                pass
                        if line.startswith("121"):
                                pass
                        if line.startswith("1PGN"):
                                columns = line.split()
                                columns[0] = columns[0]
                                columns[1] = columns[1]
                                columns[2] = ((int(columns[2]) + int(121 * (it + 1))
) % 9999) + 1
                                columns[3] = "{0:.3f}".format(float(columns[3]) + (1
.14 * (it + 1)))
                                columns[4] = columns[4]
                                columns[5] = columns[5]
                                new_line = "{0:>9s} {1:>5s} {2:>4d} {3:>7s} {4:>7s}
{5:>7s}".format(columns[0], columns[1], columns[2], columns[3], columns[4], columns[
5])
                                lst.append(new_line)
                        if line.startswith("1.14"):
                                columns = line.split()
                                columns[0] = 1.14 + (1.14 * (it + 1))
                                columns[1] = columns[1]
                                columns[2] = columns[2]
                                new_line = "{0:5f} {1:5s} {2:5s}".format(columns[0],
 columns[1], columns[2])
                                lst.append(new_line)
                                if it == number_of_monomers - 1:
                                        x_dim = columns[0]
        with open ("structure_files/PGN_grown.gro", "w") as out:
                for line in lst:
                        print(line, file=out)

#Structure code ends
lst = []
rep_counter = 0
it = 0
x_dim = 0
x_val = int(raw_input("How wide would you like your box? (in nanometres): "))
y_val = int(raw_input("How long would you like your box? (in nanometres): "))
x_box = 1.14
y_box = 1.60
#       This is to sort out the fact that you need a multiple of four monomers and a
n even number of chains because of the bonding program
implemented_x = int((float(x_val) / x_box ))
implemented_y = int((float(y_val) / y_box))
req_area = x_val * y_val
real_x = float(implemented_x) * 1.14
real_y = float(implemented_y) * 1.60
real_area = real_x * real_y
number_of_monomers = implemented_x
reset = 0
top_name = "topologies/PGN_mon.itp"
top_lst = ["topologies/PGN_mon.itp"]
while it < (number_of_monomers - 1):
        Structure(it)
        it = it + 1
        top_lst.append(top_name)
        Glue(top_lst)
        renumber("topologies/glued.itp")
        #PGN CREATOR WAS HERE BUT IS NOW FUNCTIONS BEFORE THIS LOOP
```

```
        x = 1.0
        y = 2.0
        z = 4.00

        x_dim = int(it)

        print("PEPTIDOYOUCAN STEP {}".format(x_dim))

        n = 0
        x_vir = 2
        gro_count = 1
        number_of_bonds = int(x_dim)
        iterate = 0
        c = 67
        o = 121

        while iterate < number_of_bonds:
                print(iterate)
                y = (c + (n * o))
                x = (x_vir * o)
                n = n + 1
                x_vir = x_vir + 1
                Bonder(x,y)
                iterate = iterate + 1
        gromacs.grompp(f="mdps/minim.mdp" , c="structure_files/PGN_grown.gro", o="re
sults/mini.tpr", p="topologies/system.top", maxwarn="1")
        gromacs.mdrun(deffnm="PGN_init", s="results/mini.tpr",  v=True)
        os.system("rm PGN_init.log")
        os.system("rm PGN_init.edr")
        os.system("rm PGN_init.trr")
        os.system("mv PGN_init.gro structure_files/.")



c = 67
o = 121
y = (o)
x = c + ((number_of_bonds ) * o)
Bonder(y,x)
gromacs.grompp(f="mdps/minim.mdp" , c="structure_files/PGN_init.gro", o="results/min
i.tpr", p="topologies/system.top", maxwarn="1")
gromacs.mdrun(deffnm="PGN_grown", s="results/mini.tpr")
os.system("mv PGN_grown.gro structure_files/.")



        #PGN CREATOR HAS RUN USING OUR FUNCTIONS HERE
lst = []
y_num = implemented_y
counter = 0
y_iter = 0
sht_lst = []
init_numline = None
while counter < y_num:
        with open("structure_files/PGN_grown.gro", "r") as sheet_file:
                for line in sheet_file:
                        line = line.strip()
                        columns = line.split()
                        if line.startswith("Peptidowhocan") and counter == 0:
                                sht_lst.append(line)
                        try:
                                int(line) % 121 == 0
                                init_numline = int(line)
                                init_numline = init_numline * y_num
                                if y_iter == 0:
                                        sht_lst.append(str(init_numline))
                                        y_iter = y_iter + 1
                                        #numline = str(int(line) * (int(line * count
er)))
                        except ValueError:
                                pass
                        if line.startswith("1PGN"):
                                columns[0] = columns[0]
```

```
                                columns[1] = columns[1]
                                columns[2] = int(columns[2]) % 9999
                                columns[3] = columns[3]
                                columns[4] = "{0:.3f}".format(float(float(columns[4]
) + (float(1.6 * counter))))
                                columns[5] = columns[5]
                                new_line = "{0:>9s} {1:>5s} {2:>4d} {3:>7s} {4:>7s}
{5:>7s}".format(columns[0], columns[1], columns[2], columns[3], columns[4], columns[
5])
                                sht_lst.append(new_line)
                        if line.endswith("4.00000") and counter == y_num - 1:
                                columns[0] = columns[0]
                                columns[1] = (float(float(columns[1]) * y_num))
                                columns[2] = columns[2]
                                new_line = "{0:5s} {1:5f} {2:5s}".format(columns[0],
 columns[1], columns[2])
                                sht_lst.append(new_line)
        counter = counter + 1
        with open("structure_files/Sheet_t.gro", "w") as newfile:
                for line in sht_lst:
                        print(line, file=newfile)


tran_list = []
line_counter = 0

with open("structure_files/Sheet_t.gro", "r") as infile:
        for line in infile:
                line = line.strip()
                columns = line.split()
                if line.startswith("Peptidowhocan"):
                        tran_list.append(line)
                try:
                        line = int(line)
                        tran_list.append(line)
                        continue
                except ValueError:
                        pass

                if columns[0].endswith("PGN"):
                        line_counter = line_counter + 1
                        columns[2] = str(line_counter)
                        new_line = "{0:>9s} {1:>5s} {2:>6} {3:>7s} {4:>7s} {5:>7s}".
format(columns[0], columns[1], columns[2], columns[3], columns[4], columns[5])
                        tran_list.append(new_line)
                if line.endswith("4.00000"):
                        tran_list.append(line)

with open("structure_files/transform_out.gro", "w") as outfile:
        for line in tran_list:
                print(line, file=outfile)


top_name = "topologies/bonded_clean.itp"
top_count = 0
top_lst = []

while top_count < y_num:
        top_lst.append(top_name)
        top_count = top_count + 1
Glue(top_lst)
renumber("topologies/glued.itp")
os.system("mv topologies/bonded_clean.itp topologies/bonded.itp")

it = it + 1
counter = 0
y_iter = 0
new_lst = []
del_percentage = float(raw_input("What percentage of cross linking would you like? "
))
del_float = float(del_percentage/100)

add_lst = []
```

```
total_del_lst = []
lst = []
del_lst = []
counter = 0
chain_counter = 0
atoms_per_monomer = 121
number_of_monomers = implemented_x
atoms_per_chain = number_of_monomers * 121
number_of_chains = implemented_y
percentage_delete = del_float
hydrogen = 110
oxygen = 98
current_bonds = 0
total_atoms = atoms_per_chain * number_of_chains
max_bonds = (number_of_monomers * number_of_chains) / 2
number_of_bonds = int(float(max_bonds * percentage_delete))
num_deletions = number_of_bonds


while current_bonds < int(max_bonds):
        if chain_counter % 2 == 0:
                o_del = (oxygen + (counter * 242) + (chain_counter * atoms_per_chain
)) % total_atoms
                h_del = (hydrogen + (counter * 242) + ((chain_counter + 1) * atoms_p
er_chain)) % total_atoms
                h_del2 = (hydrogen + 1 + (counter * 242) + ((chain_counter + 1) * at
oms_per_chain)) % total_atoms
                del_lst.append([o_del, h_del, h_del2])
                counter = counter + 1
        if chain_counter % 2 != 0:
                o_del = (oxygen + 121 + (counter * 242) + (chain_counter * atoms_per
_chain)) % total_atoms
                h_del = (hydrogen + 121 + (counter * 242) + ((chain_counter + 1) * a
toms_per_chain)) % total_atoms
                h_del2 = (hydrogen + 122 + (counter * 242) + ((chain_counter + 1) *
atoms_per_chain)) % total_atoms
                del_lst.append([o_del, h_del, h_del2])
                counter = counter + 1
        if counter % (number_of_monomers / 2) == 0:
                chain_counter = chain_counter + 1
                counter = 0
                current_bonds = current_bonds + (number_of_monomers/2)

guarantee_counter = 0
while guarantee_counter < number_of_bonds and guarantee_counter < number_of_chains:
        if guarantee_counter % 2 == 0:
                total_del_lst.append(del_lst[(guarantee_counter * (number_of_monomer
s/2))])
        elif guarantee_counter % 2 != 0:
                total_del_lst.append(del_lst[((guarantee_counter * (number_of_monome
rs/2)) + 1)])
        guarantee_counter = guarantee_counter +  1

number_of_bonds = number_of_bonds - guarantee_counter

bonding_counter = 0

for i in del_lst:
        if i not in total_del_lst:
                add_lst.append(i)
while bonding_counter < number_of_bonds:
        random.shuffle(add_lst)
        bond = add_lst.pop()
        total_del_lst.append(bond)
        bonding_counter = bonding_counter + 1

flat_b_list = [item for sublist in total_del_lst for item in sublist]

bonds_created = 0
max_bonding = num_deletions

#Pep_Bonder code
```

```
""" The following script contains the parameters for establising peptide bonds and p
eptidoglycan sugar linkages, prior to using this script ensure that you have deleted
 the correct atoms from the molecules that you are bonding together, if you're readi
ng this there's a good chance you're here because your bond parameters are not inclu
ded, a later comment shows how to add your own bonding parameters for a specific sys
tem"""


def Pep_Bonder(filename,x,y):
        def start(filename, lst):
                section = None
                with open(filename, "r") as file:
                        for line in file:
                                line = line.strip()
                                if line.startswith("["):
                                        section = line
                                        if line == "[ moleculetype ]":
                                                lst.append(line)
                                elif line.startswith(";") and section == "[ molecule
type ]":
                                        lst.append(line)
                                elif line.startswith("PGN"):
                                        lst.append(line)
                return lst
        list1 = start(filename, [])

        def atoms(filename, lst):
                section = None
                with open(filename, "r") as file:
                        for line in file:
                                line = line.strip()
                                if line.startswith("["):
                                        section = line
                                        if line == "[ atoms ]":
                                                lst.append(line)
                                elif line.startswith(";") and section == "[ atoms ]"
:
                                        lst.append(line)
                                elif line == "":
                                        pass
                                elif line.startswith("PGN"):
                                        pass
                                elif section== "[ atoms ]":
                                        columns = line.split()
                                        new_line = "{0:6s} {1:10s} {2:6s} {3:6s} {4:
6s} {5:6s} {6:10s} {7:10s}".format(columns[0], columns[1], columns[2], columns[3], c
olumns[4], columns[5], columns[6], columns[7])
                                        lst.append(new_line)
                return lst

        a = atoms(filename, list1)

        def bonds_sec(filename, lst):
                section = None
                with open(filename, "r") as file:
                        for line in file:
                                line = line.strip()
                                if line.startswith("["):
                                        if line == "[ bonds ]":
                                                lst.append(line)
                                        section = line
                                elif line.startswith(";") and section == "[ bonds ]"
:
                                        lst.append(line)
                                elif line == "":
                                        pass
                                        #lst.append(line)
                                elif line.startswith("PGN"):
                                        #lst.append(line)
                                        previous_section = section
                                else:
                                        if section == "[ bonds ]":
                                                columns = line.split()
```

```
                                                new_line = "{0:5s} {1:5s} {2:5s}".fo
rmat(columns[0], columns[1], columns[2])
                                                lst.append(new_line)
                return lst

        c = bonds_sec(filename, a)

        def bonding(lst, x, y):
                # y = carbon atom, x = nitrogen atom
                bond = (str(x), str(y), "1")
                new_bond = "{0:5s} {1:5s} {2:5s}".format(bond[0], bond[1], bond[2])
                lst.append(new_bond)
                return lst

        d = bonding(c, x, y)

        def pairs(filename, lst):
                section = None
                with open(filename, "r") as file:
                        for line in file:
                                line = line.strip()
                                if line.startswith("["):
                                        if line == "[ pairs ]":
                                                lst.append(line)
                                        section = line
                                elif line.startswith(";") and section == "[ pairs ]"
:
                                        lst.append(line)
                                elif line == "":
                                        pass
                                        #lst.append(line)
                                elif line.startswith("PGN"):
                                        pass
                                        #lst.append(line)
                                else:
                                        if section == "[ pairs ]":
                                                columns = line.split()
                                                new_line = "{0:5s} {1:5s} {2:5s}".fo
rmat(columns[0], columns[1], columns[2])
                                                lst.append(new_line)
                return lst

        e = pairs(filename, d)

        def angles(filename, lst):
                section = None
                with open(filename, "r") as file:
                        for line in file:
                                line = line.strip()
                                if line.startswith("["):
                                        if line == "[ angles ]":
                                                lst.append(line)
                                        section = line
                                elif line.startswith(";") and section == "[ angles ]
":
                                        lst.append(line)
                                elif line == "":
                                        pass
                                        #lst.append(line)
                                elif line.startswith("PGN"):
                                        pass
                                else:
                                        if section == "[ angles ]":
                                                columns = line.split()
                                                new_line = "{0:5s} {1:5s} {2:5s} {3:
5s}".format(columns[0], columns[1], columns[2], columns[3])
                                                lst.append(new_line)
                return lst

        f = angles(filename, e)

        def angling(lst, x, y):
                # y = carbon atom, x = nitrogen atom
```

```
                angle = (str(y), str(x), str(x+43), "5")
                new_angle = "{0:5s} {1:5s} {2:5s} {3:5s}".format(angle[0], angle[1],
 angle[2], angle[3])
                lst.append(new_angle)
                angle2 = (str(y+14), str(y), str(x), "5")
                new_angle2 = "{0:5s} {1:5s} {2:5s} {3:5s}".format(angle2[0], angle2[
1], angle2[2], angle2[3])
                lst.append(new_angle2)
                angle3 = (str(y-2), str(y), str(x), "5")
                new_angle3 = "{0:5s} {1:5s} {2:5s} {3:5s}".format(angle3[0], angle3[
1], angle3[2], angle3[3])
                lst.append(new_angle3)
                angle4 = (str(y), str(x), str(x+20), "5")
                new_angle4 = "{0:5s} {1:5s} {2:5s} {3:5s}".format(angle4[0], angle4[
1], angle4[2], angle4[3])
                lst.append(new_angle4)
                return lst

        g = angling(f, x, y)

        def dihedrals_proper(filename, lst):
                gi = 0
                dihedral_counter = 0
                section = None
                repeats = 0
                previous_section = None
                saved_line = None
                with open(filename, "r") as file:
                        for line in file:
                                line = line.strip()
                                if line.startswith("["):
                                        if line == "[ dihedrals ]":
                                                dihedral_counter = dihedral_counter
+ 1
                                                if line == "[ dihedrals ]" and dihed
ral_counter == 1:
                                                        lst.append(line)
                                        section = line
                                elif line.startswith(";") and section == "[ dihedral
s ]" and dihedral_counter < 2:
                                        lst.append(line)
                                elif line == "":
                                        pass
                                        #lst.append(line)
                                elif line.startswith("PGN"):
                                        #lst.append(line)
                                        previous_section = section
                                else:
                                        if section == "[ dihedrals ]" and dihedral_c
ounter < 2:
                                                columns = line.split()
                                                new_line = "{0:5s} {1:5s} {2:5s} {3:
5s} {4:5s}".format(columns[0], columns[1], columns[2], columns[3], columns[4])
                                                lst.append(new_line)
                return lst

        h = dihedrals_proper(filename, g)

        def dihedraling_proper(lst, x, y):
                # y = carbon atom, x = nitrogen atom
                #dihedral = (str(y-4), str(y-2), str(y), str(x), "9")
                #new_dihedral = "{0:5s} {1:5s} {2:5s} {3:5s} {4:5s}".format(dihedral
[0], dihedral[1], dihedral[2], dihedral[3], dihedral[4])
                #lst.append(new_dihedral)
                #dihedral2 = (str(y-4), str(y-2), str(y), str(x), "9")
                #new_dihedral2 = "{0:5s} {1:5s} {2:5s} {3:5s} {4:5s}".format(dihedra
l2[0], dihedral2[1], dihedral2[2], dihedral2[3], dihedral2[4])
                #lst.append(new_dihedral2)
                #dihedral3 = (str(y-2), str(y), str(x), str(x-1), "9")
                #new_dihedral3 = "{0:5s} {1:5s} {2:5s} {3:5s} {4:5s}".format(dihedra
l3[0], dihedral3[1], dihedral3[2], dihedral3[3], dihedral3[4])
                #lst.append(new_dihedral3)
                #dihedral4 = (str(y), str(x), str(x-1), str(x+2), "9")
```

```
                #new_dihedral4 = "{0:5s} {1:5s} {2:5s} {3:5s} {4:5s}".format(dihedra
l4[0], dihedral4[1], dihedral4[2], dihedral4[3], dihedral4[4])
                #lst.append(new_dihedral4)
                #dihedral5 = (str(y), str(x), str(x-1), str(x+2), "9")
                #new_dihedral5 = "{0:5s} {1:5s} {2:5s} {3:5s} {4:5s}".format(dihedra
l5[0], dihedral5[1], dihedral5[2], dihedral5[3], dihedral5[4])
                #lst.append(new_dihedral5)
                return lst
        i = dihedraling_proper(h, x, y)

        def dihedrals_improper(filename, lst):
                dihedral_counter = 0
                section = None
                with open(filename, "r") as file:
                        for line in file:
                                line = line.strip()
                                if line.startswith("["):
                                        if line == "[ dihedrals ]":
                                                dihedral_counter = dihedral_counter
+ 1
                                                if line == "[ dihedrals ]" and dihed
ral_counter == 2:
                                                        lst.append(line)
                                        section = line
                                elif line.startswith(";") and section == "[ dihedral
s ]" and dihedral_counter == 2:
                                        lst.append(line)
                                elif line == "":
                                        pass
                                        #lst.append(line)
                                else:
                                        if section == "[ dihedrals ]" and dihedral_c
ounter == 2:
                                                columns = line.split()
                                                new_line = "{0:5s} {1:5s} {2:5s} {3:
5s} {4:5s}".format(columns[0], columns[1], columns[2], columns[3], columns[4])
                                                lst.append(new_line)
                return lst

        j = dihedrals_improper(filename, i)

        def dihedraling_improper(lst, x, y):
                # y = carbon atom, x = nitrogen atom
                #dihedral = (str(y), str(y-2), str(x), str(y+1), "2", "gi_1")
                #new_dihedral = "{0:5s} {1:5s} {2:5s} {3:5s} {4:5s}".format(dihedral
[0], dihedral[1], dihedral[2], dihedral[3], dihedral[4])
                #lst.append(new_dihedral)
                #dihedral2 = (str(x), str(y), str(x-1), str(x+1), "2", "gi_1")
                #new_dihedral2 = "{0:5s} {1:5s} {2:5s} {3:5s} {4:5s}".format(dihedra
l2[0], dihedral2[1], dihedral2[2], dihedral2[3], dihedral2[4])
                #lst.append(new_dihedral2)

                with open("topologies/bonded.itp", "w") as new_file:
                        for line in lst:
                                print(line, file=new_file)

        dihedraling_improper(j, x, y)
# y = carbon , x = nitrogen



#Pep_Bonder code ends
print(flat_b_list)
while bonds_created < max_bonding:
        carbon = int(flat_b_list[0 + (3 * bonds_created)]) - 13
        nitrogen = int(flat_b_list[1 + (3 * bonds_created)]) - 44
        Pep_Bonder("topologies/bonded.itp", nitrogen, carbon)
        bonds_created = bonds_created + 1

with open("structure_files/transform_out.gro", "r") as infile:
        for line in infile:
                line = line.strip()
```

```
                columns = line.split()
                if line.startswith("Peptidowhocan"):
                        lst.append(line)
                try:
                        int(line) % 121 == 0
                        line = int(line) - int(number_of_chains) * 3 - int((len(tota
l_del_lst) - number_of_chains) * 3)
                        lst.append(line)
                        continue
                except ValueError:
                        pass
                if columns[0].endswith("PGN"):
                        if int(columns[2]) not in flat_b_list:
                                new_line = "{0:>9s} {1:>5s} {2:>4s} {3:>7s} {4:>7s}
{5:>7s}".format(columns[0], columns[1], columns[2], columns[3], columns[4], columns[
5])
                                lst.append(new_line)
                if line.endswith("4.00000"):
                        columns[0] = columns[0]
                        columns[1] = columns[1]
                        columns[2] = columns[2]
                        new_line = "{0:5s} {1:5s} {2:5s}".format(columns[0], columns
[1], columns[2])
                        lst.append(new_line)

with open("structure_files/PGN_del.gro", "w") as outfile:
        for line in lst:
                print(line, file=outfile)
joined_flat = ",".join(str(thing) for thing in flat_b_list)


#WALL DELETER BEGINS

lst = []
section = None
atom_lst = joined_flat
print(atom_lst)
atom_lst = [int(v) for v in atom_lst.split(",")]
i = int(len(atom_lst) / 3)
atom_nums = []
carbon = 85
nitrogen = 65
def Deleting_walls(atom_list):
        k = 0
        while (k < i):
                delete1 = atom_lst[(0 + 3 * k)]
                atom_nums.append(delete1)
                delete2 = atom_lst[(1 + 3 * k)]
                atom_nums.append(delete2)
                delete3 = atom_lst[(2 + 3 * k)]
                atom_nums.append(delete3)
                k = k + 1
        atom_nums.sort()
        print(atom_nums)
        with open("topologies/bonded.itp", "r") as file:
                for line in file:
                        line = line.strip()
                        if line.startswith("["):
                                lst.append(line)
                                section = line
                        elif line.startswith(";"):
                                lst.append(line)
                        elif line.startswith("#"):
                                lst.append(line)
                        elif line == "":
                                lst.append(line)
                        elif line.startswith("PGN"):
                                lst.append(line)
                        else:
                                if section == "[ atoms ]":
                                        columns = line.split()
                                        columns[0] = int(columns[0])
                                        if columns[0] not in atom_nums:
```

```
                                                counter = 0
                                                for x in atom_nums:
                                                        if columns[0] > x:
                                                                counter += 1
                                                columns[0] = columns[0] - counter

                                                new_line = "{0:6d} {1:10s} {2:6s} {3
:6s} {4:6s} {5:6s} {6:10s} {7:10s}".format(columns[0], columns[1], columns[2], colum
ns[3], columns[4], columns[5], columns[6], columns[7])
                                                lst.append(new_line)
                                if section == "[ bonds ]":
                                        columns = line.split()
                                        columns[0] = int(columns[0])
                                        columns[1] = int(columns[1])
                                        if columns[0] not in atom_nums:
                                                counter = 0
                                                for x in atom_nums:
                                                        if columns[0] > x:
                                                                counter += 1
                                                columns[0] = columns[0] - counter

                                                if columns[1] not in atom_nums:
                                                        counter = 0
                                                        for x in atom_nums:
                                                                if columns[1] > x:
                                                                        counter += 1
                                                        columns[1] = columns[1] - co
unter
                                                        new_line = "{0:5d} {1:5d} {2
:5s}".format(columns[0], columns[1], columns[2])
                                                        lst.append(new_line)
                                if section == "[ pairs ]":
                                        columns = line.split()
                                        columns[0] = int(columns[0])
                                        columns[1] = int(columns[1])
                                        if columns[0] not in atom_nums:
                                                counter = 0
                                                for x in atom_nums:
                                                        if columns[0] > x:
                                                                counter += 1
                                                columns[0] = columns[0] - counter

                                                if columns[1] not in atom_nums:
                                                        counter = 0
                                                        for x in atom_nums:
                                                                if columns[1] > x:
                                                                        counter += 1
                                                        columns[1] = columns[1] - co
unter
                                                        new_line = "{0:5d} {1:5d} {2
:5s}".format(columns[0], columns[1], columns[2])
                                                        lst.append(new_line)
                                if section == "[ angles ]":
                                        columns = line.split()
                                        columns[0] = int(columns[0])
                                        columns[1] = int(columns[1])
                                        columns[2] = int(columns[2])
                                        if columns[0] not in atom_nums:
                                                counter = 0
                                                for x in atom_nums:
                                                        if columns[0] > x:
                                                                counter += 1
                                                columns[0] = columns[0] - counter

                                                if columns[1] not in atom_nums:
                                                        counter = 0
                                                        for x in atom_nums:
                                                                if columns[1] > x:
                                                                        counter += 1
                                                        columns[1] = columns[1] - co
unter
                                                        if columns[2] not in atom_nu
ms:
```

```
                                                                counter = 0
                                                                for x in atom_nums:
                                                                        if columns[2
] > x:
                                                                                coun
ter += 1
                                                                columns[2] = columns
[2] - counter
                                                                new_line = "{0:5d} {
1:5d} {2:5d} {3:5s}".format(columns[0], columns[1], columns[2], columns[3])
                                                                lst.append(new_line)
                                        if section == "[ dihedrals ]":
                                                columns = line.split()
                                                columns[0] = int(columns[0])
                                                columns[1] = int(columns[1])
                                                columns[2] = int(columns[2])
                                                columns[3] = int(columns[3])
                                                if columns[0] not in atom_nums:
                                                        counter = 0
                                                        for x in atom_nums:
                                                                if columns[0] > x:
                                                                        counter += 1
                                                        columns[0] = columns[0] - counter

                                                        if columns[1] not in atom_nums:
                                                                counter = 0
                                                                for x in atom_nums:
                                                                        if columns[1] > x:
                                                                                counter += 1
                                                                columns[1] = columns[1] - co
unter
                                                                if columns[2] not in atom_nu
ms:
                                                                        counter = 0
                                                                        for x in atom_nums:
                                                                                if columns[2
] > x:
                                                                                        coun
ter += 1
                                                                        columns[2] = columns
[2] - counter
                                                                        if columns[3] not in
 atom_nums:
                                                                                counter = 0
                                                                                for x in ato
m_nums:
                                                                                        if c
olumns[3] > x:

    counter += 1
                                                                                columns[3] =
 columns[3] - counter
                                                                                new_line = "
{0:5d} {1:5d} {2:5d} {3:5d} {4:5s}".format(columns[0], columns[1], columns[2], colum
ns[3], columns[4])
                                                                                lst.append(n
ew_line)

        with open("topologies/Atom_deleted.itp", "w") as new_file:
                for line in lst:
                        print(line, file=new_file)

Deleting_walls(atom_lst)


#WALL DELETER ENDS

#os.system("python " + "Atom_deleter_wall.py " + joined_flat)

#Renumber function

renumber("topologies/Atom_deleted.itp")
```

```
#Renumber function

#os.system("python " + "renumber.py " + "bonded.itp")

#.GRO_deleter code ends
print("You have generated an unlinked sheet of cell wall that is made of {} strands
of {} monomers of peptidoglycan".format(y_num, it+1))
print("Let me get that Sheet linked up for you now")


tran_list = []
line_counter = 0

with open("structure_files/PGN_del.gro","r") as infile:
        for line in infile:
                line = line.strip()
                columns = line.split()
                if line.startswith("Peptidowhocan"):
                        tran_list.append(line)
                try:
                        line = int(line)
                        tran_list.append(line)
                except ValueError:
                        pass

                if columns[0].endswith("PGN"):
                        line_counter = line_counter + 1
                        columns[2] = str(line_counter)
                        if line_counter <= 9999:
                                new_line = "{0:>9s} {1:>5s} {2:>4s} {3:>7s} {4:>7s}
{5:>7s}".format(columns[0], columns[1], columns[2], columns[3], columns[4], columns[
5])
                                tran_list.append(new_line)
                        if line_counter > 9999:
                                columns[2] = columns[1] + str(line_counter)
                                new_line = "{0:>9s} {1:>10s} {2:>7s} {3:>7s} {4:>7s}
".format(columns[0], columns[2], columns[3], columns[4], columns[5])
                                tran_list.append(new_line)
                try:
                        if columns[2] == "4.00000":
                                tran_list.append(line)
                except IndexError:
                        pass

with open("results/mini_me.gro", "w") as outfile:
        for line in tran_list:
                print(line, file=outfile)




os.chdir("results")
gromacs.grompp(f="../mdps/full_min.mdp", c="mini_me.gro", p="../topologies/system.to
p", o="mini_defect.tpr", maxwarn="1")
gromacs.mdrun(deffnm="mini_defect", v=True)

gromacs.grompp(f="../mdps/cg_full_min.mdp", c="mini_defect.gro", p="../topologies/sy
stem.top", o="cg_min.tpr", maxwarn="1")
gromacs.mdrun(deffnm="cg_min", v=True)

os.system("cp ../topologies/system.top ../topologies/sol_sys.top")
os.system("gmx solvate -cp mini_defect.gro -cs -p ../topologies/sol_sys.top -o sol.g
ro")

gromacs.grompp(f="../mdps/full_min.mdp", c="sol.gro", p="../topologies/sol_sys.top",
 o="sol_min.tpr", maxwarn="1")

gromacs.mdrun(deffnm="sol_min", v=True)


gromacs.grompp(f="../mdps/charmm_nvt.mdp", c="sol_min.gro", p="../topologies/sol_sys
.top", o="nvt_stable.tpr", maxwarn="1")
```

```
gromacs.mdrun(deffnm="nvt_stable", v=True)

gromacs.grompp(f="../mdps/charmm_md.mdp", c="nvt_stable.gro", p="../topologies/sol_s
ys.top",o="cell_wall.tpr", maxwarn="1")

gromacs.mdrun(deffnm="cell_wall", v=True)
os.chdir("..")


print("You have requested a crosslinked cell wall that is %d wide and %d long" % (x_
val, y_val))
print("Peptidoyoucan can give you a crosslinked cell wall that is %f wide and %f lon
g" % (real_x, real_y))
print("For comparison purposes you have requested an area of %d nm square of cell wa
ll, given your input you will recieve an area of %f nm square of cell wall" % (req_a
rea, real_area))
print("If you are not satisfied with the size of cell wall you received, try changin
g your input parameters")
if del_percentage < 50:
        print("""
        ========================================        WARNING!        ============
============================
        YOU HAVE REQUESTED A CELL WALL WITH LESS THAN 50% CROSS LINKAGE, PEPTIDOYOUC
AN WILL DO THIS FOR YOU,
        HOWEVER BE AWARE THAT PERCENTAGE CROSS-LINKAGE IN E.COLI IS APPROXIMATELY 50
%
        PLEASE BE AWARE THIS MAY RESULT IN SOME UNREALISTIC BEHAVIOUR SUCH AS UNLINK
ED CHAINS IN THE CASES OF
        SMALL SHEETS OF CELL WALL. IF YOU WOULD LIKE UNLINKED STRANDS OF PEPTIDOGLYC
AN THERE ARE OTHER TOOLS
        IN PEPTIDOYOUCAN THAT WILL DO THIS FOR YOU.

                """)

os.system("cp structure_files/PGN_mon.gro structure_files/PGN_init.gro")
os.system("rm mdout.mdp")
os.system("rm PGN_grown.log")
os.system("rm PGN_grown.trr")
os.system("rm PGN_grown.edr")
```

7.2 Appendix B

THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Perspective

pubs.acs.org/JPCL

Progress in Molecular Dynamics Simulations of Gram-Negative Bacterial Cell Envelopes

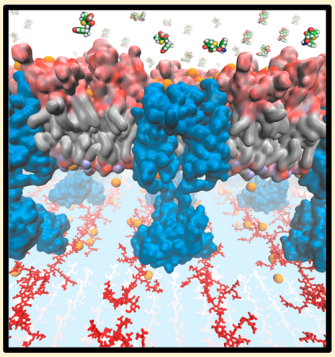
Alister Boags,[†,∥] Pin-Chia Hsu,[†,∥] Firdaus Samsudin,[†,∥] Peter J. Bond,[‡,§] and Syma Khalid*[,†]

[†]School of Chemistry, University of Southampton, Southampton, United Kingdom, SO17 1BJ

[‡]Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), Matrix 07-01, 30 Biopolis Street, 138671 Singapore

[§]Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, 117543 Singapore

ABSTRACT: Bacteria are protected by complex molecular architectures known as the cell envelope. The cell envelope is composed of regions with distinct chemical compositions and physical properties, namely, membranes and a cell wall. To develop novel antibiotics to combat pathogenic bacteria, molecular level knowledge of the structure, dynamics, and interplay between the chemical components of the cell envelope that surrounds bacterial cells is imperative. In addition, conserved molecular patterns associated with the bacterial envelope are recognized by receptors as part of the mammalian defensive response to infection, and an improved understanding of bacteria−host interactions would facilitate the search for novel immunotherapeutics. This Perspective introduces an emerging area of computational biology: multiscale molecular dynamics simulations of chemically complex models of bacterial lipids and membranes. We discuss progress to date, and identify areas for future development that will enable the study of aspects of the membrane components that are as yet unexplored by computational methods.



Bacteria are divided into two categories, Gram-negative and Gram positive, both of which include pathogens that are harmful to humans. Gram-negative bacteria have cell envelopes composed of two membranes, separated by a region known as the periplasm. The outer membrane (OM) is asymmetric in nature: the two leaflets differ in their compositions. The inner membrane contains a symmetric arrangement of phospholipids. By contrast, Gram-positive bacteria contain only one membrane, which is similar in composition to the inner membrane of Gram-negative bacteria. Both types of bacteria have a cell wall, which is composed of the biopolymer peptidoglycan. The combination of membrane plus cell wall gives rise to the characteristic semipermeable properties of the cell envelope. To be effective, antibiotics must either cause bacterial cell death or inhibit cell growth. In both cases, they must interact with the cell envelope, as they must either (i) disrupt the cell envelope, such that the cell contents leak out, or (ii) cross the cell envelope to gain access to the interior of the cell, where they may interfere with essential cellular process such as DNA replication and metabolism. The emergence of antimicrobial resistance is recognized as a major threat to human health.[1] It is thus imperative to have a detailed knowledge of the structure−dynamics−function relationships of the cell wall and membranes, in order to develop new antibiotics with reduced likelihood of resistance. Furthermore, molecules derived from the cell membrane and wall are utilized by the mammalian innate immune system to mount a defensive response.[2] Overamplification of such pathways can lead to sepsis, which remains the primary cause of death due to infection, highlighting the need for an improved understanding of the molecular mechanisms of immunostimulation. Due to the numerous molecular components involved, studying biological membranes in *in vivo* conditions remains an immensely challenging task. It is noteworthy that simulation of more complex biological membrane models is perhaps more prevalent for eukaryotic membranes, than bacterial ones, largely due to the simpler lipids in the former. In particular, both Marrink and Sansom and their research teams have independently reported eukaryotic membranes composed of multiple protein and lipid types.[3,4] In recent years, computational modeling has provided useful insights into the behavior of bacterial lipids and proteins in their native-like environment. Wonpil et al. recently reviewed current efforts in modeling and simulation of the bacterial OM, highlighting the dynamics of interactions between the components.[5] Here we discuss recent progress in the molecular simulation, in both atomic-resolution and simplified coarse-grained (CG) representations, of the envelopes of Gram-negative and their component molecules.

Gram-Negative Bacteria: The Inner Membrane. The two leaflets of the inner membrane are primarily composed of a mixture of zwitterionic (phosphatidyl-ethanolamine − PE) and anionic (cardiolipin − CL, phosphatidylglycerol − PG) phospholipids; for *Escherichia coli* they are in the ratio, 90:5:5.[6]

Simulation studies have traditionally employed one type of phospholipid to mimic the entire bilayer, with a greater emphasis placed on accurately modeling the membrane

Received: February 27, 2017
Accepted: May 3, 2017
Published: May 3, 2017

DOI: 10.1021/acs.jpclett.7b00473
J. Phys. Chem. Lett. 2017, 8, 2513−2518

It is imperative to have a detailed knowledge of the structure–dynamics–function relationships of cell walls and membranes, in order to develop new antibiotics with reduced likelihood of resistance.

protein/peptide/drug molecules that may interact with the membrane. In the past few years, the importance of the heterogeneity of the phospholipid composition has become increasingly apparent, and now often a mixture of phospholipids is used in simulation studies.[7,8] For example, Sansom and co-workers simulated the UraA H^+-uracil symporter from *E. coli* in a model of the inner membrane composed of a mixture of 1-palmitoyl 2-oleoyl-phosphatidylethanolamine (POPE), 1-palmitoyl 2-oleoyl-phosphatidylglycerol (POPG); and CL (Figure 1).

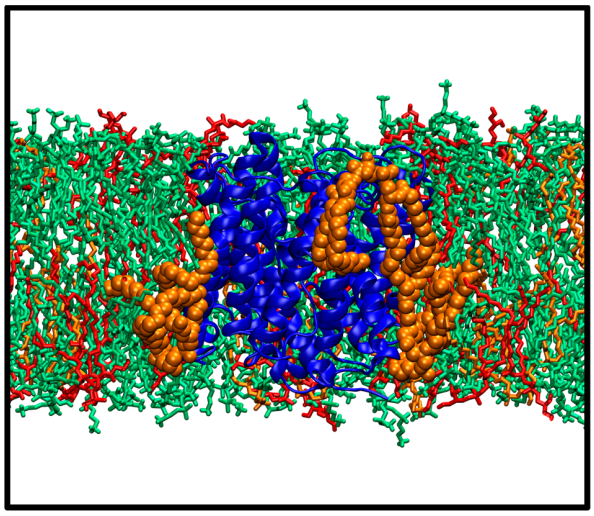


Figure 1. UraA symporter inserted in a POPE/POPG/CL bilayer.[3] UraA is shown in blue. The POPE, POPG, and CL lipids are shown in green, red, and orange, respectively. The three CL molecules that occupy the predicted CL binding sites are shown in orange van der Waals representation. The water and ions are omitted for clarity.

Despite the model membrane containing only 5% CL, preferential interaction of the protein with this lipid was observed. Simulations identified three specific cardiolipin binding sites on UraA, which led the authors to hypothesize that cardiolipin may form a "proton-trap" that channels protons to and from the protein.

The action of antimicrobial peptides at the inner membrane has been studied using atomistic molecular dynamics simulations. For example Amos et al. compared two helical antimicrobial peptides: pleurocidin and magainin 2 in inner membrane mimetic bilayers composed of a mixture of zwitterionic and ionic lipids.[9] They showed that pleurocidin has greater conformational flexibility and predicted that this was the basis of its greater potency against Gram-negative bacteria compared to magainin 2, which has less conformational lability.

It is important to note here that simulation studies of complex symmetrical membranes also focus on the membranes of Gram-positive bacteria, mitochondria, plants, and eukaryotes.[3,4,10−12] These studies are important in the progress in simulations of biological membranes, but are outside the scope of the current review, in which our focus is on Gram-negative bacteria.

Gram-Negative Bacteria: The Outer Membrane. The two leaflets of the outer membrane (OM) are rather different: the outer leaflet is composed of lipopolysaccharide (LPS) molecules, whereas the inner leaflet contains a mixture of zwitterionic and anionic phospholipids. In *E. coli* the inner leaflet of the outer membrane is composed of PE (90%), PG (5%), and CL (5%).[6] Thus, simplification of membrane models to one type of phospholipid is even more far removed from reality for the outer membrane than it is for the inner membrane. LPS molecules are large and heterogeneous, composed of the multiply acylated lipid A "membrane anchor", which contains a phosphorylated $\beta(1 \rightarrow 6)$ diglucosamine unit, attached to an extended polysaccharide that includes core regions and O-antigen. They also display two distinct phenotypes, i.e., smooth and rough; the former is characterized by the presence of a full-length O-antigen, while the latter represents a truncated or completely absent O-antigen. LPS, also known as endotoxin, provides structural integrity and chemical protection to the bacterial cell, and elicits strong immunological responses when invading mammalian hosts. In recent times, atomistic models of LPS of varying complexity have been reported in the literature,[13,14] with new CG models also emerging within the past few years.[15−17] These models have increased the scope of simulation studies, such that bacterial membrane models now incorporate important chemical details representing the *in vivo* membrane environment. The first simulation of a membrane containing LPS in the literature was that of rough PA01 LPS chemotype from *Pseudomonas aeruginosa* reported by Lins et al.[13] This was followed up by a study of the protein OprF embedded within an asymmetric outer membrane model containing rough LPS in the outer leaflet and PE in the inner leaflet.[18] Recently, more detailed models of the *E. coli* outer membrane with additional anionic phospholipids like PG and CL in the inner leaflet have been reported.[19−22] Simulations of outer membrane proteins (OMPs) in these detailed membranes have revealed protein–lipid interactions that are predicted to be key for the correct functioning of the protein. For example, the large extracellular loops of the TonB dependent transporter FecA were shown to specifically interact with LPS (Figure 2).[14] These interactions altered both the short-term fluctuations in the position of the loops, and also the longer-term loop conformations. Two different mechanisms were reported for these effects; short-lived but frequently formed nonspecific hydrogen-bonding interactions alter the local fluctuations in loop movement, and second the bulky LPS molecules provide a steric resistance to larger conformational motions of the loops. Previous simulation studies of TonB-dependent transporters were all performed with simple phospholipid membranes and thus unable to predict these functionally important membrane-protein interactions. In addition to native OMPs, insights into the interaction of antimicrobial peptides (AMPs) and a range of small chemicals with the outer membrane have also been studied at the atomistic level.[23−26] For example, extensive simulations were used to study the possible antimicrobial mechanisms of polymyxin B1 in both the IM and OM of *E. coli*, highlighting that AMPs are likely to interact with the two membranes in distinct ways.[25]

DOI: 10.1021/acs.jpclett.7b00473
J. Phys. Chem. Lett. 2017, 8, 2513−2518

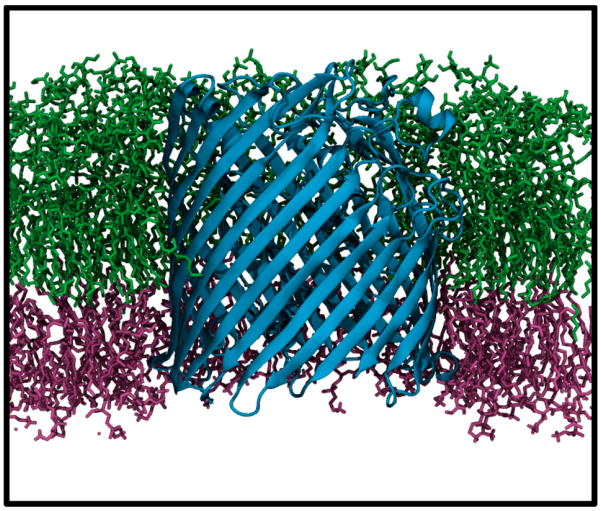


Figure 2. A snapshot of the FecA protein embedded in a model outer membrane, from Piggot et al.[14] FecA is shown in cyan, phospholipids are mauve and lipopolysaccharide molecules are lime green. Water and ions are omitted for clarity.

Atomistic models of LPS have increased the scope of simulation studies, such that bacterial membrane models now incorporate important chemical details representing the *in vivo* membrane environment.

The structure of lipid A molecules varies from bacterial species to species, both in terms of the number and length of acyl chains, phosphorylation and presence of other functional modifications on the sugar units. Chemical modification can occur under certain environmental conditions, and many of these play a role in bacterial pathogenesis and immunoevasion. To date, atomic-resolution simulations of lipid A from 12 different species of Gram-negative bacteria have been reported, as well as synthetic lipid A mimics developed for their potential therapeutic properties.[27,28] Im and co-workers provided a comparison of the properties of bilayers composed of several different lipid A types.[29]

At the CG level, we are only aware of models of LPS molecules developed within the framework of the popular MARTINI force-field.[15] In one recent study, the ability of pristine fullerenes to penetrate the outer membrane was found to be dependent upon the presence of phospholipids in the outer leaflet as well as the inner leaflet. This raises an important question; how much phospholipid is really present in the outer leaflet of Gram-negative bacteria? At the present time, there is no definitive answer to this question, yet the answer will be crucial for a full understanding of the permeability of these membranes and thus we are reliant upon experimental colleagues for data to enable us to construct more accurate models. A related and equally important question is how crowded is the outer membrane? This leads to uncertainties regarding how much "free" lipid there is in the outer membrane; in other words, how much lipid is dissociated from protein at any given time? From this we may also ask, how fluid really is the outer membrane? Addressing these questions requires complex simulations, to probe patterns of diffusion of proteins and lipids for which simplified, CG methods are ideal. Goose et al. reported one of the first large-scale simulations of membranes that simultaneously incorporated many OMPs.[30] Multiple copy numbers of the proteins OmpA, OmF, FhuA, LamB and NanC were simulated in mixtures of POPG and POPE phospholipids to mimic the outer membrane (CG models of LPS were not available at this time). The phospholipid diffusion around each protein was found to vary based on distance from protein, i.e., slower nearer to the protein. Furthermore, as the membrane models incorporated increasing numbers of proteins, thereby making them more crowded, lateral diffusion of both proteins and lipids slowed down. Vattulainen and co-workers had previously reported similar trends for eukaryotic proteins,[31] while Holdbrook et al. extended the observations to a diverse range of vesicles with varying composition to show that lipid dynamics can be coupled to proteins as far as 6 nm apart, providing evidence for long-range membrane communication.[32] Similarly, Rassam et al. employed large-scale coarse-grained simulations of BtuB and OmpF, along with fluorescence measurements, to assess protein−protein interactions in a crowded environment and their role in OM protein turnover in *E. coli*.[33]

The Bacterial Cell Wall. The cell wall of Gram-negative and Gram-positive bacteria is made up of a mesh of peptidoglycan, which is a sugar-peptide polymer. The glycan strand is conserved across all bacteria. In *E. coli* the sequence of the peptide stem of peptidoglycan is L-Ala (1) D-isoGlu (2) m-A2 pm (3) D-Ala (4) D-Ala (5), where m-A2 pm is meso-diaminopimelic acid, a derivative of lysine.[34] The width of this peptidoglycan layer in Gram-positive bacteria is generally much thicker (20−35 nm) than in Gram-negative bacteria (2−7 nm).[35,36] In the latter it is found in the periplasm, sandwiched between the two membranes, whereas in the former it is found outside the cell membrane. The precise three-dimensional structure of peptidoglycan varies between species. There have been far fewer simulation studies of peptidoglycan compared to bacterial membranes owing to the lack of structural data for these molecules and their higher-order architecture, as well as difficulties inherent to modeling extended, branched carbohydrates. However, significant recent progress has been made in this area. Gumbart et al. reported atomistic simulations of peptidoglycan from *E. coli*.[37] A single layer of peptidoglycan was constructed such that cross-linking between the peptidoglycan strands did not result in a highly ordered system, unlike those observed in previous models.[34] The peptidoglycan layer was in good agreement with experimentally measured physical properties such as thickness, elasticity and pore dimensions, and thus provided confidence in the use of atomistic simulations for modeling the periplasm. We have recently simulated an integrative model of the full-length *E. coli* OMP OmpA in its dimeric form[19,38] in the presence of a patch of OM and cell wall.[20] A peptidoglycan fragment was docked into the conserved binding site of the globular, OmpA-like C-terminal domain, using as a template the homologous protein domain from *Acinetobacter baumanii*, for which an X-ray structure cocrystallized with a peptidoglycan diaminopimelate has been solved.[39]

Simulations revealed the interaction between peptidoglycan and the protein to be rather labile, but persistent, and indicated how the domains of OmpA may serve as flexible anchors to provide adaptive mechanical support to the bacterial cell

DOI: 10.1021/acs.jpclett.7b00473
J. Phys. Chem. Lett. 2017, 8, 2513−2518

(Figure 3). These insights may offer alternative routes to the design of novel antibiotics.

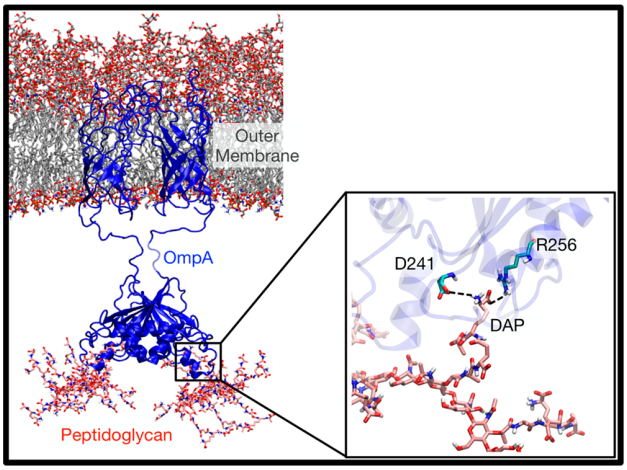


Figure 3. Model of peptidoglycan strands bound to a full-length OmpA homodimer embedded within a realistic outer membrane model containing LPS, as simulated by Samsudin et al.[20] Enlarged image shows the salt bridge interactions between diaminopimelate (DAP) on the peptidoglycan with D241 and R256 residues on the OmpA C-terminal domain.

Immunostimulation by Components of the Bacterial Cell Envelope. The evolutionarily conserved nature of the molecules associated with the bacterial cell wall and membrane—and their absence in mammalian cells—mark them out as primary indicators of a new infection. The 10 human members of the Toll-like receptor (TLR) family within the innate immune system are specialized for recognition of diverse pathogen-associated molecular patterns (PAMPs); bacterial PAMPs include, e.g., cell wall peptidoglycan, and the flagellin protein of the bacterial flagellum.[40] Another major PAMP is LPS, and its bioactive component lipid A, recognized by TLR4. This receptor has received particular interest over the years, not least because of its central role in propagation of endotoxic shock and sepsis during overwhelming Gram-negative bacterial infections.[2] A specialized relay of proteins exist to extract and transfer LPS molecules from bacterial membranes (or aggregates thereof) to the terminal cell surface receptor, TLR4, in complex with its lipid-binding coreceptor, MD-2. The great stability of the bacterial OM represents an interesting challenge for successful host recognition. Recently, atomic-resolution simulations were used to rigorously calculate the affinity of individual lipid A molecules for the lipid bilayer environment; a large barrier to lipid A extraction exists as a consequence of its extreme hydrophobicity combined with the divalent cation cross-links between phosphorylated headgroups.[28] Nevertheless, the MD-2 coreceptor was shown to have a comparable affinity for lipid A, revealing that the TLR4 system has evolved to create a membrane-like environment for endotoxin recognition. Furthermore, simulations revealed that the binding cavity of MD-2 exhibits clamshell-like dynamics,[27] which allow it to locally adapt to the size and shape of bound ligand. These dynamics, also observed in related proteins,[41,42] are allosterically coupled to the higher-order assembly of the TLR4/MD-2 complex,[27] thus ensuring that activation of TLR4 via dimerization occurs only in the presence of agonistic molecules such as hexa-acylated lipid A, but not antagonists such as lipid IVa, its biosynthetic precursor, or synthetic antagonists like Eritoran.[28] These simulations provide a platform for prediction of the activities of novel immunotherapeutic compounds, toward treatment of sepsis and other inflammatory conditions associated with bacterial infection.

Atomic-resolution simulations provide a platform for prediction of the activities of novel immunotherapeutic compounds, toward treatment of sepsis and other inflammatory conditions associated with bacterial infection.

What Next for Simulations of the Bacterial Cell Envelope? The past few years have seen tremendous progress in the complexity, time-scales, length-scales and therefore biological relevance of classical molecular simulation studies of components of the bacterial cell envelope. At the atomistic level, all of the components can now be simulated individually and in complex assemblies (Figure 4).

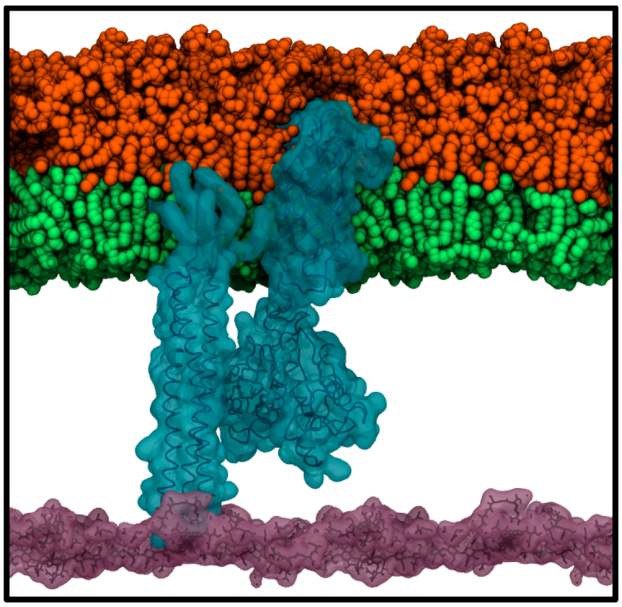


Figure 4. Snapshot from an atomistic simulation of the OM (green and orange) containing the full length OmpA dimer (cyan, right). Braun's lipoprotein (cyan, left) is held in position via a lipid anchor in the OM and covalent attachment to a sheet of peptodoglycan (pink). This is an extension of the peptidoglycan model reported in the work of Samsudin et al.[10]

However, the biology of these systems is inherently multiscale; there are time- and length-scales that are still beyond currently available models, if not methods. For example, it has long been speculated that some proteins and lipids are preferentially located at the poles of rod-shaped bacteria. To understand how and why requires simulation of larger biological membrane systems (the dimension of *E. coli* is approximately 0.5 × 2.0 μm) on relevant time scales, necessitating even coarser-grained models than those based on the ∼4:1 mapping

DOI: 10.1021/acs.jpclett.7b00473
J. Phys. Chem. Lett. 2017, 8, 2513−2518

that have recently become so popular. In addition, genome scale models of bacteria are also emerging, which further expand the system sizes that are addressable.[43,44] Diffusion of LPS is an order of magnitude slower than phospholipids, therefore extended time-lengths are required in order to probe the equilibrium behavior of these large systems. At the other end of the scale, to truly understand the chemical processes that can lead to the development of bacterial antibiotics resistance, such as the covalent modification of lipids by membrane enzymes, requires calculations at the quantum-level. For example, it is known that the outer membrane enzymes PagP and PagL modify the number of lipid tails of the lipid A portion of LPS.[45] Quantum mechanical (QM) calculations of membrane/protein systems are exceedingly rare, and have so far required a mixed QM/molecular mechanics (QM/MM) treatment.[46] Traditionally this has for the most part been due to the larger system sizes and compositional complexity of membrane protein systems compared to their soluble protein counterparts. However, recent progress in QM codes, such as the development of linear scaling density functional codes that can handle biological molecules, provide tremendous promise.[47] Attention to the development and validation of appropriate protocols to add quantum-level detail to the current arsenal of classical simulation tools will undoubtedly herald a step-change in terms of computational studies of bacterial membranes, and concurrently, rational design of antibiotics and immunotherapeutics.

AUTHOR INFORMATION

Corresponding Author

*E-mail: S.Khalid@soton.ac.uk.

ORCID

Syma Khalid: 0000-0002-3694-5044

Author Contributions

‖These authors contributed equally.

Funding

F.S. is funded by BBRSC grant number: BB/M029573.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Antreas Kalli for providing Figure 1.

REFERENCES

(1) WHO. Antimicrobial Resistance: Global Report on Surveillance. *Bull. World Health Organ.* **2014**, *61* (3), 383−394.

(2) O'Neill, L. A. J.; Bryant, C. E.; Doyle, S. L. Therapeutic Targeting of Toll-Like Receptors for Infectious and Inflammatory Diseases and Cancer. *Pharmacol. Rev.* **2009**, *61* (2), 177−197.

(3) Koldsø, H.; Sansom, M. S. P. Organization and Dynamics of Receptor Proteins in a Plasma Membrane. *J. Am. Chem. Soc.* **2015**, *137* (46), 14694−14704.

(4) van Eerden, F. J.; van den Berg, T.; Frederix, P. W. J. M.; de Jong, D. H.; Periole, X.; Marrink, S. J. Molecular Dynamics of Photosystem II Embedded in the Thylakoid Membrane. *J. Phys. Chem. B* **2017**, *121*, 3237.

(5) Patel, D. S.; Qi, Y.; Im, W. Modeling and Simulation of Bacterial Outer Membranes and Interactions with Membrane Proteins. *Curr. Opin. Struct. Biol.* **2017**, *43* (1), 131.

(6) Lugtenberg, E. J. J.; Peters, R. Distribution of Lipids in Cytoplasmic and Outer Membranes of Escherichia Coli K12. *Biochim. Biophys. Acta, Lipids Lipid Metab.* **1976**, *441* (1), 38−47.

(7) Kalli, A. C.; Sansom, M. S. P.; Reithmeier, R. A. F. Molecular Dynamics Simulations of the Bacterial UraA H+-Uracil Symporter in Lipid Bilayers Reveal a Closed State and a Selective Interaction with Cardiolipin. *PLoS Comput. Biol.* **2015**, *11* (3), e1004123.

(8) Alcock, F.; Stansfeld, P. J.; Basit, H.; Habersetzer, J.; Baker, M. A. B.; et al. Assembling the Tat Protein Translocase. *eLife* **2016**, *5*, e20718.

(9) Amos, S.-B. T. A.; Vermeer, L. S.; Ferguson, P. M.; Kozlowska, J.; Davy, M.; Bui, T. T.; Drake, A. F.; Lorenz, C. D.; Mason, A. J. Antimicrobial Peptide Potency Is Facilitated by Greater Conformational Flexibility When Binding to Gram-Negative Bacterial Inner Membranes. *Sci. Rep.* **2016**, *6*, 37639.

(10) Piggot, T. J.; Holdbrook, D. A.; Khalid, S. Electroporation of the E. Coli and S. Aureus Membranes: Molecular Dynamics Simulations of Complex Bacterial Membranes. *J. Phys. Chem. B* **2011**, *115*, 13381.

(11) Rui, H.; Lee, K. Il; Pastor, R. W.; Im, W. Molecular Dynamics Studies of Ion Permeation in VDAC. *Biophys. J.* **2011**, *100* (3), 602−610.

(12) Hedger, G.; Rouse, S. L.; Domanski, J.; Chavent, M.; Koldsoe, H.; Sansom, M. S. P. Lipid-Loving ANTs: Molecular Simulations of Cardiolipin Interactions and the Organization of the Adenine Nucleotide Translocase in Model Mitochondrial Membranes. *Biochemistry* **2016**, *55* (45), 6238−6249.

(13) Kirschner, K. N.; Lins, R. D.; Maass, A.; Soares, T. A. A Glycam-Based Force Field for Simulations of Lipopolysaccharide Membranes: Parametrization and Validation. *J. Chem. Theory Comput.* **2012**, *8* (11), 4719−4731.

(14) Piggot, T. J.; Holdbrook, D. A.; Khalid, S. Conformational Dynamics and Membrane Interactions of the E. Coli Outer Membrane Protein FecA: A Molecular Dynamics Simulation Study. *Biochim. Biophys. Acta, Biomembr.* **2013**, *1828* (2), 284−293.

(15) Hsu, P.-C.; Jefferies, D.; Khalid, S. Molecular Dynamics Simulations Predict the Pathways via Which Pristine Fullerenes Penetrate Bacterial Membranes. *J. Phys. Chem. B* **2016**, *120* (43), 11170−11179.

(16) Ma, H.; Irudayanathan, F. J.; Jiang, W.; Nangia, S. Simulating Gram-Negative Bacterial Outer Membrane: A Coarse Grain Model. *J. Phys. Chem. B* **2015**, *119* (119), 14668−14682.

(17) Jefferies, D.; Hsu, P. C.; Khalid, S. Through the Lipopolysaccharide Glass: A Potent Antimicrobial Peptide Induces Phase Changes in Membranes. *Biochemistry* **2017**, *56*, 1672.

(18) Straatsma, T. P.; Soares, T. A. Characterization of the Outer Membrane Protein OprF of Pseudomonas Aeruginosa in a Lipopolysaccharide Membrane by Computer Simulation. *Proteins: Struct., Funct., Genet.* **2009**, *74* (2), 475−488.

(19) Ortiz-Suarez, M. L.; Samsudin, F.; Piggot, T. J.; Bond, P. J.; Khalid, S. Full-Length OmpA: Structure, Function, and Membrane Interactions Predicted by Molecular Dynamics Simulations. *Biophys. J.* **2016**, *111* (8), 1692−1702.

(20) Samsudin, F.; Ortiz-Suarez, M. L.; Piggot, T. J.; Bond, P. J.; Khalid, S. OmpA: A Flexible Clamp for Bacterial Cell Wall Attachment. *Structure* **2016**, *24* (12), 2227.

(21) Wu, E. L.; Fleming, P. J.; Yeom, M. S.; Widmalm, G.; Klauda, J. B.; Fleming, K. G.; Im, W. E. Coli Outer Membrane and Interactions with OmpLA. *Biophys. J.* **2014**, *106* (11), 2493−2502.

(22) Patel, D. S.; Re, S.; Wu, E. L.; Qi, Y.; Klebba, P. E.; Widmalm, G.; Yeom, M. S.; Sugita, Y.; Im, W. Dynamics and Interactions of OmpF and LPS: Influence on Pore Accessibility and Ion Permeability. *Biophys. J.* **2016**, *110* (4), 930−938.

(23) Pino-Angeles, A.; Leveritt, J. M.; Lazaridis, T. Pore Structure and Synergy in Antimicrobial Peptides of the Magainin Family. *PLoS Comput. Biol.* **2016**, *12* (1), e1004570.

(24) Bemporad, D.; Luttmann, C.; Essex, J. W. Computer Simulation of Small Molecule Permeation across a Lipid Bilayer: Dependence on Bilayer Properties and Solute Volume, Size, and Cross-Sectional Area. *Biophys. J.* **2004**, *87* (1), 1−13.

(25) Berglund, N. A.; Piggot, T. J.; Jefferies, D.; Sessions, R. B.; Bond, P. J.; Khalid, S. Interaction of the Antimicrobial Peptide Polymyxin B1 with Both Membranes of E. Coli: A Molecular Dynamics Study. *PLoS Comput. Biol.* **2015**, *11* (4), e1004180.

DOI: 10.1021/acs.jpclett.7b00473
J. Phys. Chem. Lett. 2017, 8, 2513−2518

(26) Carpenter, T. S.; Parkin, J.; Khalid, S. The Free Energy of Small Solute Permeation through the Escherichia Coli Outer Membrane Has a Distinctly Asymmetric Profile. *J. Phys. Chem. Lett.* **2016**, *7*, 3446−3451.
(27) Paramo, T.; Piggot, T. J.; Bryant, C. E.; Bond, P. J. The Structural Basis for Endotoxin-Induced Allosteric Regulation of the Toll-like Receptor 4 (tlr4) Innate Immune Receptor. *J. Biol. Chem.* **2013**, *288* (51), 36215−36225.
(28) Paramo, T.; Tomasio, S. M.; Irvine, K. L.; Bryant, C. E.; Bond, P. J. Energetics of Endotoxin Recognition in the Toll-Like Receptor 4 Innate Immune Response. *Sci. Rep.* **2016**, *5*, 17997.
(29) Kim, S.; Patel, D. S.; Park, S.; Slusky, J.; Klauda, J. B.; Widmalm, G.; Im, W. Bilayer Properties of Lipid A from Various Gram-Negative Bacteria. *Biophys. J.* **2016**, *111* (8), 1750−1760.
(30) Goose, J. E.; Sansom, M. S. P. Reduced Lateral Mobility of Lipids and Proteins in Crowded Membranes. *PLoS Comput. Biol.* **2013**, *9* (4), e1003033.
(31) Niemelä, P. S.; Miettinen, M. S.; Monticelli, L.; Hammaren, H.; Bjelkmar, P.; Murtola, T.; Lindahl, E.; Vattulainen, I. Membrane Proteins Diffuse as Dynamic Complexes with Lipids. *J. Am. Chem. Soc.* **2010**, *132* (22), 7574−7575.
(32) Holdbrook, D. A.; Huber, R. G.; Piggot, T. J.; Bond, P. J.; Khalid, S. Dynamics of Crowded Vesicles: Local and Global Responses to Membrane Composition. *PLoS One* **2016**, *11* (6), e0156963.
(33) Rassam, P.; Copeland, N. a.; Birkholz, O.; Tóth, C.; Chavent, M.; Duncan, A. L.; Cross, S. J.; Housden, N. G.; Kaminska, R.; Seger, U.; et al. Supramolecular Assemblies Underpin Turnover of Outer Membrane Proteins in Bacteria. *Nature* **2015**, *523*, 333−336.
(34) Vollmer, W.; Blanot, D.; De Pedro, M. A. Peptidoglycan Structure and Architecture. *FEMS Microbiol. Rev.* **2008**, *32* (2), 149−167.
(35) Vollmer, W.; Seligman, S. J. Architecture of Peptidoglycan: More Data and More Models. *Trends Microbiol.* **2010**, *18* (2), 59−66.
(36) Beeby, M.; Gumbart, J. C.; Roux, B.; Jensen, G. J. Architecture and Assembly of the Gram-Positive Cell Wall. *Mol. Microbiol.* **2013**, *88* (4), 664−672.
(37) Gumbart, J. C.; Beeby, M.; Jensen, G. J.; Roux, B. Escherichia Coli Peptidoglycan Structure and Mechanics as Predicted by Atomic-Scale Simulations. *PLoS Comput. Biol.* **2014**, *10* (2), e1003475.
(38) Marcoux, J.; Politis, A.; Rinehart, D.; Marshall, D. P.; Wallace, M. I.; Tamm, L. K.; Robinson, C. V. Mass Spectrometry Defines the C-Terminal Dimerization Domain and Enables Modeling of the Structure of Full-Length OmpA. *Structure* **2014**, *22* (5), 781−790.
(39) Park, J. S.; Lee, W. C.; Yeo, K. J.; Ryu, K.-S.; Kumarasiri, M.; Hesek, D.; Lee, M.; Mobashery, S.; Song, J. H.; Kim, S. I.; et al. Mechanism of Anchoring of OmpA Protein to the Cell Wall Peptidoglycan of the Gram-Negative Bacterial Outer Membrane. *FASEB J.* **2012**, *26* (1), 219−228.
(40) Berglund, N. A.; Kargas, V.; Ortiz-Suarez, M. L.; Bond, P. J. The Role of Protein-Protein Interactions in Toll-like Receptor Function. *Prog. Biophys. Mol. Biol.* **2015**, *119* (1), 72−83.
(41) Paramo, T.; East, A.; Garzón, D.; Ulmschneider, M. B.; Bond, P. J. Efficient Characterization of Protein Cavities within Molecular Simulation Trajectories: Trj-Cavity. *J. Chem. Theory Comput.* **2014**, *10* (5), 2151−2164.
(42) Ortiz-Suarez, M. L.; Bond, P. J. The Structural Basis for Lipid and Endotoxin Binding in RP105-MD-1, and Consequences for Regulation of Host Lipopolysaccharide Sensitivity. *Structure* **2016**, *24* (1), 200−211.
(43) Liu, J. K.; O'Brien, E. J.; Lerman, J. a; Zengler, K.; Palsson, B. O.; Feist, A. M. Reconstruction and Modeling Protein Translocation and Compartmentalization in Escherichia Coli at the Genome-Scale. *BMC Syst. Biol.* **2014**, *8*, 110.
(44) Zhuang, K.; Vemuri, G. N.; Mahadevan, R. Economics of Membrane Occupancy and Respiro-Fermentation. *Mol. Syst. Biol.* **2011**, *7* (500), 500.
(45) Trent, M. S. Biosynthesis, Transport, and Modification of Lipid A. *Biochem. Cell Biol.* **2004**, *82* (1), 71−86.
(46) Lonsdale, R.; Rouse, S. L.; Sansom, M. S. P.; Mulholland, A. J. A Multiscale Approach to Modelling Drug Metabolism by Membrane-Bound Cytochrome P450 Enzymes. *PLoS Comput. Biol.* **2014**, *10* (7), e1003714.
(47) Skylaris, C. K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C. Introducing ONETEP: Linear-Scaling Density Functional Simulations on Parallel Computers. *J. Chem. Phys.* **2005**, *122* (8), 084119.

DOI: 10.1021/acs.jpclett.7b00473
J. Phys. Chem. Lett. 2017, 8, 2513−2518

7.3 Appendix C

Biophysical Journal
Article

Biophysical Society

Braun's Lipoprotein Facilitates OmpA Interaction with the *Escherichia coli* Cell Wall

Firdaus Samsudin,[1] Alister Boags,[1] Thomas J. Piggot,[1,2] and Syma Khalid[1,*]
[1]School of Chemistry, University of Southampton, Highfield, Southampton, United Kingdom and [2]CBR Division, Defence Science and Technology Laboratory, Porton Down, Salisbury, Wiltshire, United Kingdom

ABSTRACT Gram-negative bacteria such as *Escherichia coli* are protected by a complex cell envelope. The development of novel therapeutics against these bacteria necessitates a molecular level understanding of the structure-dynamics-function relationships of the various components of the cell envelope. We use atomistic MD simulations to reveal the details of covalent and noncovalent protein interactions that link the outer membrane to the aqueous periplasmic region. We show that the Braun's lipoprotein tilts and bends, and thereby lifts the cell wall closer to the outer membrane. Both monomers and dimers of the outer membrane porin OmpA can interact with peptidoglycan in the presence of Braun's lipoprotein, but in the absence of the latter, only dimers of OmpA show a propensity to form contacts with peptidoglycan. Our study provides a glimpse of how the molecular components of the bacterial cell envelope interact with each other to mediate cell wall attachment in *E. coli.*

INTRODUCTION

The cell envelope of *Escherichia coli* is composed of two membranes separated by a region known as the periplasm or the periplasmic space (1). The outer membrane (OM) is composed of lipopolysaccharide (LPS) molecules in the outer leaflet and a mixture of phospholipids, both zwitterionic and anionic, in the inner leaflet (2). Proteins that are integral to this membrane are almost invariably β-barrels in architecture (3). The periplasm contains the sugar-peptide polymer, peptidoglycan (PGN), as well as many different periplasmic proteins. The PGN network is attached to the OM and the integral membrane via both covalent and noncovalent interactions (4–6).

The only known protein that provides a covalent link to PGN is Braun's lipoprotein (BLP, also known as "Lpp" and "murein lipoprotein"), which is one of the most abundant proteins in *E. coli* (7,8). BLP is anchored in the OM via a lipidated N-terminus, whereas the C-terminus is covalently attached to the peptide chain of PGN. BLP exists in PGN-bound and PGN-unbound states, with the former representing approximately one-third of the population (9–12). Crystallographic data revealed that the *E. coli* BLP forms a stable homotrimer with a tight coiled coil motif held together by an alanine zipper unit (13). Recently, electron microscopy and electron cryomicroscopy studies showed that the length of BLP has a direct influence on the distance between the peptidoglycan layer and the outer membrane of *E. coli* (14). However, how the BLP trimer is positioned with respect to the OM and the PGN network, remains unknown at the individual molecule level.

In addition to the covalent linkage provided by BLP, PGN is also attached noncovalently to several OM and integral membrane proteins such as OmpA-like domains (15), PGN-associated lipoproteins (16), and flagella motor proteins (17). The *E. coli* outer membrane porin OmpA is a multidomain protein whose N-terminal domain (NTD) is made of a β-barrel and C-terminal domain (CTD) is a globular periplasmic unit that binds to PGN, connected by an unstructured 20-residue linker region (18). The NTD has been subject to numerous functional and structural studies (19–23), whereas the structure of the CTD has recently emerged from a NMR study (24). Experimental evidence suggests that the full-length OmpA can form a homodimer (25,26), the model of which has been proposed and validated by mass spectrometry and MD simulations (27–29). The mechanism of PGN attachment to OmpA CTD has been elucidated by crystal structures of a homolog from *Acinetobacter baumannii* bound to a short PGN peptide (30).

Although BLP and OmpA, and their interactions with PGN, have been extensively studied individually and it is likely that both proteins form simultaneous interactions with the PGN network in vivo, very little is known about

Submitted May 15, 2017, and accepted for publication August 2, 2017.
*Correspondence: s.khalid@soton.ac.uk
Editor: D. Peter Tieleman.
http://dx.doi.org/10.1016/j.bpj.2017.08.011

any interaction between the two proteins as the spatial arrangement of both proteins within the cell envelope is still largely unexplored at a resolution of individual molecules. Based on the x-ray structures, each helix of the BLP trimer is ~90 Å in length (13). If this represents the separation of the OM and the PGN layer, the OmpA linker would have to be fully extended to allow for CTD interaction with PGN. Such a conformation, however, is likely to be entropically unfavorable. Further studies including both proteins in the same environment are therefore crucial to understand the balance between covalent and noncovalent bonding between the OM and the PGN cell wall.

To this end, here we built atomistic models representing a portion of the *E. coli* cell envelope, namely the OM and the periplasm containing PGN, BLP, and OmpA. Unlike our previous study, the PGN sheet was positioned ~90 Å from the lower leaflet of the OM, unbound to the OmpA CTD, allowing us to examine how the latter can interact with PGN in the presence of BLP. This distance was specifically chosen, as it is the length of fully extended, unkinked BLP. Our simulations show that BLP lifts the PGN layer upwards by tilting and bending its helices. This in turn reduces the gap between the OM and the cell wall, thereby facilitating the initial contact between the OmpA CTD and PGN, especially in its monomeric form. OmpA dimers on the other hand are able to interact with the cell wall even in the absence of BLP by extending their linker domain. We also identify interactions between BLP and OmpA as well as showing the interaction of the latter with the cell wall.

METHODS

The models

The full-length OmpA monomer and dimer models were obtained from Carol Robinson (27); their structural stability in a model OM has been verified in our previous work (28,29). The OM model was asymmetric: the upper leaflet was made entirely of full-rough Ra LPS lipids of the R1 core type (31,32), whereas the lower leaflet comprised a mixture of phospholipids (i.e., 90% 1-palmitoyl 2-*cis*-vaccenic phosphatidylethanolamine, 5% 1-palmitoyl 2-*cis*-vaccenic phosphatidylglycerol, and 5% 1-palmitoyl 2-cis-vaccenic 3-palmitoyl 4-*cis*-vaccenic diphosphatidylglycerol, otherwise known as cardiolipin) (2,33–35). This OM model has been validated in our previous simulations studies (28,29,36). The OmpA structure was inserted into the OM model using g_membed (37), following the procedure previously described (29).

A PGN network consisting of three strands of 10 repeating NAG-NAM-peptide units was constructed and positioned ~90 Å from the surface of the lower leaflet of the OM. The BLP homotrimer was built based on the structure from Shu et al. (13) (PDB: 1EQ7) with the last residues on both the N- and C termini manually added back using PyMOL (38). The N-terminus was in turn attached to the tripalmitoyl-S-glyceryl-cysteine residues to incorporate the BLP to the inner leaflet of the OM. The parameters for tripalmitoyl-S-glyceryl-cysteine were constructed from the standard GROMOS 54A7 force field (39) with the GROMOS 53A6$_{OXY}$ (40) ether parameters used for the linkage region. PGN was then covalently linked to the Lysine on one of the C termini of the BLP trimer via its m-DAP residue. The linkage was constructed using the standard GROMOS 54A7 parameters.

Atomistic MD simulations

All simulations were performed using the GROMACS 5 code (41), the GROMOS 54A7 force field (39) with the SPC water model (42). Each simulation was run for 100 ns, and at least one independent repeat of each simulation was performed, giving at least 200 ns for each system simulated. Temperatures of 310 and 323 K were maintained using the velocity rescale thermostat (43) using a time constant of 1 ps. The pressure was maintained semiisotropically at 1 atm using the Parrinello-Rahman barostat (44) with a time constant of 1 ps. All bonds were constrained using the LINCS algorithm (45) to allow for an integration time step of 2 fs. Long-range electrostatics were described using the particle mesh Ewald method (46). The short-range electrostatic cutoff used was 1.2 nm and the short-range van der Waals cutoff was also 1.2 nm.

Short equilibration simulations were performed for each system in both the NVT and the NPT ensembles. The NVT equilibration was first run for 500 ps, followed by the NPT equilibration for another 1 ns, after which the pressure of the systems reached a plateau. No positional restraints were imposed on the proteins during these simulations. These equilibration simulations utilized the same thermostat and barostat as mentioned above.

RESULTS

Simulation systems

Four atomistic simulation systems were constructed as described in Fig. 1 and Table 1, containing either the full-length OmpA monomer or dimer (27), in the presence or absence of BLP (13). One system of only the BLP trimer in the absence of OmpA was also built as a control. The OmpA NTD was inserted into a biologically relevant model of the OM (2,33–35) described in the Methods. The binding of PGN to OmpA CTD occurs in a noncovalent fashion mediated by two residues, D241 and R256, in *E. coli* OmpA, as indicated by recent crystallographic data (30) and a simulation study (29). These residues, however, are located deep within the OmpA CTD, suggesting that it is likely that other residues on the surface of the protein might be involved in initial binding. We therefore started all our simulations with the CTD of OmpA positioned around

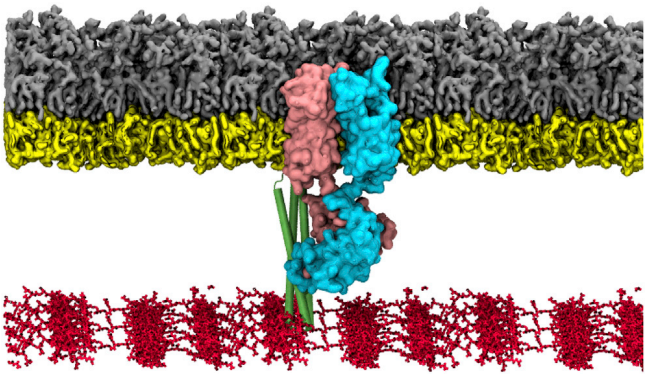


FIGURE 1 Simulation setup. Given here is a snapshot of a simulation system with the full-length OmpA dimer (*cyan* and *pink*), BLP trimer (*green*), and PGN network (*red*). OmpA and BLP are embedded within an asymmetric bilayer containing Ra LPS in the upper leaflet (*gray*), and a mixture of phospholipids in the lower leaflet (*yellow*). To see this figure in color, go online.

TABLE 1 Summary of Simulations Performed

System	OmpA	BLP	Temperature (K)	Duration (ns)	OmpA Contact with PGN
Control	no	yes	310	2 × 100	—
Control	no	yes	323	2 × 100	—
1	monomer	yes	310	2 × 100	yes
1	monomer	yes	323	2 × 100	yes
2	monomer	no	310	2 × 100	no
2	monomer	no	323	2 × 100	no
3	dimer	yes	310	2 × 100	yes
3	dimer	yes	323	2 × 100	yes
4	dimer	no	310	2 × 100	yes
4	dimer	no	323	2 × 100	yes

30 Å above a mesh of PGN network to observe the initial binding event. In the presence of BLP, the C terminus of one of the BLP trimer was covalently linked to a PGN peptide chain. For systems with both OmpA and BLP, they were separated by ~30 Å at the beginning of the simulations. Two independent simulations of each system, each for 100 ns, were performed at 310 and 323 K; the two temperatures were used as a means to enhance sampling with independent simulations. One simulation for the systems containing OmpA dimer (Systems 3 and 4) at 310 K was extended to 500 ns (Fig. S1), although we did not observe any significant changes after the first 100 ns. Therefore, for simplicity, all analyses were performed using the 100-ns simulations at the two temperatures mentioned in Table 1.

BLP tilts and kinks with respect to the membrane

At the beginning of the simulations, BLP was positioned at a right angle with respect to the plane of the membrane. Intriguingly, at the end of all simulations the BLP helices were observed to tilt, the degree of which was dependent on the presence of OmpA (Fig. 2 *A*). We measured the distribution of BLP tilt angle, and found that in the absence of OmpA, BLP stabilized at ~80°, whereas in the presence of OmpA monomer, the BLP tilted slightly more at ~75°. Adding OmpA dimer to the system resulted in BLP tilted even more at ~65°, suggesting that the BLP helices can flexibly adjust their orientation with respect to the plane of the membrane to adapt to other nearby proteins.

Inspecting the shape of the BLP trimer, we also found that each helix kinked with respect to its helical axis (Fig. 2 *B*). These helix kinks were most prominent on the C-terminal end of the BLP helices, which bent to ~30°. Although most helix kinks can be attributed to the presence of certain residues like proline and glycine (47–49), we could not find these residues on BLP. The position of the kink on each of the three helices also differed slightly, with helix 1 and 3 showing most bending around residue Y55, which was not the case with helix 2 (Fig. S2). Despite these different bending properties, BLP remains stable as a trimer throughout the entire simulations. Taken together, these

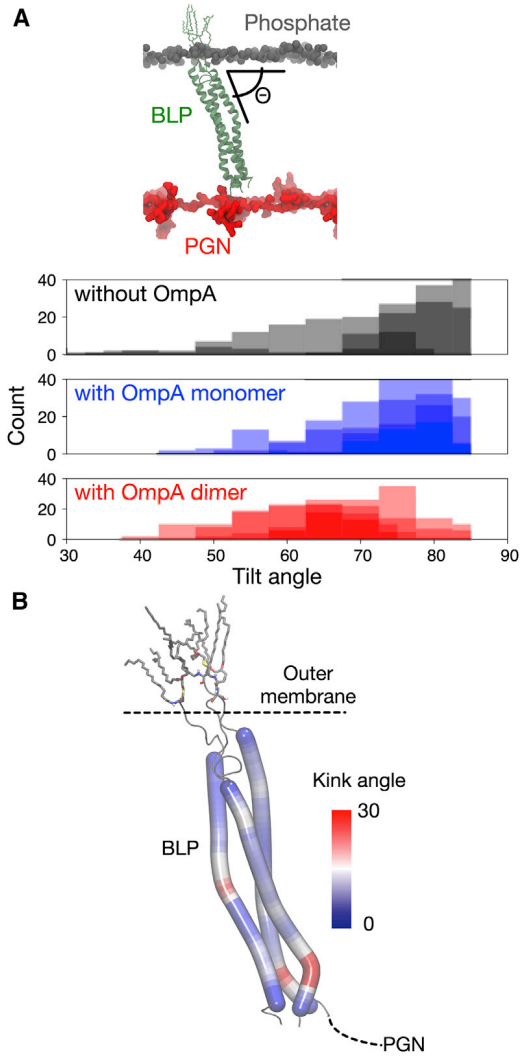


FIGURE 2 BLP tilting and bending. (*A*) Shown here is the distribution of BLP tilt angle throughout the 100-ns simulations for systems without OmpA (*black*), with OmpA monomer (*blue*), and with OmpA dimer (*red*). This is measured as the acute angle between the centers of geometry of the phosphorus atoms in the inner leaflet of the OM, the N-terminal residues of the BLP, and its C-terminal residues (illustrated at the *top of the graph*). Independent repeats were plotted separately, and a bin size of 5° was used. (*B*) BLP helix kink angles were calculated along each of the three helices using VMD Bendix plugin (54). The figure shows the final snapshot of BLP from one of the BLP-only simulations, colored based on the degree of helix kink. To see this figure in color, go online.

BLP helix kinks and tilts resulted in a lift of the PGN network closer to the OM, effectively closing the gap between PGN and OmpA CTD.

BLP facilitates interactions of OmpA monomer with PGN

Our previous simulations showed that the OmpA CTD in its monomeric state has a high propensity to interact with the OM (29). If such interaction occurs in vivo with PGN bound to the CTD, it would cause a severe distortion to the PGN

network. Missing from our previous model, however, is BLP, which can potentially maintain the PGN within a certain distance from the OM and therefore avoid such distortions. To understand how OmpA monomer behaves in the presence of BLP, we started our simulations with the OmpA CTD unbound from the PGN in the presence of one copy of the BLP trimer.

We found the distance between the OmpA CTD and the surface of the PGN network was reduced during equilibration simulations, and in all four subsequent independent production runs, the CTD contacted the PGN surface (Fig. 3, *A* and *B*). The BLP tilting and bending shifted the PGN layer toward the OmpA CTD, and concomitantly resulted in their interactions. The gap between the PGN layer and the OM was reduced to ~70 Å at the end of the simulations (Fig. 3 *C*). This upward movement of the PGN network therefore eased its interactions with the OmpA monomer. Initial examination of the contact interface pointed toward a group of basic and polar residues, specifically K294, Q295, and R296, on the OmpA CTD that interacted with the negatively charged glycan moieties. This implies that the initial binding of OmpA to PGN is mediated by a long-range electrostatic force.

We then repeated these analyses with a similar system without the BLP trimer. Interestingly, we found in all simulations the linker region connecting the OmpA NTD and CTD contracted relative to the original starting conformation, resulting in a binding to the lower leaflet of the OM instead of PGN (Fig. 4, *A* and *B*). In contrast to the simulations with BLP, the distance along the *z* axis between the PGN network and the OM remained at ~90 Å (Fig. 4 *C*). This observation agrees with Samsudin et al. (29,50), who demonstrated a similar linker contraction and membrane binding event of the OmpA monomer.

The simulations of Samsudin et al. (29,50) also revealed that the binding of OmpA CTD to the OM is mediated by residues 270–300, which intriguingly include the three key residues for initial interactions with PGN in the presence of the BLP. As this region is highly positively charged, we would expect it to form interactions with a negatively charged surface. This could either be the PGN layer underneath or the phosphate groups of the OM. Our simulations suggest that the latter is a more likely option in the absence of the BLP, perhaps due to the large energetic cost for the linker to extend and usher the CTD toward the underlying PGN network. The presence of BLP, however,

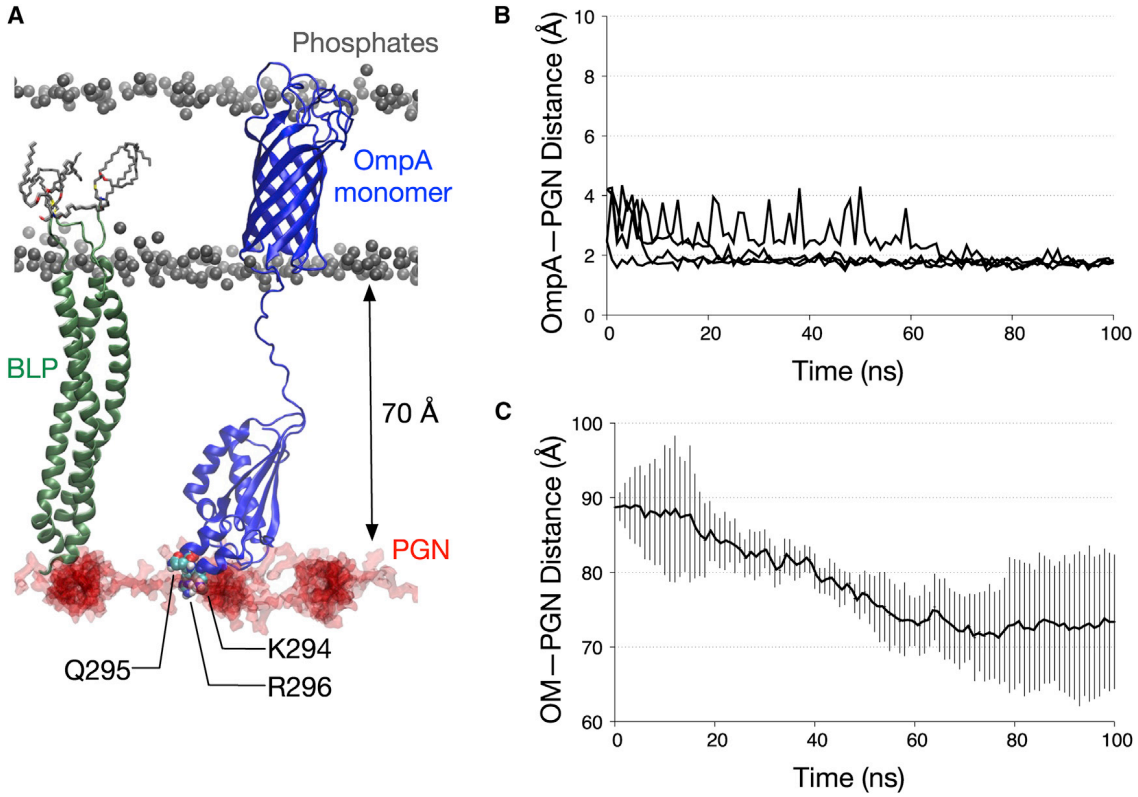


FIGURE 3 OmpA monomer interactions with PGN. (*A*) A snapshot depicts the end of one of the simulations of the OmpA monomer (*blue*) in the presence of BLP (*green*), highlighting the interactions of the CTD with the PGN network (*red*). Given here are residues involved in PGN contacts illustrated in van der Waals representation. (*B*) Shown here is the minimum distance between OmpA CTD and the PGN network for all four simulations of OmpA monomer with BLP. (*C*) Given here is the distance between the PGN network and the OM measured along the *z* axis between the centers of geometry of the PGN sugar strands and the phosphorus atoms on the lower leaflet of the OM. This is averaged over all four independent simulations and the error bars indicate SDs. To see this figure in color, go online.

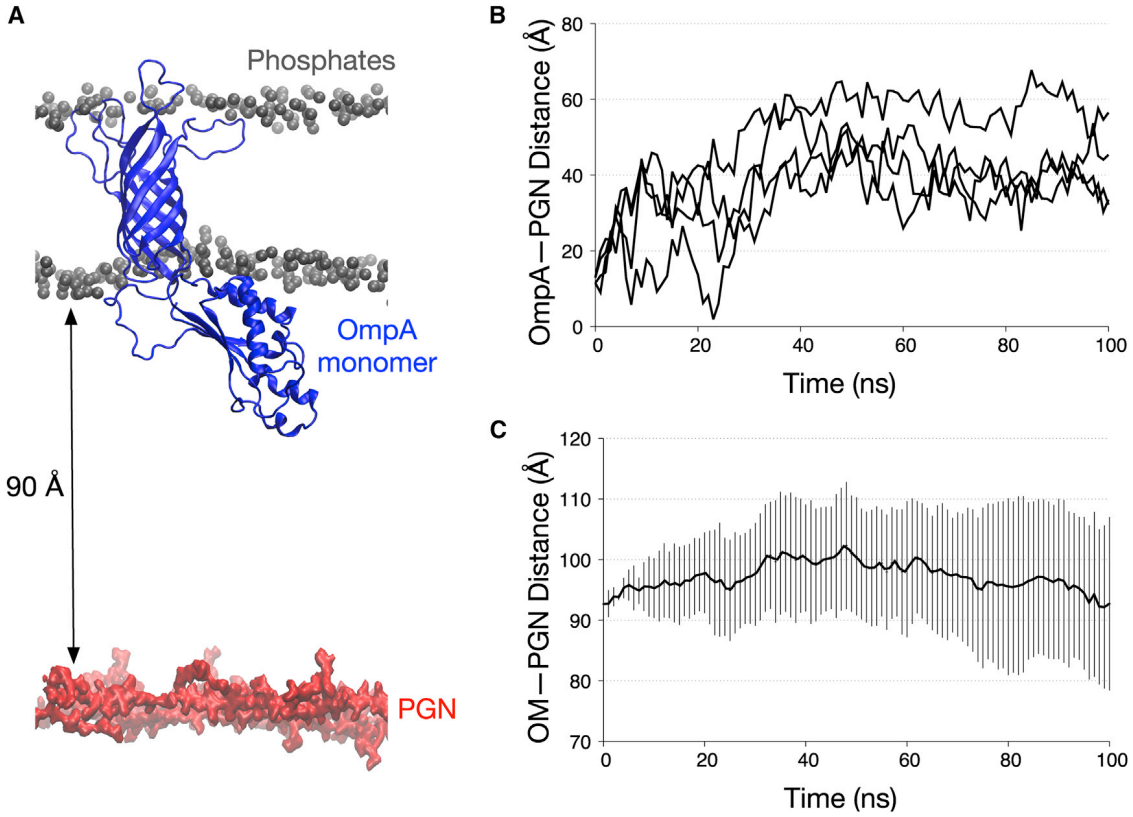


FIGURE 4 OmpA monomer interactions with the OM. (*A*) A snapshot depicts the end of one of the simulations of the OmpA monomer (*blue*) without BLP, highlighting the interactions of the CTD and the membrane (*gray spheres* represents the phosphorus atoms). (*B*) Minimum distance is given between OmpA CTD and the PGN network for all four simulations of OmpA monomer without BLP. (*C*) Distance between the PGN network and the OM is measured as described in Fig. 3, averaged over all four independent simulations. Error bars indicate SDs. To see this figure in color, go online.

lowers this energetic cost by reducing the distance between the PGN network and the CTD, making their interactions more likely to occur. Our data therefore illustrate how BLP facilitates initial binding of the OmpA monomer to the cell wall.

The OmpA dimer readily binds to PGN

Whereas earlier experimental and computational work often considered OmpA as a monomer (19–21), evidence from several recent studies suggests that full-length OmpA can form a homodimer (25–27). We therefore built a similar simulation system using a model of OmpA dimer proposed by Marcoux et al. (27). In contrast to the monomer simulations, we found that the OmpA dimer was able to bind the PGN network with or without BLP. Again, the unstructured linker connecting the NTD and the CTD played an essential role in initiating PGN binding. In simulations without BLP, the linker first extended by ~15 Å compared to its original length to reach the underlying PGN sheet (Fig. 5 *A*). Once binding occurred the linker then contracted, concomitantly shifting the PGN layer upwards by ~30 Å (Fig. 5 *B*). Although the entire PGN layer was lifted, the area around the CTD showed a more pronounced upward shift, resulting in a small undulatory pattern on the PGN surface (Fig. 5 *C*).

We have previously shown that the linker regions of the OmpA dimer in the absence of BLP are quite flexible, with the ability to extend and contract without disrupting the secondary structure of either the N- or C-terminal domains (28,29). In these simulations, comparing the systems with and without BLP, we found an intriguing difference. The presence of BLP reduced the amount of extension required by the linker to initiate interactions between OmpA CTD and PGN. In simulations with the BLP, the linker extended by only ~5 Å, compared to 15 Å in the OmpA-only simulation (Fig. S3). This is concordant with the ability of the BLP to lift the PGN networks upwards and reduce the distance between the latter and the OmpA CTD.

Next, we attempted to systematically elucidate the key residues involved in initial binding of OmpA and PGN by performing a contact analysis, combining data from simulations of both monomer and dimer (Fig. S4). Similar to the monomer simulations, a stretch of basic and polar residues, namely N203, K294, Q295, and R296, at the bottom of the CTD showed the highest degree of contacts. The latter three residues are part of a large insert only found

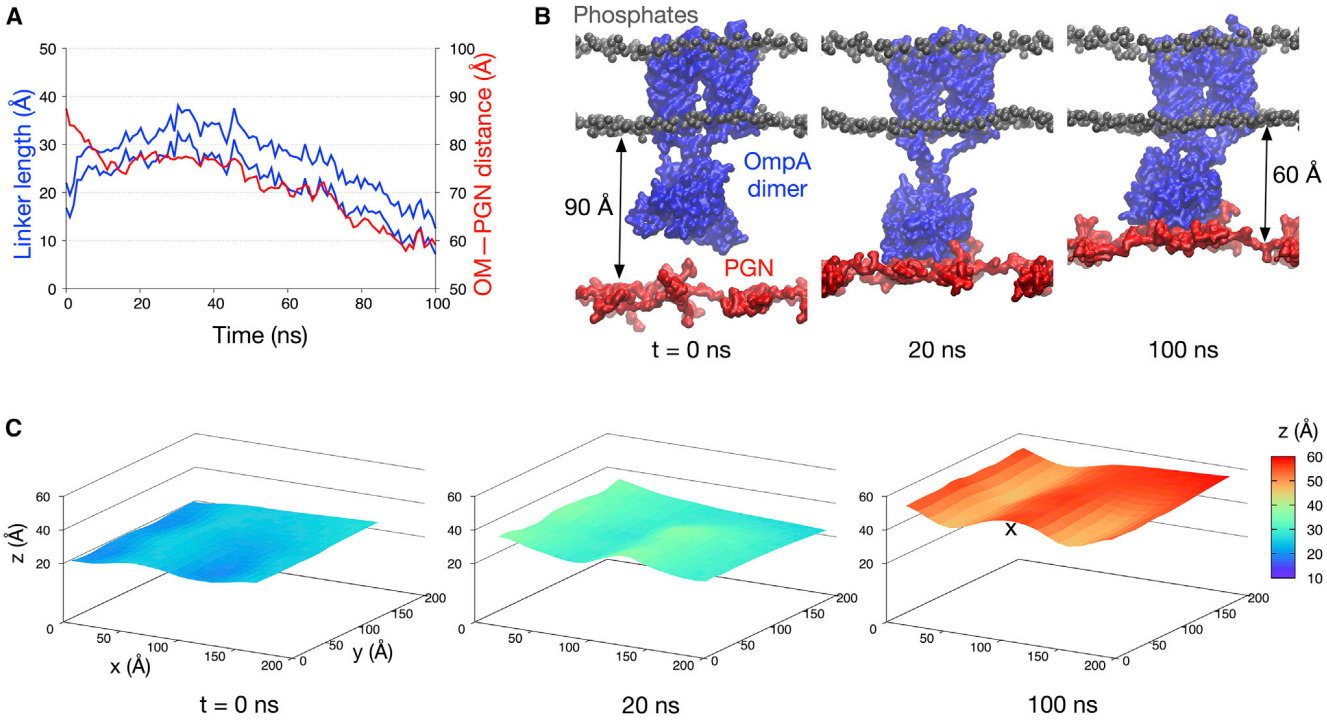


FIGURE 5 OmpA dimer interactions with PGN. (*A*) Shown here is the length of the unstructured linker connecting OmpA NTD and CTD (*blue*), plotted against the distance between the PGN network and the OM (*red*) for one of the simulations of OmpA dimer without BLP. The former is measured as previously described (28), whereas the latter is measured as in Fig. 3. (*B*) Given here are snapshots of this simulation at three different time points, highlighting the interactions between OmpA CTD (*blue*) and the PGN network (*red*). (*C*) The *z* coordinates of the PGN network are projected into a surface representation at these three time points to illustrate undulations observed during the simulations. The "x" indicates a local buckling effect induced by interactions with OmpA CTD. To see this figure in color, go online.

in OmpA from certain bacteria, which comprises the least stable region of the protein as shown by a NMR study (24) and previous computational simulations (29). The lysine and arginine residues are conserved in homologs from *Salmonella enterica* and *Neisseria meningitides* and, based on their crystal structures, are positioned similarly compared to the ones in *E. coli* (Fig. S5) (51), suggesting a potentially conserved PGN binding mechanism in these three species.

Decomposing the nonbonded energies of the OmpA and PGN interaction into their Coulombic and Lennard-Jones components revealed that the former contributes 10 times more than the latter (Fig. S6), further corroborating the role of electrostatic interactions for initial binding of OmpA to PGN. Mutations of the key residues (N203, K294, Q295, and R296) to alanine indeed altered the way OmpA interacted with PGN (Fig. S7). The timescale of interaction was longer with the mutant compared to wild-type (50 ns instead of 10 ns). Also, instead of forming a stable binding interface involving both subunits of the CTD, only one of them contacted the underlying PGN network, suggesting that these polar and basic residues play a key role in the initial binding process. That the OmpA dimer is able to bind PGN without BLP implies that the electrostatic force from these basic residues, which is stronger than the monomer due to the dimerization, is able to attract the oppositely charged PGN network over the timescale of these simulations.

OmpA CTD binds to the BLP

The periplasm is a crowded environment with myriad proteins surrounding the PGN cell wall. In addition to OmpA interacting with PGN, it is also likely that the ubiquitous BLP molecules make contact with OmpA. We therefore examined our simulation trajectories for such interactions and found that in three out of four dimer simulations and in one out of four monomer simulations, BLP interacted with OmpA. In all of these simulations, the BLP helices played a key role in contacting the OmpA CTD whereas the BLP lipid tails remained further away from the OmpA NTD (Fig. 6, *A* and *B*). This form of interactions involving only the helical part of the BLP and not the lipidated region was possible due to the tilted configuration adopted by the BLP. The higher frequency of interactions with the BLP of the OmpA dimer compared to the monomer is likely caused by the larger size of the former, which increased the possibility of OmpA to be within close proximity of the BLP.

Delineating the key residues for OmpA-BLP interactions was more challenging, due to the different ways the OmpA CTD contacts the BLP helices. In most simulations, however, BLP formed multiple salt bridges with residues on

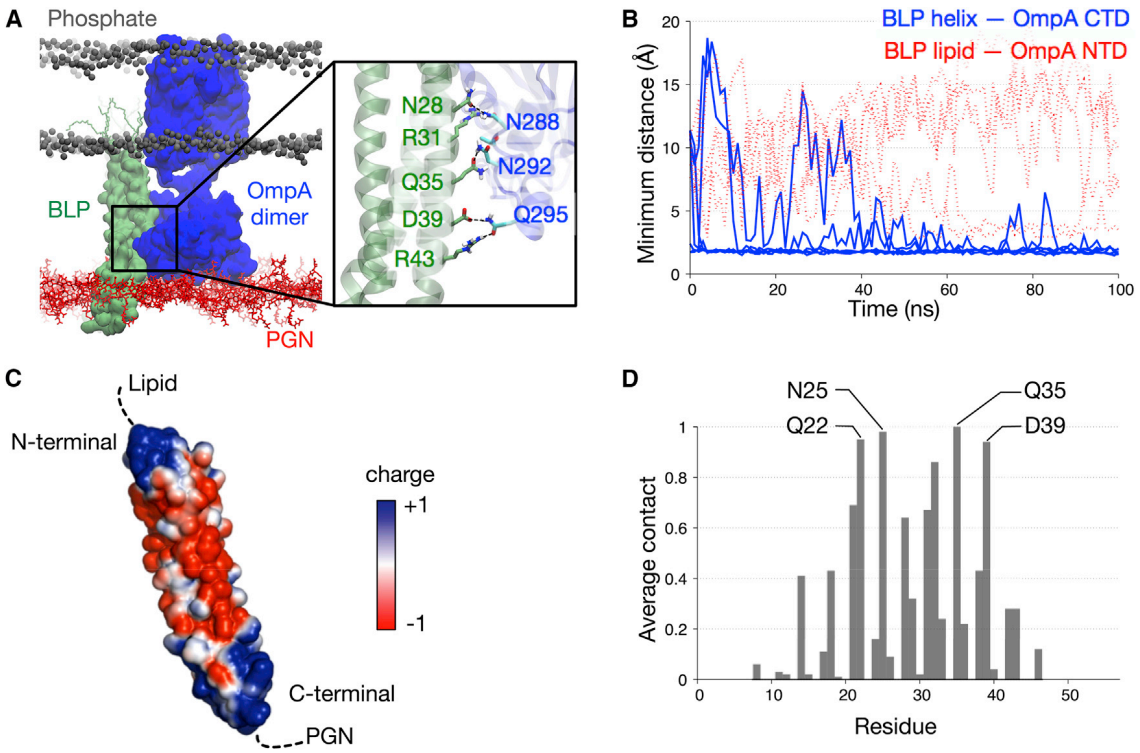


FIGURE 6 OmpA interactions with BLP. (*A*) A snapshot depicts the end of one of the simulations of OmpA dimer (*blue*) with BLP (*green*), highlighting their interactions. Enlarged image shows residues involved in these interactions from both proteins. (*B*) Minimum distance between the BLP helices and the OmpA CTD for all simulations of OmpA dimer is shown in blue, whereas the minimum distance between the BLP lipid tails and the OmpA NTD is shown in red dashed lines. (*C*) Electrostatic profile of the BLP is calculated using APBS (55) in PyMOL (18). (*D*) Shown here is contact analysis performed for each residue of the BLP averaged over all simulations where OmpA-BLP interactions were observed. A score of 1 indicates contacts throughout the entire 100-ns simulation. A distance cutoff of 4 Å was used for this analysis. To see this figure in color, go online.

OmpA. We therefore mapped the electrostatic profile of the BLP helices and found that the surface was indeed highly charged, whereby a group of basic residues clustered toward both the N- and C termini of the helices and the center of the helices was populated by acidic residues (Fig. 6 *C*). This suggests that most residues on the surface of the BLP helices are able to form electrostatic interactions with OmpA CTD. To corroborate this, we performed a contact analysis and found that there was no single prominent residue responsible for this interaction, but instead most of the polar and charged residues in the middle of the helices showed a high degree of contact with OmpA (Fig. 6 *D*). Our results therefore suggest that BLP and OmpA CTD are able to form nonspecific electrostatic interactions in the periplasm.

DISCUSSION

We have constructed an atomistic model of the *E. coli* OM bound to a network of PGN molecules via both noncovalent interactions with the outer membrane porin OmpA, and covalent linkage with the BLP. Our simulations uncover important insights into the initial binding of OmpA with the PGN cell wall and the role of BLP in facilitating these interactions. OmpA has been shown to bind PGN in a labile manner (29), and therefore it is likely that the CTD is in equilibrium between PGN-bound and PGN-unbound states. We demonstrated that from the unbound state, BLP helps the binding of OmpA monomer to the PGN network by lifting the latter closer to the former. Without BLP, the CTD of the OmpA monomer binds to the lower leaflet of the OM instead on the timescale of the simulations presented here. OmpA homodimer, on the other hand, can readily contact the PGN network by extending the linker between the NTD and the CTD, even in the absence of BLP. For both the monomer and dimer, binding is mediated by electrostatic interactions via several basic and polar residues, which are conserved in two other OmpA homologs from *S. enterica* and *N. meningitides* (51). Intriguingly these residues, and the mobile insert in which they are found, are not conserved in species like *A. baumannii* (30), and are also absent in other OmpA-like domains such as the *E. coli* PGN-associated lipoproteins (52,53) and MotB (17). This suggests that whereas PGN interactions in the binding pocket of OmpA-like domains are conserved across species (29,30), the initial interactions are likely to differ owing to the different residues found on the surface of these proteins. We acknowledge that in our simulations, PGN did not make any significant contact with the two key residues in the binding pocket (30), D241 and R256, most likely due to the relatively short timescale of the simulations. We note here that

much longer simulations or enhanced sampling methods are needed to allow the peptide chain on PGN to enter the binding pocket and form stable interactions with these residues.

To date, OmpA is the only integral membrane eight-stranded β-barrel protein that coexists as both monomers and dimers (26). Although a model of the full-length homodimer has been proposed by mass spectrometry (27), and has been shown to interact with PGN from simulation studies (29) the physiological role of dimerization is still to be confirmed. That the dimeric interface was localized within the CTD (26,27) suggests a functional importance of dimerization to the role of the CTD and its interactions with the cell wall. A homolog of OmpA CTD from *N. meningitides*, RmpM, may also exist as a dimer, as indicated by both the crystal structures and solution experiments, which the authors suggest would promote more efficient binding to PGN (51). Indeed, our OmpA dimer simulations revealed that the CTD was able to form a stable interaction with the underlying PGN layer even in the absence of the BLP. We conjecture that this is caused by the dimerization increasing the negatively charged surface area at the bottom of the CTD, and thereby strengthening the electrostatic attraction toward the PGN network. In its monomeric form, the OmpA CTD has only half as many basic residues in this region. This weaker electrostatic attraction is inadequate to surpass the energetic penalty of extending the linker connecting the NTD and the CTD, which therefore leads to contraction of the linker and subsequently interaction of the CTD and the OM. Taken together, our simulations suggest that OmpA dimerization increases the possibility of initial contact with the PGN cell wall, and therefore directly contributes toward maintaining the integrity of the cell envelope.

The BLP is one of the most abundant proteins in Gram-negative bacteria, whereby $\sim 7.2 \times 10^5$ molecules are found within each cell (10). Similarly, OmpA is one of the most ubiquitous outer membrane porins in *E. coli* (18). As both proteins play a critical structural role in preserving the robustness of the PGN cell wall, it is highly likely that BLP and OmpA function cooperatively. We show in this study that indeed BLP is required for OmpA to interact with the PGN cell wall in its monomeric state. In regions without BLP, OmpA forms homodimers to maintain this interaction.

We rationalize the differences between the behavior of OmpA monomer and dimers in the absence of BLP as follows: there is a fine balance between the energy required to extend the OmpA linker regions and the favorable electrostatic interactions formed between the C-terminal domain and PGN. When only the monomer is present, the linker will not extend; instead, the C-terminal domain forms electrostatic interactions with the lower leaflet of the outer membrane. In the case of the dimer, the combined C-terminal domains of the two monomers now provide a larger area for electrostatic interaction with PGN, and this gain in electrostatic interactions is sufficient to overcome the energy barrier required to extend the linkers. This is further augmented by the steric hindrance imposed by the dimerization interface toward interaction with the lower leaflet of the outer membrane (50).

Crucially, our results uncover some important insights into the interplay between the molecular components of the Gram-negative bacterial cell envelope, toward a better structure-function understanding of the barrier protecting the bacteria from antibiotics.

SUPPORTING MATERIAL

Seven figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(17)30869-X.

AUTHOR CONTRIBUTIONS

S.K. and F.S. designed the research and wrote the paper. F.S. and A.B. performed the research. T.J.P. developed the parameters for BLP attachment to the Om and helped to write portions of the paper.

ACKNOWLEDGMENTS

We acknowledge use of the Iridis 3 and 4 High Performance Computing Facilities at Southampton.

F.S. was supported by UK Biotechnology and Biological Sciences Research Council grant No. BB/M029573/1. This work was supported by NAMRIP, University of Southampton.

REFERENCES

1. Silhavy, T. J., D. Kahne, and S. Walker. 2010. The bacterial cell envelope. *Cold Spring Harb. Perspect. Biol.* 2:a000414.
2. Lugtenberg, E. J., and R. Peters. 1976. Distribution of lipids in cytoplasmic and outer membranes of *Escherichia coli* K12. *Biochim. Biophys. Acta.* 441:38–47.
3. Tsirigos, K. D., P. G. Bagos, and S. J. Hamodrakas. 2011. OMPdb: a database of β-barrel outer membrane proteins from Gram-negative bacteria. *Nucleic Acids Res.* 39:D324–D331.
4. Vollmer, W., and U. Bertsche. 2008. Murein (peptidoglycan) structure, architecture and biosynthesis in *Escherichia coli*. *Biochim. Biophys. Acta.* 1778:1714–1734.
5. Vollmer, W., D. Blanot, and M. A. de Pedro. 2008. Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.* 32:149–167.
6. Vollmer, W., J. V. Höltje, and J. Ho. 2004. The architecture of the murein (peptidoglycan) in Gram-negative bacteria: vertical scaffold or horizontal layer(s)? *J. Bacteriol.* 186:5978–5987.
7. Braun, V. 1975. Covalent lipoprotein from the outer membrane of *Escherichia coli*. *Biochim. Biophys. Acta.* 415:335–377.
8. Braun, V., and H. Wolff. 1970. The murein-lipoprotein linkage in the cell wall of *Escherichia coli*. *Eur. J. Biochem.* 14:387–391.
9. Bosch, V., and V. Braun. 1973. Distribution of murein-lipoprotein between the cytoplasmic and outer membrane of *Escherichia coli*. *FEBS Lett.* 34:307–310.
10. Braun, V., and U. Sieglin. 1970. The covalent murein-lipoprotein structure of the *Escherichia coli* cell wall: the attachment site of the lipoprotein on the murein. *Eur. J. B*iochem. 13:336–346.
11. Inouye, M., J. Shaw, and C. Shen. 1972. The assembly of a structural lipoprotein in the envelope of *Escherichia coli*. *J. Biol. Chem.* 247:8154–8159.

12. Lee, N., and M. Inouye. 1974. Outer membrane proteins of *Escherichia coli*: biosynthesis and assembly. *FEBS Lett.* 39:167–170.
13. Shu, W., J. Liu, …, M. Lu. 2000. Core structure of the outer membrane lipoprotein from *Escherichia coli* at 1.9 Å resolution. *J. Mol. Biol.* 299:1101–1112.
14. Cohen, E. J., J. L. Ferreira, …, K. T. Hughes. 2017. Nanoscale-length control of the flagellar driveshaft requires hitting the tethered outer membrane. *Science.* 356:197–200.
15. Koebnik, R. 1995. Proposal for a peptidoglycan-associating α-helical motif in the C-terminal regions of some bacterial cell-surface proteins. *Mol. Microbiol.* 16:1269–1270.
16. Parsons, L. M., F. Lin, and J. Orban. 2006. Peptidoglycan recognition by Pal, an outer membrane lipoprotein. *Biochemistry.* 45:2122–2128.
17. Roujeinikova, A. 2008. Crystal structure of the cell wall anchor domain of MotB, a stator component of the bacterial flagellar motor: implications for peptidoglycan recognition. *Proc. Natl. Acad. Sci. USA.* 105:10348–10353.
18. Smith, S. G. J., V. Mahon, …, R. P. Fagan. 2007. A molecular Swiss army knife: OmpA structure, function and expression. *FEMS Microbiol. Lett.* 273:1–11.
19. Arora, A., D. Rinehart, …, L. K. Tamm. 2000. Refolded outer membrane protein A of *Escherichia coli* forms ion channels with two conductance states in planar lipid bilayers. *J. Biol. Chem.* 275:1594–1600.
20. Arora, A., F. Abildgaard, …, L. K. Tamm. 2001. Structure of outer membrane protein A transmembrane domain by NMR spectroscopy. *Nat. Struct. Biol.* 8:334–338.
21. Bond, P. J., J. D. Faraldo-Gómez, and M. S. P. Sansom. 2002. OmpA: a pore or not a pore? Simulation and modeling studies. *Biophys. J.* 83:763–775.
22. Khalid, S., P. J. Bond, …, M. S. Sansom. 2008. OmpA: gating and dynamics via molecular dynamics simulations. *Biochim. Biophys. Acta.* 1778:1871–1880.
23. Pautsch, A., and G. E. Schulz. 1998. Structure of the outer membrane protein A transmembrane domain. *Nat. Struct. Biol.* 5:1013–1017.
24. Ishida, H., A. Garcia-Herrero, and H. J. Vogel. 2014. The periplasmic domain of *Escherichia coli* outer membrane protein A can undergo a localized temperature dependent structural transition. *Biochim. Biophys. Acta.* 1838:3014–3024.
25. Stenberg, F., P. Chovanec, …, D. O. Daley. 2005. Protein complexes of the *Escherichia coli* cell envelope. *J. Biol. Chem.* 280:34409–34419.
26. Zheng, C., L. Yang, …, J. E. Bruce. 2011. Cross-linking measurements of in vivo protein complex topologies. *Mol. Cell Proteomics.* 10, M110. 006841.
27. Marcoux, J., A. Politis, …, C. V. Robinson. 2014. Mass spectrometry defines the C-terminal dimerization domain and enables modeling of the structure of full-length OmpA. *Structure.* 22:781–790.
28. Ortiz-Suarez, M. L., F. Samsudin, …, S. Khalid. 2016. Full-length OmpA: structure, function, and membrane interactions predicted by molecular dynamics simulations. *Biophys. J.* 111:1692–1702.
29. Samsudin, F., M. L. Ortiz-Suarez, …, S. Khalid. 2016. OmpA: a flexible clamp for bacterial cell wall attachment. *Structure.* 24:2227–2235.
30. Park, J. S., W. C. Lee, …, H. Y. Kim. 2012. Mechanism of anchoring of OmpA protein to the cell wall peptidoglycan of the Gram-negative bacterial outer membrane. *FASEB J.* 26:219–228.
31. Appelmelk, B. J., Y. Q. An, …, J. de Graaf. 1994. Frequencies of lipopolysaccharide core types in *Escherichia coli* strains from bacteraemic patients. *Microbiology.* 140:1119–1124.
32. Vinogradov, E. V., K. van der Drift, …, O. Holst. 1999. The structures of the carbohydrate backbones of the lipopolysaccharides from *Escherichia coli* rough mutants F470 (R1 core type) and F576 (R2 core type). *Eur. J. Biochem.* 261:629–639.
33. Aibara, S., M. Kato, …, M. Kito. 1972. Changes in positional distribution of fatty acids in the phospholipids of *Escherichia coli* after shift-down in temperature. *Biochim. Biophys. Acta.* 270:301–306.
34. Kito, M., M. Ishinaga, …, S. Sawada. 1975. Metabolism of the phosphatidylglycerol molecular species in *Escherichia coli*. *Eur. J. Biochem.* 54:55–63.
35. Yokota, K., R. Kanamoto, and M. Kito. 1980. Composition of cardiolipin molecular species in *Escherichia coli*. *J. Bacteriol.* 141:1047–1051.
36. Piggot, T. J., D. A. Holdbrook, and S. Khalid. 2011. Electroporation of the *E. coli* and *S. aureus* membranes: molecular dynamics simulations of complex bacterial membranes. *J. Phys. Chem.* 115:13381–13388.
37. Wolf, M. G., M. Hoefling, …, G. Groenhof. 2010. g_membed: efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation. *J. Comput. Chem.* 31:2169–2174.
38. DeLano, W. 2002. The PyMOL Molecular Graphics System. http://www.pymol.org.
39. Schmid, N., A. P. Eichenberger, …, W. F. van Gunsteren. 2011. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* 40:843–856.
40. Horta, B. A. C., P. F. Fuchs, …, P. H. Hünenberger. 2011. New interaction parameters for oxygen compounds in the GROMOS force field: improved pure-liquid and solvation properties for alcohols, ethers, aldehydes, ketones, carboxylic acids, and esters. *J. Chem. Theory Comput.* 7:1016–1031.
41. Abraham, M. J., M. Teemu, …, L. Erik. 2015. Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX.* 1–2:19–25.
42. Berendsen, H. J. C., J. P. M. Postma, …, J. Hermans. 1981. Interaction models for water in relation to protein hydration. *In* Intermolecular Forces. B. Pullman, editor. Reidel, Boston, MA, pp. 331–342.
43. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101.
44. Parrinello, M., and A. Rahman. 1981. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52:7182–7190.
45. Hess, B., H. Bekker, …, J. G. E. M. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.
46. Essmann, U., L. Perera, …, L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.
47. Barlow, D. J., and J. M. Thornton. 1988. Helix geometry in proteins. *J. Mol. Biol.* 201:601–619.
48. Deville, J., J. Rey, and M. Chabbert. 2008. Comprehensive analysis of the helix-X-helix motif in soluble proteins. *Proteins.* 72:115–135.
49. Langelaan, D. N., M. Wieczorek, …, J. K. Rainey. 2010. Improved helix and kink characterization in membrane proteins allows evaluation of kink sequence predictors. *J. Chem. Inf. Model.* 50:2213–2220.
50. Samsudin, F., J. L. Parker, …, P. W. Fowler. 2015. Accurate prediction of ligand affinities for a peptide transporter. *Cell Chem. Biol.* 23: 299–309.
51. Grizot, S., and S. K. Buchanan. 2004. Structure of the OmpA-like domain of RmpM from *Neisseria meningitidis*. *Mol. Microbiol.* 51:1027–1037.
52. Abergel, C., A. Walburger, …, C. Lazdunski. 2001. Crystallization and preliminary crystallographic study of the peptidoglycan-associated lipoprotein from *Escherichia coli*. *Acta Crystallogr. D Biol. Crystallogr.* 57:317–319.
53. Gourlay, L. J., C. Peri, …, M. Bolognesi. 2013. Exploiting the *Burkholderia pseudomallei* acute phase antigen BPSL2765 for structure-based epitope discovery/design in structural vaccinology. *Chem. Biol.* 20: 1147–1156.
54. Dahl, A. C. E., M. Chavent, and M. S. P. Sansom. 2012. Bendix: intuitive helix geometry analysis and abstraction. *Bioinformatics.* 28:2193–2194.
55. Baker, N. A., D. Sept, …, J. A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA.* 98:10037–10041.

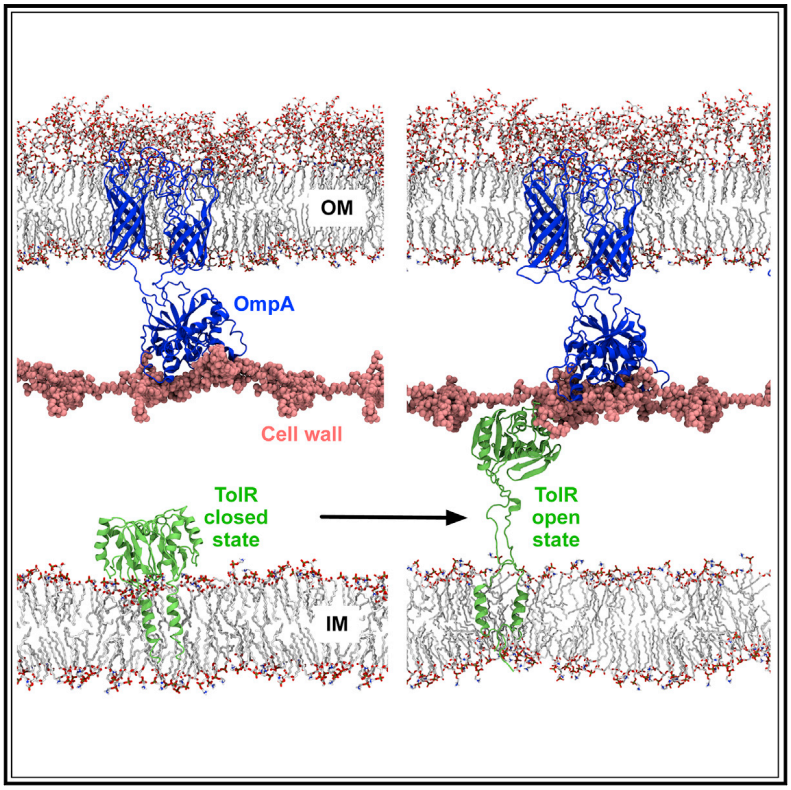
7.4 Appendix D

Theory

Structure

Binding from Both Sides: TolR and Full-Length OmpA Bind and Maintain the Local Structure of the *E. coli* Cell Wall

Graphical Abstract



Authors

Alister T. Boags, Firdaus Samsudin, Syma Khalid

Correspondence

s.khalid@soton.ac.uk

In Brief

By performing atomistic simulations of a gram-negative bacterial cell envelope, Boags et al. reveal the importance of protein binding from both the inner and the outer membrane sides in preserving the structural integrity of the peptidoglycan cell wall.

Highlights

- TolR binds peptidoglycan cell wall via electrostatic interactions
- Binding of TolR and OmpA keeps the cell wall flat
- Closed-state TolR does not interact with the cell wall
- Cell-wall-binding residues of TolR conserved across species

Boags et al., 2019, Structure *27*, 1–12
April 2, 2019
https://doi.org/10.1016/j.str.2019.01.001

CellPress

Please cite this article in press as: Boags et al., Binding from Both Sides: TolR and Full-Length OmpA Bind and Maintain the Local Structure of the *E. coli* Cell Wall, Structure (2019), https://doi.org/10.1016/j.str.2019.01.001

Structure
Theory

Binding from Both Sides: TolR and Full-Length OmpA Bind and Maintain the Local Structure of the *E. coli* Cell Wall

Alister T. Boags,[1,2] Firdaus Samsudin,[1,2] and Syma Khalid[1,3,*]
[1]School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, UK
[2]These authors contributed equally
[3]Lead Contact
*Correspondence: s.khalid@soton.ac.uk
https://doi.org/10.1016/j.str.2019.01.001

SUMMARY

We present a molecular modeling and simulation study of the *E. coli* cell envelope, with a particular focus on the role of TolR, a native protein of the *E. coli* inner membrane, in interactions with the cell wall. TolR has been proposed to bind to peptidoglycan, but the only structure of this protein thus far is in a conformation in which the putative peptidoglycan binding domain is not accessible. We show that a model of the extended conformation of the protein in which this domain is exposed binds peptidoglycan largely through electrostatic interactions. Non-covalent interactions of TolR and OmpA with the cell wall, from the inner membrane and outer membrane sides, respectively, maintain the position of the cell wall even in the absence of Braun's lipoprotein. The charged residues that mediate the cell-wall interactions of TolR in our simulations are conserved across a number of species of gram-negative bacteria.

INTRODUCTION

Gram-negative bacteria such as *E. coli* have a complex cell envelope, which protects the cell and controls influx/efflux of molecular species to ensure the normal functioning of the cell (Nikaido, 2003). The cell envelope contains an aqueous region known as the periplasm, which is sandwiched between an asymmetrical outer membrane and a symmetrical inner membrane. Contained within the periplasm is the cell wall, which is composed of a sugar-peptide polymer known as peptidoglycan (PGN) (Vollmer and Bertsche, 2008). The periplasm is host to many different proteins that are essential for the healthy growth and proliferation of gram-negative bacteria. These proteins are known to be freely moving, associated with the inner or outer membrane, or bound to the cell wall. The interactions of these proteins with one another both (1) laterally, in other words within one membrane or the periplasm, and (2) across regions, e.g., extending from one membrane to the periplasm, are important in maintaining the structural integrity and correct functioning of the cell envelope.

A number of different proteins have been shown to play a role in cross-region interactions, and others have been hypothesized to do so. Braun's lipoprotein (BLP; also known as "Lpp" and "murein lipoprotein") is an abundant protein that is lipidated at its N-terminal domain, which anchors it to the outer membrane (Braun, 1975). It is the only known protein in *E. coli* to be covalently attached to the PGN of the cell wall. It exists in two states: ~33% of the lipoprotein is covalently bound to the cell wall via a peptide bond, and ~66% is free in the periplasm. BLP is proposed to have a primarily structural function, essentially acting as a staple between the outer membrane and the PGN, which serves to maintain the required distance between the cell wall and the outer membrane (Miller and Salama, 2018). Cells that lack BLP or that have reduced amounts of BLP are viable, but they have been shown to release outer-membrane vesicles at a higher rate than normal and also suffer from cellular leakage (Schwechheimer et al., 2014; Asmar and Collet, 2018). Non-covalent interactions between the cell wall and the outer membrane are mediated through proteins such as PAL and OmpA (Parsons et al., 2006; Park et al., 2012). The latter is composed of two domains, the N-terminal domain, which is an eight-stranded β barrel that is connected via a flexible linker to the soluble C-terminal domain, which contains the PGN-binding region (Carpenter et al., 2007; Marcoux et al., 2014). We have previously shown that OmpA in its dimeric form can extend its linker region such that the C-terminal domain is able to form long-lasting interactions with PGN even in the absence of BLP, while BLP facilitates PGN binding of the monomer (Samsudin et al., 2017). We showed that BLP can tilt within the periplasm to provide some variation in the PGN-outer membrane distance.

Interactions of inner membrane proteins with the cell wall are less well understood at the molecular level than their outer-membrane counterparts. Three proteins from the Tol family, TolQ, TolR, and TolA, interact with one another via their transmembrane domains within the inner membrane (Gerding et al., 2007). TolR is proposed to interact with the cell wall in a manner similar to that of OmpA. While binding of TolR to PGN has been demonstrated, the X-ray structure of TolR from *E. coli* is of the protein in its compact form, in which the putative PGN-binding domain is not surface exposed (Wojdyla et al., 2015). Based on the X-ray structure of the closed state and biophysical and computational studies of TolR, Kleanthous and co-workers proposed a large-scale proton-motive force (PMF)-dependent conformational rearrangement in which extension of the TolR

Table 1. Summary of Simulations

OmpA Structure	TolR Structure	BLP	Membrane	Number of Atoms	Binds to PGN	Simulation Length (ns)
None	open	none	inner only	135,718	yes	3 × 200
Wild-type	open	none	inner and outer	207,787	yes	3 × 200
Wild-type	open	yes	inner and outer	275,948	yes	3 × 200
1-Truncated	open	yes	inner and outer	238,293	yes	3 × 200
Truncated	open	yes	inner and outer	236,933	yes	4 × 200
Truncated	closed	yes	inner and outer	236,789	no	3 × 200

linker enables the protein to contact the cell wall and exposure of the PGN-binding domain enables it to bind PGN in a manner similar to the structural alterations proposed for the bacterial flagellar protein MotB (Wojdyla et al., 2015). The model of the protein in this conformation was termed the "open state." The hypothesis of large-scale rearrangement is difficult to test experimentally in the absence of structures of the different states of the proteins. However, simulations offer a route to predict the behavior of the model under different scenarios.

In the last decade or so, molecular dynamics studies of the cell envelopes of gram-negative bacteria have moved beyond simple phospholipid representations of both envelopes, to incorporate the natural biochemical diversity of the lipidic components of these membranes, at both atomistic (Kirschner et al., 2012; Piggot et al., 2011a; Wu et al., 2013) and coarse-grain resolution (Hsu et al., 2016; Ma et al., 2015). Much of the setup of such systems is facilitated by tools such MARTINI-MAKER (Hsu et al., 2017). Furthermore, detailed atomistic models of the cell wall have recently emerged, too, enabling study of the biophysical properties of PGN (Gumbart et al., 2014; Hwang et al., 2018) and interactions with proteins (Samsudin et al., 2016). To test the model of the open state of TolR and compare the PGN-binding mode with that of OmpA, in the following we present an atomistic molecular dynamics and modeling study of TolR, OmpA, and BLP in a model of the cell envelope that includes both membranes and the cell wall. We note here that the structures of TolQ and TolA are not known, neither are the precise details of the way they are arranged with respect to each other and TolR; thus TolA and TolQ are omitted from the present studies. We show that the model of the open state of TolR binds PGN primarily through electrostatic interactions, whereas the closed state does not bind PGN. In the presence of full-length OmpA dimers in the outer membrane and open-state TolR in the inner membrane, the location of the cell wall is maintained between these proteins. The binding of both proteins to the cell wall also alleviates local surface distortions that are observed when only one of OmpA or TolR is bound. In contrast, if OmpA is truncated to its N-terminal domain and BLP is added to the system, then

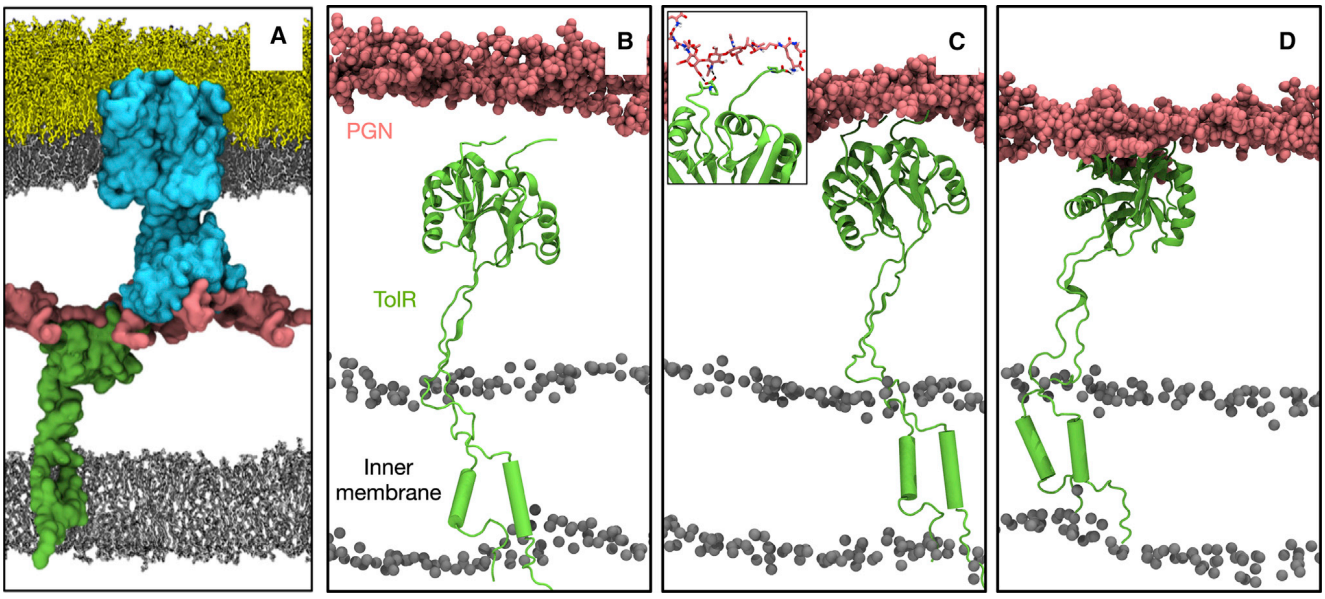


Figure 1. Initial Binding of TolR to the Cell Wall

(A) Simulation system (OmpA dimer is cyan, PGN is pink, TolR is green, phospholipids are gray, and lipopolysaccharide is yellow, water and ions are omitted for clarity).

(B–D) Mechanism of TolR binding to the cell wall. (B) TolR in an open state whereby the periplasmic domain is free to bind PGN. (C) The flexible C terminus of TolR snorkels toward PGN and the carboxyl groups interact with positively charged moieties on PGN. Inset shows example of these interactions (described in detail in Figure 3). (D) The rest of the TolR periplasmic domain binds to PGN. The linker between the periplasmic domain and the N-terminal helices is able to contract to pull the cell wall to the inner membrane.

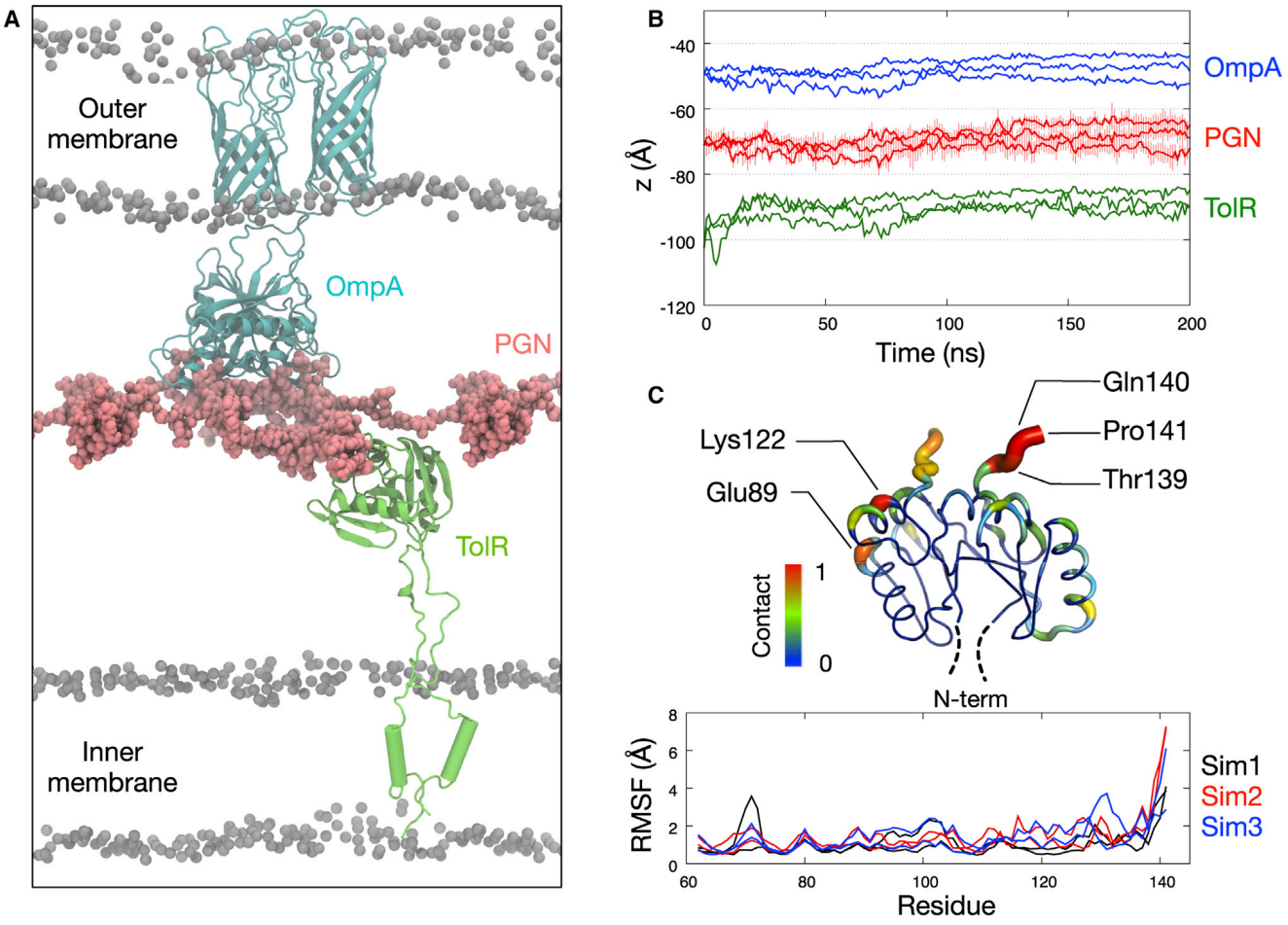


Figure 2. Simulation of TolR with the PGN Cell Wall and OmpA

(A) A snapshot from the end of a 200 ns simulation.

(B) The center of mass motion along the z axis of the OmpA C-terminal domain (residues 189–316), PGN cell wall, and TolR periplasmic domain (residues 62–141). Data from three independent simulations are shown. Error bars for the PGN plot indicate standard deviations from three adjacent strands.

(C) Root-mean-square fluctuation (RMSF) of the TolR periplasmic domain and the degree of contact each residue made with the cell wall. Two lines in the RMSF plot for each simulation indicate values from two TolR protomers. Contact analysis performed using a distance cutoff of 4 Å and a contact value of 1 indicates one interaction throughout all the simulations. Residues that made significant contacts are labeled.

the TolR linker is able to contract and, in doing so, ''pulls'' the cell wall down toward the inner membrane until BLP is fully stretched and further movement is not possible.

RESULTS

For ease of interpretation of the results the simulations described below are summarized in Table 1. The simulations of TolR and OmpA were performed with a monolayered cell wall. The reason for this is that, from test simulations of one to three layers of cell wall, we observed the thickness of three layers to be 90–100 Å, two layers to be 60–70 Å, and a single layer to be ~30 Å (Figure S1), and given that the proposed thickness of PGN in *E. coli* is 20–70 Å (Matias et al., 2003; Turner et al., 2013), a single layer was chosen.

Initial Binding of Open-State Model of TolR and Full-Length OmpA to Peptidoglycan

In simulations of TolR and full-length OmpA, both proteins were initially positioned either directly in contact with or close to the cell wall. Specifically, the OmpA-PGN complex was taken from our previous work and the TolR was positioned with the transmembrane helices embedded in the inner membrane, and the periplasmic domain was not in contact with the cell wall (Samsudin et al., 2017). The shortest distance between the TolR periplasmic domain and the cell wall was around 5 Å at the start of the simulation. This system configuration gives a periplasmic space width of around 170 Å (experimental estimates of the width vary between 100 and 250 Å; Graham et al., 1991; Vollmer and Seligman, 2010). The periplasmic domain of TolR in the open conformation was structurally stable in all simulations and showed similar root-mean-square deviation (RMSD) progressions compared with the C-terminal domain of OmpA (Figure S2). The secondary structure of this domain was also largely preserved during the simulations. In this state TolR has a long unstructured loop in the C terminus, which remained mobile throughout the simulations.

In all simulations of the wild-type TolR in the open state, binding to PGN was observed. The mechanism of cell-wall binding

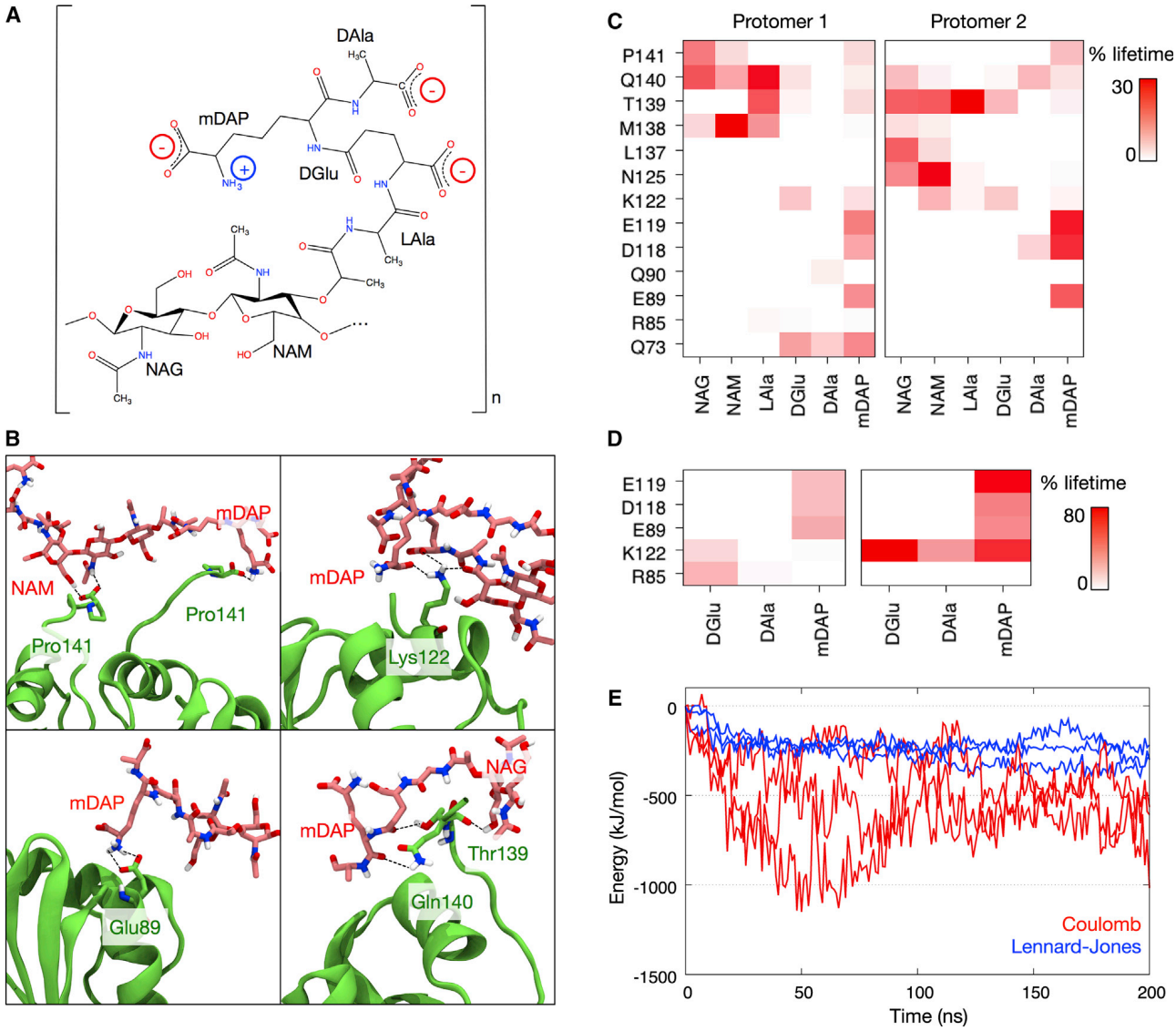


Figure 3. TolR Binding to the Cell Wall Is Driven by Electrostatic Interactions

(A) Chemical structure of the peptidoglycan cell-wall repeating units with the positions of positive and negative charges highlighted.

(B) Four snapshots showing examples of interactions between residues depicted in Figure 2C to diaminopimelate (mDAP), N-acetylmuramic acid (NAM), and N-acetylglucosamine (NAG).

(C) Average lifetime of hydrogen bonds formed between residues on both TolR protomers and the different moieties of the cell wall as a percentage of the simulation time.

(D) Average lifetime of salt bridges between TolR residues and charged moieties of the cell wall.

(E) Binding energy of TolR to the cell wall decomposed into its coulombic and Lennard-Jones components. Data were obtained from three independent simulations.

proceeded as follows: the proline (Pro141) residues at the C terminus of TolR consistently formed the first contact with PGN via its carboxyl group that interacted with either the positively charged amine group of diaminopimelate (mDAP) or the polar amide and hydroxyl moieties in adjacent sugars (Figures 1 and S3). This was immediately followed by interactions with downstream polar residues (Thr139 and Gln140). The greater flexibility of the unstructured C-terminal loop enabled this initial binding process as these residues were able to "snorkel" toward the PGN. Glu89 and Lys122, found in the more rigid globular domain of TolR, strengthened this binding; the former interacted with hydroxyl groups in N-acetylmuramic acid, while the latter formed a salt bridge with the C terminus of mDAP.

Within about 10 ns, in each simulation, the TolR linker was extended such that the periplasmic domain was in contact with PGN; in other words, PGN binding had occurred. Across all independent repeat simulations, after 200 ns, the cell wall was located about 20 Å (along the z direction, perpendicular to the plane of the membrane) from each protein, reaching a stable position after ~100 ns of simulation (Figures 2A and 2B). We extended two of these simulations to 500 ns; the binding of OmpA and TolR to the cell wall was maintained (Figure S4).

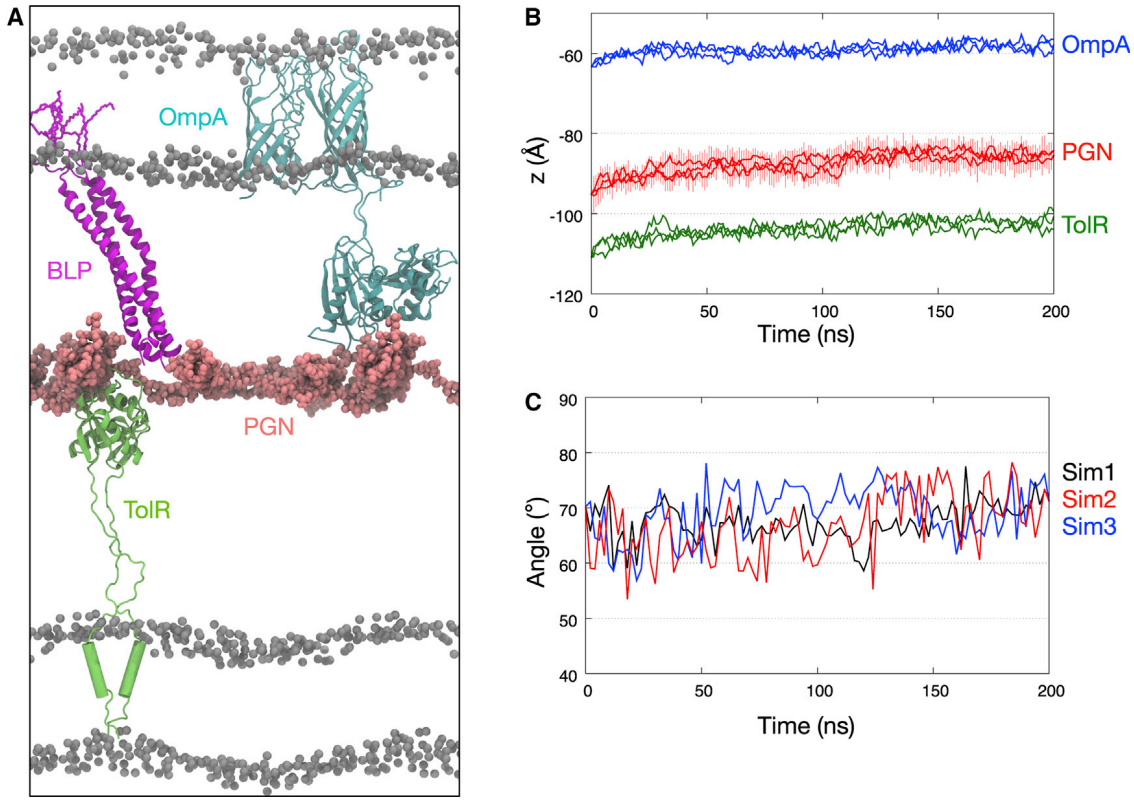


Figure 4. Simulation of TolR with BLP and Half-Truncated OmpA

(A) A snapshot from the end of a 200 ns simulation.

(B) The center of mass motion along the z axis of the remaining OmpA C-terminal domain, PGN cell wall, and TolR periplasmic domain as described in Figure 2.

(C) BLP tilt angle from three simulations measured as the angle between the centers of mass of the outer membrane, the N-terminus, and the C-terminus of BLP. See also Figure S8.

The linker regions of both proteins were only partially extended to enable the cell wall to be maintained at this position. In other words, the proteins had the potential to adopt other arrangements in terms of their location with respect to the cell wall but maintained a position in which the cell was sandwiched equidistant between the two proteins for the duration of these simulations.

Given we have previously reported details of the interactions between OmpA and PGN (Samsudin et al., 2016; Samsudin et al., 2017), here we focus on the details of the TolR-PGN interactions. Analysis of the TolR residues in contact with PGN (where contact is defined as an interatomic distance of $\leq$4 Å) reveals the five residues that made frequent contacts. These are Glu89, Lys122, Thr139, Gln140, and Pro141 (Figure 2C), and they are the same residues identified above as being key to the initial binding and stabilization process. More specifically, the arrangement of PGN in the cell wall of *E. coli* is such that numerous hydroxyl and amide groups on the sugar backbone and the peptide chains are available for hydrogen bonds (Figure 3A). In addition, there are three negatively charged carboxyl groups and one positively charged amine group on the non-cross-linked peptide chains (on residue D-glutamate, residue D-alanine, and mDAP) that can form salt bridges with TolR periplasmic domain. Examples of these interactions involving the key residues identified in Figure 2C are shown in Figure 3B. We found 13 TolR residues that formed hydrogen bonds with the different parts of the cell wall (Figure 3C). These hydrogen bonds were short-lived, with each lasting no longer than 30% of the simulation timescale. Fewer salt bridge interactions were found, as there are only five charged residues that are accessible to the cell wall. These salt bridges, however, were longer lasting, with lifetimes up to 80% of the simulation timescale (Figure 3D). Decomposition of binding free energy shows a larger coulombic contribution compared with that of van der Waals interactions, which is concordant with the numerous hydrogen bonds and salt bridges (Figure 3E). Interestingly, the coulombic contribution of the free energy is correlated to the number of hydrogen bonds formed between TolR and the cell wall, suggesting that hydrogen bond formation is key for TolR binding (Figure S5). Mapping the electrostatic surface of TolR revealed a predominantly negatively charged surface facing the cell wall contributed by the carboxyl groups on the C terminus and downstream polar residues (Figure S6). Smaller positively charged patches are found interspersed around this negatively charged surface due to basic residues like Lys122. It makes sense, therefore, that the negatively charged C-terminal region of TolR formed initial binding to the cell wall via electrostatic interactions with positive moieties of the PGN. To quantify the forces required to maintain

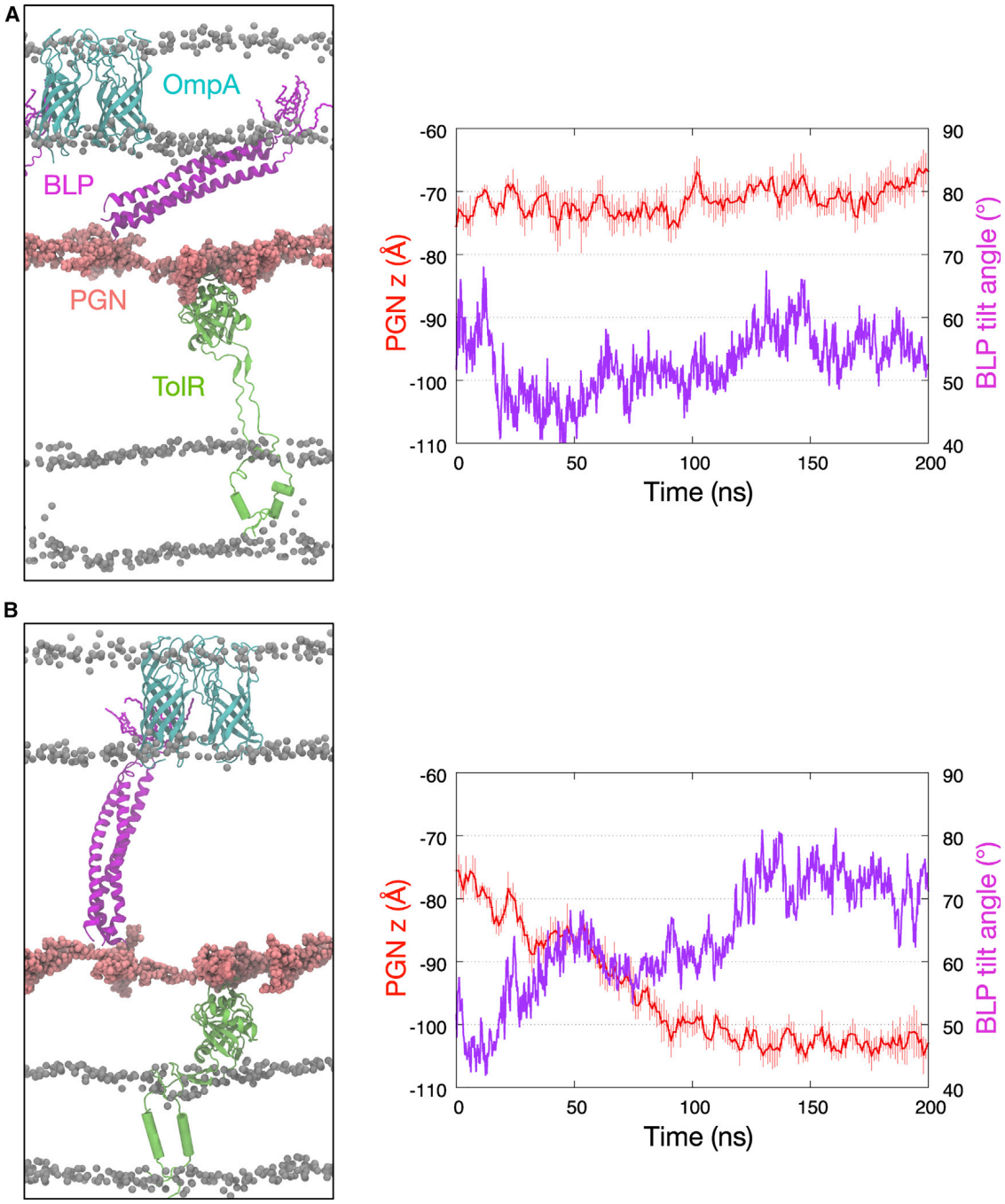


Figure 5. The Flexibility of TolR Binding to the Cell Wall

(A) A snapshot from the end of a 200 ns simulation whereby the C-terminal domains and the linker region of OmpA were truncated. In this simulation, the linker between the N-terminal transmembrane domains and the periplasmic domains of TolR was fully extended. The graph on the right shows the center of mass motion along the z axis of the PGN cell wall (red). Error bars indicate standard deviations from three adjacent stands. BLP tilt angle as measured in Figure 4 is shown in magenta.

(B) A snapshot from a simulation whereby the TolR linker was fully contracted and the TolR bound to the inner membrane, with PGN center of mass motion and BLP tilt angle shown on the right.

TolR-cell wall binding, we performed steered molecular dynamics simulations whereby the periplasmic domain of TolR was pulled away from the cell wall. An average force of around 400 kJ mol^{-1} nm^{-1} was required to detach the protein from the cell wall (Figure S7A). In comparison, a similar set of simulations performed on the OmpA periplasmic domain revealed a much higher force (~1,500 kJ mol^{-1} nm^{-1}) (Figure S7B), potentially due to the larger size of the OmpA periplasmic domain

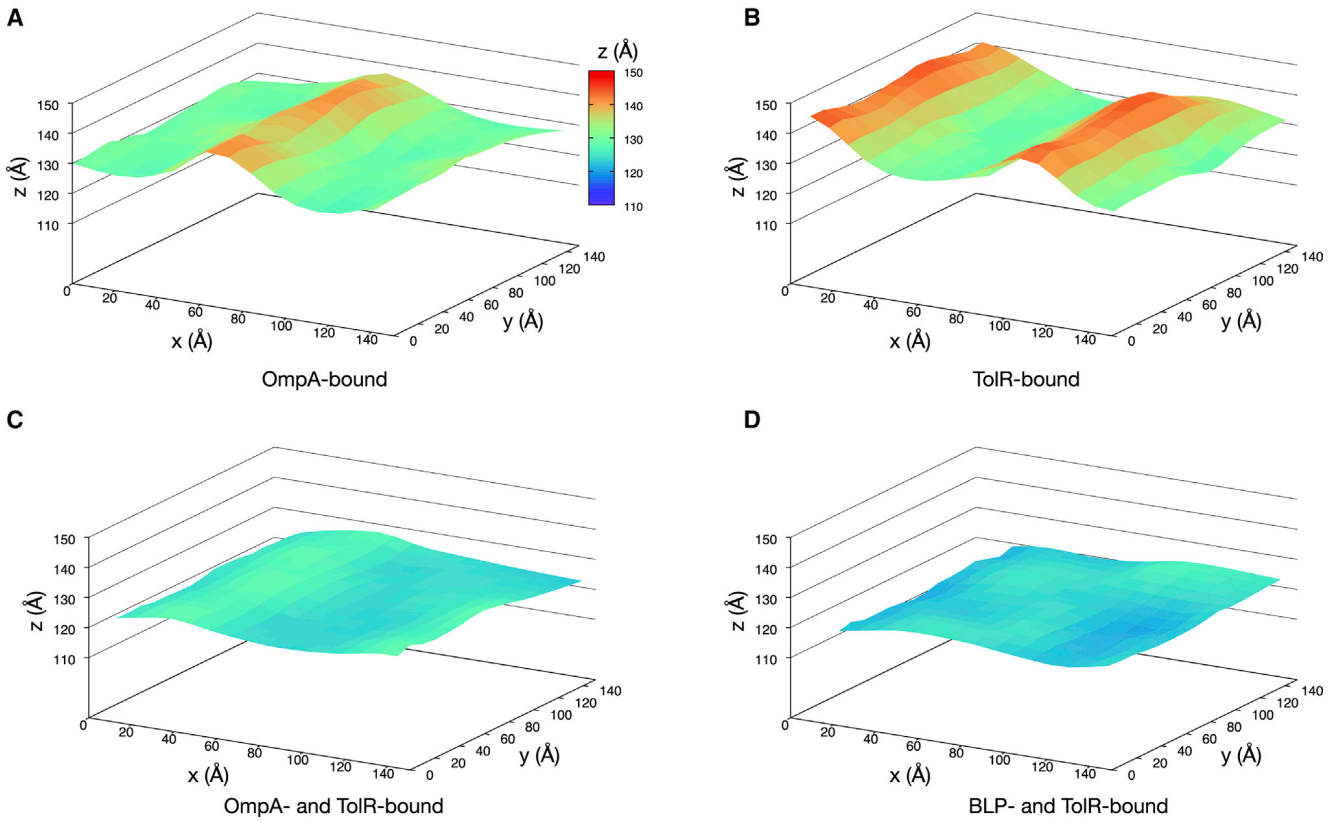


Figure 6. The z Coordinates of the PGN Cell Wall Projected onto a Surface Representation

Data were taken from (A) the beginning of wild-type simulation whereby only OmpA bound to the cell wall, (B) the end of simulation with only TolR bound to the cell wall, (C) the end of wild-type simulation whereby both OmpA and TolR bound to the cell wall, and (D) the end of truncated OmpA simulation whereby BLP and TolR bound to the cell wall.

compared with that of TolR (2,490 and 1,512 atoms, respectively). OmpA therefore has a bigger surface area for interaction with the cell wall resulting in a stronger binding.

TolR-Peptidoglycan Interactions when OmpA Is Truncated

To determine whether the aforementioned location of the cell wall, approximately equidistant from TolR in the inner membrane and OmpA in the outer membrane, is a consequence of the PGN binding strength of the protein domains being approximately equal, or simply a function of the starting position of the simulations, the C-terminal domain and the linker region of OmpA were truncated, leaving the N-terminal β barrel (residues 1–172). One BLP trimer was incorporated into these simulation systems to provide an anchor between the cell wall and the outer membrane. BLP is covalently attached to the cell wall via a peptide bond between the C-terminal lysine residue on one of its protomers and the mDAP moiety of the PGN. The remaining two BLP protomers are not covalently linked to the cell wall and their C-terminal lysine residues are able to form salt bridge interactions with negatively charged moieties of the PGN, hence providing additional support to the cell wall. In the full-length OmpA simulations and when only one of the OmpA protomers was truncated, the cell wall remained bound to both OmpA and TolR in a fashion similar to the wild-type simulations, while BLP tilted to around 60° with respect to the plane of the outer membrane (Figures 4 and S8).

When both protomers were truncated, however, two distinct behaviors were observed. In three of the simulations TolR remained extended and bound to the cell wall, with the BLP trimer tilted at 60° to enable the location of the cell wall to remain approximately equidistant between the two membranes (Figure 5A). In contrast, in one of the four simulations, the linker of TolR contracted such that the bulk of the protein moved to rest on the inner membrane, with BLP almost at right angles to the plane of the outer membrane (Figure 5B). Interestingly, TolR remained bound to PGN throughout this process. This provides compelling evidence that the PGN-binding domain of TolR proposed by Kleanthous and co-workers does indeed stably bind PGN, and that this binding can withstand contraction of the TolR linker (Wojdyla et al., 2015). Furthermore, these observations show that the balance of non-covalent interactions between proteins in both membranes and the cell wall acts like a clamp from both sides in maintaining the position of the cell wall. BLP by itself on the outer membrane side is not sufficient, given that its ability to bend and tilt enables significant deviation in the cell-wall position. This agrees with experimental studies that showed that mutations in either the *tolR* or the *ompA* gene destabilized the cell envelope, resulting in the formation of outer

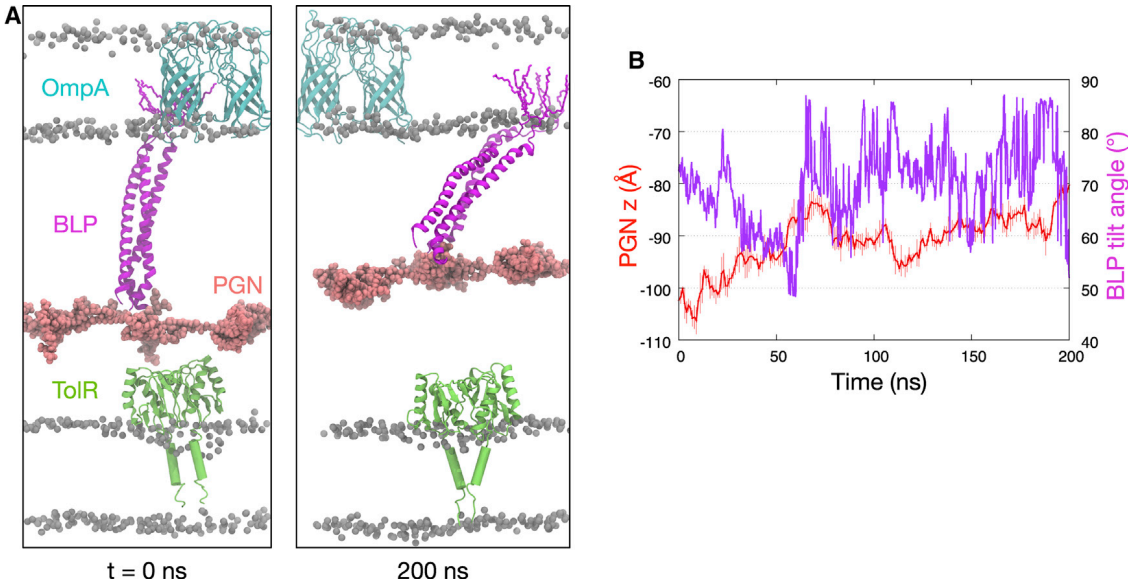


Figure 7. TolR in the Closed Conformation Does Not Interact with the Cell Wall

(A) Snapshot at the beginning (left) and at the end (right) of a 200 ns simulation whereby TolR in its open state was replaced with a closed state (PDB: 5BY4). The TolR periplasmic domain did not bind to the PGN cell wall, and BLP returned to its tilted configuration, shifting the cell wall toward the outer membrane.

(B) Center of mass motion along the z axis for the cell wall and the BLP tilt angle during the simulation as described for Figure 5.

membrane vesicles in *E. coli* (Deatherage et al., 2009; Perez-Cruz et al., 2016).

The Effects of TolR and OmpA on the Structure of the Cell Wall

Our previous study showed that binding of OmpA to the cell wall caused a local buckling effect on the latter, whereby the surface of the cell wall noticeably curved toward the outer membrane at the point of contact (Samsudin et al., 2017). Similarly here, when only TolR and the cell wall (no OmpA or BLP) were included in the simulation system, the contraction of the linker pulled the cell wall toward the inner membrane, resulting in local curvature of the cell wall. In contrast, when both OmpA and TolR were bound to the cell wall, the degree of undulation observed during the simulations was significantly reduced (Figure 6). Interestingly the distortions were also significantly reduced when TolR was bound to the cell wall in the presence of BLP (but without OmpA in the outer membrane). This is in agreement with our previous studies of the outer membrane and the cell wall in which BLP was able to prevent undulations that were otherwise present when OmpA alone was bound to the cell wall. These observations suggest that TolR and OmpA binding to the cell wall from either side of the cell envelope prevents any local distortions caused by either one of them binding alone. From the outer membrane side, BLP also plays a role to this effect; presumably, the greater the number of membrane protein interactions with the cell wall, the less distorted the cell wall is, but this hypothesis should be tested with a wider range of PGN-binding proteins.

The Closed State of TolR Does Not Bind Peptidoglycan

As a further test of whether the PGN binding of TolR is specific to the identified binding domain, or whether other regions of the protein can bind PGN too, TolR-PGN interactions when TolR is in its closed conformation were also explored. The coordinates for the protein were taken from the X-ray structure (PDB: 5BY4), with the transmembrane helices modeled in as reported by Wojdyla et al. (2015). The final snapshot was extracted from our simulation with truncated OmpA dimer in which the TolR linker had contracted to enable interaction of the protein with the inner membrane while still being bound to PGN. The TolR was replaced by the X-ray structure of the closed state. Thus, at the start of the simulation, BLP was extended and essentially at right angles to the plane of the outer membrane, and the TolR in the closed state was in contact with the inner membrane and within 5 Å of the cell wall (similar to our previous setup with open state TolR). After 200 ns of simulation BLP tilted to pull the cell wall approximately 20 Å toward the outer membrane and away from TolR (Figure 7). The electrostatic surface potential of the periplasmic domain facing the cell wall in the TolR closed state is similar to that in the open state, i.e., predominantly negatively charged surface surrounding small positively charged patches (Figure S6). In the closed state, however, the mobile C-terminal domain of TolR responsible for the initial interaction with the cell wall and the flexible linker connecting the periplasmic domain and the N-terminal helices are folded together into a β sheet buried within the dimeric structure. The lack of PGN binding of the protein in this state provides further evidence that binding of TolR requires specific domains that are not accessible in the closed state of the protein.

Conservation of Peptidoglycan-Binding Residues of TolR across Bacterial Species

Having identified Glu89 and Lys122 as key residues in the interaction of TolR and the cell wall, it is worth investigating if this binding mechanism is universally conserved. Sequence

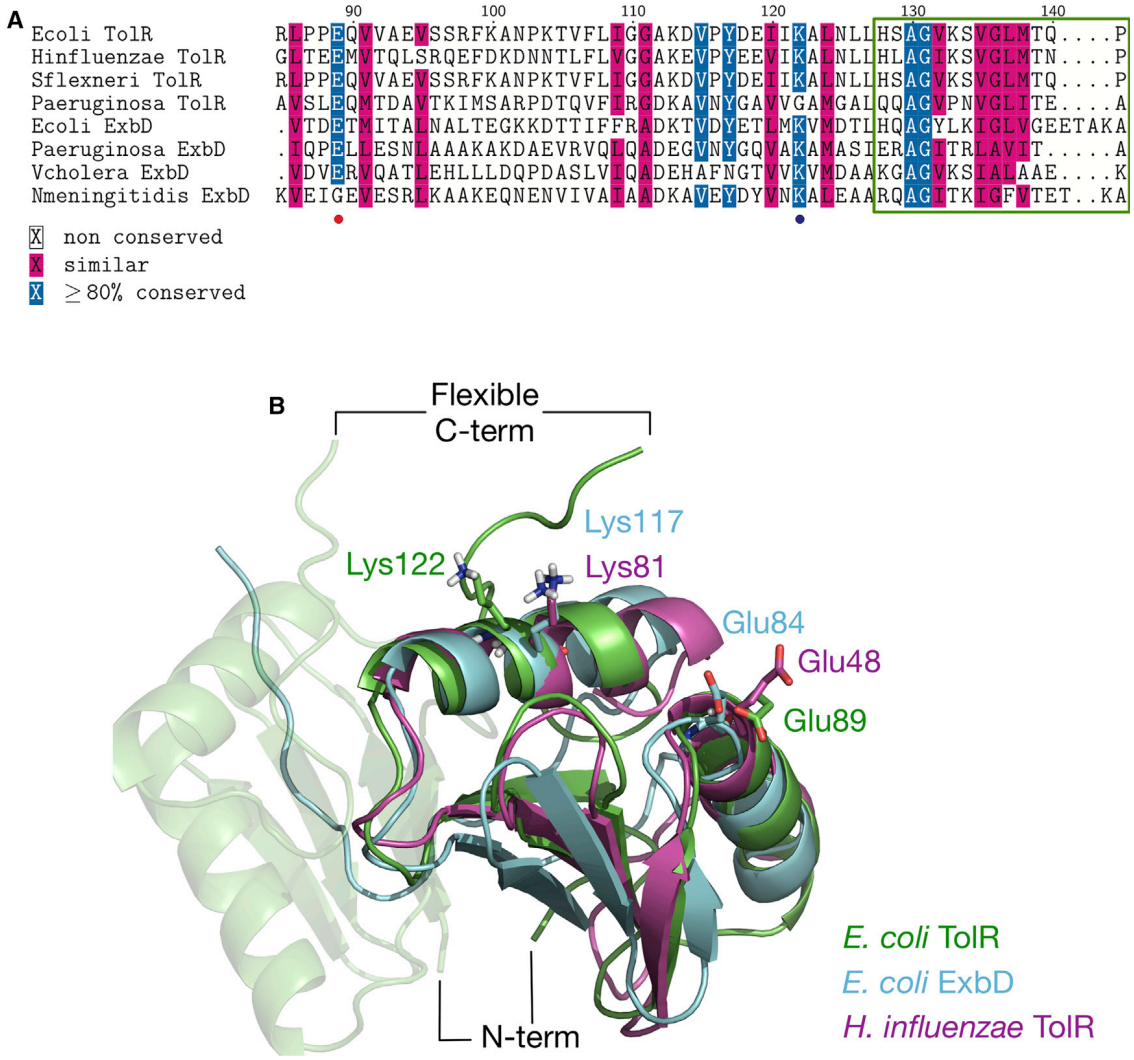


Figure 8. Key Residues in Cell-Wall Binding Are Conserved across Species

(A) Multiple sequence alignment involving TolR from *E. coli* (UniProt: P0ABV6), *Haemophilus influenzae* (P43769), *Shigella flexneri* (P0ABV9), and *Pseudomonas aeruginosa* (P50599), as well as ExbD from *E. coli* (P0ABV2), *P. aeruginosa* (Q9RMT1), *Vibrio cholerae* (Q9ZHV9), and *Neisseria meningitidis* (P0A0R8). The positions of residues equivalent to Glu89 and Lys122 in *E. coli* TolR are marked with red and blue circles, respectively. The flexible C-terminal region is highlighted in the green box. Conserved and similar residues are shaded in light blue and magenta, respectively.

(B) Structural alignment of *E. coli* TolR open state with *E. coli* ExbD (PDB: 2PFU) (Garcia-Herrero et al., 2007) and *H. influenza* TolR (2JWK) (Parsons et al., 2008). The all-atom RMSDs for these alignments are 3.0 and 2.4 Å, respectively. Residues equivalent to Glu89 and Lys122 are highlighted in stick representation. Only one of the two protomers is shown for clarity.

alignment of TolR from different gram-negative bacteria, as well as its structural homolog, ExbD from the TonB system, revealed that both residues are well preserved (Figure 8) (Garcia-Herrero et al., 2007). Structural alignment of the open-state *E. coli* TolR with *Haemophilus influenzae* TolR and *E. coli* ExbD lends further support by showing that all of these glutamates and lysines are found in regions that are solvent accessible, and therefore are able to interact with the cell wall.

Our simulations also predicted the role of the highly mobile C-terminal region in providing the first contact with the cell wall via the carboxyl group on Pro141 and two downstream polar residues (Thr139 and Gln140). A similar secondary structure is expected in this region for other homologs. Indeed, the C terminus of *E. coli* ExbD contains a large unstructured loop, while in *H. influenzae* TolR the corresponding residues could not be assigned by NMR spectroscopy, which indicates an inherent flexibility. It is worth noting that in the structure of the TolR dimer closed state, the C terminus of protomer A via residues 134–138 is a part of a five-stranded β sheet that includes the N terminus of protomer B. PMF-dependent unfolding of the N terminus to form the open state unleashes residues 134–138 to form the flexible region. That these residues are conserved in other bacteria further hints at the presence of an unstructured loop in the open state of TolR in other species. As the first contact made with the cell wall is facilitated by the carboxyl group on the C terminus instead of its side chain, we conjecture that the specific residue at this position may not be as important as the ability of the entire flexible C-terminal region to snorkel toward the PGN.

DISCUSSION

In summary, we report the first atomistic molecular dynamics simulation study to our knowledge to include both membranes and the cell wall of *E. coli*. Our simulations enable us to explore the cell-wall interactions of the putative open-state model of TolR and the impact of these interactions on components inside the periplasm and the outer membrane, namely, BLP and OmpA, respectively. We show that the non-covalent binding of the cell wall of TolR and full-length OmpA dimers maintains the location of the cell wall; in other words, the interactions balance each other. TolR binds PGN via a mechanism in which first the terminal residues of each monomer act like arms, snorkeling toward the cell wall and then binding when an appropriate moiety of the PGN is encountered. Following this, the bulk of the protein is pulled toward the PGN and then interacts via primarily electrostatic interactions. These interactions are stable over a 500 ns timescale and, together with the C termini of the OmpA dimers from the outer membrane side, are able to hold the cell wall at a position that is about equidistant from the two membranes. When the OmpA is truncated to remove its periplasmic, PGN-binding C-terminal domain, TolR is able to ''pull'' the cell wall toward the inner membrane, even when BLP is present. Thus both proteins binding PGN from both sides are required to hold the cell wall in place. Furthermore when both proteins bind simultaneously, they are able to maintain the cell wall in a flat conformation; when only one protein binds, the cell wall experiences local undulations; here BLP also plays a role in preventing undulations. We provide evidence to show that TolR is not able to bind the cell wall when in its closed state as the PGN-binding domain is inaccessible. In future work it would be informative to study mutations in the putative PGN-binding domain of TolR and explore any plausible additional orientations of the protein with respect to the membrane and cell wall. It is interesting to note that, in our simulations, while PGN can distort such that it undulates in response to the absence or presence of proteins, it does not alter its structure to accommodate TolR in the closed state. Specific interactions are clearly needed for the PGN-TolR interaction. While the emphasis has been on the TolR-PGN interactions, it is important to note that here we provide further compelling evidence that BLP is able to adjust its tilt angle to accommodate a range of outer membrane-to-cell wall distances. We show that for fine control over cell-wall position, BLP alone is not sufficient; the non-covalent interactions from proteins in both membranes play a crucial role. This is in line with experimental observations that *E. coli* mutants missing OmpA or TolR produce more vesicles than normal cells, presumably at least in part due to the easier detachment from the cell wall. Overall a picture is beginning to build up that suggests the local conformation of the cell wall is dependent upon the non-covalent interactions from proteins in both membranes. It is important to note here that other proteins that bind to the cell wall from both sides, such as PAL and MotB, are not considered in the present study; these too will undoubtedly have an impact on properties such undulations/distortions and overall position of the cell wall with respect to the two membranes (Parsons et al., 2006; Roujeinikova, 2008). It is possible and likely, based on our current findings, that the cell-wall location varies as a function of the local protein composition of both membranes, and therefore one would expect an undulating structure at the molecular level of detail. Further studies with additional inner and outer membrane proteins and varying copy numbers of BLP are needed to explore these aspects. Our current study provides an important step toward the study of the mechanical interplay between the two membranes and the cell wall of gram-negative bacteria at the atomistic level.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHOD DETAILS
  - Protein and Cell Wall Parameters
  - Outer and Inner Membrane Construction
  - Simulation Systems
  - Simulation Protocols
  - Analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information includes eight figures and can be found with this article online at https://doi.org/10.1016/j.str.2019.01.001.

ACKNOWLEDGMENTS

This research was supported by UK Biotechnology and Biological Sciences Research Council (grant BB/M029573/1) and the Institute for Life Sciences, University of Southampton. This project made use of time on Iridis V supercomputers provided by the University of Southampton and on ARCHER provided by HECBioSim through Engineering and Physical Sciences Research Council (grant EP/L000253/1). The authors thank Phillip Stansfeld for providing the coordinates of the TolR open-state model and Colin Kleanthous for useful discussions.

AUTHOR CONTRIBUTIONS

Conceptualization, S.K.; Investigation, A.T.B. and F.S.; Writing – Original Draft, A.T.B., F.S., and S.K.; Supervision, S.K.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 31, 2018
Revised: November 12, 2018
Accepted: December 31, 2018
Published: January 31, 2019

REFERENCES

Abraham, M.J., Murtola, T., Schulz, R., Pall, S., Smith, J.C., Hess, B., and Lindahl, E. (2015). GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX *1-2*, 19–25.

Asmar, A.T., and Collet, J.F. (2018). Lpp, the Braun lipoprotein, turns 50—major achievements and remaining issues. FEMS Microbiol. Lett. *365*, https://doi.org/10.1093/femsle/fny199.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U S A *98*, 10037–10041.

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., and Hermans, J. (1981). Intermolecular Forces (Reidel).

Braun, V. (1975). Covalent lipoprotein from outer membrane of Escherichia-coli. Biochim. Biophys. Acta *415*, 335–377.

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. J. Chem. Phys. *126*, 014101.

Carpenter, T., Khalid, S., and Sansom, M.S. (2007). A multidomain outer membrane protein from Pasteurella multocida: modelling and simulation studies of PmOmpA. Biochim. Biophys. Acta *1768*, 2831–2840.

Deatherage, B.L., Lara, J.C., Bergsbaken, T., Barrett, S.L.R., Lara, S., and Cookson, B.T. (2009). Biogenesis of bacterial membrane vesicles. Mol. Microbiol. *72*, 1395–1407.

DeLano, W. (2002). The PyMOL molecular graphics system (DeLano Scientific). http://www.pymol.org.

Do, C.B., Mahabhashyam, M.S.P., Brudno, M., and Batzoglou, S. (2005). ProbCons: probabilistic consistency-based multiple sequence alignment. Genome Res. *15*, 330–340.

Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G. (1995). A smooth particle mesh Ewald method. J. Chem. Phys. *103*, 8577–8593.

Garcia-Herrero, A., Peacock, R.S., Howard, S.P., and Vogel, H.J. (2007). The solution structure of the periplasmic domain of the TonB system ExbD protein reveals an unexpected structural homology with siderophore-binding proteins. Mol. Microbiol. *66*, 872–889.

Gerding, M.A., Ogata, Y., Pecora, N.D., Niki, H., and De Boer, P.A.J. (2007). The trans-envelope Tol-Pal complex is part of the cell division machinery and required for proper outer-membrane invagination during cell constriction in E. coli. Mol. Microbiol. *63*, 1008–1025.

Graham, L.L., Beveridge, T.J., and Nanninga, N. (1991). Periplasmic space and the concept of the periplasm. Trends Biochem. Sci. *16*, 328–329.

Gumbart, J.C., Beeby, M., Jensen, G.J., and Roux, B. (2014). Escherichia coli Peptidoglycan Structure and Mechanics as Predicted by Atomic-Scale Simulations. PLoS Comput. Biol. *10*.

Hess, B., Bekker, H., Berendsen, H.J.C., and Fraaije, J.G.E.M. (1997). LINCS: a linear constraint solver for molecular simulations. J. Comput. Chem. *18*, 1463–1472.

Horta, B.A.C., Fuchs, P.F.J., van Gunsteren, W.F., and Hunenberger, P.H. (2011). New interaction parameters for oxygen compounds in the GROMOS force field: improved pure-liquid and solvation properties for alcohols, ethers, aldehydes, ketones, carboxylic acids, and esters. J. Chem. Theor. Comput. *7*, 1016–1031.

Hsu, P.-C., Jefferies, D., and Khalid, S. (2016). Molecular Dynamics Simulations Predict the Pathways via Which Pristine Fullerenes Penetrate Bacterial Membranes. J. Phys. Chem. B *120*, 11170–11179.

Hsu, P.C., Bruininks, B.M.H., Jefferies, D., Cesar Telles de Souza, P., Lee, J., Patel, D.S., Marrink, S.J., Qi, Y., Khalid, S., and Im, W. (2017). CHARMM-GUI Martini Maker for modeling and simulation of complex bacterial membranes with lipopolysaccharides. J. Comp. Chem. *38*, 2354–2363.

Humphrey, W., and Dalke, A. (1996). VMD: visual molecular dynamics. J. Mol. Graph *15*, 33–38.

Hwang, H., Paracini, N., Parks, J.M., Lakey, J.H., and Gumbart, J.C. (2018). Distribution of mechanical stress in the Escherichia coli cell envelope. Biochim. Biophys. Acta *1860*, 2566–2575.

Kirschner, K.N., Lins, R.D., Maass, A., and Soares, T.A. (2012). A glycam-based force field for simulations of lipopolysaccharide membranes: Parametrization and validation. J. Chem. Theory Comput. *8*, 4719–4731.

Ma, H., Irudayanathan, F.J., Jiang, W., and Nangia, S. (2015). Simulating gram-negative bacterial outer membrane: a coarse grain model. J. Phys. Chem. B *119*, 14668–14682.

Marcoux, J., Politis, A., Rinehart, D., Marshall, D.P., Wallace, M.I., Tamm, L.K., and Robinson, C.V. (2014). Mass spectrometry defines the C-terminal dimerization domain and enables modeling of the structure of full-length OmpA. Structure *22*, 781–790.

Matias, V.R.F., Al-amoudi, A., Dubochet, J., and Beveridge, T.J. (2003). Cryo-transmission electron microscopy of frozen-hydrated sections of *Escherichia coli* and *Pseudomonas aeruginosa*. J. Bacteriol. *185*, 6112–6118.

Miller, S.I., and Salama, N.R. (2018). The gram-negative bacterial periplasm: size matters. PLoS Biol. *16*, e2004935.

Nikaido, H. (2003). Molecular basis of bacterial outer membrane permeability revisited. Microbiol. Mol. Biol. Rev. *67*, 593–656.

Ortiz-Suarez, M.L., Samsudin, F., Piggot, T.J., Bond, P.J., and Khalid, S. (2016). Full-length OmpA: structure, function, and membrane interactions predicted by molecular dynamics simulations. Biophys. J. *111*, 1692–1702.

Park, J.S., Lee, W.C., Yeo, K.J., Ryu, K.S., Kumarasiri, M., Hesek, D., Lee, M., Mobashery, S., Song, J.H., Il Kim, S., et al. (2012). Mechanism of anchoring of OmpA protein to the cell wall peptidoglycan of the gram-negative bacterial outer membrane. FASEB J. *26*, 219–228.

Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single-crystals—a new molecular dynamics method. J. Appl. Phys. *52*, 7182–7190.

Parsons, L.M., Grishaev, A., and Bax, A. (2008). The periplasmic domain of TolR from Haemophilus influenzae forms a dimer with a large hydrophobic groove: NMR solution structure and comparison to SAXS data. Biochemistry *47*, 3131–3142.

Parsons, L.M., Lin, F., and Orban, J. (2006). Peptidoglycan recognition by Pal, an outer membrane lipoprotein. Biochemistry *45*, 2122–2128.

Perez-Cruz, C., Canas, M.A., Gimenez, R., Badia, J., Mercade, E., Baldoma, L., and Aguilera, L. (2016). Membrane vesicles released by a hypervesiculating Escherichia coli Nissle 1917 tolR mutant are highly heterogeneous and show reduced capacity for epithelial cell interaction and entry. PLoS One *11*, e0169186.

Piggot, T.J., Holdbrook, D.A., and Khalid, S. (2011). Electroporation of the E. coli and S. Aureus membranes: molecular dynamics simulations of complex bacterial membranes-suppinfo. J. Phys. Chem.

Roujeinikova, A. (2008). Crystal structure of the cell wall anchor domain of MotB, a stator component of the bacterial flagellar motor: implications for peptidoglycan recognition. Proc. Natl. Acad. Sci. U S A *105*, 10348–10353.

Samsudin, F., Boags, A., Piggot, T.J., and Khalid, S. (2017). Braun's lipoprotein facilitates OmpA interaction with the Escherichia coli cell wall. Biophys. J. *113*, 1496–1504.

Samsudin, F., Ortiz-Suarez, M.L., Piggot, T.J., Bond, P.J., and Khalid, S. (2016). OmpA: a flexible clamp for bacterial cell wall attachment. Structure *24*, 2227–2235.

Schmid, N., Eichenberger, A.P., Choutko, A., Riniker, S., Winger, M., Mark, A.E., and van Gunsteren, W.F. (2011). Definition and testing of the GROMOS force-field versions 54A7 and 54B7. Eur. Biophys. J. Biophys. Lett. *40*, 843–856.

Schrodinger. (2017). The PyMOL Molecular Graphics System, Version 2.0 (Schrödinger, LLC).

Schwechheimer, C., Kulp, A., and Kuehn, M.J. (2014). Modulation of bacterial outer membrane vesicle production by envelope structure and content. BMC Microbiol. *14*, 324.

Shu, W., Liu, J., Ji, H., and Lu, M. (2000). Core structure of the outer membrane lipoprotein from Escherichia coli at 1.9 A resolution. J. Mol. Biol. *299*, 1101–1112.

Turner, R.D., Hurd, A.F., Cadby, A., Hobbs, J.K., and Foster, S.J. (2013). Cell wall elongation mode in Gram-negative bacteria is determined by peptidoglycan architecture. Nat. Commun. *4*, 1496–1498.

Vollmer, W., and Bertsche, U. (2008). Murein (peptidoglycan) structure, architecture and biosynthesis in Escherichia coli. Biochim. Biophys. Acta Biomembr. *1778*, 1714–1734.

Vollmer, W., and Seligman, S.J. (2010). Architecture of peptidoglycan: more data and more models. Trends in Microbiol. *18*, 59–66.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

Wojdyla, J.A., Cutts, E., Kaminska, R., Papadakos, G., Hopper, J.T., Stansfeld, P.J., Staunton, D., Robinson, C.V., and Kleanthous, C. (2015). Structure and function of the Escherichia coli Tol-Pal stator protein TolR. J. Biol. Chem. *290*, 26675–26687.

Wolf, M.G., Hoefling, M., Aponte-santamaría, C., Grubmüller, H., and Groenhof, G. (2010). g_membed: Efficient Insertion of a Membrane Protein into an Equilibrated Lipid Bilayer with Minimal Perturbation. J. Comp. Chem. *31*, 2169–2174.

Wu, E.L., Engström, O., Jo, S., Stuhlsatz, D., Yeom, M.S., Klauda, J.B., Widmalm, G., and Im, W. (2013). Molecular dynamics and NMR spectroscopy studies of E. coli lipopolysaccharide structure and dynamics. Biochem. J. *105*, 1444–1455.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
TolR open state	This work	https://doi.org/10.5281/zenodo.1636577
TolR closed state	(Wojdyla et al., 2015)	PDB: 5BY4
OmpA dimer	(Marcoux et al., 2014)	https://doi.org/10.5281/zenodo.1636577
Peptidoglycan cell wall	(Samsudin et al., 2017)	https://doi.org/10.5281/zenodo.1636577
Braun's lipoprotein	(Shu et al., 2000)	PDB: 1EQ7
Software and Algorithms		
GROMACS 2018	(Abraham et al., 2015)	www.gromacs.org
VMD 1.9	(Humphrey and Dalke, 1996)	www.ks.uiuc.edu/Research/vmd
PyMOL 1.8	(DeLano, 2002)	Pymol.org
JalView 2.9	(Waterhouse et al., 2009)	Jalview.org
APBS 2.1	(Baker et al., 2001)	pymolwiki.org/index.php/APBS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Syma Khalid (S.Khalid@soton.ac.uk).

METHOD DETAILS

Protein and Cell Wall Parameters

The full-length OmpA dimer model was obtained from Carol Robinson (Marcoux et al., 2014); their structural stability in a model OM has been verified in our previous work (Ortiz-Suarez et al., 2016). The model of the TolR dimer in the open state was obtained from Phillip Stansfeld (Wojdyla et al., 2015). A PGN network consisting of three strands of 10 repeating NAG-NAM-peptide units was constructed and positioned ~90 Å from the surface of the lower leaflet of the OM. The BLP homotrimer was built based on the structure from Shu et al. (PDB: 1EQ7) (Shu et al., 2000) with the last residues on both the N- and C termini manually added back using PyMOL (DeLano, 2002). The N-terminus was in turn attached to the tripalmitoyl-S-glyceryl-cysteine residues to incorporate the BLP into the inner leaflet of the OM. The parameters for tripalmitoyl-S-glyceryl-cysteine were constructed from the standard GROMOS 54A7 force field with the GROMOS 53A6OXY ether parameters used for the linkage region (Horta et al., 2011). PGN was then covalently linked to the lysine on one of the C termini of the BLP trimer via its mDAP residue. The linkage was constructed using the standard GROMOS 54A7 parameters.

Outer and Inner Membrane Construction

The OM model was asymmetric: the upper leaflet was made entirely of full-rough Ra LPS lipids of the R1 core type, whereas the lower leaflet comprised a mixture of phospholipids in the following ratio: 90% 1-palmitoyl 2-cis-vaccenic phosphatidylethanolamine, 5% 1-palmitoyl 2-cis-vaccenic phosphatidylglycerol, and 5% 1-palmitoyl 2-cis-vaccenic 3-palmitoyl 4-cis-vaccenic diphosphatidylglycerol, also known as cardiolipin. This OM model has been validated in our previously reported studies (Samsudin et al., 2017).

The IM model was symmetric, i.e. both leaflets were made entirely of phospholipid in the following ratio: 90% 1-palmitoyl 2-cis-vaccenic phosphatidylethanolamine, 5% 1-palmitoyl 2-cis-vaccenic phosphatidylglycerol, and 5% 1-palmitoyl 2-cis-vaccenic 3-palmitoyl 4-cis-vaccenic diphosphatidylglycerol, this bilayer has the same composition per leaflet as the inner leaflet of the OM model.

Simulation Systems

Simulation systems were constructed with full length dimeric OmpA in the OM and the extended conformation of TolR in the IM. OmpA was inserted into the OM using the *membed* protocol from the GROMACS package (Wolf et al., 2010). TolR was inserted into the IM manually and any overlapping lipids were removed. OmpA was positioned bound the cell wall whilst TolR was placed 5 A away so it can either interact or move away from the PGN structure. Truncation of OmpA C-terminal domain was performed, whereby residue 173–316 was removed from either one of the protomers or from both. In the truncated systems BLP is included as previously described. The cell wall was included in all simulation systems, where the cross linking was parameterized from the peptide bond parameters from the GROMOS 54A7 force field (Schmid et al., 2011). This cross linking is periodic, where the sheet is self-bonded at the end of the strands, using the existing glycan parameters from GROMOS 54A7, allowing the sheet to behave

as an infinite layer. We have described and used this approach in our previous studies (Samsudin et al., 2017). All simulation systems are run at a neutral charge with a concentration of 0.2 M sodium chloride ions in solution. All repeats of the systems are run from the same starting point, where the independent repeats are achieved by regenerating the velocities of the atoms randomly at the start point for each repeat.

Simulation Protocols

All simulations were performed using the GROMACS 2018 simulation package, the GROMOS 54A7 force field with the SPC water model (Berendsen et al., 1981; Schmid et al., 2011; Abraham et al., 2015). Each simulation was run for 200 ns, and at least three independent repeats of each simulation was performed, giving at least 600 ns for each system simulated. Temperatures of 310 K were maintained using the velocity rescale thermostat using a time constant of 1 ps (Bussi et al., 2007). The pressure was maintained semi-isotropically at 1 atm using the Parrinello-Rahman barostat with a time constant of 1 ps (Parrinello and Rahman, 1981). All bonds were constrained using the LINCS algorithm to allow for an integration time step of 2 fs (Hess et al., 1997). Long-range electrostatics were described using the particle mesh Ewald method (Essmann et al., 1995). The short-range electrostatic cutoff used was 1.4 nm and the short-range van der Waals cutoff was also 1.4 nm. The last frame from a 200 ns equilibrium simulation involving OmpA, the cell wall, TolR, the inner and outer membranes was subsequently used for steered MD simulations. A harmonic spring (force constant of 1000 kJ mol^{-1} nm^2) was attached to the center of mass of either the OmpA or TolR periplasmic domain, and these domains were pulled away from the cell wall along the z axis (perpendicular to the plane of the membrane) at a constant velocity of 0.1 nm ns^{-1}. Each steered MD simulation was repeated three times starting with different initial velocities and the average force required to pull these periplasmic domains from the cell wall was calculated.

Analysis

All sequences were fetched from the Uniprot webserver and aligned in JalView (Waterhouse et al., 2009) using the ProbCons algorithm (Do et al., 2005). The structure of TolR open state model was aligned to E. coli ExbD (PDB: 2PFU) (Garcia-Herrero et al., 2007) and H. influenza TolR (2JWK) (Parsons et al., 2008) using PyMOL (Schrodinger, 2017). Electrostatic profile of TolR was calculated using APBS (Baker et al., 2001).

QUANTIFICATION AND STATISTICAL ANALYSIS

See Method Details above.

7.5 Appendix E

The Hitchhiker's Guide to the Periplasm: Unexpected Molecular Interactions of Antibiotics Revealed by Considering Crowding Effects in *E. coli*

Conrado Pedebos[1], Iain P. S. Smith[1], Alister Boags[1,2], Syma Khalid[1*]

[1]School of Chemistry, University of Southampton, Highfield Campus, Southampton SO17 1BJ, UK
[2]Bioinformatics Institute (A*STAR), 30 Biopolis Street, #07-01 Matrix, Singapore 138671, Singapore

*To whom correspondence should be addressed. E-mail: S.Khalid@soton.ac.uk

Abstract

The periplasm of Gram-negative bacteria is a highly crowded environment comprised of many different molecular species. Antibacterial agents that causes lysis of Gram-negative bacteria by their action against the inner membrane must cross the periplasm to arrive at their target membrane. Very little is currently known about their route through the periplasm, and the interactions they experience. To this end, here atomistic molecular dynamics simulations are used to study the path taken by the antibiotic polymyxin B1 through a number of models of the periplasm which are crowded with proteins and osmolytes to different extents. The simulations reveal that PMB1 forms transient and long-lived interactions with proteins and osmolytes that are free in solution as well as lipoproteins anchored to the outer membrane and bound to the cell wall. We show that PMB1 may be able to 'hitchhike' within the periplasm by binding to lipoprotein carriers. Overall our results show that PMB1 is rarely uncomplexed within the periplasm; an important consideration for interpretations of its therapeutic mechanism of action. It is likely that this observation can be extended to other antibiotics that rely on diffusion to cross the periplasm.

57

58

59

60

Introduction

The periplasm of Gram-negative bacteria is a crowded aqueous compartment bounded by the inner and outer membranes. The cell wall is contained within the periplasm as well as hundreds of proteins including chaperones, transporters proteases and nucleases[1,2]. The solution also contains a range of osmolytes, including urea, sugars, spermidine and putrescine. This makes for a complex and crowded environment for any molecular species to negotiate when moving across the periplasm towards either membrane.

Very little is known about the spatial arrangement of these myriad molecules within the periplasm. In other words, it is not known if the proteins and osmolytes are evenly distributed, or if is there some degree of organization and if so, to what extent. This makes it very difficult to predict the interactions experienced by molecules within the periplasm. This extends to molecules that are not native to the bacteria, such as antibacterial agents. Thus, we have little information regarding which moieties of antibiotics are available to carry out the desired functions, and which are unavailable as they are involved in interactions with native proteins/osmolytes/cell wall. To this end we have conducted a study of polymyxin B1 (PMB1) within models of the *E. coli* periplasm. PMB1 is a lipopeptide antibiotic used as a “last resort” drug for the treatment of infections caused by Gram-negative bacteria[3]. PMB is composed of a cyclic, cationic polypeptide ring connected to a branched fatty acid tail. The cationic ring contains five residues of the irregular amino acid α,γ-Diamino Butyric acid (DAB), each of which contributes a charge of +1 *e* giving PMB1 an overall charge of +5 *e*. The cationic ring enables solubility in aqueous solvents, whereas the lipid tail facilitates insertion into bacterial membranes[4-7]. While PMB1 along with colistin (polymyxin E) were for many years, last resort antibiotics, in recent years bacterial strains that are resistant to both antibiotics have emerged in a number of countries[8,9]. Thus, in order to either modify these drugs or to develop completely novel antibiotics, it is timely to establish a thorough, molecular-level understanding of each stage of the process *via* which they bring about bacterial cell death. To date, mechanistic studies of PMB1 have focused almost entirely on the two

membranes of Gram-negative bacteria[7,10], leaving unaddressed the question, how does PMB1 cross the periplasm to get from the outer membrane to the inner membrane?

Here, a series of atomistic molecular dynamics simulations (Table 1) were performed of models of portions of the *E. coli* cell envelope. The simulation systems contain an asymmetric model of the outer membrane composed of LPS and phospholipids, a single-layered cell wall, various proteins/lipoproteins, osmolytes and PMB1, with systems sizes ranging from 200,000 to 750,000 atoms. The proteins are a combination of Braun's lipoprotein (BLP), LolA, LolB, OmpA, and Pal (Fig1). BLP is the most abundant protein in *E. coli* (there are an estimated 10^5 copies of BLP in each *E. coli*[11]). It exists as a coiled-coil trimer that is essential for compartment stability[12]. It is anchored in the outer membrane *via* a lipidated moiety at its N-terminus, whereas it is covalently bound to peptidoglycan via its C-terminus. LolA and LolB are small soluble proteins that carry lipoproteins[13,14], they are largely similar in structure, although LolB is anchored to the OM *via* a lipidated moiety whereas LolA is free to diffuse across the cell envelope. OmpA is composed of an eight-stranded barrel which is located in the OM, and is connected *via* a linker to the soluble domain that can bind peptidoglycan in the periplasm[15,16,17]. Pal also has a lipidated anchor in the OM like LolB, while its C-terminal domain resembles the OmpA soluble domain. Like OmpA, Pal also has a linker that can extend into the periplasm enabling the protein to bind non-covalently to peptidoglycan, thereby assisting with maintaining compartment integrity[18,19]. Where present, in each system there are 4 x BLPs and 1 x each of OmpA, Pal, LolA and LolB. The most compositionally complex system studied here also contained a range of osmolytes, in order to better represent the crowded environment that these molecules encounter in the periplasm.

The osmolytes incorporated into our periplasmic model were selected on the basis of a combination of their abundance and chemical diversity. Importantly, all of these osmolytes have their concentrations in the periplasm either documented or estimated in the literature[20-25] and these concentrations are reproduced here: osmoregulated periplasmic glucans (OPG) (20 mM), trehalose (10 mM), putrescine (30 mM), spermidine (3 mM), glycerol (36 mM) and urea (20 mM). Both OPG and trehalose are widely distributed in Bacteria, with OPG having a prominent role on regulating osmotic pressure and virulence[26], whereas trehalose is mainly involved in response to stress conditions[27]. The

polyamines, putrescine and spermidine, are the two most common in all bacteria, with functions that includes supporting bacterial growth, incorporation into the cell wall, and biosynthesis of siderophores[28]. Glycerol is metabolized in *E. coli* cells for different applications, both aerobically and anaerobically[29,30]. Urea is a source of nitrogen, after its breakdown[31].

Simulations were initiated by placing PMB1 molecules randomly in the aqueous region between the outer membrane and the cell wall. The osmolyte concentrations are derived from literature values and the number of proteins is selected to reproduce crowding volume fraction of ϕ ~0.21 as estimated from experimental studies[20]. A set of simulations of PMB1 in just water and ions was also performed for comparison.

Results

Table 1 provided a summary of the simulations performed in this study. Initial observations focused on general mobility and aggregation of PMB1 followed by in depth analyses probing the causes of these observations.

Table 1. Summary of all simulated systems.

		Simulations			
System	Proteins	PMB1 Molecules	Osmolytes and Ions[a]	Length	Force field
PMB_{only}	None	30	Na^+, Cl^-	1 μs	GROMOS 54a7
PMB_{dil}	BLP (x4)	30	Na^+, Cl^-	2 x 500 ns	GROMOS 54a7
PMB_{prot}	BLP (x4), LolA, LolB, Pal, OmpA	30	Na^+, Cl^-	2 x 500ns	GROMOS 54a7
PMB_{crowd}	BLP (x4), LolA, LolB, Pal, OmpA	30	Na^+, Cl^-, Glycerol, Urea, Trehalose, Spermidine, Putrescine, OPG	2 x 500 ns	GROMOS 54a7

[a] Molar concentration: Na+, Cl- (200 mM), Glycerol (35mM), Urea (30mM), Trehalose (10 mM), Spermidine (0.2 mM), Putrescine (30 mM), Osmoregulated Periplasmic Glucans (OPG - 20 mM).

The crowded nature of the systems had a clear impact upon the solvent accessible surface area (SASA) of PMB1 (Fig. 1). The SASA is lower when PMB1 molecules are just in water and counter ions ($PMB1_{only}$), compared to when in the protein-containing systems

(PMB1$_{dil}$, PMB1$_{prot}$ and PMB1$_{crowd}$). Tracking the PMB1 motion within the XY plane (Fig. S1) of the protein-containing systems shows the movement of polymyxins in the crowded systems (PMB1$_{crowd}$ and PMB1$_{prot}$) is more confined compared to PMB1$_{dil}$ in which BLP is the only protein. Additionally, in the latter system, more PMB1 molecules moved towards the outer membrane and the cell wall rather than remaining in the solution area between these two large structures, compared to PMB1$_{crowd}$ and PMB1$_{prot}$. Another effect observed with increasing system complexity is the slower diffusion of PMB1 (Fig. 2A and Fig. S2), by calculation of the translational diffusion coefficients (D_t) from two different time regimes (1-10ns; and 50-100ns). For the PMB$_{only}$ system, the D_t from the longer time regime was estimated to be 4.0 ± 0.3 x 10^{-6} cm^2/s, while for PMB$_{dil}$, PMB$_{prot}$, and PMB$_{crowd}$ systems, the values were 3.8 ± 0.4 x 10^{-8} cm^2/s, 2.9 ± 0.3 x 10^{-8} cm^2/s and 2.0 ± 0.4 x 10^{-8} cm^2/s, respectively, demonstrating a major reduction when compared to PMB$_{only}$ (100-fold). The slowest diffusion is recorded for the most crowded system.

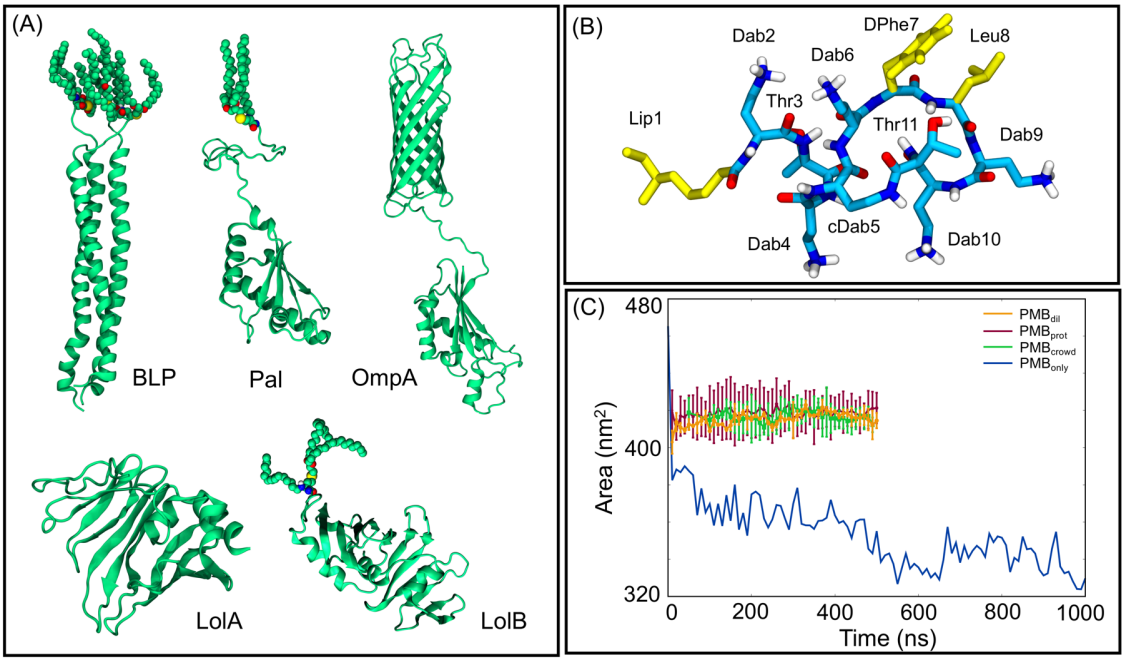


Figure 1. Summary of proteins studied and SASA data. Panel (A) shows the structures of the five proteins simulated in this study. Panel (B) shows the structure of PMB1 and panel (C) provides a summary of SASA versus time for each system.

The dynamics of water was also impacted by crowding (Fig. 2B), with a D_t rate of 4.93 ± 0.4 x 10^{-5} cm^2/s in PMB_{only} compared to 2.49 ± 0.1 x 10^{-5} cm^2/s in PMB_{crowd} for the 50-100 ns time period. The values found for systems in the presence of the outer membrane are similar to a previous report[32] of simulations of the outer membrane in water using the SPC water model[33] and with crowded simulations[34] using a different water model. Protein diffusion rates were also calculated for the PMB_{dil}, PMB_{prot}, and PMB_{crowd} systems, showing D_t values that also decrease with increasing crowding volume fraction ϕ. Although LolA is neither bound to the cell wall nor anchored/embedded in the membrane, its calculated D_t falls in the same range as the other proteins indicating that overall protein motion is quite restricted in the crowded systems for all proteins. While the environment we have simulated is more complex due to the presence of membrane and cell wall, the diffusion rates for proteins calculated here are comparable to previous reports involving simulations of crowded environments[35] and cytoplasm models[36,37], as well as with experimental data from GFP proteins at the periplasm and cytoplasm[38].

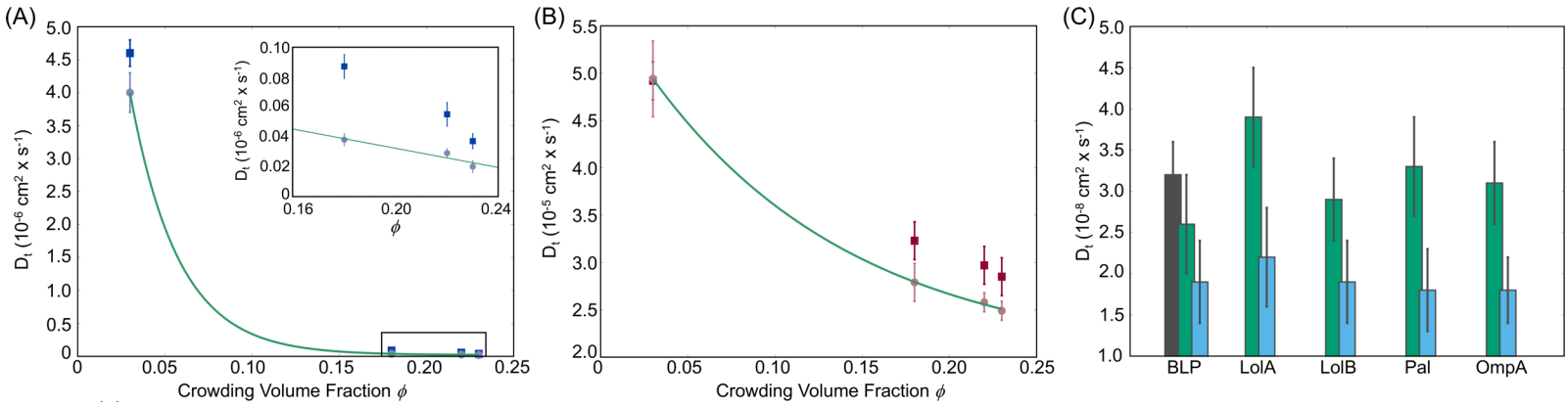


Figure 2: **Translational diffusion coefficients (D_t) obtained for PMB1 (left panel), water (middle panel), and proteins (right panel). (A): Dt values calculated for two different time regimes, 1-10 ns (blue) and 10-100 ns (grey), as a function of the crowding volume fraction ϕ of each system (PMB_{only} = 0.03; PMB_{dil} = 0.18; PMB_{prot} = 0.22; PMB_{crowd} = 0.23). (B): D_t values obtained for water molecules in the same time regimes as above, 1-10 ns (red) and 10-100 ns (pink), as a function of the crowding volume fraction ϕ. Exponential fits were applied for the long-time scale regimes. (C)*:* Histogram showing D_t values for each protein in each system (PMB_{dil}**

= black; PMB_{prot} = green; PMB_{crowd} = blue). Error bars indicate standard error for all molecules across all repeat simulations.

In this study we seek to characterize the molecular interactions that underpin the aforementioned SASA, lateral motion and translational diffusion profiles calculated from our simulations. The complexity of the system composition is such that a vast amount of data regarding molecular interactions is generated from these simulations. To facilitate interpretation of the observations we have presented the results from the perspective of PMB1 interactions, namely PMB1 interactions with itself, osmolytes, proteins, and the cell wall.

PMB1 self-interactions

In systems containing only PMB1 in solution (PMB_{only}), differently sized aggregates (dimers to pentamers) formed during the simulations. The lifetimes of interactions between PMB1 molecules ranged from short periods (a few nanoseconds) to longer term interactions (200-400 ns) leading to formation of aggregates, as shown in the example in Fig. 3A-B. A range of different configurations were observed during the simulations. The majority of the interactions occurred *via* the hydrophobic portions of PMB1, namely Lip1, DPhe7, and Leu8, while the charged sites remained largely exposed to water and ions (Fig. 3A-B). In the example of a tetrameric association as shown in Fig. 3A, four of the PMB1 molecules had their Lip1 tails buried in the middle of the aggregate along with two DPhe7 and three Leu8 moieties, thus forming a structure that resembled a micelle. Due to the exposure to the aqueous environment of the positive charges and polar residues in this tetramer, the surface of the micelle-like structure was decorated by Cl^- ions, which interacted mostly with the NH_3^+ groups from Dab residues. The center of the micelle was mostly protected from exposure to water (Fig. 3D), with only one constant water molecule present at 0.5 nm (Fig. S3). This self-assembly behavior has previously been reported for other similar amphiphilic antibiotics, such as colistin and colistin methanesulfonate (CMS), but shown not to occur for the non-amphiphilic polymyxin B nonapeptide, an analogue that lacks the hydrophobic tail[39]. In the cases previously reported, aggregate

diameters were calculated to have a z-average of around 2 nm ± 0.3, which correlates well with the tetrameric aggregate observed in our simulations (2.2 nm ± 0.5). Thus, as predicted for colistin and its analogue[39], PMB1 micelle formation followed a "closed association" model, where the number of monomers per micelle does not exceed five in our simulations. In the PMB_{dil}, PMB_{prot}, and PMB_{crowd} systems, interactions between PMB1s resulted in smaller aggregates, generally involving dimerization (but with the additional participation of other molecular species, as discussed in the next section).

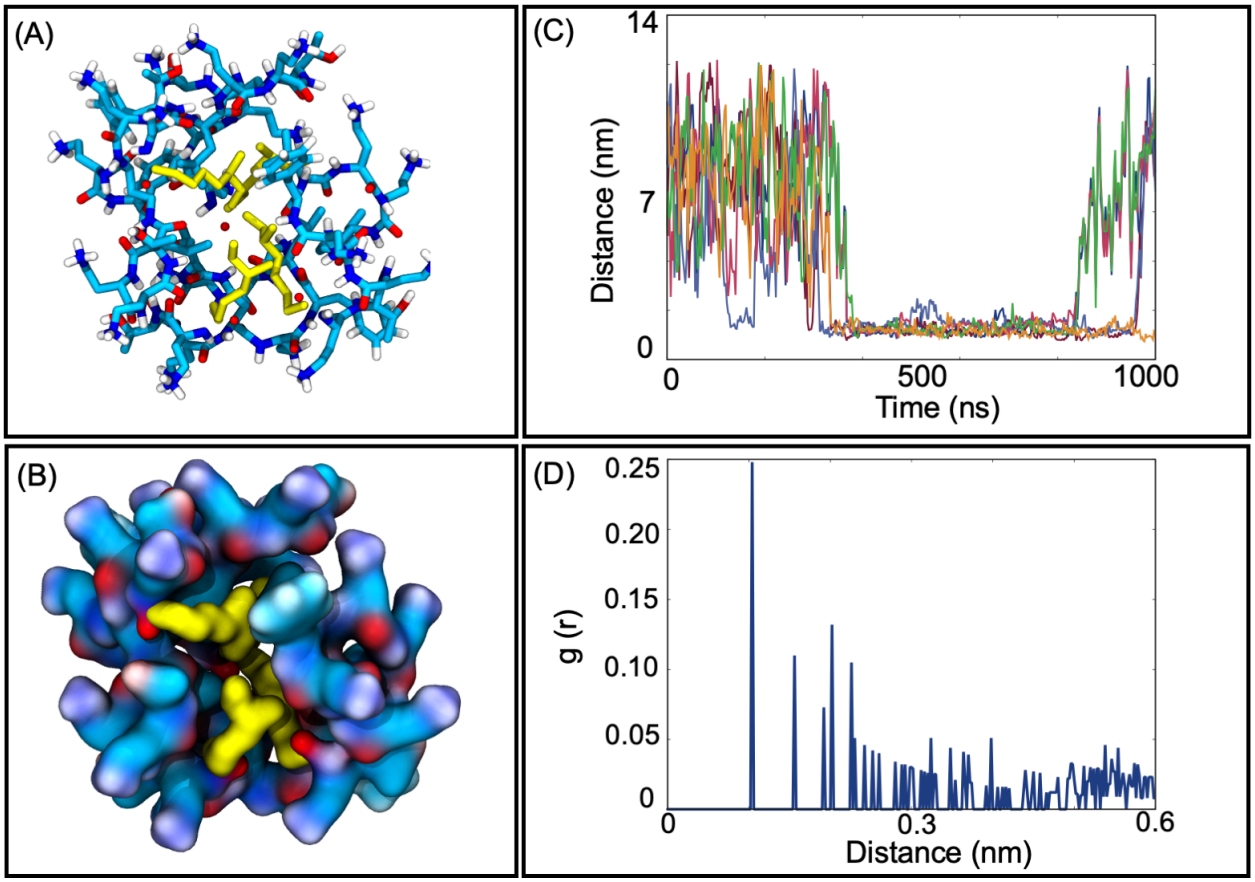


Figure 3: Micelle-like associations observed during the simulations. Sticks (A) and surface (B) representations of a tetrameric micelle, with the hydrophobic portions pointing inwards the aggregate colored in yellow. (C) Distances calculated between the center of mass (COM) of each monomer composing the micelle structure. Each curve describes the distance between two different monomers, with values below the 2 nm threshold indicating an association. (D) Radial Distribution Function (RDF) for water molecules calculated using the COM of the whole aggregate as a reference point.

PMB1 interaction with osmolytes

The interaction of PMB1 with osmolytes and ions was firstly characterized by measuring the proximity of each type of osmolyte to PMB1 molecules. The radial distribution function (RDF) of each osmolyte with PMB1s as a reference (Fig. 4) showed a clear preference for glycerol and OPG. This is reasonable considering the number of polar groups available for interactions on both osmolytes and the negative charge (-1 *e*) on the phosphate group of OPGs. Putrescine, spermidine and Na^+ ions were found furthest from PMB1, which correlates with both being positively charged (putrescine = +2 *e*, spermidine = +4 *e*).

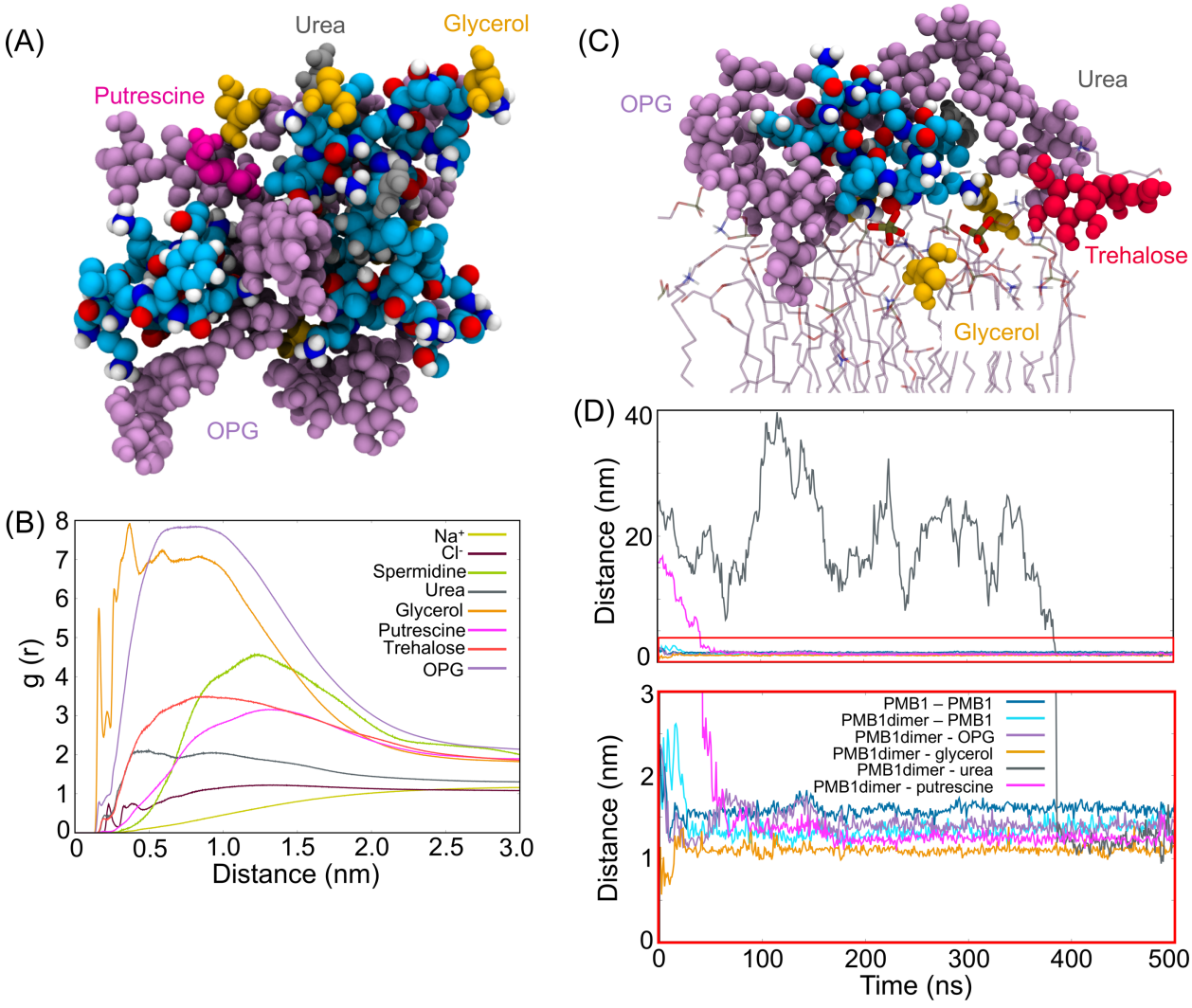


Figure 4: Osmolyte distribution and cluster formation in PMB_{crowd}. (A) Cluster formed by three PMB1 molecules (cyan, white, red, blue), five OPG molecules (violet), four glycerol molecules (orange), two urea molecules (grey) and one putrescine molecule (magenta). (B) Radial distribution function (RDF) using PMB1 as a reference point with Glycerol (orange), OPG (violet) and Cl^- (maroon), Na^+ (yellow), Putrescine (magenta) and Spermidine (light green), urea (grey) and trehalose (red). (C) Cluster formed at the surface of the outer membrane involving one PMB1 molecule, three OPG molecule, two glycerol molecules, one urea molecule and one trehalose molecule (colors as in (A)). Phosphate groups (brown and red sticks) form salt bridge interactions with Dab6 and Dab10 residues of PMB1. (D) Distances between representative molecules forming clusters are shown in panel (A), with a zoomed-in area marked with a red rectangle. Colored curves correspond to PMB1-PMB1, PMB1dimer-PMB1, PMB1dimer-OPG, PMB1dimer-glycerol and PMB1dimer-urea.

It has been discussed previously[35,40-44] that in crowded cellular environments, non-specific binding occurs constantly, generating transient clusters that affect the structure and dynamics of the molecules in this environment. In the simulations performed here, we observed formation of small osmolyte-PMB1 clusters which had an average size ~2.5 - 3.0 nm (slightly larger than the PMB1 micelles described in the previous section). These clusters generally contained PMB1 monomers interacting directly with OPG (*via* -OH groups and cyclohexane rings) and glycerol (*via* -OH groups), although participation of other osmolytes such as putrescine and urea was also observed, but usually without these molecules directly interacting with PMB1. In particular, the association between PMB1 molecules and OPG was prevalent (as shown in the RDF in Fig 4). For example, in one case, four OPG molecules bound around the surface of a PMB1 dimer (Fig. 4A), while a fifth OPG molecule mediated the interaction between the PMB1 dimer and a third PMB1 molecule. Four additional molecules of glycerol, one putrescine and two urea molecules also participated in this cluster, effectively bridging the PMB1 dimer to the third PMB1 (Fig. 4D), stabilizing the complex. This cluster took ~ 100 ns to stabilize in terms of number of components, apart from one urea molecule which only joined the cluster

after 400 ns (Fig. 4D and 4E). The final cluster shape was achieved at around 420 ns and remained stable until the end of the simulation. The largest cluster in all simulations was ~ 4.2 nm in diameter and was composed of four PMB1 molecules and ~20 osmolytes (one trehalose, five putrescine, seven glycerol and seven OPG). In this cluster, only two of the PMB1 molecules are directly associated with each other, interacting *via* their DPhe7 residues. The formation of the cluster was initiated by many of the molecules binding to the cell wall (within 30 ns of the start of the simulation), while the full cluster had formed after ~100 ns and lasted for around 240 ns. Despite showing a higher preference for cluster formation in the cell wall area, a few aggregates were also observed on the surface of the inner leaflet of the outer membrane (Figure 4C). For example, in one cluster, one PMB1 molecule is surrounded by three OPG molecules, two glycerol molecules, one trehalose molecule and one urea molecule. Glycerol not only intermediates interactions between PMB1 and OPG, but also with 1-palmitoyl,2-cis-vaccenyl-phosphatidyl ethanolamine (PVPE) lipids, in this aggregate. PMB1 also interacts with PVPE *via* Dab6 and Dab10 - phosphate salt bridges. Additionally, the cluster was visited by two putrescine molecules – one remaining in the aggregate for 160 ns and the other for only 50 ns.

PMB1 interaction with proteins

The number of interactions between PMB1 molecules and proteins were calculated based on intermolecular contacts (distances < 0.4 nm) during the course of the simulations, values concatenated over all trajectories for each protein are provided in Table S1. Interactions of PMB1 were observed with all of the different proteins in the systems. We consider the lipoprotein carriers, LolA and LolB first. In PMB_{prot} and PMB_{crowd} PMB1 molecules were found interacting both near to and at the entrance to the hydrophobic cavities of both proteins (Fig. 5 and 6). It is a useful reminder here that the normal function of LolA and LolB requires the lipid tails of cargo lipoproteins to bind in their hydrophobic cavities. A number of different PMB1 to LolA/B binding events were observed in our simulations. For example, in the PMB_{prot} system three molecules of PMB1 were observed to interact with the entrance to the cavity of LolA simultaneously (Fig. 5A), throughout one of the 500 ns simulations, with one partially inserted into the cavity. The LolA residues

involved in theses interactions range from hydrophobic to charged: Trp49 (51%), Met51 (95%), Thr52 (69%), Gln53 (72%), Pro54 (46%), Asp55 (45%), Phe72 (32%) and Glu74 (25%), where parentheses indicate percentage of simulation time for which the interactions existed, reflecting the chemical diversity of PMB1. Interestingly, in PMB_{prot}, one molecule of PMB1 bound to the entrance of the LolB cavity with its Lip1 residue inserted into the cavity (Fig. 6). Due to the extended conformation adopted by this PMB1, several LolB residues participated in long-lasting interactions (more than 60% of total simulation time) (Fig. 6B-C) with PMB1, including residues previously predicted as important for the binding of acyl chains[45], namely Phe37, Val46, Met107 and Ile109. This indicates that PMB1 can bind in the LolB cavity in a manner that resembles normal the binding of acyl chains of lipoproteins to LolB.

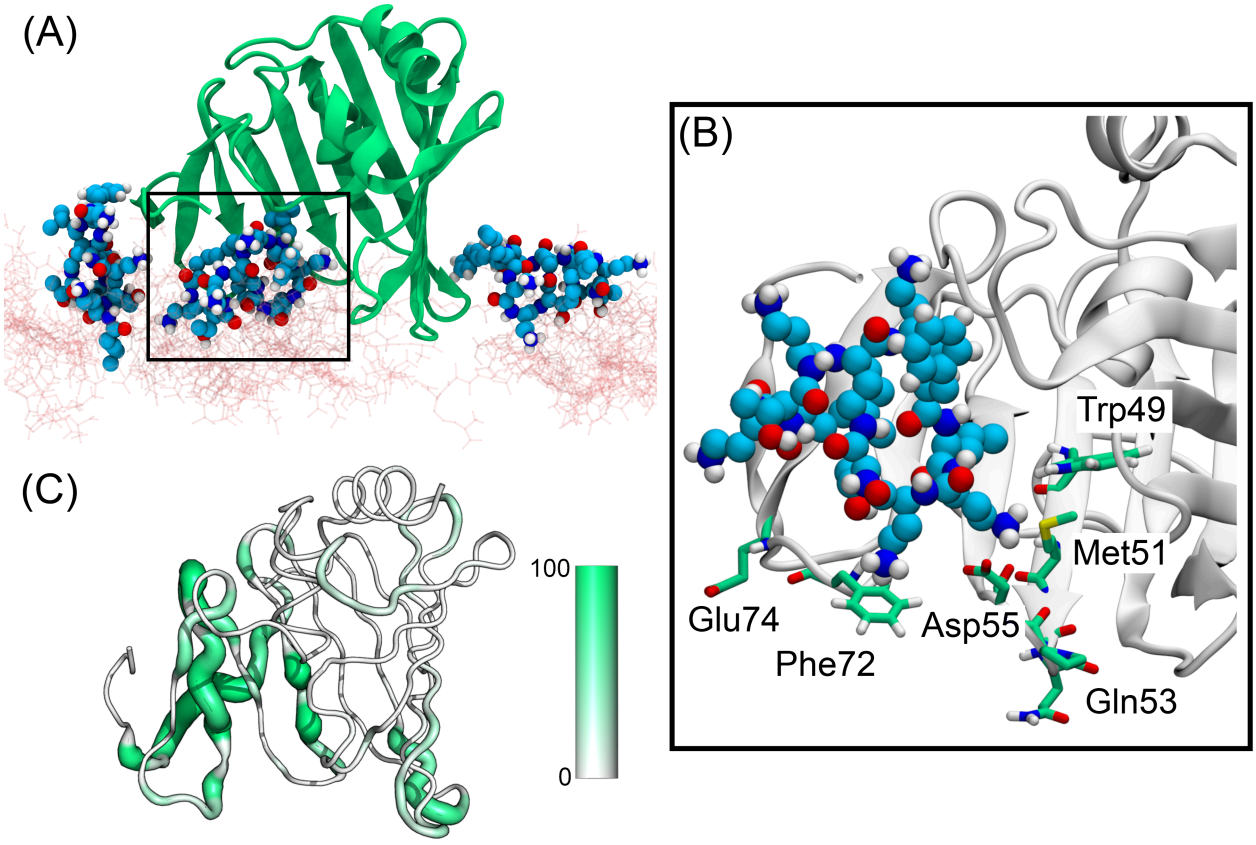


Figure 5: PMB1 binding modes to LolA. In PMB_{prot}, three PMB1 bind to LolA at the same time (A) LolA = green, PMB1 = as previously, cell wall = pink sticks). One PMB1 is partially inside the hydrophobic cavity. (B) Zoomed in region where PMB1 is bound to the hydrophobic cavity of LolA, interacting mainly with residues Trp49,

Met51, Gln53, Asp55, Phe72, and Glu74. (C) Sausage representation of LolA with respect to PMB1 interactions. Regions of the protein with higher percentage of time spent interacting with PMB1 are shown as enlarged tube, while regions with fewer interactions are shown as narrower tubes.

Next we consider PMB1 - Pal interactions. In both the $PMB1_{crowd}$ and $PMB1_{prot}$ simulations, PMB1 was observed sandwiched in between the C-terminal domain (CTD) of Pal, the linker domain and the outer membrane. The interactions lasted for the entirety of the simulations. Pal residues involved in the interactions are provided in Table S1. Furthermore, upon PMB1 binding, the motion of the linker region of Pal become more restricted (Fig. S4), but the initial non-covalent interaction of Pal with peptidoglycan seems to be unaffected (Fig. S5). Despite that, this binding to the linker appears to limit the increase in number of contacts between Pal and the cell wall, when compared to systems that do not have a PMB1 attached to the linker (Fig. S5). More transient PMB1-Pal binding events also occurred in simulations of each system. In PMB_{prot}, a PMB1 molecule entered the area in between the cell wall and the CTD of Pal (residues Ala109, Asp110, Arg112, Thr114, Tyr117 and Gly149), after initially being bound to a BLP for 370 ns (Fig. S6A). Other examples of PMB1 exchanging binding partners were also observed from Pal to LolA (Fig. S6B), from BLP to LolB (Fig. S6C), and from BLP to BLP (Fig. S6D). Thus, showing that within 500 ns PMB1 can move from interacting with one protein to another. In PMB_{crowd}, after intermittently interacting with Leu176 and Lys185 in the Pal CTD for 160 ns, one PMB1 moved slightly away from Pal and associated with another PMB1 forming a dimer (Fig. S7). While the same region of Pal that was previously bound to the original PMB1 formed an interaction with a small cluster containing one molecule of OPG, one molecule of Glycerol and one molecule of putrescine. This small cluster also simultaneously interacted with the newly formed PMB1 dimer, which at this stage was not directly interacting with Pal.

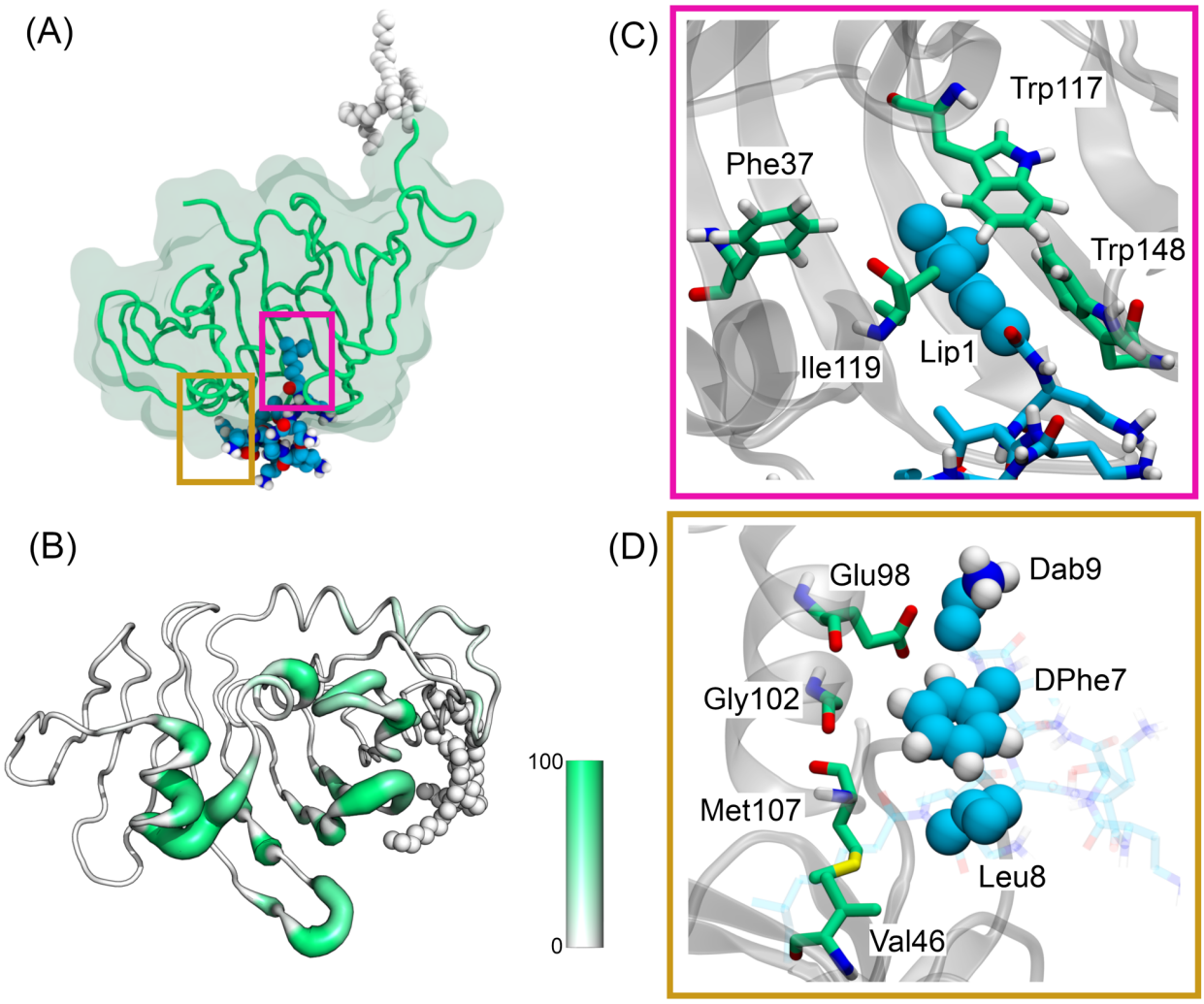


Figure 6: PMB1 binding mode to LolB. (A) PMB1 inserts Lip1 inside the hydrophobic cavity of LolB, reaching hydrophobic residues that usually interacts with acyl chains from lipoprotein ligands. (B) Sausage plot representation of the structure of LolB. Regions of the protein with higher percentage of time (from 0 to 1, coloured from white to green) spent interacting with PMB1 are shown as enlarged tube, while regions with less interactions are shown in thinner tubes. (C) Zoomed-in region showing binding at the hydrophobic cavity of LolB. In this area, Lip1 interacted with Phe37, Ile109, Trp117, and Trp148. (D) Zoomed-in region showing binding at the exterior part of LolB. In this region DPhe7, Leu8 and Dab9 interacted with residues Val46, Glu98, Gly102 and Met107.

PMB1 interactions with OmpA were mainly with the CTD. Each system had two or three PMB1 molecules binding to OmpA simultaneously. At least one molecule in each system was bound at the interface between the OmpA CTD and the cell wall, mediating the interaction between both structures. This binding region was located between the two main helices (composed of residues Glu212 to Asn226 and Ser253 to Lys267) from the CTD, with a prominent role of residues Gln214 (92%) and Tyr263 (87%). Interactions were hydrogen-bonding (Dab9 - Gln214) and hydrophobic (DPhe7 and Leu8 – Tyr263) in nature. Pal and OmpA have some structural similarities in their C-terminal domains (similarity of 35%), and both are known to bind to the cell wall[46-48]. Analysis of the contact data between PMB1-Pal and PMB1-OmpA revealed two long-lived interactions involving a specific helix from the CTD of each protein. This helix is composed of residues H_{112}ANFLRSNPS$_{122}$ in Pal and Y$_{244}$SQLSNLDP$_{252}$ in OmpA. Interestingly this region forms part of the dimerization interface of OmpA, thus would only be available for interaction when the protein is in its monomeric state[16,48]. Additionally, PMB1 was observed to bind to these regions while simultaneously interacting with adjacent motifs (Fig. S8) in both proteins: Q_{63}MQQLQ$_{68}$ in Pal (a short helix) and K_{290}GIPADKIS$_{298}$ (a loop connecting an α-helix to a β-strand). Interestingly only one PMB1 across all simulations was observed binding directly to the linker region of OmpA (in PMB_{prot}), in PMB_{crowd}, the linker area is largely occupied by osmolytes.

Comparison of data from PMB1 binding to BLP and Pal, revealed a short motif comprising the sequence S-S-E/D/N-X-Q/N (Fig. S9). Serine and acidic residues have particular propensity to interact with PMB1 (Fig. S10) due to the possibility of interacting *via* hydrogen bonding or salt bridges. Interactions between these residues are PMB1 lasted from 20 ns up to 400 ns. In simulations containing proteins, the four BLPs were the main target of binding events, with an average of 6 molecules of PMB1 binding to the four BLPs per system simulated. Regions of longest interactions (over 70% of simulation time) include residues S_{11}SDVQTLNA$_{19}$ (at the vicinity of the cell wall) and Asp34, Asp41, Ala42, Ala43, Arg48 (adjacent to the outer membrane). A number of the molecules interacting with BLP were also inserted in the interface with either the cell wall or the outer membrane in all of the simulations. Interestingly, in the PMB_{crowd} system most of the PMB1s that interacted with BLP were located in the solution region. The interfacial

regions favoured by PMB1 in other systems (PMB_{dil} and PMB_{prot}) were occupied by osmolytes in PMB_{crowd}. Thus, the presence of osmolytes seems to force PMB1 into the solution to some extent, by occupying interfacial binding regions.

PMB1 interactions with the cell wall

A number of PMB1 molecules (at least 14) reached the peptidoglycan layer area very rapidly (within the first 10 ns) in all simulations (Fig. S1 and S11). The negative charges presented by D-Glu and meso-diaminopimelate (m-DAP) residues from the peptide portion of the cell wall interact with PMB1. Other interactions are also present, such as PMB1 forming hydrogen bonds to the hydroxyl groups attached to the pyranosidic rings of the peptidoglycan glycan strands (Fig. S12). None of the PMB1 molecules were observed to go through the pores of the cell wall and dissociate from it during the total time of 3 μs of all simulated systems. Most osmolytes also did not cross through the pores easily: only putrescine, trehalose and urea were able to cross multiple times (Fig. S13). Generally, two modes of association between PMB1 and the cell wall were observed. In one mode, PMB1 inserts itself between glycan strands, as seen in one example from PMB_{dil} system (Fig. 7A). In some cases, it acted similarly to a peptide linkage, as it was able to form salt bridge interactions with both glycan strands simultaneously for more than 200 ns (Fig. 7C), decreasing the local distance between the strands from ~2.7 nm to ~1.9 nm. The other observed mode of association is PMB1 attaching to the surface of the cell wall, not inserted between the strands, but rather located around peptide linkages (Fig. 7B). A common aspect from both binding modes is that the Dab residues from PMB1 attract the loose peptide portions (not connected to > 1 glycan strand), forming salt bridges. The insertion mode of interaction involved the formation of multiple long-lived (~ 200 ns) salt bridges with the Dab - m-DAP and Dab - D-Glu, in contrast, in the surface binding mode the salt bridges had a lifetime of ~ 100 ns. (Fig. 7D).

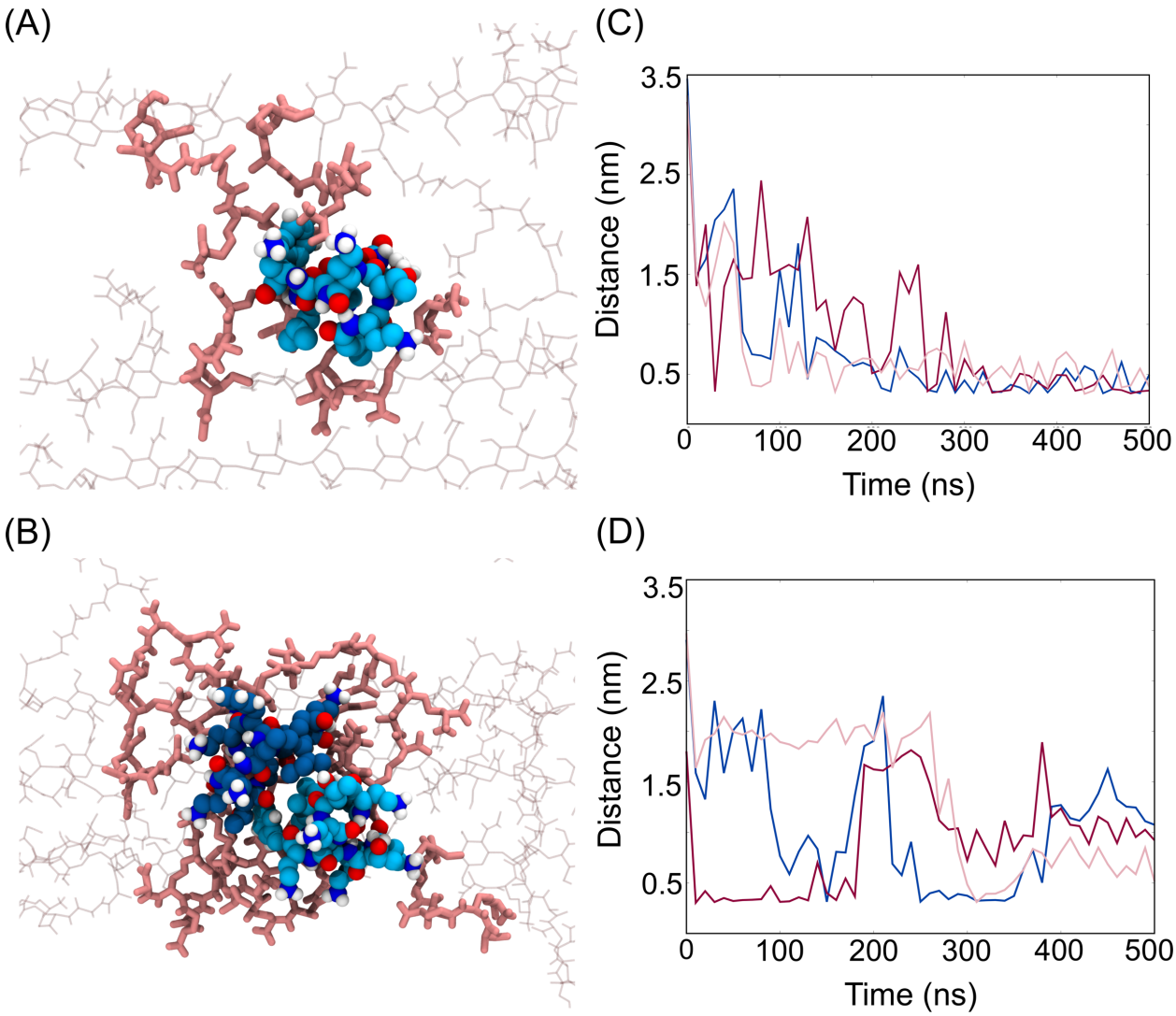
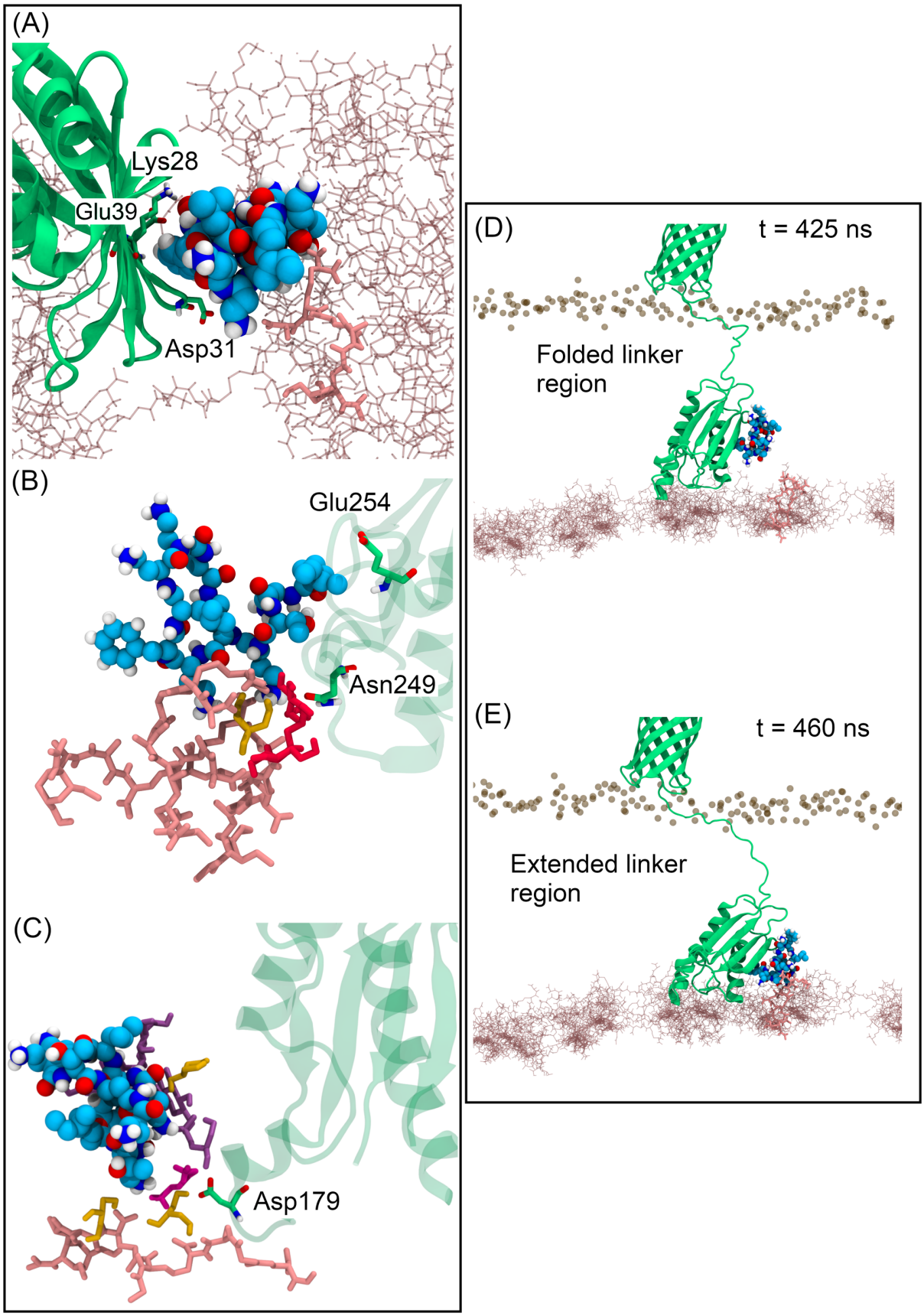


Figure 7: PMB1 binding modes to the cell wall extracted from PMB_{dil} and PMB_{prot}. The inserted binding mode is depicted on (A), showing PMB1 (blue carbon spheres) attached inside of the pores of the cell wall (pink sticks) and interacting with many negatively charged residues. The surface binding mode is depicted in (B), showing a PMB1 dimer (two shades of blue) bound to the surface of three glycan strands. (C) and (D) shows examples of distances of salt bridges interactions between different Dab residues (PMB1) and m-DAP and D-Glu (cell wall). In the inserted binding mode, interactions seem to last longer overall than in the surface binding mode.

In the PMB_{prot} system, a few of the PMB1 molecules mediated protein binding to the cell wall (Fig. 8). For example, PMB1 bound to the surface of LolA, interacting with residues Lys23, Asp26 and Glu34, while at the same time interacting with m-DAP and D-Glu from the cell wall. An observation from both simulations of the PMB_{prot} system, was an extension of the OmpA linker region that appeared to be induced by PMB1 which is bound to the CTD region. By around 420 ns, there is an increase in the number of residues from OmpA contacting the cell wall, increasing from 5 to 10 (Fig. S14). This appeared to occur spontaneously, but subsequently the PMB1 Dab residues form salt bridges with three charged groups (two D-Glu and one m-DAP) of the loose peptide ends of the cell wall (Fig. 8 and S14). These interactions of PMB1 which is bound to both OmpA and the cell wall, seemed to have acted as a driving force to push the OmpA CTD further towards the cell wall, stabilizing and increasing the OmpA-peptidoglycan binding interface from 10 residues to almost 20. Similarly, in the PMB_{crowd} system, some associations between the PMB1, proteins (OmpA, LolA and Pal) and the cell wall were also observed. For example, initially, a PMB1 is associated with the OmpA CTD mediated by two trehalose molecules and one glycerol molecule (Fig. 8). At the same time, the same PMB1, one of the trehalose, glycerol and residue Asn249 of OmpA were in the vicinity of the cell wall (within 0.6 nm), displaying a few hydrogen bonding interactions. By the end of the simulation, this cluster had dissociated. Another example is the PMB1-Pal association mentioned in the previous section with the presence of a small cluster of osmolytes (Fig. 8). Residue Asp179 of Pal interacted with OPG, glycerol, and putrescine molecules while they were bound to PMB1 and/or the cell wall, the interaction partners in this cluster changed frequently over time, indicating the non-specificity of these intermolecular associations.



455

Figure 8: PMB1 mediating interactions between LolA, Pal, OmpA (green) and the cell wall (pink sticks). Panel (A) shows highlighted residues Asp26, Lys23 and Glu34 (green) from LolA interacting with Dab (Asp26) and DPhe7 (Lys23 and Glu34,) from PMB1 (colored in blue spheres), while another Dab from PMB1 interacts with a negative charged residue from the cell wall (pink licorice sticks). In Panel (B), PMB1 and osmolytes (glycerol – orange; trehalose – red) are shown mediating interactions for OmpA-cell wall. In (C), a bigger cluster containing OPG, putrescine, glycerol and urea is shown mediating interactions with Pal and the cell wall. Panels (D) and (E) depict two different states of OmpA and PMB1, before (D) and after (E) enhancing the interaction with the cell wall. In (B) and (C), only cell wall residues that are in contact with PMB1 and osmolytes are shown to improve clarity.

Discussion

Currently there is a great need to find novel therapeutic agents to address the problem of antimicrobial resistance to antibiotics[49,50]. In order to do so in a rational manner requires a thorough understanding of the environment faced by antibiotics such as PMB1 as they negotiate the bacterial cell envelope. In this work, we have simulated an atomistic model of the periplasmic space to study the fate of PMB1 in this region once it has already crossed the OM. Our results predict that PMB1, and likely other drugs relying on diffusion alone to cross the periplasm, face a complex path, full of molecular obstacles which hinder their movement through the periplasm. The presence of structures from subcellular compartments in our model systems had a major effect on the diffusion coefficients of PMB1 molecules. This was observed by the 100-fold reduction in values in the periplasm models compared to PMB1 just in solution. Similarly, diffusion of native proteins was also affected by increased crowding. Previous studies of diffusion rates in the cytoplasm and the periplasm showed that the GFP proteins have a slightly slower diffusion rate in the periplasm[38] when compared to the cytoplasm and there have been some discussions in the literature regarding the nature of the periplasm being a gel-like environment[51] or a fluid environment[38]. At the level of crowding we have in our simulations the system is clearly still fluid. Crowding up to 30% of volume has been reported previously to have

modest effects on water properties[34], showing alterations in the self-diffusion coefficient of water that are in-line with our simulations (~20% Volume with ~2.0 x 10^{-5} cm^2/s).

PMB1 is an amphipathic molecule with considerable conformational flexibility. In just water and ions, we observed PMB1 tetramers forming micelles with sizes comparable to experimental data[30] for micelle formation in solution of colistin (polymyxin E). In addition, as predicted for colistin and its analogue[39], our micelle formation followed a "closed association" model, in which the number of monomers is discrete, limiting micelles to a certain size (pentamers with 2.6 nm). Thus, our model of PMB1 in water provided a 'baseline' reference system which gave aggregate sizes and dynamical behavior that reproduced experimental observables for similar molecules.

We observed a wide range of associations of PMB1 with other molecules. One particularly interesting phenomenon observed here is PMB1 insertion into the hydrophobic cavities of the lipoprotein carriers LolA and LolB. These cavities have previously been shown to be non-specific binders of hydrophobic molecules[52-54]. Our results suggest it is possible that some PMB1s may be carried through the cell wall by hijacking the lipoprotein carrying functionalities of LolA/LolB. Given we observe PMB1 adhering to the cell wall this 'hitchhiking' mechanism would be advantageous in providing an easier route through the cell wall. To our knowledge this spontaneous phenomenon is described here for the first time. It is important to consider here other potential consequences of PMB1 binding to the lipoprotein carriers. LolA and LolB play important roles in avoiding toxicity due to accumulation of BLP in the inner membrane[55,56]. Binding of PMB1 into their hydrophobic cavities may serve to inhibit their natural functions and lead to mislocalization of lipoproteins in the inner membrane, in a similar manner to small hydrophobic inhibitors such as MAC13243[52-54]. Interestingly, LolA transcription is triggered with increasing concentrations of PMB, a mechanism connected to the activation of the stress regulator Rcs phosphorelay system[57] which provides indirect evidence to support our hypothesis.

Osmolyte - PMB1 interactions varied depending upon the chemistry of the osmolyte. We observed fast formation of small clusters of molecules, with PMB1 usually binding to, and often becoming partially coated with the polar OPG and glycerol molecules. The OPG concentration becomes slowly diluted when bacteria are moved to concentrated media[58]. Given the high propensity for these molecules to bind to PMB1 in our simulations, we

518 suggest that the changes in the OPG concentration may also impact on dynamics of
519 PMB1 in the periplasm. As discussed in previous simulation studies[34,35,37], molecular
520 crowders can promote a range of effects, including excluded volume effects and replacing
521 interactions. In our simulations, osmolytes mediate interactions between other molecules
522 and also replace some interactions. For example in the absence of osmolytes, there is a
523 greater propensity for PMB1-BLP interactions to occur at the BLP/cell wall and BLP/OM
524 interfaces, however these regions are occupied by osmolytes in the most crowded
525 system, and consequently PMB1 interactions with BLP are largely with the region of the
526 protein in 'bulk' solution in the periplasm. This suggests that the non-specific binding of
527 osmolytes to cell envelope components may have local consequences for available
528 binding modes for antibiotics.
529 Two main peptidoglycan binding modes were observed; one in which PMB1 inserts
530 in between glycan strands acting as a pseudo cross-link and one in which it is surface-
531 bound close to the peptide linkages. Atomic Force Microscopy (AFM)[59] studies have
532 suggested that the cell wall interactions formed by colistin may be responsible for
533 rigidifying the cell envelope.
534 A limitation of our results involves the correction of the diffusion coefficients for
535 finite-size effects. Given the complexity of our models, with the presence of the outer
536 membrane and the cell wall, that our simulation boxes are non-cubic, and that the
537 correction becomes smaller with increasing box size, we opted for not applying it in these
538 D_t values. Our results for proteins and water diffusion correlate well with previous
539 atomistic crowding models[34-37]. Experimental studies[38,60] reported D_t rates for GFP that
540 are in a similar range to ours. We note here that experimental value for the OmpA N-
541 terminal domain (transmembrane) alone, $D_{exp} = 4.9 \pm 0.09 \times 10^{-7}$ cm^2/s is faster than the
542 value obtained for the complete protein from our simulations ($D_t = 3.1 \pm 0.6 \times 10^{-8}$
543 cm^2/s in PMB_{prot} and $D_t = 1.8 \pm 0.5 \times 10^{-8}$ cm^2/s in PMB_{crowd}). The C-terminal domain of
544 OmpA is bound to the peptidoglycan which is highly likely to be the cause of the slower
545 diffusion in the simulations.
546 It is also worth mentioning that the SPC water model (which works well with the
547 GROMOS54a7 force field[61]) overestimates experimental values of self-diffusion for water
548 molecules[62], so D_t values for solutes possibly are affected by this. Finally, crowding

549 systems simulations are very complex and could lead to intense aggregation when using
550 standard additive force fields[63,64], with several methods being proposed to solve this[65,66].
551 From our perspective, GROMOS54a7 was a reasonable choice, since it is one of the
552 force fields that shows a lesser preference for the aggregated state[70], while also having
553 validated parameters for the complex mixture of lipids that compose the bacterial outer
554 membrane and the bacterial cell wall. Our results where comparable with other
555 experimental and simulation studies are in-line with those, providing further confidence in
556 the predictions from the complex simulations which go beyond previous studies in terms
557 of the resolution and complexity of the simulations studied.
558

559 # Conclusions

560 In conclusion, atomistic molecular dynamics simulations of the antibiotic PMB1 in a
561 number of models of the periplasm of Gram-negative bacteria with differing levels of
562 crowding have revealed slower diffusion of the antibiotic as the periplasm becomes more
563 crowded. PMB1 forms complexes with osmolytes, the cell wall and native cell envelope
564 proteins which can be short-lived or long-lived. PMB1 is rarely uncomplexed within the
565 periplasm, therefore its functional groups are occupied in interaction with other species
566 more often than not. We feel this is an important factor to consider in future development
567 of antibiotics (and may be extended to drugs that target other organisms too). The *in vivo*
568 environment is not a chemistry experiment in which ones controls the type and number
569 of molecules involved, and the complexity of the former may impact upon foreign
570 molecules such as drugs in many unexpected ways. The simulations described here
571 show that incorporation of the chemical details of the local environment can predict likely
572 interactions with other species and highlight potential mechanistic pathways that may
573 have been originally unintended (such as the ability of PMB1 to bind to LolA and LolB).
574

575 # Methods

576

577 ## System Preparation

We constructed the template model based on previously published works from our group[17,48,67]. This was composed by an asymmetric outer membrane (OM) of an identical composition as seen in previous works[17,67-69], a one-layer peptidoglycan cell wall (PGN) formed by 12 glycan strands of 17 repeating NAG-NAM-peptide units, four Braun Lipoproteins (BLP) covalently attached to PGN and inserted in the membrane by tripalmitoyl-S-glyceryl-cisteine residues[67,69]. Models for BLP(PDB: 1EQ7)[70], LolA (PDB: 1IWL) and LolB (PDB: 1WLM)[45], OmpA[16], Pal (PDB: 2W8B)[71] were taken from previous works from the group[72,48,54,67]. Pal and OmpA proteins were also included in PMB_{prot} and PMB_{crowd} systems. Crystallographic structures of LolA and LolB GROMOS 54A7 force field with the GROMOS 53A6OXY ether parameters[73] were used for the construction of the tripalmitoyl-S-glyceryl-cysteines. After setting up this initial template system, we added the other components: Pal bound and unbound to the cell wall, PMB1 molecules, and ions concentrations. PMB1 GROMOS 54a7 parameters were obtained by using the Automated Topology Builder (ATB) server[74]. For OmpA insertions into the OM, we employed the *gmx membed* tool[75], similarly to a previous report[67].

For the construction of the PMB_{crowd} system, parameters for the molecular crowders were obtained using the ATB server, with the exception of OPG and trehalose, in which the GROMOS 56a6 (CARBO)[76] parameters were employed, which are compatible with GROMOS 54a7. We adapted the “droplet methodology” from Bortot *et al*[37] to deal with the insertion of osmolytes, by adding each osmolyte with a water shell obtained from 100 ns molecular dynamics simulations.

Atomistic Molecular Dynamics Simulations

We performed molecular dynamics simulations employing the GROMACS simulation suite (version 2018.3)[77] along with GROMOS54a7 force field[61] and SPC water model[33]. We divided the simulations in two parts: equilibration simulations in NVT and NPT ensembles with position restraints in proteins, cell wall, and PMB1, which lasted for 1 and 100 ns, respectively; and production simulations in NPT ensemble, which ran for 500 ns. Simulations were performed at 310 K, which was maintained by employing the velocity rescale thermostat[78] with a coupling constant of $\tau = 0.1$. Pressure was maintained semi-isotropically at 1 atm by employing the Parrinello-Rahman barostat[79] with a time constant

of 2 ps. The particle mesh Ewald method treated long-range electrostatics[80]. LINCS algorithm[81,82] constrained the covalent bonds, which allowed an integration step of 2 fs. Both long-range electrostatics and van der Waals cutoffs were set to 1.4 nm. To neutralize charges, we added the correct number of counterions together with an extra salt concentration of 0.2 M of sodium chloride ions for all simulations. For the replicates, starting positions of the proteins and PMB1 molecules were changed, along with re-solvation of the system, equilibration and production phases. In addition, we modified the starting velocities to ensure the difference between runs and improve conformational sampling. For molecular manipulation, visualization, and analysis, we employed the VMD software[83].

Analysis

Translational diffusion coefficients, D_t, were obtained by using the *gmx msd* analysis tool from the GROMACS tool set to calculate the mean square displacements (MSD). MSD plots (Figure S2) were calculated for time delays The linear fit (where the slope was obtained) was performed in different time regimes (1-10 ns; 50-100 ns) for PMB1 and water (1-10 ns; 10-100 ns) aiming to capture the slowdown in D_t due to crowding effects, while for proteins, the linear fit was performed in the range of 5-15 ns. Error estimates were obtained by averaging over all and by averaging over replicate simulations of each system. For the XYZ motion analysis, the *gmx trajectory* tool was employed to obtain the coordinates of the center of mass of each PMB1 in the X, Y and Z axis. Radial distribution function values (RDF) were obtained using the *gmx rdf* tool, while *gmx sasa* was used for SASA calculations. In-house scripts were employed for the intermolecular contact analysis.

ASSOCIATED CONTENT
Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Further analysis of the data including Figures S1-S14 and Table S1 (PDF file).

641

642 Author Contributions

643 The manuscript was written through contributions of all authors. SK conceived the project,
644 CP conducted the simulations, IPSS and AB helpeed with simulation setup and analysis,
645 CP and SK also performed analysis, CP and SK wrote the paper. All authors have given
646 approval to the final version of the manuscript.

647

648 Acknowledgements

649 The authors acknowledge the use of the IRIDIS High Performance Computing Facility,
650 and associated support services at the University of Southampton, in the completion of
651 this work. This project made use of time on ARCHER granted via the UK High-End
652 Computing Consortium for Biomolecular Simulation, HECBioSim (http://hecbiosim.ac.uk),
653 supported by EPSRC through grant number EP/R029407/1, which also supports CP.

654

655 References

656 1. Weiner JH, Li L. 2008. Proteome of the Escherichia coli envelope and technological
657 challenges in membrane proteome analysis. Biochim Biophys Acta - Biomembr
658 1778:1698–1713.
659 2. Goemans C, Denoncin K, Collet J-FF. 2014. Folding mechanisms of periplasmic
660 proteins. Biochim Biophys Acta - Mol Cell Res 1843:1517–1528.
661 3. Vaara, M. 2019. Polymyxins and Their Potential Next Generation as Therapeutic
662 Antibiotics. Front. Microbiol. 10, 1689.
663 4. Morrison, D. C., Jacobs, D. M. 1976. Binding of polymyxin B to the lipid A portion of
664 bacterial lipopolysaccharides. Immunochemistry. 13, 813-818.
665 5. Bader, J., Teuber, M. 1973. Action of polymyxin B on bacterial membranes, I. Binding to
666 the O-antigenic lipopolysaccharide of Salmonella typhimurium. Z Naturforsch [C] 28,
667 422-430.
668 6. Evans, M. 1999. Polymyxin B sulfate and colistin: Old antibiotics for emerging
669 multiresistant Gram-negative bacteria. Ann. Pharmacother. 33, 960–967.
670 7. Trimble, M. J., Mlynarcik, P., Kolar, M. & Hancock, R. E. W. 2016. Polymyxin: alternative
671 mechanisms of action and resistance. Cold Spring Harb. Perspect. Med. 6, a025288.
672 8. Gales, A. C., Jones, R. N., Sader, H. S. 2011. Contemporary activity of colistin and
673 polymyxin B against a worldwide collection of Gram-negative pathogens: results from
674 the SENTRY Antimicrobial Surveillance Program (2006–09). J. Antimicrob. Chemother.
675 66, 2070–2074.

9. Li, Z., Cao, Y., Yi, L., Liu, J. H., Yang, Q. 2019. Emergent Polymyxin Resistance: End of an Era? Open Forum Infec. Dis. 6, ofz368.
10. Li Z & Velkov T. 2019. Polymyxins: Mode of Action. Adv. Exp. Med. Biol. 1145, 37-54.
11. Vollmer, W. & Holtje, J. V. 2004. The architecture of the murein (peptidoglycan) in gram-negative bacteria: vertical scaffold or horizontal layer(s)? J Bacteriol. 186, 5978–5987.
12. Hirota Y, Suzuki H, Nishimura Y, Yasuda S. 1977. On the process of cellular division in Escherichia coli: A mutant of *E. coli* lacking a murein-lipoprotein. Proc. Natl. Acad. Sci. USA. 74 (4), 1417-1420.
13. 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. EMBO J. 14, 3365–3372.
14. Matsuyama SI, 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*. EMBO J. 16, 6947–6955.
15. Smith SG, Mahon V, Lambert MA, Fagan RP. 2007. A molecular Swiss army knife: OmpA structure, function and expression. FEMS Microbiol Lett. 273 (1), 1-11.
16. Marcoux J, Politis A, Rinehart D, Marshall DP, Wallace MI, Tamm LK2, Robinson CV. 2014. Mass spectrometry defines the C-terminal dimerization domain and enables modeling of the structure of full-length OmpA. Structure. 22 (5), 781-90.
17. Boags, A., Samsudin, F., Khalid, S. 2019. Binding from both sides: TolR and OmpA bind and maintain the local structure of the *E. coli* cell wall. Structure 27, 713-724
18. Mizuno T. 1979. A novel peptidoglycan-associated lipoprotein found in the cell envelope of Pseudomonas aeruginosa and Escherichia coli. J Biochem. 86 (4), 991-1000.
19. Lazzaroni JC, Portalier R. 1992. The excC gene of Escherichia coli K-12 required for cell envelope integrity encodes the peptidoglycan-associated lipoprotein (PAL). Mol Microbiol. 6 (6), 735-42.
20. Bontemps-Gallo S, Lacroix JM. 2015. New Insights into the Biological Role of the Osmoregulated Periplasmic Glucans in Pathogenic and Symbiotic Bacteria. Environ Microbiol Rep. 7 (5), 690–697.
21. Ruhal R, Kataria R, Choudhury B. 2013. Trends in bacterial trehalose metabolism and significant nodes of metabolic pathway in the direction of trehalose accumulation Microb. Biotechnol. 6(5), 493–502.
22. Wortham BW, Oliveira MA, Patel CN. 2007. Polyamines in Bacteria: Pleiotropic Effects Yet Specific Mechanisms, Adv Exp Med Biol, 603, 106-15.
23. Murarka A, Dharmadi Y, Yazdani SS, Gonzalez R. 2008. Fermentative Utilization of Glycerol by Escherichia coli and Its Implications for the Production of Fuels and Chemicals. Appl Environ Microbiol. 74 (4), 1124–1135.
24. Martínez-Gómez K. et al. 2012. New insights into Escherichia coli metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol. Microb Cell Fact. 11:46.
25. Beckers G, Bendt AK, Krämer R, Burkovski A. 2004. Molecular Identification of the Urea Uptake System and Transcriptional Analysis of Urea Transporter- and Urease-Encoding Genes in Corynebacterium glutamicum. J Bacteriol. 186 (22), 7645–7652.

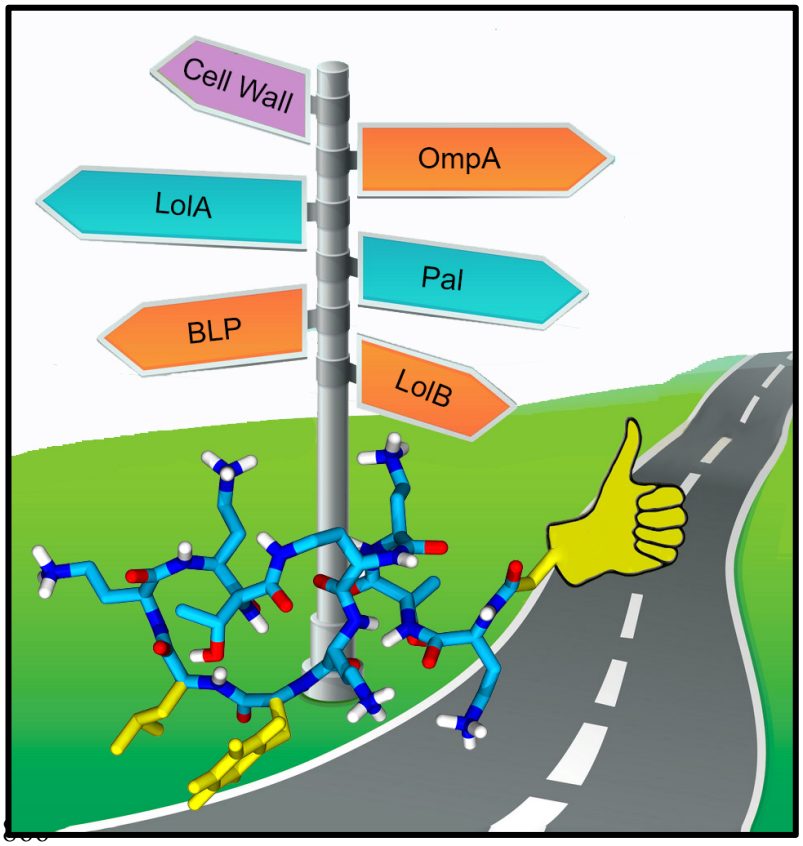
26. Cayley DS, Guttman HJ, Record Jr MT. 2000. Biophysical Characterization of Changes in Amounts and Activity of Escherichia Coli Cell and Compartment Water and Turgor Pressure in Response to Osmotic Stress. Biophys J, 78 (4), 1748-64.
27. Boos, W. Ehmann U, Bremer E, Middendorf A, Postma P. 1987. Trehalase of Escherichia coli. Mapping and cloning of its structural gene and identification of the enzyme as a periplasmic protein induced under high osmolarity growth conditions. J Biol Chem, 262 (27), 13212-8.
28. Cohen, S.S. 1997. A Guide to the Polyamines. Oxford University Press, New York, USA.
29. Shah P & Swiatlo E. 2008. A Multifaceted Role for Polyamines in Bacterial Pathogens. Mol Microbiol, 68 (1), 4-16.
30. Wang D, Weng J, Wang W. 2019. Glycerol transport through the aquaglyceroporin GlpF: bridging dynamics and kinetics with atomic simulation. Chem Sci, 10 (29), 6957-6965.
31. Krishnamurthy P. et al. 1998. Helicobacter pylori Containing Only Cytoplasmic Urease Is Susceptible to Acid. Infect Immun. 66 (11), 5060–5066.
32. Lima MPM, Nader M, Santos DES, Soares TA. 2019. Compatibility of GROMOS-Derived Atomic Parameters for Lipopolysaccharide Membranes with the SPC/E Water Model and Alternative Long-Range Electrostatic Treatments Using Single Nonbonded Cutoff and Atom-Based Charge Schemes. J. Braz. Chem. Soc. 30 (10), 2219-2230.
33. Berendsen, H. J., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. 1981. In Intermolecular Forces (ed. Pullman, B.) 331–342 (Reidel Publishing, 1981).
34. Harada R, Sugita Y, Feig M. 2012. Protein Crowding Affects Hydration Structure and Dynamics. J. Am. Chem. Soc. 134, 4842−4849.
35. von Bülow S, Siggel M, Linke M, Hummer G. 2019. Dynamic cluster formation determines viscosity and diffusion in dense protein solutions. Proc Natl Acad Sci U S A. 116 (20), 9843-9852.
36. Yu I, Mori T, Ando T, Harada R, Jung J, Sugita Y, Feig M. 2016. Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. Elife. 5, e19274.
37. Bortot LO, Bashardanesh Z, van der Spoel D. 2020. Making Soup: Preparing and Validating Models of the Bacterial Cytoplasm for Molecular Simulation. J. Chem. Inf. Model. 60, 322−331.
38. Mullineaux CW, Nenninger A, Ray N, Robinson C. 2006. Diffusion of green fluorescent protein in three cell environments in Escherichia coli. J. Bacteriol. 188 (10), 3442-3448.
39. Wallace et al. 2011. Self-assembly behaviour of colistin and its prodrug colistin methanesulfonate: implications for solution stability and solubilization. J Phys Chem B. 114 (14), 4836–4840.
40. Ando T, Skolnick J. 2010. Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. Proc Natl Acad Sci USA. 107, 18457–18462.
41. Minton AP. 1980. Excluded volume as a determinant of protein structure and stability. Biophys J. 32, 77–79.
42. Muramatsu N, Minton AP. 1989. Hidden self-association of proteins. J Mol Recognit. 1, 166–171.
43. Stradner A, et al. 2004. Equilibrium cluster formation in concentrated protein solutions and colloids. Nature. 432, 492–495.

763 44. Paulino J, et al. 2019. Influenza A M2 channel clustering at high protein/lipid ratios:
764 Viral budding implications. Biophys J. 116, 1075–1084.
765 45. Takeda et al. 2003. Crystal structures of bacterial lipoprotein localization factors, LolA
766 and LolB. EMBO J. 22 (13), 3199–3209.
767 46. Koebnik R. 1995. Proposal for a peptidoglycan-associating *α*-helical motif in the C-
768 terminal regions of some bacterial cell-surface proteins. Mol. Microbiol. 16, 1269–1270.
769 47. Parsons L.M., Lin F., Orban J. 2006. Peptidoglycan recognition by Pal, an outer
770 membrane lipoprotein. Biochemistry. 45, 2122–2128.
771 48. Samsudin F, Boags A, Piggot TJ, Khalid S. 2017. Braun's Lipoprotein Facilitates OmpA
772 Interaction with the Escherichia coli Cell Wall. Biophys. J. 113, 1496–1504.
773 49. Zaman, S. B. et al. 2017. A review on antibiotic resistance: alarm bells are ringing.
774 Cureus, 9 (6), e1403.
775 50. Laxminarayan, R. et al. 2013. Antibiotic resistance: the need for global solutions. Lancet
776 Infect. Dis. 13, 1057–1098.
777 51. Hobot JA, Carlemalm E, Villiger W, Kellenberger E. 1984. Periplasmic gel: new concept
778 resulting from the reinvestigation of bacterial cell envelope ultrastructure by new
779 methods. J Bacteriol. 160 (1), 143–152.
780 52. Pathania R. et al. 2009. Chemical genomics in Escherichia coli identifies an inhibitor of
781 bacterial lipoprotein targeting. Nat. Chem. Biol. 5, 849–856.
782 53. Barker, C. A. et al. 2013. Degradation of MAC13243 and studies of the interaction of
783 resulting thiourea compounds with the lipoprotein targeting chaperone LolA. Bioorg.
784 Med. Chem. Lett. 23, 2426–2431.
785 54. Boags A, Samsudin F, Khalid S. 2019. Details of hydrophobic entanglement between
786 small molecules and Braun's lipoprotein within the cavity of the bacterial chaperone
787 LolA. Sci. Rep. 9, 3717.
788 55. Grabowicz M, Silhavy TJ. 2017. Redefining the essential trafficking pathway for outer
789 membrane lipoproteins. Proc. Natl. Acad. Sci. U.S.A. 114, 4769–4774.
790 56. Caro F, Place NM, Mekalanos JJ. 2019. Analysis of lipoprotein transport depletion in
791 Vibrio cholerae using CRISPRi. Proc Natl Acad Sci U S A. 116 (34), 17013-17022.
792 57. Tao K, Narita SI, Tokuda H. 2012. Defective Lipoprotein Sorting Induces lolA Expression
793 through the Rcs Stress Response Phosphorelay System. J Bacteriol. 194 (14), 3643–
794 3650.
795 58. Bontemps-Gallo S, Bohin JP, Lacroix JM. 2017. Osmoregulated Periplasmic Glucans.
796 EcoSal Plus. 7 (2).
797 59. Mortensen, N. P. et al. 2009. Effects of colistin on surface ultrastructure and
798 nanomechanics of Pseudomonas aeruginosa cells. Langmuir. 25, 3728-3733.
799 60. Konopka MC, Shkel IA, Cayley S, Record MT, Weisshaar JC. 2006. Crowding and
800 confinement effects on protein diffusion in vivo. J Bacteriol. 188 (17), 6115-6123.
801 61. Schmid, N. *et al*. 2011. Definition and testing of the GROMOS force-field versions 54A7
802 and 54B7. Eur. Biophys. J. 40, 843–856.
803 62. van der Spoel D, van Maaren PJ, Berendsen HJC. 1998. A systematic study of water
804 models for molecular simulation: Derivation of water models optimized for use with a
805 reaction field. J. Chem. Phys. 108 (24), 10220-10230.

63. Petrov D, Zagrovic B. 2014. Are Current Atomistic Force Fields Accurate Enough to Study Proteins in Crowded Environments? PLoS Comput Biol. 10 (5), e1003638.
64. Abriata LA, Dal Peraro M. 2015. Assessing the potential of atomistic molecular dynamics simulations to probe reversible protein-protein recognition and binding. Scientific Reports. 5, 10549.
65. Best, R. B.; Zheng, W.; Mittal, J. 2014. Balanced protein−water interactions improve properties of disordered proteins and nonspecific protein association. J. Chem. Theory Comput. 10, 5113−5124.
66. Carballo-Pacheco, M.; Ismail, A. E.; Strodel, B. 2018. On the Applicability of Force Fields to Study the Aggregation of Amyloidogenic Peptides Using Molecular Dynamics Simulations. J. Chem. Theory Comput. 14, 6063−6075.
67. Samsudin F, Ortiz-Suarez ML, Piggot TJ, Bond PJ, Khalid S. 2016. OmpA: A Flexible Clamp for Bacterial Cell Wall Attachment. Structure 24, 2227–2235.
68. Piggot TJ, Holdbrook DA, Khalid S. 2011. Electroporation of the E. coli and S. Aureus membranes: molecular dynamics simulations of complex bacterial membranes. J. Phys. Chem. 115, 13381–8.
69. Ortiz-Suarez ML, Samsudin F, Piggot TJ, Bond PJ, Khalid S. 2016. Full-Length OmpA: Structure, Function, and Membrane Interactions Predicted by Molecular Dynamics Simulations. Biophys. J. 111, 1692–1702.
70. Shu W, Liu J, Ji H, Lu M. 2000. Core structure of the outer membrane lipoprotein from Escherichia coli at 1.9 A resolution. J. Mol. Biol. 299, 1101–1112.
71. Bonsor, D. A. et al. 2009. Allosteric beta-propeller signalling in TolB and its manipulation by translocating colicins. EMBO J. 28, 2846–2857.
72. Szczepaniak J. et al. 2020. The lipoprotein Pal stabilises the bacterial outer membrane during constriction by a mobilisation-and-capture mechanism. Nat Comm. 11, 1305.
73. Horta, B.A.C., Fuchs, P.F.J., van Gunsteren, W.F., and Hunenberger, P.H. 2011. New interaction parameters for oxygen compounds in the GROMOS force field: improved pure-liquid and solvation properties for alcohols, ethers, aldehydes, ketones, carboxylic acids, and esters. J. Chem. Theor. Comput. 7, 1016–1031.
74. Koziara KB, Stroet M, Alpeshkumar MK, Mark AE. 2014. Testing and validation of the Automated Topology Builder (ATB) version 2.0: prediction of hydration free enthalpies. J. Comput. Aided Mol. Des. 28, 221–233.
75. Wolf MG. *et al*. 2010. g_membed: efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation. J. Comput. Chem. 31, 2169–2174.
76. Plazinski W, Lonardi A, Hünemberger PH. 2016. Revision of the GROMOS 56A6(CARBO) Force Field: Improving the Description of Ring-Conformational Equilibria in Hexopyranose-Based Carbohydrates Chains. J Comput Chem. 37 (3), 354-365.
77. Abraham, M. J. *et al*. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 1, 19–25.
78. Bussi, G., Donadio, D. & Parrinello, M. 2007. Canonical sampling through velocity rescaling. J. Chem. Phys. 126, 014101.
79. Parrinello, M., Rahman, A. 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. J. Appl. Phys. 52, 7182.

80. Darden, T., York, D. & Pedersen, L. 1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J. Chem. Phys. 98, 10089–10092.
81. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. 1997. LINCS: A linear constraint solver for molecular simulations. J. Comp. Chem. 18, 1463–1472.
82. Hess, B. 2008. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. J. Chem. Theory Comput. 4, 116-122.
83. Humphrey, W. & Dalke, A. 1996. VMD: visual molecular dynamics. J. Mol. Graph. 15, 33–38.

For Table of Contents Only



7.6 Appendix F

SCIENTIFIC REPORTS

OPEN

Details of hydrophobic entanglement between small molecules and Braun's lipoprotein within the cavity of the bacterial chaperone LolA

Received: 26 July 2018
Accepted: 1 February 2019
Published online: 06 March 2019

Alister Boags, Firdaus Samsudin & Syma Khalid

The cell envelope of Gram-negative bacteria is synthesized and maintained via mechanisms that are targets for development of novel antibiotics. Here we focus on the process of moving Braun's lipoprotein (BLP) from the periplasmic space to the outer membrane of *E. coli*, via the LolA protein. In contrast to current thinking, we show that binding of multiple inhibitor molecules inside the hydrophobic cavity of LolA does not prevent subsequent binding of BLP inside the same cavity. Rather, based on our atomistic simulations we propose the theory that once inhibitors and BLP are bound inside the cavity of LolA, driven by hydrophobic interactions, they become entangled with each other. Our umbrella sampling calculations show that on the basis of energetics, it is more difficult to dislodge BLP from the cavity of LolA when it is uncomplexed compared to complexed with inhibitor. Thus the inhibitor reduces the affinity of BLP for the LolA cavity.

The cell envelope of Gram-negative bacteria is a complex chemical architecture composed of three distinct regions; the outer membrane, the inner membrane and the periplasmic space[1,2]. The latter is aqueous in nature, whereas the membranes are amphipathic: low dielectric hydrophobic cores surrounded by polar moieties on either side. All three regions contain proteins, which perform a variety of functions including maintaining the functional integrity of the cell envelope, and importantly, also synthesize the different molecules that constitute the cell envelope and localize them at the appropriate position within the whole architecture of the cell envelope[3]. This requires exquisite molecular choreography given the crowded environment of the cell envelope, and consequently a number of different pathways exist for synthesis and sorting of molecules.

Lipoproteins are abundant within the cell envelope, indeed Braun's lipoprotein (BLP) is the most abundant protein in *E. coli*[4,5]. It is anchored in the outer membrane via a lipid moiety, which has three hydrocarbon tails at its N-terminus and is covalently bound to the cell wall peptidoglycan at its C-terminus[6,7]. Thus far it is the only known protein to be covalently bound to the cell wall. It provides stability to the cell envelope by linking the outer membrane and the cell well. Braun's lipoprotein is synthesized at the inner membrane and then delivered to the outer membrane via a pathway involving the five Lol proteins, LolA, LolB, LolC, LolD, and LolE[8]. These proteins play key roles in the outer membrane-directed lipoprotein localization. The chaperone protein LolA has been shown to deliver BLP and other lipoproteins to the outer membrane-anchored protein LolB which then localizes them by a yet undetermined pathway[9–12]. The absence of LolA and LolB could result in toxic accumulation of mislocalized lipoproteins[13]. It has previously been shown that small hydrophobic molecules including MAC13243 (MAC) and its degradation products, S-(4-chlorobenzyl) isothiourea and 3,4-dichlorobenzyl carbamimidothioate, are able to bind to LolA and partially inhibit the protein *in vivo*[14,15].

A previous study suggests that inhibitor molecules like MAC bind within the hydrophobic cavity of LolA and physically block access to BLP in a classic competitive inhibition manner[16]. Unable to bind, BLP therefore cannot be delivered to LolB and be correctly localized. Support for this theory comes from the observation that

University of Southampton, Southampton, SO17 1BJ, United Kingdom. Alister Boags and Firdaus Samsudin contributed equally. Correspondence and requests for materials should be addressed to S.K. (email: S.Khalid@soton.ac.uk)

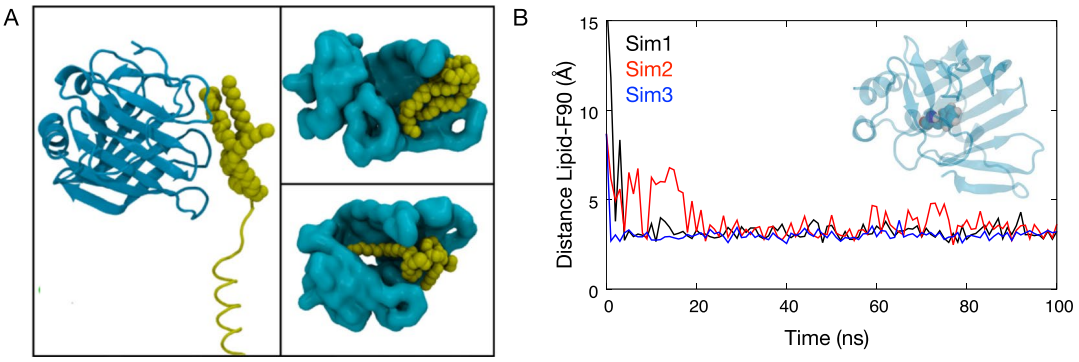


Figure 1. BLP binding modes within the hydrophobic cavity of LolA. (**A**) Left: the starting configuration, whereby BLP (yellow) is located just outside the LolA (cyan) cavity. Top right: binding mode in which three lipid tails of BLP bind near the mouth of the cavity. Bottom right: one lipid tail is deep within the cavity whereas the other two are bound at the mouth of the cavity. (**B**) Minimum distance between BLP lipid tails and residue F90 (shown in van der Waals representation in inset) from three independent simulations.

lipoproteins are partially retained at the inner membrane when *E. coli* strains are treated with MAC. However as yet there is no structure of either LolA nor LolB bound to BLP or to inhibitor molecules. The mode of BLP binding is therefore unknown, and important questions remain unanswered, for example how deep into the cavity of LolA does it bind, and how many acyl chains are accommodated in the cavity? Furthermore, there is no direct structural or molecular basis for the proposed mechanism of inhibition of LolA by blocking access to the hydrophobic cavity by small molecules. To address these questions, we have investigated the binding of BLP to LolA in the presence and absence of the inhibitor molecules, MAC and its degradation products, using atomistic molecular dynamics simulations. The first objective being to identify BLP binding modes and the second to identify the molecular processes that effect inhibition. Here, we propose a new molecular mechanism of inhibition of LolA by these small molecules. We find that both MAC and BLP bind LolA at the same time, whereby the lipid tails of the latter wrap around the former. Free energy calculations reveal that the MAC-BLP complex is easier to remove from LolA than BLP alone.

Results

Multiple binding modes of BLP to LolA. Firstly LolA and BLP were simulated in water and counterions by initiating the simulations with the BLP positioned just outside the hydrophobic cavity that is the proposed binding site of LolA[17]. The structure of LolA remained stable during the simulation as demonstrated by the similar RMSD progression compared to a simulation without BLP (Fig. S1). The main secondary structural motifs were also well preserved during these simulations (Fig. S2). We observed BLP moving into the cavity within 100 ns, after which the β sheet surrounding the cavity tilted to wrap around BLP, providing better interactions. The BLP lipid tails interacted with hydrophobic residues inside the binding cavity such as F90. There were 4 or 5 water molecules inside the cavity throughout the simulation; they were able to move in and out of the cavity even with BLP bound. We did not observe a single binding mode, but rather multiple LolA-BLP configurations such that at least one of the lipid tails of the lipoprotein was deep inside the cavity, stabilized by hydrophobic contacts (Fig. 1). In some simulations 2 or 3 lipid tails were inside the cavity. It is perfectly logical that binding within the cavity is general, as a single specific binding mode would likely restrict the LolA to chaperoning only BLP and not the other myriad lipoproteins within *E. coli*[12]. Furthermore, specific binding within LolA would likely stabilize the complex to an extent that would prohibit delivery of BLP to LolB. Indeed, it was previously shown that in the presence of LolB, a LolA-lipoprotein complex readily dissociated to form a LolB-lipoprotein complex, indicating a weaker, non-specific binding[11].

BLP binds to LolA in the presence of inhibitor molecules. Having established that 100 ns of simulation is sufficient to observe the movement of BLP to deep within the hydrophobic cavities of LolA, we next sought to characterize the effect of the inhibitors and to identify the mechanism of inhibition. Simulations were conducted in which 2 or 3 molecules of MAC were placed around the hydrophobic cavity of LolA with BLP placed just a bit further outside the cavity. Three independent 100 ns simulations were performed for each scenario. The rationale was to investigate (a) if the inhibitor would bind in the cavity and (b) if BLP could bind even if the inhibitor is present. From our previous simulations without the inhibitor, we identified residue F90 located in the LolA binding site to be one of the residues that interacted with BLP lipid tails; we therefore used a distance measurement between this residue and BLP as a metric to determine whether BLP binds to LolA.

Intriguingly even with 2 or 3 molecules of MAC within the cavity of LolA, the BLP lipid moiety can still bind inside the cavity, with the MAC molecules arranged around it (Fig. 2). Hydrophobic interactions stabilize the different molecules inside the cavity. The acyl chains of BLP are able to slide beyond the inhibitor molecules to reach deeper into the cavity. The acyl chains become entangled with the inhibitor molecules and the hydrophobic bulk of the inhibitor-BLP complex becomes wedged inside the cavity. We then replace the MAC molecules with its degradation products, S-(4-chlorobenzyl) isothiourea and 3,4-dichlorobenzyl carbamimidothioate, which are also expected to bind to LolA and inhibit BLP binding[15]. Similarly we observed binding of BLP to LolA in the

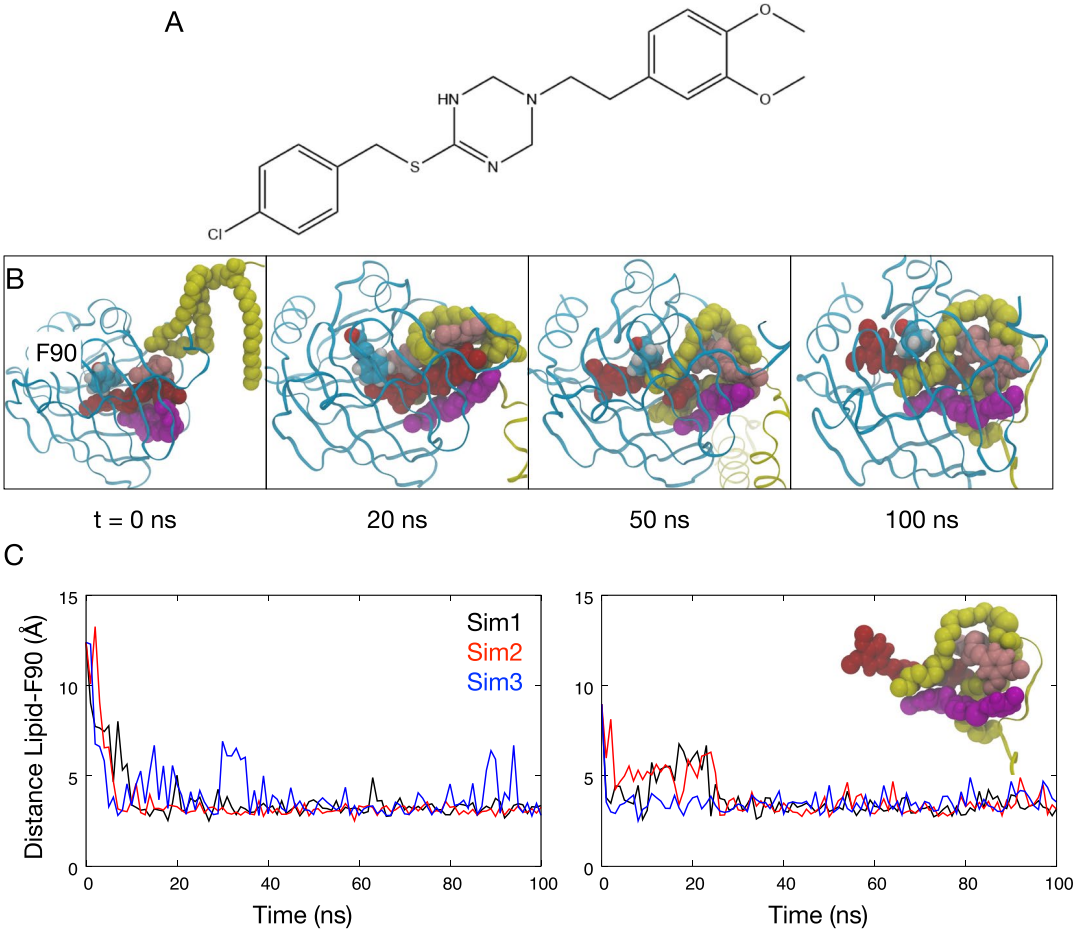


Figure 2. BLP binds to LolA even when two or three inhibitor molecules are present. (**A**) The chemical structure of MAC13243 inhibitor. (**B**) Snapshots at different time points during a simulation of LolA and BLP with three MAC molecules (red, pink, magenta) bound in the hydrophobic cavity. Residue F90 is labelled. (**C**) Minimum distance between the BLP lipids and F90 from simulations with two (left) and three (right) MAC inhibitors. Results are from three independent simulations. The insert illustrates hydrophobic entanglement of BLP lipids and three MAC molecules from one of the simulations.

presence of 2 or 3 molecules of the MAC degradation products within 100 ns (Fig. S3), further corroborating our hypothesis that the binding of these hydrophobic inhibitors does not occlude lipoprotein binding to LolA.

Our results are in contrast to the suggestions in the literature that binding of these molecules inside the hydrophobic cavity prevents binding of BLP;[16] in other words our simulations show that these molecules are not conventional competitive inhibitors, but instead uncompetitive inhibitor. It is worth noting that a system with two MAC molecules translates to a concentration of MAC of around 750 µg/ml, which is much higher than its minimum inhibitory concentration (16 µg/ml)[14], and even at this higher concentration binding of BLP was observed.

Inhibitor molecules reduce BLP interactions with LolA. To elucidate the mechanism of inhibition of MACs and its degradation products we looked for any structural differences between the LolA-BLP complexes in the absence and presence of MACs. BLP binding to LolA was found to elicit a larger conformational change of the latter in the absence of MAC, compared to when MACs were also bound. There was a pronounced tilt of the β sheets, particularly involving strands β3-β6, which allowed the binding site to 'wrap around' the BLP lipid tails (Fig. 3A). We performed principal component analysis and found the motion along the first eigenvector to involve the β sheets tilting towards the position of the BLP lipid tails (Fig. S4). In contrast this conformational change was absent in the simulations in which MACs were bound in the LolA cavity (Fig. 3C). LolA retained its original structure throughout in these simulations. This explains the lower RMSD progression from the X-ray structure in these simulations compared to those in which MAC was absent (Fig. S1).

The conformational change that occurred when BLP alone is bound to LolA resulted in an increased number of interactions between the BLP lipids and residues in the LolA binding site. Contact analysis using a distance cutoff of 4 Å revealed that in the absence of MACs the BLP lipid tails interacted with more hydrophobic residues located deep within the binding cavity such as L9, I23, W95 and I131 (Fig. 3B). On the other hand, when MACs were present the BLP lipid moieties contacted residues that were mostly found around the opening of the binding site (Fig. 3D). The results of the conformational and contact analyses together suggest that the MAC-BLP acyl

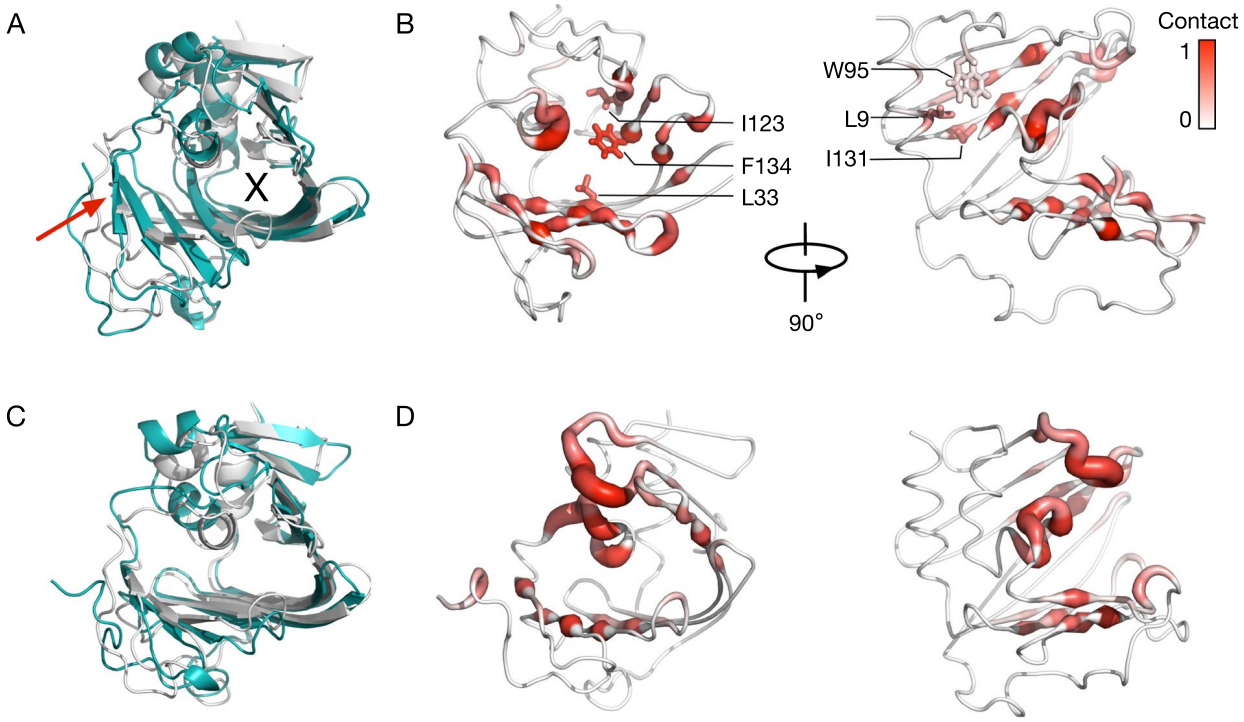


Figure 3. LolA binding site conforms to the shape of BLP lipid. (**A**) Structural comparison of LolA at the end of a simulation with zero MAC (cyan) to the crystal structure (grey). The tilt of the β-sheet is highlighted by the red arrow and the binding cavity is marked "X". (**B**) Contact analysis with a distance cut-off of 4 A performed for each residue of LolA with BLP lipid averaged over three simulations. A score of 1 indicates contact made throughout the entire simulations. Residues that show more contact in simulations with zero MAC compared to that with three MACs are labelled. (**C**,**D**) Same analyses as (**A**,**B**), respectively, performed for simulations with 3 MACs.

chain entanglement in the LolA binding site restricts conformational changes of LolA that are necessary for BLP interactions with key hydrophobic residues, and therefore results in shallower and potentially weaker binding.

While the BLP lipid tails occupy the LolA binding site, the BLP helix was found to bind to the outer solvent-exposed surface of LolA. To understand the conformational dynamics experienced by this helix during the simulations with and without MACs, we performed cluster analysis using the algorithm of Daura *et al.*[18]. Each conformation is assigned to a cluster based on an RMSD cutoff of 2 Å. The cluster size as well as the representative structures of the most frequently sampled conformation are shown in Fig. S5. Comparing two independent repeat simulations for each system we found the BLP helix to show very different dynamics: for example, in one simulation with 3 MACs the helix mostly sampled an elongated conformation, whilst in another simulation it preferred an L-shape conformation. We note that this suggests that our 100 ns duration simulations may not have converged vis-à-vis the conformational dynamics of the BLP helix and longer simulations are therefore needed to study the dynamics of the full BLP protein with LolA, however this is not the focus of the present study.

Inhibitor molecules weaken LolA-BLP binding. To quantify the strength of BLP binding to LolA in the absence and presence of 2 or 3 MACs, we performed a potential of mean force (PMF) calculation along the dissociation pathway of BLP. Our cluster analysis shows that long-timescale simulations are required to adequately sample the conformational dynamics of the BLP helix when it binds to the LolA outer surface; we therefore removed the helix for our PMF calculation to help achieve convergence. We first performed a steered MD simulation to pull the bound BLP lipid into solution (Fig. 4A), and subsequently used snapshots from this simulation as windows for a series of 100 ns umbrella sampling MD simulations. We achieved adequate sampling along the reaction coordinate as indicated by the histogram overlap (Fig. S6). PMF profiles constructed from increasing amount of simulation time suggest that convergence was reached after around 30 ns (Fig. S7).

Our PMF profiles shows that the BLP lipid bound most strongly to LolA when there was no MAC present, with the free energy of dissociation estimated to be around 45 kT (Fig. 4B). In the presence of two and three MACs these free energy values are reduced to around 40 kT and 37 kT respectively. For the systems with zero and three MACs, two additional PMF profiles were constructed using independent sets of steered and umbrella sampling MD from different starting coordinates. Encouragingly we found all three repeats to agree with each other within statistical errors and point towards the higher binding affinity of BLP lipid in the absence of MAC inhibitors (Fig. S8). This is concordant with our contact analysis, which shows that in the absence of MACs more interactions were formed between the BLP lipid moieties and hydrophobic residues in the LolA binding site. These additional interactions with residues found deep within the binding cavity therefore are crucial for a strong BLP binding. It is worth noting that even in the presence of MAC inhibitors binding is still likely, albeit weaker,

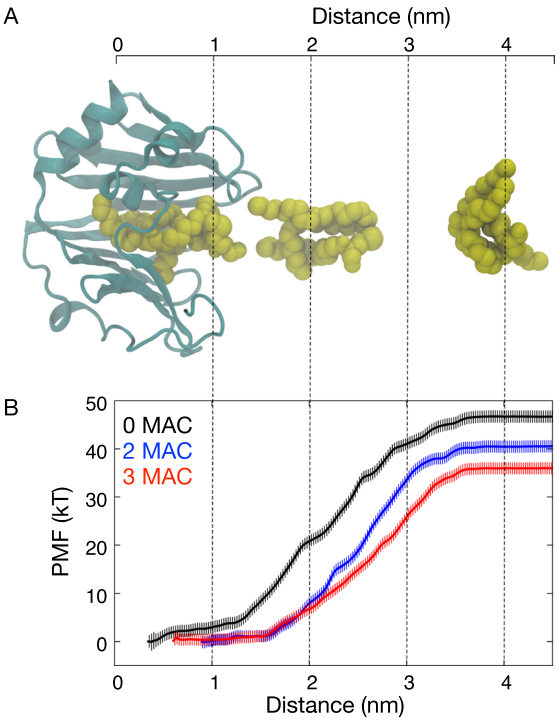


Figure 4. Free energy profiles of BLP lipid unbinding from LolA. (**A**) Snapshots taken from a steered MD simulation whereby BLP lipid is pulled out from LolA binding site. The reaction coordinate is the distance between centers of mass of the Sulphur atom on BLP lipid and the protein. (**B**) PMF profiles for LolA unbinding from umbrella sampling calculations. Error bars indicate standard deviations estimated from bootstrapping.

given the positive value of the free energy of dissociation. Again, this suggests that MAC and potentially its degradation products are not competitive inhibitors of LolA, but rather reduce the binding affinities of BLP.

Inhibitor molecules diffuse freely across the cell wall. To put our study in the larger context of the bacterial cell envelope, we then built a system incorporating the outer membrane[19] with LolB embedded, a single layer of peptidoglycan cell wall[20,21], and LolA positioned on the cell wall on the outer membrane side of the periplasm. Twenty MAC molecules were placed randomly in the solvent at the beginning of the simulation. We performed three independent 100 ns simulations to study how the MACs behave in the cell envelope environment and whether they readily bind to either LolA and LolB. Unsurprisingly we found most MACs spontaneously inserted into the hydrophobic core of the outer membrane (Fig. 5A). We did not observe MAC binding to either LolA or LolB within the 100 ns timescale of the simulation, which is largely a consequence of the greater distance between the MACs and the proteins here than in the simulations described above. More crucially several MAC molecules were observed to diffuse freely across the cell wall (Fig. 5B). This suggests that the inhibitor molecules can bind LolA even when this protein is located on the inner membrane side of the periplasm.

Discussion

In summary, we have shown that the lipid moieties of BLP binds deep into the hydrophobic cavity of LolA via non-specific binding modes. The small molecule, MAC13243 and its degradation products readily bind into the same cavity of LolA. We have shown that BLP lipid tails are able to bind into the cavity even when the cavity is already occupied by 3 MAC molecules, albeit not as deeply into the cavity as when the small molecules are absent. Inside the cavity, the BLP tails and MAC molecules become entangled with each other. Interestingly LolA undergoes a conformational change when BLP alone binds into the cavity, this provides additional stabilizing interactions between BLP and LolA. Free energy calculations reveal that dissociation of the BLP-inhibitor complex from LolA is energetically more favorable than dissociation of BLP alone from LolA. Based on these observations we propose that MAC and its degradation products inhibit LolA by reducing the binding affinity of BLP for LolA. It has been shown that in the presence of LolB (the protein to which BLP is delivered by LolA), the LolA-BLP complex dissociates, indicating that the affinity of BLP for LolB is higher than its affinity for LolA[11]. It is possible that MAC reduces the LolA-BLP affinity such that the MAC-BLP complex can become dislodged from LolA before reaching BLP. We would like to state here that while the free energy differences come directly from our simulations, the hypothesis regarding inhibition mechanism is not a conclusion based on our own results, but some thoughts on the implications of the results, which need further work (experimental and simulation) to confirm.

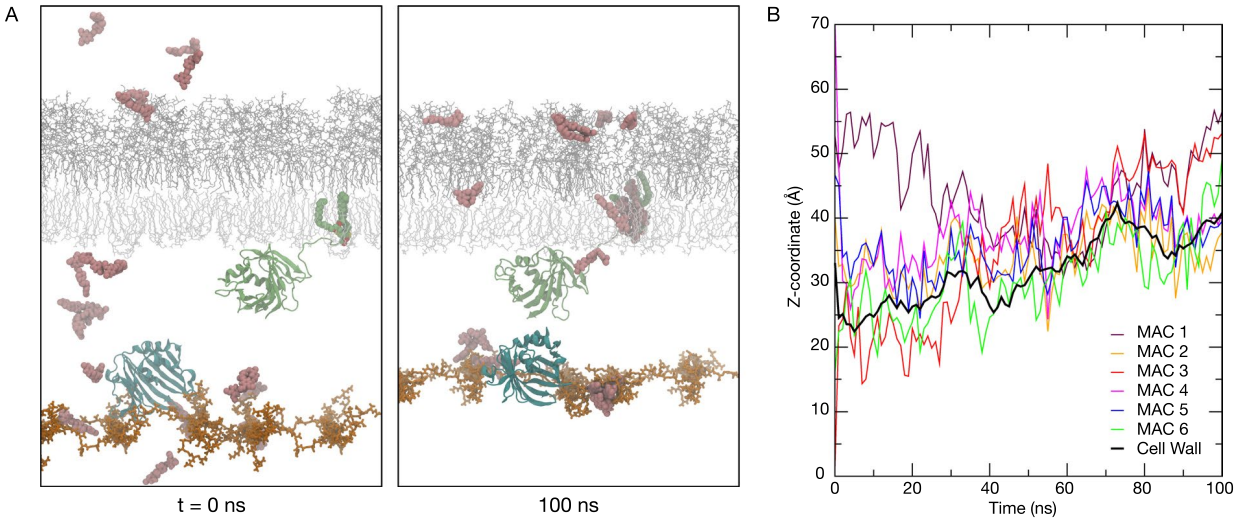


Figure 5. Simulation of MAC inhibitors with LolA and LolB. (**A**) Snapshots taken at the beginning (left) and at the end (right) of a simulation with LolA (cyan), LolB (green) embedded in an outer membrane model (grey), cell wall (orange), and 20 MAC molecules (pink). (**B**) The Z-coordinates of the centers of mass of six different MACs with respect to that of the cell wall throughout a 100 ns simulation, to highlight the ability of the inhibitor to diffuse through the cell wall layer.

Methods

The models. The BLP monomer was constructed from the X-ray the structure reported by Shu *et al.* (PDB: 1EQ. 7)[7]. The N-terminus was attached to the tripalmitoyl-S-glyceryl-cysteine residue as previously used in our simulations of BLP and OmpA[21]. The structures of LolA and LolB were obtained from the protein database (PDB: 1IWL and 1IWM respectively)[17]. The first nine amino acids of LolB were added in using Modeller 9.19[22] and the modelled protein was functionalised with a tripalmitoyl-S-glyceryl-cysteine residue (same lipid moiety as for BLP). MAC13243 and its degradation products, S-(4-chlorobenzyl) isothiourea and 3,4-dichlorobenzyl carbamimidothioate, were parameterised using the Automated Topology Builder (ATB)[23] with parameters for the GROMOS 54A7 forcefield. The outer membrane model comprises Ra LPS lipids of the R1 core type[24,25], in the upper leaflet, whilst the lower leaflet is composed of the following phospholipids: 90% 1-palmitoyl 2-cis-vaccenic phosphatidylethanolamine, 5% 1-palmitoyl 2-cis-vaccenic phosphatidylglycerol and 5% 1-palmitoyl 2-cis-vaccenic 3-palmitoyl 4-cis-vaccenic diphosphatidylglycerol)[26–29]. This model of the OM has previously been used and validated in our simulation studies[20,21,30,31].

Atomistic MD simulations. All simulations were performed using the GROMACS 2018 code[32], the GROMOS 54A7 force field with the SPC water model[33]. Each simulation was run for 100 ns and three independent repeats were performed. The temperature was maintained at 310 K using the velocity rescale thermostat with a time constant of 1 ps[34]. The pressure was maintained isotropically in simulations that did not contain a membrane, and semi-isotropically in membrane-containing simulations, at 1 atm using the Parrinello-Rahman barostat with a time constant of 1 ps[35]. All bonds were constrained using the LINCS algorithm to allow for an integration time step of 2 fs[36]. Long-range electrostatics were described using the particle mesh Ewald method[37]. The long-range electrostatic cut-off was 1.4 nm and the short-range van der Waals cut-off was also 1.4 nm. All systems were charge neutral. The LPS molecules were neutralised by Ca^{2+} ions, and the simulation systems contained additional 0.2 M NaCl ions. The trajectories were visualised in VMD[38] and analyses were performed using in-house scripts and GROMACS tools.

PMF calculation. Steered and umbrella sampling MD simulations were performed to generate inputs for the calculation of PMFs along a reaction coordinate parallel to the BLP dissociation pathway from LolA binding site. Snapshots at the end of the 100 ns simulations described above were used as starting configurations. For systems with zero and three MACs three independent sets of steered and umbrella sampling MD simulations were performed starting from three different snapshots from the equilibrium simulations. To aid convergence the BLP helix was removed in all systems leaving only the lipid moieties in the LolA binding site. A short 1 ns simulation was performed to re-equilibrate each system. Constant-velocity (0.1 nm ns^{-1}) steered MD simulation was performed to pull the BLP lipid away from the binding site along the y-axis using an elastic spring (force constant of 100 kJ mol^{-1} nm^2) applied to the Sulphur atom on its headgroup. The backbone atoms of LolA was positionally restrained with a force constant of 1000 kJ mol^{-1} nm^2 during this simulation. From this trajectory 50 umbrella sampling windows were then selected based on the distance between the centers of mass of the Sulphur atom and LolA along the reaction coordinate with a separation of 1 Å between windows. For each window a 100 ns simulation was performed with the center of mass of the Sulphur atom restrained in the vector of the reaction coordinate using a harmonic force constant of 1000 kJ mol^{-1} nm^2. No restraint was imposed on LolA. The weighted histogram analysis method (WHAM)[39] incorporated in the GROMACS *gmx wham* tool[40] was used to compute the PMF from the umbrella sampling data. Histogram overlaps were plotted (Fig. S6A), and extra sampling windows were

performed at missing coordinates to ensure adequate sampling. Integrated autocorrelation time (IACT) was computed for each umbrella sampling window (Fig. S6B); as IACTs computation is potentially inaccurate due to limited sampling, the IACTs along the reaction coordinate were subsequently smoothed with a Gaussian filter as described in Hub *et al.*[40]. The PMF profiles were then generated by WHAM taking these IACTs into account. Statistical errors were estimated using bootstrap analysis whereby new random trajectories were generated with data points distributed according to the given histograms and properly autocorrelated based on the previously computed IACT values. For each PMF profile, 100 bootstrap trials were performed. To confirm that the PMF has converged within the 100 ns umbrella sampling simulations, PMF profiles were constructed from increasing amount of simulation time (Fig. S7). We found in all systems and repeats PMF profiles converged after around 30 ns; the first 30 ns of the simulations was therefore regarded as equilibration time and excluded from the final PMF calculations.

References

1. Nikaido, H. Molecular basis of bacterial outer membrane permeability revisited. *Microbiol. Mol. Biol. Rev.* **67**, 593–656 (2003).
2. Silhavy, T. J., Kahne, D. & Walker, S. The bacterial cell envelope. *Cold Spring Harb. Perspect. Biol.* **2**, 1–17 (2010).
3. Silhavy, T. J., Ruiz, N. & Kahne, D. Advances in understanding bacterial outer-membrane biogenesis. *Nat. Rev. Microbiol.* **4**, 57–66 (2006).
4. Braun, V. & Wolff, H. The murein-lipoprotein linkage in the cell wall of Escherichia coli. *Eur. J. Biochem.* **14**, 387–391 (1970).
5. Braun, V. Covalent lipoprotein from the outer membrane of escherichia coli. *Biochim. Biophys. Acta.* **415**, 335–377 (1975).
6. Bosch, V. & Braun, V. Distribution of murein-lipoprotein between the cytoplasmic and outer membrane of *Escherichia coli*. *FEBS Lett.* **34**, 307–310 (1973).
7. Shu, W., Liu, J., Ji, H. & Lu, M. Core structure of the outer membrane lipoprotein from Escherichia coli at 1.9 A resolution. *J. Mol. Biol.* **299**, 1101–1112 (2000).
8. Tokuda, H. & Matsuyama, S. I. Sorting of lipoproteins to the outer membrane in *E. coli*. *Biochim. Biophys. Acta.* **1693**, 5–13 (2004).
9. Matsuyama, S., Tajima, T. & Tokuda, H. A novel periplasmic carrier protein involved in the sorting and transport of Escherichia coli lipoproteins destined for the outer membrane. *EMBO J.* **14**, 3365–3372 (1995).
10. Matsuyama, S. I., Yokota, N. & Tokuda, H. A novel outer membrane lipoprotein, LolB (HemM), involved in the LolA (p20)-dependent localization of lipoproteins to the outer membrane of Escherichia coli. *EMBO J.* **16**, 6947–6955 (1997).
11. Yokota, N., Kuroda, T., Matsuyama, S. I. & Tokuda, H. Characterization of the LolA-LolB system as the general lipoprotein localization mechanism of Escherichia coli. *J. Biol. Chem.* **274**, 30995–30999 (1999).
12. Masuda, K., Matsuyama, S. & Tokuda, H. Elucidation of the function of lipoprotein-sorting signals that determine membrane localization. *Proc. Natl. Acad. Sci. USA* **99**, 7390–7395 (2002).
13. Grabowicz, M. & Silhavy, T. J. Redefining the essential trafficking pathway for outer membrane lipoproteins. *Proc. Natl. Acad. Sci. USA* **114**, 4769–4774 (2017).
14. Pathania, R. *et al.* Chemical genomics in Escherichia coli identifies an inhibitor of bacterial lipoprotein targeting. *Nat. Chem. Biol.* **5**, 849–856 (2009).
15. Barker, C. A. *et al.* Degradation of MAC13243 and studies of the interaction of resulting thiourea compounds with the lipoprotein targeting chaperone LolA. *Bioorg. Med. Chem. Lett.* **23**, 2426–2431 (2013).
16. Muheim, C. *et al.* Increasing the permeability of Escherichia coli using MAC13243. *Sci. Rep.* **7**, 1–11 (2017).
17. Takeda, K. *et al.* Crystal structures of bacterial lipoprotein localization factors, LolA and LolB. *EMBO J.* **22**, 3199–3209 (2003).
18. Daura, X. *et al.* Peptide Folding: When Simulation Meets Experiment. *Angew. Chemie* **38**, 236–240 (1999).
19. Piggot, T. J., Holdbrook, D. A. & Khalid, S. Electroporation of the *E. coli* and S. Aureus membranes: molecular dynamics simulations of complex bacterial membranes-suppinfo. *J. Phys. Chem.* (2011).
20. Samsudin, F., Ortiz-Suarez, M. L., Piggot, T. J., Bond, P. J. & Khalid, S. OmpA: A Flexible Clamp for Bacterial Cell Wall Attachment. *Structure* **24**, 2227–2235 (2016).
21. Samsudin, F., Boags, A., Piggot, T. J. & Khalid, S. Braun's Lipoprotein Facilitates OmpA Interaction with the Escherichia coli Cell Wall. *Biophys. J.* **113**, 1496–1504 (2017).
22. Sali, A. & Blundell, T. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1994).
23. Koziara, K. B., Stroet, M., Alpeshkumar, M. K. & Mark, A. E. Testing and validation of the Automated Topology Builder (ATB) version 2.0: prediction of hydration free enthalpies. *J. Comput. Aided Mol. Des.* **28**, 221–233 (2014).
24. Appelmelk, B. J. *et al.* Frequencies of lipopolysaccharide core types in Escherichia coli strains from bacteraemic patients. *Microbiology* **140**, 1119–1124 (1994).
25. Vinogradov, E. V. *et al.* The structures of the carbohydrate backbones of the lipopolysaccharides from Escherichia coli rough mutants F470 (R1 core type) and F576 (R2 core type). *Eur. J. Biochem.* **261**, 629–639 (1999).
26. Lugtenberg, E. J. & Peters, R. Distribution of lipids in cytoplasmic and outer membranes of Escherichia coli K12. *Biochim. Biophys. Acta* **441**, 38–47 (1976).
27. Aibara, S., Kato, M., Ishinaga, M. & Kito, M. Changes in Positional Distribution of Fatty Acids in the Phospholipids of Escherichia coli after Shift-Down in Temperature. *Biochim. Biophys. Acta* **270**, 301–306 (1972).
28. Kito, M., Ishinaga, M., Nishihara, M., Kato, M. & Sawada, S. Metabolism of the phosphatidylglycerol molecular species in Escherichia coli. *Eur. J. Biochem.* **54**, 55–63 (1975).
29. Yokota, K., Kanamoto, R. & Kito, M. Composition of cardiolipin molecular species in Escherichia coli. *J. Bacteriol.* **141**, 1047–1051 (1980).
30. Piggot, T. J., Holdbrook, D. A. & Khalid, S. Electroporation of the *E. coli* and S. Aureus membranes: molecular dynamics simulations of complex bacterial membranes. *J. Phys. Chem.* **115**, 13381–8 (2011).
31. Ortiz-Suarez, M. L., Samsudin, F., Piggot, T. J., Bond, P. J. & Khalid, S. Full-Length OmpA: Structure, Function, and Membrane Interactions Predicted by Molecular Dynamics Simulations. *Biophys. J.* **111**, 1692–1702 (2016).
32. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19–25 (2015).
33. Berendsen, H. J., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. In *Intermolecular* Forces (ed. Pullman, B.) 331–342 (Reidel Publishing, 1981).
34. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
35. Parrinello, M. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182 (1981).
36. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.* **18**, 1463–1472 (1997).
37. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577 (1995).
38. Humphrey, W. & Dalke, A. VMD: visual molecular dynamics. *J. Mol. Graph.* **15**, 33–38 (1996).
39. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).
40. Hub, J. S., De Groot, B. L. & Van Der Spoel, D. G-whams-a free Weighted Histogram Analysis implementation including robust error and autocorrelation estimates. *J. Chem. Theory Comput.* **6**, 3713–3720 (2010).

Author Contributions

A.B. and F.S. performed and analysed simulations. F.S. and S.K. wrote the paper. All of the authors have given approval to the final version of the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-40170-z.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.