UNIVERSITY OF SOUTHAMPTON

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

SCHOOL OF CHEMISTRY

Molecular Dynamics Studies of Transmembrane Proteins within Complex Lipid Environments

Daniel Holdbrook

For the title of Doctor of Philosophy

October 2013

ABSTRACT

The interactions between lipids and proteins are crucial for many cellular processes. Typically, the nature of these interactions is studied in simple model lipid bilayers, which lack the complexity and heterogeneity of *in vivo* systems. Thus, this thesis investigates the impact of the lipid bilayer composition on protein dynamics and function. Both coarse grain and atomistic molecular dynamics simulations have been used to model membranes that contain lipid compositions approximating those found *in vivo*. The influence of these complex lipid environments on the dynamics of α-helical and β-barrel membrane protein is investigated. In particular, coarse grained simulations of a bilayer composed of a complex mixture of lipids, representing the Golgi apparatus, were used to identify preferential interactions of a helical transmembrane peptide with PIP2 lipids. Furthermore, atomistic molecular dynamics simulations have been used to identify several behaviour altering interactions between lipopolysaccharide, which is a key component of the Gram-negative bacterial outer membrane, and two outer membrane proteins, Hia and FecA. Lastly, coarse grained unilamellar vesicles, containing a complex mixture of phospholipids, were simulated in order to investigate protein aggregation and the short-term anomalous diffusion of lipids.

Contents

List of tables

List of figures

DECLARATION OF AUTHORSHIP

I, Daniel Aitor Holdbrook

declare that the thesis entitled

Molecular Dynamics Studies of Transmembrane Proteins within Complex Lipid Environments

and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- 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;
- I have acknowledged all main sources of help;
- where the 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;
- parts of this work have been published as:
    - Holdbrook DA, Leung YM, Piggot TJ, Marius P, Williamson PT and Khalid S. Stability and membrane orientation of the fukutin transmembrane domain: a combined multiscale molecular

dynamics and circular dichroism study. *Biochemistry*. 2010, 49(51):10796-10802.

- Piggot TJ, Holdbrook DA, Khalid S. Conformational dynamics and membrane interactions of the E. coli outer membrane protein FecA: a molecular dynamics simulation study. *Biochim Biophys Acta*. 2013, 1828(2):284-293.
- Holdbrook DA, Piggot TJ, Sansom MS, Khalid S. Stability and membrane interactions of an autotransport protein: MD simulations of the Hia translocator domain in a complex membrane environment. *Biochim Biophys Acta*. 2013, 1828(2):715-723.

Siged: ……..

Date:……

Acknowledgements

There are many people that deserve thanks for their assistance during the years of work that it took to create this thesis. On the note of funding, I would like to express my gratitude to the EPSRC. In addition, acknowledgements must be made to the people who have helped me to publish the work that is presented in the following chapters. In particular, I would like to thank Dr Thomas Piggot, with whom I have worked closely on a number of projects. I would also like to thank Dr Joseph Goose and Professor Mark Sansom for the many useful scientific discussions, particularly relating to the simulations of the outer membrane vesicles. A special acknowledgement must be made to my supervisor, Dr Syma Khalid, for the many years of guidance. In this regard, I would like to thank all of the SSBM group, past and present, for their support. Finally, I would like to thank my parents, Isabel Holdbrook and Dr Stephen Holdbrook, my brother, Adam Holdbrook, and my fiancée, Dr Cristina Crisostomo, who have all provided encouragement throughout.

Definitions and Abbreviations

AT – atomistic

ATMD – atomistic molecular dynamics

CG – coarse-grained

CGMD – coarse-grained molecular dynamics

CHOL – cholesterol

COM – centre of mass

DLPC – 1,2-dilauroyl phosphatidylcholine

DMPC – 1,2-dimyristoyl phosphatidylcholine

DPPC – 1,2-dipalmitoyl phosphatidylcholine

DPPE – 1,2-dipalmitoyl phosphatidylethanolamine

DPPG – 1,2-dipalmitoyl phosphatidylglycerol

DPPS – 1,2-dipalmitoyl phosphatidylserine

DSPC – 1,2-distearoyl phosphatidylcholine

DSSP – Dictionary of protein secondary structure

ER – endoplasmic reticulum

FK1TMD – fukutin transmembrane domain

GA – golgi apparatus

H-bond – hydrogen bond

IM – inner membrane

LPS – lipopolysaccharide

MD – molecular dynamics

MPoPE - 1-myristol 2-palmitoleoyl phosphatidylethanolamine

OM – outer membrane

OMP – outer membrane protein

OMV – outer membrane vesicles

PC – phosphatidylcholine

PDB – protein data bank

PE – phosphatidylethanolamine

PG – phosphatidylglycerol

PI – phosphatidylinositol

PIP2 – 1,2-dipalmitoyl phosphatidylinositol 4,5-bisphosphate

POPC – 1-palmitoyl,2-oleoyl phosphatidylcholine

PS – phosphatidylserine

SM – sphingomyelin

wt – wild type

Chapter 1: Introduction

1.1 Lipid Membranes and Membrane Proteins

Phospholipid membranes are one of the fundamental building blocks of life. They serve to protect biological cells and compartmentalise cellular processes. However, this barrier to the movement of polar molecules, while essential for cells to maintain a stable intracellular chemical environment, poses problems for cell-to-cell communication, waste disposal of unwanted intracellular components, and nutrient retrieval from the extracellular medium. Thus, cells have developed sophisticated mechanisms to transport molecules across phospholipid bilayers.

Biological membranes are not simply a sea of lipids that surround membrane proteins. Indeed, membrane bilayers can have a broad spectrum of chemical and physical properties as a result of the large diversity in lipid structures that can be created. This diversity can allow membranes to regulate protein behaviour. Protein function, for instance, can be dramatically altered, by interactions with different lipid membranes. This is especially evident for the diacylglycerol kinase from *E. coli*, where its activity was shown to be dependent on the hydrophobic thickness of the surrounding phospholipid bilayer (*1*).

In addition to changing the hydrophobic thickness, lipids can alter the gross properties of membrane in other ways to exert their influence. Indeed, the lateral pressure of a membrane is key to regulating the activity of mechanosensitive proteins. For example, the mechanosensitive channel from *E. coli* called MscL relieves osmotic pressure within the cell by sensing changes to membrane tension (*2*), acting like an emergency release valve.

It is believed that membrane proteins account for 25 % of all the proteins encoded by the human genome (*3, 4*). Membrane proteins are often important drug targets, not only due to prevalence in cells, but as a result of the key role that they play in signalling pathways. By forming microdomains, membrane bilayers can also provide a platform for membrane protein organisation within cells, and thereby regulate signalling cascades. Microdomains form as a result of a complex interplay between a combination of different lipid species.

Homogenous lipid bilayers, which contain only a single lipid species, undergo a range of structural transitions when they are cooled and heated. At low temperatures the homogenous bilayer will form gel-like structure. As the temperature is raised the bilayer structure transitions to a liquid disordered state. The temperature at which this occurs is called the phase transition temperature. Every lipid has a characteristic temperature at which it will undergo a phase transition. Increasing the complexity of bilayers, by mixing two different lipids together, can dramatically alter the membrane behaviour. Indeed, membrane bilayers containing a binary mixture of lipids, each with very different phase transition temperatures, can be observed to organise into phase separated domains, each with a uniform composition of lipids. In a similar process, but involving by many more factors, microdomains are thought to form in the complex lipid mixtures of biological membranes. Some proteins may be preferentially included or excluded from these microdomains. Signal transduction proteins, for example, have been shown to localise into microdomains (*5*), which could favour the activation of signalling cascades.

In Gram-negative bacteria, membrane-spanning proteins of the outer membrane (OM), called outer membrane proteins (OMPs), facilitate nutrient collection. These proteins can act either as simple selectivity filters or as energy driven machines, which actively move molecules into cells against their concentration gradients. Other OMPs, called the autotransporters, are part of a simple secretion mechanism, which allows them to transport one of their domains into the extracellular medium, and thereby facilitate bacterial colonisation and biofilm creation (*6*).

The structural diversity of lipids enables cells to construct membrane bilayers with many different compositions. The unique, and complex, combinations of lipids may alter the localisation and the function of membrane proteins. This ability of different lipids to modulate protein behaviour has been investigated in the following thesis. The overall aim is to provide predictions of membrane protein behaviour in a number of different lipid bilayer environments. In all cases herein, an attempt has been made to replicate *in vivo* membrane compositions as accurately as possible. This thesis improves on previous studies of membrane protein dynamics that use lipid bilayers with homogenous compositions.

The study of membrane and protein dynamics is a unifying theme throughout this thesis. In particular, the following chapters present Molecular Dynamics (MD) simulations of membrane bilayers that contain a complex composition of lipids. As a subsidiary theme, several of the studies concern proteins or elements of secretory pathways, and, therefore, the effects of lipid composition on protein diffusion have also been investigated.

In Chapter 3, the dynamics of a short membrane spanning peptide is investigated in number of different membrane environments. The peptide corresponds to the transmembrane α-helix of a human Golgi-located glycosyltransferase called fukutin, which is involved in the post-translational modification of secreted proteins. In Chapter 4, MD simulations of the translocator domain from an autotransporter called Hia, within a complex membrane, provide details about novel interactions that may anchor the protein within the membrane. In Chapter 5, the effect of membrane bilayer composition on the dynamics of an *E. coli* OMP called FecA is investigated. Finally, in Chapter 6, the diffusion of lipids and proteins in OM vesicles is examined. These vesicles are secreted from the surface of Gram-negative bacteria in order to export insoluble material, including membrane proteins.

1.2 Lipid Compositions of Biological Membranes

1.2.1 Eukaryotic Cells

Biophysical and biochemical analyses have provided detailed information about the composition and structure of membranes. Polar glycerophospholipids, or glycerol-based phospholipids, are the major structural constituents of cell membranes. They consist of a hydrophilic headgroup and a hydrophobic diacylglycerol portion. Different combinations of hydrophilic and hydrophobic regions can alter the physicochemical characteristics of the glycerophospholipids. A number of different phospholipid headgroups are found in eukaryotic cells. Phospholipids with headgroups composed of either phosphatidylcholine (PC) or phosphatidylethanolamine (PE) are zwitterionic, while those with headgroups composed of phosphatidylglycerol (PG), phosphatidylinositol (PI) or phosphatidylserine (PS) are anionic (Figure 1). In addition, an unusual anionic lipid called cardiolipin is found in the mitochondrial inner membrane. Cardiolipin is created by the covalent attachment of two PG phospholipids.

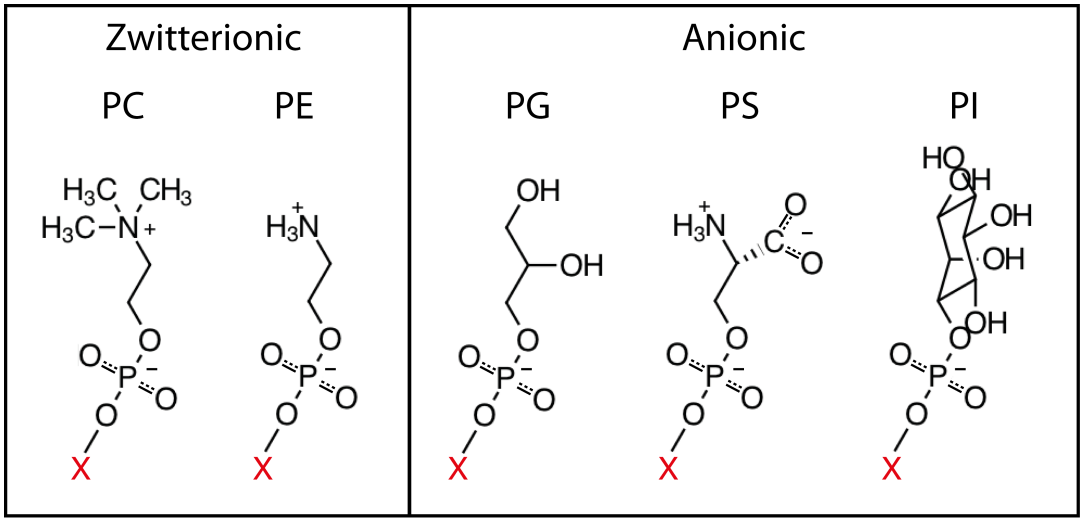


Figure 1. The headgroup of common phospholipids: phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylglycerol (PG), phosphatidylserine (PS), and phosphatidylinositol (PI). The X denotes the position of the glycerol backbone (not shown), to which the hydrophobic fatty acid tails are also attached.

Eukaryotic membranes also contain a number of other important constituents. The sphingolipids, which have a hydrophobic backbone of ceramide, represent another class of structural phospholipid, of which the major type in mammalian cells is sphingomyelin (SM). Sterols, such as cholesterol (CHOL), are also present at high concentrations within eukaryotic membranes.

The organelles of eukaryotic cells have been found to contain different proportions of the glycerophospholipids (*7*). The unique compositions, in part, may result from local lipid metabolism. Both the Endoplasmic Reticulum (ER) and Golgi apparatus (GA) are major sites of lipid synthesis. The ER specialises in the production of phospholipids and sterols, while the GA is a significant site of sphingolipid production (*8*).

The starkest difference in bilayer composition is between the ER and plasma membranes (*7, 9*). The ER is enriched with PC and PI, while the plasma membrane contains much more PS and SM. Starting from the ER, the membranes of the GA gradually become more similar in composition to the plasma membrane the further they are along the secretory pathway (*10*). Parts of the plasma membrane may also be compartmentalised to perform special functions. Indeed, epithelial cells are polarised, having unique apical (environment facing) and basolateral (adjacent cell facing) domains. The apical domain is enriched with sphingolipids, while the basolateral membrane contains more PC (*7*).

In addition to differences in lipid composition, organelles can have asymmetries in the distribution of lipids across their two leaflets. The asymmetric distribution of lipids is created by protein pores, called flippases and floppases, which flip certain lipids from one side of the membrane to other (*11-13*). In addition, differences in the site of production of some lipids can also contribute to the asymmetry (*12*). Once formed, the asymmetric distribution of lipids is preserved by the slow rate of natural exchange, or flip-flop, of phospholipids between the leaflets (*14, 15*). Nearly all of the PI, PS and PE are found in the inner cytosolic leaflet of the cells. The functional importance of this asymmetry in lipid composition is most clearly demonstrated in programmed cell death, or apoptosis, in higher eukaryotes. During programmed cell death, the negatively charged PS lipids are presented

on the outer extracellular leaflet of the cell (*16, 17*), signalling to the immune system to remove the unwanted cell.

1.2.2 Gram-negative Bacteria

The precise composition of lipids in the bacterial inner membrane (IM) differs from one bacterial species to another. In Gram-negative *E. coli*, the composition of the IM phospholipids is approximately 70% PE, 20% PG, and 10% cardiolipin. In comparison, the Gram-positive *B. subtilis* plasma membrane contains approximately 20% PE, 40% PG, 25% cardiolipin and the remaining 15% is a zwitterionic lipid called lysyl-phosphatidylglycerol. However, it must be noted that the compositions of the IM/plasma membrane phospholipids can change during a bacterium's life cycle, as has been demonstrated for both *E. coli* and *B. subtilis* (*18*). Furthermore, the fatty acid composition of the phospholipids is altered in response to temperature changes in the environment, with cis-vaccenic and palmitoyl becoming more prevalent at lower temperatures (*19, 20*).

The phospholipids of the Gram-negative IM are evenly distributed between the two leaflets, forming a symmetric bilayer. Meanwhile, the OM contains an asymmetric distribution of lipids between each of the leaflets. The inner leaflet of the OM consists of glycerolphospholipids: PE, PG, and cardiolipin, while the outer leaflet almost entirely consists of an unusual lipid called lipopolysaccharide (LPS). Compared to the IM phospholipids, the inner leaflet of the OM is enriched in PE and saturated fatty acids (*21*). However, the exact proportions of PE, PG and cardiolipin in the inner leaflet vary during the cell cycle. Similarly to the IM, the OM composition of fatty acids is dependent on the environmental temperature (*20*).

The LPS in outer leaflet of the OM can be divided in three distinct regions (Figure 2): i) the hydrophobic anchor of LPS is called Lipid A, ii) the core region, which consists of a phosphorylated non-repeating oligosaccharide, is attached to the polar end of the Lipid A anchor, and iii) an O-antigen of repeating oligosaccharide is also attached to the core sugars in wild-type *E. coli*, but it is absent in *E. coli* K-12 strains.

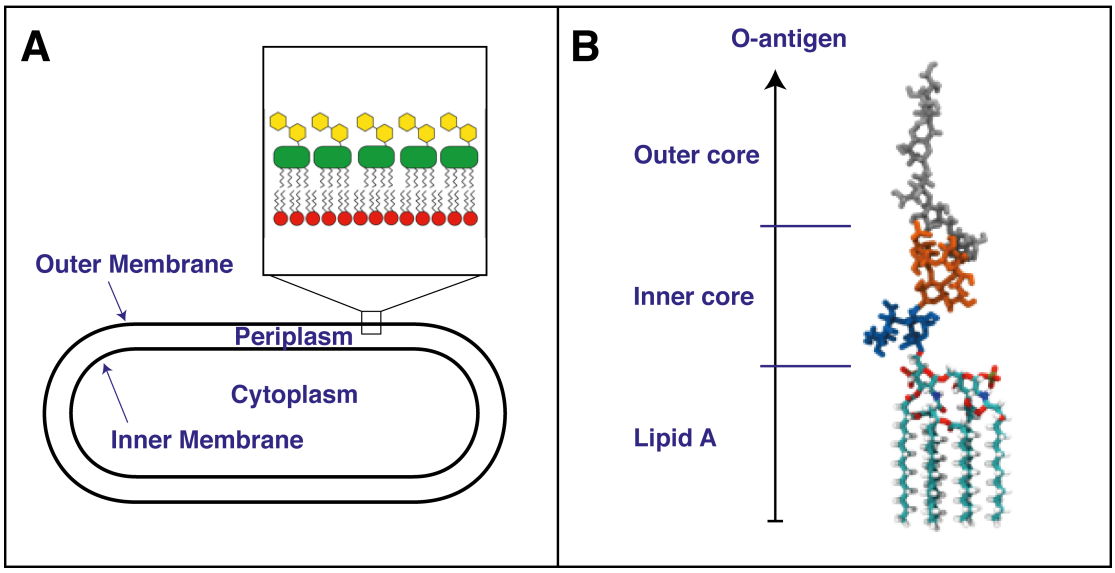


Figure 2. A) A schematic representation of a Gram-negative bacterium showing the position of the inner and outer membranes. A small section of the outer membrane has been enlarged, showing the inner leaflet lipids in red, and the lipopolysaccharide in green (lipid A) and yellow (core sugars). B) A single LPS molecule displaying the levels of oligosaccharide attached to lipid A.

1.3 Glycosyltransferases of The Conventional Eukaryotic Secretion Pathway

In the conventional secretion pathway, proteins are targeted to the plasma membrane in a regulated and organised process. Conventional protein secretion refers to the ER-GA mediated secretory pathway, in which an N-terminal signal peptide directs a protein to the ER. Proteins that use this pathway are fed co-translationally across the membrane of the ER. During the translocation, proteins may be modified by the covalent attachment of a preformed oligosaccharide at specific asparagine residues within the polypeptide chain, a process called N-linked glycosylation. After folding is completed in the lumen of the ER, the glycosylated proteins are transported in COPII transport vesicles to the GA. The proteins enter the GA when the transport vesicle fuses with the cis membrane sac (Figure 3A). In many cases, the proteins will then be transported through the cis, medial and trans cisternae. How this occurs is a matter of contention. Proteins may transit between individual cisterna in the way described by the cisternal maturation and/or rapid partitioning models (*22*). Conversely, proteins transiting the GA may do so in forward moving COPI-generated transport vesicles. In all cases, retrograde vesicles may be required to retrieve mislocalised proteins.

During passage through the GA, the N-linked glycan is modified. Additional oligosaccharides can be attached to serine or threonine residues in a process called O-glycosylation. The addition and modification of N- and O-linked oligosaccharides is catalysed by glycosyltransferases (*23*), which are the major integral membrane proteins of GA membranes. Glycosyltransferases are a single-pass membrane protein with an N-terminal cytoplasmic tail (Figure 3B). Their catalytic domain resides in the lumen of the GA.

Electron microscopy has shown the glycosyltransferases to have a non-uniform distribution within the GA (*24, 25*), possibly assisting the sequential modification of N- and O-linked oligosaccharides. The mechanisms that organise glycosyltransferases in the Golgi are not understood. The organisation of proteins in the Golgi is believed be the consequence of iterative cycles of anterograde and retrograde transport. No single mechanism

can account for the retention of all glycosyltransferases in specific cisternae of the GA.

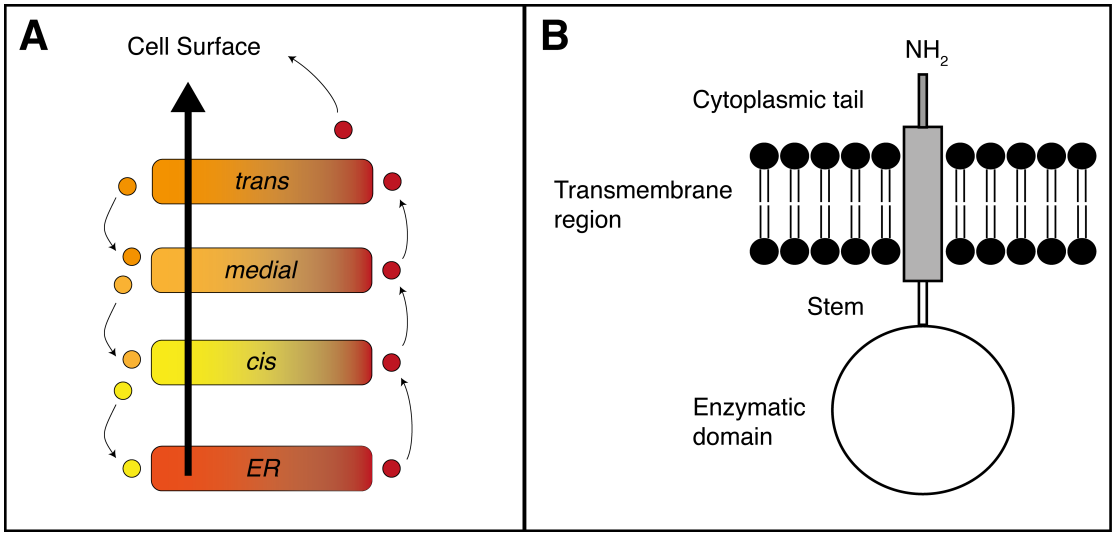


Figure 3. A) The possible mechanisms of membrane trafficking in the GA. Anterograde vesicles on the right hand side of the diagram are coloured red, while retrograde vesicles are on the left hand side. The large black arrow demonstrates the direction of travel of secreted proteins. B) The domain structure of glycosyltransferases. In shaded grey are the regions of fukutin that were modelled and simulated in Chapter 3.

Several studies have highlighted the importance of motifs in the cytoplasmic tail of glycosyltransferases (*26*). The hydrophobic region of the plasma membrane is thicker than in membranes of the early stages of secretory pathway. This is reflected in the transmembrane domains of GA resident proteins, which have thinner hydrophobic regions than their plasma membrane counterparts (*27*). Indeed, changing the length of the transmembrane spanning region of certain glycosyltransferases affects their retention in the GA (*26*). The correct localisation of some glycosyltransferases requires both the cytoplasmic tail and transmembrane domains together, or the transmembrane domain and stem region together. The localisation of other glycosyltransferases is mediated by protein-protein interactions. The majority of glycosyltransferases, however, appear to find their correct place of residence using the cytoplasmic tail and transmembrane domain (*28*).

1.4 Bacterial Type V Secretion Pathway

The cell wall of Gram-negative bacteria contains two membrane bilayers. The space sandwiched between these two membranes is known as the periplasmic space, or periplasm (Figure 2). In Gram-negative bacteria, a protein that crosses the inner membrane can become either a resident of the periplasm, an integral part of the OM, or be exported into the extracellular milieu. At least six distinct protein export pathways have been characterised in Gram-negative bacteria, reviewed in (*29-34*). Perhaps, the simplest of these is the type V secretion pathway, which transports proteins across both the IM and OM. Five subtypes of the type V secretion pathway are known to exist, though they share many mechanistic features in common (*30*). The apparent simplicity of the type V secretion pathway has made it popular as a possible protein expression system (*35, 36*). Efforts to exploit this pathway have been hindered, however, by an inadequate understanding of the final stages of secretion.

The type V secretion pathway is used by a family of proteins called the autotransporters. Two subtypes of autotransporter are known to exist, the monomeric and trimeric autotransporters, though they have a unifying domain structure. The autotransporters are characterised by three structural domains: i) an N-terminal signal sequence, ii) a passenger domain, and iii) a C-terminal translocator domain.

The passenger domain is the functional part of the protein, and it typically has a virulence related activity, such as cell adhesion, proteolysis or lipolysis (*6, 30, 37*). The autotransporter passenger domain must, therefore, be transported to the outside of cell in order to function. Both the N-terminal signal peptide and C-terminal translocator domain help with the transport of the passenger. The signal peptide directs the protein for secretion across the IM, while the translocator domain enables the passenger to cross the OM.

The first stage of the type V pathway involves targeting the nascent autotransporter polypeptide to the inner cytoplasmic membrane. This is also a component of two of the other bacterial export pathways. Indeed, proteins that integrate with, or cross, the cytoplasmic membrane often contain a N-terminal signal sequence. This signal directs the nascent polypeptide, as it emerges from the ribosome, to an inner membrane channel, called the Sec translocon

(*30, 38*). The channel of the Sec translocon is able to detect residents of the inner cytoplasmic membrane and release their transmembrane segments laterally into the bilayer, but it ignores the transmembrane segments of the OMPs. The OMPs pass through the Sec translocon unnoticed because they are structurally distinct to integral IM proteins, with OMPs having much shorter membrane spanning segments. As a rule, IM proteins have a α-helical topology; while, OMPs are almost always β-barrel in structure. Perhaps, one reason for this difference is that proteins destined for the OM would likely be retained incorrectly in the cytoplasmic membrane if they were α-helical in structure.

The N-terminal signal sequence of autotransporters is often longer than that of other Sec-dependant proteins. Indeed, alterations to the length of the N-terminal signal sequence results in misfolded aggregates in the periplasm (*39*). The longer signal sequence appeared to slow translocation across the inner membrane, presumably to allow time for chaperones of the periplasm to bind and prevent premature folding. Like many OMPs, the autotransporters are delivered to the OM by the chaperones, SurA and Skp (*40, 41*). Currently, it is unclear how the autotransporter translocator domain aids in the transport of the passenger domain across the OM. However, a number of models have been proposed (Figure 4) (*42*).

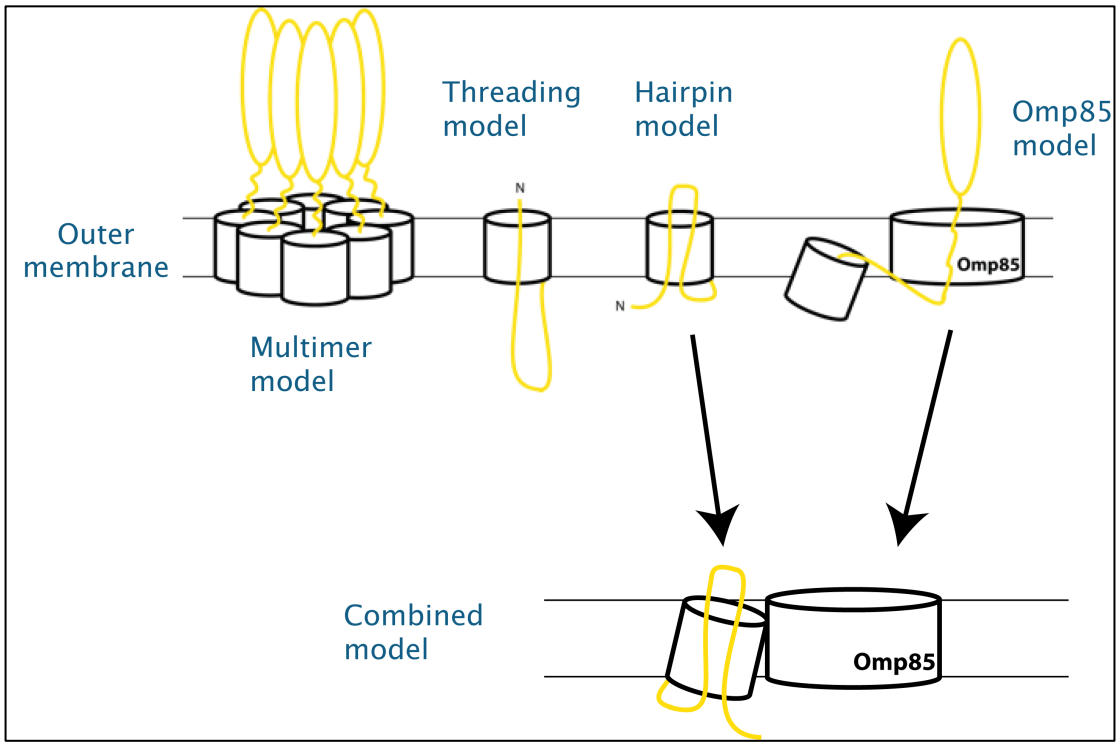


Figure 4. The different models that have been proposed to explain the translocation of the autotransporter passenger domain across the OM. The multimer model proposes that several translocator domains can associate to create a large pore through which several passenger domains can be transported, possibly in a fully folded state. In the threading model the unfolded passenger domain is threaded through a pore in the translocator domain in an N- to C-terminal direction. Similarly, the hairpin model proposes that the unfolded passenger threads through a pore in the translocator domain, but in a C- to N-terminal with a hairpin on the extracellular side of the membrane. The Omp85 model proposes that the passenger could be transported through the pore of another protein called Omp85. Experimental evidence supports both, or a combination, of the hairpin and Omp85 models.

The structures of several autotransporter translocator domains have been elucidated: EspP and Hbp from *E. coli (43-45)*, Hia from *H. influenzae (46)*, EstA from *P. aeruginosa* (*47*), BrkA from *B. pertussis* (*48*), and NalP from *N. menigitidis* (*49*). All of these translocator domains consist of a 12-stranded β-barrel structure (Figure 5). Indeed, the translocator domain of the trimeric autotransporters is also a 12-stranded β-barrel, but it is constructed from 3 identical subunits, each providing 4 β-strands.

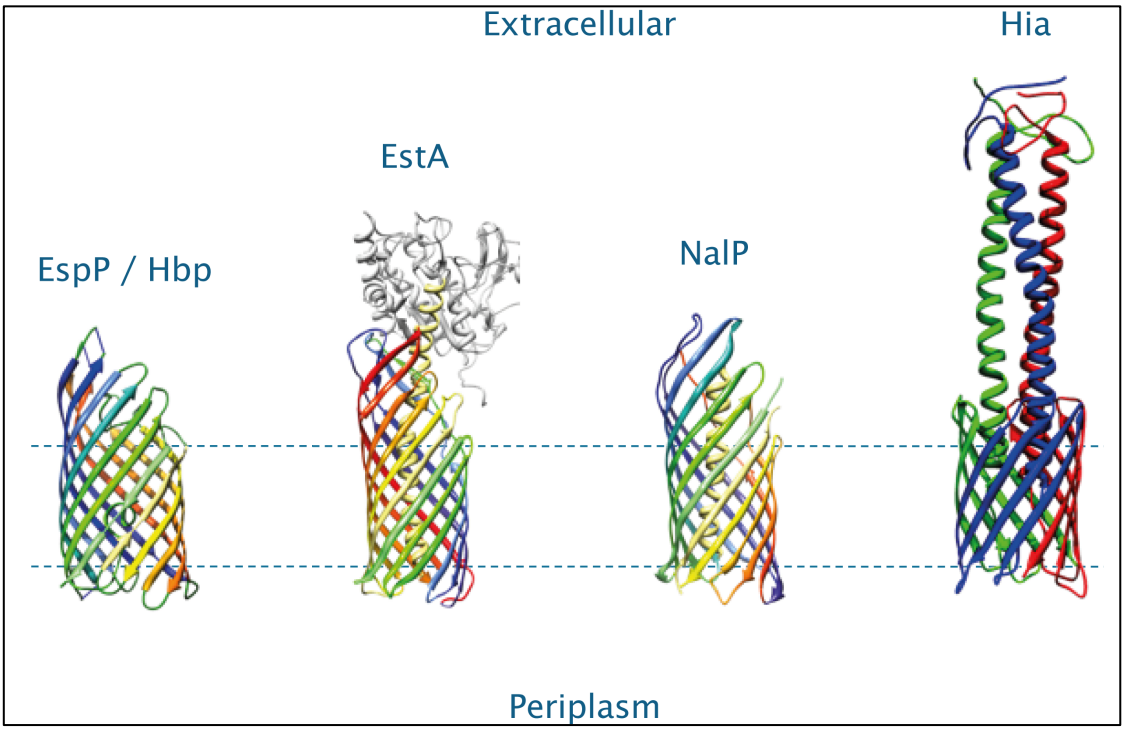


Figure 5. Structures of the autotransporter translocator domain. Each translocator domain forms a 12-stranded β-barrel with either one or three central α-helices in the pore. The β-strands are coloured separately in the monomeric structures of EspP/Hbp, EstA and NalP. In the structure of the trimeric autotransporter, Hia, each subunit is coloured separately in red, green or blue. EstA is currently the only autotransporter to have been characterized with the complete passenger domain attached (shown in grey).

The multimer model proposes that many translocator domains could assemble into a large pore through which the passenger domain could pass into the extracellular medium. This model was suggested as result of an observation that the IgA protease from *E. coli* could form ring like aggregates (*50*). However, most autotransporters have little propensity to aggregate into complexes (*49, 51, 52*), and therefore the multimer model may not represent a general mechanism for translocation of the passenger domain.

The threading model suggests that the passenger is extruded through the centre of the β-barrel in a N- to C-terminal direction. Similarly, in the hairpin model the passenger is extruded through the centre of the β-barrel (*53*), but in the C- to N-terminal direction, with a hairpin turn on extracellular side of the membrane. Evidence from translocation-stalled mutants of the autotransporters has demonstrated that the passenger translocation occurs in the C- to N-terminal direction (*40, 54*), supporting the hairpin model.

It has been suggested that a key member of the OMP insertion complex called BamA, also known as Omp85, could translocate the autotransporter passenger domain through its pore. Indeed, crosslinking studies have demonstrated that BamA and other key members of the OMP insertion complex are involved in autotransporter maturation (*40, 55*). Thus, a new model combining elements of both the hairpin and Omp85 models is likely to provide an overall mechanism for passenger domain transport.

1.5 Molecular Dynamics Simulations of Membranes with Complex Compositions

The behaviour of complex lipid mixtures in membranes has been investigated using both atomistic (AT) and coarse-grained (CG) simulation methods. However, full atomistic studies are rarely used due to the limited timescales that can be achieved in the simulations. Typically, AT simulations focus on the molecular rearrangements of pre-formed domains. Meanwhile, the simplicity and speed of CG simulations allows for study of long timescale organisation within membranes, and therefore CG techniques have dominated the study of lateral lipid domain formation.

A bilayer containing a mixture of two lipids, both with different phase transition temperatures, can separate into domains of homogeneous lipid compositions. Indeed, mixtures of di-lauroyl phosphatidylcholine (DLPC) and di-stearoyl phosphatidylcholine (DSPC) are observed to form phase separated domains in experimental systems. The phospholipids both contain unsaturated acyl tails, but they differ in length: having 12 carbons in DLPC and 18 in DSPC. This difference in length results in individual phase transition temperatures of 268 K for DLPC and 328 K for DSPC. However, in early CG simulations a bilayer composed of a mixture of DLPC and DSPC was not observed to phase separate. Interestingly, increasing the concentrations of DLPC had the effect of lowering the phase transition temperature of the whole bilayer (*56*). This general pattern is observed in the experimental phase transition temperatures, though the exact transition temperatures were not reproduced in the CG simulations. Using the same CG model, preferential separation of lipids has been observed in simulations of DLPC and DSPC mixtures. The simulations demonstrated demixing of DLPC and DSPC, with a distinct preference for the lipids to cluster into small groups of the same type (*57*). Limited evidence of a two-phase coexistence also was present in the simulations. In another study, a bilayer containing equal mixtures of DPPE and DPPC showed phase coexistence in CG simulations (*58*). The phase separation was confirmed by slower lipid diffusion coefficients, more ordered acyl tails, and a change to the area per lipid in the gel region.

Mixtures of saturated and unsaturated lipids may also exhibit phase coexistence in presence of CHOL. Initial studies reported domain formation for ternary mixtures of lipids where one of the lipids had two polyunsaturated acyl tails. Indeed, using the MARTINI CG force field, coexistence of gel-like and liquid crystalline phases were observed in mixtures of DPPC/dilinoleyl-phosphatidylcholine (DUPC)/CHOL and DPPC/diarachidonoyl-phosphatidylcholine (DAPC)/CHOL (*59, 60*). The spontaneous separation was also observed in unilamellar vesicles. Phase separation has also been observed in ternary mixtures with one lipid containing a double mono unsaturated tail. However, mixtures containing a saturated lipid, DPPC, mono unsaturated DOPC lipids and cholesterol were seen to spontaneously separate only after the original MARTINI parameter set was systematically modified (*61*).

The incorporation of model peptides into membranes with coexisting gel and liquid crystal phase has been investigated (*62*). The peptides partitioned into the liquid-disordered phase regardless of the hydrophobic mismatch. These observations matched confocal microscopy images, which also showed the peptides to concentrate in the liquid crystal phase. A membrane bound Ras protein, which contains a lipid anchor, was shown to cluster at the interface between liquid crystal and liquid ordered domains (*63*).

Mitochondria can be selectively stained using a cardiolipin-specific fluorescent dye called N-nonyl-acridine orange (NAO) (*64*). The same dye has been used to identify cardiolipin-rich membrane domains in E. coli (*65*). The cardiolipin domains are localised to the cell poles and the site of cell division. Further evidence for the polar localisation of cardiolipin has come from the analysis of the lipid composition of minicells. Minicells are achromosomal products of bacterials strains that are aberrant in cell division. The minicells are produced at the poles of the bacterium, and are enriched for with cardiolipin at the expense of PG (*66*).

Cardiolipin has been shown in MD simulations to increase membrane order and decrease lipid mobility (*67-69*). More recently, CG simulations of cardiolipin have shown hints at domain formation (*70*). The study examined the effect of reducing the charge of the two phosphates of cardiolipin, reflecting the effect of local changes in pH. The addition of a modified cardiolipin to planar lipid bilayers containing a mixture of DOPE and DOPC

resulted in aggregation of cardiolipin and subsequent induction of local negative curvature. Therefore, electrostatic screening of the phosphates of cardiolipin may allow more molecules to pack closer to form localised domains. Furthermore, an overall charge of -1 may also be more realistic at pH 7 as, between them, the two phosphates of cardiolipin have a propensity to trap a proton (*71*).

Asymmetric models of the bacterial OM have also been developed. They display a distinct crosslinking of LPS by Ca^{2+} ions in outer leaflet (*72*). In addition, subsequent MD simulations demonstrated the rigidity of the outer leaflet that resulted from the large size of the LPS molecules and the networks of crosslinking Ca^{2+} ions (*73*). Perhaps unsurprisingly, the complex simulations of the bacterial OM showed that negatively charged molecules were excluded from the surface of the outer leaflet, while the LPS showed greater affinity for positively charged molecules (*74*).

1.6 Molecular Dynamics Simulations of Model Transmembrane Helices Derived from Membrane Proteins

Although many membrane protein structures that have been discovered in recent years, only ~800 out of ~94,000 total proteins structures in the Protein Data Bank (PDB) are of membrane proteins. Furthermore, less than 420 of these are unique membrane protein structures (http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html). Due to their relative paucity in the PDB, membrane proteins have become a primary target for the development of structure prediction methods. Using simple membrane spanning peptides, MD simulations have provided insights into the folding and assembly of helical membrane proteins. In many of the studies described below, the mapping of AT particles back onto CG beads has allowed the synergistic use of ATMD for specific details and CG models for improved sampling.

A two-stage model has been suggested for α-helical membrane protein folding, where transmembrane α-helices are inserted into the lipid bilayer independently of one another prior to assembling into a functional bundle (*75*). Coarse-grained MD simulations of the glycophorin A transmembrane domain have provided supporting evidence for this two-stage model. Indeed, structurally constrained α-helices were able to assemble into a phospholipid bilayer and dimerise, forming a helix bundle that was in good agreement with the NMR determined structure (*76*).

The M2 channel from the influenzae A virus is a simple membrane protein that is formed from a symmetrical tetramer of parallel transmembrane α-helices. The assembly of individual α-helical subunits was observed in a DPPC bilayer using CGMD simulations. The resulting structures corresponded to the closed form of the channel (*77*). Using a similar methodology, the lateral association and dimerization was studied for several different transmembrane α-helices, all of which contained the GxxxG or a similar motif. Destabilizing mutations were observed to modify the distribution of the crossing angles of the dimers. Once the dimers had formed in CGMD simulations, the structures were also refined in full atomistic MD simulations, resulting in a closer agreement with

experimental structures. This demonstrates the utility of a joint CG and atomistic approach in the study of helix-helix interactions (78).

Multiscale simulation approaches have been used to gain insights into the disruption of transbilayer signalling by mutations affecting the transmembrane helices of an integrin heterodimer. The mutations adversely affected the crossing angles of the helices, reducing the conformational stability of the dimers (79, 80).

In ATMD simulations of the epidermal growth factor receptor, ErbB2, a systematic search was performed in order to predict the crossing angle and handedness of two transmembrane helices (81). In CGMD simulations, ErbB2 peptides were shown to dimerise by a different sequence of events in phospholipid bilayers that contain CHOL. Curiously, high concentrations of CHOL favoured a dimerization pathway similar to that observed in pure phospholipid bilayers. The CHOL also altered the tilt angle of the helices in a dose-dependant manner (82). Subtle changes in tilt angle were also observed in mutated variants of the transmembrane helix 2 from Tar chemoreceptor (18). The mutations were in key residues that cause Tar to either be inactive or constitutively active.

In addition to providing insights into the molecular assembly of transmembrane α-helices, MD simulations can be used to predict the structure of the transmembrane regions of proteins. The histidine kinase, PhoQ, is a transmembrane protein that is part of a two-component signalling complex, but the structure of its transmembrane region is not known. A model of this region was generated, however, using a combination of *ab-initio* modelling of the transmembrane α-helices, ATMD simulations, and guidance from crosslinking disulphide scanning data (83).

1.7 Molecular Dynamics Simulations of Bacterial Outer Membrane Proteins

The primary goal of many MD studies, particularly in protein research, is to understand functional movements at an AT level of detail. While CG approaches can be used to enhance sampling, it is an AT level of detail about a proteins dynamics that provides most information for targeted engineering and knowledge based drug design. In addition, AT simulations provide a unique opportunity to study the process of selective uptake of ions and molecules by OMPs. Nevertheless, CG methods have provided valuable details about specific interaction sites for lipids on the membrane exposed surface of OMPs (*84*). The following is a summary of the OMP simulations that have been conducted by other investigators, prior and concurrent to the investigations described in later chapters.

1.7.1 OmpF

Outer membrane protein F (OmpF) is a major porin that allows the passage of water, ions, sugars, polar nutrients and waste across the OM. The X-ray structures of OmpF show the protein to exist in a homo trimeric configuration (*85, 86*). Each OmpF monomer forms a complete β-barrel with a pore in the centre. The pore in each of the monomers is lined by parts of the β-barrel and extracellular loop, L3. The interior position of L3 is the major obstruction in the pore, and it has a predominately negative electric charge. A patch of positive charge is located on the face of the β-barrel opposite L3, and this creates a transverse electric field at the constriction site.

In short, 100 to 300 ps, simulations of OmpF, it was shown that MD could, in principle, describe the correct behaviour of OMPs, when compared to experimental data. As expected, the extracellular loops are many times more flexible than the secondary structural elements (*87*). In addition, these simulations provided clear evidence for the necessity to simulate membrane proteins in their correct membranous environments, as opposed to a vacuum. Indeed, a number of MD studies have investigated the structural dynamics of the OmpF pore in a pseudo-vacuum, where the solvent and bilayer lipids were not explicitly included in the system (*87-89*). These simulations showed large

conformational rearrangements in L3 that were thought to control the voltage gating of the channel. However, the inclusion of explicit water (*90*), and later membrane lipids, detected no large-scale motions in L3 (*91*).

Ion permeation through OmpF was investigated in simulations with an applied transmembrane potential. The translocation of a single ion was observed after 1.3 ns, with a transmembrane electric field applied at 500 mV. A mechanism for channel voltage gating was proposed in which the configuration of counter ions could block the channel (*90*). Insights into the water structure around the pore, fluctuations in the pore radius, and membrane interactions were obtained from 1 ns AT simulations of OmpF (*91*). This simulation study of OmpF was the first to incorporate explicit solvent and membrane lipids. The strong electric field at the pore constriction was found to restrict the movements and orientation of water molecules. Longer simulations of OmpF demonstrated the distinct pathways and solvation characteristics of ions that pass through the pore. Contributions from both water and the protein side chains were required to keep the ions properly solvated (*92*). The ion selectivity of the pore can be modified by mutations of charged residues at the constriction site (*93*). The formation of stable ion pairs within the OmpF pore may also be partly responsible for the gating phenomenon (*94*).

A number of β-lactam antibiotics have been shown to utilise OmpF as a gateway into *E. coli*. There is evidence to suggest that these antibiotics may traverse the OM by passing through the pore of the β-barrel. Indeed, mutations that narrow the pore radius of the β-barrel also inhibit antibiotic uptake (*95*). The passage of various antibiotics through OmpF has been studied using metadynamics simulations. These studies have revealed a potential high affinity site for zwitterionic antibiotics at the constriction site within the OmpF pore that improves the ability of some drugs to traverse the pore (*96, 97*). Flexibility in the molecule was suggested to allow the antibiotic to pass through the constriction site (*98*). Subsequently, these observations of antibiotic binding to OmpF were found to be broadly consistent with X-ray structures of OmpF with ampicillin and carbencillin bound in the pore (*99*). Accelerated MD simulations of OmpF have also identified important residues in OmpF that may maintain the selectivity of the pore to the fluoroquinolone class of antibiotics (*100*).

The diffusion of OmpF trimers was investigated in crowded environment using both atomic force microscopy and MD simulation. The OmpF trimer was found to form stable assemblies with two minimum energy modes of interaction, termed base-to-base and tip-to-tip. The specific protein-protein interactions reduced the overall lateral diffusivity of OmpF, accounting for its unusually slow diffusion when compared to other proteins of similar size (*101*). A preference for tip-to-tip aggregation was observed in CG simulations of OmpF in planar bilayers (*102*). The reduced lateral mobility of lipids in the vicinity of OmpF was also demonstrated.

1.7.2 OmpA

Outer membrane protein A (OmpA) is a general porin of the *E. coli* OM. Molecular dynamics simulations of OmpA have provided insight into a salt-bridge gating mechanism, which may regulate the protein's permeability to water (*103*). These observations were supported by subsequent mutagenesis studies and growth assays, which suggested that OmpA could play a role in the osmoprotection of *E. coli* cells *(104)*. Lipids in direct contact with OmpA have been shown to be motionally restricted in their acyl tails (*105*). A comparative study, using OmpA as model OMP, revealed further details of the reduced lateral mobility of annular lipids when compared to that of bulk lipids (*106*). This reduce mobility has since been observed in simulations of a number of OMPs and IM proteins. Specific lipid protein interactions are key to function of many proteins. For example, the OM protein OmpT from *E. coli* is a protease that requires LPS to be active. Analyses of the contacts made between OmpT and a DMPC bilayer in MD simulations have revealed two potential LPS binding sites (*107*), one that is similar to a site proposed for other OMPs, and second site that also had a high affinity for negatively charged phospholipid moieties.

Further functional insights into bacterial OMPs have been gained through simulations in membrane bilayers that are more representative of the proteins *in vivo* environments. Indeed, the general porin, OprF, from *P. aeruginosa*, which is homologous to OmpA, displayed different dynamics depending on the composition of the surrounding membrane environment. Strong interactions between the extracellular loops of OprF and the saccharide core of LPS were observed in a complex asymmetric membrane, which incorporated LPS in the

outer leaflet of the bilayer (108). Comparing these simulations to those performed in a phospholipid bilayer (109), the strong interactions with LPS appeared to cause the extracellular mouth of the β-barrel to close. These differences in the dynamics of OprF between the two membrane environments may have implications for investigations into the gating of the pore.

1.7.3 TonB-dependent Transporters

The bioavailability of iron is low in the environment due to its insolubility. Therefore, bacteria release scavenger molecules called siderophores to chelate, and thereby solubilise, the iron at its source. High affinity proteins that reside in the bacterial OM bind the iron-chelated siderophores, but they are unable to internalise the molecule without an external supply of energy. Thus, translocation of the siderophore into the periplasm can only occur once an integral protein from the IM called TonB binds to the complex. TonB utilises the proton electrochemical gradient across the IM to facilitate the import of the siderphore.

The structures of three TonB dependent siderophore transporters from *E. coli* have been characterised by X-ray crystallography. These are FhuA, FepA and FecA, which mediate the uptake of ferrichrome, ferric citrate, and enterobactin, respectively. The structures of the three *E. coli* TonB dependent transporters are structurally similar, all forming 22-stranded β-barrels. The crystal structures of FhuA and FecA, with and without the siderophore bound, reveal tantalising insights into the initial stages of siderphore internalisation. In particular, two extracellular loops, L7 and L8, appear to undergo substantial structural reorganisations, with L8 covering the ligand in binding site (Figure 6). Presumably, L8 blocks the escape of the ferric-citrate. Molecular simulations have provided evidence that the dynamics of extracellular L8 are linked to the liganded state of FhuA (*110*). The later stages of siderophore internalisation have been studied using a combination of protein modelling and normal modes analysis. These models have suggested that communication may occur between the loop regions at both ends of the protein (*111*).

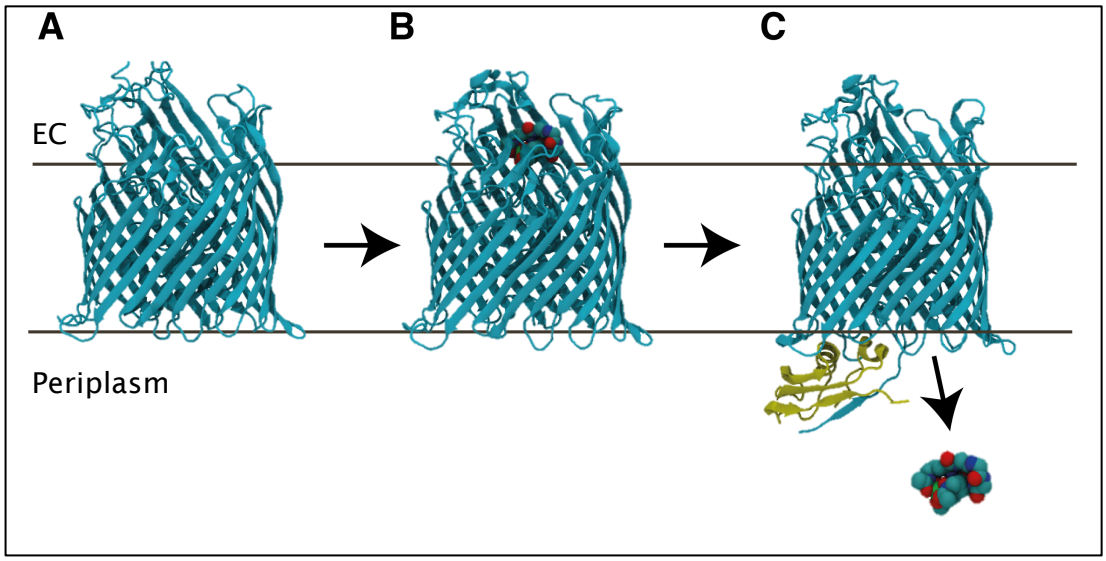


Figure 6. A) apo-FecA. B) holo-FecA with siderophore bound, promoting extracellular loops L7 and L8 to undergo a conformational change, where they block the entrance to the binding site. C) TonB, bound to periplasmic side of FecA, provides the energy for the internalization of the siderophore. The black lines indicate the approximate hydrophobic width of the outer membrane.

1.7.4 The Autotransporters

The autotransporter secretion method is poorly understood. However, the functions of some structural features of the autotransporters have been revealed in MD simulations. Indeed, the extent of the flexibility of the β-barrel translocator domain was revealed in simulations of NalP (*112*). Furthermore, following the removal of the α-helical plug domain, the NalP β-barrel remained structurally stable, with water filling the space previously occupied by the α-helix. More recent evidence suggests that deletion of the α-helical plug impairs autotransporter folding and OM insertion (*113*). However, the structural stability of NalP minus the α-helical plug in (*112*) may demonstrate the stability of final fully folded β-barrel following OM insertion (*49*). Indeed, site-specific cross-linking studies of EspP suggest that the β-barrel is inserted into the OM in a semi-folded state (*114*). Subsequent to OM insertion and folding of the β-barrel, the role of the α-helix may be to restrict access to the pore, and it may not contribute much to the structural stability of the β-barrel.

After release of their passenger domain, some autotransporters, like EspP, have a short α-helix occupying the intracellular entrance of the pore. In these instances, the long extracellular loops may occlude the opposite side of the pore. In the X-ray structure of the post-cleavage EspP translocon, an extracellular loop, L5, is enclosed within the mouth of the β-barrel (*44*). The structural stability of this loop was assessed in steered molecular dynamics simulations. A pulling force, normal to the plane of the bilayer, was applied to L5 in attempt to remove it from the pore of the β-barrel (*115*). Remarkably, L5 was stably inserted into the pore, and it was resistant to removal by both high pulling forces and thermal motion.

An autocatalytic mechanism, which in some autotransporters allows the passenger to be released into the extracellular medium, involves the cyclization of an asparagine inside the pore of the β-barrel (*116*). Using MD simulations of the pre-cleavage structures of Hbp and EspP, it was possible to identify the position of a catalytic water molecule that may help promote asparagine cyclization and cleavage (*45*).

Chapter 2: Computational Methods

2.1 Particle-Particle Interactions in Molecular Dynamics Simulations

Molecular dynamics simulations provide the time-evolved motions of a system of particles with respect to an initial configuration. At its simplest, the MD methodology involves the numerical integration of Newton's equation of motion. The method takes advantage of the relationship between force, $\vec{F}$, mass, m, and acceleration, $\vec{a}$, where the force and acceleration are vector quantities (Equation 1).

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

Equation 1

A series of equations and parameters are used to describe the pairwise interaction energy between particles. The typical form of this set of equations is shown in Equation 2. These describe a potential, $V(r^N)$, which is a function of the positions (r) of N particles.

$$\begin{aligned} V(r^N) = \quad & \sum_{bonds} \frac{1}{2} k_i (l_i - l_{i,0})^2 + \\ & \sum_{angles} \frac{1}{2} k_i (\theta_i - \theta_{i,0})^2 + \\ & \sum_{torsions} \frac{1}{2} k_x \left[1 + \cos(nx - \sigma)\right] + \\ & \sum_{impropers} \frac{1}{2} k_i (\phi_i - \phi_{i,0})^2 + \\ & \sum_{i}^{N} \sum_{j}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right) \end{aligned}$$

Equation 2

The force that acts on each particle, $F(r)$, has a direction opposite (the negative) to the change in potential energy (Equation 3).

$$\frac{dV}{dr} = -F(r)$$

Equation 3

2.1.1 Bonded Interaction Potentials

Many of the bonded properties of molecules, such as the length of the bonds, the bond angles, and out of plane bending (impropers), are modelled using a simple harmonic potential (Figure 7) with the form shown in Equation 4,

$$V(y) = \frac{1}{2}k(y - y_0)^2$$

Equation 4

where k is a stiffness constant, y is the current displacement, and y_0 is the equilibrium position. The potential energies involved in bond stretching over short displacements from the equilibrium position are modelled well with this function. However, at large displacements, which in practice do no occur frequently, the harmonic approximation breaks down. In these circumstances it is necessary to use other potential energy functions, such as the Morse potential, which models the energies of bond stretching more accurately.

The AT simulations described in this thesis have been performed with the bond lengths of non-water molecules constrained to their ideal length using the LINear Constraints Solver (LINCS) algorithm (117, 118). Bonds in water molecules have been constrained using the SETTLE algorithm (119). Constrained bonds allow for an increase in the time step, up to 2 fs, for the calculation of the time dependent evolution of the system, since the high frequency bond and angle vibrations no longer need to be calculated explicitly.

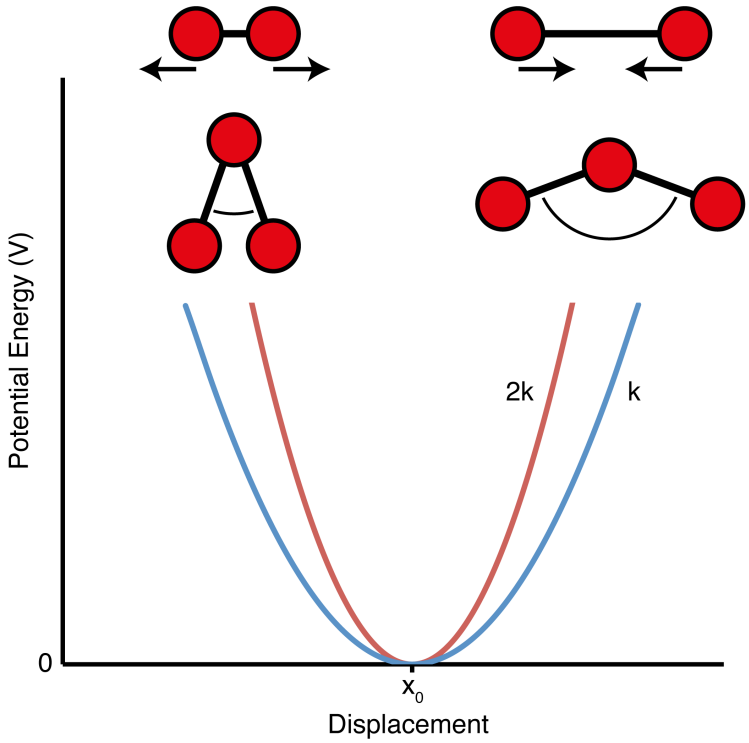


Figure 7. Harmonic potential used to model bond stretching, bond angles and improper dihedrals. The blue and red lines show the effect of changing the force constant.

The potential energy of 360-degree turns (torsions) around chemical bonds is periodic (Figure 8). Thus, these motions are described by a cosine expression with the form shown in Equation 5,

$$V(x) = \frac{1}{2} k_x [1 + \cos(nx - \sigma)]$$

Equation 5

where k_x defines the height of the barrier between minima, σ is the phase shift, and n is the periodicity. The angle, x, is measured between two planes. One plane is formed by the first three particles of a quartet, and the other by the last three particles.

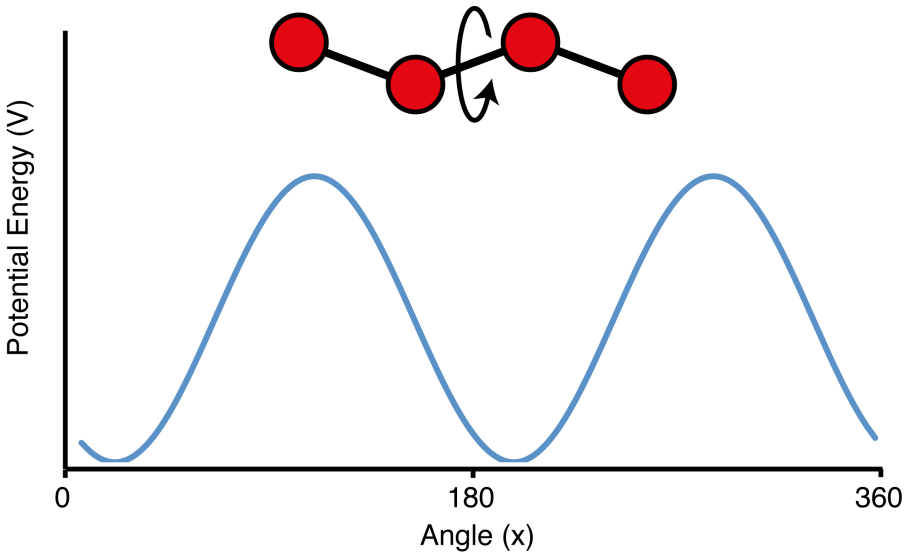


Figure 8. The torsion potential to model rotations around bonds.

2.1.2 Non-bonded Interaction Potentials

The non-bonded interactions between particles are often split into two terms, where one describes the influence of Pauli repulsion and London dispersion (or van der Waals interactions), and the other describes electrostatic interactions.

The van der Waals interactions can be modelled with the Lennard-Jones 12-6 function (Equation 6) (120).

$$V(r_{ij}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$

Equation 6

where ε_{ij} dictates the depth of the potential energy minima, σ_{ij} determines the interaction distance at which the intermolecular potential is equal to zero, and r_{ij} is the distance between particle i and j (Figure 9). When the distance between particles is small, the dominant term in the equation is $\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}$ and its effect is to repel the particles. At intermediate distances, the term, $-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}$,

predominates, which leads to an attraction between the particles. And, at large values of r_{ij} both the terms tend towards zero. However, the repulsive term approaches zero slightly faster than the attractive term, and the function asymptotes at zero but remains slightly negative. The Lennard-Jones potential is attractive from a computational perspective, as r^{12} can be obtained by simply squaring r^6.

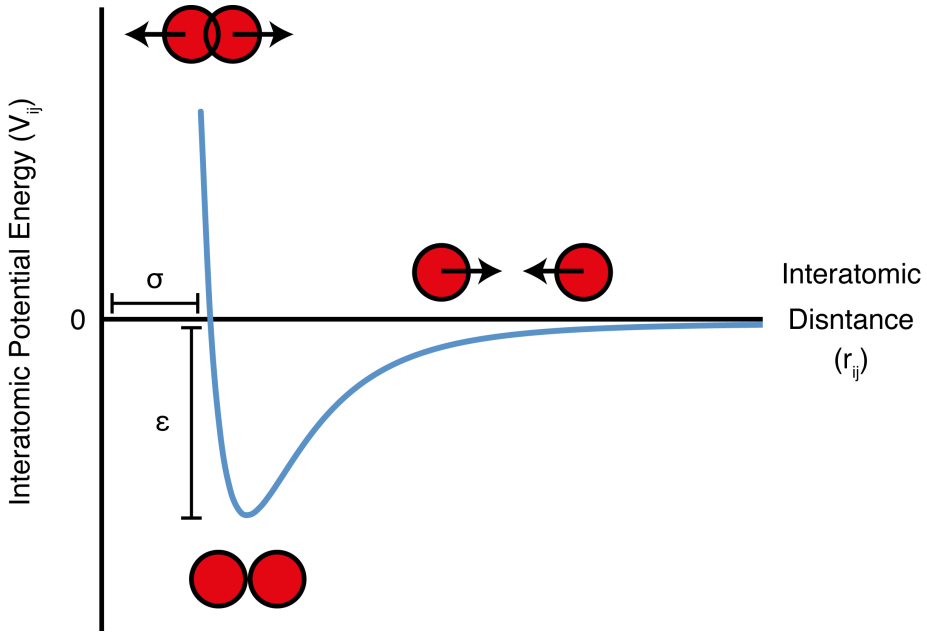


Figure 9. The Lennard-Jones 12-6 potential to model van der Waals interactions. The σ term determines the position at which the intermolecular potential is zero, and the ε term determines the depth of the potential energy well.

The electrostatic potential between charged molecules is calculated from point charges, using Coulombs law (Equation 7)

$$V_{ij}^{qq} = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}}$$

Equation 7

where q is the point charge of either particle i or j, ϵ_0 is the permittivity of electrical charge in a vacuum, ϵ_r is the relative permittivity of electrical charge, and r_{ij} is distance between the point charges of the particles (Figure 10).

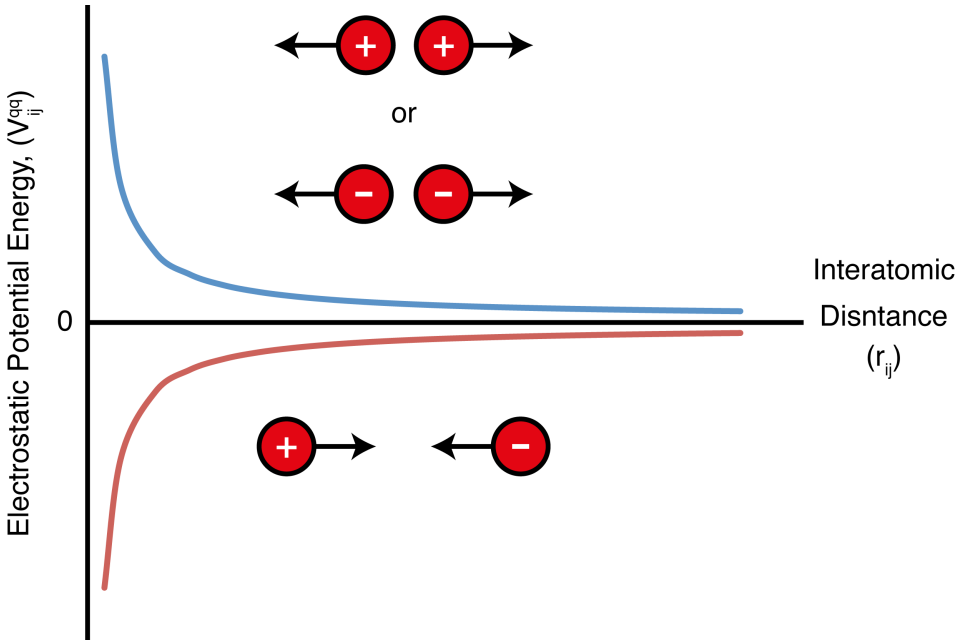


Figure 10. Electrostatic potential for point charges. The blue line shows the intermolecular potential for two like charges, and the red line shows the potential for oppositely charged particles.

2.1.3 Boundary Conditions

The typical number of particles represented in MD simulations is between one thousand and one million atoms, which represents a small volume. To avoid effects caused by interaction with the edges of the simulation box, MD simulations are performed using periodic boundary conditions to simulate the effect of bulk material. Particles exist within a primary cell, but are surrounded by cells that are exact replicas of the primary cell. The cells have open boundaries and particles can move freely from one cell to the next. When a particle exits the primary cell, another particle, identical to the particle that exited, enters from the opposite replica cell.

2.1.4 Long-Range Interactions

The calculation of the non-bonded interactions is a computationally intensive task. It involves the sum over all the pairwise interactions in the system. In order to reduce the computational cost, some potential energy functions are truncated so that certain interactions are not considered beyond a cut-off distance. For example, the Lennard-Jones potential approaches zero quickly after 1.0 to 1.2 nm, and it is often inefficient to calculate the interactions beyond this point. The calculation of electrostatic interactions also can be limited to particles within a certain distance of one another. However, the use of such a cut-off can cause large artefacts, since the interactions beyond an arbitrary cut-off can still have a substantial influence on the potential energy. Indeed, the rate of decay of the electrostatic potential is proportional to the distance, r_{ij}, while the number of interactions with a particle can increase at a rate proportional to the distance squared, r_{ij}^2.

While the influence of individual long-range dispersion interactions is small, it is not negligible when summed over all pairwise interactions. A dispersion correction has been developed to account, partially, for the dispersion energy neglected beyond the cut-off in homogeneous liquids. The correction is applied to the energy and pressure-volume virial at each time step (121).

The Ewald summation method can be used for the calculation of electric potentials (122), and it takes into account the effects of distant electric charges. The method was developed originally to calculate the long-range electrostatic interactions in crystals. In the method, all the point charges are summed across an infinite array of periodic images. The goal is to turn this slowly converging summation into two more rapidly converging sums and a constant. The first sum is performed in real space, while the second is performed in reciprocal space.

The computational cost of the reciprocal part of the Ewald summation increases at the rate, N^2. Therefore, the calculation is not efficient for large systems. The Particle-mesh Ewald (PME) method improves the speed of the reciprocal sum, and it scales as $N\,log(N)$ (123, 124). In the PME method the charges are interpolated onto a grid. The grid spacing determines the accuracy of the calculated electrostatic forces acting on the particles.

2.2 System Configuration and Energy Minimization

2.2.1 Initial Coordinates

The atomic coordinates for proteins are in most cases obtained from experimental data. It is also possible to model entire proteins, and small parts of proteins, using information derived from homologous or similar proteins. In addition, small peptides with a mostly α-helical structure, determined by circular dichroism spectroscopy, can also be modelled by assuming ideal α-helical phi and psi angles around the peptide bond.

In contrast, the lipid coordinates for membrane protein simulations are normally obtained in the form of a homogeneous, pre-equilibrated bilayer from the final snapshot of a MD simulation. The membrane can also be assembled by adding lipids one by one around the protein (125). When a pre-equilibrated bilayer is used, the coordinates of protein must be inserted into the membrane. The simplest method to achieve this is to superimpose the two sets of coordinates, with protein in the appropriate orientation, and then to delete any lipids and solvent that overlap with the protein. However, the deletion of an excessive number of lipids can lead to a sizeable hole in the lipid bilayer, with the lipids loosely packed around the protein. A long equilibration run is then required to close the gap between the protein and lipid bilayer. Thus, a number of techniques have been developed to combine a protein and lipid bilayer with the aim of disrupting the protein structure and the integrity of the membrane as little as possible (126-128).

The program g_membed, which is part of the GROMACS simulation package, is able to insert a protein into a membrane with minimal disruption to the lipid bilayer (126). The program must be supplied with the protein coordinates, in the correct orientation and position, superimposed on those of an intact, pre-equilibrated lipid bilayer. Initially, the protein is compressed along the dimensions orthogonal to the Cartesian z-axis, which has to be the plane of the bilayer in the coordinate file. A small number of lipids and solvent molecules are deleted to create a hole around the compressed protein. Over a series of molecular dynamics steps the protein is expanded to its original size, allowing the lipids to pack closely around the protein, and, thus, shortening the time required to equilibrate the membrane.

2.2.2 Energy Minimization

Overlapping or bad contacts in the structure files are removed using the steepest descent energy minimisation method. The method moves the particles in each step of the algorithm in the direction of the net force vector, the negative gradient of the potential energy. The algorithm is usually implemented so that the particles are moved along the force vector by an arbitrary step size. The first iteration of the algorithm uses a default step size. In subsequent iterations of the algorithm the step size is increased, so long as the step reduces the potential energy of the system. Conversely, if a step creates an increase in the potential energy then the step size is shortened. The algorithm is stopped when either the number of user-defined iterations has been reached or the maximum force component is smaller than the tolerance value.

2.2.3 Initial Velocities

Once the starting positions of the particles have been defined they must be given an initial velocity. These velocities are randomly derived from the Maxwell-Boltzman distribution for a particular reference temperature. The initial velocities are adjusted so that the total momentum of the system is zero.

2.3 Integration Algorithm

In practice, the properties of a large number of particles cannot be calculated analytically. The particles display chaotic behaviour and it is difficult to predict the time-evolved configuration of a system from the starting positions and velocities. Therefore, finite difference methods are used to integrate the equations of motion to generate molecular dynamics trajectories.

All integration algorithms calculate the new atomic positions, velocities and accelerations by approximation with a Taylor series expansion, and a small change in time. The first MD integrator to be developed was the Verlet algorithm. It uses the positions of the particles at time, t and $t - \delta t$, to calculate the new positions, $t + \delta t$. The algorithm is arrived at after deriving a Taylor series expansion for particle positions, r, at both $r(t - \delta t)$ and $r(t + \delta t)$. The sum of the two expressions produces Equation 8, where a is acceleration.

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + a(t)\delta t^2$$

Equation 8

One major disadvantage of the algorithm is that it does not calculate velocities explicitly, as the velocity terms are cancelled out in the summation. This can have implications for the calculation of particle kinetic energies, since the particle velocities have to be estimated. In addition, the Verlet algorithm is not a self-starting algorithm. That is, the positions at $r(t - \delta t)$ are unknown at the start, and thus, in order to begin the integration algorithm, the first time step has to be approximated.

The leapfrog integration method has been used for all the molecular simulations described in this thesis. The algorithm calculates velocities, v, at half time intervals, $\frac{1}{2}\delta t$, $\frac{3}{2}\delta t$, $\frac{5}{2}\delta t$, etc, and new positions at full time intervals δt, $2\delta t$, $3\delta t$, etc. The leapfrog algorithm can be derived from the Verlet algorithm by defining the velocity at half time steps, at $v\left(t - \frac{\delta t}{2}\right)$ and at $v\left(t + \frac{\delta t}{2}\right)$, and

then combining these with Equation 8. The result is a new expression for the positions and velocities (Equation 9 and Equation 10).

$$r(t+\delta t) = r(t) + \delta t v\left(t + \frac{1}{2}\delta t\right)$$

Equation 9

$$v\left(t+\frac{1}{2}\delta t\right) = v\left(t - \frac{1}{2}\delta t\right) + \delta t a(t)$$

Equation 10

The leapfrog algorithm is also not a self-starting algorithm, but it has two advantages over the Verlet algorithm: i) the leapfrog algorithm explicitly calculates the velocities of the particles, and ii) the leapfrog algorithm, unlike the Verlet algorithm, does not require the calculation of the difference of two large floating point numbers, which can introduce a rounding error.

2.4 Thermodynamic Ensembles

The simulations described in this thesis are performed in the NPT ensemble, using a constant number of particles, N, pressure, P, and temperature, T. The temperature and pressure are maintained at a reference value by coupling the particles to a thermostat and barostat, respectively. The thermostats function by altering the particle velocities, while the barostats change box volume and, in some cases, the particle positions.

2.4.1 Thermostats

A number of thermostats have been developed and each have their advantages and disadvantages. The Berendsen thermostat is a robust method that couples all the particles in the system to an external heat bath (129). The bath is fixed at the desired temperature, and it can provide or remove kinetic energy from the system, as required. The heat exchanged is proportional to the difference in temperature between the bath and the system.

$$\frac{dT(t)}{dt} = \frac{1}{\tau}\left(T_{bath} - T\left(t\right)\right)$$

Equation 11

$$\Delta T = \frac{\delta t}{\tau}\left(T_{bath} - T\left(t\right)\right)$$

Equation 12

A coupling parameter, τ, determines how strongly the system is coupled to the heat bath between successive time steps (Equation 11 and Equation 12). The system is weakly coupled to the heat bath when τ is large, and strongly coupled when τ is small.

The actual velocities of the particles in the system are scaled every time step, or series of steps, by a factor, λ (Equation 13 and Equation 14).

$$\left(\lambda^2 - 1\right)T\left(t\right) = \frac{\delta t}{\tau}\left(T_{bath} - T\left(t\right)\right)$$

Equation 13

$$\lambda^2 = 1 + \frac{\delta t}{\tau}\left(\frac{T_{bath}}{T\left(t\right)} - 1\right)$$

Equation 14

The factor, λ, emerges after considering the relationship between the kinetic energy and the system temperature described by the kinetic theory of gases (Equation 15, Equation 16 and Equation 17), where m is mass, v is velocity and k_B is the Boltzmann constant. From this relationship it is clear that the same average kinetic energy between the heat bath and the system can be obtained by scaling all the particle velocities in one step (Equation 18). However, the Berendsen thermostat allows the temperature to be modified over a series of small steps rather than one quick jump. An Instantaneous rescaling of the particle velocities would not allow for fluctuations about the desired temperature, which is necessary to more accurately represent the canonical ensemble.

$$\frac{1}{2}\sum_{i=1}^{N} m_i v_i^2 = \frac{3}{2} N k_B T$$

Equation 15

$$\frac{1}{3Nk_B}\sum_{i=1}^{N} m_i v_i^2 = T$$

Equation 16

$$\Delta T = \frac{1}{3Nk_B}\sum_{i=1}^{N} m_i \left(\lambda v_i\right)^2 - \frac{1}{3Nk_B}\sum_{i=1}^{N} m_i v_i^2$$

Equation 17

$$\Delta T = \lambda^2 T\left(t\right) - T\left(t\right),$$

$$\lambda = \sqrt{\frac{T_{new}}{T\left(t\right)}}$$

Equation 18

The Berendsen thermostat functions by uniformly scaling the velocity of each particle in the system, without any consideration for how the velocities are distributed among the particles of the system. If the solvent is not able to exchange kinetic energy efficiently with the solute, the solvent can become excessively hot and the solute, too cold, but without altering the average temperature. Grouping some of particles together and coupling them to a separate thermostat can alleviate this problem.

The instantaneous velocities created by the Berendsen thermostat do not form a strict canonical distribution. Temperature is a macroscopic property, and it can only be measured as a time averaged property. Therefore, a variation in the instantaneous temperature about a central value is expected. The Velocity-Rescaling thermostat (V-Rescale) is an extension to the Berendsen thermostat in which a stochastic term is added to create artificial fluctuations that are derived from the correct canonical distribution (*163*).

2.4.2 Barostats

The amount of volume fluctuation in an isobaric system is related to the isothermal compressibility, κ (Equation 19)

$$\kappa = -\frac{1}{V}\left(\frac{\delta V}{\delta P}\right)_T$$

Equation 19

where V is the volume and P is the pressure. This coefficient represents how much the volume of a substance is able to change as a result of a compressive force. Fluids are relatively incompressible, and have small values for κ. The isothermal compressibility of lipids, $\kappa = 0.00005\ bar^{-1}$, is similar to that of water, $\kappa = 0.000045\ bar^{-1}$, in the 0 to 100 °C range.

A constant pressure can be maintained with the Berendsen barostat (129), which functions by scaling the box volume and particle positions, in a way comparable to the scaling of particle velocities in the Berendsen thermostat. In analogy to the thermostat, the Berendsen barostat couples the system to an external pressure bath, with a coupling constant, τ, to ensure that the change in pressure happens gradually, over a series of steps.

$$\frac{dP(t)}{dt} = \frac{1}{\tau}\left(P_{bath} - P(t)\right)$$

Equation 20

$$\lambda = 1 - \kappa\frac{\delta t}{\tau}(P - P_{bath})$$

Equation 21

The volume of the box is scaled by a factor, λ, (Equation 20 and Equation 21). In an isotopic coupling scheme each box vector will be scaled by a factor, $\lambda^{1/3}$. Since the pressure is related to the virial, where the virial tensor is a product of atomic positions and the force acting on them, the atomic coordinates can also be scaled. As such, the new atomic positions are given by $r_i' = \lambda^{1/3} r_i$.

In membrane simulations it is useful to scale the *xy* box vectors, which is typically the plane of the bilayer, independently of the *z*. This is known as semi-isotropic pressure coupling. This allows the membrane to equilibrate independently of the solvent.

The Berendsen barostat will provide the correct average pressure during a simulation. However, it does not generate the correct canonical ensemble of pressure and volume fluctuations. The correct fluctuations can be achieved using the Parrinello-Rahman pressure-coupling scheme (130). In this scheme the equations of motion are modified, but there is no instantaneous change in the positions of atoms. Another advantage of the Parrinello-Rahman pressure-coupling scheme is that it can be used for any shape of simulation cell.

2.5 Coarse Grained Simulations

The desire to simulate events that occur on longer timescales than can be reached with AT simulations has led to the development of CG molecular force fields. These force fields group together selected clusters of atoms into a single interaction centre, thereby reducing the number of interactions calculated at each time step. Furthermore, the absence of high-frequency bond oscillations and a smoothed potential energy landscape allows for increases in the time step between force calculations to be in the range of 10 to 50 fs.

A number of CG models have been developed for phospholipid membranes (*131-133*). The advantages of the Martini CG model over its counterparts are its speed, easy of use and extensibility. The MARTINI model, in addition to phospholipids, contains parameters describing molecules such as amino acids (*76, 134*), detergents (*134*) and DNA (*135*).

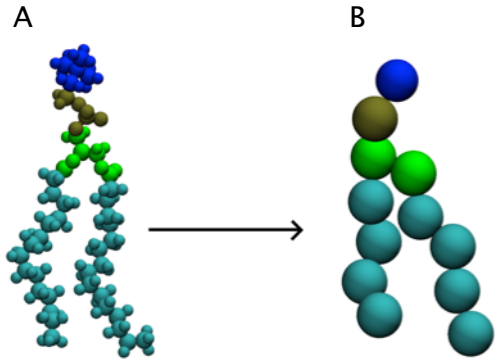


Figure 11. Mapping of DPPC from an (A) AT representation to a (B) CG representation. The choline group is coloured blue, the phosphate moiety is coloured brown, the glycerol is coloured green, and the acyl tails are coloured cyan.

The MARTINI CG model combines four heavy atoms into one interaction centre (Figure 11). Though, it is not a strict mapping, and three or five heavy atoms can be represented by a single particle (*134, 136*). Thus, a typical lipid molecule such as DPPC is reduced to 12 particles in MARTINI, down from 130

particles in an all atom model. Soft harmonic potentials between CG particles are used to maintain the approximate shape of the AT molecule. These harmonic potentials are equivalent in form to those for bond lengths and bond angles in atomistic simulations (Figure 7).

The solvent in CG models can be modelled implicitly or explicitly. The CG models that use an explicit representation for water often group a number a number of water molecules into a single interaction site. In the MARTINI model, four water molecules are grouped into a single interaction site. The new particle has zero overall charge. Thus, the solvent is represented in the MARTINI model as a simple apolar Lennard-Jones fluid. The lack of partial charges in the water model means that it does not screen electrostatics as water would normally. Consequently, a dielectric constant, $\epsilon_r = 20$, for electrostatic interactions is required to represent the permittivity of electric charge through water. The Martini force field has been extended to include a polarisable water model that resolves some of the issues with the apolar water model (*137*).

The validity of martini model has been assessed by reproduction of experimental observables. The MARTINI water model has been parameterised so that its density and compressibility closely match that of normal water. A similar approach was taken for the parameterisation of aliphatic carbons, but, additionally, energies of hydration, vaporization and portioning were used to calibrate the particle interaction potentials. Furthermore, other properties of phospholipids are used to adjust the force field parameters, such as the rate of diffusion, the average area occupied in a bilayer and the crystalline phase behaviour.

The protein model in MARTINI has been through a number of development iterations. In addition, an independent protein model has been developed that use the original MARTINI particle types, but with a modified elastic network to maintain the secondary structure of the protein (76). The Elastic Network Model (ENM) consists of harmonic potentials connecting CG particles. In effect the harmonic potentials serve as distance constraints between structural elements of the protein. The force constant and equilibrium position used in the harmonic potential is dependent on the structural component being modelled. For example, a force constant of $35\ kJ\ mol^{-1}\ rad^{-2}$ and equilibrium

position of 90° is used for bond angles between backbone particles in α-helices, while a weaker force constant of $25\ kJ\ mol^{-1}\ rad^{-2}$ and equilibrium position of 120° is used for residues in a random coil. The stability of secondary structural elements is enhanced further by a set of relatively strong, $1000\ kJ\ mol^{-1}\ nm^{-2}$, distance constraints that mimic hydrogen bonds.

New molecules can be parameterised for the MARTINI force field by using information obtained from AT simulations (*138*). In this method, the CG particles are mapped on to the AT molecule, following the examples set out in the development of the force field. The distribution of bond lengths and angles can then be matched by empirical refinement to the distributions obtained from AT simulations.

Chapter 3: Stability and Membrane Orientation of the Fukutin Transmembrane Domain

3.1 Abstract

The N-terminal transmembrane domain of fukutin is implicated in the localisation of the protein within the Golgi apparatus. It has been proposed that the transmembrane domain mediates retention by recognising the thinner lipid bilayer of the Golgi. Molecular dynamics simulations have been used to assess the structure, stability, and orientation of the short 36-residue N-terminal peptide, representing the fukutin transmembrane domain, in lipids with differing tail lengths. These simulations demonstrate that the N-terminal peptide adopts a stable α-helical conformation in phosphatidylcholine lipids when oriented with its principal axis perpendicular to the bilayer plane. The stability of the helix is largely insensitive to the lipid tail length, preventing hydrophobic mismatch by virtue of ability to tilt within the lipid bilayers. This suggests that changes in the tilt of the fukutin transmembrane domain in response to bilayer properties may be responsible for trafficking of the full protein. Coarse-grained simulations of the complex Golgi membrane suggest the N terminal domain may also induce the formation of microdomains in the surrounding membrane by preferentially interacting with 1,2-dipalmitoyl-sn-glycero-3-phosphatidylinositol 4,5-bisphosphate (PIP2) lipids.

The study presented in this chapter has been published as part of a journal article in Biochemistry (*173*). A full copy of this article is included in Appendix 8.2. The contributions of two authors to this article are discussed briefly in the following chapter. Of note, Yuk Ming Leung performed the initial atomistic simulations of the FK1TMD in a DMPC bilayer. These simulations were later extended, and repeated. Additionally, Dr Thomas Piggot developed the coarse-grained parameters for the PIP2 lipids.

3.2 Introduction

3.2.1 The fukutin Glycosyltransferase

Fukutin is a single pass transmembrane (TM) protein that is normally found associated with the Golgi apparatus (GA) (*139*). The topology of the fukutin protein is such that its N-terminal tail is located in the cytosol while its C-terminal globular domain is located in the lumen of the GA. The precise function of the fukutin globular domain remains unclear but there is strong evidence to suggest that it acts as a glycosyltransferase involved in the O-linked glycosylation of α-dystroglycan, a major structural muscle protein. Indeed, it has been found that patients with Fukuyama-type muscular dystrophy contain a much higher proportion of unglycosylated α-dystroglycan (*140, 141*). In addition, the presence of a conserved DxD motif in the amino acid sequence of the fukutin globular domain further supports the view that fukutin operates directly in catalysis (*142, 143*).

The N-terminus of fukutin is highly hydrophobic, and is enriched with aromatic residues. Indeed, secondary structure prediction algorithms suggest that the N-terminal residues, 7 to 33, of fukutin exist as a transmembrane α-helix. Furthermore, a peptide representing the first 36 N-terminal amino acids of fukutin has been shown by Circular Dichroism to adopt a predominantly α-helical conformation in membrane mimetic environments (*144*).

3.2.2 Glycosylation and Sorting within the Golgi Apparatus

Glycosylation of proteins begins in the Endoplasmic Reticulum (ER) when specific asparagine (N) residues are covalently linked to a branched oligosaccharide that consists mostly of mannose. The O-glycosylation of serine and threonine hydroxyl groups, and most of the subsequent assembly and modification of N- and O-linked oligosaccharides, occurs within the GA. Originally, it was believed that the many membrane enclosed compartments (cisternae) of the GA allowed enzymatic events to be spatially separated. This compartmentalization of processes was able to explain the directed way in which oligosaccharide chains are assembled. Indeed, enzymes that acted earlier in oligosaccharide assembly pathways were found to exist in separate compartments to later acting enzymes (*145*). Other studies using antibodies

also showed the segregation of glycosylation enzymes into different cisternae (*146*). More recent work, however, has shown that glycosylation enzymes are not completely compartmentalized and, instead, they have a non-uniform distribution across the cisternae from cis to trans(*24*). This is to say that each cisterna was observed to contain a unique mixture rather than a unique set of enzymes. The GA also appears to contain a unique phospholipid composition when compared to the compositions of the ER and plasma membrane (*10*).

Protein retention within the GA is a highly dynamic process. Both antero-and retrograde transport steps occur between different membrane enclosed compartments (*147*). The mislocalisation of glycosylation enzymes within the GA can result in the aberrant processing of glycoproteins, and this is a feature of certain muscular dystrophies (*148*).

Retrograde transport within the ER/GA involves receptor-mediated recognition of an extramembranous motif, a C-terminal KDEL/HDEL sequence (*149*). In contrast, the correct sorting of membrane proteins within the GA is thought, in part, to involve interactions of transmembrane domain with the membranes of the Golgi cisternae. Indeed, the lipid composition varies between the ER and plasma membrane, with cholesterol and sphingomyelin becoming more prevalent towards the plasma membrane (*150*). Furthermore, many GA resident proteins are observed to have transmembrane domains with much shorter hydrophobic regions than plasma membrane proteins (*151*). Thus, the interactions of glycosylation enzymes with the unique lipid and protein environment in the GA could lead to retention within specific compartments.

Glycosyltransferases attached to synthetic transmembrane domains have been used to probe the effect on localisation when the hydrophobic length is varied. In these studies, it was demonstrated that the correct localisation within the GA appeared to depend on the thickness of the hydrophobic region (*149, 152, 153*). These results, however, have to be interpreted with caution since the effect of the synthetic transmembrane domain on the globular domain is unknown.

Retention of ER/GA resident proteins may result from the interaction of the glycosyltransferases transmembrane domain with the distinctive lipid composition found within the ER/GA. Oligomerisation of transmembrane

domains may also follow subsequent to partitioning into the correct lipid environment (*147, 154*).

3.2.3 Simulations of FK1TMD

The following describes a molecular dynamics (MD) study into the structure and membrane orientation of the fukutin transmembrane domain. The N-terminal transmembrane domain of fukutin has been shown to be sufficient to retain the protein within the GA (*155*). A short peptide corresponding to the first 36 N-terminal amino acids of fukutin (FK1TMD) was modelled as an idealized α-helix. The stability of this model was subsequently assessed in atomistic molecular dynamics (ATMD) simulations. The conformational stability of the FK1TMD was assessed in phospholipid bilayers of varying hydrophobic thickness and head group composition.

Coarse-grained molecular dynamics (CGMD) simulations have been used to study the tilt angle of the FK1TMD within phospholipid bilayers. The amino acids that make specific contacts with the phospholipid headgroups particles are also determined. The tilt angle and protein lipid contacts are also investigated for membranes of differing hydrophobic thickness and head group composition. Finally, the FK1TMD was simulated in a membrane with a headgroup composition more representative of the GA. Examples of both the AT and CG representations of the FK1TMD can be seen in Figure 12A and Figure 12B, respectively.

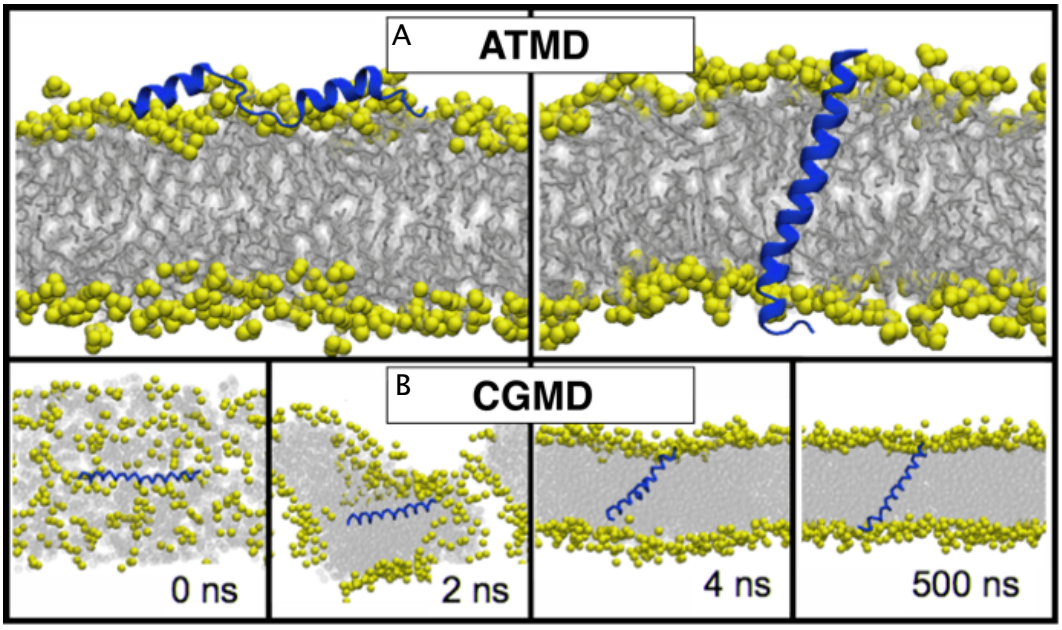


Figure 12. A) The top two panels display the final snapshots (after 50 ns of simulation) from ATMD simulations of the FK1TMD in DMPC. The top left panel displays the simulation of FK1TMD in which it was initially positioned with its helical axis parallel to the surface of the bilayer. The top right panel displays a simulation in which the FK1TMD was inserted initially in a transmembrane arrangement within the bilayer. B) DPPC bilayer assembly and protein insertion from CGMD simulations (bottom). Lipid headgroups are yellow, the tails are grey and the FK1TMD is blue. Water and ions are omitted for clarity.

3.3 Methods

3.3.1 Environments

The ATMD simulations of FK1TMD were conducted in (i) water, (ii) 1 M NaCl, (iii) a 1,2-dimyristoyl phosphatidylcholine (DMPC) bilayer, and (iv) a 1,2-dipalmitoyl phosphatidylcholine (DPPC) bilayer. The insertion of FK1TMD into (i) 1,2-dilauroyl phosphatidylcholine (DLPC) (ii) DPPC, (iii) POPC, (iv) POPS, and (v) mixed lipid bilayers was studied using CGMD simulations.

3.3.2 Generation of the FK1TMD Model

The model of the 36-residue transmembrane domain of fukutin (FK1TMD) was created using Modeller 9v7 (*156*) by constraining the FK1TMD sequence to an idealized α-helix. The resulting helix was evaluated for stereochemical integrity using Procheck (*157*). This model was used for all of the atomistic simulations and also for creating the CG version of FK1TMD.

3.3.3 Details of Atomistic Simulations

The AT simulations were performed using GROMACS 4.0.7 (www.gromacs.org) (*158-160*). The simulations used an extended united atom version of the GROMOS96 force field (*161*). Berger parameters were used for phospholipids (*161*). Explicit water molecules were modelled using the SPC water model (*162*). The charge of the simulation systems was neutralised with Na^+ and Cl^- ions. The LINCS algorithm was used to constrain the lengths of all lipid and protein bonds (*117*), while the SETTLE algorithm was used for water molecules (*119*). This allowed the use of a 2 fs time step. The system temperature was maintained with the velocity-rescaling thermostat, with a time constant for coupling of 0.1 ps (*163*). The system pressure was maintained at 1.0 bar using the Berendsen barostat, with a time constant of 1.0 ps (*129*). The potential energy of electrostatic interactions were calculated using the particle mesh Ewald (PME) method, with a cutoff of 1.0 nm for the calculation of direct space interactions (*123*). The energies of van der Waals interactions were calculated using the Lennard-Jones potential, also using a cut-off of 1.0 nm. A long-range dispersion correction was applied to the energy and pressure to account for van der Waals interactions beyond the cut-off. The neighbour list was updated

every 10 steps. A summary of the system components of each simulation is shown in Table 1. All simulations were run for 50 ns. The conformational properties of FK1TMD were analysed using GROMACS tools and DSSP (*164*). Visualization was conducted with VMD (*165*).

Table 1. Summary of the AT simulations. The Cα RMSD values reported are the mean values calculated for the last 25 ns of each simulation.

Simulation	System components				Temp (K)	Cα RMSD (nm)
	FK1TMD	Lipids	water	Ions		
A1_wat	1	0	39,654	4 Cl-	310	1.36 ± 0.04
A2_wat	1	0	38,208	727 Cl-, 723 Na+	310	0.60 ± 0.10
A3_Sfdmpc	1	512 DMPC	23,323	4 Cl-	310	0.43 ± 0.03
A4_25dmpc	1	497 DMPC	33,423	4 Cl-	298	0.37 ± 0.03
A5_37dmpc	1	497 DMPC	33,423	4 Cl-	310	0.33 ± 0.04
A6_43dmpc	1	497 DMPC	33,423	4 Cl-	316	0.13 ± 0.03
A7_25dppc	1	502 DPPC	31,611	4 Cl-	298	0.13 ± 0.02
A8_37dppc	1	502 DPPC	31,611	4 Cl-	310	0.28 ± 0.02
A9_43dppc	1	502 DPPC	31,611	4 Cl-	316	0.26 ± 0.03
A10_43popc	1	501 POPC	28,352	4 Cl-	316	0.30 ± 0.03

3.3.4 Details of Coarse Grained Simulations

All the CG simulations were performed using GROMACS 4.0.7 (www.gromacs.org) (*158-160*). The simulations were initiated from a random configuration of protein, lipids, ions, and water, which were left to assemble into a lipid bilayer in the presence of one FK1TMD molecule (*76, 77, 134*). The CG parameters for DPPC, DLPC, POPC, ions, and water molecules were obtained from the MARTINI CG force field (*134*). The parameters for the CG 1,2-dipalmitoyl-sn-glycero-3-phosphatidylinositol 4,5-bisphosphate (PIP2) lipid headgroup were obtained by matching the bond and angle distributions of a CG inositol 1,4,5-trisphosphate (IP3) (*138*) to the distributions obtained from AT simulations using the GROMOS carbohydrate parameters (*166*). The glycerol and fatty acid portions of the PIP2 lipid parameters were the same as those used for the CG DPPC lipid. Parameters for amino acids are given in (*76*), (*167*), and (*168*). The α-helical structure of the FK1TMD was maintained using an elastic network, as described in (*76*). For simulations of FK1TMD in the more complex GA membrane, the MARTINI2.0 force field was used due to the greater number of lipid models available (*136*). The complex GA membrane was composed of 50% PC, 20% phosphatidylethanolamine (PE), 12% PIP2, 8% phosphatidylserine (PS), and 10% sphingomyelin (SM). All phospholipids had palmitoyl fatty acid tails (four coarse grain beads) in both the *sn*-1 and *sn*-2 positions of the glycerol backbone. The ratio of phospholipid to cholesterol was 16:1 (0.06). The ratio of lipids was chosen to be similar to that of the GA (*169*). As with the AT simulations, the analyses of the CG simulations were performed using GROMACS tools and locally written code. Molecular images were produced with VMD (*165*).

Table 2. Summary of FK1TMD CG simulations.

Simulation	System components				Temp (K)
	FK1TMD	lipids	Water	Ions	
CG1_dlpc	1	266 DLPC	3140	-	323
CG2_dppc	1	266 DPPC	3140	-	323
CG3_popc	1	266 POPC	3140	-	323
CG4_pops	1	266 POPS	3127	370 Na 3 Na+	323
CG4_real	1	200 DPPC 80 DPPE 32 DPPS 48 PIP2 40 SM 66 CHOL	6736	264 Na+	323

3.4 Results

3.4.1 Atomistic Simulations

Using ATMD, the structural stability of the FK1TMD model was assessed in an aqueous environment. Much of the α-helical structure of the FK1TMD was lost within the first 10 ns of simulation. By the end of the 50 ns simulation, only residues V8 to L13 and L25 to H30 were in a α-helical conformation (Figure 14A and Figure 14B). A similar pattern α-helical structure loss is also observed for the FK1TMD in aqueous simulations containing a 1 M concentration of NaCl (Figure 14B). Therefore, it appears that the burying of hydrophobic residues, and not electrostatic and hydrogen bond interactions, results in the loss of α−helical structure in water. The NaCl ions seemed to slow the rate of structural drift of the FK1TMD from the initial ideal α-helical conformation, with an RMSD after 20 ns of 1.30 nm in pure water and 0.5 nm in 1 M NaCl (Figure 13). This may suggest that polar interactions between amino acids also help to destabilise the α-helix. However, this cannot be concluded from the current data, as there is only one example of the transition.

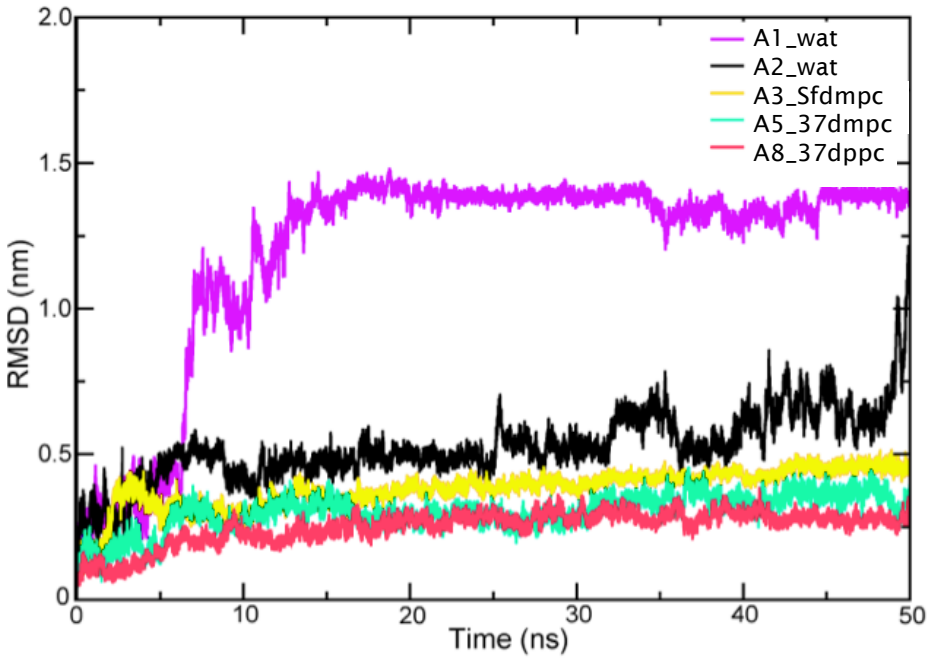
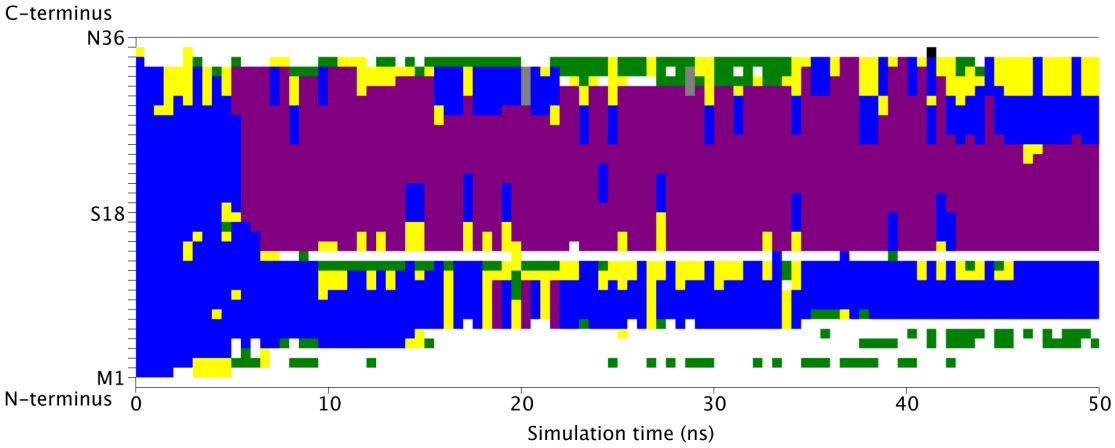


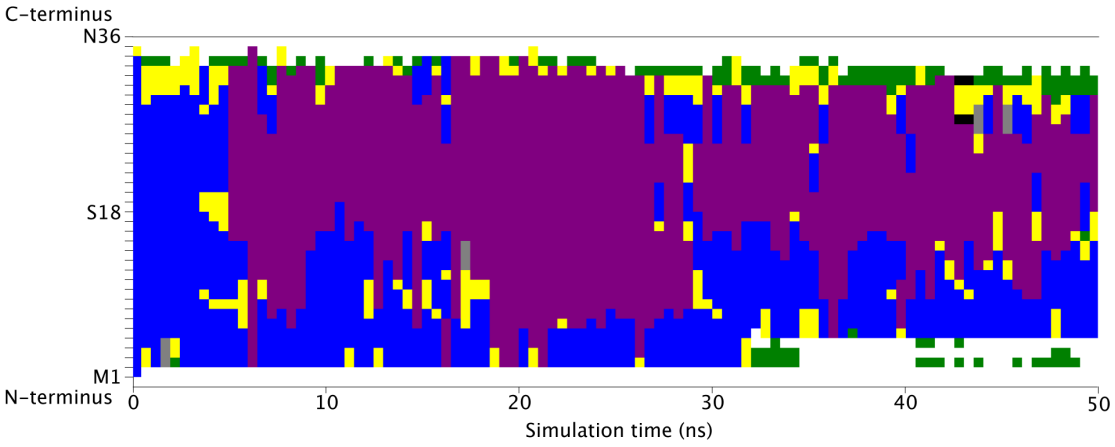
Figure 13. RMSD of the backbone (Cα atoms) of FK1TMD after a least squares fit to the initial, idealized α-helix model. The details of the simulations referred to in the legend, are provided in Table 1.

Figure 14. DSSP matrices of the FK1TMD in a) water at 310K, b) 1M NaCl at 310K, c) the water-headgroup interface of a DMPC bilayer at 310K, d) a transmembrane orientation within a DMPC bilayer at 310K, and e) a transmembrane orientation within a DPPC bilayer at 310K. In the DSSP matrices (a to c), each square represents the structure assignment of a particular residue at an instance in time. Blue coloured squares represent residues in an α-helical conformation; mauve represents residues in a π-helix conformation; grey represents residues in a 3_{10}-helix conformation; black represents residues in an isolated β-bridge conformation; green represents residues in a bend; yellow represents residues in a hydrogen bonded turn; and, white represents residues that have adopted an undefined structure.

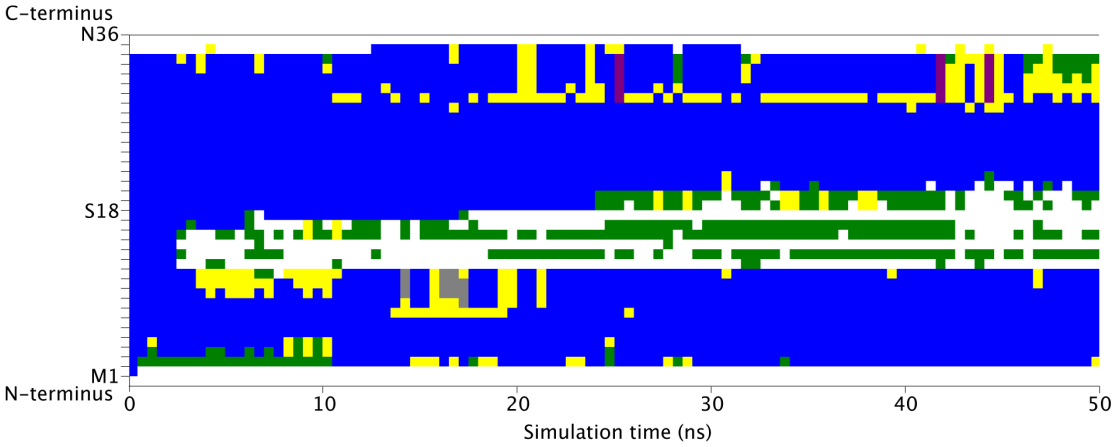
A)



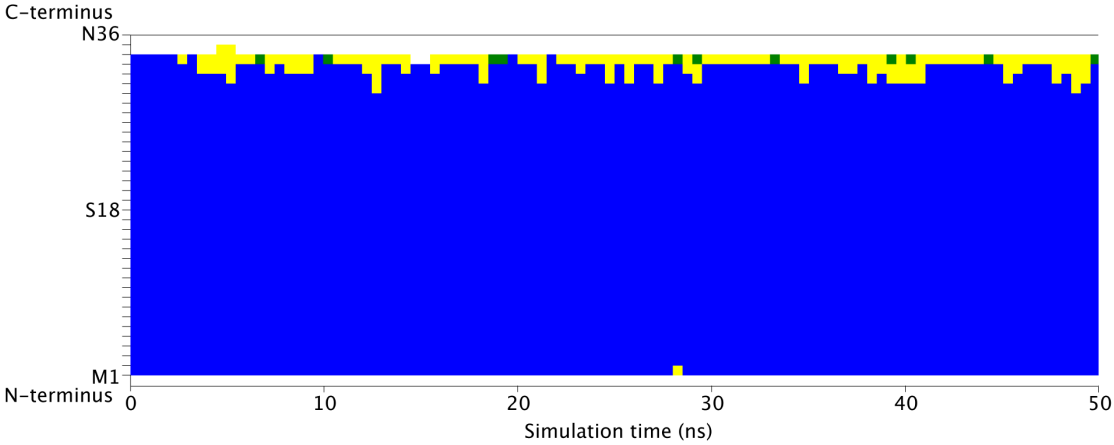
B)



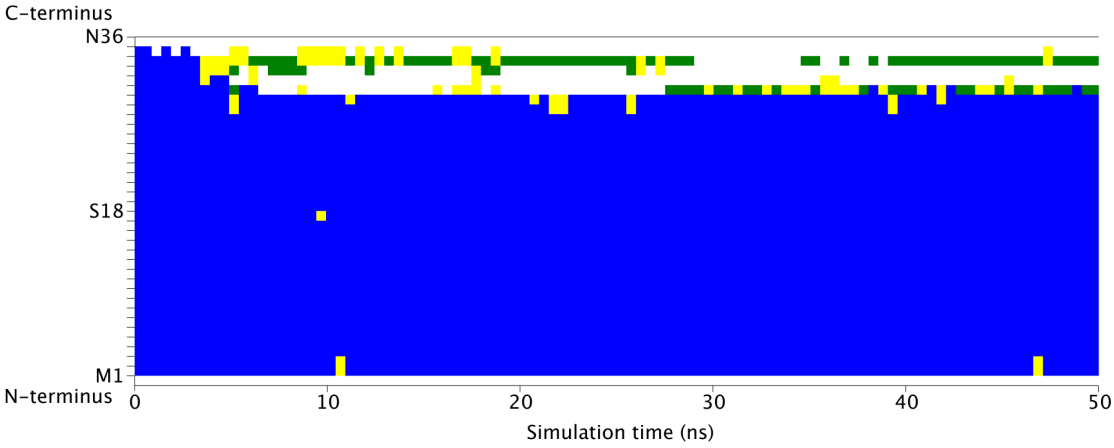
C)


C-terminus
N36
S18
M1
N-terminus
0
10
20
30
40
50
Simulation time (ns)

D)


C-terminus
N36
S18
M1
N-terminus
0
10
20
30
40
50
Simulation time (ns)

E)


C-terminus
N36
S18
M1
N-terminus
0
10
20
30
40
50
Simulation time (ns)

The stability of the FK1TMD model was also evaluated in the vicinity of a phospholipid bilayer. The simulations were initiated with the FK1TMD model in two positions relative to the phospholipid bilayer (i) in the water-lipid interface just above the headgroups of a DMPC bilayer and (ii) in a transmembrane orientation, in which the principal axis of the α-helix was perpendicular to the bilayer plane. Perhaps unsurprisingly, given that the FK1TMD exists *in vivo* in the ER/GA membranes, the secondary structure of the model was more stable near a phospholipid bilayer (Figure 14C) compared to that in just water or the 1 M NaCl environments (Figure 14A and Figure 14B). Indeed, when positioned just above the lipid headgroups with the helical axis parallel to the bilayer plane, the FK1TMD retained its α-helical structure in two short regions at either end of the peptide (Figure 12A and Figure 14C). Meanwhile, the region between these two α-helices (L13 to F20) became unstructured (Figure 14C). In contrast, when inserted into the hydrophobic core of the bilayer in a transmembrane orientation, the FK1TMD generally remained in a stable α-helical conformation (between R3 and H30), although some unfolding was observed at the C-terminal end (Y31 to N36) (see Figure 14D and Figure 14E). The FK1TMD model was also simulated in a DPPC lipid bilayer to assess the effect of the bilayer thickness on the structural stability. The hydrophobic core region of DPPC bilayers is thicker, as a result of its longer fatty acid tails, than that of DMPC. The pattern of unfolding at the C-terminal end observed in the DMPC simulations was also present in the DPPC simulations (Figure 14E). The structural drift of the FK1TMD was measured by calculating the root-mean-square deviation (RMSD) of the backbone Cα atoms from their initial ideal α-helical conformation. The RMSD values were similar in both lipid environments, with an average, over the last 25 ns of simulation, of 0.33 nm in DMPC compared to 0.28 nm in DPPC (Figure 13).

3.4.2 Coarse Grained Simulations

The preferred membrane localization and orientation of the FK1TMD was studied on longer time scales using CGMD, in which four heavy atoms are replaced with a single, spherical particle, to reduce the complexity of the system. A number of examples exist where CGMD has permitted the study of bilayer self-assembly and the insertion of peptides into membranes (*76, 77, 170*). Here, an elastic network was used to constrain the secondary structure of the FK1TMD model to an idealized α-helix. Each of the 1 μs simulations was initiated from a random arrangement of lipids, peptide, water, and counter ions to allow an unbiased insertion of the FK1TMD into a self-assembled lipid bilayer. The effect of the bilayer thickness on the orientation and dynamics of FK1TMD was assessed by simulating the idealized α-helical model in DLPC (three particles in both tails), DPPC (four particles in both tails) and POPC (four particles in one tail and five in the other tail) phospholipids. The bilayer self-assembly proceeded in all cases via the stalk formation mechanism that has been reported previously for DPPC lipids (*76, 134*). By the end of each CG simulation, the FK1TMD adopted a transmembrane orientation within the membrane, with the principle axis of the α-helix oriented approximately perpendicular to the plane of the bilayer (Figure 12B). In a few cases, the FK1TMD was oriented parallel to the plane of the bilayer at the start of the simulations, particularly in DLPC lipids. However, one end of the α-helix would insert into the hydrophobic core of the bilayer after a short period of time (less than 75 ns), followed by full insertion of the rest of the α-helix, so that the FK1TMD was traversing the bilayer. Once the FK1TMD had inserted into the bilayer, it remained in a transmembrane orientation for the rest of the simulation, although its precise tilt with respect to the bilayer plane was observed to fluctuate. The tilt angle of the FK1TMD, calculated for the final 200 ns of the CG simulations, was 39 ± 1.5° in DLPC and 30 ± 1.5° in DPPC. Presumably, the greater average tilt angle in the shorter DLPC lipid tails is a result of a degree of mismatch between the hydrophobic region of the protein and the lipid tails. The standard error was estimated by fitting either a single or double exponential to the block-averaged tilt-angle data (*171*). In block analysis the time sequence data are divided into blocks of equal size, and the number of divisions is reduced sequentially until the standard error of the block averages reaches a plateau.

Additional CGMD simulations were performed using POPC phospholipids in order to investigate, further, the effects of the lipid tail composition on the tilt angle of the FK1TMD. The POPC CG lipid model implemented within MARTINI has five particles for the oleoyl tail and four particles for the palmitoyl tail, compared to the four particles for each palmitoyl tail in DPPC. The oleoyl tail also contains an unsaturated bond, which causes the tail to bend outwards at the end. In agreement with the trend observed in the DPPC and DLPC simulations, of a smaller tilt angle in longer lipid tails, the mean tilt angle of FK1TMD calculated during the final 200 ns of the CG simulations was 17.6 ± 0.9° (Figure 15).

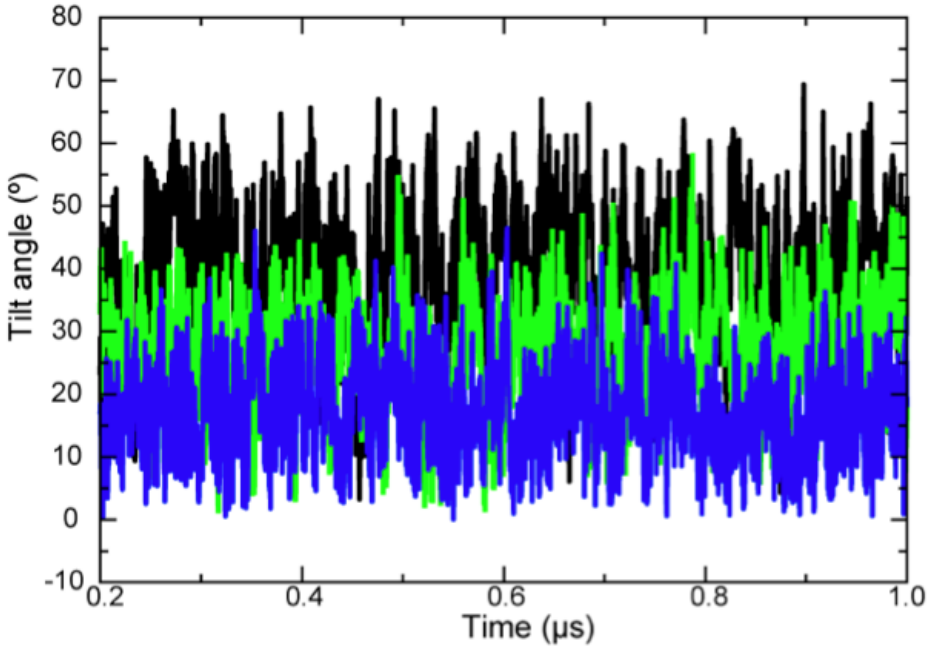


Figure 15. The tilt of the FK1TMD with respect to the bilayer plane in DLPC (black), DPPC (green) and POPC (blue) bilayers.

The correct localization of some glycosyltransferases within the GA requires that a conserved motif, R/K-x-x-R/K, is present at their N-terminus (*172*). The R/K-x-x-R/K motif of FK1TMD (R-I-N-K) is located between residues R3 and K6. The role of this motif in membrane localization and orientation of the FK1TMD was investigated by examining its interactions with the polar headgroup particles of the bilayer lipids. The number of contacts was calculated (contact

defined as an interatomic distance of less than 6 Å) between the protein and the phospholipid headgroup particles by sampling the last 100 ns of the 1 μs CGMD trajectories at 0.4 ns intervals. The protein-lipid contacts for DPPC and DLPC are presented in Figure 16. In all three of the lipid bilayers (DLPC, DPPC, and POPC), the N-terminal residues between M1 and N7 made regular contacts with the phospholipid headgroup particles. In particular, R3 and K6 had a strong propensity to interact with lipid headgroup particles, and, together, they accounted for >20% of all headgroup-protein contacts in all three lipid bilayer simulations combined (Figure 16). Presumably, these charged residues help to anchor the N-terminal end of the FK1TMD in the zwitterionic headgroup region of the lipid bilayer.

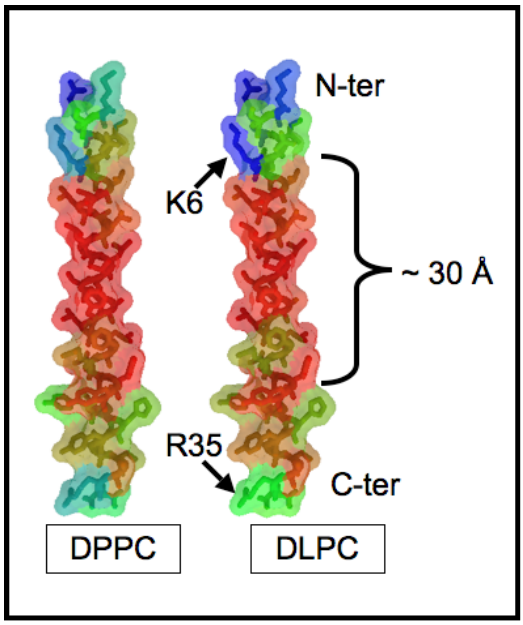


Figure 16. Surface representation of FK1TMD coloured according to number of contacts, defined as an interatomic distance of less than 6 Å, with lipid headgroup particles. The contacts have been summed for each residue over the last 100 ns of the CG simulations and normalised by the total number of contacts. This is represented pictorially by an RGB scale - red for no contacts, and blue the highest number of contacts.

In the other leaflet of the bilayer, at C-terminal end of the FK1TMD, the residues between Y28 and N36 made regular contacts with the lipid headgroup

particles. In particular, residues K29, R35, and N36 show a particular propensity to interact with the lipid headgroups, and, together, they make up 15% of the total headgroup-protein contacts in the DLPC simulations and 20% in the DPPC and POPC simulations. These observations suggest that the charged residues at both N- and C-terminal ends of the helix play a key role in anchoring the protein in a transmembrane orientation in the lipid bilayer.

In addition to the protein-lipid contacts, the lateral motion of the FK1TMD was also measured. The position of the centre of mass of the terminal residues was monitored in the *xy* plane during the last 100 ns of the simulations. In all three of the bilayer simulations, the lateral movement of the helix in the *xy* plane was up to 7 nm in either the *x* or *y* dimension. This lateral movement in the plane of bilayer, in addition to the fluctuations observed in the tilt angles of the helix, suggest that there may be some mismatch between the width of the hydrophobic patch on FK1TMD and the hydrophobic core of the bilayer defined by the lipid tails.

3.4.3 In Vivo Environment of FK1TMD

The GA membrane is composed of a complex mixture of lipids, which includes anionic and polyanionic lipid headgroups. In the previous sections, the influence of the lipid tail length on the orientation and mobility of the FK1TMD has been investigated in uniform zwitterionic PC lipid bilayers. Given the high positive charge at both termini of the FK1TMD, the affinity of the protein for other lipid headgroups was investigated, in particular anionic lipids. Firstly, the role of anionic lipids was investigated in a self-assembly simulation of FK1TMD, in a bilayer composed of only POPS lipids. Although the charge distribution of a 100% POPS bilayer does not reflect the *in vivo* environment of GA, these simulations were useful in determining the effect of anionic lipid headgroups on the dynamics of the FK1TMD, as direct comparisons could be made with the POPC simulations. The change of headgroup had a large effect on the protein tilt angle. Indeed, an increase was observed in the tilt angle from 17.6 ± 0.9˚ in POPC to 25.5 ± 2.2˚ in POPS. This is presumably driven by the increase in the number of contacts between the POPS headgroup atoms and the N-terminal R3 and K6 and C-terminal K29 residues. The lateral mobility of the protein in the two different lipid types was similar.

In addition, a more complex bilayer was simulated to investigate the effect of the different headgroups in an environment more representative of the GA membrane. This complex bilayer consisted of a mixture of DPPC, DPPE, PIP2, SM, and DPPS lipids in a 50:20:12:10:8 ratio and also cholesterol (CHOL) at a 16:1 phospholipid:CHOL ratio (*169*). In the complex GA membrane, the residues of the R/K-x-x-R/K motif, R3 and K6, at the N-terminus of FK1TMD made a large number of contacts with the headgroups of all lipid types. These two positively charged amino acids made the most contacts with the highly negatively charged PIP2. Indeed, the interactions of PIP2 with R3 and K6 accounted for 12% of the total contacts between FK1TMD and the phospholipid headgroups. At the C-terminus of the FK1TMD, in contrast, the interactions of K29 and R35 with PIP2 accounted for only 5% of the total headgroup contacts. Due to its position within the bilayer (Figure 17), the hydroxyl group of CHOL makes the most contacts with an isolated region between I4 and L10 at the N-terminus of FK1TMD. This region accounted for approximately 55% of the total contacts between CHOL and FK1TMD. Conversely, the C-terminal amino acids make fewer contacts with CHOL, and the interactions are spread evenly across a large number of residues. The interactions of SM with the FK1TMD were similar to those observed for DPPC. The tilt angle of FK1TMD was 23.6 ± 1.8˚ in the complex GA membrane. This value is between the tilt angles observed for the DPPC, DLPC and POPC bilayers, and is similar to the simulations of the FK1TMD in a pure POPS bilayer, where a tilt angle of 25.5˚ was measured.

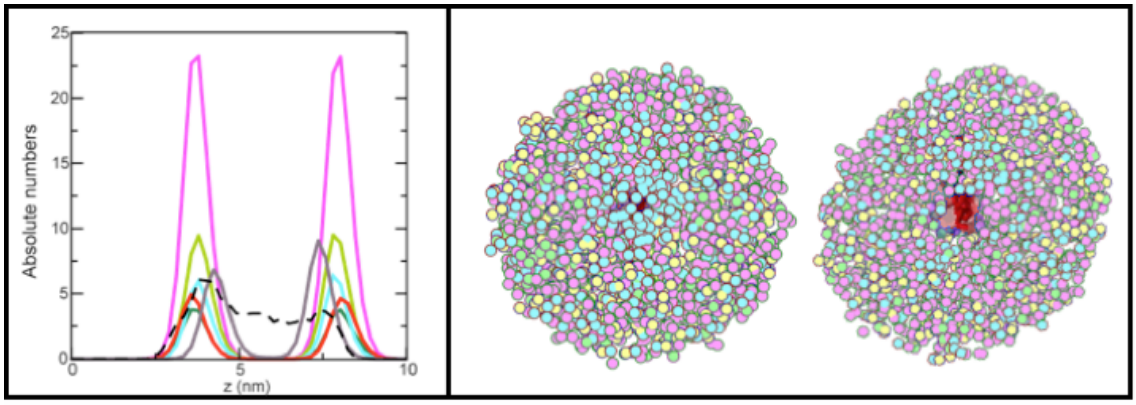


Figure 17. The panel on the left shows the number density of protein and lipid headgroup particles along the *z* dimension. The curves representing the different lipids are coloured as follows: DPPC is pink, DPPE is yellow, DPPS is green, PIP2 is cyan, SM is dark orange, CHOL is grey and the black dashed line represents FK1TMD. The panel on the right displays the lipid headgroup particles within 3 nm of the R and K residues in N- (left) and C- (right) termini (colours are as in the left panel, cholesterol and SM are omitted for clarity and the FK1TMD is red). The trajectory is fitted on the protein using a least squares procedure and the frames are superimposed at 40 ns intervals over the 1.5 μs trajectory. Clustering of the PIP2 lipids near the N-terminus is clear. Such clustering does not occur in the bilayer leaflet interacting with the C-terminus, which has a more random distribution of lipids (*173*).

The density profiles of the FK1TMD and the various lipid headgroups were calculated along an axis normal to the plane of the bilayer (Figure 17). The distribution indicates that DPPC, DPPE, and DPPS lipids occupy similar positions with respect to the bilayer normal. This is perhaps expected given that all the lipids contain the same acyl tails. The positional distributions for SM lipids are slightly further apart than for the other lipids, thus they are displaced slightly toward the solvent. This appears to be a result of lipid-lipid contacts within the bilayer. In contrast, the distributions for CHOL are closer together than those for the phospholipids, and so CHOL is located deeper into the hydrophobic core of the bilayer. The slight asymmetry in the CHOL density is a result of one molecule flipping into the opposite leaflet during the equilibration procedure.

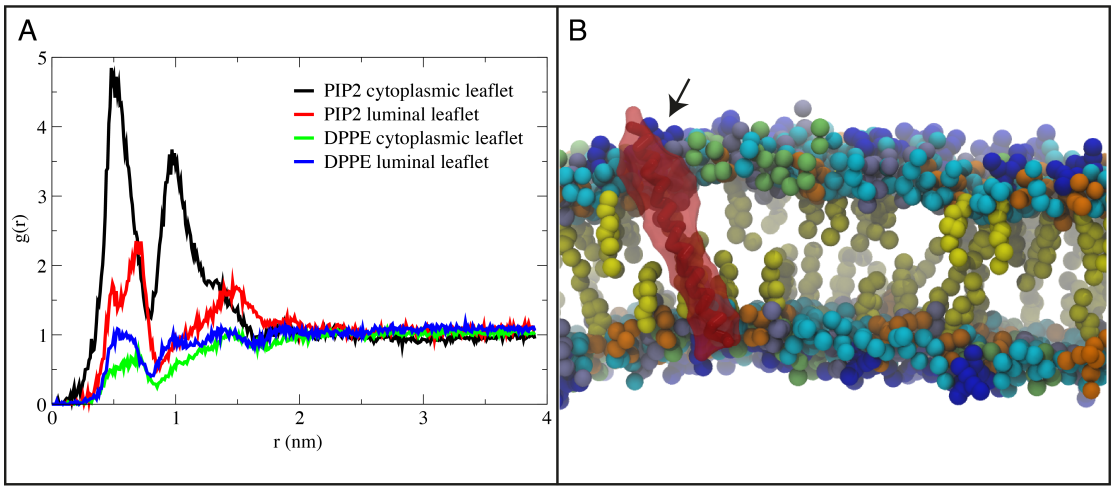


Figure 18. (A) Radial distribution of the glycerol linked phosphates in PIP2 and DPPE. The reference points for the radial distribution are K6 for the cytoplasmic leaflet and K29 for the luminal leaflet. (B) The GA model membrane. The lipid tails are omitted for clarity. PE headgroups are coloured cyan, PIP2 headgroups are coloured dark blue, PS headgroups are coloured green, SM is coloured mauve, and CHOL is coloured yellow. The interaction of K6 with PIP2 is indicated with an arrow.

A clustering of PIP2 lipids around the N-terminus of the FK1TMD was discovered by visual inspection of the simulation trajectories (Figure 17). This clustering was investigated further by calculating the radial distribution for the glycerol-linked phosphates of PIP2 relative to the K6 and K29 residues of FK1TMD (Figure 18). The distributions show that PIP2 is found more often in the proximity (within 1 nm) of K6 than in the proximity of K29. The density of the DPPE phosphates within 1 nm of FK1TMD was much lower than that of PIP2. Indeed, compared to PIP2, the radial distribution of the DPPE phosphates was similar in both leaflets of the bilayer, but fell below the bulk average density close (< 2 nm) to the FK1TMD, suggesting some exclusion at this distance.

3.5 Discussion

ATMD simulations have been used in this study to assess the structural stability of a α-helical model peptide, FK1TMD, representing the N-terminal transmembrane spanning region of fukutin. The simulations have demonstrated that FK1TMD forms a stable α-helix in DPPC, POPC and DMPC bilayers when in a transmembrane orientation, but it partially unfolds when lying parallel to the plane of the bilayer, at the headgroup-water interface. The differences in fatty acid tail length of the phospholipids did not affect the stability of the α-helix. Circular Dichroism measurements have also confirmed that the helical content of the peptide is constant irrespective of the bilayer thickness (*173*). The structural stability of the FK1TMD suggests that partitioning into lipid domains is unlikely to be accompanied by a conformational change in the transmembrane domain of fukutin.

The FK1TMD was demonstrated to preferentially adopt a transmembrane orientation in self-assembly CGMD simulations. To be accommodated in bilayers of different hydrophobic thickness, the FK1TMD transmembrane domain adjusts its tilt relative to the plane of the bilayer. Positively charged residues anchor both termini of FK1TMD in the headgroup region of the lipid bilayer. These residues are responsible for much of the variation in tilt angles that was observed in bilayers of differing compositions.

In addition to changing the headgroup composition, CGMD simulations of the FK1TMD were performed in three different lipid tails, and in a mixture of headgroups that closely mimics the *in vivo* Golgi membrane composition. The tilt angle of the FK1TMD was found to be sensitive to the lipid tail composition of surrounding lipid bilayer. Thus, it seems that the tilt angle is determined by an interplay between the fluidity and headgroup composition of the bilayer. Furthermore, the N-terminal residues of the FK1TMD exhibit a propensity to interact with PIP2 lipids, and this may play a role in the formation of microdomains by inducing clustering of these lipids. A possible limitation of this study is that only the transmembrane domain of the fukutin protein is considered. Similar studies of isolated transmembrane domains have previously been used to explore membrane protein interactions within bilayers (*77, 170*), though it is noted that the inclusion of neighbouring domains would provide further insights into the dynamics of the functional protein.

In summary, the transmembrane domain of fukutin exists as a stable α-helix in a bilayer environment. The changes observed in the tilt angle of the α-helix as a result of changes to the bilayer composition may suggest a method for the localisation within the cell of the full-length protein. The FK1TMD may be able to sense the bilayer compositions found in different intracellular compartments, which has implications for protein packing and the formation of higher oligomeric structures. While some aspects of the Golgi membrane, such as varied lipid tail composition, have not been explored in the present study, the simulations of the complex Golgi membrane represent a step towards linking *in silico* experiments with real biological systems.

Chapter 4: Stability and Membrane Interactions of the Hia Translocator Domain in a Complex Membrane Environment

4.1 Abstract

Hia is a trimeric autotransporter found in the outer membrane of *Haemphilus influenzae*. Each monomer of the Hia translocator consists of an α-helix connected via a loop to a 4-stranded β-sheet. Thus, the topology of the trimer is a 12-stranded β-barrel containing 3 α-helices that protrude from the mouth of the β-barrel into the extracellular medium. Molecular dynamics simulations of the Hia monomer and trimer have been performed to explore the interactions between the α-helices, β-barrel and connecting loops that may contribute to the stability of the trimer. In simulations of the Hia monomer, the central α-helix is shown to stabilise the fold of the 4-stranded β-sheet. In simulations of the Hia trimer, a hydrogen bond network has been identified involving residues in the β-barrel, α-helices and connecting loops. In this network, glutamine residues located in the connecting loops are orientated in a triangular arrangement such that each forms 2 hydrogen bonds with a glutamine in a neighbouring connecting loop. In the absence of the loops, the β-barrel becomes distorted. Thus, this network may provide stability for the trimeric arrangement of the monomers. The simulations also reveal that the trimeric translocator domain, although inherently flexible, is unlikely to accommodate the passenger domain in a folded conformation. In addition, simulations were also performed of Hia in an asymmetric bilayer representing the outer membrane of *Haemphilus influenzae*. These simulations showed interactions between Hia and lipopolysaccharide that may anchor the protein within outer membrane.

The study presented in this chapter has been published as part of a journal article in Biochimica et Biophysica Acta (*191*). A full copy of this article is included in Appendix 8.3. In contribution to the article, Dr Syma Khalid performed the simulations of the Hia trimer in a DMPC bilayer, under the

guidance of Professor Mark Sansom. These Hia simulations were later extended, and additional simulations were performed. In addition, Dr Thomas Piggot contributed to the article by creating the *H. influenzae* outer membrane model, including the parameters for lipopolysaccharide.

4.2 Introduction

Haemphilus influenzae is a Gram-negative bacterium and a human pathogen responsible for a wide range of clinical diseases including sinusitis, bronchitis and conjunctivitis (*174*). The initial step in infection involves adherence to epithelial cells and colonization of the upper respiratory tract. A number of adhesive factors facilitate the attachment of the bacterium to host epithelial cells. One such factor is the trimeric autotransporter called Hia. In order to mediate cell adherence, Hia must present its two protein binding domains on the surface of *H. influenzae* (*42, 175*). The cell-surface display of the binding domains requires that part of Hia be translocated across both the inner membrane (IM) and the outer membrane (OM). To achieve this, Hia utilises the autotransporter secretion pathway, also known as the type V secretion system (30, 42).

The autotransporters can be divided into two subfamilies (42), monomeric and trimeric autotransporters. However, both subfamilies share a common structure. Indeed, like the monomeric autotransporters, Hia has three characteristic structural features: i) an N-terminal signal sequence that directs the nascent polypeptide to the Sec machinery for translocation across the IM, ii) a passenger domain that contains the secreted cell surface binding domains, and iii) a β-barrel translocator domain that is inserted into the OM to assist in the translocation of the passenger domain (Figure 19A). A schematic representation of the domain organisation of the autotransporters is shown in Figure 19B.

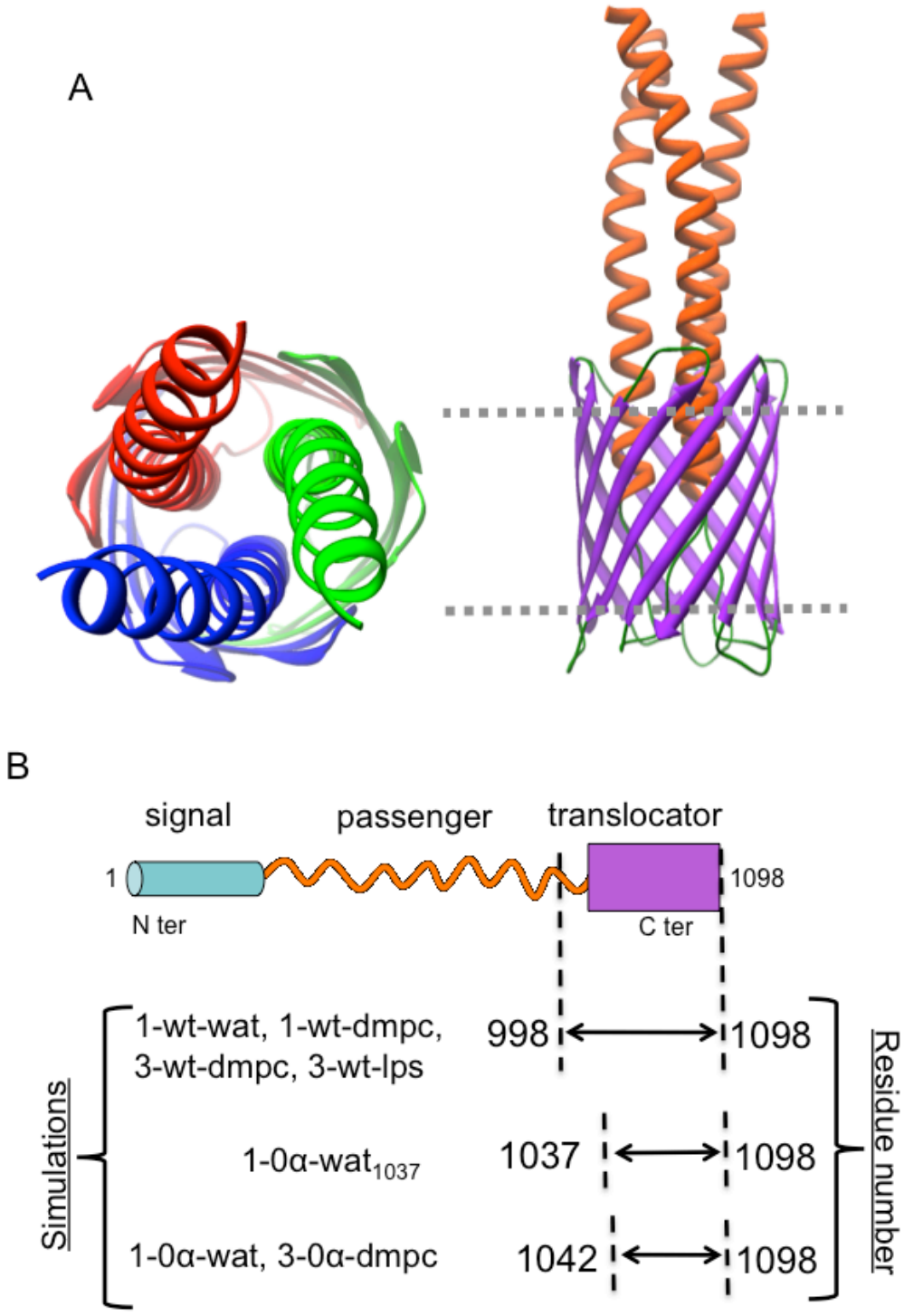


Figure 19. A) Two views of the X-ray structure of Hia. Looking down the principal axis from the extracellular end (left), here the three monomers are coloured individually (red, green and blue) and side view (right), here the β-barrel is purple, the loops are green and the α-helices are red. B) Schematic of domain organisation. The boundaries and residue numbers indicate the portion of the Hia structure that was used in each simulation of this study.

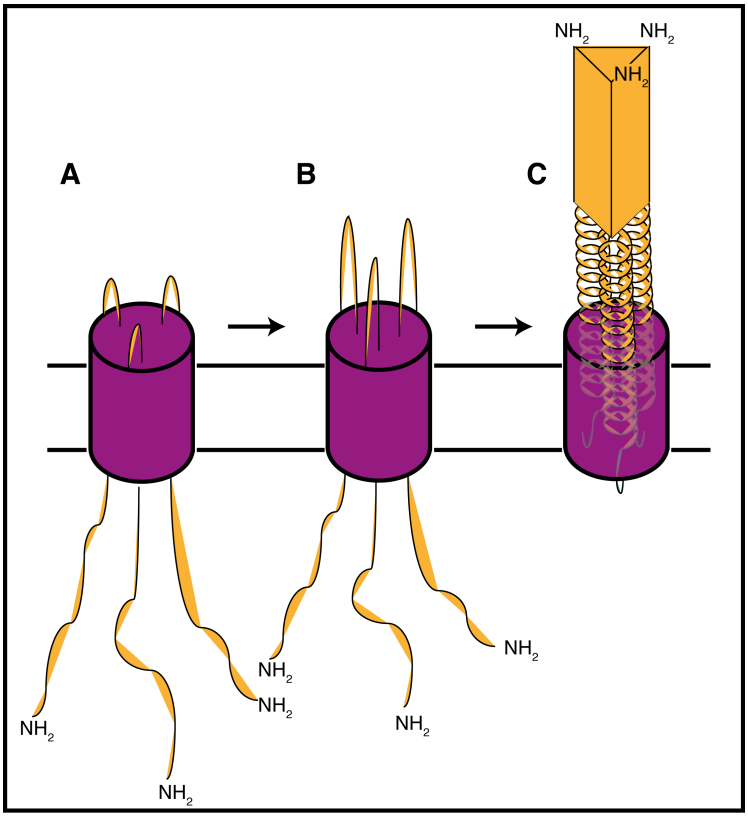


Figure 20. The hairpin model extended to the trimeric autotransporters. Translocation of the passenger (orange) proceeds in the C- to N- direction through the β-barrel translocator (purple). In A and B, six polypeptide chains are in the β-barrel pore simultaneously (these have been omitted from the figure for clarity).

A number of models have been proposed to explain the translocation of passenger domains across the OM. The trimeric autotransporters, like their monomeric cousins, presumably translocate their passengers in the C- to N-terminal direction (30), and the process has been shown to involve the Bam complex (*176*), which is also called the outer membrane protein (OMP) insertion complex. If the translocation of the passenger follows a similar pathway to the hairpin model (Figure 20) then there may need to be substantial and complex conformation changes in the translocator domain. The final stage of translocation is likely to be the full insertion and release of the folded Hia into the OM. In contrast to many of the monomeric autotransporters such as EspP and NalP (Figure 21), Hia does not release its passenger via autoproteolytic cleavage after transocation, and it remains cell

surface bound. The structure of the Hia translocator domain has been revealed by X-ray crystallography (*46, 177*). It forms a β-barrel with 12 transmembrane β-strands, with each subunit in the trimer contributing four β-strands. One α-helix from each of the subunits occupies the central channel of the β-barrel. The α-helices are connected to the first β-strand of each monomer by a long loop (α1-β1 loop), which extends two thirds of the way into the pore. The interface between the monomers is formed largely through symmetrical interactions between elements of α1-β1 loop. The interface between the α1-β1 loops is principally polar, and it is therefore hydrogen-bonding and electrostatic interactions that are likely to contribute to the stability of the trimer.

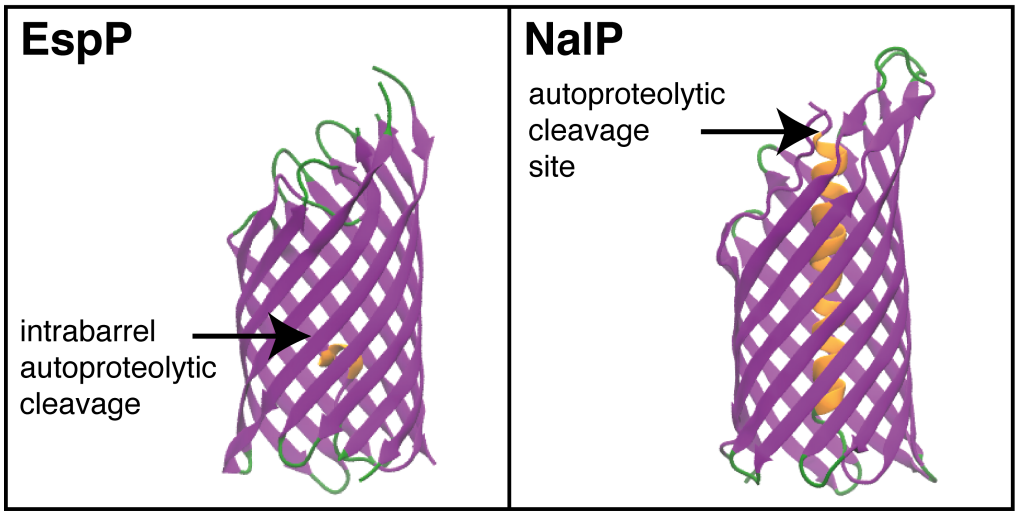


Figure 21. The translocator domains of the monomeric autotransporters, EspP and NalP. The site of autoproteolytic cleavage of the passenger domain is shown. The β-sheet residues are coloured purple, the pore α-helical residues are coloured orange, and loops between secondary structural elements are coloured green.

The Hia β-barrel has a 1.8-nm-diameter central pore that contains three α-helices, one from each monomer. The pore diameter, although slightly larger, is similar to that of the monomeric autotransporters, EspP (1.1 nm) (44) and NalP (1.2 nm) (49). Given the similarity in their pore dimensions, it might

be expected that the autotransporters share a common translocation mechanism. However, experimental and simulation studies have suggested a plug-like role for the α-helix in NalP, as well as promoting the correct fold of the β-barrel (113). Though, the α-helix is not implicated in maintaining the integrity of the β-barrel once it has formed (49). In contrast, experimental studies have shown that the α-helices located within the Hia β-barrel and in particular the loops that connect the α-helices to the β-barrel (the α1-β1 loop) are essential to maintain the structural integrity of the trimer (46).

The following is an atomistic molecular dynamics (ATMD) investigation into the structural stability and function of the Hia translocator domain. The α1-β1 loop is found to play a key role in maintaining the stability of the β-barrel. In the absence of the α1-β1 loops, the β-barrel becomes distorted. The simulations show that, while the trimeric translocator domain β-barrel is inherently flexible, it is unlikely to accommodate the passenger domain in a folded conformation.

The conformational dynamics of the Hia translocator domain is explored further using an OM model membrane containing a realistic mixture of lipopolysaccharide (LPS), also known as lipooligosaccharide, and phosphatidylethanolamine (PE) lipids. The analysis of these simulations reveals that a ring of positive residues located at the base of the α-helices, near the extracellular mouth of the β-barrel, interacts with the negatively charged moieties of the LPS inner core. These specific interactions in the OM may help to anchor Hia within its native membrane environment.

4.3 Methods

4.3.1 General Simulation Details

All simulations of the Hia translocator domain were performed using the GROMACS simulation package, version 4.5.1 (*158-160*). The SPC water model was used in all the simulations (*162*), unless stated otherwise. The simulations were undertaken in the NPT ensemble. The Nosé-Hoover thermostat, with a time constant of 0.5 ps, was used to maintain the system temperature at 310 Kelvin (*178, 179*). The Parrinello-Rahman barostat (*130, 180*) with a time constant of 5.0 ps, was used to maintain a system pressure of 1 bar. Long-range electrostatic interactions were calculated using the smooth particle mesh Ewald method (123). A long-range dispersion correction was applied to the energy and pressure to account for interactions beyond the cut-off used for the Lennard-Jones calculations. The neighbour list was updated every 5 steps during the simulations. The LINCS algorithm was used to constrain the lengths of all bonds (117), allowing a 2 fs time step to be used in all simulations. The steepest descent method was used for energy minimisations, and performed for 1000 steps in all cases. Simulations in which position restraints were used involved the application of a harmonic restraint force of 1000 $kJmol^{-1}nm^{-2}$ to non-hydrogen atoms. The Hia crystal structure used in the simulations was the 2.0 Å model described by Meng et al. (PDB code 2GR7) (*46*). The protonation states of all titratable residues of Hia were assigned using the standard protonation states at pH 7. These protonation states were subsequently checked with the H++ webserver (*181-183*). The repeats of all of the simulations used randomly generated initial velocities. A summary of the simulations described in this work is provided in Table 3 and Figure 19B.

4.3.2 Aqueous Simulations

The Hia structure from Meng et al. (PDB code 2GR7) (46) was solvated in a cubic periodic box so that the protein was at least 1.5 nm from the edge of the box. Water molecules were randomly replaced by Cl^- to neutralise the system. The resultant system was energy minimized. A 1 ns simulation was performed with position-restraints applied to the protein. This was to allow the water to equilibrate around the protein. The restraints were removed and a further

200 ns production simulation was performed. The protein was modelled using the GROMOS 43A1 force field (*184*). The short-range cut-off for both the van der Waals and electrostatic interactions was 1.0 nm.

4.3.3 Aqueous AMBER Simulations

The same initial system configuration to that of the aqueous GROMOS simulations was used in the AMBER simulations. However, the TIP3P water model was used (*185*) instead of SPC, and the protein was modelled using the AMBER 99SB-ILDN force field. The short-range cut-off for both the van der Waals and electrostatic interaction was 1.0 nm. The neighbour list cut-off was 1.05 nm. These simulation parameters match those used by Lindorff-Larsen et al., (*186*). The AMBER simulations were also undertaken in the NPT ensemble. The Nosé-Hoover thermostat (*178, 179*) with a time constant of 1.0 ps, was used to maintain a system temperature of 310 Kelvin, while the Parrinello-Rahman barostat (*130, 180*), with a time constant of 5.0 ps, was used to maintain the system pressure at 1 bar. Steric conflicts between the protein and solvent were relaxed by energy minimisation. Subsequently, a 1 ns position-restrained simulation was performed to allow an equilibration of the solvent around the protein. These restraints were removed and a 200 ns production simulation was performed.

4.3.4 Model Membrane Simulations

In the wild type simulations, the complete Hia X-ray structure was embedded in a 1, 2-dimyristoyl phosphatidylcholine (DMPC) lipid bilayer. In order to create the truncated variant of Hia, residues were deleted from the X-ray structure, prior to inserting the protein into a DMPC bilayer. This variant of Hia corresponds to one that was used by Meng et al. (*177*). The Hia model in each simulation was inserted into a DMPC membrane by positioning the aromatic belts of Hia at the water-membrane interface regions. Any overlapping lipids were removed. The force field parameters for DMPC are described in (*161*). The system was solvated by superimposition of a box of pre-equilibrated water. Any waters that were too close to either the protein or lipid were removed. Water molecules were randomly replaced by Cl^- ions to neutralise the system. More than 30 water molecules per lipid were added to the system, consisting of the protein in a bilayer of around 250 DMPC lipids. The simulation systems

were energy minimized to relax any steric conflicts between protein, lipid and solvent. Simulations of 1 ns in length with position-restraints applied were performed to equilibrate the membrane around the protein. These restraints were removed and a 100 ns production simulation was performed.

4.3.5 Outer Membrane Model Simulations

The simulations of the realistic *H. influenza* outer membrane used parameters derived from the GROMOS 53A6 force field (*187*). A snapshot of this complex membrane is shown in Figure 22A. The force field parameters for the *H. influenza* LPS were adapted from the *E. coli* LPS parameters (*188*). The LPS in the complex bilayer has a minimal level of glycosylation, and it represents the variant from the I-69 Rd-/b+ deep rough mutant strain. The LPS model is composed of a lipid A plus one phosphorylated 3-deoxy-D-manno-octulosonic acid (Kdo) sugar (Figure 22B) (*189*). The phospholipid inner leaflet contains a 1-myristol 2-palmitoleoyl phosphatidylethanolamine (MPoPE) lipid, which contains a tail composition similar to that found in the inner leaflet lipids of the OM (Figure 22C) (*190*). The force field parameters for MPoPE were taken from the GROMOS-CKP PE lipid parameters (*188*). The complex membrane was constructed from two individually equilibrated, symmetric LPS and MPoPE bilayers. The LPS bilayer was equilibrated for 500 ns, and the MPoPE bilayer was equilibrated for 100 ns. One leaflet of the symmetric LPS bilayer was combined with another from the symmetric MPoPE bilayer to create the complex *H. influenza* membrane. The area per chain of the LPS and MPoPE lipids were used to guide the construction of the complex membrane, so that the correct numbers of lipids were in each leaflet of the bilayer. The complex model was equilibrated for 100 ns (*191*). The wild type Hia protein was inserted into the complex membrane using the GROMACS program g_membed (*126*). Neutralizing Cl^- ions, in addition to the Mg^{2+} ions that are required to maintain the stability of LPS leaflet, were added to the system. Steric conflicts between protein, lipid and solvent were relaxed by energy minimisation. Subsequently, a 20 ns position-restrained simulation was performed to allow the bilayer to equilibrate around the protein. The restraints were removed and a 200 ns production simulation was performed. The short-range cut-off for the electrostatic interactions was 0.9 nm, with van der Waals interactions truncated at 1.4 nm.

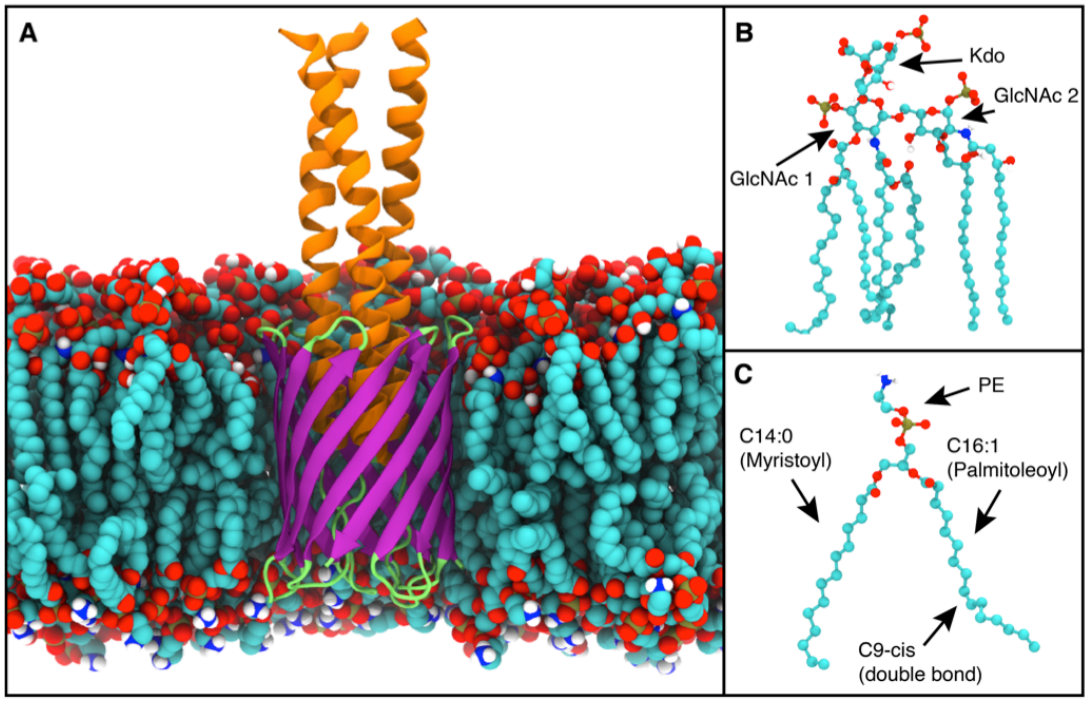


Figure 22. (A) Snapshot of Hia from simulation 3-wt-lps, the protein (β-barrel is purple, the loops are green and the α-helices are orange) is surrounded by a bilayer composed of (B) lipopolysaccharide (LPS) and (C) 1-myristol, 2-palmitoleoyl-phosphatidylethanolamine (MQPE), where Carbon is cyan, Oxygen is red, Nitrogen is blue, Hydrogen is white and Phosphorus is tan. Water and ions are omitted for clarity. In this picture the LPS is in the upper leaflet and the MQPE lipids are in the lower leaflet. Abbreviations: 3-deoxy-D-manno-Octulosonic Acid (Kdo) and Glucosamine sugar (GlcNAc).

4.3.6 General Analysis

All analyses were performed using GROMACS tools, MDAnalysis (*192*), and locally written code. Secondary structure analyses used DSSP (*164*). The pore within the β-barrel interior was analysed and visualised using HOLE (*193, 194*). Molecular graphics images were generated using VMD (*165*). In the H-bond analysis, a hydrogen bond was considered to have formed between a donor and acceptor where there was an interatomic distance of less than 0.35 nm, and an angle of less than 30˚ between the acceptor-donor-hydrogen.

Table 3. Summary of Hia simulations

Simulation	Hia model	Lipids	Solvent	length (ns)	Cα RMSD (nm)		
					All	β-barrel	α-helices
1-wt-wat	1 x $Hia_{998-1098}$	-	7 Cl^- 40,620 H_2O	3 x 200	1.05	0.27	1.14
1-0α-wat	1 x $Hia_{1042-1098}$	-	4 Cl^- 45,103 H_2O	3 x 200	0.65	0.45	-
1-0α-wat_{1037}	1 x $Hia_{1037-1098}$	-	4 Cl^- 40,620 H_2O	4 x 200	0.61	0.35	-
1-wt-dmpc	1 x $Hia_{998-1098}$	194 DMPC	7 Cl^- 14,294 H_2O	2 x 200	0.89	0.15	0.80
3-wt-dmpc	3 x $Hia_{998-1098}$	426 DMPC	21 Cl^- 49,778 H_2O	2 x 100	0.27	0.12	0.25
3-0α-dmpc	3 x $Hia_{1042-1098}$	426 DMPC	12 Cl^- 26,917 H_2O	2 x 100	0.34	0.26	-
3-wt-lps	3 x $Hia_{998-1098}$	59 LPS 161 MPoPE	219 Mg^{2+} 46 Cl^- 22,432 H_2O	2 x 200	0.30	0.11	0.24

4.4 Results

4.4.1 Monomer Stability

Trimerization of the Hia translocator domain is necessary for the efficient translocation of the N-terminal passenger across the OM (*195*). However, it is still not clear whether trimerisation of the Hia translocator domain occurs before or after membrane insertion. Therefore, the stability of a monomeric component of the Hia translocator domain was investigated in both aqueous and membrane environments. Any loss of secondary structure or disruption to the local membrane environment would indicate that the simulated systems were unlikely to represent real *in vivo* configurations. Thus, changes to the membrane thickness and the protein conformation were used as measures of the conformational stability of Hia in each environment. Three different systems were constructed that were composed of (i) a wildtype (wt) Hia monomer in water (1-wt-wat), (ii) a wt Hia monomer without any α-helices or α1-β1 loops in water (1-0α-wat), (iii) a wt Hia monomer that includes the α1-β1 loop but not the α-helices in water (1-0α-wat_{1037}), and (iv) a wt Hia monomer in a DMPC bilayer (1-wt-dmpc).

The Cα root mean-squared deviation (RMSD) provides a measure of the protein drift from the initial starting conformation derived from X-ray structure (Table 3). The overall plateau RMSD values were 1.05 nm in 1-wt-wat, 0.65 nm in 1-0α-wat and 0.89 nm in 1-wt-dmpc. These large RMSD values indicate the monomer is undergoing substantial structural changes. Much of the structural instability of the monomers was the result of bending and unwinding of the α-helices at the N-terminal end of the protein. However, the β-sheet residues were substantially more stable, and had plateau RMSD values of 0.27 nm in 1-wt-wat, 0.45 nm in 1-0α-wat and 0.15 nm in 1-wt-dmpc.

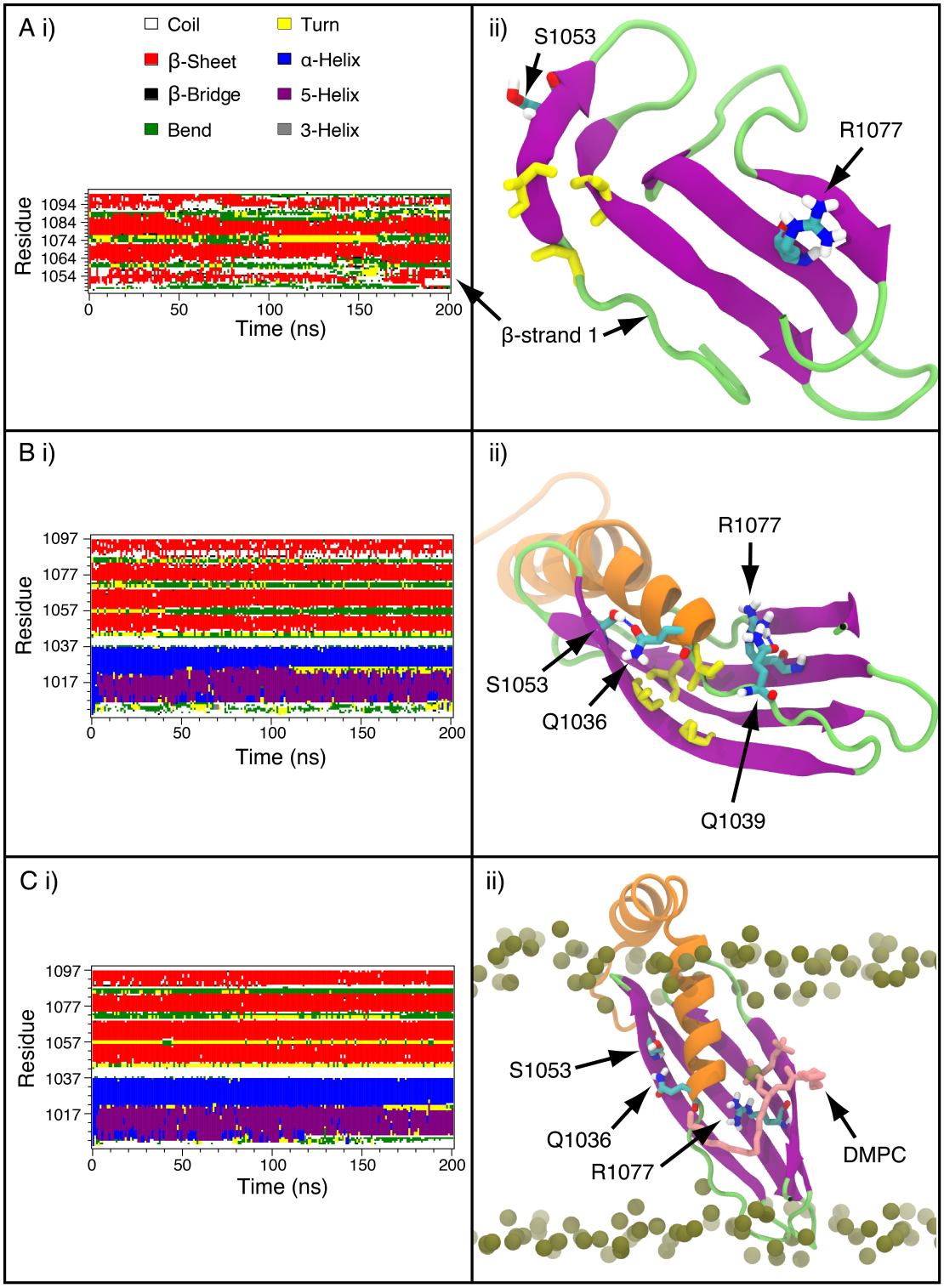


Figure 23. The DSSP plot (i) and final simulation snapshots (ii) of (A) 1-0α-wat, (B) 1-wt-wat and (C) 1-wt-dmpc. The β–sheet residues highlighted in yellow are A1049, A1051, and A1062. These residues form the boundaries of a pocket into which L1037, from the N-terminal end of the α1-β1 loop, is inserted.

The structural stability of both 1-wt-wat and 1-0α-wat was examined further using the DSSP secondary structure classification method. Interestingly, the β-sheet had a noticeably less stable secondary structure when the central α-helix and α1-β1 loops were removed (Figure 23A). Indeed, the average number of residues in a β-sheet conformation during the simulations was calculated to be 34 in the 1-wt-wat simulations and 26 in the 1-0α-wat simulations. These differences appeared mostly to be the result of the partial unfolding of β-strand 1 in the 1-0α-wat monomer. Such subtle changes in the stability of secondary structure can sometimes be the result of force field inaccuracies. Therefore, the 1-0α-wat and 1-wt-wat systems were simulated again using the AMBER99SB-ILDN all atom force field (*186*) in order to verify that the differences in stability of the β-sheet were force field independent. Encouragingly, these simulations showed similar trends to the GROMOS force field simulations (see Figure 24).

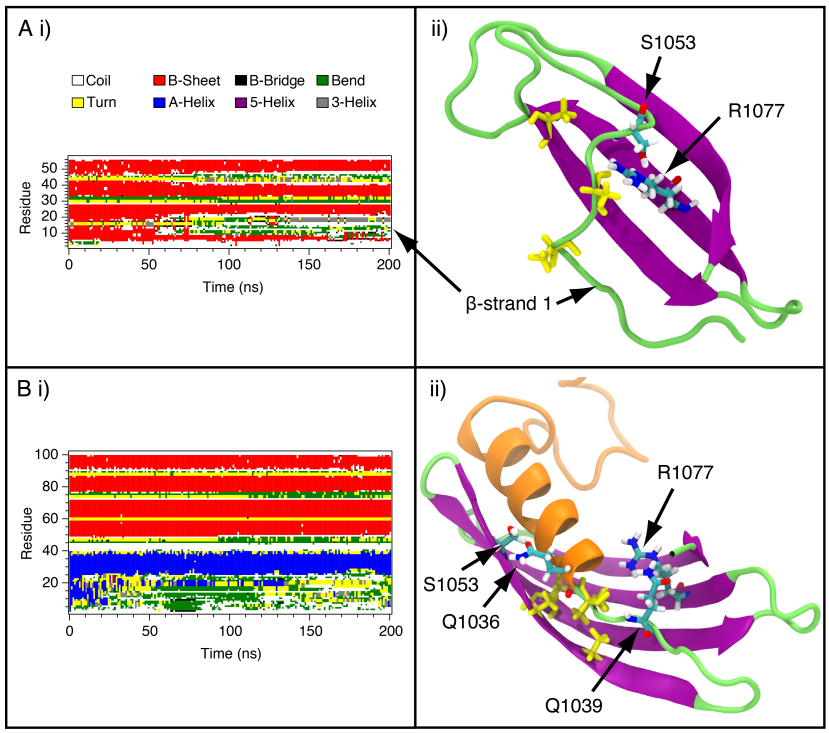


Figure 24. The DSSP plot (i) and final simulation snapshots (ii) of (A) 1-0α-wat and (B) 1-wt-wat. These simulations were performed using the AMBER 99SB-ILDN force-field. The β-sheet residues highlighted in yellow are A1049, A1051, and A1062. These residues form the boundaries of a pocket into which L1037, from the N-terminal end of the α1β1 loop, is inserted.

The greater stability of β-strand 1 in 1-wt-wat did not appear to result from polar interactions with the α-helix. Indeed, there were relatively few specific interactions between β-strand 1 and the α-helix, except for one hydrogen bond (H-bond), which was present in 40% of the trajectory frames, between residue S1053 of β-strand 1 and residue Q1036 from the α-helix. The low number of H-bonds is surprising, given that there were approximately 20 barrel-helix H-bonds reported in simulations of NalP (112). However, a greater number of specific interactions occurred between the α1-β1 loop and β-strands 2 and 3. In particular, H-bonds were formed between residue Q1039 in the α1-β1 loop and residue R1077 in β-strand 3. This H-bond was present in 45% of the trajectory frames. It is interesting to note that neither the H-bond between Q1036 and S1053 nor that between Q1039 and R1077 are present in the X-ray derived structure, where the residues are oriented in such a way that they do not appear to interact.

It was observed from visual inspection of the Hia X-ray structure that residue L1037, at the N-terminal end of the α1-β1 loop, fits into a small pocket bordered by A1049, A1051, and A1062 from β-strands 1 and 2 (Figure 24). It seems likely that the favourable hydrophobic packing in this region may help to stabilise the α1-β1 loop in its position on the surface of β-strands 1 and 2. In order to investigate the packing of L1037 further, a series of 200 ns simulations were performed of a truncated Hia monomer containing residues from L1037 to W1098 (1-0α-wat$_{1037}$). The Hia structure in 1-0α-wat$_{1037}$ contains the α1-β1 loop but not the α-helix. This Hia variant was less stable than the intact monomer, with an RMSD of 0.35 nm for the β-sheet residues. In contrast to 1-0α-wat, however, the β-strand 1 of 1-0α-wat$_{1037}$ retained most of its the β-sheet conformation (Figure 25A and Figure 25B), as assessed by DSSP, for the full 200 ns. It must be noted that, in a few of the repeat simulations, the residues in β-strand 1 showed signs of conformational instability, though the β-sheet conformation did not completely disappear within 200 ns, as it did in 1-0α-wat.

It seems that L1037, by itself, is unable to maintain the α1-β1 loop in its usual position on the surface of β-strands 1 and 2. Indeed, L1037 became detached from its hydrophobic pocket after about 80 ns of simulation. The point of separation can be observed in Figure 25C, which plots, as a function of time, the distance between the COM of L1037 and the COM of the hydrophobic

pocket formed by residues, A1049, A1051, and A1062. Indeed, the distance remained stable at 0.5 nm for the first 80 ns of the simulations, but increased markedly, becoming greater than 2.0 nm, for the remaining 120 ns of the simulations.

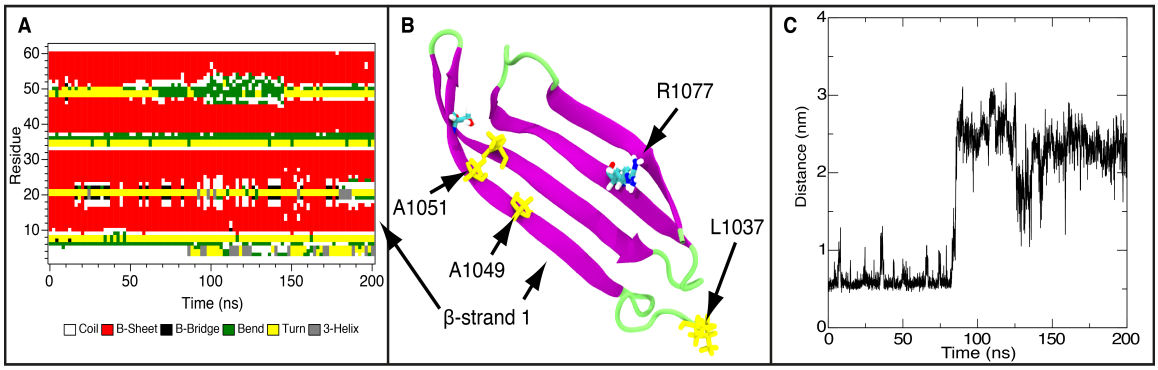


Figure 25. Simulation 1-0α-wat$_{1037}$: (A) A DSSP structure analysis of Hia (B) final simulation snapshot, and (C) the distance of L1037 from the hydrophobic pocket formed by A1049, A1051, and A1062 (yellow).

Since the Hia translocator exists as an integral OM protein *in vivo*, it is of interest to examine the structural stability of a single monomer within a lipid environment. Thus, a single Hia monomer was embed within a DMPC bilayer in a transmembrane orientation (1-wt-dmpc). In this simulation, the secondary structure was mostly conserved throughout the 200 ns (Figure 23C), and less secondary structure variability was observed compared to 1-wt-wat and 1-0α-wat. However, the hydrophilic face, which would normally line the interior of the β-barrel in the trimeric arrangement, caused a substantial distortion of the lipid bilayer. The distortions were the result of the negatively charged phosphate group from the DMPC lipids interacting with the positively charged R1077 in the centre of β-strand 3 (Figure 23C). In addition, the hydrophilic face was further accommodated within the hydrophobic core of the bilayer by the formation of a water pore that spanned both leaflets of the membrane and passed through the central cavity of the Hia monomer. These results suggest that the insertion of a Hia monomer into the OM would need to be accompanied by a substantial shielding of its central hydrophilic face.

4.4.2 Barrel Flexibility and Stability of the Trimer

The flexibility and stability of the full trimeric β-barrel from Hia was investigated in two simulations, each 100 ns in length. The two simulation systems contained (i) a wt Hia trimer within a DMPC bilayer (3-wt-dmpc) and (ii) a wt Hia trimer without any of the α-helices or α1-β1 loops within a DMPC bilayer (3-0α-dmpc). An illustrative snapshot of the Hia trimer is shown in Figure 19.

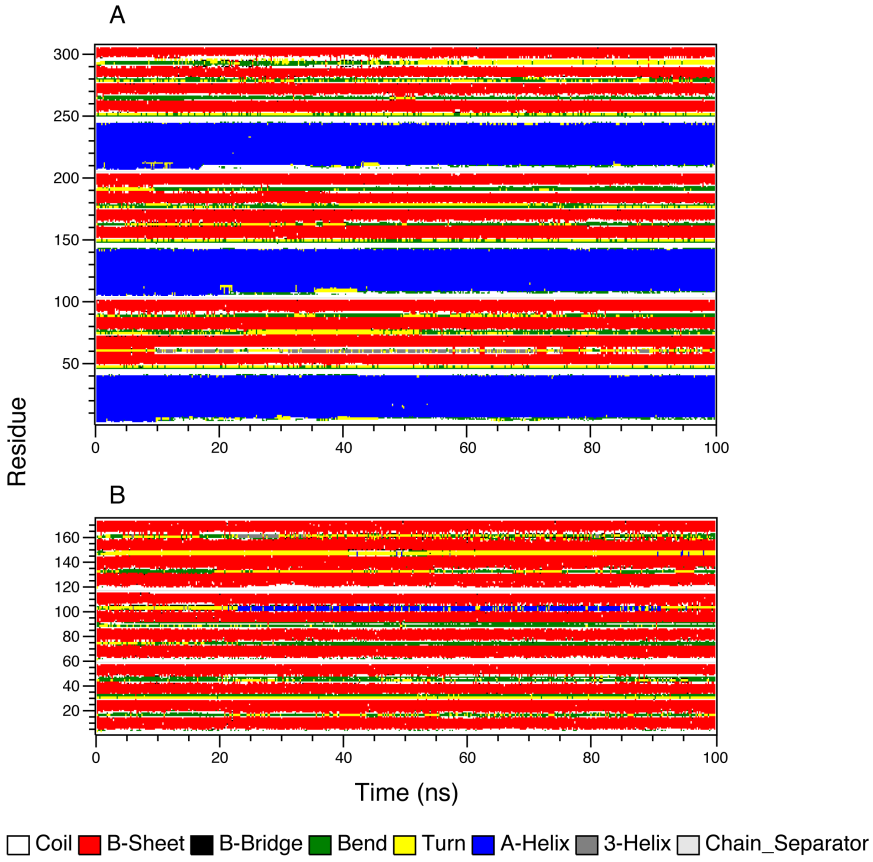


Figure 26. A DSSP secondary structure analysis of Hia in (A) 3-wt-dmpc and (B) 3-0α-dmpc. The secondary structure of the β-barrel remains intact for the full 100 ns in both simulations. In 3-wt-dmpc, all three helices unwind slightly at the N-terminus.

In contrast to the monomeric (1-wt-wat and 1-wt-dmpc) simulations, the α-helices were stable in simulations of the intact, trimeric configuration of Hia (3-wt-dmpc), although some unwinding of the α-helices was observed in the N-terminal residues. Indeed, the unwinding of the α-helices was confined to the regions exposed to the extracellular medium, outside the pore of the β-barrel (Figure 26A). The combined RMSD value for the three α-helices was 0.25 nm (Figure 27) for 3-wt-dmpc. A similar level of structural drift was reported for the single α-helix of NalP, which had a plateau RMSD of 0.20 nm (112).

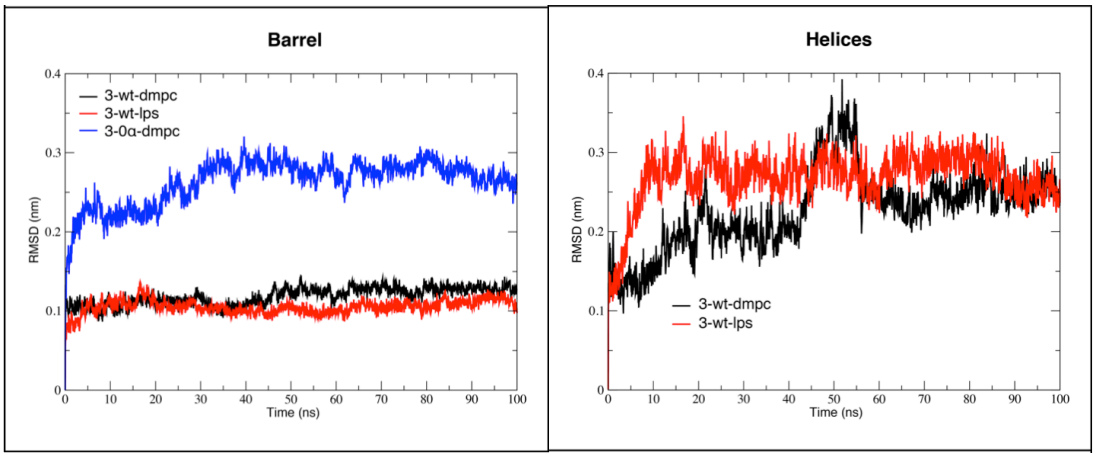


Figure 27. The root mean squared deviation (RMSD) of Cα atoms in the α−helical and the β−barrel portions of Hia.

A marked difference in the RMSD of the β-barrels was observed between 3-wt-dmpc and 3-0α-dmpc, with values of 0.12 nm to 0.25 nm, respectively (Table 3). The difference in RMSD values resulted from a large change in the conformation of the β-barrel during 3-0α-dmpc (Figure 28C). Indeed, at the end of 3-0α-dmpc, the β-barrel resembled the triangular prism conformation of the 12-stranded outer membrane protein, OprM (*196*). In contrast, the overall cylindrical shape of the β-barrel was maintained in 3-wt-dmpc throughout the 100 ns simulations (Figure 28B).

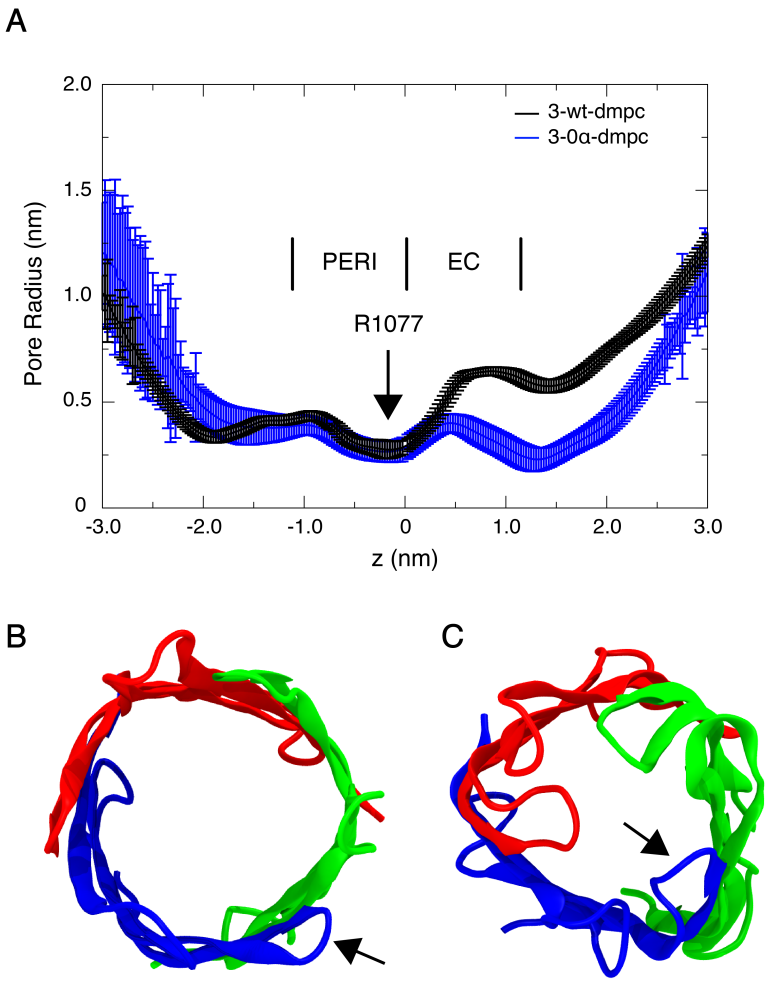


Figure 28. A – Pore radius profiles for 3-wt-dmpc and 3-0α-dmpc simulations. The location of the β-barrel residue R1077 that participates in the H-bond network is marked. The positions of the periplasmic (PERI) and extracellular (EC) leaflets are also indicated. B and C – A top view (from extracellular side) of the protein conformation at the end of (B) 3-wt-dmpc and (C) 3-0α-dmpc. An arrow marks the position of G1057 on one of the subunits. A distinct narrowing of the extracellular mouth of the β-barrel can be observed in 3-0α-dmpc. The α-helices and α1-β1 loops have been removed for clarity.

It is interesting to note that the plateau RMSD value of the β-barrel in 3-0α-dmpc is only slightly higher than in 3-wt-dmpc, even though there is substantial distortion at the extracellular mouth of the β-barrel in 3-0α-dmpc. Furthermore, the pattern of H-bonds between the backbone elements of the β-barrel was mostly unaffected in 3-0α-dmpc when compared to the 3-wt-dmpc. Thus, the removal of the α-helices and loops does not lead to loss of the secondary structure in the β-barrel on the timescale of the simulations (Figure 26B); rather, the deletion of the α-helices and α1-β1 loops causes substantial changes in the conformation of the β-barrel.

All-to-all RMSD comparisons were calculated in order to examine in more detail the conformational rearrangement in 3-0α-dmpc (Figure 29). Each monomer was compared with every other monomer across all the frames in a trajectory. In 3-0α-dmpc the monomers were observed to diverge rapidly in structure from one another, reaching a maximum RMSD of 0.40 nm, almost immediately after the start of the simulations. In contrast, the monomers of 3-wt-dmpc remained in a similar conformation, with a maximum RMSD of 0.25 nm, across the length of the trajectory. This demonstrates that there is a clear loss of symmetry in 3-0α-dmpc, which may indicate an increased flexibility in the β-barrel. Thus, the β-barrel is not a rigid cylinder, but, instead, may have the ability to adapt its conformation in response to the number of α-helices and α1-β1 loops contained within the pore region. Indeed, a similar opening and closing motion has been reported previously for the 12-stranded β-barrel of the type I secretion protein, TolC (*197*).

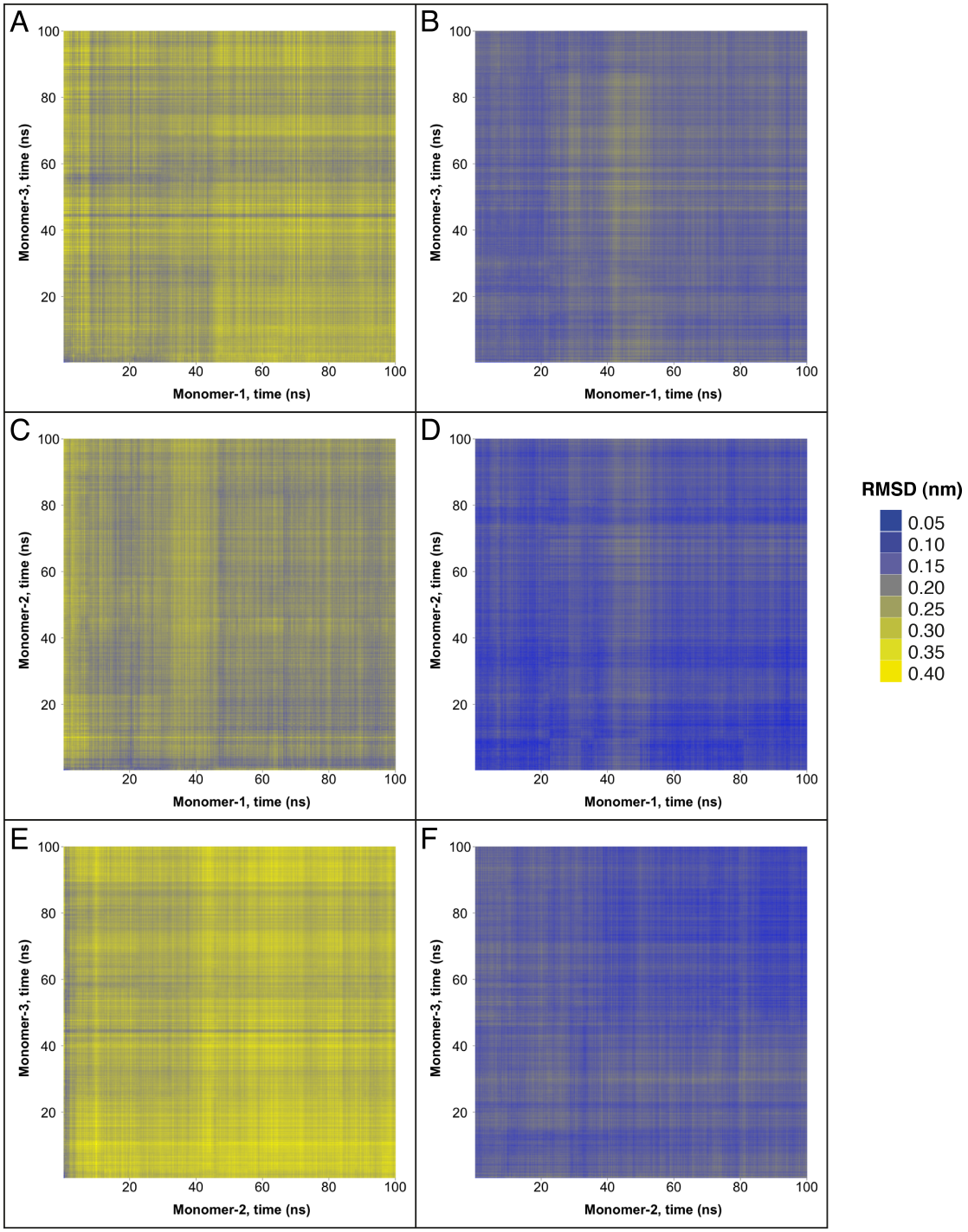


Figure 29 All-to-all RMSD plots comparing the β-barrel Cα atoms of each monomer with every other monomer for 3-0α-dmpc (A,C and E) and 3-wt-dmpc (B,D and F) simulations. The all-to-all RMSD plots compare all the frames from one trajectory with all the frames in another. A colour scale is used to indicate the value of the RMSD.

The extent of the conformational re-arrangement in the Hia translocator domain in 3-wt-dmpc and in 3-0α-dmpc can be quantified by monitoring the dimensions of the pore within the β-barrel. The HOLE code was used to calculate the average pore radii in all simulations (*198*). In the 3-wt-dmpc simulations, the α-helices and α1-β1 loops were removed prior to calculation of the pore dimensions. In general the pore was narrower in 3-0α-dmpc than in 3-wt-dmpc, particularly at the extracellular mouth of the β-barrel. A constriction point is observed in both simulations around R1077 (Figure 28A). The average width of the pore in this region was similar in both 3-wt-dmpc and 3-0α-dmpc, indicating the R1077 residues remain pointing into the β-barrel even in the absence of the α-helices and α1-β1 loops. The pore is narrower in 3-wt-dmpc than 3-0α-dmpc at the periplasmic mouth β-barrel. This is a result of the α1-β1 loops being forced into a tight arrangement by the bulk of the three α-helices.

The narrowest point within the pore of 3-0α-dmpc is at the extracellular mouth of the β-barrel. In this region, the β-barrel adopts a distinctly triangular conformation in 3-0α-dmpc (Figure 28), and the extracellular loops partially fold into the pore and form H-bonds to each other. This leads to the overall narrowing of the extracellular mouth of the pore observed in the pore radius profile of 3-0α-dmpc. Clearly, this H-bonding between the loops at the extracellular end of the β-barrel is not possible in the presence of the α-helices. A similar folding of the loops into the β-barrel was reported for the monomeric autotransporter NalP upon removal of its central α-helix (112).

The conformations of the β-barrel at extracellular mouth can be studied further by measuring the inter-monomer distances, across the mouth of the β-barrel, of the Cα particles from G1057. In 3-wt-dmpc, these distances were ~23.6 Å after equilibration. During the simulation, these distances decreased 8 % on average, indicating a slight narrowing. In contrast, these distances decreased by 42 % in 3-0α-dmpc.

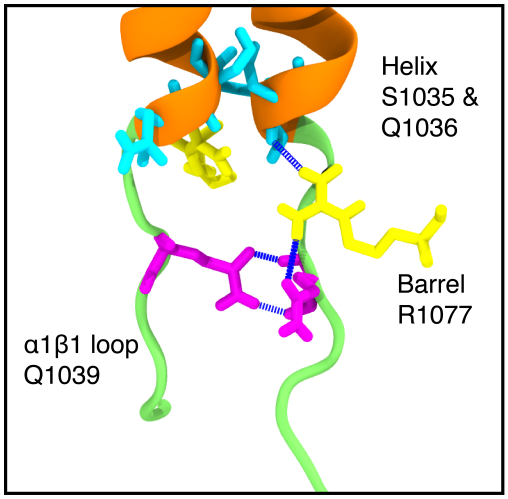


Figure 30. A snapshot from simulation 3-wt-dmpc showing the H-bond network proposed to be key in maintaining the structural integrity of the trimer. Helix, S1035 and Q1036, residues are cyan, β-barrel, R1077, residues are yellow and Q1039 residues in the α1-β1 loops are magenta. One monomer is completely omitted for clarity. H-bonds are shown in blue.

A search for H-bonds within 3-wt-dmpc revealed an extended network to exist between residues from the α1-β1 loops, the β-barrel and the α-helices (Figure 30 and Figure 31). The most notable aspect of this network is the triangular arrangement of Q1039 residues, allowing them to form 2 H-bonds with a corresponding Q1039 residue on a neighbouring loop. These residues also form H-bonds to R1077 residues located in the β-barrel, which in turn form H-bonds to S1035 and Q1036 residues in the α-helix. The side chains of R1077 residues also adopt a distinctly triangular arrangement near the ends of the loops, connecting the base of the α-helices to the Q1039 residues.

The H-bond network observed within 3-wt-dmpc does not exist in the X-ray derived structure (PDB ID: 2GR7), though R1077 is within 0.35 nm of S1035, Q1036, and Q1039. However, there were 9.0 H-bonds present on average between these residues in each frame of 3-wt-dmpc. In addition, the residues were able to interact with water molecules, which tended to cross-link the H-bonding groups (Figure 31, A and B). The water molecules were able to exist in several pockets within the β-barrel (Figure 31C), though the highest density of water was found in a pocket at the base of the α-helices between the α1-β1 loops, where the H-bond network exists between the monomers.

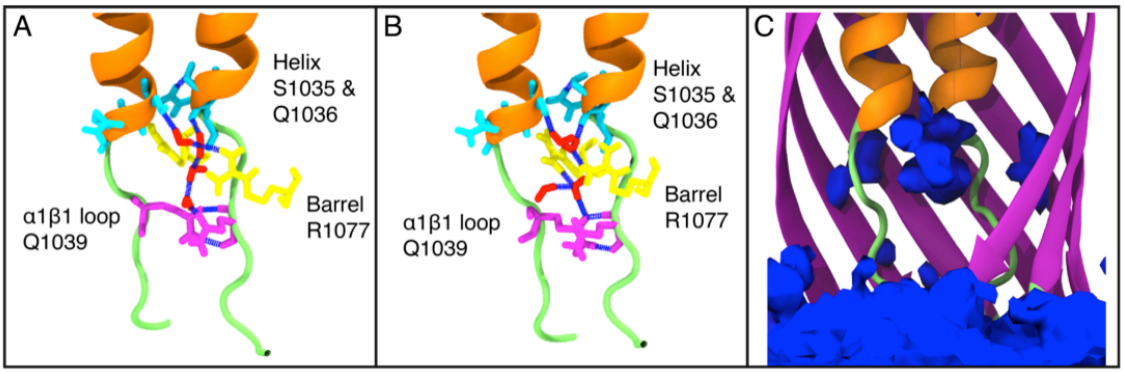


Figure 31. A) and B) Snapshots of 3-wt-dmpc showing the H-bond network with cross-linking water interactions. Helix, S1035 and Q1036 residues are cyan, β−barrel R1077 residues are yellow, Q1039 residues in the α1-β1 loops are magenta, and water molecules are red. One monomer is completely omitted for clarity. H-bonds are shown in blue. C) The water density around the β−barrel has been averaged over the final 100 ns trajectory of 3-wt-dmpc. The water density is represented as a blue isosurface. One subunit is completely omitted for clarity.

4.4.3 Specific Protein-Lipid Interactions

The dynamics of Hia are likely to be influenced by the unique environment created by the OM lipids. Therefore, an asymmetric bilayer was constructed with components more closely resembling those found *H. influenzae*. This bilayer contained a deep-rough variant of LPS in the outer leaflet and 1-myristoyl 2-palmitoleoyl phosphatidylethanolamine (MPoPE) in the inner leaflet (Figure 22). The simulations of Hia embedded within this membrane are referred to as 3-wt-lps from here on.

The structural drift of Hia in 3-wt-lps from the initial conformation was similar to that observed in the DMPC bilayer simulations. Indeed, the backbone RMSD values for the α-helices reached plateau values of 0.24 nm in 3-wt-dmpc and 0.18 nm in 3-wt-lps after 100 ns, while values for the β-barrel reached a plateau of 0.12 nm in both environments.

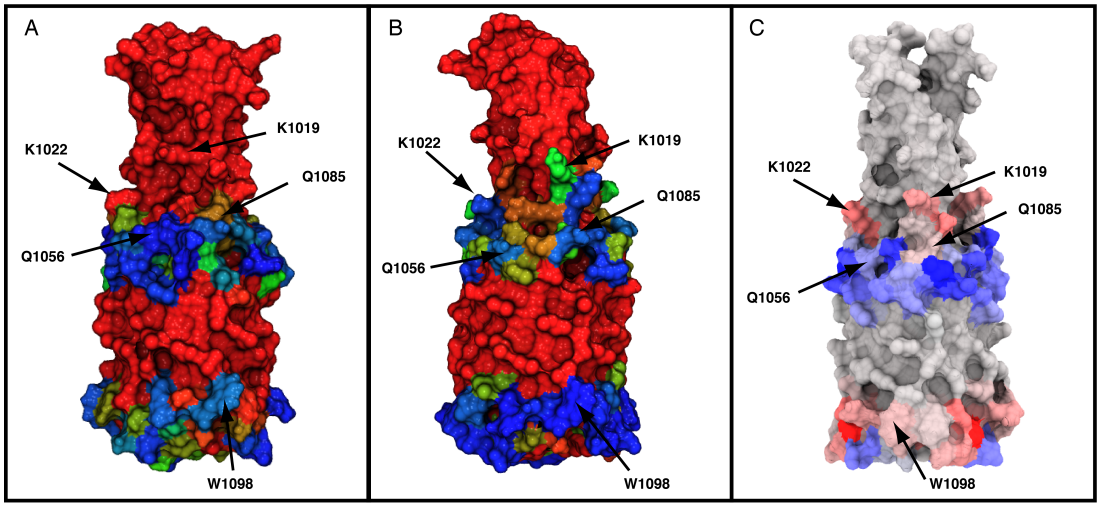


Figure 32. The surface of the protein is coloured according to number of contacts (where contact is defined as an inter-atomic distance of ≤ 0.35 nm) between (A) residues of Hia and DMPC headgroup atoms, and between (B) residues of Hia and the polar groups of LPS. The numbers of contacts are represented on a RGB scale, where red indicates 0 contacts with the lipid groups and dark blue indicates that the residue makes contacts in 100% of the frames analysed. The difference in the number of headgroup contacts (C) is shown on a red-grey-blue scale where dark red means more contacts with the polar regions of LPS, dark blue means more contacts with the DMPC headgroups, and grey means that there were equal numbers of contacts in both sets of simulation.

The protein-lipid interactions, defined as an interatomic distance of ≤ 0.35nm (109), produced the most noticeable differences between the two membrane environments. In general, Hia made fewer contacts with the sugar and phosphate atoms of LPS in 3-wt-lps than it did with the headgroup atoms of DMPC in 3-wt-dmpc (Figure 32). In addition, the sugar and phosphate moieties of LPS predominantly made contact with the larger residues, Q1056, Q1058, K1083 and Q1085, in the extracellular loops of Hia. Together, these observations demonstrate that the LPS molecules were not able to approach Hia over the 200 ns timescale of the simulations, hence the LPS could not gain access to interact with the side chains of the smaller amino acids. Presumably, the pattern of protein-LPS interactions are a result of both the slow diffusion of

LPS compared to DMPC and the steric restrictions imposed by the larger amino acids.

The residues of Hia exhibit broadly similar headgroup contacts in the inner leaflet of 3-wt-lps to those observed in 3-wt-dmpc (Figure 32). Indeed, the interaction of the large aromatic residue, W1098, with the phospholipid headgroup atoms was observed in greater than 80% of the simulations frames in both 3-wt-lps and 3-wt-dmpc. However, Hia made more contacts in general with the PE headgroups in the inner leaflet of 3-wt-lps than with PC headgroups in the corresponding leaflet of 3-wt-dmpc. While the partial charges of the PE and PC headgroups are identical in the force fields used for 3-wt-lps and 3-wt-dmpc, a smaller van der Waals radius allows the NH_3 group of PE to approach the protein side chains more closely to form tighter electrostatic interactions.

In visual observations of the Hia structure, it has been noted that a ring of basic residues (K1019, K1022, R1023 and K1087) is positioned close to the extracellular mouth of the β-barrel (*177*). It is likely that this positive patch may help to anchor Hia within the outer membrane via interactions with the negatively charged LPS inner core. The simulations of 3-wt-lps show that K1019 and K1022, in particular, make regular interactions with the carboxylic acid, phosphate and C5 hydroxyl groups of the Kdo sugar of LPS (Figure 33). The H-bond and electrostatic interactions were observed to occur frequently. Indeed, in one of the 3-wt-lps trajectories, K1022 interacted with the LPS inner core in greater than 90% of the frames. The electrostatic interactions between the lysine residues on the α-helices with the negative moieties of the Kdo sugar (carboxylic acid and phosphate) had occupancies in the range of 6% to 18% of the simulation trajectory. H-bonds between the lysine residues and the Kdo sugar had occupancies in the range of 1% to 39% of the simulation trajectory. In addition, there was no obvious screening of electrostatic and H-bond interactions as a result of the Mg^{2+} ions, which typically cross-link the negatively charged moieties of LPS.

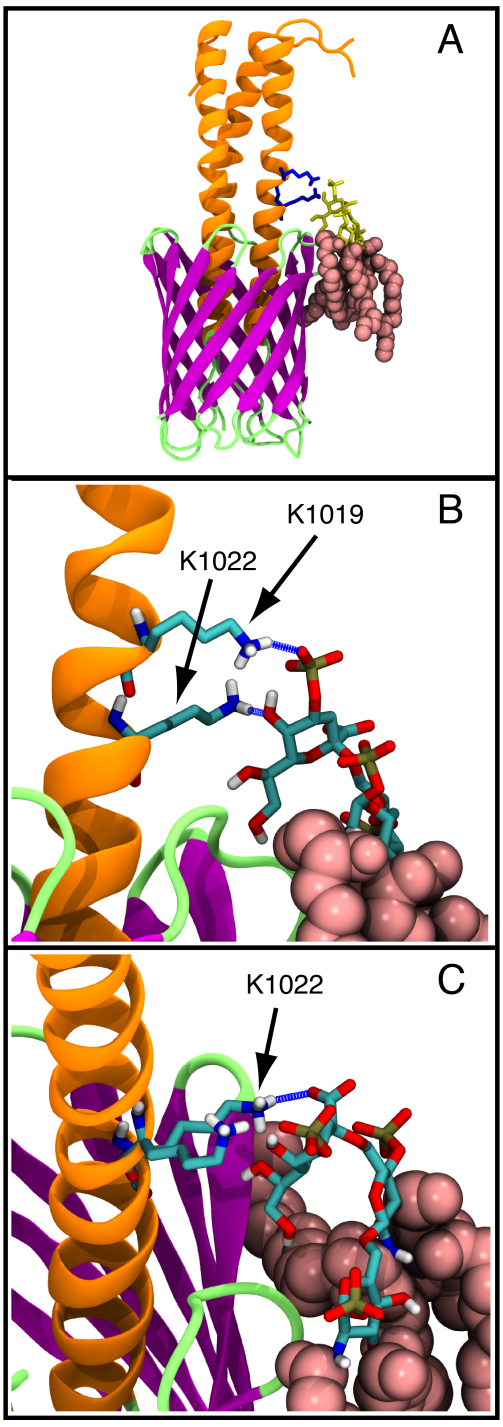


Figure 33. Interaction of LPS with Hia. (A) The relative position of Hia within the sugar and phosphate groups of LPS. K1019 and K1022 are able to form electrostatic and H-bond interactions with the (B) C4 phosphate, C5 hydroxyl and the (C) C1 carboxylic acid of LPS. Electrostatic and H-bond interactions are displayed as a blue dotted line in B and C. The LPS molecule is shown with the acyl tails in pink in a space filling representation, while the Kdo and glucosamine sugars are shown in a licorice representation.

The LPS molecules diffuse many times more slowly than phospholipids (*188*). This could mean that the 200 ns timescale of the 3-wt-lps simulations is not sufficient to correct the membrane position of Hia, if its initial position were not at the appropriate height within the bilayer. Therefore, a further two simulations of 3-wt-lps were performed with Hia placed 0.5 nm higher within the bilayer, so that the correct starting position within the model membranes could be determined. The new position of Hia within the membrane exposed the β-barrel to more of the extracellular side of the bilayer. In these extra simulations, Hia was observed to move by 0.2 nm towards the inner phospholipid leaflet after 200 ns of simulation (Figure 34A). In addition, the new position of Hia resulted in a thinning of the inner leaflet, of approximately 0.3 nm, around the protein (Figure 34B). In contrast, no distortion of the inner leaflet or change in position of Hia was observed in the original 3-wt-lps simulations. This suggests that Hia was positioned at the appropriate depth within the membrane in the original 3-wt-lps simulations.

In the hairpin model of the autotransporter mechanism, it is suggested that passenger domain to passed through the pore of the β-barrel. Therefore, it might be expected that the α-helices would be free to move up and down within the β-barrel pore. The movement of the α-helices relative to the β-barrel was investigated by measuring the distance between the COM of the β-sheet residues and the COM of Cα atoms from S1034 of each monomer. While the distance oscillated between 0.2 nm and 0.3 nm, no net movement of S1034 in the z-coordinate was observed in either 3-wt-dmpc or 3-wt-lps.

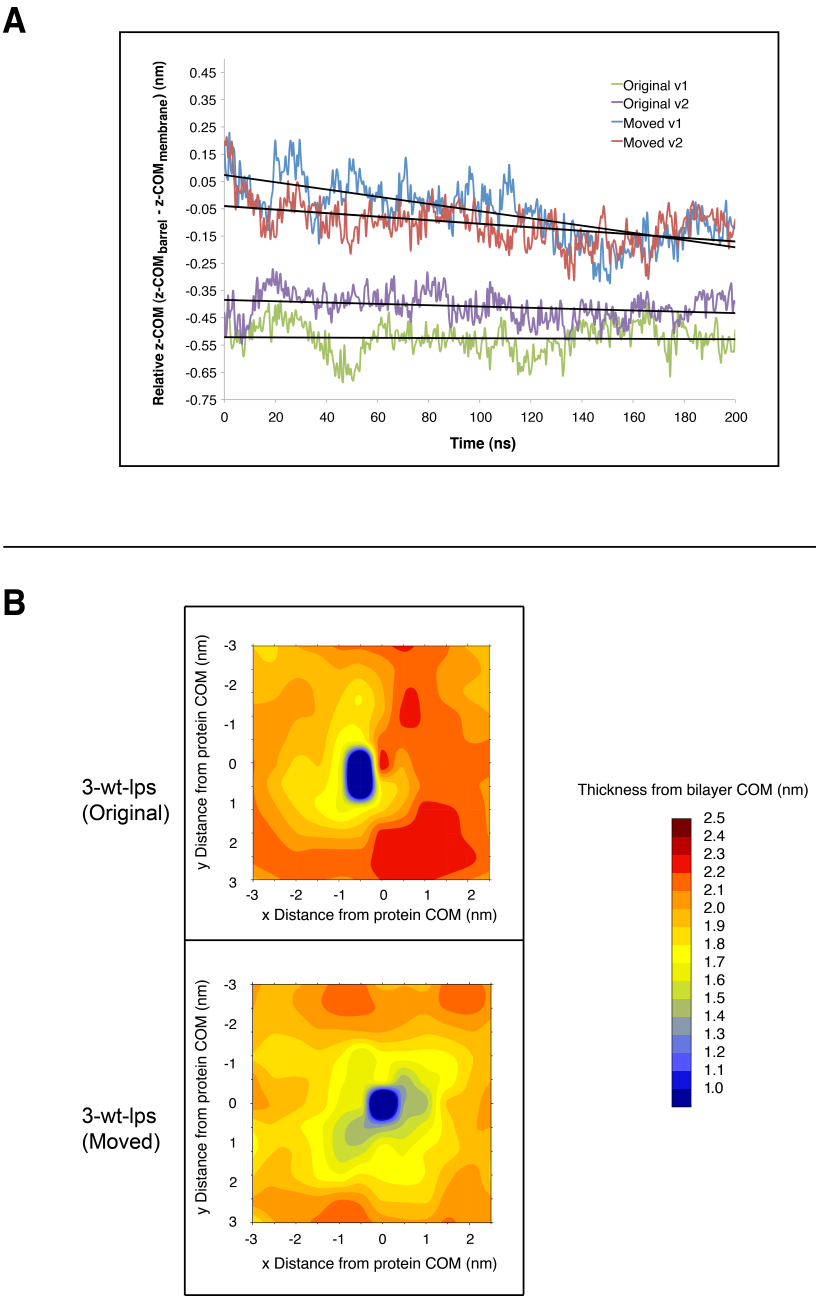


Figure 34. (A) The difference in the centre of mass of the β-barrel and the centre of mass of the bilayer in the z-coordinate for the original simulations (green and purple) and simulations where the initial position of the protein was altered (blue and red). (B) Bilayer thickness based on the average of snapshots taken at 0.5 ns intervals from a 100 ns trajectory. The space occupied by Hia appears as a dark blue patch in the centre of the plots. A clear thinning of the bilayer can be observed around Hia in the inner leaflet of 3-wt-lps (moved) and this is not apparent in 3-wt-lps (original). The translational and rotational motion of Hia in the xy plane was removed prior to performing all the analyses.

4.5 Discussion

Several molecular dynamics simulations have been performed of the intact Hia translocator domain, and also of this domain with various structural elements removed. In simulations of a single Hia monomer in isolation, it was demonstrated that the α1-β1 loop and α-helix stabilise the β-sheet through key H-bond and hydrophobic interactions. It has been suggested that the central α-helix of a monomeric autotransporter, EspP, from *E. coli* could be positioned within the proto-β-barrel domain at an early stage in the protein's biogenesis (*199*). While it is possible that the process of membrane integration may differ for trimeric autotransporters and between bacterial species, it is interesting to note that the simulations of the Hia monomer showed the α1-β1 loop and the α-helix to have a stabilising effect on the β-sheet residues. Presumably, polar interactions play a role in positioning the α1-β1 loop and α-helix so that they can pack tightly onto β-strands 1 and 2. The tight binding of the β-strands residues to the α1-β1 loop and α-helix may restrict the movement of the α1-β1 loops and thereby stabilise β-strand 1. Any disturbance of this binding could reduce the stability of the monomer. In Hia, G1064 provides part of the base to the hydrophobic pocked into which a α1-β1 loop residue, L1037, is inserted. In YadA, mutations of G389, which is synonymous to G1064 in Hia, resulted in a reduced stability of the trimeric form of the protein (*200*). Overall, the monomer is observed to be relatively unstable in an aqueous environment, and more stable when it is inserted in the membrane, where it causes local distortion of the lipid bilayer. Taken together these observations suggest that trimerisation of the Hia is likely to occur prior to membrane insertion.

In simulations of the full wild type translocator domain, the Hia trimer was shown to be structurally stable in a DMPC bilayer on a 100 ns timescale. Indeed, the plateau RMSD values were similar to those reported from simulations of the monomeric autotransporter, NalP. In contrast, the plateau RMSD value was larger when the α-helices were removed from the β-barrel. Removal of all three α-helices at once did not cause destabilisation of the β-barrel on a 100 ns timescale. However, large conformational changes were observed, including distortion of the β-barrel and a loss of symmetry between the monomers. Thus, the simulations suggest that the α-helices and α1-β1 loops play a key role in maintaining the conformational integrity of the trimer.

A number of H-bonds were also identified that form between residues of the β-barrel, α-helices and α1-β1 loops. These interactions, which are not present in the X-ray structure, may also contribute to the stability of the trimer. Using immunoblot analysis, Meng et al. examined constructs in which (a) the α-helix had been removed and (b) the α-helix and α1-β1 loop had been removed (*46, 177*). Whereas (a) was only detected in a monomeric form, the construct in (b) was not detected at all. These experimental observations suggest that any extension to the simulations of the translocator domain without the α-helices or α1-β1 loops may lead to the eventual collapse of the trimeric configuration.

The relatively narrow diameter of the central pore of Hia, as observed in the X-ray structure, suggests the β-barrel would need to be flexible in order to accommodate a passenger domain passing through. Encouragingly, the simulations of translocator domain show that even the β-barrel of the wildtype trimer exhibits a natural flexibility, particularly at the intracellular and extracellular mouths. In the absence of α-helices and α1-β1 loops, the β-barrel adopts a triangular conformation at the extracellular mouth. The flexibility of the β-barrel in this region may have important implications for the mechanism of passenger translocation. Indeed, the ability of the β-barrel to expand and contract in response to the α-helices located within the pore could also be extended to the presence or absence of the passenger domain. Even after taking into account the ability of the β-barrel to expand, however, it is difficult to conceive of a situation where the diameter of the pore could increase sufficiently to allow the passenger domain to pass through in anything other than an extended or unfolded conformation. Furthermore, the simulations of the truncated monomer and trimer suggest that the central α-helices and α1-β1 loops enhance the structural stability to the β-barrel. Given this information, it is difficult to envisage a mechanism where the three central α-helices were not inside the β-barrel during the transport of the passenger. Thus, the translocation would be especially difficult for the trimeric autotransporters following the hairpin model, as 4 to 6 polypeptides would have to be accommodated within the pore at the same time. In this context it seems unlikely that the passenger domain would be transported through the central cavity of the β-barrel.

In order to explore some of the specific membrane-protein interactions that would occur *in vivo*, Hia was simulated within a complex membrane model that

incorporates the asymmetry and some of the unique lipid components of the *H. influenzae* OM. While Hia made certain specific interactions with the more complex environment of the asymmetric membrane, it is important to note that the nature of the dynamics of Hia were similar in both the realistic OM and DMPC bilayers. The realistic bilayer, however, allowed the exploration of certain features of Hia that could otherwise not have been investigated. For example, the role of the positive patch, formed by lysine and arginine residues from the extracellular loops and α-helices, at the extracellular mouth of the β-barrel could be investigated. In the simulations of realistic OM model, Hia residues, K1019 and K1022, located in the α-helices were observed to interact regularly with the phosphate and sugar moieties of the LPS molecule. The simulations suggest that these interactions serve to anchor Hia within the outer leaflet of the OM. However, it is noted that the minimal LPS model, located within the outer leaflet of the realistic membrane simulations, is found only in an atypical strain of *H. influenza*. Though, it seems reasonable to suggest that longer variants of the LPS molecule, which are more representative of infectious strains of *H. influenzae*, would likely interact with the α-helices to a greater extent, leading to stronger anchoring of Hia within the OM. This is more important for the role of Hia as an adhesion than, perhaps, for the monomeric autotransporters that autoproteolytically release their passengers after translocation.

In summary, a number of key protein-protein interactions have been identified that stabilise the both the Hia monomer and trimer. The simulations indicate that the β-barrel of the Hia translocator domain is inherently flexibility, but it is unlikely that the passenger domain could be accommodated in its pore in anything other than an extended conformation. In addition, the simulations of Hia in a complex model of the OM have enabled the characterisation of key interactions that anchor the protein within its native membrane environment.

4.6 Future Work

It remains to be seen whether the lipids of the OM play a significant role in the translocation of the autotransporter passenger domain. Of particular interest is extracellular loop 4 in the monomeric autotransporter that appears to play a role in the transport of the passenger domain of BrkA (48), and has a similar structure to the loop 4 of EspP and Hbp. This loop is likely to have modified dynamics in a realistic environment that contains a full length LPS. Thus, a logical next step is to simulate the translocator domain of the monomeric autotransporters in a complex lipid environment. Currently, simulations of Hbp and EspP are being performed, and the early indications are that the extracellular loops have different dynamics in the complex membrane compared to simulations in simple phospholipid bilayers.

Chapter 5: The Dynamics of FecA in the Complex Environment of the Outer *Escherichia coli* Membrane

5.1 Abstract

Bacteria survive low iron environments by secreting siderophores, which solubilise otherwise inaccessible sources of the metal. FecA is an *E. coli* outer membrane protein that binds and internalises the siderophore, di-ferric di-citrate. Experimental studies have identified loop regions that are essential for the correct functioning of FecA. A number of these loops have been implicated in ligand binding. Thus, atomistic molecular dynamics simulations have been performed to investigate the impact of lipopolysaccharide and ligand binding on FecA. Simulations were performed with FecA embedded within two different bilayers: i) an asymmetric model membrane that incorporates lipopolysaccharide molecules and ii) a simple phospholipid membrane. In addition, comparative simulations of the apo and holo forms of FecA were performed. The simulations provide insight into the influence of the environment and ligand on the dynamics of the functionally important loops of FecA. In addition, some evidence is obtained of ligand induced conformational changes within the β-barrel of FecA.

The study presented in this chapter has been published as part of a journal article in Biochimica et Biophysica Acta (*221*). A full copy of this article is included in Appendix 8.4. The contributions of Dr Thomas Piggot, the lead author, to the article include the construction of the outer membrane model and development of the lipid parameters. Additionally, Dr Thomas Piggot performed the analysis of the ligand protein interactions and conformational changes in the switch helix.

5.2 Introduction

5.2.1 Membrane Composition and Protein Function

Changes to the composition of cellular membranes can influence the function of membrane proteins. The behaviours of enzymes, molecular pumps, and ion channel proteins can be altered by variations in lipid headgroup compositions and fatty acid chain length. Integral membrane proteins are also sensitive to solutes, such as cholesterol and general anaesthetics, which alter the physical properties of the membrane. Variations in the lateral pressure profile of the bilayer may provide a general mechanism for the influence of lipid components on membrane protein dynamics. In addition to changes in the general bulk properties of the membrane, specific protein-lipid interactions may stoichiometrically and allosterically modify protein structure and function.

The bacterial outer membrane (OM) contains an asymmetric distribution of lipids, with phospholipids on the inner leaflet and a unique lipid called lipopolysaccharide (LPS) in the outer leaflet. Typically, LPS has either 5 or 6 acyl acid tails, depending on the species of bacteria. In *E. coli*, the LPS molecules contain six acyl tails. The acyl tails combined with two glucosamine sugars represent the Lipid A, or membrane anchor, region of LPS. The Lipid A sugars are attached to a "core" set of sugars, which can be subdivided into the inner and outer regions. In *E. coli* with the R1 type core, the inner core region contains two KDO and three heptose sugars, while the outer core sugars consist of predominantly hexose sugars (*201-203*). The Rd mutants of *E. coli* lack the outer core sugars and one heptose sugar, while Rd_1 strains, in addition, lack a phosphate group on the inner heptose sugar compared to Rd mutants (*203*).

The integrity of the *E. coli* OM relies on the crosslinking of LPS molecules with divalent cations. Indeed, exposure to EDTA results in the release of LPS from the bacterial membrane (*204*). The LPS in the outer leaflet of the OM has been associated with protein folding, assembly and sorting. While it is still a point of contention, LPS has been shown to be useful, if not entirely necessary, for the folding of denatured OMPs *in vitro* (*205, 206*). In bacterial species that produce two types of LPS, with either a neutral or negative O-antigen, only the negatively charged O-antigen is found in outer membrane vesicles (*207*).

Mutations that disrupt the production of the negatively charged variant of LPS affect the sorting of proteins into outer membrane vesicles (*208*).

LPS is an important ancillary molecule in the function of some enzymes and porins, such as the Omptins and PhoE. The Omptins are a family of catalytic OMPs, and they require the presence of LPS to be catalytically active. The LPS binds to a specific site on the outside of the β-barrel (*209, 210*). How LPS activates the Omptins is still unclear. However, LPS may alter the behaviour of several extracellular loops, resulting in the widening the active site of the enzyme (*211*). In the case of PhoE, a phosphoporin from *Pseudomonas aeruginosa*, LPS has been implicated in the gating and trimerisation of the protein pore. Interestingly, the gating activity of PhoE was observed to be dependent on the length and composition of the sugar moiety attached to the lipid A anchor of LPS (*212*), indicating that interactions between the LPS and the extracellular loops of PhoE may be functionally important.

5.2.2 FecA: Ferric Citrate Transporter

Protein pores called porins mediate the transport of key nutrients across the OM. Large ions and molecules, which are often scarce in the environment, are actively transported across OM in a process that requires proteins located in both the inner membrane (IM) and OM (*213, 214*). Like the other members of the TonB-dependent transporters, the ferric citrate iron uptake transporter, FecA, is a 22-stranded β-barrel, with short turns connecting the β-sheets on the periplasmic side and the long loops on the extracellular side. In the centre of the β-barrel there is a globular plug domain (Figure 35A). Upon ligand binding, it is thought that FecA undergoes a large conformational change, allowing TonB to bind to the C-terminus of the plug domain. This in turn leads to another conformational change and passage of the ligand into the periplasm. A number of extracellular loops in FecA are considered essential to its function (*215-218*). Deletion of loops 3, 7, 8 and 11 of FecA results in a non-functional protein, while deletion of loops 5, 9 and 10 reduce the activity of FecA. It seems likely that residues located within the long extracellular loops of FecA will form electrostatic and hydrogen bond interactions with the sugars of the LPS in the outer leaflet of the OM. The interactions may have functional implications. A hint at how LPS might interact with FecA comes from a related

protein, FhuA, which has been crystallised with a bound LPS molecule at a putative binding site (*219, 220*).

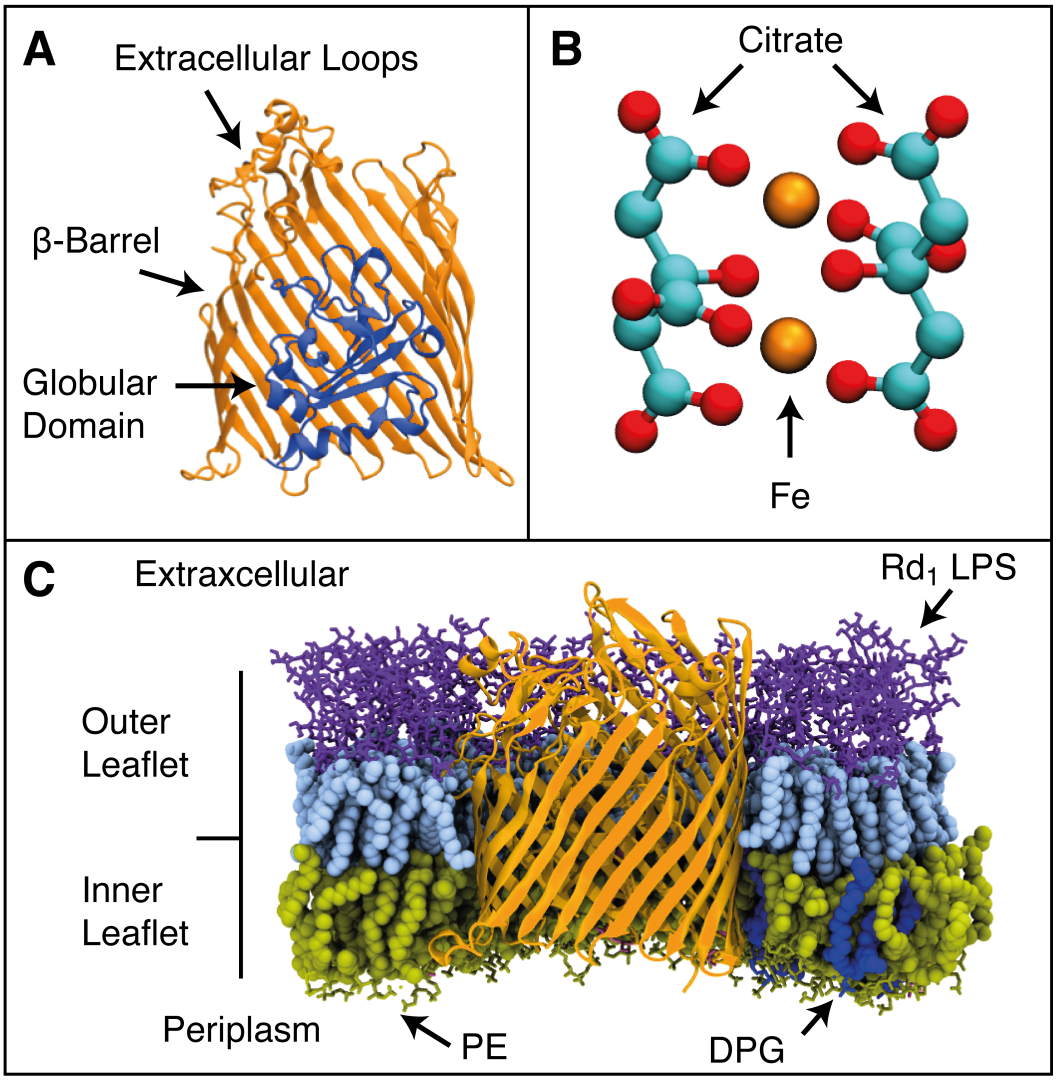


Figure 35. (A) The domain structures of FecA, (B) the structure of di-ferric-di-citrate (DFDC), and (C) FecA in the OM model.

5.2.3 Atomistic Molecular Dynamics Simulations of FecA

Previous simulation studies of bacterial OMPs in LPS membranes have indicated that extracellular loops can display altered dynamics as a result of the unique OM environment (*108*). Therefore, it is of interest to investigate how the OM environment affects the dynamics of FecA, given the functional significance of its extracellular loops. In addition, the conformational rearrangements that occur in FecA upon ligand binding are key to understanding the how di-ferric-di-citrate (DFDC) (Figure 35B) is transported into the periplasm.

The effects of ligand binding and specific membrane interactions on the dynamics of FecA are explored in the following comparative ATMD simulation study. Simulations of FecA were performed in both a simple phosphatidylcholine (PC) membrane and an *E. coli* OM model that contains Rd_1 LPS molecules in the outer leaflet and a mixture of phosphatidylethanolamine (PE), phosphatidylglycerol (PG) and diphosphatidylglycerol (DPG) in the inner leaflet (Figure 35C). In addition, MD simulations were performed of an apo and a pseudo-holo (DFDC bound) form of FecA in both the OM and PC membrane, so that the combined effects of the membrane and ligand on the protein dynamics could be investigated.

5.3 Methods

5.3.1 The Simulations of FecA

Full details of the simulation protocols and starting structures are provided in the Methods section of (*221*), and the full journal article is included in Appendix 8.4. For reference, the simulations of FecA were performed using the apo conformation of the structure (PDB ID: 1KMO). Eight simulations each of 200 ns duration were performed (Table 4). The simulations performed with the di-ferric-di-citrate (DFDC) ligand also used 1KMO as a starting structure, and it therefore represents a pseudo-ligand bound structure.

Table 4. Summary of FecA simulations.

Simulation	**Membrane Composition**	**Ligand (DFDC)**
apo_OM_v1	Outer Membrane Model	No
apo_OM_v2	Outer Membrane Model	No
DFDC_OM_v1	Outer Membrane Model	Yes
DFDC_OM_v2	Outer Membrane Model	Yes
apo_POPC_v1	POPC	No
apo_POPC_v2	POPC	No
DFDC_POPC_v1	POPC	Yes
DFDC_POPC_v2	POPC	Yes

5.3.2 Statistical Analysis

A two-way Analysis of Variance (ANOVA) was used to assess the effect of the ligand and membrane composition on the centre of mass (COM) displacements of FecA Cα atoms within extracellular loops. The COM displacements were calculated after every frame in a trajectory was least squares fitted to the β-barrel Cα atoms of the initial X-ray derived structure. Block averaging was performed in order to identify the longest correlation time in COM displacements for each extracellular loop of FecA. Correlation times of up to 100 ns were identified from the start of the 200 ns trajectories. Thus, frames from the first 100 ns of the trajectory were discarded and the block averaging repeated. The COM displacements in the final 100 ns of the trajectory had correlation times of less than 40 ns in duration. Thus, the COM displacements of the final 100 ns of each trajectory were averaged in 2 x 50 ns blocks. The averages obtained from each of the 50 ns blocks were used as a data points in the ANOVA. The assumptions of the ANOVA require that the response variables have homogeneity of variance and that they be normally distributed. These properties of the data were confirmed by visual inspection of box-plots.

5.3.3 General Analysis

Analyses were performed using GROMACS tools and locally written code. The COM movements of individual loops were calculated using the MDAnalysis python tools (*192*). Statistical tests were performed using the program, R (*222*). Molecular graphics images were generated using VMD (*165*).

5.4 Results

5.4.1 Structural Drift

Information regarding the conformational stability and overall motion of FecA in each of the simulations can be obtained by calculating the structural drift from the starting configuration. The structural drift is typically measured using the root mean square deviation (RMSD) after a least squares fit between two structures of interest. The RMSD from the initial FecA structure as a function of time was calculated independently for the extracellular loop and β-barrel Cα atoms (Figure 36). In all simulations, the RMSD initially increased while the system was relaxing, reaching a plateau value after 50 to 100 ns. The β-barrel showed the least deviation from the starting structure (Figure 36C and Figure 36D), and the greatest deviation was in the extracellular loop regions of the protein (Figure 36A and Figure 36B). Interestingly, differences in the plateau RMSDs were observed between the POPC and OM simulations and also between apo and DFDC-bound simulations.

The plateau RMSD values of the extracellular loops tended to differ depending upon the membrane environment in which FecA was embedded. In the POPC simulations the loop RMSD reached a value between 0.45 and 0.65 nm, while in the OM simulations the loops RMSD reached a value between 0.30 and 0.40 nm (Figure 36A and Figure 36B; Table 5 and Table 6). Block analysis of the RMSD values over the final 50 ns of the simulations was performed (*223*) to ensure that these plateau RMSD values were converged. This analysis showed that the standard error of the block means converged for block sizes of greater than 10 ns. The mobility of the loops, and the impact of the membrane environment, can be visualised in the average density plots of loop Cα atoms (Figure 37). Indeed, the loop Cα atoms appear as small points in the OM simulations (Figure 37A), while the densities are more dispersed in the POPC membrane simulations (Figure 37B). Thus, there is an increase in the mobility and conformational variability of the extracellular loops in the POPC membrane compared to the OM.

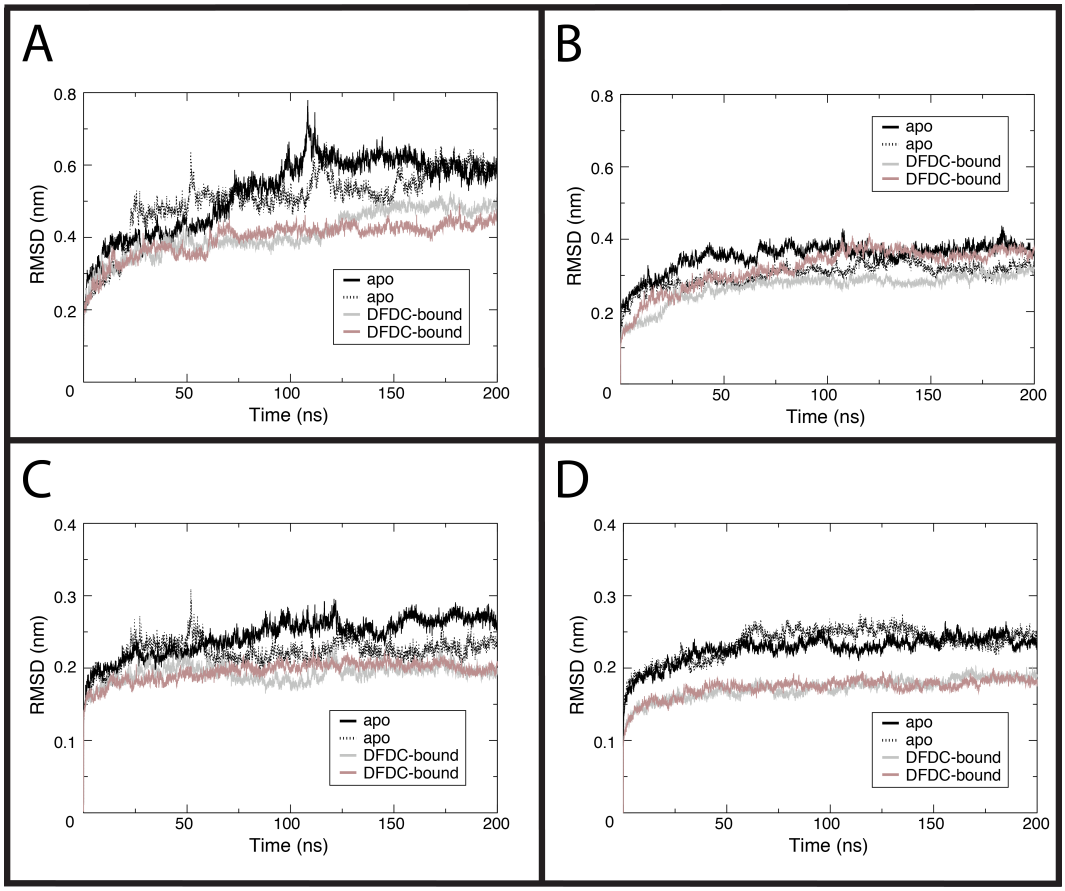


Figure 36. Root mean square deviations of A) the FecA loops in the POPC membrane, B) the FecA loops in the OM, C) the FecA β-barrel in the POPC membrane and D) the FecA β-barrel in the OM during apo (black solid and dotted lines) and DFDC-bound (grey and brown solid lines) simulations. The loops or β-barrel were fitted, using the least squares method, to the starting conformations prior to calculating the RMSD.

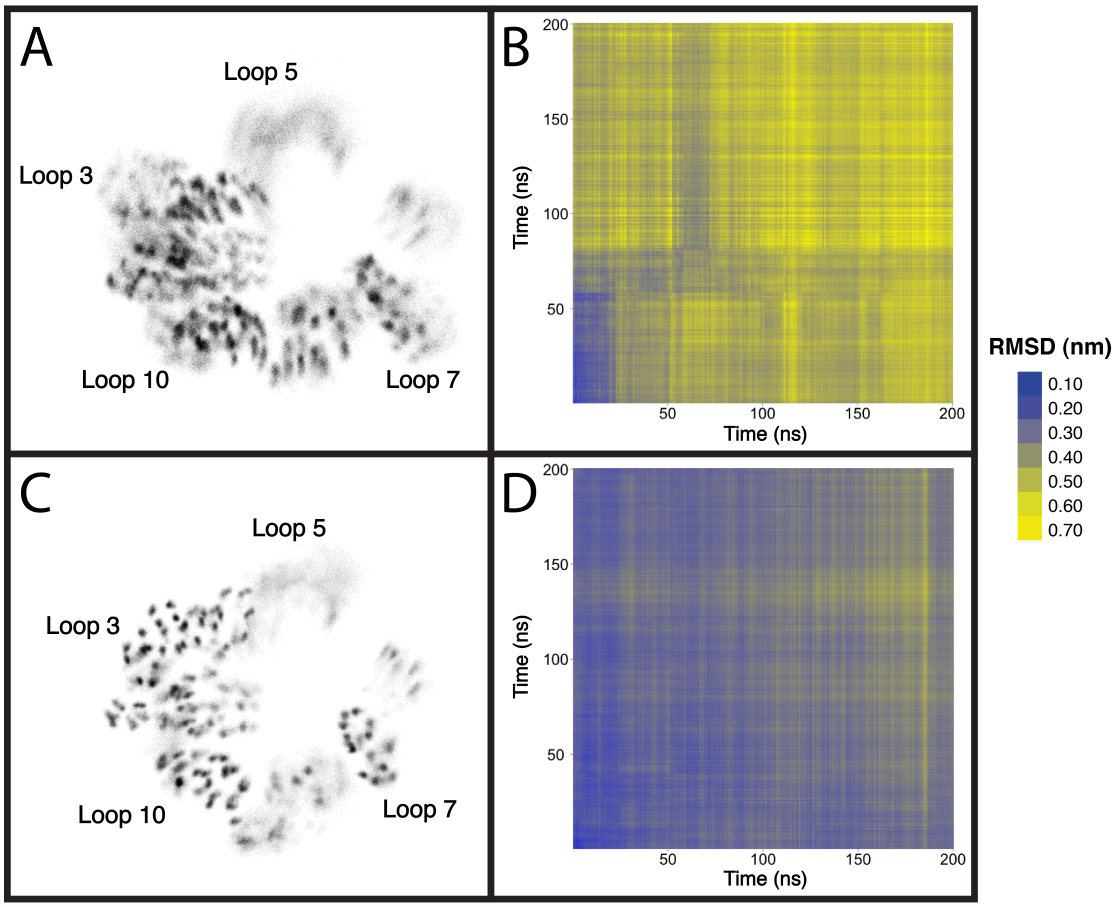


Figure 37. A top down view of the time averaged density of the Cα atoms from the extracellular loops of FecA for A) POPC and C) OM simulations. The Cα atoms from 400 sample frames were fitted, using the least squares method, to the starting loop conformations. The Cα densities are displayed on a grey scale, where black is high density and white is low density. The levels in this image have been altered equally so that regions with lower Cα densities can be seen more clearly. The differences in loop conformations are also shown in the all-to-all loop RMSD plots of the two apo POPC (B) and the two apo OM (D) simulations. The RMSD colour scale used in B and C is also shown in this figure.

Table 5. A table displaying the average extracellular loop RMSD values (nm) for the final 50 ns of each simulation of FecA. Each simulation is split by two factors: the membrane and the ligand state of FecA. The RMSD values for repeat simulations are provided.

	Membrane	
Ligand State	**POPC**	**OM**
Apo	0.595, 0.586	0.376, 0.317
DFDC bound	0.485, 0.433	0.357, 0.298

Table 6. A table displaying the average β-barrel RMSD values (nm) for the final 50 ns of each simulation of FecA. Each simulation is split by two factors: the membrane used and the ligand state of FecA. The RMSD values for repeat simulations are provided.

	Membrane	
Ligand State	**POPC**	**OM**
Apo	0.268, 0.225	0.240, 0.238
DFDC bound	0.197, 0.201	0.181, 0.187

The plateau RMSD values of the β-barrel also appeared to be lower if the DFDC ligand was bound to FecA. This difference was similar in both membrane environments, but was amplified in the OM. The RMSD of the β-barrel in the DFDC bound simulations reached a plateau value between 0.18 and 0.21 nm (Figure 36, C and D), while in the apo simulations the plateau value is higher, between 0.23 and 0.27 nm.

5.4.2 All-to-All RMSD Comparisons

Differences in the plateau RMSD values can suggest that the simulated proteins are sampling different regions of conformational space. However, similar plateau RMSD values do not necessarily mean that the two proteins are sampling the same region of conformational space (*224*). Indeed, the plateau RMSD is a simple measure of the average distance from the initial structure, where the distance is measured in a multidimensional space with 3N (three times the number of particles) dimensions. Therefore, the similarities in the conformational space explored by FecA in separate simulations can be examined, further, using all-to-all RMSDs (Figure 37, B and D, and Figure 38, Figure 39). The all-to-all RMSDs avoid the issues associated with using an arbitrary reference (such as the starting structure) in the standard RMSD calculation.

The all-to-all comparison of the extracellular loop RMSDs between two independent POPC simulations showed a large and fast divergence from the starting structure and between simulations, with a maximum RMSD value of 0.70 nm reached after 25 ns (Figure 37B). In contrast, the all-to-all comparisons of two independent OM simulations displayed much lower RMSD values in general. Indeed, the loop structures of two independent OM simulations reached a maximum RMSD value of 0.6 nm after 180 ns, but, thereafter, returned to a configuration more alike the starting structure, with RMSD values between 0.20 and 0.30 nm (Figure 37D). The lower divergence of the loop structures from the starting configuration in OM simulations is clear in all-to-all RMSD plots with POPC simulations (Figure 38B). The plots demonstrate that the loops in the POPC simulations diverge rapidly from the starting structure, while the loops in OM simulations remain much closer in structure to their initial configuration over the whole trajectory. This gives the appearance of a gradient across one of the axes in the plot.

The all-to-all plots of the β-barrel RMSDs showed fewer differences between simulations than the plots of the extracellular loops. The largest RMSD value between simulations was 0.25 nm, though the RMSDs remained below 0.20 nm in general. The smallest differences in the β-barrel RMSDs were seen in simulations where the DFDC ligand was bound to FecA. In these simulations the RMSDs stayed mostly below 0.15 nm (Figure 39A). In a similar way, though less dramatically, to the differences between the OM and POPC simulations in the extracellular loop RMSDs, a gradient in the all-to-all plot can be observed along one of the axis when DFDC simulations are contrasted with apo simulations (Figure 39E). This indicates that the ligand may restrict the conformational space that the β-barrel is able to explore.

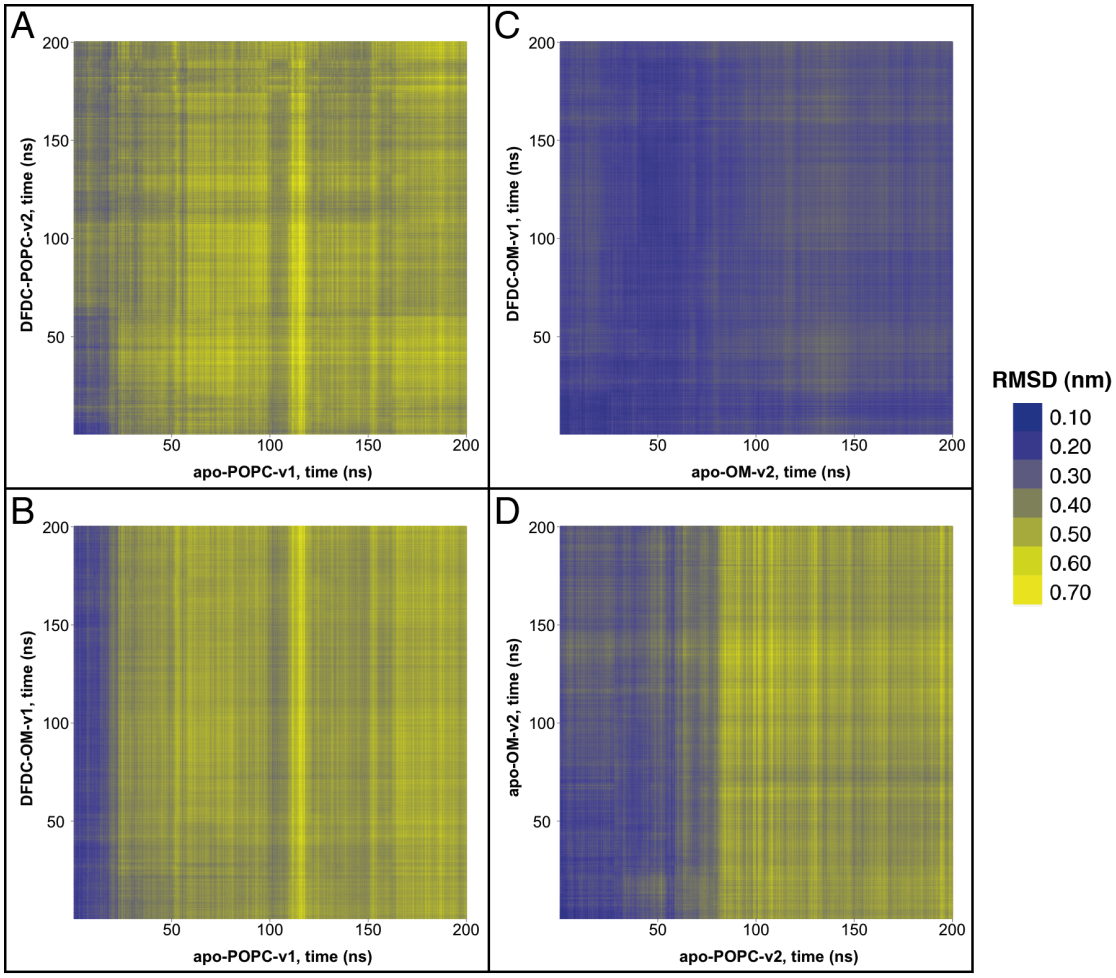


Figure 38. A selection of the Cα all-to-all RMSD plots of the extracellular loops, highlighting the impact of the membrane upon the extracellular loop RMSD. The colour scale for the RMSD is also shown in this figure (please note the difference in scale compared to Figure 39). The all-to-all comparisons shown are A) DFDC-bound in POPC vs apo in POPC; B) DFDC-bound in OM vs apo in POPC; C) DFDC-bound in OM vs apo in OM; and D) apo in OM vs apo in POPC.

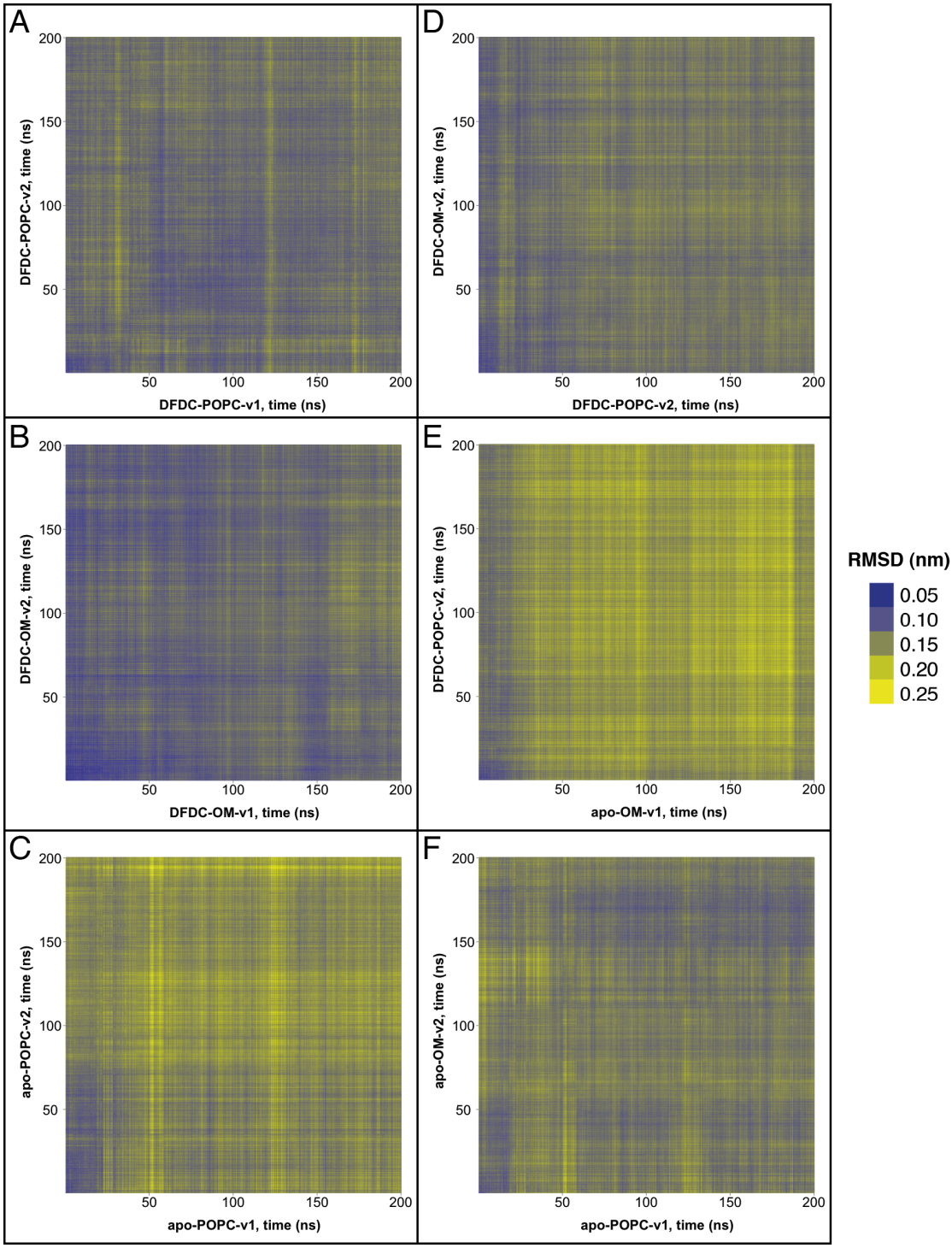


Figure 39. A selection of the all-to-all RMSD plots of the β-barrel, highlighting the impact of the ligand and the membrane upon the barrel RMSD. The colour scale for the RMSD is also shown in this figure (please note the difference in scale compared to Figure 38). The all-to-all comparisons shown are A) DFDC-bound in POPC (v1) vs DFDC-bound in POPC (v2); B) DFDC-bound in OM (v1) vs DFDC-bound in OM (v2); C) apo in POPC (v1) vs apo in POPC (v2); D) DFDC-

bound in OM vs DFDC-bound in POPC; E) DFDC-bound vs apo in OM; and F) apo in OM vs apo in POPC.

5.4.3 Essential Dynamics

The influence of the membrane and ligand on the conformational space explored by FecA was examined further using a Principal Component Analysis (PCA), which was performed on a 1.6 μs trajectory constructed from the 8 independent 200 ns trajectories. A separate PCA was performed for the Cα atoms of the β-barrel and extracellular loops. In both cases the original trajectories were projected onto the first two principal components (Figure 40). The first two principal components accounted for 30.8 % of the total variance in the extracellular loops and 31.1 % of the total variance in the β-barrel.

The projections suggest that the extracellular loops of FecA are conformationally restricted in the OM simulations, since their structures only occupy a small region on the first two principal components (Figure 40B). Meanwhile, the extracellular loop structures from POPC simulations are spread across the first two principle components, suggesting a they able adopt more, and different, conformations.

In contrast to the extracellular loop structures, there is a much greater similarity between the simulations in the first two principal components of the β-barrel structures. Of note, there are four distinct clusters for β-barrel structures, and only one of these is occupied by the structures from simulations of FecA with the DFDC ligand bound (Figure 40A). However, it may be necessary to improve the sampling, by performing a larger number of independent simulations, in order to gain conclusive evidence that DFDC binding alters the dynamics of the β-barrel.

The covariance overlap compares all of the eigenvectors and eigenvalues, not just the first two components, between individual simulations (*225-228*). A value closer to 1.0 indicates greater similarity in the eigenvectors and eigenvalues, and, hence, greater similarity in their conformational dynamics. The subspace overlap of the extracellular loop motions appears to be influenced by the membrane environment surrounding FecA (Table 7B). Indeed, the subspace comparisons of independent OM simulations had values between 0.32 and 0.37, while the comparisons of POPC simulations were

mostly below 0.33. In contrast, the ligand appears to have little impact on the β-barrel subspace movements, with a similar covariance overlap in all of the simulations (Table 7A).

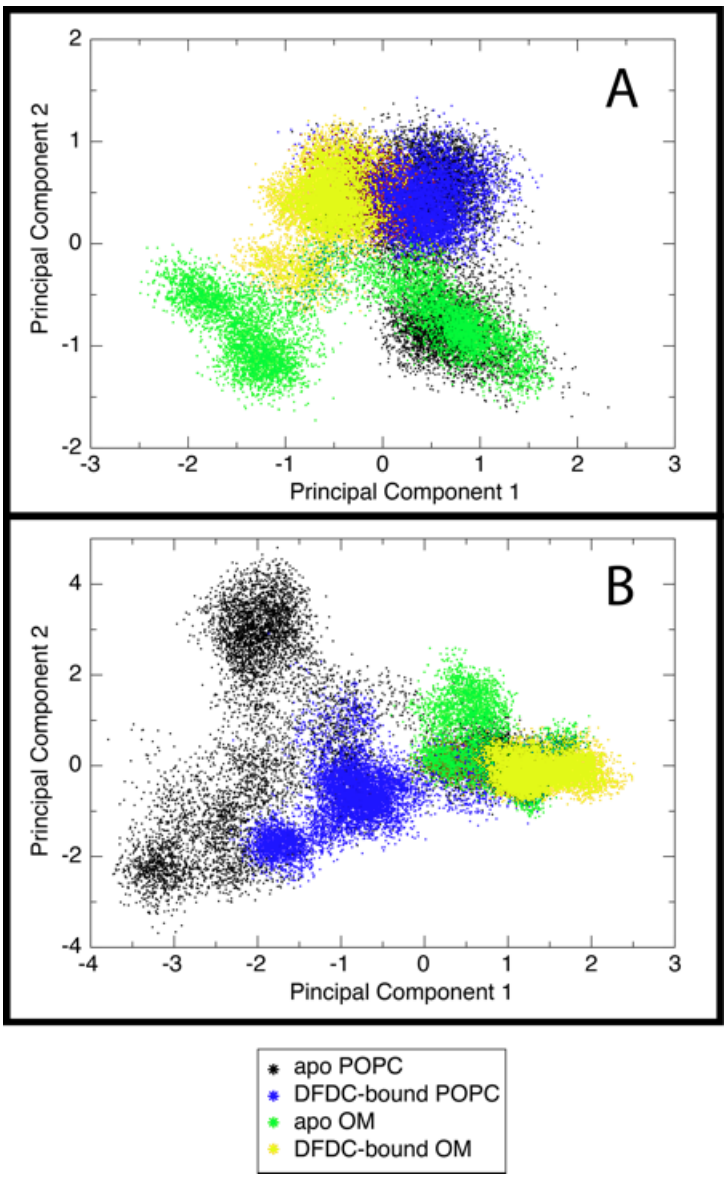


Figure 40. Projections of the A) barrel and B) loop structures from different simulations on the first two principal components determined from the combined principal component analysis. The apo POPC simulations are shown in black, the apo OM simulations in green, the DFDC-bound POPC simulations in blue and the DFDC-bound simulations in yellow.

Table 7. Overlap of the subspace defined by all of the eigenvectors and eigenvalues for (A) the β-barrel and (B) the extracellular loops from the different POPC and OM simulations.

A	apo POPC v1	apo POPC v2	apo OM v1	apo OM v2	DFDC POPC v1	DFDC POPC v2	DFDC OM v1	DFDC OM v2
apo POPC v1	-	-	-	-	-	-	-	-
apo POPC v2	0.484	-	-	-	-	-	-	-
apo OM v1	0.460	0.399	-	-	-	-	-	-
apo OM v2	0.484	0.411	0.454	-	-	-	-	-
DFDC POPC v1	0.534	0.469	0.430	0.457	-	-	-	-
DFDC POPC v2	0.549	0.459	0.439	0.468	0.517	-	-	-
DFDC OM v1	0.476	0.406	0.475	0.469	0.447	0.477	-	-
DFDC OM v2	0.478	0.431	0.507	0.484	0.460	0.482	0.506	-

B	apo POPC v1	apo POPC v2	apo OM v1	apo OM v2	DFDC POPC v1	DFDC POPC v2	DFDC OM v1	DFDC OM v2
apo POPC v1	-	-	-	-	-	-	-	-
apo POPC v2	0.319	-	-	-	-	-	-	-
apo OM v1	0.278	0.29	-	-	-	-	-	-
apo OM v2	0.223	0.243	0.344	-	-	-	-	-
DFDC POPC v1	0.294	0.267	0.300	0.264	-	-	-	-
DFDC POPC v2	0.325	0.306	0.323	0.286	0.356	-	-	-
DFDC OM v1	0.208	0.233	0.334	0.366	0.273	0.293	-	-
DFDC OM v2	0.248	0.249	0.320	0.317	0.257	0.299	0.331	-

5.4.4 Fast Fluctuations of the Extracellular Loops

The plateau RMSD values from the initial X-ray derived structure (Figure 36), the all-to-all RMSD plots (Figure 37) and the PCA (Figure 40) have all demonstrated that the LPS molecules reduce the overall conformation space available to the extracellular loops. In order to explore the influence of LPS on the local fluctuations of the loops, the short timescale (50 ps) centre of mass (COM) displacements of the individual FecA loops, relative to the Cα atoms in the β-barrel, were calculated. Block analysis was used to determine the longest correlation times for the 50 ps COM movements. In nearly all cases the correlations lasted for less than 100 ns (Figure 41A). Therefore, the frames from the first 100 ns were discarded from each trajectory in order to remove any influence from the system relaxing. The block averaging was repeated for all the loops in the final 100 ns of the trajectories, and in this instance, the longest correlation times for the 50 ps COM movements were less than 40 ns (Figure 41B).

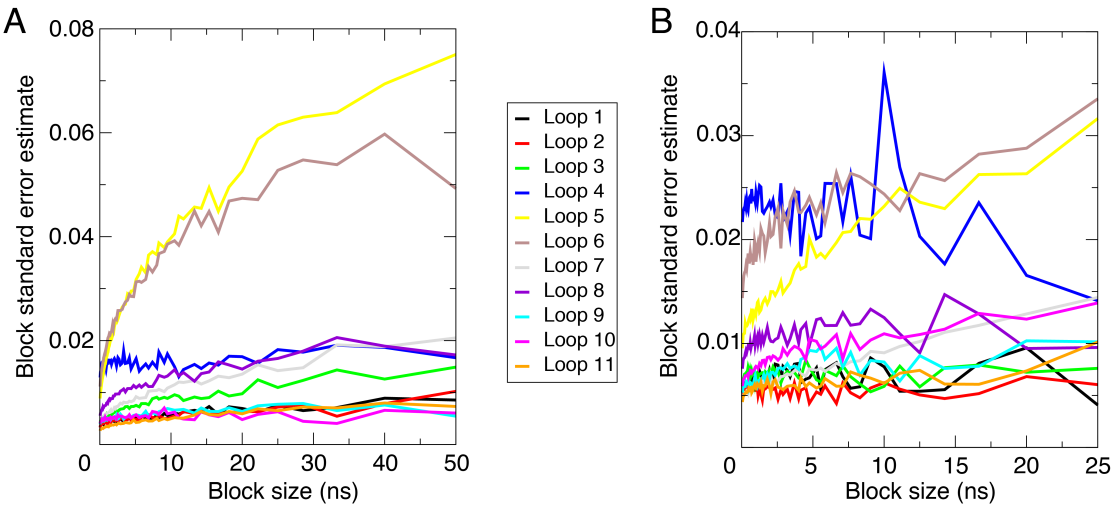


Figure 41. Block analysis showing the estimate of the standard errors of the block means for each extracellular loop of FecA over A) the complete 200 ns and B) the final 100 ns from the same DFDC-bound POPC simulation. These results are typical of all the POPC and OM simulations.

In general, the COM displacements of the loops were smaller in the OM simulations than in the POPC simulations. The statistical significance of differences in COM loop displacements between simulations was evaluated using an ANOVA. The ANOVA was performed using data averaged over two 50 ns blocks from the final 100 ns of each trajectory. The 50 ns blocks were chosen so that the data were average over a time longer than the longest correlation time (Figure 41). The movements of the loop COM over 50 ps intervals were significantly different between the POPC and OM simulations at the 5% level for extracellular loop 4 ($p = 0.015$), loop 5 ($p = 0.016$) and loop 6 ($p = 0.012$) and at the 1% level for loop 1 ($p < 0.001$), loop 2 ($p = 0.002$), loop 3 ($p < 0.001$), loop 7 ($p = 0.006$), loop 8 ($p = 0.004$), loop 9 ($p < 0.001$), loop 10 ($p < 0.001$) and loop 11 ($p = 0.001$) (Table S5). The ligand also had a statistically significant effect at the 5% level on loop 5 ($p = 0.039$) and loop 9 ($p = 0.039$) and at the 10% level on loop 11 ($p = 0.055$). These results suggest that differences in membrane environment, and also binding of the DFDC ligand, can alter the short timescale displacements of the FecA loops.

5.5 Discussion

Simulations have been performed of FecA in both POPC and OM bilayers. They have provided details regarding the nature of the interactions between the FecA and its native OM environment. Perhaps as expected, the inner core sugars of the LPS molecules interact extensively with the extracellular loop residues, these interactions, understandably, are absent in the POPC simulations. The novel interactions between FecA and LPS had a significant impact upon the loop dynamics, increasing their conformational stability compared to the POPC simulations. The short-term fluctuations in the extracellular loop positions were significantly altered by interactions with LPS. In addition, the number of conformations that could be adopted by the extracellular loops was restricted in the LPS membrane compared POPC. Thus, the bulky LPS molecules appear to provide a steric hindrance to large conformational motions in the extracellular loops. These results highlight the need for complex membrane models in simulations of proteins containing long, functional extracellular loops.

It is noted that due to the slow diffusion rate of LPS, any influence on the dynamics of FecA, resulting from specific protein–LPS interactions, is to some extent influenced by aspects of the initial simulation setup, such as the positioning of the protein within the membrane. Given that multiple independent simulations, each of which is 200 ns in length, it is reasonable to suggest that the level of sampling in the present study is sufficient to observe the influence of LPS on the extracellular loop dynamics. However, with many more simulations of greater length, it is expected that the differences between environments will become more apparent, as the conformation space available to FecA would be sampled more completely.

Chapter 6: Coarse Grained Simulations of Outer Membrane Vesicles: Lipid Diffusion and Protein Aggregation

6.1 Abstract

Outer membranes vesicles are spheroid particles secreted by all Gram-negative bacteria. Their composition is similar to that of the outer membrane from which they are derived. The functions ascribed to the outer membrane vesicles include long-range toxin delivery, sequestration of harmful agents and communication between bacteria. Many features of outer membrane vesicle production, and the selective uptake of protein and lipid cargo, are still unknown. Therefore, the diffusion and aggregation of proteins, which may lead to selective incorporation, has been investigated in coarse grained molecular dynamics simulations of unilamellar vesicles that mimic the size and composition of bacterial outer membrane vesicles. Indeed, the inner leaflet of the simulated vesicles contains a mixture of phosphatidylethanolamine, phosphatidylglycerol and cardiolipin lipids, closely resembling the composition of inner leaflet of the *E. coli* outer membrane. A high protein concentration within the vesicles is shown to slow the lipid diffusion, as a result of long-timescale interactions between the lipids and proteins. These interactions result in highly correlated lipid and protein motions at distances of less than 5 nm. A lower level of correlation is also observed between the lipids at the same distance. Finally, the aggregation characteristics of the *E. coli* porin, OmpF, is investigated, leading to the observation that changes in membrane thickness around the protein may determine the orientation of interactions.

6.2 Introduction

6.2.1 Outer Membrane Vesicles

Gram-negative bacteria continually discharge Outer Membrane Vesicles (OMVs) from the cell surface during growth. While being formed, the OMVs capture some of the periplasm within their interior. The natural, asymmetric arrangement of the bacterial outer membrane components is maintained in OMVs, and they are, in essence, small pieces of the bacterial cell wall. The OMVs have a spherical structure with an average diameter of between 20 and 250 nm, depending on the strain of bacteria from which they are derived. The OMVs produced by *E. coli* are typically 25 to 100 nm in diameter, with 75 % of those being between 25 and 50 nm in size (*229*). The OMVs allow bacteria to secrete lipids, membrane proteins and other insoluble material.

The biological functions ascribed to the OMVs are diverse. Mostly, the OMVs are a means by which a bacterium can interact with its environment. OMVs are produced constitutively in many bacteria, though the rate production of OMVs can vary depending on environmental stimuli. Importantly, OMVs can serve as virulence mediators for pathogens by carrying toxins, proteases and other degradative factors. Thus, the OMVs allow bacteria to behave in a predatory manner, through secretion, binding and lysis of host cells or other bacterial species. Indeed, the OMVs can cause a severe inflammatory response, known as sepsis, in host organisms (*229-231*). The ability of OMVs to activate the immune system have made them potential vehicles for immunization against bacterial pathogens (*231-236*). The production of OMVs may also function as a waste disposal pathway by collecting misfolded and unfolded proteins from the periplasm, and then expelling them from the cell (*237*). It also seems likely that OMVs provide short term protection for the bacteria against antimicrobials by providing decoys to absorb the harmful agents (*238*).

The OMVs may play an important role in the communication and coordination between bacteria, a process called quorum sensing. The *P. aeruginosa* Quinolone-Signal (PQS) molecule is a hydrophobic molecule, and it is found to preferentially intercalate into the OM, inducing membrane curvature. Indeed, the PQS molecule appears necessary and sufficient for OMV formation in *P. aeruginosa*. The finding that PQS cosediments with OMVs is further evidence

for a role in quorum sensing. Indeed, the removal of OMVs from bacterial populations results in the abolition of the group behaviour facilitated by PQS (*239*).

The mechanics of OMVs production are still undetermined. It is also unknown whether OMV production is a stochastic or regulated process. However, there is mounting evidence suggesting that OMV production is a regulated process. The production of OMVs is increased under OM or periplasmic stress, for example. Furthermore, the stress response genes, rseA, degP and degS, cause constitutively high levels of OMV production. In addition, certain proteins are found to be in higher concentrations in OMVs than in the bacterial envelope, suggesting a selective incorporation of cargo and virulence factors (*208*). In some cases there also seems to be selective exclusion of some of the more abundant periplasmic proteins (*240*). A stochastic process for OMV is thought likely due to the inherent curvature stress induced by lipopolysaccharide in the outer leaflet of the OM. Mutant bacteria lacking an OM anchor to the peptidoglycan also show a marked increase in OMV production.

6.2.2 Molecular Simulations of Vesicles

Much like planar membrane simulations, small vesicles can be assembled through self-assembly (*241-243*). In practice, however, it is simpler to create the initial coordinates by either biasing phospholipids to form a vesicle with artificial particles (*244*) or to generate loosely packed lipids with a lipid-packing algorithm (*245, 246*).

Early CG vesicle simulations studied the tightly regulated process of vesicular fusion, using two 15 nm vesicles with heterogeneous lipid compositions (*247*). In other MD studies, a reverse mapping procedure has been used to gain atomistic insights into the process of a vesicle fusion (*248*). Budding and fission of DPPC vesicles has been observed from oversaturated DPPC lipid bilayers.

Vesicle shape changes have been observed in molecular dynamics simulations. Indeed, altering the composition of vesicles, to include lipids that introduce spontaneous curvature, can generate a number of vesicle shapes (*249*). Proteins have also been observed to induce shape changes in lipid vesicles.

Small unilamellar vesicles have been used to study protein aggregation. Indeed, a mismatch in the hydrophobic thickness between proteins and the lipid bilayer was shown to promote protein aggregation. The nature of the protein aggregation was markedly different in vesicles compared to that observed in planar membranes (*244*). Simulations of a bacterial chemoreceptor dimer, Tsr, in 70 nm vesicles, containing a biologically relevant mixture of lipids, have revealed key lipid-protein interactions and oligomerisation events (*245*).

6.2.3 The OMV Simulations

Key questions remain about how proteins can be selectively sorted into OMVs. The OMVs produced by *E. coli* have been found to contain proteins such as OmpF, FhuA and OmpA (*240, 250*). In contrast, the EspP translocator domain has been found not to be associated with OMVs, and instead remains OM bound (*39*). In order to better understand the dynamics of OMPs in the bacterial OMVs, a number of CGMD simulations have been performed of vesicles. The vesicles are approximately 50 nm in size and they contain a complex mixture of lipids and OMPs. The inner leaflet of the vesicle membrane approximates the composition of the inner leaflet of the *E. coli* OM.

6.3 Methods

6.3.1 Outer Membrane Vesicles

The OMVs were constructed by modifying a 50 nm DPPC vesicle. The OMVs have a complex composition of inner leaflet phospholipids (Table 8), similar to that found in the inner leaflet of the *E. coli* outer membrane (*19, 21, 251*), comprising a phospholipid headgroup composition of 90% PE and 5% PG (*20, 21*). Similar to the *E. coli* OM (*252*), the OMVs also contain approximately 5% cardiolipin, a phospholipid created by the fusion of two phosphatidylglycerols. The cardiolipin molecules, therefore, can have any combination of tail pairs from the PG phospholipids in Table 8. In order of their abundance in the OMVs, the cardiolipin molecules, abbreviated by the fatty acid tail types, are: PVPV, PVPQ, PVVQ, PVDV, PVDP, PQVQ, PQDV, VQVQ, VQDV, VQDP, DVDV, DVDP, PQPQ, PQDP and DPDP. Currently, there is not a lipopolysaccharide model available for the MARTINI force field. Therefore, the outer leaflet of the OMVs contains DLPC to approximate the thickness of the corresponding leaflet of the OM.

Due to the coarse nature of the MARTINI force field, the same molecular model can be used to represent more than one lipid tail. For example, the model for DVPE is equivalent to DOPE. The double bond in the DVPE fatty acids tails occurs at a different position to DOPE, but the coarse grain bead for the double bond is mapped onto same position in both models.

A number of different proteins were inserted into the OMVs in varying concentrations (Table 9). The proteins were inserted manually into the vesicles so that they were equally spaced around the vesicle. Any overlapping lipids were deleted. In addition to the OMVs, two other 50 nm vesicle simulations were performed both containing no proteins. One of these contained only DPPC (NoPro_DPPC), and the other contained a mixture of POPE to POPG in a ratio of 3:1 (NoPro_POPE_POPG).

Table 8. Approximate two-tailed phospholipid composition of the *E. coli* OM, inner leaflet.

Chemical name	**% of total OMV lipids**	**Abbrev.**
1-palmitoyl 2-palmitoleoyl phosphatidylethanolamine	30.6	PQPE
di-palmitoyl phosphatidylethanolamine	12.6	DPPE
1-palmitoyl 2-cis-vaccenoyl phosphatidylethanolamine	25.2	PVPE
1-cis-vaccenoyl 2-palmitoleoyl phosphatidylethanolamine	17.1	VQPE
di-cis-vaccenoyl phosphatidylethanolamine	4.5	DVPE
1-palmitoyl 2-cis-vaccenoyl phosphatidylglycerol	1.6	PVPG
1-cis-vaccenoyl 2-palmitoleoyl phosphatidylglycerol	0.9	VQPG
di-cis-vaccenoyl phosphatidyglycerol	0.9	DVPG
1-palmitoyl 2-palmitoleoyl phosphatidylglycerol	1.2	PQPG
di-palmitoyl phosphatidyglycerol	0.4	DPPG

Table 9. A list of OMV simulations.

Simulation ID	Proteins	Repeats	Total time (μs)
NoPro_OMV	None	2	2.1
2_OmpA_OMV	2 x OmpA	2	2.5
2_OmpF_OMV	2 x OmpF	2	2.3
2_FhuA_OMV	2 x FhuA	2	3.8
2_EspP_OMV	2 x EspP	2	4.1
4_OmpA_OMV	4 x OmpA	2	2.3
4_OmpF_OMV	4 x OmpF	2	2.3
4_FhuA_OMV	4 x FhuA	2	4.6
4_EspP_OMV	4 x EspP	2	4.8
4_Mix_OMV	2 x OmpA, 2 x OmpF	1	1.4
8_EspP_OMV	8 x EspP	2	2.5
8_Mix_OMV	2 x OmpA, 2 x OmpF, 2 x FhuA, 2 x EspP	2	6.9
32_Mix_OMV	8 x OmpA, 8 x OmpF, 8 x FhuA, 8 x EspP	2	4.6
32_OmpF_OMV (outer leaflet DLPE)	32 x OmpF	1	1.4

6.3.2 Simulation Protocols and Analysis

All vesicles were simulated using the GROMACS package, version 4.0.7. The MARTINI force field version 1.4 was used for the lipid, water and ions, while the protein was modelled as described in (*76*). All the vesicles, apart from NoPro_DPPC, were maintained at constant temperature of 313 Kelvin using the Berendsen thermostat with a time constant of 10 ps. The NoPro_DPPC vesicle was maintained at 323 Kelvin to be above its phase transition temperature of 314 K. A pressure of 1 bar was maintained using anisotropic pressure coupling with the Berendsen barostat with a time constant of 10 ps. The van der Waals interactions were shifted to zero between 0.9 and 1.2 nm. Electrostatic interactions were truncated at 1.2 nm. The neighbour list was updated every 10 steps during the simulations. A rhombic dodecahedron simulation cell was employed to reduce the number of solvent molecules in the system. The analysis of the simulations was performed using GROMACS tools and locally written scripts. The figures were generated using VMD and R.

6.3.3 Diffusion Coefficients of Lipids in a Vesicle

The diffusion coefficients, D, for lipids were obtained using the Einstein-Stokes relation (Equation 22),

$$D = \lim_{t\to\infty} \frac{\langle \big(r(t) - r(0)\big)^2 \rangle}{2dt}$$

Equation 22

where, d, is the number of dimensions, $r(t)$, represents the position of the particle at time, t, and $r(0)$ is initial position of the particle The average of the time displacements is taken over all particles in the system to lessen the statistical error. In addition, the averaging is performed over multiple time origins. In planar membrane simulations the random walk occurs in just 2 dimensions, typically the xy dimensions, so the value of d is set to 2.

In a vesicle, the lateral displacement of lipids occurs with a magnitude in all three spatial dimensions, x, y, and z. Despite this, it is not acceptable simply to

use the 3-dimensional form of the Einstein-Stokes relation to calculate the lateral diffusion coefficient of lipids in a vesicle. The diffusion is still restricted to a surface, which is not planar, but one that is approximately spherical. Thus, the displacement of the lipids must be calculated in a way that accounts for their restricted movement.

The spherical surface can be defined from the vesicles centre of mass (COM). The radius of the sphere is calculated as the mass-weighted average radius, r_v, from the vesicles COM,

$$r_v = \frac{\sum_{i=1}^{N} m_i r_i}{\sum_{i=1}^{N} m_i}$$

Equation 23

where m_i is the mass of particle i, and r_i is the distance of particle i from the vesicle's COM. In practice, the average radius from the geometric centre of the vesicle is calculated, rather than the mass-weighted radius from the COM, since all the particle types in MARTINI version 1.4 have equivalent mass.

It is important to note that the spherical surface, which restricts the movement of the lipids, is different for the inner and outer leaflets of a bilayer. In addition, there are more particles in the outer leaflet than the inner leaflet, due to the differences in the surface area. Thus, a sphere created using an average radius of all the particles would not be midway between two spheres created independently for each leaflet. In addition, a "midway" spherical surface for both leaflets would not be desirable. A midway sphere would amplify lateral displacements for inner leaflet lipids and contract those for outer leaflet lipids. Therefore, the radius and COM need to be calculated separately for each leaflet. In principle, this allows for direct comparisons to be made between the inner and outer leaflets, since the spherical surface is placed at equivalent depths in both leaflets.

In planar membrane simulations it is commonly assumed that the lipids move a series of straight lines within the plane, since a straight line between two end points is the shortest of all possible paths. If the spherical approximation of

the vesicle is appropriate then for consistency with the planar approach, the lipids can be assumed to move in small arcs across the surface of the sphere following the path of the great circle though the end points of the motion, since this represents the shortest path between two points on a sphere. An important consequence of this assumption, also consistent with the planar approach, is that it implies no rotation of the lipid about the surface normal during the motion (since the surface normal is orthogonal to the axis of rotation). This preserves the assumption, implicit in the planar approach, that the particle motion is irrotational (an important concept in classical fluid mechanics theory). Thus, the total distance travelled will be the arc of the great circle between a lipid's COM at time, t, and, $t + \Delta t$. It is possible to calculate the great circle arc by first calculating the chord of a circle between two lipid's COM. However, this would only be accurate if the lipid COM were to move exclusively within the surface of the sphere. In reality, a lipids COM will move above and below the spherical surface, which is simply an average distance of the particles from the leaflet's COM. This out of surface movement of the lipids would result in an overestimate of the great circle distances, if the arc length were to be computed from the distance between lipids. Instead, the length of the great circle arc, s, has been calculated using relationship, $s = r\theta$, where, θ, in radians, is the angle of the segment defined by two points on a circle and the centre of the circle and, r, is the radius of the circle. This method allows the lipid COM position vectors, originating from the leaflet COM, at, t, and, $t + \Delta t$, to be treated as though the lipid were located on an ideal circle, obviating errors due to out of surface movements.

The angle between the lipid COM position vectors at, t, and, $t + \Delta t$, is calculated using the scalar product.

$$\theta = cos^{-1}\left(\frac{\vec{a} \cdot \vec{b}}{|a||b|}\right)$$

Equation 24

The great arc length, s, between the positions of a lipid at, t, and, $t + \Delta t$, is, thus the angle, θ, multiplied by the leaflet radius. This is simply the projection

of a lipid's displacement onto a spherical surface. To account for variability in the radius of the leaflet, the average radius is calculated between t, and, $t + \Delta t$.

The translational displacement of the entire vesicle between, t, and, $t + \Delta t$, is removed before calculating the great arc between the two position vectors. This is achieved by subtracting the COM movement from every particle position at time $t + \Delta t$.

An overall rotation of the vesicle also can cause error in the calculation of lipid displacements on a spherical surface. Before the lateral diffusion coefficient can be calculated, it is important to remove the displacement as a result of the vesicles angular momentum. This is achieved by repositioning the particles at time, $t + \Delta t$, so that there has been no net rotation of the vesicle between time t and $t + \Delta t$.

In planar bilayer simulations, the system has zero linear momentum if the sum of all the mass–weighted particle velocities is zero. Before calculating the diffusion coefficient, the effect of the systems linear momentum is removed from particle displacements in planar bilayer simulations by a rigid body correction, the subtraction of the system COM movement between trajectory frames from every particle.

The angular momentum, L, of a body is related to the product of its moment of inertia, I, and its angular speed, ω (Equation 25).

$$L = I\omega$$

Equation 25

In the same manner as planar bilayer simulations, the angular momentum can be removed from a vesicle with a rigid body adjustment by subtracting the mass-weighted angular velocity of all the vesicle particles.

The rotation of a position vector of a particle about an axis can be calculated using the vector product between its position vectors, $\vec{a}$ and $\vec{b}$, relative to the leaflet COM, at time t, and, $t + \Delta t$.

$$\vec{a} \times \vec{b} = |a||b| \sin\theta \, \vec{n}$$

Equation 26

$\vec{n}$, is a unit vector normal to the plane created by the origin (the leaflet COM) and the particle positions at t, and, $t + \Delta t$ and θ the magnitude of the angle of rotation of the particle's position vector about the origin. The $\vec{n}$ vectors of two particles moving in opposite directions around the same great circle will point in opposite directions. The expression can be rearranged so that the magnitude of the normal vector is θ. This transforms the length of the normal axis to be equivalent to the total angle traversed by a particle about the axis, in radians, between t, and, $t + \Delta t$.

$$\theta\vec{n} = \sin^{-1}\left(\frac{\vec{a} \times \vec{b}}{|a||b| \sin\theta}\right)\vec{n}$$

Equation 27

Summing all the scaled normal axes for the individual particles, and normalising for the number of particles, provides the angular rotation of the leaflet between t and $t + \Delta t$. Dividing by the change in time in seconds gives the average angular velocity, in radians per second.

$$\omega_v = \frac{1}{\Delta t N} \sum_{i=1}^{N} \theta_i \, \vec{n}_{i,t \to t+\Delta t}$$

Equation 28

Though not necessary for systems described by the MARTINI 1.4 parameters, it is more accurate to calculate the angular velocity as a mass-weighted average, which is the angular equivalent to the linear changes in the COM.

$$\omega_v = \frac{1}{\Delta t \sum_{i=1}^{N} m_i} \sum_{i=1}^{N} m_i \theta_i \vec{n}_{i,t \to t+\Delta t}$$

Equation 29

It is possible, though unlikely, that a particle could move more than $\frac{\pi}{2}$ radians (90°) around the leaflet COM between two frames. Since the sine function is symmetrical about $\frac{\pi}{2}$ radians, in practice the scalar product is used to calculate θ. This expression produces unambiguous results for $0 < \theta < \pi$, while the vector product was used to determine the axis of the rotation vector (Note that θ is always assumed to be positive; the direction of rotation is determined by the direction of the unit vector $\vec{n}$ which defines the axis of rotation).

The true displacement of each particle is its position minus the rigid body rotation of the whole leaflet. Knowing the effective rotation axis of the leaflet and the velocity at time, t, the particle positions at time, $t + \Delta t$, can be corrected by rotating the particles around the leaflet axis of rotation by an amount equivalent to $-\Delta t \omega_v$. The new position, $r'_{i,t+\Delta t}$, of a particle can be calculated using the Rodrigues rotation formula, Equation 30.

$$r'_{i,t+\Delta t} =$$

$$\cos(\Delta t|\omega_v|)\, r_{i,t+\Delta t} +$$

$$\sin(-\Delta t|\omega_v|)\left(\frac{\omega_v}{|\omega_v|}\times r_{t+\Delta t}\right) +$$

$$(1-\cos(\Delta t|\omega_v|))\frac{\omega_v}{|\omega_v|}\left(\frac{\omega_v}{|\omega_v|}\cdot r_{t+\Delta t}\right)$$

Equation 30

All the expressions presented here assume that the leaflet COM is centred on the origin. In practice, this requires that the particle be first repositioned by subtraction of the vector for the leaflet COM. The calculations are then performed, and the new coordinates can be obtained by adding the vector for the leaflet COM.

6.3.4 Using GROMACS to Calculate the Mean Squared Displacement of Particles in a Vesicle

Within the GROMACS molecular dynamics package contains a program called g_msd, which can be used to calculate the diffusion coefficient for a group of molecules. In order to calculate the lateral diffusion of particles on a spherical surface, the program needs to be modified to include additional subroutines. One of the new subroutines calculates the average angular displacement of the particles in the vesicle between two frames (Appendix Figure 50A), another calculates the vesicle's COM (Appendix Figure 50B), and one more subtracts the average angular displacement from every particle's total displacement (Appendix Figure 50C).

The g_msd program calculates the diffusion coefficient by fitting a straight line to the mean squared displacement against time. Averaging is performed over multiple molecules and start positions to reduce the statistical error. The program loops over all frames in the trajectory, and stores the coordinates for each of the new start positions. For each frame in the trajectory, the displacement of each particle from each starting position is calculated. Therefore, the angular displacement from each of the start positions is

calculated separately for the particle positions in the frame currently being read into the program. The angular displacement is then subtracted from the particle positions in the current frame for each of the start positions, yielding several separate sets of coordinates for the current frame, one for each start position. This removes the extra displacement that results from the vesicles angular velocity between each start position and the current frame.

Rather than calculate the radius for each leaflet separately, the average radius of a particles position at t and $t + \Delta t$ is used to calculate the arc length between the two points. This is, perhaps, more realistic as each particle will have its preferred position within the bilayer, a spherical surface with its own radius.

6.3.5 Lipid Direction Correlation

The lateral direction in which lipids travel can show a distance dependent correlation with other nearby lipids. The direction vector in this case is considered to be the unitised displacement vector between a lipid's position at t, and, $t + \Delta t$. The correlation between two direction vectors can be measured using the scalar product, Equation 31. The angle between two unit vectors pointing in same direction is 0 rad, and thus, $\cos 0 = 1$, while the angle between two vectors that point in opposite directions is π rad, and $\cos \pi = -1$.

$$\cos \theta = \frac{\vec{a} \cdot \vec{b}}{|a||b|}$$

Equation 31

In planar bilayer simulations, only the angle between the lateral components of the displacement vector, typically the Cartesian x and y dimensions, is calculated. For similar reasons as those stated for the calculation of the lateral diffusion coefficient, it is not acceptable simply to incorporate the third Cartesian dimension into the calculation of correlation between the direction vectors of lipids in a vesicle. Indeed, the curvature of the vesicle would introduce errors. For example, two lipids travelling in opposite directions would have to follow the curvature of the vesicle, and the angle between their direction vectors could never equal π rad, except in one unusual scenario: if the lipids were at antipodal points at time, t, and had swapped positions at time, $t + \Delta t$.

It could be supposed that the calculation of lateral diffusion in a planar bilayer is a special case of lateral diffusion on a sphere with an infinite radius. Indeed, a small patch on a large sphere would show little curvature, and would appear approximately flat. The displacement of a particle in the patch could be calculated with $s = r\theta$, where r, the radius, would be large and θ, the angle, would be small. If the size of a particles displacement is kept constant and the radius of a sphere is increased then the angle rotated becomes increasingly small, and the length of the arc and the linear distance approach equivalency.

In the calculation of lateral diffusion coefficients of lipids in a vesicle, the displacements of particles on the surface of a sphere were considered to be great arcs of a great circle. In this case, a particles direction of movement on the vesicle surface is defined by the great circle. Two particles that are traveling in opposite directions would move around the same great circle, one turning clockwise and the other counter clockwise. Conversely, two particles that are travelling in the same direction will move around the great circle with the same rotation. Therefore, it could be proposed that if two lipids are moving in the same direction then they are also rotating about the same axis of a great circle. And, if two lipids are moving in opposite directions then the axes of their great circles will point in opposite directions to one another. Thus, it seems reasonable that the correlation of lipids in a vesicle can be calculated from their rotation axis vectors that describe great circles. Indeed, the angle between two great circles at the surface of a sphere, where they intersect, is equivalent to the angle between their axes of rotation.

The rotation axis for a great circle between a molecules COM at t, and, $t + \Delta t$ is calculated using the vector product relation, where $\vec{a}$ and $\vec{b}$ are the two vectors representing the position of a particle at t and $t + \Delta t$, θ is the angle of rotation, and $\vec{n}$ is the unit vector parallel to the axis of rotation.

$$\frac{\vec{a} \times \vec{b}}{|a||b| \sin \theta} = \vec{n}$$

Equation 32

The direction correlation between two molecules, i and k, is then the dot product between their two axes of rotation, represented by unit vectors $\vec{n}_i$ and $\vec{n}_k$.

$$\cos \theta = \vec{n}_i \cdot \vec{n}_k$$

Equation 33

The distance dependent direction correlation is a subtle phenomenon, and it would not be observed with a single observation. The true influence of correlated motion on lipid diffusion can only be observed by averaging over many observations. Thus, the direction correlation between all pairs of molecules in all frames is calculated. Each value for the direction correlation is binned based on the great arc distance between the molecules. The average of the values in each bin is calculated.

6.3.6 Visualizing Correlated Motion

On a vesicle, the lipids are considered to move in arcs of a great circle across the surface of a sphere, as described earlier. In order to visualise this motion over different time periods, a number of lipids in the vesicle must first be grouped together, so that an average of their displacement on the surface of the sphere can be calculated. The lipids can be grouped together based on their closest vertex of a tessellated sphere. Evenly spaced vertices are obtained by tessellating the sphere with icosahedrons. The direction of motion of the group of lipids is then displayed as an arrow, which is tangential to the surface of the sphere. The arrow originates from the vertex of the group and it points in the direction of the great circle that describes the irrotational motion of the group. The size of the arrow is varied according to the magnitude of the angle, θ, of the arc of the great circle, as described previously. The angular velocity of the entire vesicle was removed in each calculation.

6.3.7 Membrane Thickness

The thickness of a planar membrane can be estimated by calculating the average height of a particle, typically the headgroup phosphate particle, with respect to an axis normal to the plane of the bilayer. In a vesicle, there is no fixed axis with which to compare the height of particles. Indeed, the vector normal to a spherical surface is different at each point on the sphere. The axis normal to the bilayer plane at any point within spherical vesicle will pass through the COM of the vesicle. The thickness of the bilayer can be determined, thus, by the difference in the radii of particles from the COM of the vesicle. This can be calculated separately for each leaflet.

The measurement of the bilayer thickness around a protein requires that the protein's irrotational and rotational motion about the vesicles surface normal be removed between frames of a trajectory. A square grid is then projected onto the surface of the vesicle, with the protein at the centre of the grid. The lipids closest to each vertex are selected, and the distances of the headgroup particles from the vesicle COM are summed. The average is taken over all observations in all frames in a 1.1 μs trajectory, with a 2 ns time interval between fames.

6.4 Results

6.4.1 The Spherical Surface for Lipid Displacements in Vesicles

The diffusion coefficients of lipids on a vesicle are calculated based on the assumption that the vesicles remain approximately spherical throughout the simulations. The RMSD can be used to monitor a vesicles average distance from an ideal sphere. In each frame, the average radius of every particle is determined from the vesicle COM. The RMSD is calculated using the mass weighted squared difference of each lipid COM from the ideal leaflet radius.

The RMSD of both the inner and outer leaflets of NoPro_OMV reach a plateau after 15 ns at approximately 0.30 nm (Figure 42A). The RMSD from an ideal sphere remains stable at 0.28 nm for NoPro_DPPC and for NoPro_POPE_POPG. The values for NoPro_DPPC and NoPro_POPE_POPG are similar to NoPro_OMV, suggesting that the complex lipid mixtures do not cause an increase in the deviation from a spherical surface for lateral diffusion.

There is some variability in the RMSD at the start of the NoPro_OMV that is absent from NoPro_DPPC and NoPro_POPE_POPG. The initial variability in NoPro_OMV could result from differences in the methods to generate the initial coordinates, but it may also result from positional rearrangements in the complex lipid mixtures to satisfy area constraints. Indeed, the OMVs are often observed to contain water pores at the start of the simulations, which can allow a few lipids to flip between leaflets. The water pores are typically closed after 10 ns of simulation. After 200 ns of simulation time, only one DLPC molecule was found to be in the inner leaflet of NoPro_OMV, and no lipids that are normally found in the inner leaflet were found to be in the outer leaflet.

Proteins have been shown to distort the shape of vesicles in simulation studies. Similarly, the distance of lipids from an ideal spherical surface appears to be affected at high protein concentrations in the OMVs. Indeed, the RMSD from an ideal sphere is 0.30 nm in both leaflets of 8_EspP_OMV, similar to that of NoPro_OMV, while the RMSD plateaus higher in 32_Mix_OMV, at 0.43 nm (Figure 42A). Similarly, the RMSD plateaus at 0.37 nm in 32_OmpF_OMV, slightly higher than NoPro_OMV. Thus, the type of protein, not just total protein, in the OMVs may have an impact the shape of the vesicle.

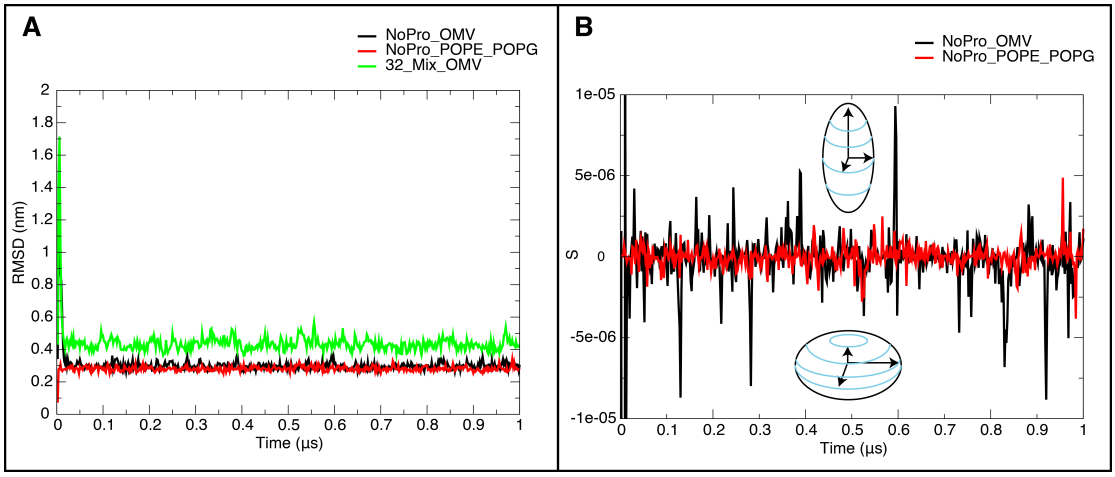


Figure 42. A) The RMSD of lipid positions from an ideal sphere over a 1 μs trajectory. Only the outer leaflet values are displayed in the figure, although the inner leaflet lipids show similar trends. B) The shape of the vesicles over a 1 μs trajectory. Values below zero correspond to oblate spheroids, while positive values correspond to prolate spheroids.

6.4.2 The Shape of the Vesicles

Although the RMSD describes how far, on average, a lipid is from the ideal radius of the vesicle, it does not provide information about the shape changes occurring within the vesicle. Indeed, a measure of asphericity, originally developed to assess the shape of proteins (*253*), can be used to examine how a vesicle's shape fluctuates about its three principle axes. The method compares the eigenvalues of three principal radii of gyration to assess how far atomic coordinates are from an ideal sphere. The parameter calculated, Δ, has values in the range $0 \leq \Delta \leq 1$. A value for $\Delta > 0$ represents deviation from a sphere, while the shape of the vesicle is perfectly spherical if $\Delta = 0$. Membrane proteins in the vesicles are likely to distort this measure in an unpredictable way, so only a direct comparison of NoPro_DPPC, NoPro_POPE_POPG, and NoPro_OMV can be performed.

In general, the value for Δ was close to zero, $\Delta < 1 \times 10^{-4}$, for each of the NoPro_DPPC, NoPro_POPE_POPG and NoPro_OMV simulations, suggesting that the vesicles have a close to spherical shape. As with the measure of RMSD from an ideal sphere, the Δ parameter shows that NoPro_OMV deviates from an ideal sphere, $\Delta = 1.06 \times 10^{-2}$, at 6 ns into the simulation, but then returns to a more spherical shape after 15 ns, $\Delta = 5.78 \times 10^{-5}$.

Still using the three principal radii of gyration, another shape parameter, S, can provide details about the approximate shape of the vesicle. The parameter, S, has values in the range, $(-1/4) \leq S \leq 2$. Positive values suggest prolate shapes, while negative values correspond to oblate shapes. Similarly to Δ, the vesicle is exactly spherical if $S = 0$.

The value of S remains close to zero in all the simulation trajectories of NoPro_DPPC, NoPro_POPE_POPG, and NoPro_OMV. However, there is a small amount of fluctuation around $S = 0$, with a high value 9.2×10^{-6} and low value -8.86×10^{-6} in NoPro_OMV, which suggest some oscillation between prolate and oblate spheroids. The oscillation was most pronounced in NoPro_OMV, which had larger extreme values than either NoPro_DPPC or NoPro_POPE_POPG (Figure 42B).

6.4.3 Diffusion Rate of Lipids in NoPro_DPPC and NoPro_POPE_POPG

The NoPro_DPPC vesicle was simulated at 323 K (50 $°C$), above the phase transition boundary, which is at approximately 314 K (41 $°C$). The lateral diffusion coefficient of all DPPC lipids in NoPro_DPPC is $13.12\ (\pm\ 0.62) \times 10^{-7}\ cm^2 s^{-1}$. Interestingly, a small difference between leaflets was observed for the rate of DPPC diffusion, with inner leaflet lipids appearing to diffuse more slowly than those in the outer leaflet. The lateral diffusion coefficient was $13.51\ (\pm\ 0.73) \times 10^{-7}\ cm^2 s^{-1}$ for the outer leaflet lipids and $12.65\ (\pm\ 0.49) \times 10^{-7}\ cm^2 s^{-1}$ for the inner leaflet. With a factor 4 scaling for time, as has been used in other studies (*134*), the diffusion coefficients are 3.16 and $3.38 \times 10^{-7}\ cm^2 s^{-1}$. These values are similar to those reported in previous simulation studies. In planar bilayers the diffusion coefficient for DPPC, using the MARTINI force field, is reported to be $3.00 \times 10^{-7}\ cm^2 s^{-1}$ (with the simulation time multiplied by a factor 4) at 323 K (*134*) . In vesicles, the reported lateral diffusion coefficients for DPPC is different for each leaflet, with a value of $2.50 \times 10^{-7}\ cm^2 s^{-1}$ for the outer leaflet and $4.80 \times 10^{-7}\ cm^2 s^{-1}$ for the inner leaflet, both recorded at 323 K and using the scaling factor of 4 for time (*241*). Experimental studies of the lateral diffusion coefficient of DPPC provide estimates of $1.50 \times 10^{-7}\ cm^2 s^{-1}$ (*254*), using fluorescence quenching at 325 K, and between 0.99 and $1.20 \times 10^{-7}\ cm^2 s^{-1}$ at 321 K (*255*).

In NoPro_POPE_POPG, the lateral diffusion coefficient for all lipids in the vesicle is $9.40\ (\pm\ 0.05) \times 10^{-7}\ cm^2 s^{-1}$. Little difference was observed in the diffusion coefficients between leaflets, though inner leaflet diffusion appears to be slower, as was observed in NoPro_DPPC. The diffusion coefficient for the outer leaflet is $9.63\ (\pm\ 0.05) \times 10^{-7}\ cm^2 s^{-1}$and the inner leaflet is $9.11\ (\pm\ 0.18) \times 10^{-7}\ cm^2 s^{-1}$. These values suggest a faster rate of diffusion compared to the reported rates for planar membranes of similar composition, where $D_L = 8.50 \times 10^{-7}\ cm^2 s^{-1}$ was calculated for long timescales at 313 K (*102*). It has been reported that lipid lateral diffusion is, in general, slower in vesicles (*244*). The discrepancy here may be attributed to anomalous diffusion at short timescales and long time correlations between restarts, both of which were accounted for in (*102*). Other factors may also be involved, such as the area per lipid or the difference in the size of systems (*256*).

6.4.4 Diffusion Rate of Lipids in the OMVs

The diffusion coefficients, along with error estimates, for each lipid type in NoPro_OMV are displayed in Table 10. A non-zero value for the vesicle's angular velocity is likely to increase the distance travelled for lipids between some time points, and reduce the distance travelled at others. Therefore, it might be expected that the error estimate for diffusion coefficients would be lower when the effect of the vesicle's angular velocity is subtracted from the particle displacements. Indeed, the error estimates are lower for many of the lipids in the vesicle, such as DLPC, DPPE, and PVPE. However, the error values are not consistently lower for all lipid types, and, therefore, it seems unlikely that the angular velocity of the entire vesicle is one of the major sources of statistical inaccuracy.

The rate of DLPC lateral diffusion, $D_L = 18.48\ (\pm\ 0.61) \times 10^{-7} cm^2 s^{-1}$, is substantially faster than the other lipid species in the vesicles. This is perhaps not unexpected due to its short fatty acid tails. The slowest diffusing lipids were the cardiolipins, $D_L = 7.53\ (\pm\ 0.54) \times 10^{-7}\ cm^2 s^{-1}$, presumeably due to their large size. Slower rates of diffusion have been reported for the cardiolipin in simulation studies, with $D_L = 3.00 \times 10^{-7}\ cm^2 s^{-1}$, in planar membranes (*257*). The higher rate of diffusion in NoPro_OMV may be attributed to complex mixture containing cardiolipin lipids with saturated and unsaturated fatty tails.

Table 10. The diffusion coefficients (D_L) of lipids, with error estimates in parentheses[1], from the simulation NoPro_OMV. All values for D_L are provided in $\times 10^{-7}\ cm^2 s^{-1}$.

Lipid type	**Vesicle angular velocity removed**	**Vesicle angular velocity not removed**
DLPC	18.48 (± 0.61)	18.53 (± 0.95)
PQPE	10.11 (± 0.44)	10.13 (± 0.08)
DPPE	10.46 (± 0.02)	10.51 (± 0.45)
PVPE	9.78 (± 0.31)	9.84 (± 0.76)
VQPE	9.69 (± 0.36)	9.73 (± 0.01)
DVPE	10.22 (± 0.55)	10.30 (± 0.11)
PQPG	9.80 (± 0.73)	9.88 (± 1.13)
DPPG	9.33 (± 1.83)	9.34 (± 2.35)
PVPG	10.57 (± 1.46)	10.50 (± 1.07)
VQPG	11.93 (± 1.62)	12.03 (± 1.43)
DVPG	11.76 (± 1.33)	11.66 (± 1.48)
Cardiolipin (all types)	7.53 (± 0.54)	7.56 (± 0.94)

[1] Standard errors are estimated from the difference in the diffusion coefficients over the two halves of the full fit interval.

It has been noted in previous studies that proteins can slow the rate of lipid diffusion (*102, 244*). Similarly, the rate of lipid diffusion is slowed in the OMVs by integral membrane proteins, though the diffusion rate was largely unaffected if the protein concentration in the OMV was low. The protein content of the OMV simulations ranges from 0 to 15% of the total mass. All protein containing OMVs had lower values than NoPro_OMV for the diffusion coefficients of lipids, though the differences were often marginal for OMVs with only 2 proteins. The diffusion coefficients of the lipids in the OMV show moderate correlation to the fractional protein content of the OMVs, with a R^2 value of 0.62. This value must be treated with caution, however, as there is a considerable difference between the OMV with the highest protein content, 32_Mix_OMV at 15% of the vesicle's total mass, and second highest, 8_Mix_OMV at 4 % of the vesicle's total mass. All the other OMV simulations have a protein content of less than 3.5 % of the total mass of the vesicle. The R^2 value is 0.13 for the subset of the dataset that does not contain the diffusion coefficients from 32_Mix_OMV, suggesting that little or no correlation exists at low protein concentrations. The lateral diffusion coefficient of all lipids combined was $13.60\ (\pm 0.24) \times 10^{-7}\ cm^2 s^{-1}$ in NoPro_OMV. This compares to $D_L = 10.29\ (\pm 0.28) \times 10^{-7}\ cm^2 s^{-1}$ in 32_Mix_OMV and $D_L = 11.10\ (\pm 0.36) \times 10^{-7}\ cm^2 s^{-1}$ in 8_Mix_OMV.

In order to investigate whether some lipids are affected more than others by the protein content of the vesicle, the diffusion coefficient was calculated for subsets of phospholipids with the same headgroup particles, i.e. PE, PG and PC, from both NoPro_OMV and 32_Mix_OMV. The high protein content of 32_Mix_OMV affected the PE and PG phospholipid diffusion equivalently, with the diffusion coefficient of both being $3.00 \times 10^{-7}\ cm^2 s^{-1}$ lower in 32_Mix_OMV than in NoPro_OMV. In the outer leaflet, the diffusion coefficient of PC lipids was $4.50 \times 10^{-7}\ cm^2 s^{-1}$ lower in 32_Mix_OMV than in NoPro_OMV.

6.4.5 Correlated Motion in Vesicles

The lateral diffusion of lipids is driven by a concerted mechanism, involving large groups of lipids. Two adjacent lipids are more likely than random chance to have moved in the same or similar direction after a small change in time, between 0 to 20 ns. The average great circle displacement of lipids after different time periods are shown for NoPro_OMV in Figure 44. The motions of lipids appear to be correlated over distances greater than 10 nm, and over time periods of greater than 10 ns. There also appears to be large correlated motions between leaflets (Figure 44). Proteins can also be observed to move in the same direction as the bulk lipids (Figure 43).

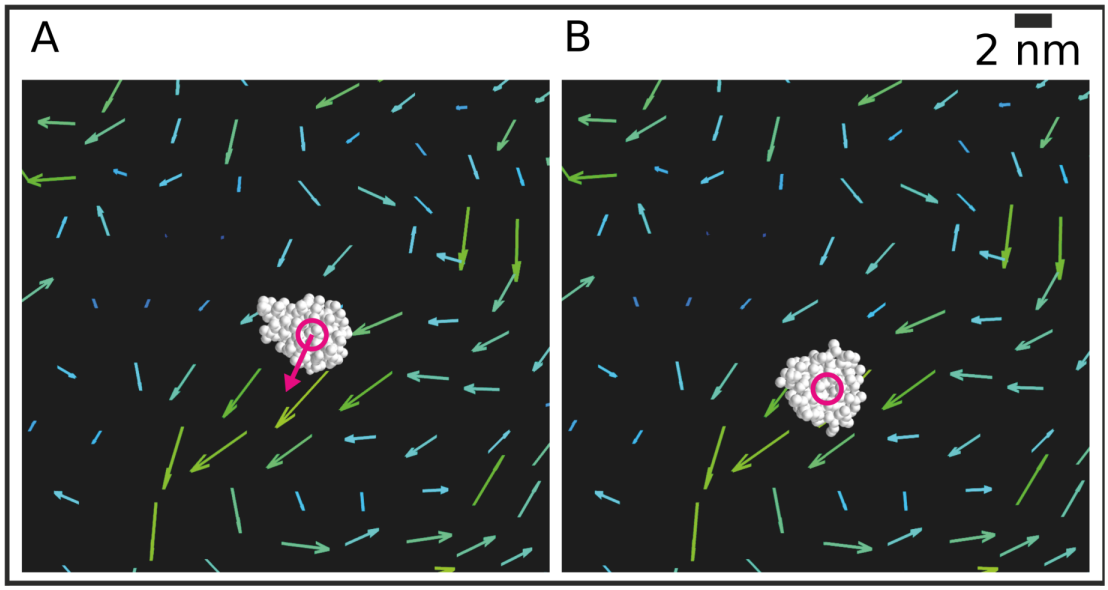


Figure 43. The displacement of OmpA (white) after 10 ns. The direction of lipid displacements after 10 ns are displayed in both panels, A and B, as blue and green arrows.

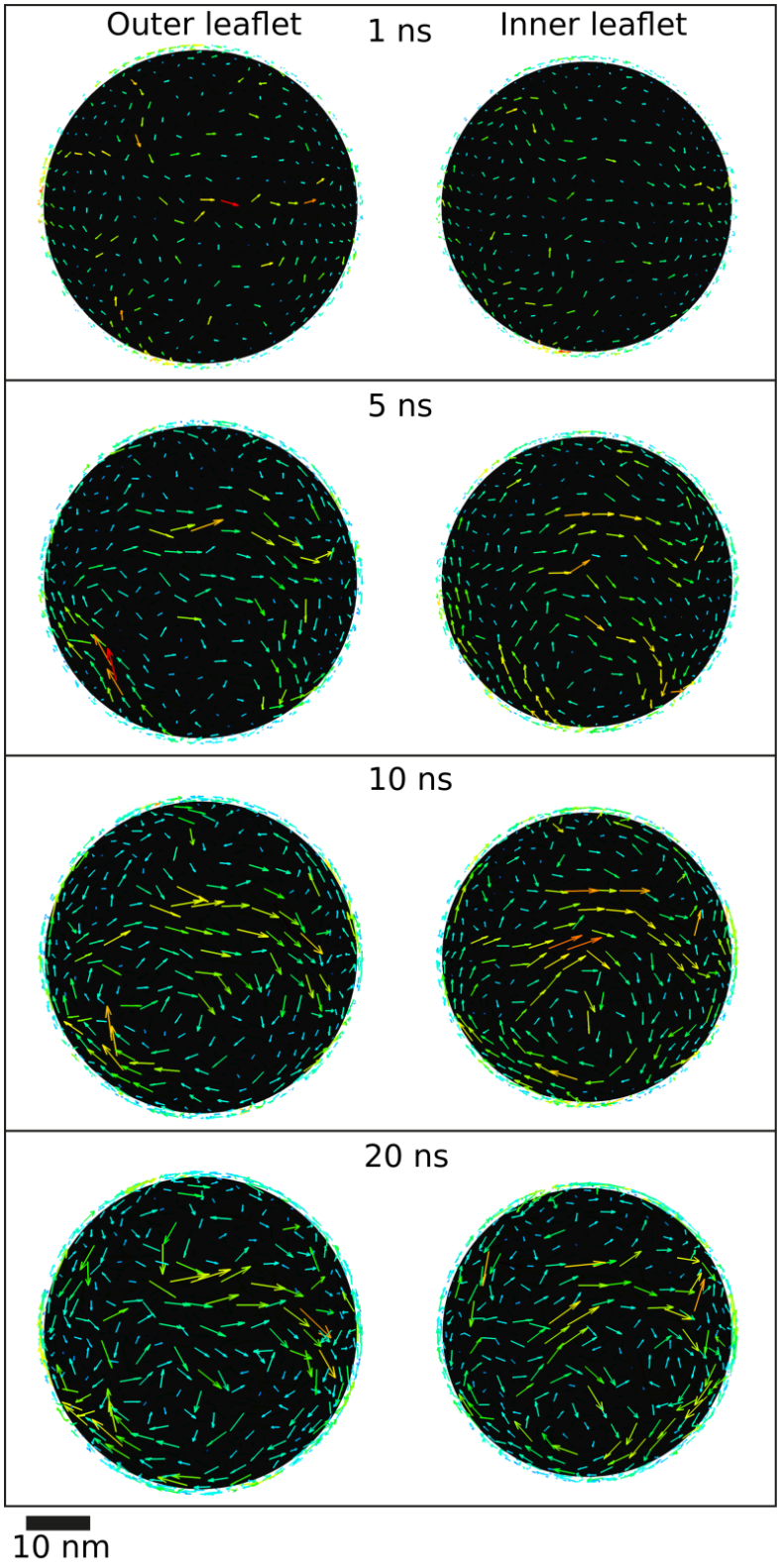


Figure 44. The average great circle displacement of groups of lipids in NoPro_OMV after different time periods. The arrows show the direction of the displacement and are projected onto the surface of an ideal sphere. The arrows are coloured blue (smallest displacement) to red (largest displacement).

A more quantitative assessment of the lipid-lipid direction correlations shows that their great circle axes are correlated for up to 20 nm (Figure 45E and Figure 45F), and for times up to 16 ns. A similar degree of correlation, between 0.30 and 0.40, at short distances is observed for inter and intra leaflet lipids.

There is a greater statistical uncertainty with the protein-lipid direction correlation data compared to the lipid-lipid direction correlation, due to a small number of proteins in the vesicles. However, it is clear that close range protein-lipid interactions result in a very high correlation, > 0.75, in the direction of protein and lipid displacements (Figure 45A to Figure 45D). Similarly to the lipid-lipid correlations, there is a positive direction correlation between proteins and lipids up to 20 nm, beyond which there appears to be no correlation between the great circle axes. In addition, a high level of direction correlation of proteins and lipids existed up to 10 ns time periods. The direction correlations may have been present for longer time periods, though there was insufficient data to perform the analyses for a time period greater 10 ns. Interestingly, the protein direction vectors showed marginally less correlation with the lipids for 0.2 ns time periods than for longer time periods.

Of note, the close range direction correlation of the molecules in the vesicles means that some anticorrelated regions must exist elsewhere in the vesicle, as there can be no overall angular velocity. Indeed, anticorrelation can be observed at distances between 20 nm and 50 nm, where the plots in Figure 45 are all a fraction below zero. Whereas positive correlation is a large effect on a small number of molecules, the negative correlation is a small effect on a large number of molecules.

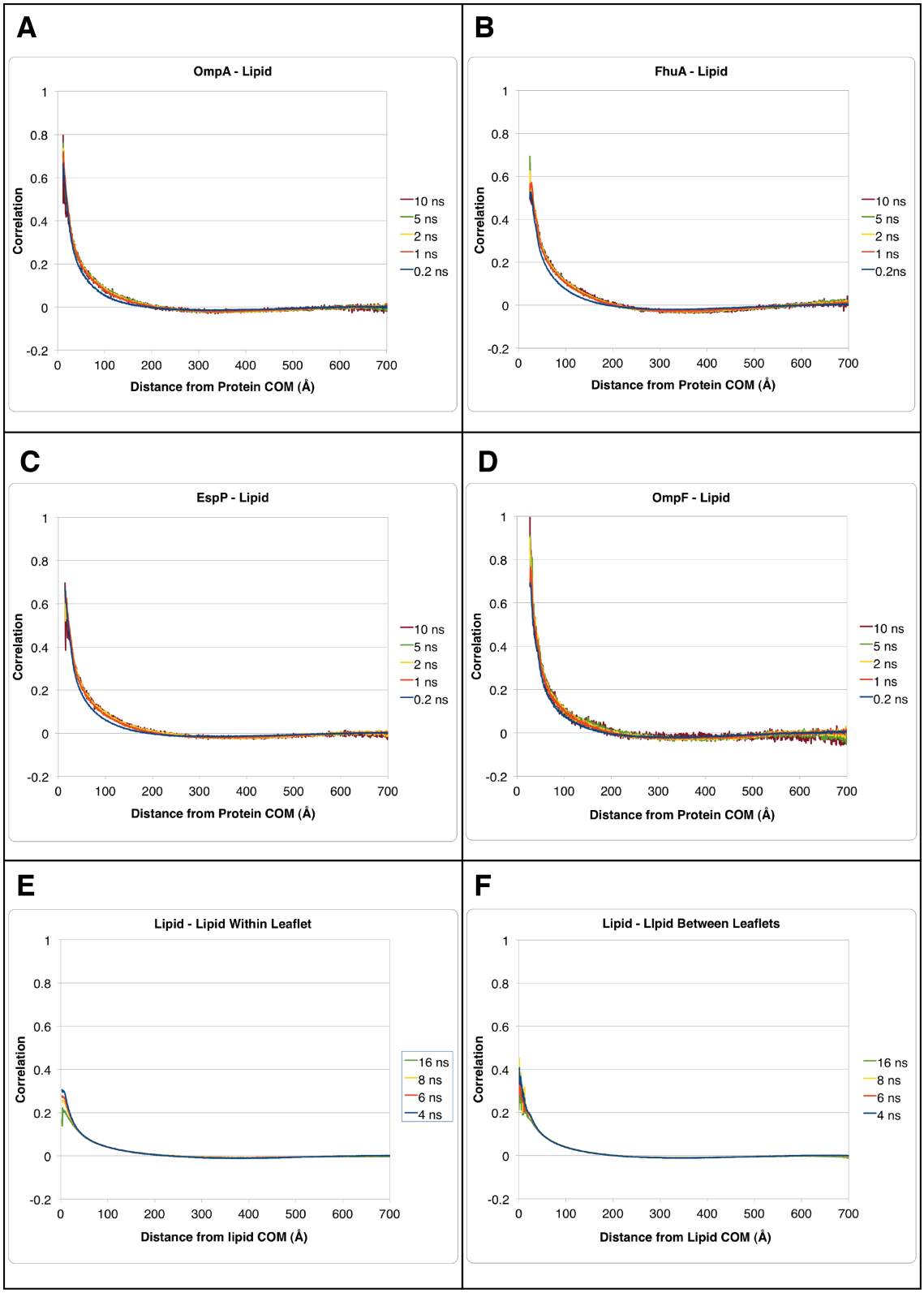


Figure 45. The direction correlation of A) OmpA, B) FhuA, C) EspP and D) OmpF with lipids is plotted as a function of the lipids distance from the protein COM. The direction correlation of E) lipids within the same leaflet and F) lipids between leaflets are plotted as function of the distance between lipids.

6.4.6 Membrane Thickness Measurements

Proteins are able to distort the shape of phospholipid bilayers, altering their curvature and thickness. The effect of the proteins on the OMV bilayer was examined by measuring the phosphate distance of the lipids from the average radius of the vesicle. The average height of the phosphate molecules of each leaflet of the OMVs around OmpA and EspP are displayed in Figure 46. Compared to OmpF and FhuA, the integral membrane domains of OmpA and EspP are relatively small, at 19 kDa (β-barrel) and 30 kDa (β-barrel), respectively. Both proteins have little impact on the relative height of the lipid phosphate moieties within each leaflet of the OMVs. Of note, EspP, for its size, appeared to thin a large area of the outer leaflet, with the lowest point being 0.5 nm below the bulk average phosphate height. A similar level of thinning is seen for OmpA, but over a smaller area.

The average height of the phosphate molecules in each leaflet around FhuA and OmpF are displayed in Figure 47. FhuA appeared to cause very little distortion of either the outer or inner leaflets. The most notable changes in the height of the lipid phosphates were caused by the OmpF trimer. Interestingly, the inner leaflet phosphates were raised, by 0.2 nm, above the bulk height near the interface of the three proteins in the trimer. In addition, OmpF caused a substantial thinning of the inner leaflet lipids at the points on the protein that are furthest away from the trimer interface. At these points, the phosphates were lower by 0.7 nm than the average bulk height.

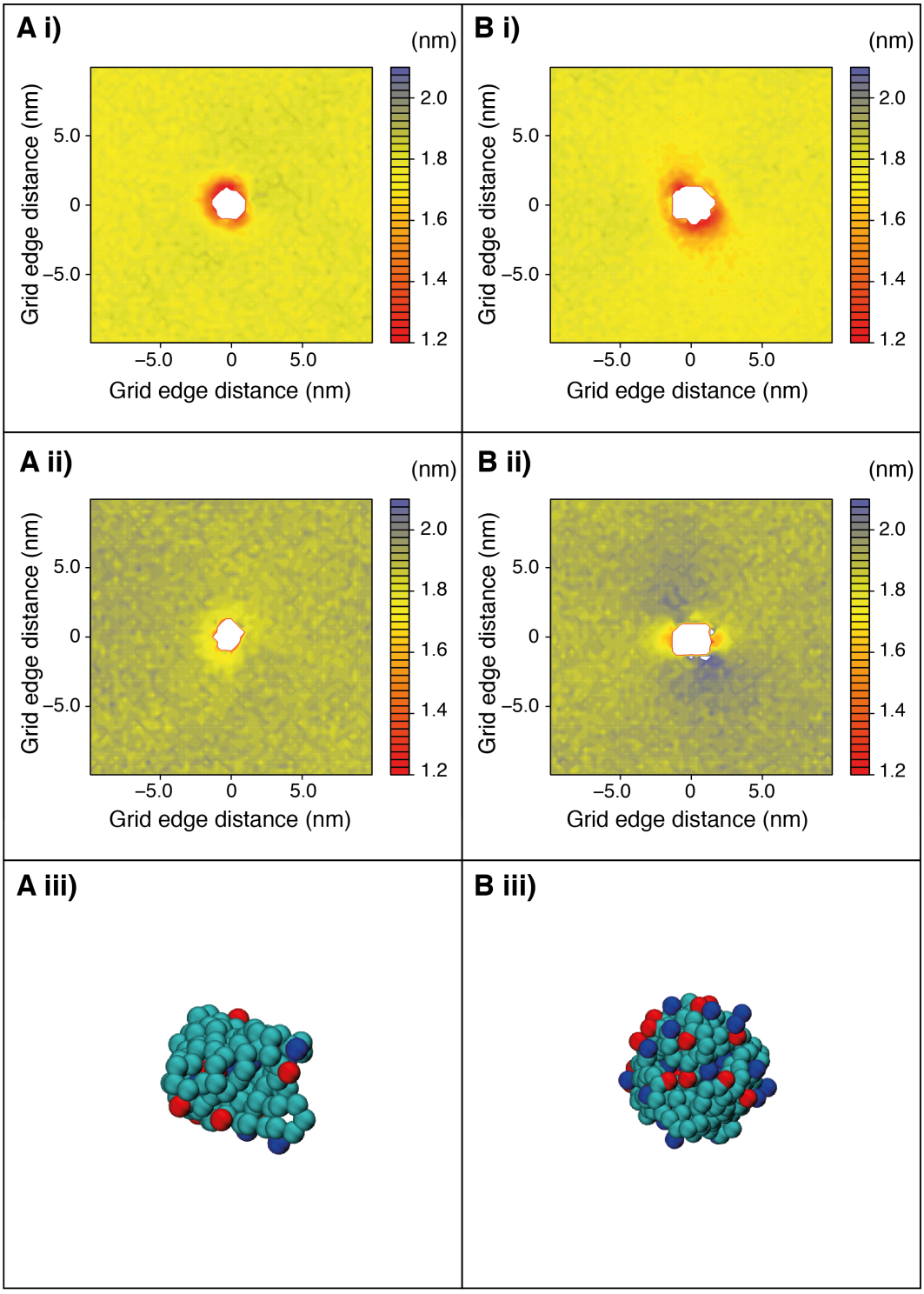


Figure 46. The phosphate height of lipids from the centre of the bilayer around A) OmpA and B) EspP. The outer leaflet phosphate heights are shown in (i), and the inner leaflet phosphates are shown in (ii). In (iii), the positions of the Cα particles from both proteins are shown in the same orientation as the plots. The position of charge residues, -ve in red and +ve in blue, are also shown.

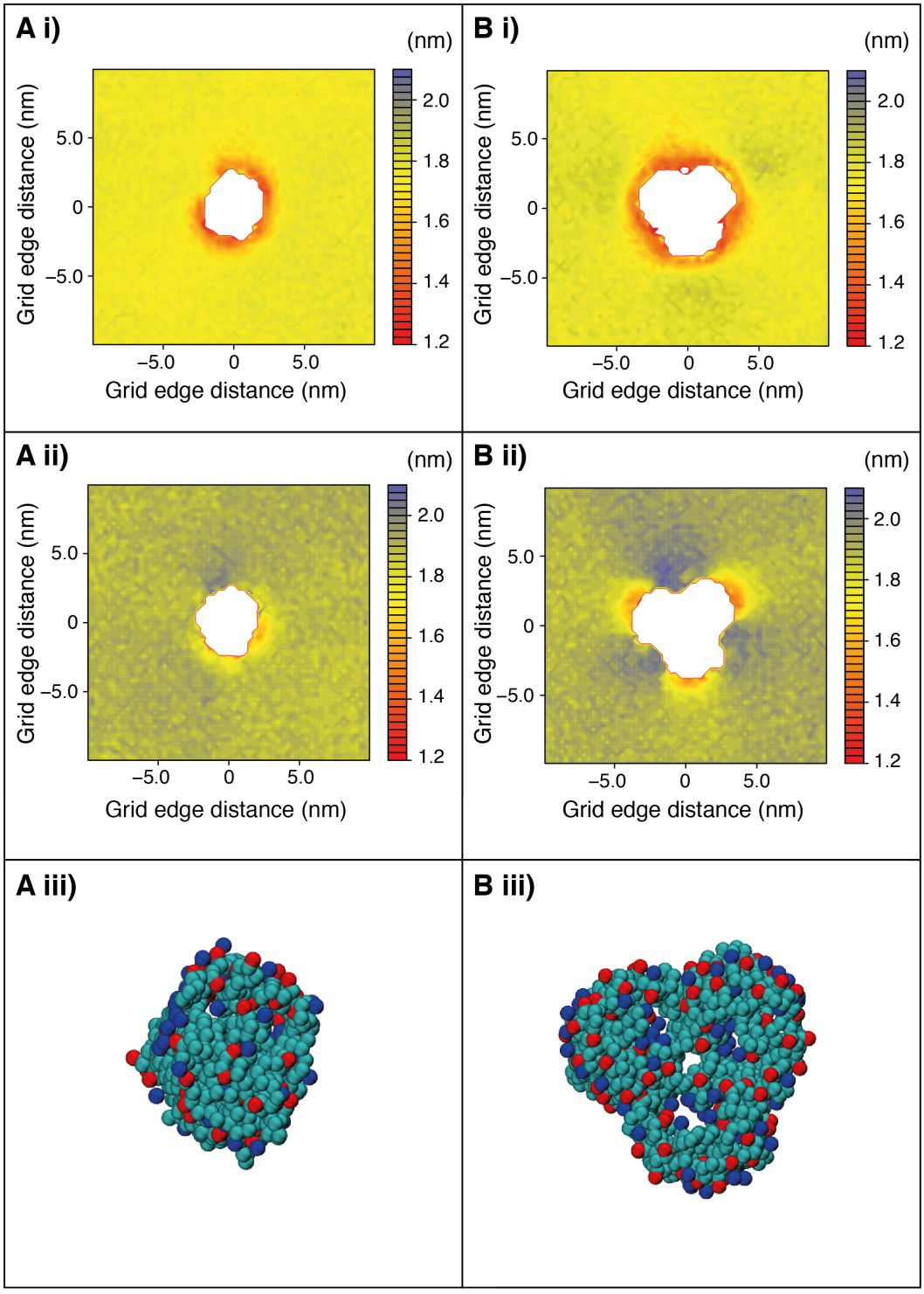


Figure 47. The phosphate height of lipids from the centre of the bilayer around A) FhuA and B) OmpF. The outer leaflet phosphate heights are shown in (i), and the inner leaflet phosphates are shown in (ii). In (iii), the positions of the Cα particles from both proteins are shown in the same orientation as the plots. The position of charge residues, negative in red and positive in blue, are also shown.

6.4.7 Protein-Protein Interactions

The OmpF trimer is known to aggregate in two orientations, termed tip-tip and base-to-base (*101*). No base-to-base aggregation was observed in any of the OMV simulations. Three events of tip-to-tip aggregation were observed in 32_OmpF_OMV, and one event was also observed in 32_Mix_OMV. The tip-to-tip aggregation would seem to occur at the regions OmpF that thin the inner leaflet of the OMVs, while base to base aggregation would occur at regions around OmpF that appear to thicken the inner leaflet.

The aggregation of two OmpF trimers was examined over a 60 ns period (Figure 48 and Figure 49). A degree of inner leaflet thinning was observed at the interface between the two trimers of OmpF (Figure 48). The phosphates in the areas of thinning were approximately between 0.3 nm and 0.4 nm below the bulk average. Interestingly, a large thickened region, with the phosphates about 0.5 nm higher than the bulk, was observed at 30 ns, presumably hindering base-to-base aggregation. At 40 ns the protein is seen to reorientate, and the thickened region disappears. The thickened region at 30 ns may result from a stream of lipids colliding with the proteins (Figure 49). The lipids appear to dissipate at 40 ns.

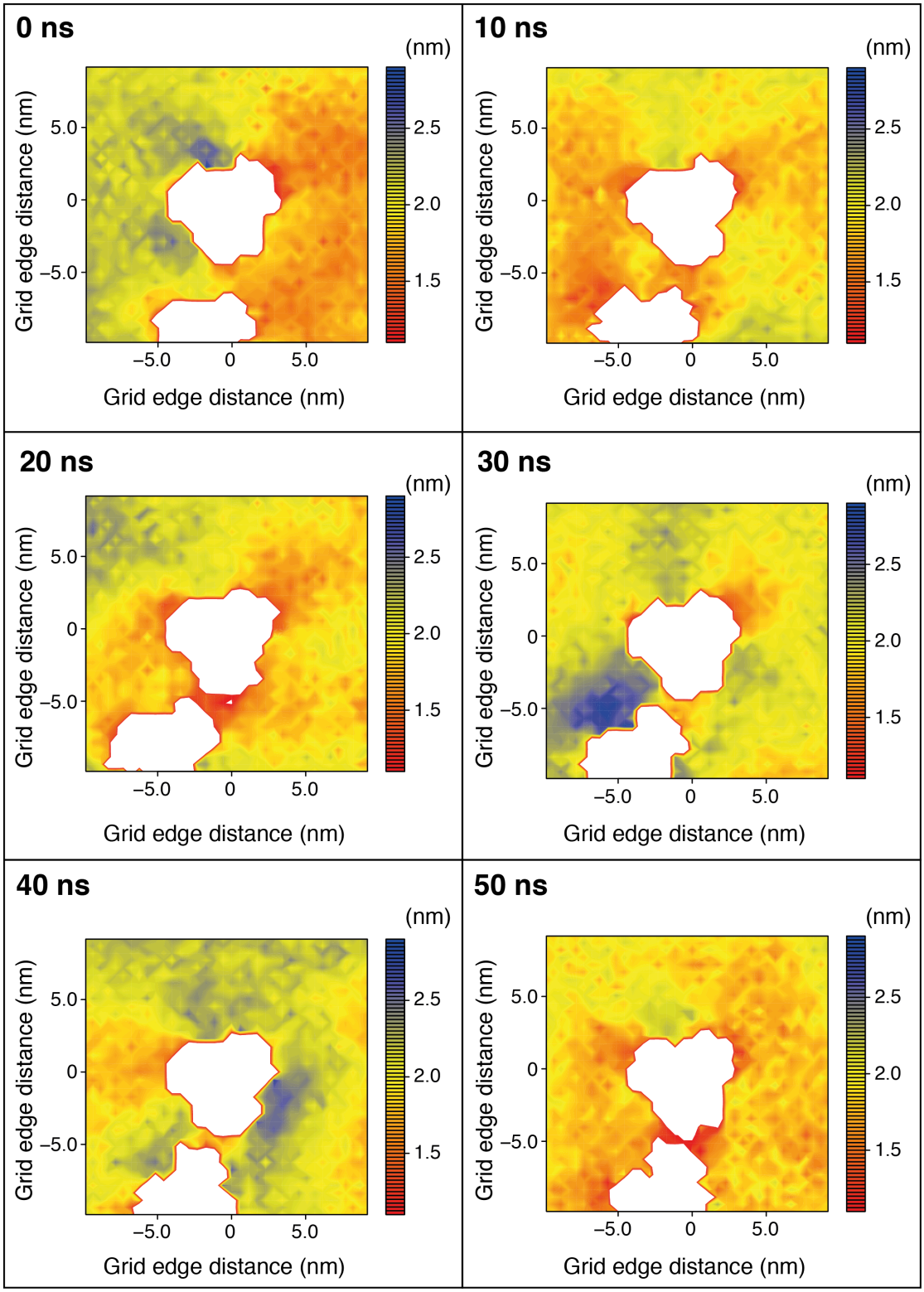


Figure 48. The dimerization of two OmpF trimers in the tip-to-tip orientation over 50ns. The heights of the phosphates from the centre of the bilayer are displayed in each of the panels. The plots were produced by averaging frames over 5 ns, spaced at 0.2 ns, from the start time indicated in the panel.

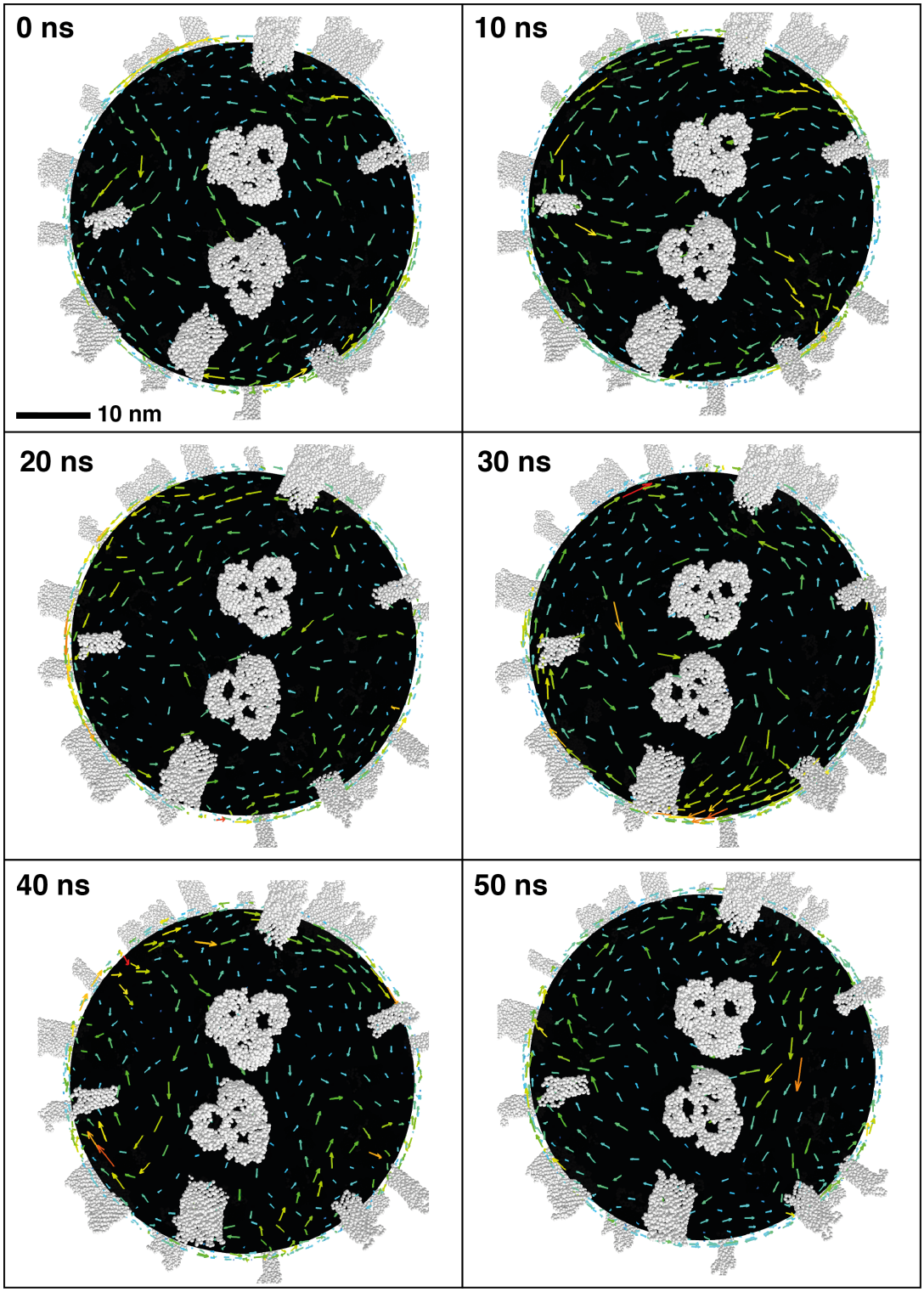


Figure 49. The dimerization of two OmpF trimers in the tip-to-tip orientation over 50ns. The direction of lipid displacements of inner leaflet lipids over 10 ns are displayed as arrows, coloured blue, for no displacement, to red, for the largest displacement. The final position of the proteins (in white) after the lipid displacements is also shown. The start time of the lipid displacements is stated for each panel.

6.5 Discussion

Removing the angular velocity, as a rigid body rotation, from the vesicles had little effect on the calculation of lateral diffusion coefficients on the vesicles surface. It is perhaps expected that the vesicle should not spin around an axis during the simulation, since the friction with the vesicle's periodic image would tend to prevent such an event from occurring.

Contrary to previous studies, the rate of lipid diffusion has been found to be similar between the two leaflets of a vesicle with symmetrical leaflet compositions, though the diffusion of inner leaflet lipids was found to be marginally slower. However, the observed differences may be due to area per lipid (*256*) or the difference in the surface area. The lateral diffusion coefficient calculations of lipids in the vesicles have assumed that the particles travel on great circles, on the surface of a sphere. While the vesicles remain approximately spherical during the simulations, lipids are able to move above and below the surface of the sphere. This is not a unique problem associated with spherical bilayer systems. Indeed, undulations or random movements normal to the plane of planar bilayer systems may also affect the calculations of lateral diffusion coefficients and bilayer thickness.

The OmpF trimer is observed to aggregate in a specific pattern (*101*). In the OMVs the inner leaflet was observed to thin at the tips of the OmpF trimer, and thicken at the base (the interface between the three OmpFs of one trimer). In addition, the inner leaflet membrane was observed to thin at the interface of the tip-to-tip aggregation of two OmpF trimers. Suggesting that membrane deformations may help to drive protein aggregation.

Lipid and protein direction correlation appears to exist at long time periods, greater than 10 ns. Interestingly, the correlation of protein and lipid direction vectors was consistently lower at short time periods, 0.2 ns. This indicates that the diffusion of proteins and lipids may be the result of a predominately long-term concerted process. At close range, there appears to be an annulus of lipids that move almost exclusively in the same direction as the proteins to which they are attached, on the timescale of 0 to 10 ns. This close association of lipids to proteins may be important in the process of protein-protein interaction by mediating bilayer distortions. Indeed, more loosely associated

lipids would likely not translate the effects of bilayer distortions to other lipids around the protein.

The pattern of bilayer distortions caused by proteins would be affected by the curvature stress already present within a bilayer. Indeed, this may account for the differences in protein aggregation between planar membranes and vesicles observed in (*244*). It is proposed that the pattern of lipid deformation around a protein may be involved in the process of protein aggregation within vesicles and bilayers, and differences in the curvature of the membrane may alter the energetics of protein aggregation.

It is important to reiterate that the OMVs simulated here have an unnatural outer leaflet. It is not known to what extent the results would change if the outer leaflet were to contain LPS rather than DLPC. While the DLPC approximates the hydrophobic thickness of the outer leaflet of the OM, it has dissimilar physicochemical properties to LPS, which is a large, slow diffusing, and heterogeneous molecule with unusual phase properties.

Chapter 7: Concluding Remarks

The membranes of cells and cellular organelles are composed of a complex mixture of lipids and protein. The correct sorting and function of membrane proteins requires that they interact extensively with the complex membrane environment. In this thesis, the role of bilayer composition on protein function has been investigated using molecular simulation methods. Simulations of membranes with complex mixtures of lipids have provided a number of interesting outcomes, particularly in the dynamics of the short peptide representing the transmembrane domain of fukutin. In more recent work, this transmembrane domain has been shown to form homodimers in bilayers, driven by interactions of a characteristic TXXSS motif (*154*). Future studies will undoubtedly investigate the effect of the large globular glycosyltransferase domain, of which there is currently no solved structure, on the dimerization and tilt angle of the fukutin transmembrane domain.

The bacterial outer membrane contains an unusual lipid called lipopolysaccharide, which has the potential to modify the behaviour of proteins. Indeed, the presence of lipopolysaccharide in membrane simulations of FecA resulted in substantial modifications of the extracellular loop dynamics. The lipopolysaccharide may also help to anchor some proteins within the outer membrane through electrostatic interactions, such as those observed in Hia. Future studies will characterise the interaction of lipopolysaccharide with the monomeric autotransporters, since, in contrast to Hia, they contain a functional extracellular loop, L4. This loop may show modified dynamics in the presence of lipopolysaccharide. In addition to L4, another loop, L5, has been suggested to occlude the extracellular mouth of the EspP translocator domain after the passenger has been cleaved. Thus, it may be interesting to investigate whether L5 can insert into the translocator in the presence of lipopolysaccharide.

Often it is difficult to calculate a lateral protein diffusion coefficient in small simulations, as there is insufficient data to obtain a good estimate of the mean squared displacement. Therefore, large coarse-grained planar bilayer systems with many proteins can be used to improve the statistics. These systems, however, are affected by the periodic boundaries and pressure coupling of the simulation cell. Indeed, work, currently in progress, is revealing that the size of undulations in planar membranes is subject to the size of the simulation cells, with larger bilayers having larger undulations. This has an impact on the estimate of lateral diffusion coefficients. In contrast, simulations of vesicle systems offer a perspective on protein and lipid diffusion that is unbiased by periodic boundary conditions. The vesicles have been shown to remain predominantly spherical in shape, in μs timescale simulations, as long as they enclose a sufficient amount of water. Future work will investigate whether undulations differ for vesicles with different radii. The early indications are that the distributions of lipids around the mean radius are similar for a range of vesicle radii, with standard deviations much lower than those observed in planar membranes.

Reference list

1. Pilot, J. D., East, J. M., and Lee, A. G. (2001) Effects of bilayer thickness on the activity of diacylglycerol kinase of Escherichia coli, *Biochemistry 40*, 8188-8195.
2. Martinac, B., Adler, J., and Kung, C. (1990) Mechanosensitive ion channels of E. coli activated by amphipaths, *Nature 348*, 261-263.
3. Hopkins, A. L., and Groom, C. R. (2002) The druggable genome, *Nat Rev Drug Discov 1*, 727-730.
4. Russ, A. P., and Lampel, S. (2005) The druggable genome: an update, *Drug Discov Today 10*, 1607-1610.
5. Brown, D. A., and Rose, J. K. (1992) Sorting of GPI-anchored proteins to glycolipid-enriched membrane subdomains during transport to the apical cell surface, *Cell 68*, 533-544.
6. Henderson, I. R., and Nataro, J. P. (2001) Virulence functions of autotransporter proteins, *Infect Immun 69*, 1231-1243.
7. van Meer, G. (1989) Lipid traffic in animal cells, *Annu Rev Cell Biol 5*, 247-275.
8. van Meer, G., Voelker, D. R., and Feigenson, G. W. (2008) Membrane lipids: where they are and how they behave, *Nat Rev Mol Cell Biol 9*, 112-124.
9. Zambrano, F., Fleischer, S., and Fleischer, B. (1975) Lipid composition of the golgi apparatus of rat kidney and liver in comparison with other subcellular organelles, *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism 380*, 357-369.
10. Keenan, T. W., and Morre, D. J. (1970) Phospholipid class and fatty acid composition of golgi apparatus isolated from rat liver and comparison with other cell fractions, *Biochemistry 9*, 19-25.
11. Sebastian, T. T., Baldridge, R. D., Xu, P., and Graham, T. R. (2012) Phospholipid flippases: Building asymmetric membranes and transport vesicles, *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids 1821*, 1068-1077.
12. van Meer, G. (2011) Dynamic transbilayer lipid asymmetry, *Cold Spring Harb Perspect Biol 3*.
13. Daleke, D. L. (2007) Phospholipid flippases, *The Journal of biological chemistry 282*, 821-825.

14. Kornberg, R. D., and McConnell, H. M. (1971) Inside-outside transitions of phospholipids in vesicle membranes, *Biochemistry 10*, 1111-1120.
15. van Meer, G., Poorthuis, B. J., Wirtz, K. W., Op den Kamp, J. A., and van Deenen, L. L. (1980) Transbilayer distribution and mobility of phosphatidylcholine in intact erythrocyte membranes. A study with phosphatidylcholine exchange protein, *Eur J Biochem 103*, 283-288.
16. Fadok, V. A., Voelker, D. R., Campbell, P. A., Cohen, J. J., Bratton, D. L., and Henson, P. M. (1992) Exposure of phosphatidylserine on the surface of apoptotic lymphocytes triggers specific recognition and removal by macrophages, *Journal of immunology 148*, 2207-2216.
17. Vermes, I., Haanen, C., Steffens-Nakken, H., and Reutelingsperger, C. (1995) A novel assay for apoptosis. Flow cytometric detection of phosphatidylserine expression on early apoptotic cells using fluorescein labelled Annexin V, *J Immunol Methods 184*, 39-51.
18. Gidden, J., Denson, J., Liyanage, R., Ivey, D. M., and Lay, J. O. (2009) Lipid compositions in Escherichia coli and Bacillus subtilis during growth as determined by MALDI-TOF and TOF/TOF mass spectrometry, *International Journal of Mass Spectrometry 283*, 178-184.
19. Aibara, S., Kato, M., Ishinaga, M., and Kito, M. (1972) Changes in positional distribution of fatty acids in the phospholipids of Escherichia coli after shift-down in temperature, *Biochimica et biophysica acta 270*, 301-306.
20. Ishinaga, M., Kanamoto, R., and Kito, M. (1979) Distribution of phospholipid molecular species in outer and cytoplasmic membrane of Escherichia coli, *J Biochem 86*, 161-165.
21. Lugtenberg, E. J., and Peters, R. (1976) Distribution of lipids in cytoplasmic and outer membranes of Escherichia coli K12, *Biochimica et biophysica acta 441*, 38-47.
22. Jackson, C. L. (2009) Mechanisms of transport through the Golgi complex, *Journal of cell science 122*, 443-452.
23. Lehle, L., Strahl, S., and Tanner, W. (2006) Protein glycosylation, conserved from yeast to man: a model organism helps elucidate congenital human diseases, *Angew Chem Int Ed Engl 45*, 6802-6818.
24. Rabouille, C., Hui, N., Hunte, F., Kieckbusch, R., Berger, E. G., Warren, G., and Nilsson, T. (1995) Mapping the distribution of Golgi enzymes

involved in the construction of complex oligosaccharides, *Journal of cell science 108 (Pt 4)*, 1617-1627.

25. Rottger, S., White, J., Wandall, H. H., Olivo, J. C., Stark, A., Bennett, E. P., Whitehouse, C., Berger, E. G., Clausen, H., and Nilsson, T. (1998) Localization of three human polypeptide GalNAc-transferases in HeLa cells suggests initiation of O-linked glycosylation throughout the Golgi apparatus, *Journal of cell science 111 (Pt 1)*, 45-60.
26. Munro, S. (1991) Sequences within and adjacent to the transmembrane segment of alpha-2,6-sialyltransferase specify Golgi retention, *The EMBO journal 10*, 3577-3588.
27. Munro, S. (1995) A comparison of the transmembrane domains of Golgi and plasma membrane proteins, *Biochem Soc Trans 23*, 527-530.
28. Tu, L., and Banfield, D. K. (2010) Localization of Golgi-resident glycosyltransferases, *Cell Mol Life Sci 67*, 29-41.
29. Silverman, J. M., Brunet, Y. R., Cascales, E., and Mougous, J. D. (2012) Structure and regulation of the type VI secretion system, *Annu Rev Microbiol 66*, 453-472.
30. Leo, J. C., Grin, I., and Linke, D. (2012) Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane, *Philos Trans R Soc Lond B Biol Sci 367*, 1088-1101.
31. Chatterjee, S., Chaudhury, S., McShan, A. C., Kaur, K., and De Guzman, R. N. (2013) Structure and biophysics of type III secretion in bacteria, *Biochemistry 52*, 2508-2517.
32. Korotkov, K. V., Sandkvist, M., and Hol, W. G. (2012) The type II secretion system: biogenesis, molecular architecture and mechanism, *Nat Rev Microbiol 10*, 336-351.
33. Voth, D. E., Broederdorf, L. J., and Graham, J. G. (2012) Bacterial Type IV secretion systems: versatile virulence machines, *Future Microbiol 7*, 241-257.
34. Omori, K., and Idei, A. (2003) Gram-negative bacterial ATP-binding cassette protein exporter family and diverse secretory proteins, *J Biosci Bioeng 95*, 1-12.
35. Nicolay, T., Lemoine, L., Lievens, E., Balzarini, S., Vanderleyden, J., and Spaepen, S. (2012) Probing the applicability of autotransporter based surface display with the EstA autotransporter of Pseudomonas stutzeri A15, *Microb Cell Fact 11*, 158.

36. Ko, H. J., Park, E., Song, J., Yang, T. H., Lee, H. J., Kim, K. H., and Choi, I. G. (2012) Functional cell surface display and controlled secretion of diverse Agarolytic enzymes by Escherichia coli with a novel ligation-independent cloning vector based on the autotransporter YfaL, *Applied and environmental microbiology 78*, 3051-3058.
37. Dautin, N., and Bernstein, H. D. (2007) Protein secretion in gram-negative bacteria via the autotransporter pathway, *Annu Rev Microbiol 61*, 89-112.
38. Sijbrandi, R., Urbanus, M. L., ten Hagen-Jongman, C. M., Bernstein, H. D., Oudega, B., Otto, B. R., and Luirink, J. (2003) Signal recognition particle (SRP)-mediated targeting and Sec-dependent translocation of an extracellular Escherichia coli protein, *The Journal of biological chemistry 278*, 4654-4659.
39. Szabady, R. L., Peterson, J. H., Skillman, K. M., and Bernstein, H. D. (2005) An unusual signal peptide facilitates late steps in the biogenesis of a bacterial autotransporter, *Proceedings of the National Academy of Sciences of the United States of America 102*, 221-226.
40. Ieva, R., and Bernstein, H. D. (2009) Interaction of an autotransporter passenger domain with BamA during its translocation across the bacterial outer membrane, *Proceedings of the National Academy of Sciences of the United States of America 106*, 19120-19125.
41. Ruiz-Perez, F., Henderson, I. R., Leyton, D. L., Rossiter, A. E., Zhang, Y., and Nataro, J. P. (2009) Roles of periplasmic chaperone proteins in the biogenesis of serine protease autotransporters of Enterobacteriaceae, *Journal of Bacteriology 191*, 6571-6583.
42. Henderson, I. R., Navarro-Garcia, F., Desvaux, M., Fernandez, R. C., and Ala'Aldeen, D. (2004) Type V protein secretion pathway: the autotransporter story, *Microbiol Mol Biol Rev 68*, 692-744.
43. Tajima, N., Kawai, F., Park, S. Y., and Tame, J. R. (2010) A novel intein-like autoproteolytic mechanism in autotransporter proteins, *J Mol Biol 402*, 645-656.
44. Barnard, T. J., Dautin, N., Lukacik, P., Bernstein, H. D., and Buchanan, S. K. (2007) Autotransporter structure reveals intra-barrel cleavage followed by conformational changes, *Nat Struct Mol Biol 14*, 1214-1220.
45. Barnard, T. J., Gumbart, J., Peterson, J. H., Noinaj, N., Easley, N. C., Dautin, N., Kuszak, A. J., Tajkhorshid, E., Bernstein, H. D., and

Buchanan, S. K. (2012) Molecular basis for the activation of a catalytic asparagine residue in a self-cleaving bacterial autotransporter, *Journal of molecular biology 415*, 128-142.

46. Meng, G., Surana, N. K., St Geme, J. W., 3rd, and Waksman, G. (2006) Structure of the outer membrane translocator domain of the Haemophilus influenzae Hia trimeric autotransporter, *The EMBO journal 25*, 2297-2304.
47. van den Berg, B. (2010) Crystal structure of a full-length autotransporter, *Journal of molecular biology 396*, 627-633.
48. Zhai, Y., Zhang, K., Huo, Y., Zhu, Y., Zhou, Q., Lu, J., Black, I., Pang, X., Roszak, Aleksander W., Zhang, X., Isaacs, Neil W., and Sun, F. (2011) Autotransporter passenger domain secretion requires a hydrophobic cavity at the extracellular entrance of the β-domain pore, *Biochemical Journal 435*, 577-587.
49. Oomen, C. J., van Ulsen, P., van Gelder, P., Feijen, M., Tommassen, J., and Gros, P. (2004) Structure of the translocator domain of a bacterial autotransporter, *The EMBO journal 23*, 1257-1266.
50. Veiga, E., Sugawara, E., Nikaido, H., de Lorenzo, V., and Fernandez, L. A. (2002) Export of autotransported proteins proceeds through an oligomeric ring shaped by C-terminal domains, *The EMBO journal 21*, 2122-2131.
51. Hritonenko, V., Kostakioti, M., and Stathopoulos, C. (2006) Quaternary structure of a SPATE autotransporter protein, *Molecular membrane biology 23*, 466-474.
52. Skillman, K. M., Barnard, T. J., Peterson, J. H., Ghirlando, R., and Bernstein, H. D. (2005) Efficient secretion of a folded protein domain by a monomeric bacterial autotransporter, *Molecular Microbiology 58*, 945-958.
53. Pohlner, J., Halter, R., Beyreuther, K., and Meyer, T. F. (1987) Gene structure and extracellular secretion of Neisseria gonorrhoeae IgA protease, *Nature 325*, 458-462.
54. Junker, M., Besingi, R. N., and Clark, P. L. (2009) Vectorial transport and folding of an autotransporter virulence protein during outer membrane secretion, *Molecular Microbiology 71*, 1323-1332.
55. Pavlova, O., Peterson, J. H., Ieva, R., and Bernstein, H. D. (2013) Mechanistic link between beta barrel assembly and the initiation of

autotransporter secretion, *Proceedings of the National Academy of Sciences of the United States of America 110*, E938-947.

56. Faller, R., and Marrink, S. J. (2004) Simulation of domain formation in DLPC-DSPC mixed bilayers, *Langmuir : the ACS journal of surfaces and colloids 20*, 7686-7693.
57. Bennun, S. V., Longo, M., and Faller, R. (2007) Phase and mixing behavior in two-component lipid bilayers: a molecular dynamics study in DLPC/DSPC mixtures, *The journal of physical chemistry. B 111*, 9504-9512.
58. Shi, Q., and Voth, G. A. (2005) Multi-scale modeling of phase separation in mixed lipid bilayers, *Biophysical Journal 89*, 2385-2394.
59. Apajalahti, T., Niemela, P., Govindan, P. N., Miettinen, M. S., Salonen, E., Marrink, S. J., and Vattulainen, I. (2010) Concerted diffusion of lipids in raft-like membranes, *Faraday Discuss 144*, 411-430; discussion 445-481.
60. Risselada, H. J., and Marrink, S. J. (2008) The molecular face of lipid rafts in model membranes, *Proceedings of the National Academy of Sciences of the United States of America 105*, 17367-17372.
61. Davis, R. S., Sunil Kumar, P. B., Sperotto, M. M., and Laradji, M. (2013) Predictions of phase separation in three-component lipid membranes by the MARTINI force field, *The journal of physical chemistry. B 117*, 4072-4080.
62. Schafer, L. V., de Jong, D. H., Holt, A., Rzepiela, A. J., de Vries, A. H., Poolman, B., Killian, J. A., and Marrink, S. J. (2011) Lipid packing drives the segregation of transmembrane helices into disordered lipid domains in model membranes, *Proceedings of the National Academy of Sciences of the United States of America 108*, 1343-1348.
63. Janosi, L., Li, Z., Hancock, J. F., and Gorfe, A. A. (2012) Organization, dynamics, and segregation of Ras nanoclusters in membrane domains, *Proceedings of the National Academy of Sciences of the United States of America 109*, 8097-8102.
64. Septinus, M., Berthold, T., Naujok, A., and Zimmermann, H. W. (1985) [Hydrophobic acridine dyes for fluorescent staining of mitochondria in living cells. 3. Specific accumulation of the fluorescent dye NAO on the mitochondrial membranes in HeLa cells by hydrophobic interaction. Depression of respiratory activity, changes in the ultrastructure of

mitochondria due to NAO. Increase of fluorescence in vital stained mitochondria in situ by irradiation], *Histochemistry 82*, 51-66.

65. Mileykovskaya, E., and Dowhan, W. (2000) Visualization of phospholipid domains in Escherichia coli by using the cardiolipin-specific fluorescent dye 10-N-nonyl acridine orange, *Journal of bacteriology 182*, 1172-1175.
66. Koppelman, C. M., Den Blaauwen, T., Duursma, M. C., Heeren, R. M. A., and Nanninga, N. (2001) Escherichia coli Minicell Membranes Are Enriched in Cardiolipin, *Journal of Bacteriology 183*, 6144-6147.
67. Dahlberg, M., and Maliniak, A. (2008) Molecular dynamics simulations of cardiolipin bilayers, *The journal of physical chemistry. B 112*, 11655-11663.
68. Rog, T., Martinez-Seara, H., Munck, N., Oresic, M., Karttunen, M., and Vattulainen, I. (2009) Role of cardiolipins in the inner mitochondrial membrane: insight gained through atom-scale simulations, *The journal of physical chemistry. B 113*, 3413-3422.
69. Poyry, S., Rog, T., Karttunen, M., and Vattulainen, I. (2009) Mitochondrial membranes with mono- and divalent salt: changes induced by salt ions on structure and dynamics, *The journal of physical chemistry. B 113*, 15513-15521.
70. Dahlberg, M., and Maliniak, A. (2010) Mechanical Properties of Coarse-Grained Bilayers Formed by Cardiolipin and Zwitterionic Lipids, *J Chem Theory Comput 6*, 1638-1649.
71. Haines, T. H., and Dencher, N. A. (2002) Cardiolipin: a proton trap for oxidative phosphorylation, *FEBS Lett 528*, 35-39.
72. Lins, R. D., and Straatsma, T. P. (2001) Computer simulation of the rough lipopolysaccharide membrane of Pseudomonas aeruginosa, *Biophysical journal 81*, 1037-1046.
73. Shroll, R. M., and Straatsma, T. P. (2002) Molecular structure of the outer bacterial membrane of Pseudomonas aeruginosa via classical simulation, *Biopolymers 65*, 395-407.
74. Lins, R. D., Vorpagel, E. R., Guglielmi, M., and Straatsma, T. P. (2008) Computer simulation of uranyl uptake by the rough lipopolysaccharide membrane of Pseudomonas aeruginosa, *Biomacromolecules 9*, 29-35.
75. Popot, J. L., and Engelman, D. M. (1990) Membrane protein folding and oligomerization: the two-stage model, *Biochemistry 29*, 4031-4037.

76. Bond, P. J., and Sansom, M. S. (2006) Insertion and assembly of membrane proteins via simulation, *Journal of the American Chemical Society 128*, 2697-2704.
77. Carpenter, T., Bond, P. J., Khalid, S., and Sansom, M. S. (2008) Self-assembly of a simple membrane protein: coarse-grained molecular dynamics simulations of the influenza M2 channel, *Biophysical journal 95*, 3790-3801.
78. Psachoulia, E., Marshall, D. P., and Sansom, M. S. (2010) Molecular dynamics simulations of the dimerization of transmembrane alpha-helices, *Acc Chem Res 43*, 388-396.
79. Kalli, A. C., Hall, B. A., Campbell, I. D., and Sansom, M. S. (2011) A helix heterodimer in a lipid bilayer: prediction of the structure of an integrin transmembrane domain via multiscale simulations, *Structure 19*, 1477-1484.
80. Psachoulia, E., Fowler, P. W., Bond, P. J., and Sansom, M. S. (2008) Helix-helix interactions in membrane proteins: coarse-grained simulations of glycophorin a helix dimerization, *Biochemistry 47*, 10503-10512.
81. Beevers, A. J., and Kukol, A. (2006) The transmembrane domain of the oncogenic mutant ErbB-2 receptor: a structure obtained from site-specific infrared dichroism and molecular dynamics, *Journal of molecular biology 361*, 945-953.
82. Prakash, A., Janosi, L., and Doxastakis, M. (2011) GxxxG motifs, phenylalanine, and cholesterol guide the self-association of transmembrane domains of ErbB2 receptors, *Biophysical journal 101*, 1949-1958.
83. Lemmin, T., Soto, C. S., Clinthorne, G., DeGrado, W. F., and Dal Peraro, M. (2013) Assembly of the transmembrane domain of E. coli PhoQ histidine kinase: implications for signal transduction from molecular simulations, *PLoS Computational Biology 9*, e1002878.
84. Chetwynd, A. P., Scott, K. A., Mokrab, Y., and Sansom, M. S. (2008) CGDB: a database of membrane protein/lipid interactions by coarse-grained molecular dynamics simulations, *Molecular membrane biology 25*, 662-669.
85. Jeanteur, D., Schirmer, T., Fourel, D., Simonet, V., Rummel, G., Widmer, C., Rosenbusch, J. P., Pattus, F., and Pages, J. M. (1994) Structural and functional alterations of a colicin-resistant mutant of OmpF porin from

Escherichia coli, *Proceedings of the National Academy of Sciences of the United States of America 91*, 10675-10679.

86. Cowan, S. W., Garavito, R. M., Jansonius, J. N., Jenkins, J. A., Karlsson, R., Konig, N., Pai, E. F., Pauptit, R. A., Rizkallah, P. J., Rosenbusch, J. P., and et al. (1995) The structure of OmpF porin in a tetragonal crystal form, *Structure 3*, 1041-1050.
87. Watanabe, M., Rosenbusch, J., Schirmer, T., and Karplus, M. (1997) Computer simulations of the OmpF porin from the outer membrane of Escherichia coli, *Biophysical journal 72*, 2094-2102.
88. Soares, C. M., Björkstén, J., and Tapia, O. (1995) L3 loop-mediated mechanisms of pore closing in porin: a molecular dynamics perturbation approach, *"Protein Engineering, Design and Selection" 8*, 5-12.
89. Björkstén, J., Soares, C. M., Nilsson, O., and Tapia, O. (1994) On the stability and plastic properties of the interior L3 loop in R.capsulatus porin. A molecular dynamics study, *"Protein Engineering, Design and Selection" 7*, 487-493.
90. Suenaga, A., Komeiji, Y., Uebayasi, M., Meguro, T., Saito, M., and Yamato, I. (1998) Computational observation of an ion permeation through a channel protein, *Bioscience Reports 18*, 39-48.
91. Tieleman, D. P., and Berendsen, H. J. C. (1998) A Molecular Dynamics Study of the Pores Formed by Escherichia coli OmpF Porin in a Fully Hydrated Palmitoyloleoylphosphatidylcholine Bilayer, *Biophysical Journal 74*, 2786-2801.
92. Im, W., and Roux, B. (2002) Ions and counterions in a biological channel: a molecular dynamics simulation of OmpF porin from Escherichia coli in an explicit membrane with 1 M KCl aqueous salt solution, *Journal of molecular biology 319*, 1177-1197.
93. Pezeshki, S., Chimerel, C., Bessonov, A. N., Winterhalter, M., and Kleinekathofer, U. (2009) Understanding ion conductance on a molecular level: an all-atom modeling of the bacterial porin OmpF, *Biophysical Journal 97*, 1898-1906.
94. Chimerel, C., Movileanu, L., Pezeshki, S., Winterhalter, M., and Kleinekathofer, U. (2008) Transport at the nanoscale: temperature dependence of ion conductance, *Eur Biophys J 38*, 121-125.

95. Simonet, V., Mallea, M., and Pages, J. M. (2000) Substitutions in the eyelet region disrupt cefepime diffusion through the Escherichia coli OmpF channel, *Antimicrob Agents Chemother 44*, 311-315.
96. Danelon, C., Nestorovich, E. M., Winterhalter, M., Ceccarelli, M., and Bezrukov, S. M. (2006) Interaction of zwitterionic penicillins with the OmpF channel facilitates their translocation, *Biophysical Journal 90*, 1617-1627.
97. Kumar, A., Hajjar, E., Ruggerone, P., and Ceccarelli, M. (2010) Molecular Simulations Reveal the Mechanism and the Determinants for Ampicillin Translocation through OmpF, *The Journal of Physical Chemistry B 114*, 9608-9616.
98. Ceccarelli, M., Danelon, C., Laio, A., and Parrinello, M. (2004) Microscopic Mechanism of Antibiotics Translocation through a Porin, *Biophysical Journal 87*, 58-64.
99. Ziervogel, B. K., and Roux, B. (2013) The binding of antibiotics in OmpF porin, *Structure 21*, 76-87.
100. Kumar, A., Hajjar, E., Ruggerone, P., and Ceccarelli, M. (2010) Structural and dynamical properties of the porins OmpF and OmpC: insights from molecular simulations, *Journal of Physics: Condensed Matter 22*, 454125.
101. Casuso, I., Khao, J., Chami, M., Paul-Gilloteaux, P., Husain, M., Duneau, J.-P., Stahlberg, H., Sturgis, J. N., and Scheuring, S. (2012) Characterization of the motion of membrane proteins using high-speed atomic force microscopy, *Nature Nanotechnology 7*, 525-529.
102. Goose, J. E., and Sansom, M. S. (2013) Reduced lateral mobility of lipids and proteins in crowded membranes, *PLoS Computational Biology 9*, e1003033.
103. Bond, P. J., Faraldo-Gomez, J. D., and Sansom, M. S. (2002) OmpA: a pore or not a pore? Simulation and modeling studies, *Biophysical journal 83*, 763-775.
104. Hong, H., Szabo, G., and Tamm, L. K. (2006) Electrostatic couplings in OmpA ion-channel gating suggest a mechanism for pore opening, *Nat Chem Biol 2*, 627-635.
105. Ramakrishnan, M., Pocanschi, C. L., Kleinschmidt, J. H., and Marsh, D. (2004) Association of spin-labeled lipids with beta-barrel proteins from the outer membrane of Escherichia coli, *Biochemistry 43*, 11630-11636.

106. Deol, S. S., Bond, P. J., Domene, C., and Sansom, M. S. (2004) Lipid-protein interactions of integral membrane proteins: a comparative simulation study, *Biophysical journal 87*, 3737-3749.
107. Baaden, M., and Sansom, M. S. (2004) OmpT: molecular dynamics simulations of an outer membrane enzyme, *Biophysical journal 87*, 2942-2953.
108. Straatsma, T. P., and Soares, T. A. (2009) Characterization of the outer membrane protein OprF of Pseudomonas aeruginosa in a lipopolysaccharide membrane by computer simulation, *Proteins 74*, 475-488.
109. Khalid, S., Bond, P. J., Deol, S. S., and Sansom, M. S. (2006) Modeling and simulations of a bacterial outer membrane protein: OprF from Pseudomonas aeruginosa, *Proteins 63*, 6-15.
110. Faraldo-Gomez, J. D., Smith, G. R., and Sansom, M. S. (2003) Molecular dynamics simulations of the bacterial outer membrane protein FhuA: a comparative study of the ferrichrome-free and bound states, *Biophysical journal 85*, 1406-1420.
111. Sen, T. Z., Kloster, M., Jernigan, R. L., Kolinski, A., Bujnicki, J. M., and Kloczkowski, A. (2008) Predicting the complex structure and functional motions of the outer membrane transporter and signal transducer FecA, *Biophysical Journal 94*, 2482-2491.
112. Khalid, S., and Sansom, M. S. (2006) Molecular dynamics simulations of a bacterial autotransporter: NalP from Neisseria meningitidis, *Molecular membrane biology 23*, 499-508.
113. Marin, E., Bodelon, G., and Fernandez, L. A. (2010) Comparative analysis of the biochemical and functional properties of C-terminal domains of autotransporters, *Journal of Bacteriology 192*, 5588-5602.
114. Ieva, R., Tian, P., Peterson, J. H., and Bernstein, H. D. (2011) Sequential and spatially restricted interactions of assembly factors with an autotransporter beta domain, *Proceedings of the National Academy of Sciences of the United States of America 108*, E383-391.
115. Tian, P., and Bernstein, H. D. (2010) Molecular basis for the structural stability of an enclosed beta-barrel loop, *Journal of molecular biology 402*, 475-489.

116. Dautin, N., Barnard, T. J., Anderson, D. E., and Bernstein, H. D. (2007) Cleavage of a bacterial autotransporter by an evolutionarily convergent autocatalytic mechanism, *The EMBO journal 26*, 1942-1952.
117. Hess, B. (2008) P-LINCS: A parallel linear constraint solver for molecular simulation, *J Chem Theory Comput 4*, 116-122.
118. 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.
119. Miyamoto, S., and Kollman, P. A. (1992) Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models, *J Comput Chem 13*, 952-962.
120. Jones, J. E. (1924) On the Determination of Molecular Fields. II. From the Equation of State of a Gas, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 106*, 463-477.
121. Shirts, M. R., Mobley, D. L., Chodera, J. D., and Pande, V. S. (2007) Accurate and efficient corrections for missing dispersion interactions in molecular simulations, *The journal of physical chemistry. B 111*, 13052-13063.
122. Ewald, P. P. (1921) Die Berechnung optischer und elektrostatischer Gitterpotentiale, *Annalen der Physik 369*, 253-287.
123. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method, *The Journal of Chemical Physics 103*, 8577.
124. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems, *The Journal of Chemical Physics 98*, 10089.
125. Woolf, T. B., and Roux, B. (1996) Structure, energetics, and dynamics of lipid-protein interactions: A molecular dynamics study of the gramicidin A channel in a DMPC bilayer, *Proteins 24*, 92-114.
126. Wolf, M. G., Hoefling, M., Aponte-Santamaria, C., Grubmuller, H., and Groenhof, G. (2010) g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation, *J Comput Chem 31*, 2169-2174.
127. Kandt, C., Ash, W. L., and Tieleman, D. P. (2007) Setting up and running molecular dynamics simulations of membrane proteins, *Methods 41*, 475-488.

128. Schmidt, T. H., and Kandt, C. (2012) LAMBADA and InflateGRO2: efficient membrane alignment and insertion of membrane proteins for molecular dynamics simulations, *J Chem Inf Model 52*, 2657-2669.
129. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath, *The Journal of Chemical Physics 81*, 3684.
130. Parrinello, M. (1981) Polymorphic transitions in single crystals: A new molecular dynamics method, *J Appl Phys 52*, 7182.
131. Shelley, J. C., Shelley, M. Y., Reeder, R. C., Bandyopadhyay, S., and Klein, M. L. (2001) A coarse grain model for phospholipid simulations, *J Phys Chem B 105*, 4464-4470.
132. Izvekov, S., and Voth, G. A. (2005) A multiscale coarse-graining method for biomolecular systems, *The Journal of Physical Chemistry B 109*, 2469-2473.
133. Orsi, M., and Essex, J. W. (2011) The ELBA force field for coarse-grain modeling of lipid membranes, *PLoS One 6*, e28637.
134. Marrink, S. J., de Vries, A. H., and Mark, A. E. (2003) Coarse Grained Model for Semiquantitative Lipid Simulations, *The Journal of Physical Chemistry B 108*, 750-760.
135. Khalid, S., Bond, P. J., Holyoake, J., Hawtin, R. W., and Sansom, M. S. (2008) DNA and lipid bilayers: self-assembly and insertion, *J R Soc Interface 5 Suppl 3*, S241-250.
136. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., and de Vries, A. H. (2007) The MARTINI Force Field: A Coarse Grained Model for Biomolecular Simulations, *The Journal of Physical Chemistry B 111*, 7812-7824.
137. Levitt, M., Yesylevskyy, S. O., Schäfer, L. V., Sengupta, D., and Marrink, S. J. (2010) Polarizable Water Model for the Coarse-Grained MARTINI Force Field, *PLoS Computational Biology 6*, e1000810.
138. López, C. A., Rzepiela, A. J., de Vries, A. H., Dijkhuizen, L., Hünenberger, P. H., and Marrink, S. J. (2009) Martini Coarse-Grained Force Field: Extension to Carbohydrates, *J Chem Theory Comput 5*, 3195-3210.
139. Kobayashi, K., Nakahori, Y., Miyake, M., Matsumura, K., Kondo-Iida, E., Nomura, Y., Segawa, M., Yoshioka, M., Saito, K., Osawa, M., Hamano, K., Sakakihara, Y., Nonaka, I., Nakagome, Y., Kanazawa, I., Nakamura, Y., Tokunaga, K., and Toda, T. (1998) An ancient retrotransposal insertion

causes Fukuyama-type congenital muscular dystrophy, *Nature 394*, 388-392.

140. Hayashi, Y. K., Ogawa, M., Tagawa, K., Noguchi, S., Ishihara, T., Nonaka, I., and Arahata, K. (2001) Selective deficiency of alpha-dystroglycan in Fukuyama-type congenital muscular dystrophy, *Neurology 57*, 115-121.
141. Beedle, A. M., Turner, A. J., Saito, Y., Lueck, J. D., Foltz, S. J., Fortunato, M. J., Nienaber, P. M., and Campbell, K. P. (2012) Mouse fukutin deletion impairs dystroglycan processing and recapitulates muscular dystrophy, *J Clin Invest 122*, 3330-3342.
142. Wiggins, C. A., and Munro, S. (1998) Activity of the yeast MNN1 alpha-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases, *Proceedings of the National Academy of Sciences of the United States of America 95*, 7945-7950.
143. Esapa, C. T., Benson, M. A., Schroder, J. E., Martin-Rendon, E., Brockington, M., Brown, S. C., Muntoni, F., Kroger, S., and Blake, D. J. (2002) Functional requirements for fukutin-related protein in the Golgi apparatus, *Human molecular genetics 11*, 3319-3331.
144. Marius, P., Wright, J. N., Findlow, I. S., and Williamson, P. T. (2010) Expression and purification of the transmembrane domain of Fukutin-I for biophysical studies, *Protein Expr Purif 72*, 107-112.
145. Dunphy, W. G., Fries, E., Urbani, L. J., and Rothman, J. E. (1981) Early and late functions associated with the Golgi apparatus reside in distinct compartments, *Proc Natl Acad Sci U S A 78*, 7453-7457.
146. Berger, E. G., and Hesford, F. J. (1985) Localization of galactosyl- and sialyltransferase by immunofluorescence: evidence for different sites, *Proc Natl Acad Sci U S A 82*, 4736-4739.
147. Opat, A. S., van Vliet, C., and Gleeson, P. A. (2001) Trafficking and localisation of resident Golgi glycosylation enzymes, *Biochimie 83*, 763-773.
148. Torelli, S., Brown, S. C., Brockington, M., Dolatshad, N. F., Jimenez, C., Skordis, L., Feng, L. H., Merlini, L., Jones, D. H., Romero, N., Wewer, U., Voit, T., Sewry, C. A., Noguchi, S., Nishino, I., and Muntoni, F. (2005) Sub-cellular localisation of fukutin related protein in different cell lines and in the muscle of patients with MDC1C and LGMD2I, *Neuromuscul Disord 15*, 836-843.

149. Gleeson, P. A. (1998) Targeting of proteins to the Golgi apparatus, *Histochem Cell Biol 109*, 517-532.
150. Ansell, G. B. H. J. N. D. R. M. C. (1973) *Form and function of phospholipids*, Elsevier Scientific Pub. Co., Amsterdam; New York.
151. Bretscher, M. S., and Munro, S. (1993) Cholesterol and the Golgi apparatus, *Science 261*, 1280-1281.
152. Pelham, H. R., and Munro, S. (1993) Sorting of membrane proteins in the secretory pathway, *Cell 75*, 603-605.
153. Munro, S. (1995) An investigation of the role of transmembrane domains in Golgi protein retention, *EMBO J 14*, 4695-4704.
154. Marius, P., Leung, Y. M., Piggot, T. J., Khalid, S., and Williamson, P. T. (2012) Probing the oligomeric state and interaction surfaces of Fukutin-I in dilauroylphosphatidylcholine bilayers, *Eur Biophys J 41*, 199-207.
155. Esapa, C. T., McIlhinney, R. A., and Blake, D. J. (2005) Fukutin-related protein mutations that cause congenital muscular dystrophy result in ER-retention of the mutant protein in cultured cells, *Hum Mol Genet 14*, 295-305.
156. Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000) Comparative protein structure modeling of genes and genomes, *Annu Rev Biophys Biomol Struct 29*, 291-325.
157. Laskowski, R. A., Moss, D. S., and Thornton, J. M. (1993) Main-chain bond lengths and bond angles in protein structures, *Journal of molecular biology 231*, 1049-1067.
158. van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005) GROMACS: fast, flexible, and free, *J Comput Chem 26*, 1701-1718.
159. Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. (1995) GROMACS: A message-passing parallel molecular dynamics implementation, *Computer Physics Communications 91*, 43-56.
160. Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation, *J Chem Theory Comput 4*, 435-447.
161. Berger, O., Edholm, O., and Jahnig, F. (1997) Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature, *Biophysical Journal 72*, 2002-2013.

162. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., and Hermans, J. (1981) Interaction models for water in relation to protein hydration, *Intermolecular Forces*, 331-342.
163. Bussi, G., Donadio, D., and Parrinello, M. (2007) Canonical sampling through velocity rescaling, *The Journal of Chemical Physics 126*, 014101.
164. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers 22*, 2577-2637.
165. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics, *J Mol Graph 14*, 33-38, 27-38.
166. Lins, R. D., and Hunenberger, P. H. (2005) A new GROMOS force field for hexopyranose-based carbohydrates, *J Comput Chem 26*, 1400-1412.
167. Bond, P. J., Holyoake, J., Ivetac, A., Khalid, S., and Sansom, M. S. (2007) Coarse-grained molecular dynamics simulations of membrane proteins and peptides, *J Struct Biol 157*, 593-605.
168. Bond, P. J., Wee, C. L., and Sansom, M. S. P. (2008) Coarse-Grained Molecular Dynamics Simulations of the Energetics of Helix Insertion into a Lipid Bilayer†, *Biochemistry 47*, 11321-11331.
169. van Meer, G. (1998) Lipids of the Golgi membrane, *Trends in Cell Biology 8*, 29-33.
170. Rouse, S. L., Carpenter, T., Stansfeld, P. J., and Sansom, M. S. (2009) Simulations of the BM2 proton channel transmembrane domain from influenza virus B, *Biochemistry 48*, 9949-9951.
171. Hess, B. (2002) Determining the shear viscosity of model liquids from molecular dynamics simulations, *The Journal of Chemical Physics 116*, 209-217.
172. Keramaris-Vrantsis, E., Lu, P. J., Doran, T., Zillmer, A., Ashar, J., Esapa, C. T., Benson, M. A., Blake, D. J., Rosenfeld, J., and Lu, Q. L. (2007) Fukutin-related protein localizes to the Golgi apparatus and mutations lead to mislocalization in muscle in vivo, *Muscle Nerve 36*, 455-465.
173. Holdbrook, D. A., Leung, Y. M., Piggot, T. J., Marius, P., Williamson, P. T., and Khalid, S. (2010) Stability and membrane orientation of the fukutin transmembrane domain: a combined multiscale molecular dynamics and circular dichroism study, *Biochemistry 49*, 10796-10802.

174. Murphy, T. F., and Apicella, M. A. (1987) Nontypable Haemophilus influenzae: a review of clinical aspects, surface antigens, and the human immune response to infection, *Rev Infect Dis 9*, 1-15.
175. Jacob-Dubuisson, F., Fernandez, R., and Coutte, L. (2004) Protein secretion through autotransporter and two-partner pathways, *Biochimica et biophysica acta 1694*, 235-257.
176. Lehr, U., Schutz, M., Oberhettinger, P., Ruiz-Perez, F., Donald, J. W., Palmer, T., Linke, D., Henderson, I. R., and Autenrieth, I. B. (2010) C-terminal amino acid residues of the trimeric autotransporter adhesin YadA of Yersinia enterocolitica are decisive for its recognition and assembly by BamA, *Molecular Microbiology 78*, 932-946.
177. Meng, G., St Geme, J. W., 3rd, and Waksman, G. (2008) Repetitive architecture of the Haemophilus influenzae Hia trimeric autotransporter, *Journal of molecular biology 384*, 824-836.
178. Hoover, W. G. (1985) Canonical dynamics: Equilibrium phase-space distributions, *Phys Rev A 31*, 1695-1697.
179. Nosé, S. (1984) A molecular dynamics method for simulations in the canonical ensemble, *Molecular Physics 52*, 255-268.
180. Nosé, S., and Klein, M. L. (1983) Constant pressure molecular dynamics for molecular systems, *Molecular Physics 50*, 1055-1076.
181. Anandakrishnan, R., Aguilar, B., and Onufriev, A. V. (2012) H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations, *Nucleic Acids Research 40*, W537-541.
182. Myers, J., Grothaus, G., Narayanan, S., and Onufriev, A. (2006) A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules, *Proteins 63*, 928-938.
183. Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., and Onufriev, A. (2005) H++: a server for estimating pKas and adding missing hydrogens to macromolecules, *Nucleic Acids Research 33*, W368-371.
184. van Gunsteren, W. F., Billeter, S. R., Eising, A. A., H√ºnenberger, P. H., Kr√ºger, P., Mark, A. E., Scott, W. R. P., and Tironi, I. G. (1996) *Biomolecular Simulation: The {GROMOS96} manual and userguide*, Hochschuleverlag AG an der ETH Z√ºrich.

185. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *The Journal of Chemical Physics 79*, 926.
186. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field, *Proteins 78*, 1950-1958.
187. Oostenbrink, C., Villa, A., Mark, A. E., and van Gunsteren, W. F. (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6, *J Comput Chem 25*, 1656-1676.
188. 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, *The journal of physical chemistry. B 115*, 13381-13388.
189. Helander, I. M., Lindner, B., Brade, H., Altmann, K., Lindberg, A. A., Rietschel, E. T., and Zahringer, U. (1988) Chemical structure of the lipopolysaccharide of Haemophilus influenzae strain I-69 Rd-/b+. Description of a novel deep-rough chemotype, *Eur J Biochem 177*, 483-492.
190. Wang, H., and Cronan, J. E. (2003) Haemophilus influenzae Rd lacks a stringently conserved fatty acid biosynthetic enzyme and thermal control of membrane lipid composition, *Journal of Bacteriology 185*, 4930-4937.
191. Holdbrook, D. A., Piggot, T. J., Sansom, M. S., and Khalid, S. (2013) Stability and membrane interactions of an autotransport protein: MD simulations of the Hia translocator domain in a complex membrane environment, *Biochimica et biophysica acta 1828*, 715-723.
192. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., and Beckstein, O. (2011) MDAnalysis: A toolkit for the analysis of molecular dynamics simulations, *J Comput Chem*.
193. Smart, O. S., Goodfellow, J. M., and Wallace, B. A. (1993) The pore dimensions of gramicidin A, *Biophysical Journal 65*, 2455-2460.
194. Smart, O. S., Neduvelil, J. G., Wang, X., Wallace, B. A., and Sansom, M. S. (1996) HOLE: a program for the analysis of the pore dimensions of ion channel structural models, *J Mol Graph 14*, 354-360, 376.

195. Mikula, K. M., Leo, J. C., Lyskowski, A., Kedracka-Krok, S., Pirog, A., and Goldman, A. (2012) The translocation domain in trimeric autotransporter adhesins is necessary and sufficient for trimerization and autotransportation, *J Bacteriol 194*, 827-838.
196. Akama, H., Kanemaki, M., Yoshimura, M., Tsukihara, T., Kashiwagi, T., Yoneyama, H., Narita, S., Nakagawa, A., and Nakae, T. (2004) Crystal structure of the drug discharge outer membrane protein, OprM, of Pseudomonas aeruginosa: dual modes of membrane anchoring and occluded cavity end, *The Journal of biological chemistry 279*, 52816-52819.
197. Vaccaro, L., Scott, K. A., and Sansom, M. S. (2008) Gating at both ends and breathing in the middle: conformational dynamics of TolC, *Biophysical Journal 95*, 5681-5691.
198. Smart, O. S., Breed, J., Smith, G. R., and Sansom, M. S. (1997) A novel method for structure-based prediction of ion channel conductance properties, *Biophysical Journal 72*, 1109-1126.
199. Ieva, R., Skillman, K. M., and Bernstein, H. D. (2008) Incorporation of a polypeptide segment into the beta-domain pore during the assembly of a bacterial autotransporter, *Molecular Microbiology 67*, 188-201.
200. Grosskinsky, U., Schutz, M., Fritz, M., Schmid, Y., Lamparter, M. C., Szczesny, P., Lupas, A. N., Autenrieth, I. B., and Linke, D. (2007) A conserved glycine residue of trimeric autotransporter domains plays a key role in Yersinia adhesin A autotransport, *Journal of Bacteriology 189*, 9011-9019.
201. Heinrichs, D. E., Yethon, J. A., and Whitfield, C. (1998) Molecular basis for structural diversity in the core regions of the lipopolysaccharides of Escherichia coli and Salmonella enterica, *Molecular Microbiology 30*, 221-232.
202. Erridge, C., Bennett-Guerrero, E., and Poxton, I. R. (2002) Structure and function of lipopolysaccharides, *Microbes Infect 4*, 837-851.
203. Yethon, J. A., Vinogradov, E., Perry, M. B., and Whitfield, C. (2000) Mutation of the lipopolysaccharide core glycosyltransferase encoded by waaG destabilizes the outer membrane of Escherichia coli by interfering with core phosphorylation, *Journal of Bacteriology 182*, 5620-5623.
204. Leive, L., Shovlin, V. K., and Mergenhagen, S. E. (1968) Physical, chemical, and immunological properties of lipopolysaccharide released

from Escherichia coli by ethylenediaminetetraacetate, *The Journal of biological chemistry 243*, 6384-6391.

205. Visudtiphole, V., Thomas, M. B., Chalton, D. A., and Lakey, J. H. (2005) Refolding of Escherichia coli outer membrane protein F in detergent creates LPS-free trimers and asymmetric dimers, *Biochem J 392*, 375-381.
206. de Cock, H., and Tommassen, J. (1996) Lipopolysaccharides and divalent cations are involved in the formation of an assembly-competent intermediate of outer-membrane protein PhoE of E.coli, *The EMBO journal 15*, 5567-5573.
207. Kadurugamuwa, J. L., and Beveridge, T. J. (1995) Virulence factors are released from Pseudomonas aeruginosa in association with membrane vesicles during normal growth and exposure to gentamicin: a novel mechanism of enzyme secretion, *Journal of bacteriology 177*, 3998-4008.
208. Haurat, M. F., Aduse-Opoku, J., Rangarajan, M., Dorobantu, L., Gray, M. R., Curtis, M. A., and Feldman, M. F. (2011) Selective sorting of cargo proteins into bacterial membrane vesicles, *Journal of Biological Chemistry 286*, 1269-1276.
209. Kramer, R. A., Brandenburg, K., Vandeputte-Rutten, L., Werkhoven, M., Gros, P., Dekker, N., and Egmond, M. R. (2002) Lipopolysaccharide regions involved in the activation of Escherichia coli outer membrane protease OmpT, *Eur J Biochem 269*, 1746-1752.
210. Kukkonen, M., Suomalainen, M., Kyllonen, P., Lahteenmaki, K., Lang, H., Virkola, R., Helander, I. M., Holst, O., and Korhonen, T. K. (2004) Lack of O-antigen is essential for plasminogen activation by Yersinia pestis and Salmonella enterica, *Molecular Microbiology 51*, 215-225.
211. Eren, E., and van den Berg, B. (2012) Structural Basis for Activation of an Integral Membrane Protease by Lipopolysaccharide, *Journal of Biological Chemistry 287*, 23971-23976.
212. Hagge, S. O., de Cock, H., Gutsmann, T., Beckers, F., Seydel, U., and Wiese, A. (2002) Pore formation and function of phosphoporin PhoE of Escherichia coli are determined by the core sugar moiety of lipopolysaccharide, *The Journal of biological chemistry 277*, 34247-34253.

213. Raymond, K. N., Dertz, E. A., and Kim, S. S. (2003) Enterobactin: an archetype for microbial iron transport, *Proceedings of the National Academy of Sciences of the United States of America 100*, 3584-3588.
214. Braun, V., and Braun, M. (2002) Iron transport and signaling in Escherichia coli, *FEBS Lett 529*, 78-85.
215. Endriss, F., and Braun, V. (2004) Loop deletions indicate regions important for FhuA transport and receptor functions in Escherichia coli, *Journal of bacteriology 186*, 4818-4823.
216. Fuller-Schaefer, C. A., and Kadner, R. J. (2005) Multiple extracellular loops contribute to substrate binding and transport by the Escherichia coli cobalamin transporter BtuB, *Journal of bacteriology 187*, 1732-1739.
217. Sauter, A., and Braun, V. (2004) Defined inactive FecA derivatives mutated in functional domains of the outer membrane transport and signaling protein of Escherichia coli K-12, *Journal of bacteriology 186*, 5303-5310.
218. Ferguson, A. D., Chakraborty, R., Smith, B. S., Esser, L., van der Helm, D., and Deisenhofer, J. (2002) Structural basis of gating by the outer membrane transporter FecA, *Science 295*, 1715-1719.
219. Ferguson, A. D., Hofmann, E., Coulton, J. W., Diederichs, K., and Welte, W. (1998) Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide, *Science 282*, 2215-2220.
220. Ferguson, A. D., Welte, W., Hofmann, E., Lindner, B., Holst, O., Coulton, J. W., and Diederichs, K. (2000) A conserved structural motif for lipopolysaccharide recognition by procaryotic and eucaryotic proteins, *Structure 8*, 585-592.
221. Piggot, T. J., Holdbrook, D. A., and Khalid, S. (2013) Conformational dynamics and membrane interactions of the E. coli outer membrane protein FecA: a molecular dynamics simulation study, *Biochimica et biophysica acta 1828*, 284-293.
222. Development Core Team. (2005) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
223. Flyvbjerg, H., and Petersen, H. G. (1989) Error estimates on averages of correlated data, *The Journal of Chemical Physics 91*, 461-466.

224. Grossfield, A., and Zuckerman, D. M. (2009) Quantifying uncertainty and sampling quality in biomolecular simulations, *Annu Rep Comput Chem 5*, 23-48.
225. Hess, B. (2002) Convergence of sampling in protein simulations, *Phys Rev E Stat Nonlin Soft Matter Phys 65*, 031910.
226. Faraldo-Gomez, J. D., Forrest, L. R., Baaden, M., Bond, P. J., Domene, C., Patargias, G., Cuthbertson, J., and Sansom, M. S. (2004) Conformational sampling and dynamics of membrane proteins from 10-nanosecond computer simulations, *Proteins 57*, 783-791.
227. Grossfield, A., Feller, S. E., and Pitman, M. C. (2007) Convergence of molecular dynamics simulations of membrane proteins, *Proteins 67*, 31-40.
228. Romo, T. D., and Grossfield, A. (2011) Block Covariance Overlap Method and Convergence in Molecular Dynamics Simulation, *J Chem Theory Comput 7*, 2464-2472.
229. Park, K. S., Choi, K. H., Kim, Y. S., Hong, B. S., Kim, O. Y., Kim, J. H., Yoon, C. M., Koh, G. Y., Kim, Y. K., and Gho, Y. S. (2010) Outer membrane vesicles derived from Escherichia coli induce systemic inflammatory response syndrome, *PLoS One 5*, e11334.
230. Shah, B., Sullivan, C. J., Lonergan, N. E., Stanley, S., Soult, M. C., and Britt, L. D. (2012) Circulating bacterial membrane vesicles cause sepsis in rats, *Shock 37*, 621-628.
231. Alaniz, R. C., Deatherage, B. L., Lara, J. C., and Cookson, B. T. (2007) Membrane vesicles are immunogenic facsimiles of Salmonella typhimurium that potently activate dendritic cells, prime B and T cell responses, and stimulate protective immunity in vivo, *Journal of immunology 179*, 7692-7701.
232. Kim, O. Y., Hong, B. S., Park, K. S., Yoon, Y. J., Choi, S. J., Lee, W. H., Roh, T. Y., Lotvall, J., Kim, Y. K., and Gho, Y. S. (2013) Immunization with Escherichia coli Outer Membrane Vesicles Protects Bacteria-Induced Lethality via Th1 and Th17 Cell Responses, *Journal of immunology 190*, 4092-4102.
233. Haneberg, B., Dalseg, R., Wedege, E., Hoiby, E. A., Haugen, I. L., Oftung, F., Andersen, S. R., Naess, L. M., Aase, A., Michaelsen, T. E., and Holst, J. (1998) Intranasal administration of a meningococcal outer membrane vesicle vaccine induces persistent local mucosal antibodies and serum

antibodies with strong bactericidal activity in humans, *Infect Immun 66*, 1334-1341.

234. McConnell, M. J., Rumbo, C., Bou, G., and Pachon, J. (2011) Outer membrane vesicles as an acellular vaccine against Acinetobacter baumannii, *Vaccine 29*, 5705-5710.
235. Saunders, N. B., Shoemaker, D. R., Brandt, B. L., Moran, E. E., Larsen, T., and Zollinger, W. D. (1999) Immunogenicity of intranasally administered meningococcal native outer membrane vesicles in mice, *Infect Immun 67*, 113-119.
236. Schild, S., Nelson, E. J., Bishop, A. L., and Camilli, A. (2009) Characterization of Vibrio cholerae outer membrane vesicles as a candidate vaccine for cholera, *Infect Immun 77*, 472-484.
237. McBroom, A. J., and Kuehn, M. J. (2007) Release of outer membrane vesicles by Gram-negative bacteria is a novel envelope stress response, *Molecular Microbiology 63*, 545-558.
238. Manning, A. J., and Kuehn, M. J. (2011) Contribution of bacterial outer membrane vesicles to innate bacterial defense, *BMC Microbiol 11*, 258.
239. Mashburn, L. M., and Whiteley, M. (2005) Membrane vesicles traffic signals and facilitate group activities in a prokaryote, *Nature 437*, 422-425.
240. Lee, E.-Y., Bang, J. Y., Park, G. W., Choi, D.-S., Kang, J. S., Kim, H.-J., Park, K.-S., Lee, J.-O., Kim, Y.-K., Kwon, K.-H., Kim, K.-P., and Gho, Y. S. (2007) Global proteomic profiling of native outer membrane vesicles derived fromEscherichia coli, *Proteomics 7*, 3143-3153.
241. Marrink, S. J., and Mark, A. E. (2003) Molecular dynamics simulation of the formation, structure, and dynamics of small phospholipid vesicles, *Journal of the American Chemical Society 125*, 15233-15242.
242. de Vries, A. H., Mark, A. E., and Marrink, S. J. (2004) Molecular dynamics simulation of the spontaneous formation of a small DPPC vesicle in water in atomistic detail, *Journal of the American Chemical Society 126*, 4488-4489.
243. Risselada, H. J., Mark, A. E., and Marrink, S. J. (2008) Application of mean field boundary potentials in simulations of lipid vesicles, *The journal of physical chemistry. B 112*, 7438-7447.
244. Parton, D. L., Klingelhoefer, J. W., and Sansom, M. S. (2011) Aggregation of model membrane proteins, modulated by hydrophobic mismatch,

membrane curvature, and protein class, *Biophysical journal 101*, 691-699.

245. Hall, B. A., Armitage, J. P., and Sansom, M. S. (2012) Mechanism of bacterial signal transduction revealed by molecular dynamics of Tsr dimers and trimers of dimers in lipid vesicles, *PLoS Computational Biology 8*, e1002685.
246. Martinez, L., Andrade, R., Birgin, E. G., and Martinez, J. M. (2009) PACKMOL: a package for building initial configurations for molecular dynamics simulations, *J Comput Chem 30*, 2157-2164.
247. Marrink, S. J., and Mark, A. E. (2003) The mechanism of vesicle fusion as revealed by molecular dynamics simulations, *Journal of the American Chemical Society 125*, 11144-11145.
248. Knecht, V., and Marrink, S. J. (2007) Molecular dynamics simulations of lipid vesicle fusion in atomic detail, *Biophysical journal 92*, 4254-4261.
249. Markvoort, A. J., van Santen, R. A., and Hilbers, P. A. (2006) Vesicle shapes from molecular dynamics simulations, *The journal of physical chemistry. B 110*, 22780-22785.
250. Wai, S. N., Lindmark, B., Soderblom, T., Takade, A., Westermark, M., Oscarsson, J., Jass, J., Richter-Dahlfors, A., Mizunoe, Y., and Uhlin, B. E. (2003) Vesicle-mediated export and assembly of pore-forming oligomers of the enterobacterial ClyA cytotoxin, *Cell 115*, 25-35.
251. Kito, M., Ishinaga, M., Nishihara, M., Kato, M., and Sawada, S. (1975) Metabolism of the phosphatidylglycerol molecular species in Escherichia coli, *Eur J Biochem 54*, 55-63.
252. Yokota, K., Kanamoto, R., and Kito, M. (1980) Composition of cardiolipin molecular species in Escherichia coli, *Journal of bacteriology 141*, 1047-1051.
253. Dima, R. I., and Thirumalai, D. (2004) Asymmetry in the Shapes of Folded and Denatured States of Proteins†, *The Journal of Physical Chemistry B 108*, 6564-6570.
254. Meyer, R., Sonnen, A. F., and Nau, W. M. (2010) Phase-dependent lateral diffusion of alpha-tocopherol in DPPC liposomes monitored by fluorescence quenching, *Langmuir : the ACS journal of surfaces and colloids 26*, 14723-14729.
255. Sheats, J. R., and McConnell, H. M. (1978) A photochemical technique for measuring lateral diffusion of spin-labeled phospholipids in

membranes, *Proceedings of the National Academy of Sciences of the United States of America 75*, 4661-4663.

256. Wohlert, J., and Edholm, O. (2006) Dynamics in atomistic simulations of phospholipid membranes: Nuclear magnetic resonance relaxation rates and lateral diffusion, *The Journal of Chemical Physics 125*, 204703.

257. Karo, J., Peterson, P., and Vendelin, M. (2012) Molecular dynamics simulations of creatine kinase and adenine nucleotide translocase in mitochondrial membrane patch, *The Journal of biological chemistry 287*, 7467-7476.

Chapter 8: Appendix

8.1 Subroutines added to g_msd for the calculation of the diffusion coefficients of molecules in a vesicle membrane

A

```
static void calc_angular_displacement(t_corr *this,int nmol,int *molindex,
        rvec xcur[],rvec nx0[],rvec vcom1,rvec vcom2,rvec angvel)
{
  int m,mol,d;
  rvec xprod,ppos,cpos,xprodt,angveltot;
  real mass,mtot,linvc,linvp,theta,sintheta,costheta;

  clear_rvec(angvel);
  clear_rvec(angveltot);
  mtot=0;

  for(m=0; m<nmol; m++) {
    clear_rvec(ppos);
    clear_rvec(cpos);
    clear_rvec(xprod);
    clear_rvec(xprodt);
    mol = molindex[m];
    mass = this->mass[mol];
    for(d=0; d<DIM; d++) {
      ppos[d] = nx0[mol][d] - vcom2[d];
      cpos[d] = xcur[mol][d] - vcom1[d];
    }
    cprod(ppos,cpos,xprod);
    linvp = norm(ppos);
    linvc = norm(cpos);
    for(d=0; d<DIM; d++) {
      xprodt[d] = xprod[d]/(linvp*linvc);
    }
    sintheta = norm(xprodt);
    costheta = cos_angle(ppos,cpos);
    if (sintheta > 0) {
      if (costheta >= 0){
        theta = asin(sintheta);
      } else {
        theta = acos(costheta);
      }
      for(d=0; d<DIM; d++) {
        angveltot[d] += mass*theta*(xprodt[d]/sintheta);
      }
    } else {
        clear_rvec(angveltot);
    }
    mtot += mass;
  }
  svmul(1/mtot,angveltot,angvel);
}
```

B

```
static void calc_rot_rodrig(rvec avrot, rvec radt0)
{
    int d;
    rvec xprod,uavrot,a,b,c;
    real linv,cost,sint,dprod;

    clear_rvec(uavrot);
    clear_rvec(xprod);
    clear_rvec(a);
    clear_rvec(b);
    clear_rvec(c);
    linv = norm(avrot);
    cost = cos(linv);
    sint = sin(-linv);

    for(d=0;d<DIM;d++) {
      uavrot[d] = avrot[d]*(1/linv);
    }
    cprod(uavrot,radt0,xprod);
    dprod = iprod(uavrot,radt0);
    for(d=0;d<DIM;d++) {
      a[d] = cost*radt0[d];
      b[d] = xprod[d]*sint;
      c[d] = uavrot[d]*dprod*(1-cost);
      radt0[d] = a[d]+b[d]+c[d];
    }
}
```

C

```
static void calc_ves_com(t_corr *this,int nmol,
                         int *molindex,rvec xa[],
                         rvec vcom)
{
  int  m,mol,d;
  rvec xm;
  real mass,mtot;

  clear_rvec(vcom);
  clear_rvec(xm);
  mtot = 0;

  for(m=0; m<nmol; m++) {
    mol = molindex[m];
    mass = this->mass[mol];
    for(d=0; d<DIM; d++)
      xm[d] += mass*xa[mol][d];
    mtot += mass;
   }
  svmul(1/mtot,xm,vcom);
}
```

Appendix Figure 50. A) This subroutine calculates the total angular displacement of the vesicle between two frames. B) This subroutine uses the Rodrigues rotation equation to correct the positions of the particles at time $t + \Delta t$. C) This subroutine calculates the vesicles COM.

8.2 Stability and membrane orientation of the fukutin transmembrane domain: a combined multiscale molecular dynamics and circular dichroism study

8.3 Stability and membrane interactions of an autotransport protein: MD simulations of the Hia translocator domain in a complex membrane environment

8.4 Conformational dynamics and membrane interactions of the E. coli outer membrane protein FecA: a molecular dynamics simulation study