Structural and functional studies of perforin-like proteins implicated in zoonotic parasitic diseases



Sophie Williams
Exeter College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Trinity 2019

Abstract

MACPF/CDCs are an ancient superfamily of pore forming proteins that have evolved a critical role in both cellular defence and attack. Many copies of MACPF/CDC proteins are released by a cell before binding to target membranes, oligomerising and undergoing a large structural rearrangement to punch a hole into the membrane, typically killing the target cell. In this thesis the steps of pore formation from membrane recognition through to oligomerisation and membrane insertion have been investigated for several recently discovered MACPF/CDC proteins using a combination of structural and computational biology techniques.

Perforin-like proteins (PLPs) belong to the MACPF/CDC superfamily and are unique to apicomplexan parasites including *Plasmodium* and *Toxoplasma*. PLPs have a role in facilitating parasite escape from host cells following replication, however their precise mechanism of action is unknown.

Here, molecular simulation approaches were used to investigate the stages of pore formation of the *Toxoplasma gondii* PLP1 (TgPLP1). The mode of membrane binding of TgPLP1 was determined, while investigation of oligomerisation states of the MACPF domain led to the generation of a model for the TgPLP1 transmembrane pore structure. *Plasmodium* have 5 PLPs (PPLPs1-5), which are expressed at different points in their lifecycles. The crystal structure of the *Plasmodium* PLP1 APCβ domain is presented here, and molecular dynamics (MD) was used to investigate PPLP1 and PPLP2 membrane binding, showing that these proteins have distinct mechanisms of membrane binding.

Another recently discovered MACPF/CDC protein is perforin-2, a close relative to perforin, which is produced by both immune and parenchymal cells in the body to defend against intracellular bacteria. The structures of the perforin-2 monomer and pre-pore were recently solved, and here we used MD to investigate the initial steps of membrane recognition and oligomerisation of perforin-2. A key finding was that perforin-2 binds to the membrane with its MACPF domain facing away from the bilayer, therefore identifying a unique mechanism of binding and activation amongst the MACPF/CDC family.

Acknowledgements

My thanks firstly go to my supervisors Phill Stansfeld and Rob Gilbert for their exceptional support and guidance over the last few years and for welcoming me into their labs during my project. I would like to thank the DTP and BBSRC for providing me the opportunity and funding to undertake this work, and thanks to Esther Becker at the DTP for her guidance as my tutor.

I would like to thank members of both the Stansfeld and Gilbert labs, and the members of SBCB as a whole, for making both labs a great place to work. In particular, my thanks go to Tao Ni and Xiulian Yu for their teaching, helpful discussion and collaboration.

I also thank my colleagues at Oxford Nanopore Technologies, especially my managers Stephen Pascoe and Richard Carter, for providing me the chance to further my software development skills as part of their fantastic team during my internship, and now in my subsequent employment.

Thanks to Sarah and Jacquie for all their support and many much-needed lunch breaks, dinners and game nights. Thanks to my family for their encouragement over the years. And finally, thanks to my fiancé John, for always being there through the highs and lows, for his endless support, kindness, patience and encouragement that have helped me along every step of the way.

Contents

List of figures v

List of tables viii

List of acronyms and abbreviations ix

Chapter 1 1

1.1 Membranes and membrane proteins 2

1.1.1 Peripheral membrane proteins 5

1.2 Evolution and molecular functions of pore forming proteins 7

1.3 The MACPF/CDC superfamily of pore forming proteins 7

1.3.1 The MACPF/CDC Superfamily 7

1.3.2 MACPF/CDC membrane binding domains 9

1.3.3 MACPF oligomerization and pore formation 12

1.4 The Apicomplexan life cycle and perforin-like proteins 14

1.4.1 Perforin-like proteins 14

1.4.2 Apicomplexans 17

1.4.3 Toxoplasma 18

1.4.4 The Plasmodium lifecycle 20

1.5 Perforin-2 22

1.5.1 The role of perforin-2 in the immune system 22

1.5.2 The Bacterial cell walls 23

1.6 Thesis Outline 26
Chapter 2 27
2.1 Molecular Dynamics 28
2.1.1 Molecular dynamics and membrane proteins 28
2.1.2 Molecular Dynamics simulations 28
2.1.3 Coarse grained simulations 32
2.1.4 Simulating bilayers, lipids and protein-lipid interactions 35
2.2 Protein Production and X-ray crystallography 37
2.3 Biophysical techniques 39
2.3.1 Lipid binding assay 39
Chapter 3 40
3.1 Introduction 41
3.1.1 *Toxoplasma gondii* is the causative agent of toxoplasmosis 41
3.1.2 TgPLP1 facilitates membrane disruption during cell egress 41
3.1.3 The MACPF and C-terminal domain structures of TgPLP1 41
3.1.4 Regulation of TgPLP1 activity 44
3.1.5 Proposed mechanisms of TgPLP1 mediated egress 45
3.2 Results and Discussion 45
3.2.1 Oligomerisation of TgPLP1 MACPF domain 45
3.2.2 Stability of MACPF oligomers in atomistic simulations 46
3.2.3 Analysis of hydrogen bonding at MACPF oligomer interfaces 48
3.2.4 Large scale fluctuations of the two oligomeric forms 52
3.2.5 Discussion of TgPLP1 MACPF oligomerisation interface preferences 55
3.2.6 Membrane association of TgPLP1 APCβ 57
3.2.7 PMF calculations to investigate lipid preferences 59
3.2.8 Detailed atomistic investigation of lipid binding characteristics of APCβ 67
3.2.9 APCβ membrane binding discussion 69
3.2.10 Modelling the TgPLP1 transmembrane pore 72
3.3 Conclusions 85

Chapter 4 88
4.1 Introduction 89
4.1.1 Apicomplexan Perforin-like proteins 89
4.1.2 *Plasmodium* PLPs function at different points in its lifecycle 89
4.1.3 PPLPs 3, 4 and 5 act in cell traversal across the mosquito midgut 90
4.1.4 Evidence for membrane disruption and pore formation in PLPs 91
4.1.5 Phylogenetic analysis of PLPs 92
4.2 Results and Discussion 96
4.2.1 Structural Determination of the PvPLP1 APCβ domain 96
4.2.2 The PvPLP1 APCβ domain crystal structure 103
4.2.3 Membrane binding characteristics of PvPLP1 and PvPLP2 APCβ domains 108
4.2.4 PvPLP2 membrane binding 118
4.2.5 Lipid specificity of PvPLP2 APCβ 127
4.3 Conclusion 128
Chapter 5 131
5.1 Introduction 132
5.1.1 The role of perforin-2 in the immune system 132
5.1.2 Molecular insights into perforin-2 activity 132
5.1.3 The perforin-2 structure 134
5.2 Results and Discussion 136
5.2.1 Atomistic simulation of P2 domain monomer in solvent 137
5.2.2 P2 domain simulations 137
5.2.3 Monomer membrane binding 151
5.2.4 Pre-pore membrane binding 152
5.2.5 PFN2 membrane binding discussion 156
5.2.6 Investigating the mechanism of pre-pore to pore transition 157
5.2.7 Discussion and further work to investigate the pre-pore to pore transition 160
5.3 Conclusions 161
Chapter 6 163

6.1 Summary 164
6.2 Future Directions 165
References 168

List of Figures

Figure 1.1 The eukaryotic plasma membrane 3
Figure 1.2 Phospholipid structure 4
Figure 1.3 Comparison of MACPF and CDC structures 8
Figure 1.4 Mechanism of CDC pore formation 11
Figure 1.5 The Complement membrane attack complex 13
Figure 1.6 Apicomplexan PLP domain architecture 16
Figure 1.7 Apicomplexan cell morphology 18
Figure 1.8 The Toxoplasma gondii lifecycle 19
Figure 1.9 The plasmodium lifecycle 21
Figure 1.10 Gram-negative and Gram-positive bacterial cell walls 25
Figure 1.11 The structure of LPS 25
Figure 2.1 A typical potential energy function for biomolecular force fields 30
Figure 2.2 Martini forcefield coarse grain mapping 32
Figure 2.3 Membrane association simulation setup 34
Figure 2.4 Membrane self-assembly 36
Figure 3.1 Crystal structures of TgPLP1 MACPF domain. 43
Figure 3.2 The TgPLP1 APCβ domain crystal structure. 44
Figure 3.3 Analysis of structural stability of TgPLP1 MACPF monomeric crystal structure during MD simulations. 46
Figure 3.4 Structural stability of TgPLP1 MACPF hexameric helix and ring structures in simulations 47
Figure 3.5 Hydrogen bonding in TgPLP1 MACPF oligomers. 50
Figure 3.6 Comparison of hydrogen bonding at MACPF oligomer interfaces. 51
Figure 3.7 Assessing large scale stability of TgPLP1 MACPF ring. 53

Figure 3.8 Vertical motions in MACPF helix simulations..... 54
Figure 3.9 Helix motions in CG simulations..... 55
Figure 3.10 Membrane binding of TgPLP1 APCβ domain. 58
Figure 3.11 Lipid binding preferences and mutational studies of TgPLP1 APCβ domain..... 60
Figure 3.12 Comparison of TgPLP1 APCβ domain binding to PC compared to PE bilayers. 62
Figure 3.13 CG and AT APCβ lipid contacts..... 65
Figure 3.14 Electrostatic interactions and lipid clustering of TgPLP1 APCβ 66
Figure 3.15 Stability of TgPLP1 membrane bound and soluble APCβ domain. 68
Figure 3.16 TgPLP1 APCβ domain protein-lipid hydrogen bonds..... 68
Figure 3.17 TgPLP1 β-barrel modelling procedure. 76
Figure 3.18 Membrane self-assembly simulations of the TgPLP1 β-barrel model. 79
Figure 3.19 Stability of the TgPLP1 β-barrel in CG and AT simulations 81
Figure 3.20 Protein-lipid contacts for the TgPLP1 β-barrel model..... 83
Figure 4.1 PLP domain architecture and phylogenetic tree. 94
Figure 4.2 PPLP sequence alignment..... 95
Figure 4.3 PvPLP1 APCβ small scale expression tests..... 97
Figure 4.4 Small scale mammalian expression tests of PvPLP1 MACPF and NTD constructs..... 97
Figure 4.5 Large Scale purification and crystallisation of PvPLP1 APCβ. 98
Figure 4.6 Large Scale purification of PvPLP2 APCβ..... 99
Figure 4.7 Stability of PvPLP1, PvPLP2 and TgPLP1 101
Figure 4.8 PvPLP1 APCβ crystal structure..... 104
Figure 4.9 Structural stability of PvPLP1 crystal structure in MD simulations..... 105
Figure 4.10 Sequence conservation across PPLPs. 107
Figure 4.11 Structural comparison of PvPLP1, 2, 4, 5 and TgPLP1 APCβ structures. 108
Figure 4.12 Lipid binding of PvPLP1 and PvPLP2 APCβs..... 110
Figure 4.13 Membrane association simulations of PvPLP1 APCβ on PC:PE:PS bilayer. 113
Figure 4.14 Membrane association simulations of PvPLP1 APCβ on PC:PE:PS:PI(3,4,5)P_3 bilayer. 115
Figure 4.15 Free energy of binding of PvPLP1 APCβ at pH 7 and pH 5.5 117
Figure 4.16 Simulations of PvPLP1 APCβ protonated to pH5.5 on PC:PE:PS bilayers. 117
Figure 4.17 PvPLP2 APCβ membrane association simulations. 120
Figure 4.18 Stability of PvPLP2 APCβ CG2AT conversion of membrane binding in upright and sideways orientation. 122
Figure 4.19 Free energy of binding of PvPLP2 APCβ on membranes. 126
Figure 4.20 CG simulations show clustering of PI(4)P around PvPLP2 APCβ. 129
Figure 5.1 The Perforin-2 P2 domain crystal structure..... 133
Figure 5.2 The perforin-2 pre-pore structure 135

Figure 5.3 Stability of P2 domain in solution 138
Figure 5.4 Perforin-2 P2 domain membrane association simulations on prokaryotic and eukaryotic membranes 141
Figure 5.5 PG:PE lipid contact analysis of PFN2 P2 domain 142
Figure 5.6 P2 domain contact with LPS 146
Figure 5.7 AT simulation of the P2 domain on PG:PE bilayers 146
Figure 5.8 Membrane contact of P2 domain with PG:PE bilayers in AT simulations 147
Figure 5.9 Perforin-2 monomer binding to PG:PE bilayers 151
Figure 5.10 Lipid contacts and membrane binding of PFN2 monomers 153
Figure 5.11 Pre-pore membrane binding 155
Figure 5.12 A model of the PFN2 pre-pore bound to the membrane 157
Figure 5.13 Stability of PFN2 monomer and pre-pore 158
Figure 5.14 PFN2 domain interfaces 160

List of Tables

Table 3.1 Counts of main chain hydrogen bonds between each adjacent pair of MACPF chains in the ring and helical crystal structures. 48
Table 3.2 Table indicating the total number of simulations out of 20 in which TgPLP1 APCβ bound to the membrane for each bilayer composition. 56
Table 3.3 Results of transmembrane β-barrel prediction. Residues highlighted in red are within TMH regions. 74
Table 3.4 Results of 5 repeats of PC membrane self-assembly around the final β-barrel model. 80
Table 4.1 PvPLP1 APCβ data collection and refinement statistics..... 102
Table 4.2 Summary of Simulations..... 111
Table 4.3 Summary of PvPLP2 mutations. 124
Table 5.1 Lipid composition of membrane association simulations performed with PFN2 P2 domain 139

Acronyms and Abbreviations

APCβ	ApiPLP C-terminal β-pleated sheet domain
α-PFP	α-pore forming protein
ApiPLP	Apicomplexan perforin-like protein
AT	Atomistic
β-PFP	β-pore forming protein
CD	Circular dichroism
CDC	Cholesterol dependent cytolysin
CG	Coarse grained
CL	Cardiolipin
COM	Centre of mass
CTD	C-terminal domain
CTT	C-terminal tail
EM	Erythrocyte membrane
ER	Endoplasmic reticulum
IMP	Integral membrane protein
LPA	Lysophosphatidic acid
LPS	Lipopolysaccharide

MAC	Membrane attack complex
MACPF	Membrane attack complex/perforin superfamily
MD	Molecular dynamics
MPEG1	Macrophage expressed gene 1
NTD	N-terminal domain
OMP	Outer membrane transmembrane β-barrel protein
PFN2	Perforin-2
PFP	Pore forming protein
PH	Pleckstrin homology domain
PIP	Phosphatidylinositol phosphate
PI(3,4,5)P_3	Phosphatidylinositol (3,4,5)-trisphosphate
PI(4,5)P_2	Phosphatidylinositol (4,5)-bisphosphate
PLP	perforin-like protein
PM	Plasma membrane
PMF	Potential of mean force
PMP	Peripheral membrane protein
PC	Phosphatidylcholine
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
PS	Phosphatidylserine
PPLP	*Plasmodium* perforin-like protein
PtdIns/PI	Phosphatidylinositol
PV	Parasitophorous vacuole
PVM	Parasitophorous vacuole membrane
Ra-LPS	Rough LPS
RBC	Red blood cell

RDF	Radial distribution function
Re-LPS	Smooth LPS
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
TgPLP1	*Toxoplasma gondii* perforin-like protein
TMH	Transmembrane hairpin
TV	Transient vacuole
WHAM	Weighted histogram analysis method

Chapter 1

Introduction

1.1 Membranes and membrane proteins

Biological membranes perform vital membrane boundaries that enable cellular life

A key feature of biological cells is their envelope of membranes surrounding cell contents to achieve separation and protection from the extracellular environment. In order for a cell to function it must keep its contents separate from its environment to efficiently perform the reactions and processes required for life, as well as to exclude environmental toxins that would interfere with those processes. Eukaryotic cells have then evolved to compartmentalise themselves further to form a series of discrete organelles with different molecular identities and functions to cluster together or separate proteins and molecules required for different reactions (1).

Membranes are composed of lipids which assemble into a planar bilayer with an outer extracellular leaflet, and an inner cytoplasm facing leaflet. This is facilitated by the amphipathic nature of lipids, with hydrophilic headgroups facing the surrounding solution, and hydrophobic acyl chains which face into the center of the lipid bilayer (**Figure 1.1**). The resulting bilayer is on average 60 Å thick, with the hydrocarbon core spanning 30 Å, and the polar interfacial regions contributing 15 Å each (2).

The hydrophobic barrier is critical for maintaining the boundary functions of the cell membrane. Small non-polar molecules are able to freely diffuse through membranes, while the membrane is impermeable to larger and polar molecules that must be transported selectively either by active or passive mechanisms to cross the barrier into or out of the cell. Disruption of homeostasis via damage to the barrier function of the membrane, or perturbation of molecular mechanisms facilitating transport, can result in damage to the cell or trigger apoptotic cell death.

In addition to their barrier function, membranes have a variety of other critical roles, including cellular signaling, providing a platform for molecular reactions and to facilitate cell division and other dynamic processes of exchange of material between organelles, cells and the environment. They perform these functions in concert with proteins which are either embedded within the membrane as integral membrane proteins (IMPs) or on the surface of the membrane as peripheral membrane proteins (PMPs).

The membrane contains hundreds of different lipid molecules with different functions

Membranes are composed of hundreds of distinct lipid species, which lend different characteristics and functions to the membrane (3). Different lipid types are not dispersed throughout the membrane evenly, but have characteristic spatial and temporal distributions across the membrane relating to their specific functions (1). Cells utilise large amounts of resources to generate the many different lipid species, and then to maintain their asymmetric distribution across both the different membranes in the cell and across the two bilayers. This indicates the importance of lipids in the function of the cell outside of a simple barrier function, which includes the clustering and regulation of membrane proteins.

Membranes are composed of several lipid types, including phospholipids, which play key structural roles, as well as sphingolipids and steroids. Lipids can define structural properties of the membrane such as bilayer thickness, fluidity and stiffness, act as specific ligands for protein membrane recognition or activation as well as acting in signaling cascades (1, 4).

Phosphatidycholine (PC) is a major structural component of the eukaryotic membrane, along with the phospholipids phosphatidylserine (PS), phosphatidylethanolamine (PE) and phosphatidylinositol (PtdIns), which are enriched in the cytosolic leaflet of eukaryotic plasma membranes (5) (Figure 1.2). Phosphatidylglycerol (PG) is an anionic phospholipid that has a major role in many bacterial membranes, while it is mainly found in mitochondria in eukaryotes.

In addition to its structural role, PE has a conical geometry which causes it to induce curvature in the membrane, particularly in processes involving membrane remodeling, fusion and fission. PS is required for membrane localization by some proteins, with its anionic charge attracting proteins such as the C2 domain of protein kinase C and Raf1, and for activity of other proteins such as the Na^+/K^+ ATPase (4, 6). PS exposure to the outer leaflet of the cell is a key indicator of membrane disruption, with a high density of PS outer leaflet exposure triggering apoptosis (4).

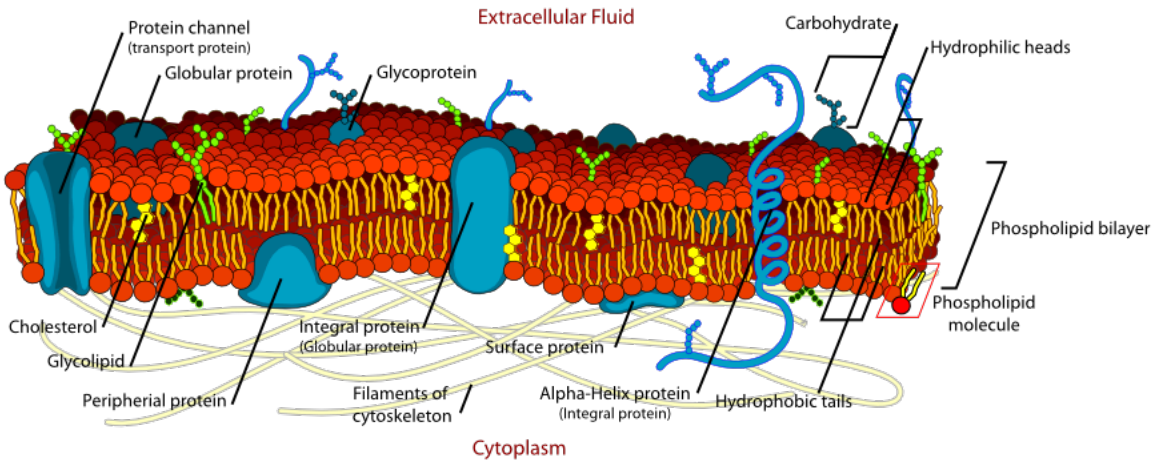


Figure 1.1 The eukaryotic plasma membrane

Representation of the eukaryotic plasma membrane. Lipid headgroups are shown with red spheres while their tails are coloured orange. Various membrane associated proteins are also shown in blue. Image is an original by Mariana Ruiz[a]

a)https://en.wikipedia.org/wiki/Cell_membrane#/media/File:Cell_membrane_detailed_diagram_en.svg

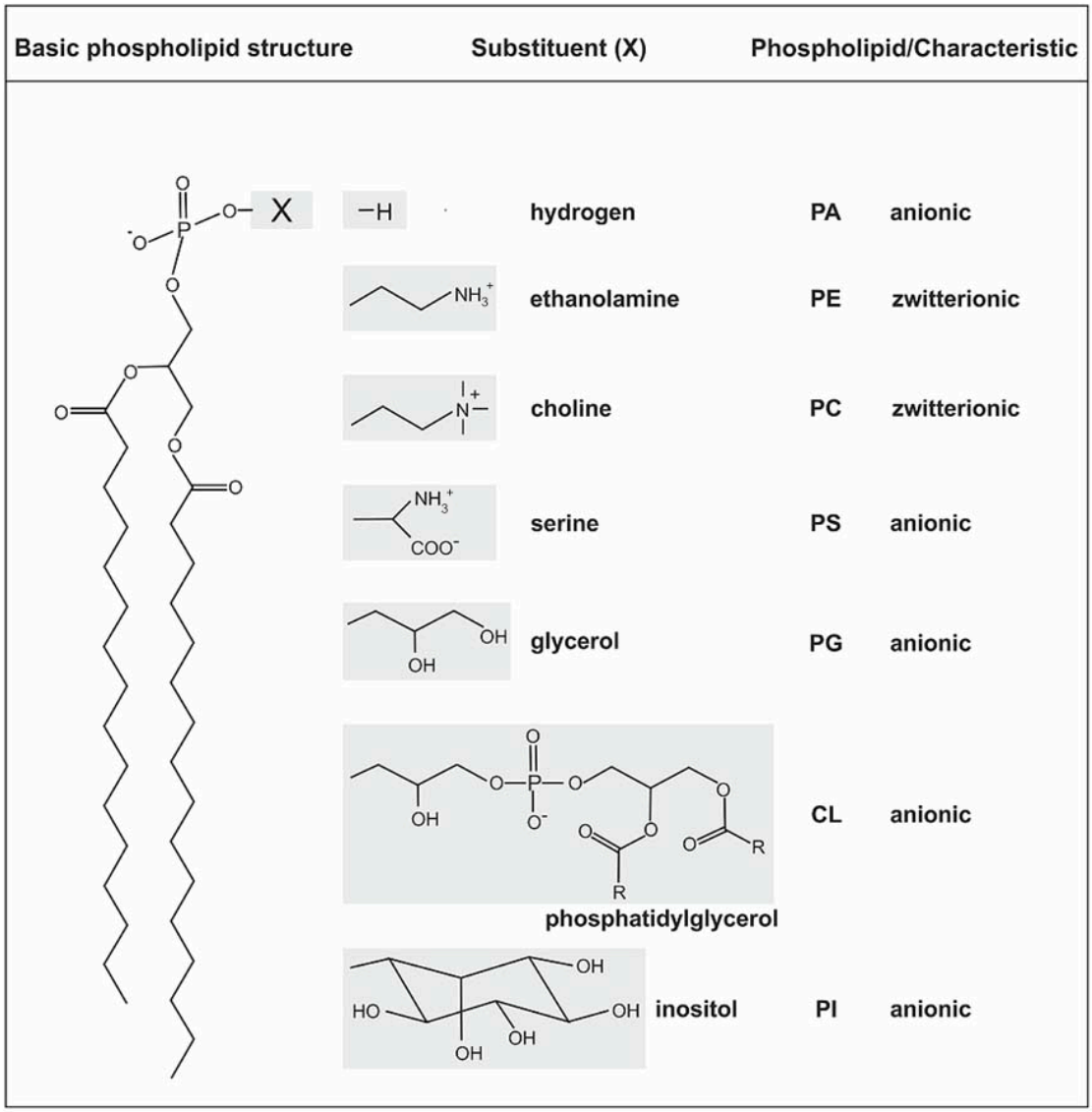


Figure 1.2 Phospholipid structure

The basic phospholipid structure is shown on the left, along with common headgroups on the right, indicating characteristics of each head group. Image taken from (7).

Phosphoinositides can be phosphorylated at various sites on their inositol ring headgroup, forming phosphatidyl-inositol phosphate molecules (PIPs), which can be differentially recognised and so are used to confer organelle identity, and for molecular recognition and recruitment of cytosolic proteins. They are located at the cytoplasmic leaflet of the plasma membrane, and also enriched in endosomes and the trans-golgi network (1, 8).

The lipids sphingomyelin and cholesterol are also important components of eukaryotic membranes. Sphingomyelin is enriched in the outer leaflet of the plasma membrane. It has longer, more saturated acyl chains than PC, allowing it to pack more densely into a gel phase, and along with cholesterol to contribute to the strength and low permeability of the plasma membrane (1, 9). Cholesterol is thought

to segregate with sphingomyelin, which helps to solubilise the more hydrophobic steroid rings of cholesterol. Together these have been implicated in the formation of lipid rafts which are used to cluster particular functional lipids and proteins together (10, 11).

1.1.1 Peripheral membrane proteins

Peripheral membrane proteins (PMPs) are defined by their surface association with membranes. Rather than containing regions that span the hydrophobic core of the membrane, PMPs contain specific patterns on their surface that allow them to associate with lipids, and in many cases they are able to leave membranes transiently and exist as fully soluble proteins. In contrast, integral membrane proteins either fully or partially span the membrane. PMPs have a variety of roles including in lipid transport and metabolism, cellular signaling, and cytoskeleton assembly (12, 13). The transition from soluble to membrane bound means these proteins can be highly dynamic. Binding to membranes is dictated by the membrane environment, and can be influenced by lipid composition, presence of ions such as Ca^{2+}, pH and bilayer characteristics such as curvature. Regulation of binding in this way means these proteins are able to function with temporal and spatial specificity within the cell, such as localizing to specific subcellular regions or responding to specific signaling cues.

Many PMPs contain specialised membrane binding domains, including protein kinase C conserved 1 (C1); protein kinase C conserved 2 (C2); pleckstrin homology (PH); Fab1, YOTB, Vac1 and EEA1 (FYVE); and Bin amphiphysic Rvs (BAR) domains. Other PMPs do not contain a specialised domain and bind to membranes through surface characteristics of the protein, or through covalent attachment to a lipid anchor within the membrane (2). Mechanisms by which these PMPs recognise lipids can be classified into electrostatic interactions and non-specific hydrophobic interactions. Cho & Stahelin proposed that PMPs can be classified into three basic types depending in their lipid interactions (2). Surface, or S-type, PMPs interact only with the surface portion of the membrane, mainly with the lipid headgroups. I-type proteins penetrate into the interfacial region of the membrane, below the level of the phosphate backbone and can interact with the hydrocarbon lipid chains, while H-type proteins penetrate further into the hydrophobic core of the membrane. Each of these interactions is governed by the propensity of surface exposed residues to interact with hydrophobic core and interfacial regions of the bilayer (14).

The classic model of PMP membrane binding suggests that initial association with the membrane is driven by long-range electrostatics between clusters of cationic residues and anionic lipids in the membrane. Membrane association then restricts the dimensionality of binding and increases surface concentration of the protein, allowing it to find more specific electrostatic and hydrophobic interactions for tighter binding (15).

Electrostatic membrane interactions

Initial association of PMPs is often driven by long-range electrostatic interactions between positively charged residues on the protein and anionic lipids in the bilayer. This interaction is non-specific, and once associated with the membranes, specific ligand interactions may form between proteins and lipids in the membrane. These are often driven by specific electrostatic interactions with charged lipids. Several signaling pathways are controlled by phosphorylation of phosphoinositides, which is specifically recognised by a range of PMPs, including the PH domain and the FYVE domain (16, 17).

Hydrophobic protein-lipid interactions

As discussed above, hydrophobic interactions involve partial penetration of the binding protein into the hydrophobic core of the bilayer. These interactions are particularly important in binding uncharged lipids such as PC, and the outer leaflet of eukaryotic plasma membranes (2). Hydrophobic membrane interactions can achieved by PMPs in several different ways, including partial unfolding of the protein to expose buried hydrophobic residues, covalent attachment of lipid anchors, and the use of amphipathic helices and exposed loops (13). Aside from the use of lipid anchors, these interactions are governed by surface exposed hydrophobic residues. The hydrophobic residues tryptophan, phenylalanine, tyrosine and leucine are known to play important roles in hydrophobic membrane anchoring (18). Tryptophan and tyrosine residues are found in high proportion at the lipid-water interface of integral membrane proteins and peripheral membrane proteins, tending to preferentially sit at the boundary between acyl chains and the polar region of the bilayer, about 15 Å from the centre of the bilayer (2, 19). The tendency of these residues to sit at this region, despite their hydrophobicity, is thought to be due to their large disk-shaped side chains whose interaction with the acyl chains of the bilayer may be entropically unfavourable, as well as the presence of a quadrupole moment which may favour the more complex interfacial region (20).

Amphiphathic helices are also often used by proteins to facilitate membrane binding. These typically involve a upper surface of the helix decorated with polar residues, while the opposite face of the helix typically contains hydrophobic residues that interact with the bilayer's hydrophobic core allowing the helix to lay parallel to the membrane surface (21). Examples of the use of amphipathic helices to bind to membranes include antimicrobial peptides (22), as well as the short N-terminal amphipathic helix of the mammalian regulator of G-protein signaling 4 (RGS4) which confers specificity for anionic lipids (23), and the widely conserved bacterial protein MinD which uses a conserved 8-12 residue C-terminal amphipathic helix to recognise the bacterial inner membrane (24). Surface exposed loops are also often involved in penetrating the lipid bilayer and anchoring the protein to the membrane.

1.2 Evolution and molecular functions of pore forming proteins

From single celled prokaryotes up to large mammals including humans, almost all organisms share an early mechanism of cellular immunity and attack. Pore forming proteins (PFPs) form a diverse group of proteins which use a variety of different mechanisms to achieve membrane disruption of target cells. Formation of a transmembrane pore typically, although not always, aims to kill the target cells through disruption of homeostasis or by aiding delivery of further effector molecules into the cell (25). These are used by organisms typically either as a potent toxin, including up to 30% of bacterial toxins (26), or as a component of the innate immune response against invading pathogens. PFPs can be divided into two broad classes: β-pore forming toxins (β-PFTs) and α-pore forming toxins (α-PFTs), which are defined based on formation of transmembrane pores by either beta-barrel like assemblies or alpha-helical assemblies (27).

While the precise mechanism of action varies, PFPs are typically secreted by cells as soluble monomers, which then associate with target membranes, oligomerise and undergo structural rearrangement to insert into the membrane to form a transmembrane pore.

Members of the α-PFTs include the Colicins (28), actinoporins (29), and Diptheria toxin (30). These proteins tend to contain pore forming domains with high α-helical content, have a membrane binding structural motif or domain that mediates initial membrane contact, and at least one α-helix per monomer that inserts into the membrane to form the pore (31). β-pore forming proteins include aerolysin, produced by the *aeromonas* family of Gram-negative bacteria, α-hemolysin produced by *Staphylococcus aureus* (32), the anthrax toxin protective antigen (33) and the membrane attack/perforin MACPF/CDC superfamily (34). As with the α-PFTs, β-PFTs also contain a membrane binding domain. Some bind to specific ligands in the target membrane and oligomerise on the membrane before pore formation. They typically contain either flexible loops or bundles of α-helices that form transmembrane β-hairpins upon pore formation. These hairpins then undergo intermolecular hydrogen-bonding to form large β-barrels spanning the membrane in the final pore structure.

1.3 The MACPF/CDC superfamily of pore forming proteins

1.3.1 The MACPF/CDC Superfamily

The MACPF/CDCs form a large superfamily of β-PFPs that span both eukaryotic and prokaryotic organisms. Cholesterol dependent cytolysins (CDCs) are toxins secreted by Gram-positive bacteria, some of the best studies of which include Listeriolysin O (LLO) produced by *Listeria monocytogenes*, Perfringolysin O (PFO) produced by *Clostridium perfringens*, and Pneumolysin (PLY) produced by *Streptococcus pneumoniae*. CDCs are mostly secreted by extracellular bacteria and act to damage target membranes causing necrosis and tissue death. Several CDCs also function through different

mechanisms, an example of which is LLO which is used by the *Listeria* during intracellular replication to disrupt phagosomal membranes among other functions (35).

The first CDC structure solved was that of PFO, and showed a molecule consisting of two domains (36). The first domain, which was originally identified as three separate domains (Domains 1-3), is responsible for pore formation and oligomerisation, and forms a flat rectangular domain consisting primarily of beta-sheets (Figure 1.3). A central 4-stranded bent beta-sheet is characteristic of these domains, along with two clusters of alpha-helices on either side of the beta-sheet named transmembrane helices (TMH) 1 and 2, which unfurl to form the transmembrane beta-barrel (34). TMH1 is packed into the core of the domain, partially shielded between the central β-sheet and domain 2, while TMH2 is more exposed and packed against the β-sheet. The final Domain 4 (D4) forms a β-sandwich composed of 4- and 5-stranded β-sheets and is responsible for membrane binding.

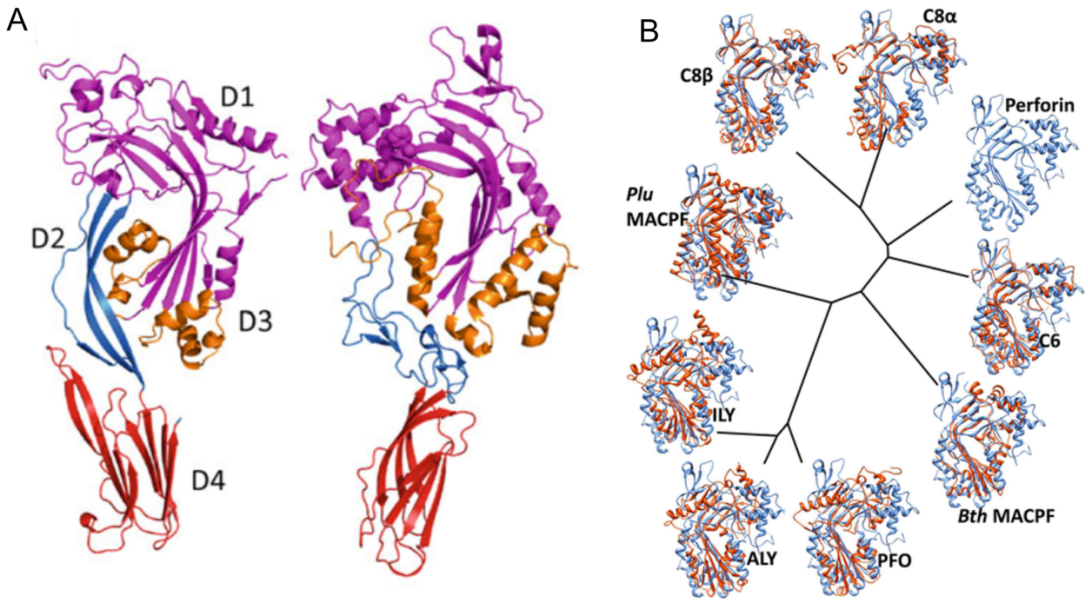


Figure 1.3 Comparison of MACPF and CDC structures

(A) Crystal structure of the CDC perfringolysin O (PDB ID: 1PFO) with each domain labelled (left) and the crystal structure of perforin (PDB ID; 3NSJ) (right). The membrane binding domain is labelled in red in each, linking regions between CTD and MACPF/CDC domain are coloured in blue, the MACPF/CDC domain is coloured magenta, with TMH helices coloured in orange. (B) A structural phylogenetic tree comparing both CDCs and MACPFs. Perforin is coloured in blue, with each additional structure fitted for RMSD and coloured red. Image taken from (37) and reproduced with permission from Springer Nature.

The membrane attack complex/perforin (MACPF) family of pore forming proteins was originally identified through a shared MACPF domain found in components of the membrane attack complex (MAC) and perforin (38). With their roles in the mammalian immune system, the MAC forms heterooligomeric pores to kill Gram-negative bacterial membranes and is composed of proteins C5b-C9 (39). The role of perforin in the immune system is to target and kill virus-infected and tumour cells, forming pores that facilitate the targeted delivery of granzyme molecules into cells and triggering cell death (40). The MACPF domain is widespread throughout all kingdoms of life, and while primarily thought of as a eukaryotic family, is also found in some bacteria including the human pathogen *Chlamydia* and insect pathogen *Photorhabdus* (41). The primary role of MACPF proteins is in cellular attack and defence as described, however some members appear to have a non-lytic role, including Astrotactin which has a role in neural migration and the *Drosophila* gene Torso-like involved in embryonic development (42), while the *Apicomplexan* MACPF family of perforin-like proteins facilitate parasite escape from within cells rather than direct lysis of cells (43).

The first structures of the MACPF domain included a structure of Plu-MACPF from the bacteria *Photorhabdus luminescens* (44) and the C8α component of MAC (45). These structures revealed that the MACPF domain has structural homology with the N-terminal domain (originally, domains 1-3) of the CDCs, in particular they share the central 4-stranded antiparallel beta-sheet with α-helical TMH clusters packed on either side, as well as forming a similar characteristic flat, box-shaped domain (Figure 1.3) (46). These findings suggested that while the two domains do not share sequence homology, they use similar mechanisms for pore formation, including deploying the two TMH bundles to form transmembrane beta-barrels.

1.3.2 MACPF/CDC membrane binding domains

The first stage of pore formation for all PFPs is membrane association, which for MACPF/CDCs is typically achieved using a membrane binding domain (Figure 1.4). The membrane binding stage can perform several functions. Firstly, recognition of specific ligands within the membrane ensures targeting of the PFP to the correct membrane and therefore avoids damage to self. Membrane binding also increases the surface concentration of monomers, therefore facilitating oligomerisation (31), and has been proposed to allosterically cause conformational changes that trigger oligomerisation and pore formation (47, 48).

The membrane binding domain of most CDCs is the C-terminal domain originally named Domain 4 (D4) (Figure 1.3). This consists of a beta-sandwich structure containing several critical membrane binding loops at its base. One of these loops contains a highly conserved tryptophan-rich undecapeptide sequence motif ECTGLAWEWWR (48, 49). This loop is critical in membrane anchoring of D4, containing tryptophan residues that penetrate into the bilayer (50, 51). In this way, D4 acts as a PMP, with hydrophobic loops at its tip facilitating membrane binding. CDCs require cholesterol in the

membrane to recognise the membrane as well as to trigger conformational changes for pore formation (52, 53), which is in high concentration in the outer leaflet of eukaryotic plasma membranes. In order to recognise cholesterol, D4 also contains a conserved Thr-Leu cholesterol-recognition motif (CRM) (53, 54). Some CDCs also require an additional ligand, the human CD59 glycosyl phosphatidylinositol anchored protein in the membrane (55).

In contrast to CDCs, while most membrane binding domains of MACPFs are found at the C-terminal of the protein they are far more structurally diverse. Perforin contains an EGF-like domain followed by a C2 domain at its C-terminal end. This is most structurally similar to the CDC D4 and contains a conserved Trp residue in the equivalent region to the undecapeptide (49), however it lacks the CRM of CDCs, not requiring cholesterol in the membrane to bind. The perforin C2 domain requires a high concentration of calcium to bind, meaning it is activated outside the cell where calcium concentrations are higher as a means of regulation (56, 57), however it is relatively non-specific when it comes to membrane lipid composition (58). Binding of calcium to the C2 domain leads to conformational change that exposes aromatic side chains for efficient membrane binding (59, 60).

Pleurotolysin (PLY), a fungal toxin, is an unusual MACPF in that it is composed of two separate components: PlyA and PlyB (61, 62). This has the effect of separating the MACPF domain and membrane binding domain, with PlyA being a smaller protein with a β-sandwich structure that associates with the membrane and has specificity for sphingomyelin, and PlyB containing the MACPF domain (63). PlyB must associate with membrane bound PlyA for pore formation.

Pore formation of the terminal pathway of the Complement system, formation of the MAC, involves a variation on the canonical mechanism of membrane binding and pore formation. Rather than forming a homo-oligomer, the MAC consists of an initial membrane bound protein complex, containing proteins C5b, C6, C7 and C8 forming the C5b8 complex, onto which multiple copies of C9 attack and oligomerise into a pore. Membrane binding is not achieved using the typical C-terminal domain. Formation of the C5b8 complex occurs in a step-by-step manner, in which C6 associates with C5b, followed by C7 and C8. C7 mediates initial membrane anchoring (64), while C8 contains a MACPF domain, which inserts into the membrane upon association with the C5b67 complex (65). The C5b67 complex contains a hydrophobic site which appears to partially bury into the outer leaflet of the membrane, while further membrane association and stability is lent by C8 and C9 MACPF insertion into the membrane (66, 67).

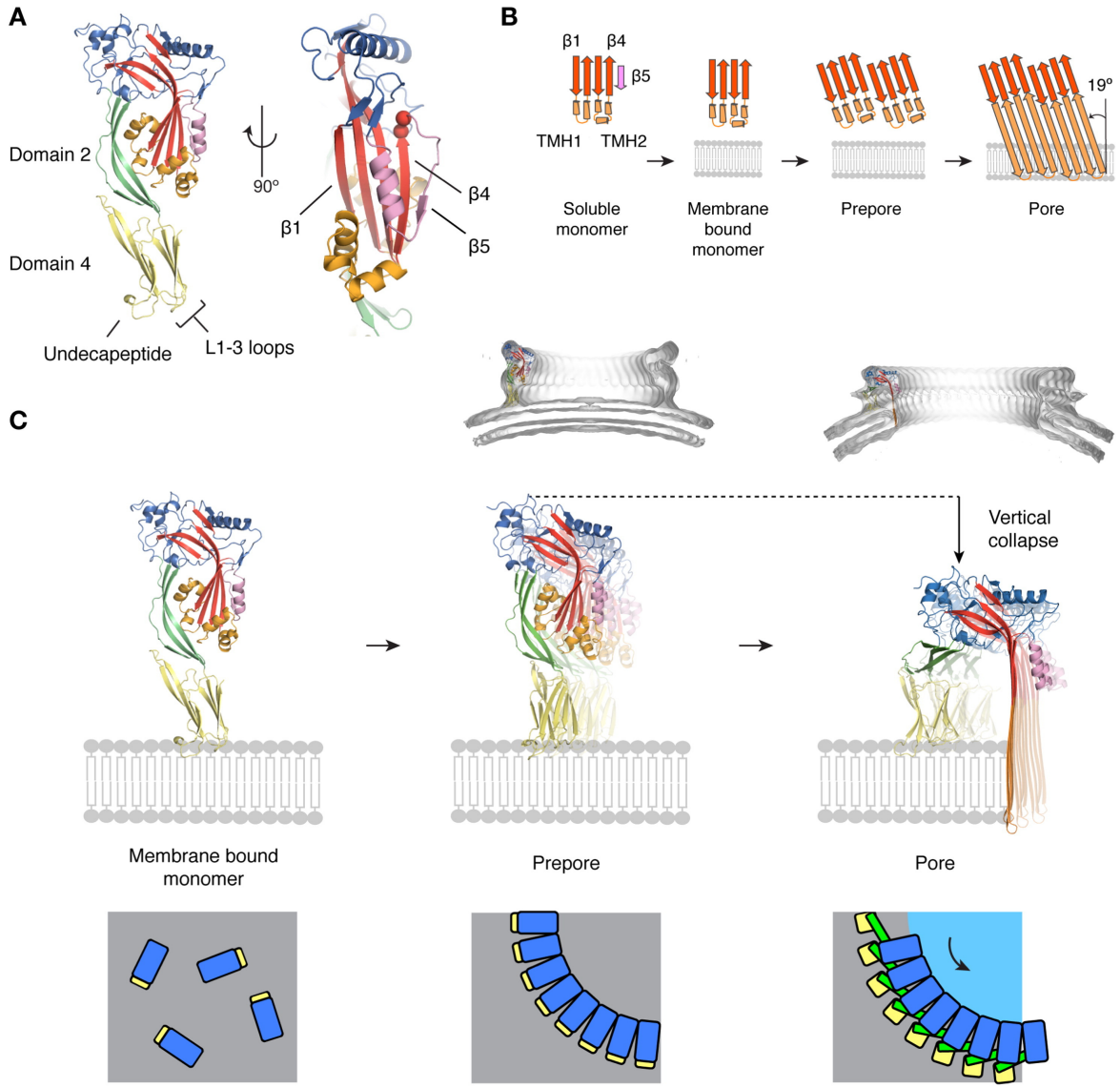


Figure 1.4 Mechanism of CDC pore formation

Schematic showing the mechanisms of CDC pore formation. (A) Crystal structure of PFO (PDB ID: 1PFO) with Domains 2 and 4 labelled. The CTD is coloured in yellow. The MACPF/CDC domain sits atop Domain 2. Its central twisted β-sheet is coloured in red, with flanking TMH helices in orange. α-helices at the top of the domain are coloured blue. (B) Shows a schematic of the MACPF central 4-β-strands (red) and TMH (orange) transitioning from soluble to transmembrane pore. TMH1 and 2 start in an α-helical structure and become transmembrane β-sheets upon transition from pre-pore to pore. (C) The PFO structure is shown initially as a membrane bound monomer (left), oligomerisation then occurs to form a pre-pore structure (centre). Finally, pore formation occurs in which Domain 2 collapses down bringing the MACPF/CDC domain nearer to the membrane, while TMH1 and TMH2 unfurl and insert into the membrane to form a β-barrel structure (right). The panel below shows the same events from a top-down view. Image taken from (68) and reproduced with permission from Elsevier.

1.3.3 MACPF oligomerization and pore formation

Following association with the membrane, oligomerisation and structural rearrangements to facilitate pore formation take place. Currently, the mechanisms of pore formation by CDCs is better understood, with all known CDCs following a similar pathway. Firstly, oligomerisation of monomers occurs. Membrane binding and cholesterol recognition of CDCs transmits structural changes from D4, along D2 and up to the MACPF domain (54, 65). Alongside the central 4 beta-strands of the domain, a fifth capping beta-strand, β5, forms a continuous β-sheet alongside β4 inhibiting association of monomers in solution. The hydrogen bonds between β4 and β5 are disrupted by membrane binding, allowing β5 to rotate away and the free edge of β4 to associate with β1 of an adjacent monomer (69, 70). Two conserved glycines at the turn between β4 and β5 are critical for this motion. CDCs can oligomerise into rings up to 35-50 monomers (71), and evidence suggests they are able to either form a closed circular pore, or incomplete arcs on the membrane, both of which are capable of pore formation (72).

Oligomerisation leads to the formation of a distinct pre-pore structure in CDCs, in which all pore subunits sit atop of the membrane before transition to a transmembrane pore state (73, 74). Once the pre-pore has formed, the two TMH helical clusters deploy from their helical form either side of the central β-sheet and unfurl into amphipathic β-hairpins that insert into the membrane (48, 75, 76). Each monomer contributes two hairpins, which penetrate the membrane in concert and bond together to form the transmembrane β-barrel pore. While the precise mechanism of membrane insertion is not well understood, it is thought that the β-hairpins insert together, possibly forming a β-barrel prior to insertion to overcome the energetic cost of membrane insertion (77). Insertion of TMH segments into the membrane is accompanied by a vertical collapse in CDCs (70, 78). This is achieved by a 90° rotation of domain 2, causing translation of the rest of the MACPF/CDC domain nearer the membrane followed by structural rearrangement of the alpha-helical bundles into the transmembrane beta-barrel (78-80).

While pores are typically visualized as closed circles, negative stain electron microscopy and atomic force microscopy studies of perforin, listeriolysin, perfringolysin amongst other MACPF/CDCs suggest that alongside formation of complete pores, CDC and MACPF proteins are able to form incomplete arcs in the membrane, which are also capable of cell damage (81-83). Size of pores, and formation of arcs over complete pores, appears to be kinetically driven, with low concentration of monomers kinetically trapping arcs which are then able to form functional pores (84).

In contrast to CDCs, less is known about pore formation in MACPFs, the mechanisms of which also appear to be more diverse than CDCs. Due to the structural similarity between the CDC and MACPF pore forming domains, these are now considered to belong to the same family and thus have similar mechanisms of action (44, 82, 85). Perforin forms transmembrane pores and arcs which vary in size, typically between 19 and 24 subunits for closed pores (56, 86). Atomic force microscopy and electron microscopy show that perforin forms pre-pore assemblies before membrane insertion similar to CDCs.

Studies suggest that in contrast to CDC which do not grow once they have formed pores, perforin pore assembly may begin through formation of smaller pre-pores, which insert into the membrane and act as nucleation sites for recruitment of further subunits to increase pore size (86).

Recent structural work has shown that the MAC forms unique asymmetric hetero-oligomeric pores (87). The recent cryo-EM structure of the MAC reveals that after pore formation, the C5b, C6, C7 and C8 subunits form the C5b8 complex of the MAC form a stalk protruding from the pore, while 18 C9 subunits forms the rest of the pore. The C6, C7, C8α and C8β subunits contain MACPF domains, which form an asymmetric segment of the pore (Figure 1.5). The membrane inserted β-hairpins of C6, C7 and C8β are shorter in length than those of C9, with C6 and C7 only inserting into the outer leaflet of the membrane. Together these are proposed to induce curvature in the membrane and disrupting the lamellar nature of the bilayer in a similar mechanism to those used by arcs. This in turn is thought to reduce the energy required for subsequent oligomerisation and provide a platform for subsequent C9 oligomerisation (88, 89). C9 subunits oligomerise after binding to C8α to form the rest of the pore. The poly-C9 portion of the pore is asymmetric in nature, and formation does not appear to involve a pre-pore stage like perforin or CDCs (89). The final C9 subunit does not form significant contacts with C6, meaning that altogether the MAC pore forms an open asymmetric pore that resembles a split-washer (87, 88).

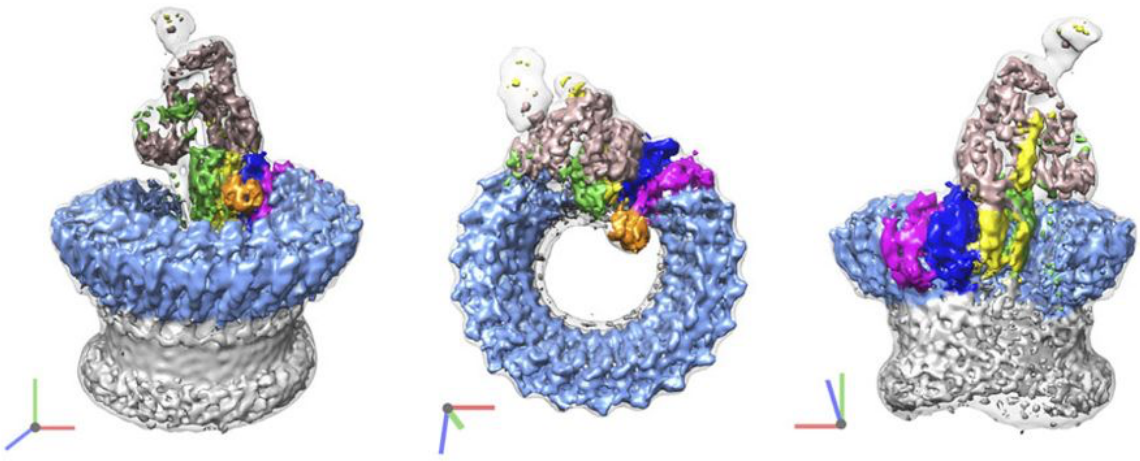


Figure 1.5 The Complement membrane attack complex

A sharpened cryo-EM reconstruction of the MAC showing the split washer pore structure. The map is coloured according to the complement protein. C5b is coloured tan, C6 is green, C7 is yellow, C8α is magenta, C8β in dark blue, C8γ in orange and C9 in light blue. The transmembrane segment is shown below in grey. Image taken from (87).

Altogether, these recent advances have shown that while they share similar domains and mechanisms of pore formation, MACPF proteins differ from CDCs in both in the size of their pores, and kinetics of association and membrane insertion. Another key difference is that while CDCs collapse towards the membrane during transition from pre-pore to pore, all MACPF proteins studied to date do not collapse in the same way, remaining at the same height above the membrane and containing longer TMH segments to achieve membrane interaction.

1.4 The Apicomplexan life cycle and perforin-like proteins

Perforin-like proteins (PLPs) were first identified as a family of MACPF domain containing proteins by Kaiser *et al* in 2004 (90). Many pathogens use PFPs to invade, egress or otherwise manipulate host cell membranes for the purpose of infection and replication. Therefore, Kaiser *et al* reasoned that because *Plasmodium* have multiple complex interactions with cell membranes through cell traversal, replication and egress, it is highly likely that they may use PFPs to facilitate some of these interactions. They identified 5 MACPF domain containing proteins, naming them *Plasmodium* perforin-like proteins (PLPs) 1-5.

1.4.1 Perforin-like proteins

While some apicomplexans with more restricted hosts, such as *Toxoplasma*, which replicate solely within mammals have fewer PLPs, *Plasmodium* has 5, while *Theileria* and *Babesia* have 6 PLPs each (43). The expansion of PLPs among species with more complex lifecycles involving both arthropods and mammals indicates that more PLPs are required to adapt to different parts of the lifecycle.

Each PLP contains 3 domains, and an N-terminal signal peptide (Figure 1.6). The N-terminal domain has the highest variability, varying significantly in length between PLPs, and much of the sequence is predicted to be disordered. The central MACPF domain contains a conserved MACPF motif. Highest conservation is seen at the region predicted to form amphipathic helices which become transmembrane hairpins (TMHs) upon pore formation (90). The recently solved structure of TgPLP1 MACPF domain (91) (also see Chapter 3) is consistent with these findings, and contains the MACPF canonical core central twisted beta-sheet flanked by two pairs of alpha helices. Hidden Markov modelling shows a signature pattern W-x(2)-[FL]-[FI]-x(2)-[FY]-G-T-H-x(7)-G-G shared by ApiPLP MACPF domains which is similar to the canonical MACPF signature (43). Conservation of the shared ApiPLP MACPF signature motif and cysteine residues indicate that the MACPF domain structure will be conserved among *Plasmodium* PLPs also.

The C-terminal domain contains a triple repeat of a conserved cysteine pattern and beta-sheet content, indicating a shared domain structure. Secondary structure prediction indicates a repeat of 5-stranded beta-sheets, leading to its designation as the ApiPLP C-terminal β-pleated sheet (APC-β) domain (43). Originally each of the three repeated sequences were designated as separate domains, however the first

structure of the APCβ domain showed that they are arranged in a novel prism structure forming one compact domain with highly hydrophobic contacts buried in the centre of the domain between each repeat (91). A homology search using the APCβ sequence identified no homologous sequences to the C-terminal domain outside of Apicomplexans indicating this domain is unique to the apicomplexan PLPs.

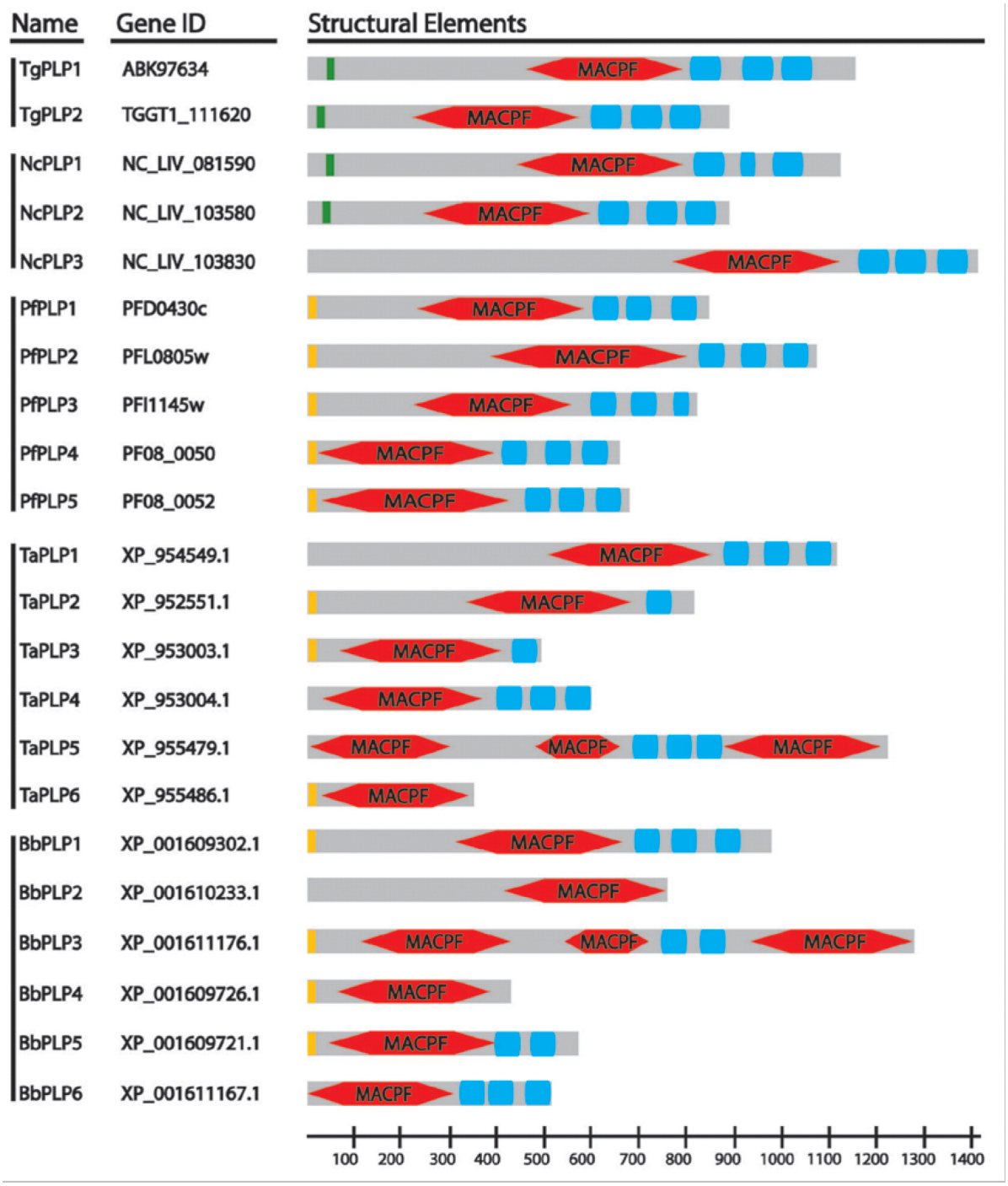


Figure 1.6 Apicomplexan PLP domain architecture

Domain architecture of apicomplexan PLPs from Toxoplasma gondii, Neospora caninum, Plasmodium falciparum, Theileria annulate and Babesia bigemina. Predicted Signal peptides are shown (yellow) along with signal anchors (green), MACPF domains (red), and APC-β pseudo-repeats (blue). Image taken from (43).

1.4.2 Apicomplexans

Apicomplexans are a phylum of eukaryotic protozoan parasites that are important to human health and agriculture. Parasites within this phylum include human infectious agents *Plasmodium* spp. and *Toxoplasma gondii* which cause malaria and toxoplasmosis respectively. Other lesser known members include *Babesia* spp. which causes infections in cattle and opportunistic infections in humans, cattle-infecting *Theileria* spp., *Eimera* spp. which infects both cattle and chickens, and *Cryptosporidium* which infect animals and causes opportunistic infections in humans.

These single-celled parasites have complex lifecycles, which involve both an asexual and sexual stage of replication. While lifecycle stages vary considerably between organisms, the asexual stage of infection typically involves haploid parasites called sporozoites that are highly motile and invasive. This is the form of parasite that is often transmitted between hosts and is able to rapidly invade and replicate within new hosts. At some point, a subset of parasites differentiate into sexual stage male and female gametes, which fuse together to form diploid zygotes. The sexual stage of replication often occurs in a different organism, such as mosquitos and ticks, that act as vectors of the disease. Diploid zygotes rapidly develop and differentiate back into haploid sporozoites to continue to invade new warm-blooded hosts.

Apicomplexa are obligate intracellular parasites, and their replicative cycle is characterised by cycles of cell invasion, replication and lysis, before re-invading new cells. This lysis causes tissue damage which is the main cause of disease across the phylum. Sporozoites are often the main invasive and replicative stage of parasites. Sporozoite motility allows them to quickly move through cells, termed cell traversal, to travel from point of host entry to tissue in which they can replicate. They then invade cells, termed cell invasion, and replicate within them. Replication occurs within a Parasitophorous vacuole (PV). The PV is a non-phagosomal membrane bound compartment that is formed during parasite invasion of host cells. The PV membrane is derived from the host plasma membrane, however it is extensively remodeled by the parasite during growth and replication (92, 93).

Morphology of the different lifecycle stages of apicomplexa can vary, however the motile sporozoite stage has a characteristic elongated shape. A specialised apical region is found at the tip of the cell, which contains several specialized organelles including secretory organelles named rhoptries and micronemes. These contain proteins that are critical for cell motility, adhesion and invasion of host cells. They are also triggered during cell egress following replication, and there is a growing appreciation of the complexity and diversity of different secretory organelles involved in these different processes (94). The cell membrane of Apicomplexa also varies from other eukaryotes. The plasma membrane forms part of a cellular envelope named the pellicle, in which the PM is closely apposed to an inner membrane complex.

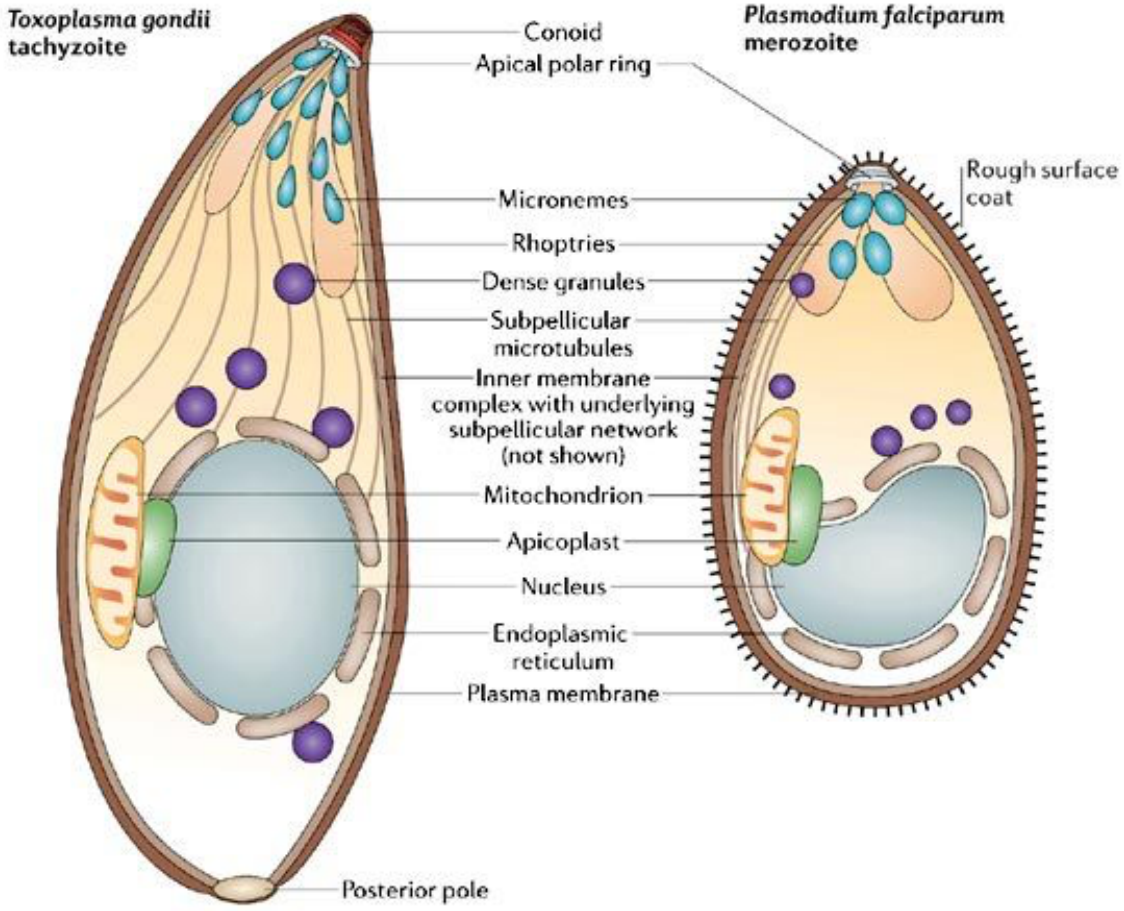


Figure 1.7 Apicomplexan cell morphology

Morphologies of Toxoplasma gondii tachyzoite (left) and Plasmodium falciparum merozoite (right) showing an elongated cellular shape tipped with an apical conoidal structure used for cell attachment and invasion. Several types of secretory vesicles are organised around the apical conoid including micronemes, rhoptries and dense granules. Image taken from (95) and reproduced with permission from Springer Nature.

1.4.3 Toxoplasma

The *Toxoplasma gondii* apicomplexan causes toxoplasmosis in humans. Toxoplasmosis can cause serious acute disease, including up to 24% of all food-related deaths and costing nearly $3 billion in the USA alone (96, 97). Toxoplasma has two PLPs (TgPLP1 and TgPLP2), however TgPLP1 is the best studied and so far only copy known to be expressed. Disruption of TgPLP1 has been shown to lead to a fivefold reduction of toxoplasmosis in mice (98), indicating it could be an effective target for vaccination or treatment. In addition, *Toxoplasma gondii* is the model organism for the apicomplexan family (99). Development in understanding TgPLP1 role in egress could impact our understanding of PLPs across the Apicomplexa phylum.

Toxoplasma lifecycle and toxoplasmosis

Toxoplasmosis is contracted through ingestion of *T. gondii* contaminated food or water. Oocysts originate from the definitive feline host following a sexual lifecycle phase in the small intestine and can pass on infection from any water or food in contact with feline faeces. The asexual stage of infection on the other hand can occur in almost any vertebrate tissue, resulting in cysts that can be transmitted in raw meat (100). In the digestive system sporozoites from oocysts, and bradyzoites from tissue cysts, are released and penetrate the intestinal epithelium, where they differentiate to tachyzoites and spread throughout the body (Figure 1.8).

During acute infection tachyzoites replicate in many tissues throughout the body, before inducing a strong immune response resulting in a self-limiting infection. The immune response promotes differentiation to bradyzoites. Bradyzoites encyst themselves within host cells and enter a dormant, largely benign, chronic stage of infection. Impairment of the immune system can then allow re-activation of the infection and differentiation back to tachyzoites resulting in a rapid replication and a potentially fatal infection. A particular characteristic of *T. gondii* is its ability to cross the blood:brain barrier, forming cysts within the brain. Re-activation of these cysts and subsequent replication within the brain can result in development of multiple abscesses causing toxoplasmic encephalitis.

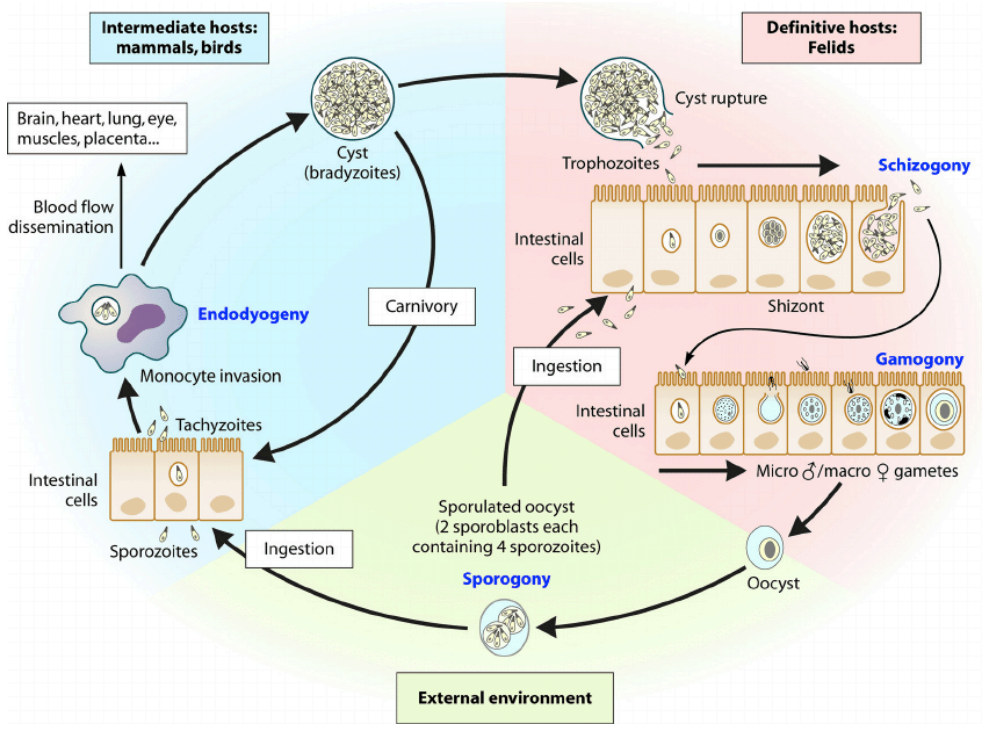


Figure 1.8 The Toxoplasma gondii lifecycle

The Toxoplasma gondii lifecycle is shown in both the definite feline host as well as the intermediate host that can include humans. Image taken from (101) and reproduced with permission from the American Society for Microbiology.

Replication and the lytic cycle

The *Toxoplasma* lytic cycle is the fundamental disease process required to perpetuate the *Toxoplasma* lifecycle. The lytic cycle comprises three steps; invasion, replication and egress. Parasites enter host cells through cell attachment, invasion and formation of the PV. Once inside the PV, membrane remodelling is performed along with recruitment of host cell ER and mitochondria to facilitate access to nutrients required for replication (102). The parasite replicates via endodyogeny, with synchronous replication of daughter cells until a critical mass is reached and egress, escape from the PV and host cell, is initiated.

Cell egress is initiated by environmental cues that are sensed by the parasite, including acidification of the PV, accumulation of abscisic acid released by replicating cells, depletion of ATP, and calcium flux (100, 103, 104). Cell signalling then triggers secretory processes which activate cell motility and release many effector molecules, including TgPLP1, into the PV. Egress then occurs through sequential perforation of the PV membrane (PVM) and host cell plasma membranes allowing parasite escape. Tachyzoites are then free to disseminate through the host tissue, quickly invading adjacent cells and initiating further cycles of replication (103).

The mechanisms involved in cell egress and membrane perforation are many and complex. While there had been a greater focus on mechanisms of parasite egress in recent years, the complex interplay of different factors, and signalling cascades required, are still not well understood (102). Disruption of the PVM and host PM are clearly both critical factors allowing parasite release from the cell, and TgPLP1 has been found to be a vital effector molecule for membrane perforation. Other mechanisms involved in egress include parasite motility and secretion of other effector molecules. Parasite transition from the intracellular replicative stage to egress and the extracellular stage involves activation of gliding motility. It is hypothesised that as well as promoting movement out of the damaged cell, this motion also mechanically aids in disrupting the PVM and PM, with mutations in gliding machinery blocking escape (105-107).

1.4.4 The Plasmodium lifecycle

In comparison to the *Toxoplasma* lifecycle, the *Plasmodium* lifecycle is more complex, involving replicative stages in multiple tissues across both a mammalian and arthropod host. Symptoms of malaria in humans are a result of replication within the red blood cells (RBCs), which destroys cells and causes cyclical flu-like symptoms including fever, muscle pains and headaches (Figure 1.9). The noticeable symptoms occur after an incubation period of 7-18 days post infection during which the parasite travels from an injection site in the skin towards the liver for an initial stage of replication, after which it enters the blood stage of infection.

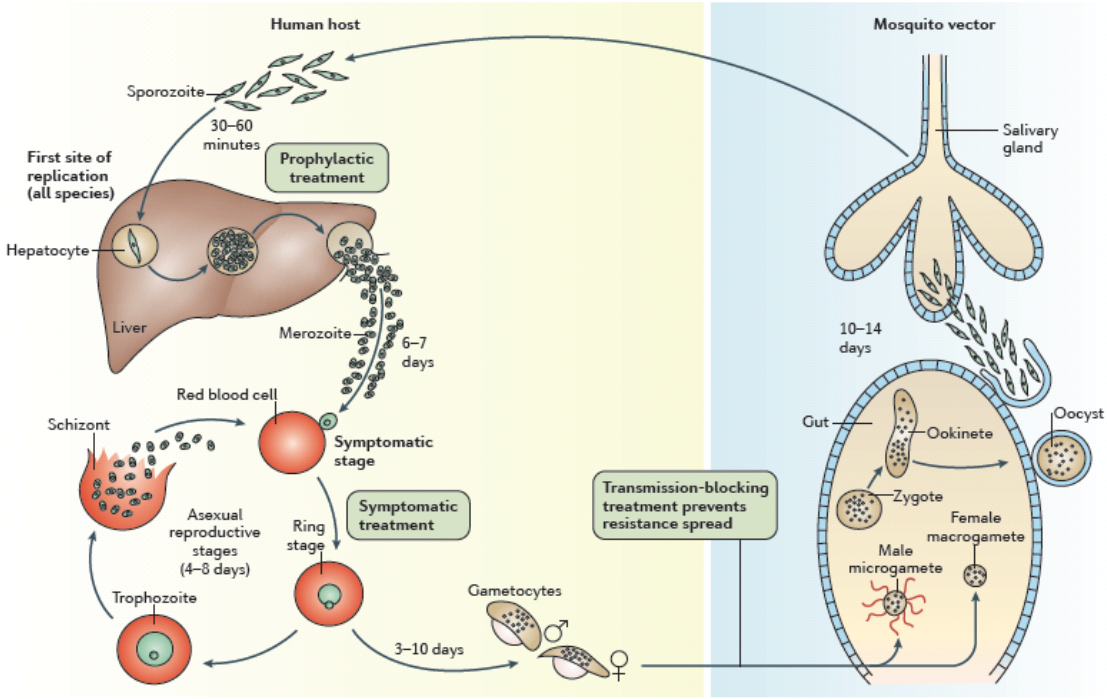


Figure 1.9 The plasmodium lifecycle

Schematic showing the lifecycle of the Plasmodium parasite across both the mosquito vector and human host. Sporozoites are transmitted from the mosquito salivary gland to human dermis during a blood meal. Traversal from the dermis to liver then occurs, followed by replication within liver hepatocytes, release and infection of erythrocytes follows. Finally, a proportion of merozoites differentiate to gametocytes in the blood and are taken up by the mosquito where they undergo sexual replication in the mosquito midgut. Image taken from (108). Image reproduced with permission from Springer Nature.

Plasmodium PLP1 facilitates cell traversal in the dermis and liver

Infected female *Anopheles* mosquitos inject *Plasmodium* sporozoites into host dermis during blood feeding. This motile sporozoite stage goes on to replicate within the host liver before releasing thousands of erythrocyte infecting merozoites. The first process that sporozoites must accomplish is the traversal across host tissue to gain access to hepatocytes in the liver. Cell traversal involves parasite entry into a cell, followed by traversal across the cell cytoplasm and escape through the opposing membrane (109). PLP1 is critical for traversal through the dermis (110, 111), and liver sinusoidal membranes (112, 113). PLP1 has also been implicated in escape from the RBC during blood stage infection (114), however subsequent reports however have not identified any PLP1 expression during blood stage infection (113).

Erythrocyte cell egress in the mosquito midgut is mediated by Plasmodium PLP2

Following release from the liver, thousands of merozoites then travel into the blood stream to infect erythrocytes. Merozoites mostly replicate asexually within RBCs, where they divide within a PV then escape, bursting the RBC, and go on to re-infect another erythrocyte. A small proportion of merozoites differentiate into the sexual stage of replication, forming male and female gametocytes within red blood cells.

Following a blood meal and ingestion by the *Anopheles* mosquito, gametocytes egress from the RBC within the mosquito midgut. This egress is facilitated by PLP2 and is triggered by environmental cues such as a drop in temperature within the midgut (115, 116). Egress follows an 'inside-out' model, involving activation of gametocytes in the mosquito midgut followed by rupture and escape firstly from the PV membrane followed by the RBC membrane (117, 118).

PPLPs 3, 4 and 5 act in cell traversal across the mosquito midgut

Following escape from blood cells in the midgut, gametocytes form a zygote which develops into the ookinete stage which, once formed, rapidly traverses the mosquito midgut. On the basal side, ookinetes then form oocysts ultimately dividing into new sporozoites for another round of mammalian infection. The final PLPs, PLPs3, 4 and 5, all act at this stage, facilitating traversal across the mosquito midgut wall.

1.5 Perforin-2

1.5.1 The role of perforin-2 in the immune system

Mammalian immune systems have evolved multiple mechanisms of defence against invading pathogens, including several MACPF proteins each targeted towards a different range of pathogens. Natural killer and T-killer cells of the immune system secrete perforin-1, which attacks virally infected and tumour cells. Perforin-1 forms pores allowing non-selective transport of water and ions as well as granzymes

into the cell, disrupting cellular homeostasis and inducing apoptosis (119, 120). Perforin-1 is critical to the immune system, with some mutations affecting perforin-1 function causing profound immune deficiencies or dysregulation. The membrane attack complex forms part of the Complement system and is the second MACPF complex of proteins within the immune system. The Complement system is tailored to target extracellular gram-negative bacteria. Recently, a third pore forming protein of the human immune system was discovered named perforin-2 (PFN2) (121). Found in both professional and non-professional phagocytes, PFN2 is able to target and kill intracellular bacteria, which cannot otherwise be reached by the MAC.

Given their non-specific ability to attack pathogens, pore forming proteins evolved as an early innate mechanism of cellular defense. The widespread existence of PFN2 across both invertebrates and vertebrates reflects the ancient origins of this protein (122). Phylogenetic studies show that perforin, perforin-2 and MACPF complement proteins are related, arising with the evolution and increasing complexity of the immune system. Studies suggest that perforin arose from an early duplication of PFN2, indicating that PFN2 is the more ancient of the two (123).

PFN2 was originally named macrophage expressed gene 1 (MPEG1) as it was identified within macrophages, however it is now known that it is expressed by both professional phagocytes including macrophages as well as in many other parenchymal cell types (124). Expression of PFN2 in professional phagocytes is constitutive, while expression in other cell types is only induced upon exposure to bacterial patterns or cytokines (124). PFN2 then mediates intracellular killing of bacteria through formation of pores on the bacterial membrane, and is critical for killing of a range of bacteria including methicillin-resistant *Staphylococcus aureus* (MRSA), *Salmonella typhimurium*, *Chlamydia trachomatis* and *Listeria monocytogenes* (124-126).

Uniquely for MACPF proteins, PFN2 exists as a transmembrane protein, containing a MACPF domain, C-terminal P2 domain, the transmembrane segment and then a short cytosolic domain. Upon cellular infection, the proposed orientation for PFN2 suggests that the MACPF domain faces the endosomal or phagosomal lumen where it would be exposed to phagocytosed bacteria (124). Disruption of the PFN2 gene in mice has been shown to dramatically increase susceptibility to certain infections in mice including MRSA and *Salmonella* (124). Less is known about the effects of mutation in humans, however several point mutations have been identified that increase susceptibility to pulmonary non-tuberculous mycobacterial infections (127).

1.5.2 The Bacterial cell walls

Perforin-2 has broad spectrum bactericidal activity, killing both Gram-negative and Gram-positive bacteria. In contrast to eukaryotic membranes, bacterial cells have a more extensive protective cell wall around an inner membrane that is more similar to the eukaryotic plasma membrane. Gram negative and Gram-positive bacteria differ primarily in the type of cell wall (Figure 1.10). The Gram-positive

bacterial cell wall consists primarily of a thick layer of peptidoglycan that lends strength and shape to the cell. Peptidoglycan consists of long strands of sugar chains, interlocked by short peptide bridges. The sugars are either N-acetylglucosamine or N-acetylmuramic acid, and together with peptide bridges form a very strong matrix-like structure that can withstand high pressure while allowing diffusion of small molecules across. Embedded within the peptidoglycan in Gram-positive cells are teichoic acids. Teichoic acid is a glycopolymer, consisting of polyribitol or polyclycerol phosphate. Teichoic acid may either exist embedded at the surface of the peptidoglycan wall, or as lipoteichoic acid which are connected to the cell membrane with a lipid anchor. These contribute to the negative charge of the cell wall and contribute to its overall strength. Underneath the peptidoglycan layer exists the bacterial cell membrane, existing as a typical phospholipid bilayer, containing primarily PG, PE and cardiolipin lipids in a composition of roughly 70% PE, 20% PG and 10% Cardiolipin (128).

Gram negative bacteria in contrast contain two cell membranes, an inner and outer plasma membrane sandwiched around a peptidoglycan layer in the periplasmic space between the membranes. The inner membrane is similar to that of the Gram-positive cell wall. In this case however the peptidoglycan is much thinner, while the outer plasma membrane has a distinct composition from the inner membrane. The outer membrane is highly asymmetric in nature. The inner leaflet contains mostly PG and PE, similar to the inner membrane, while the outer leaflet is composed mostly of lipopolysaccharide (LPS). LPS varies with the type of bacteria, but typically consists of a lipid A molecule made of two glucosamine sugars that has 6 acyl chains that anchor it in the bilayer, and two phosphate groups, one on either sugar. The lipid A is then attached to the core polysaccharide. The core polysaccharide is separated into the inner and outer core, consisting of a minimally-branched chain of sugars that is reasonably uniform in character between bacterial species (Figure 1.11). Phosphate groups bonded to core sugars are thought to contribute significantly to the stability of the outer membrane by forming electrostatic interactions and attracting divalent cations (129).

Finally, the most exposed segment of LPS is the O-antigen. This is a further branched sugar structure extending from the polysaccharide core which accounts for most of the heterogeneity of LPS between species. Different species produce distinctive O-specific chains, which can differ in their length, branching and composition.

Proteins found in outer membrane are categorised into two groups. Outer membrane transmembrane β-barrel proteins (OMPs) form β-barrels in the outer membrane and have multiple important functions including in nutrient transport, cell adhesion and cell signaling. An additional category of proteins includes lipoproteins tethered to the membrane.

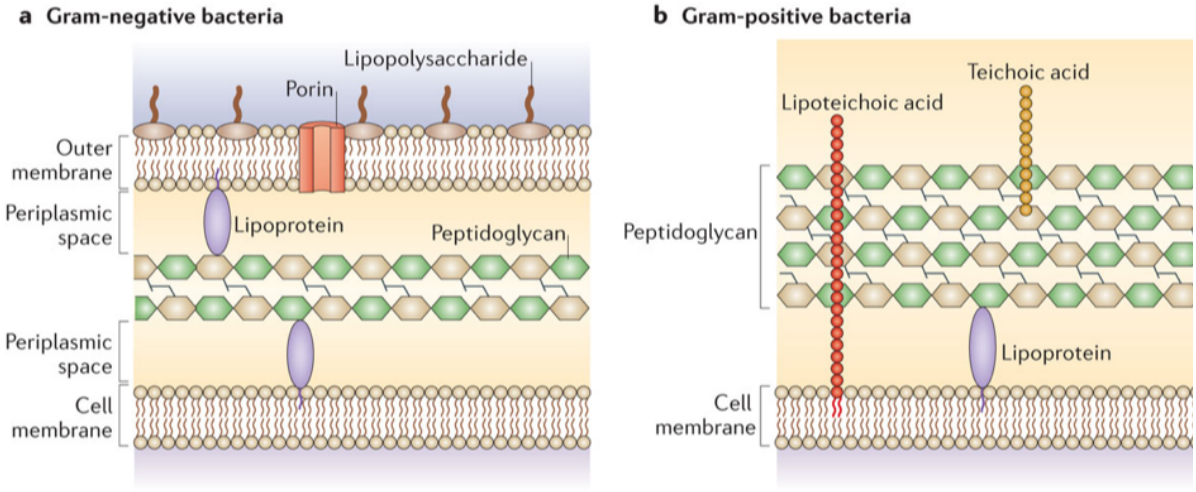


Figure 1.10 Gram-negative and Gram-positive bacterial cell walls

Gram negative bacterial cell walls contain two membrane layers, sandwiching peptidoglycan between, along with a diverse lipopolysaccharide layer on the outer membrane. In contrast, Gram-positive bacteria have a single cell membrane, protected by a thick peptidoglycan wall. Image from (130) reproduced with permission from Springer Nature.

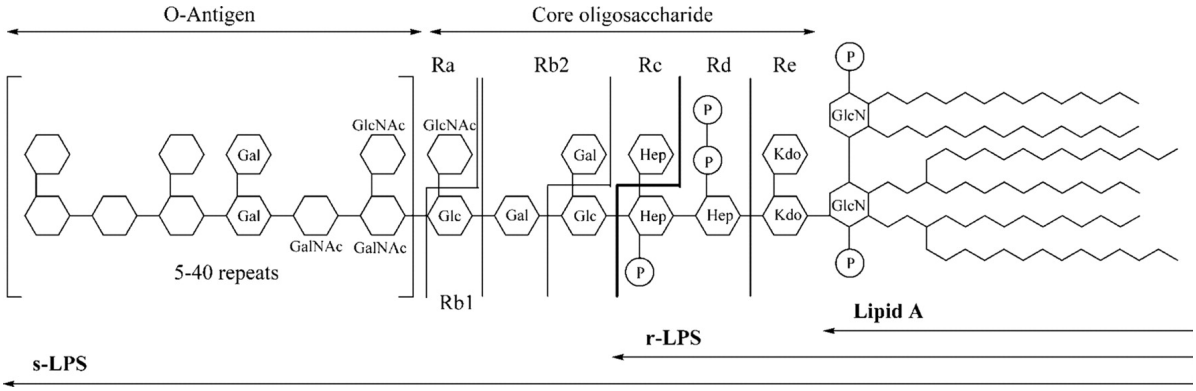


Figure 1.11 The structure of LPS

Diagram showing the general structure of E coli LPS. Ra to Re indicate different variations of rough LPS, whereby only the core oligosaccharide and not O-antigen exists, and the core antigen is shortened by a number of sugars. The core oligosaccharide is well conserved in E. coli, while the O-antigen varies. GlcN: glucosamine*, GalNAc: N-acetyl* glucosamine*, Glc: glucose, Gal: galactose, Hep: L-glycerol-D-manno-heptose, Kdo: 2-keto-3deoxyoctonic acid, P:phosphate. Image taken from (131) and reproduced with permission from the American Society for Microbiology.*

1.6 Thesis Outline

This thesis primarily addresses the use of molecular dynamics (MD) and structural experimental approaches to investigate membrane interactions of MACPF pore forming proteins. In Chapter 2 the practical applications and theoretical underpinnings of the computational approaches used herein are discussed, including the use of multi-scaled MD to investigate lipid-protein interactions.

Chapters 3, 4 and 5 then go on to investigate several MACPF proteins. The apicomplexan MACPFs TgPLP1 and *Plasmodium* PLPs 1 and 2 are studied in chapters 3 and 4 respectively. Work previously done in the Gilbert lab to solve the structures of TgPLP1 MACPF and APCβ domains is continued with a detailed study of the dynamics of membrane binding and lipid interactions to address the initial stage of pore formation of TgPLP1. Crystal structures were also simulated to investigate oligomerisation of TgPLP1 and predict the it's transmembrane β-barrel pore structure, aiming to produce a more complete model of pore formation in Chapter 3.

In Chapter 4, the X-ray crystallography was used to solve the structure of the *Plasmodium* PLP1 APCβ domain, which was used to generate a homology model of PPLP2 APCβ for comparison. Multi-scaled simulations were performed to assess membrane binding and lipid specificity of these proteins to investigate functional differences between *Plasmodium* PLPs.

The final results chapter applies these approaches to the recently solved structures of perforin-2 from the Gilbert group, to address the questions of PFN2 membrane association, oligomerisation and functional states. To conclude, Chapter 6 follows with future perspectives and further questions and challenges to this work on apicomplexan and immune system pore forming proteins.

Chapter 2

Theory and Methods

2.1 Molecular Dynamics

2.1.1 Molecular dynamics and membrane proteins

Molecular dynamics (MD) simulations use a classical mechanics approach to give life to biochemical structures. MD takes into account the forces between individual atoms in a structure and calculates the interactions between those atoms over time to simulate an overall picture of the dynamics of the molecule. While structural and biophysical approaches are able to elucidate a protein's structure and molecular functions across a larger timescale, computational techniques can access narrower timeframes on the picosecond to microsecond timescale, at an atomistic level of detail. These resolutions are difficult to achieve using alternative methods and may be improved further by the use of a combination of atomistic and coarse-grained simulations. Coarse-grained simulations reduce the level of structural detail in order to achieve timescales of multiple microseconds of simulation, in many cases this is long enough to observe larger scale conformational changes and assess the dynamics of protein-lipid interactions in membranes.

MD has been particularly valuable to the study of membrane proteins where it is able to provide visualization of protein interactions and dynamics with lipids in their natural membrane environment that are difficult to study using alternative experimental methods. MD has been used to study protein-lipid interactions of both transmembrane and peripheral membrane proteins. Examples of study of peripheral membrane proteins include studies of PH domains (132-134) and C2 domains (135, 136). Simulations were used to identify the role of non-specific electrostatic protein-lipid interactions in driving the GRP1-PH domain towards membranes before binding its $PI(3,4,5)P_3$ ligand (137). These simulations identified contacts between $PI(3,4,5)P_3$ and the PH domain that were comparable to those seen in X-ray structures. Further MD work on this system has used potential of mean force (PMF) calculations to calculate the free energy of binding of the PH domain to PIPs (132), while further simulations have identified several distinct PIP binding sites within the family of PH domains, and identified instances of binding of multiple PIPs to a PH domain (138).

2.1.2 Molecular Dynamics simulations

In order to study biological molecules using MD, a representation of the atoms within those molecules, and the bonds between them, must be produced. Starting from a crystal structure which provides the locations and identities of the atoms within a molecule, the first step to produce an MD simulation involves applying a suitable forcefield to that molecule.

In order to represent the motions of the molecule, the forcefield is used to define the physical relationships between its individual atoms to calculate how they will move together. To allow simulation of biologically relevant molecules across useful timescales, atomistic simulations treat each atom as a single particle. In contrast to quantum-mechanical simulations in which electron spin within

an atom is simulated, the explicit motion of electrons in these simulations is ignored and the Born-Oppenheimer approximation is used to assume that electrons adjust instantly to nuclear motion. With this assumption, single particles can be treated using classical mechanics to solve Newton's law of motion.

The forcefield breaks down interactions between particles in the system into several components (Figure 2.1). These include the stretching of inter-atomic bonds, bending of angles between bonds, torsional rotations and non-bonded interactions. Together these functions describe the molecular forces within the system and their changes in energy as the system moves.

Terms for bond length and angles are both treated using harmonic potential, treating the bond more or less as a simple spring which inherits an equilibrium bond length/angle, and requires exponentially more energy to deviate further away from this minimum energy structure. Torsional terms differ, in that as a bond is rotated the energy varies sinusoidally, having minima at several points. This is usually expressed as a cosine series expansion.

Two terms are typically used for non-bonded interactions; one for van der Waals interactions and one for Coulombic interactions. Van der Waals interactions are typically modelled using the Lennard-Jones potential, while electrostatic interactions follow Coulomb's law, generally reducing in energy at an inverse power of distance up to a certain cut-off distance beyond which interactions are not considered.

Forcefields are parametrised based on their intended applications, and many more or less specialised forcefields exist for different uses. As mentioned above, atomistic forcefields are often used in MD simulations of biological molecules, and these consider each atom of the molecule to be a separate particle and consider the bonds between them. Common atomistic forcefields include AMBER (139), GROMOS (140), OPLS (141) and CHARMM (142). In this thesis, atomistic simulations were performed with Amber 99SB-ILDN for chapter 3 and 4 (143), and GROMOS96 53a6 for chapter 5 (140).

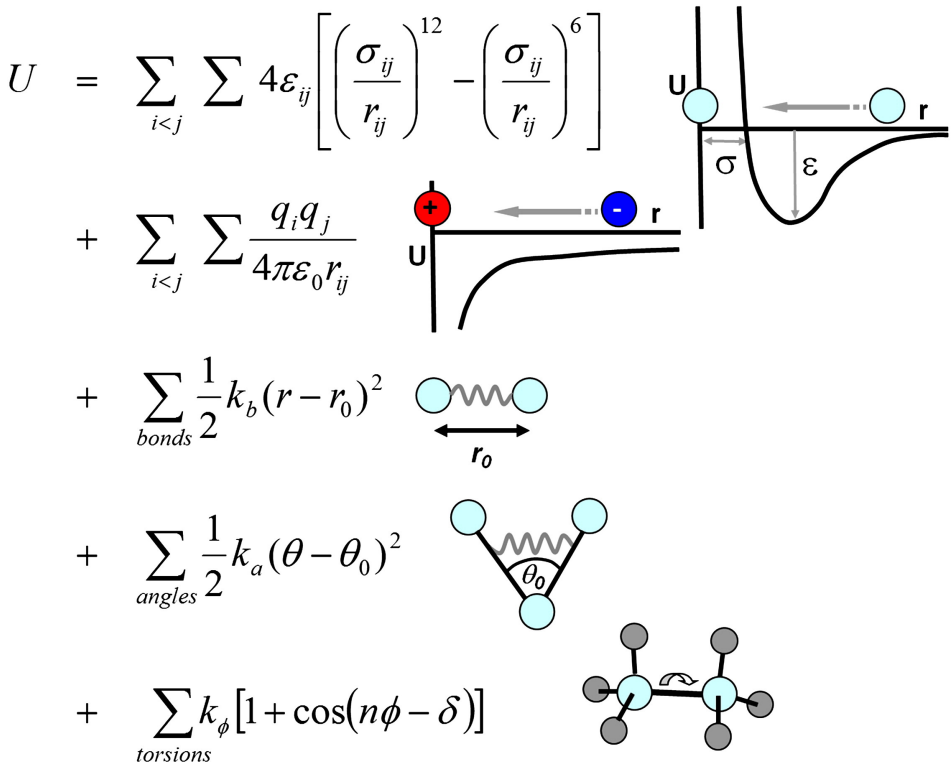


Figure 2.1 A typical potential energy function for biomolecular force fields.

The forcefield is broken down into 5 main components describing the bonded and non-bonded interactions between particles in the system. The first two terms here represent non-bonded Van der Waals and coulombic interactions, with i and j representing two particles. Next, bond stretching is considered, followed by angle bending and torsional rotations. Image taken from (144) and reproduced with permission from Elselvier.

Each forcefield has a set of atom types that it then assigns to each atom within a system. These vary depending on the specific forcefield; however, they typically take into account carbon hybridization states, protonation state as well as other factors in the atom's environment. These provide information on the expected bond lengths and angles. Simulation of water constitutes a special case. Due to the polar nature of water and its unusual properties, it has proven a challenge to generate a water model that accurately represents the physical properties of the molecule. Most biological simulations are performed in water, meaning that a large number of molecules must be simulated and therefore computational efficiency must be a primary consideration. While more complex models exist which account for the polarizability and flexibility of water molecules, in this thesis we use the SPC water model which assigns three partial charges to the water molecule, and treats the molecule as a rigid body (145, 146).

Once the system is defined, MD calculates the trajectory of the system by applying Newton's laws of motion to calculate the velocities of particles within the system. The force on each particle is calculated based on its bonded and non-bonded interactions with nearby particles and using Newton's second law of motion. This represents a many-body problem as each particle is coupled together to make the

calculation exponentially more complex. Once the forces between particles are calculated, acceleration of each particle can be determined to produce a new configuration. The LINCS algorithm was also applied to constrain bond lengths and angles (147).

Successive configurations of the system are calculated, with each frame in the trajectory representing a fixed time step. In the work described in this thesis time steps for atomistic simulations were 0.002 ps, and atomic coordinates were written to the trajectory every 10 ps.

Several different software packages are available to produce MD simulations. Gromacs v5.1 was used in this thesis (148, 149). Several further simulation parameters can be set that further define the temperature, pressure and interactions of particles. Coulombic forces and the Lennard-Jones potential decay exponentially with distance between particles. Therefore, to reduce computational burden, a cut-off distance beyond which electrostatics and Van der Waal's forces are not calculated is defined. Here the cut-off of 1 nm was used for both calculations. Coulombic interactions were also treated using Particle-Mesh Ewald (PME) electrostatics (150).

These simulations are performed with a constant number of particles, and under constant pressure and temperature (NPT) conditions as this is most relevant to experimental conditions. Constant temperature is achieved here using velocity rescaling. This couples the system to an external heat bath and re-scales velocities of particles within the system to simulate kinetic interactions and exchange of energy with an external environment (151). Protein, lipid and solvent (water and ions) were separately coupled to an external heat bath in these simulations, and the temperature was maintained at 310 K.

Pressure was maintained at 1 bar using isotropic pressure coupling. The Parrinello-Rahman barostat was used for atomistic simulations, allowing the simulation system to vary in shape and size to accommodate changes in pressure, and a coupling constant of 1 ps was used (152).

Energy Minimization

Following set up of a system, before simulation, an energy minimisation step is performed to optimize inter-atomic bonds to find an arrangement corresponding to a stable set of energy minima and reduce steric clashes. The steepest descent method of energy minimisation was used here. Given the coordinates of the system, the energy of the system is calculated including the force on each atom. Each atom is then moved a distance corresponding to the magnitude of that force, and the energy is calculated again. This process is repeated iteratively over a number of steps, aiming to find a local energy minimum as a stable starting point for simulations.

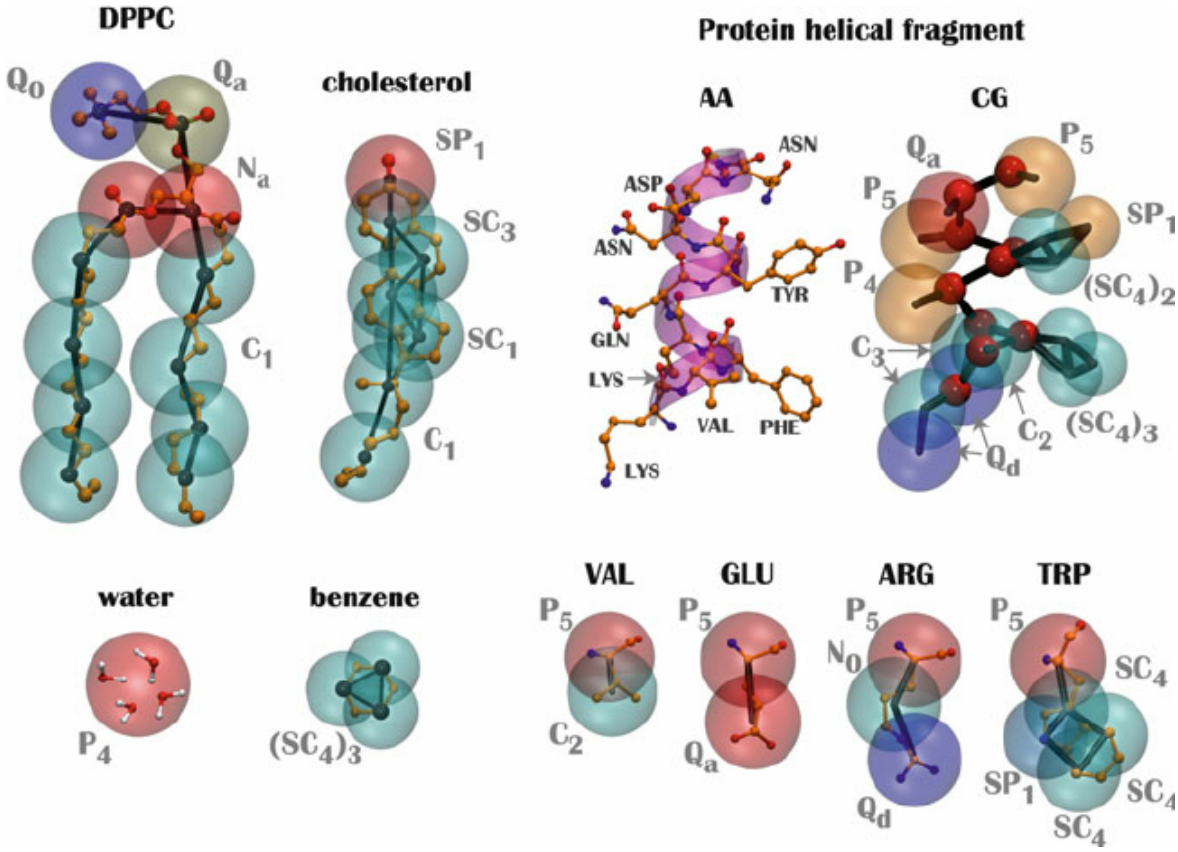


Figure 2.2 Martini forcefield coarse grain mapping

The chemical structure of several molecules is shown in atomistic resolution, with coarse grained beads mapped on top as transparent spheres. Colours of beads represent different particle types. The protein helical fragment AA and CG representations are shown side by side for clarity. Image taken from (153) and reproduced with permission from Springer Nature.

Equilibration

Once energy minimized, the systems were then equilibrated. A short simulation was performed in which the positions of heavy atoms were restrained using a force constant of 1000 kJ mol^{-1} nm^{-3}, allowing the solvent to equilibrate around the protein and lipid molecules.

Restraints were then removed for final production runs, which unless stated otherwise, were produced in triplicate, with each simulation started with random velocities.

2.1.3 Coarse grained simulations

Atomistic simulations produce an accurate representation of the motions of biological molecules, however coarse-grained approaches have been developed as more complex systems and larger molecules are now being explored. Atomistic MD simulation of such systems would require either a system size or timescale unfeasible given current constraints on computational power available. In order to overcome this problem, coarse grained forcefields have been developed that reduce the resolution of

the system to access much larger timescales, and allowing quick and easy production of many simulations. A popular coarse grain (CG) forcefield for biomolecules is the Martini forcefield, v2.2 which was used in this thesis (154, 155).

To reduce system complexity, the Martini forcefield uses one particle to represent four heavy atoms and their corresponding hydrogens. Four different particle types are defined based on polarity, each of which can have subtypes of particles depending on hydrogen bonding capacities (Figure 2.2). Mapping onto proteins, one bead represents the protein backbone, while different beads are then used to represent the different amino acid side chains. Four water molecules are represented by one particle, making water particles larger in Martini systems but further reducing computational cost.

Due to the reduced representation of molecules in CG, protein secondary structure is predefined based on a starting atomistic structure to preserve its higher order structure. To achieve this, extra harmonic bonds are placed between backbones of nearby non-bonded particles. This forms an elastic network around the protein and can be parametrized during set up to define the strength of the harmonic bonds, and the cut-off distance beyond which they are not applied. In this thesis, the elastic network applied used a force constant of 1000 kJ mol^{-1} nm^{-2}. The lower elastic bond cut-off used was 0.5 nm, and the upper cut-off was 1.0 nm. Lennard-Jones interactions were shifted to zero between 0.9 and 1.1 nm. The electrostatic potential energy was shifted to zero between 0 and 1.1 nm

To set up an atomistic simulation, the protein atomic structure is first converted into a Martini CG representation using the Martinize script publicly available from Martini (156). Water and ions were added to solvate the system, at an ion concentration of 150 mM NaCl. The system was kept at a constant temperature of 310 K coupled using velocity rescaling (151), with protein, water and ions, and lipids coupled separately. The pressure was coupled semi-isotropically using the Berendsen algorithm (157), maintained at 1 bar with a compressibility of 5 x 10^{-6} bar^{-1}. Semi-isotropic coupling maintains the pressure separately for *xy* and *z* dimensions of the simulations box which is useful for membrane simulations and reduces distortions in the bilayer. The coupling constant for temperature and pressure was 1 ps and 4 ps respectively. The time-step for integration was 20 fs, and particle coordinates were written to the trajectory file every 200 ps.

Multiscaled simulations of membrane proteins through conversion from coarse grained to atomistic

Multiscaled simulations of large systems have proven extremely effective at generating accurate system conformations and enabling study of large-scale dynamics over longer timescales followed by a more detailed molecular analysis. This approach is often used for study of integral and peripheral membrane proteins. In these cases, the structure of the protein is rarely solved with an intact membrane, meaning accurate knowledge of the membrane and lipid interaction site is not always known. In coarse-grained simulations, a protein can be placed near its predicted binding position and simulated to allow

exploration of binding interactions before a stable system is generated. This can then be converted to an atomistic system and simulated again in detail (158, 159).

The protein atomic structure is first converted into a Martini CG representation, and for peripheral membrane proteins (PMPs), the protein can then be allowed to associate with a membrane to identify its binding orientation, as described under membrane association simulations below. Further CG simulation can be performed to assess membrane binding, lipid interactions and membrane dynamics, which required longer timescales to allow full diffusion and equilibration of the membrane system. Once an interesting interaction is identified, or if the protein is in a stably bound state, the system can then be converted back to atomic resolution using a fragment-based approach and further simulated using AT simulations (158).

In comparison to other methods of predicting a protein's orientation within its membrane environment such as OPM (160), this method includes explicit lipid molecules to more accurately model interactions, and can be modified to include different lipid species depending on the proteins subcellular localisation and function. This takes into account charge differences in lipid species and hydrophobic matching between bilayer thickness and protein hydrophobicity in the case of integral membrane proteins (IMPs). Detailed interactions and lipid modification of protein structure and function can then be studied using both CG and AT simulations.

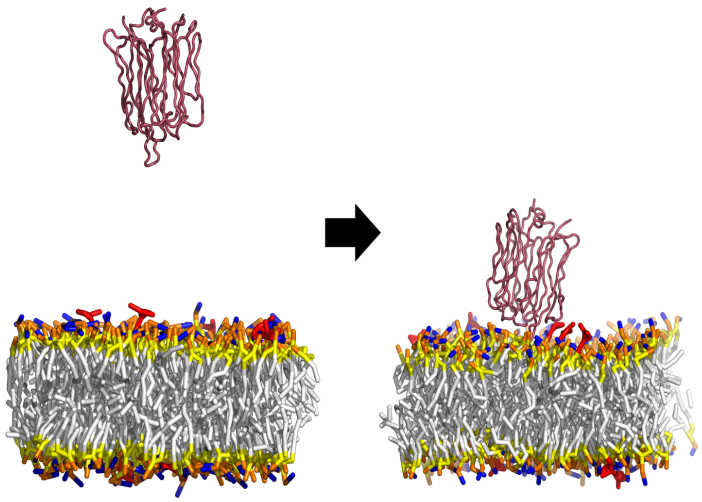


Figure 2.3 Membrane association simulation setup

The protein is placed 10 nm away from a preformed bilayer (left panel) and allowed to freely associate with the membrane over the course of a 1 μs simulation (right panel). A cartoon representation of TgPLP1 APCβ is shown, with lipids shown as a stick representation. Lipid tails are coloured in grey, while glycerol is coloured in yellow, phosphate in orange and PC headgroups in blue. PIP headgroups are coloured in red.

2.1.4 Simulating bilayers, lipids and protein-lipid interactions

Membrane association simulations.

High throughput coarse grained membrane association simulations were performed here to investigate membrane binding of peripheral membrane proteins. This method was adapted from the method used in Yamamoto *et al*. where the binding of pleckstrin homology (PH) domains with PIP lipids in the membrane was studied (133). This study found good agreement between experimental and computational methods, and additionally identified non-canonical binding of some PH domains that had been identified by other experimental means.

The protein was placed 10 nm away from a preformed bilayer and 20-25 simulations were set up in which the protein was rotated sequentially around its *x*, *y* and *z* axis to reduce directional bias. Each was then simulated for at least 1 μs to allow the protein to associate with the membrane (Figure 2.3). Length of simulation and membrane compositions are reported alongside results for each ensemble.

Typically, analysis for each ensemble of association simulations involved production of a 2-dimensional histogram plotting distance between the protein residues and membrane to determine binding orientations. Lipid contacts were also calculated, in which the frequency of contacts between lipids and protein residues were calculated for each simulation and normalised to produce a fingerprint plot of lipid contacts. Contacts were defined as any lipid and residue within 0.6 nm proximity. This distance was reduced to 0.4 nm in the case of lipid contact calculations in atomistic simulations.

Membrane self-assembly

To identify the membrane bound position of integral membrane proteins, a standard approach has been to use prediction software such as OPM to predict the orientation of the protein within a membrane, and then insert the protein structure into a membrane and delete any interfering lipids to produce a starting configuration for simulation. Similar to the above methods, this method has the disadvantages of not explicitly representing lipids and oversimplifying the membrane. Here a CG membrane self-assembly method was used in which lipids are randomly placed around a protein, and a simulation is run to allow the lipids to form a membrane around the protein (Figure 2.4) (161). Simulations are typically run for 100 ns however membrane assembly usually occurs within 10 ns. This method allows both structural changes in the protein to settle into the membrane, as well as changes in the bilayer to adapt to the protein such as induction of membrane curvature, changes in thickness and density of lipids. This method has proven successful in accurately predicting membrane interactions of IMPs, and has been further utilized to generate an automatic pipeline for generation of membrane inserted IMP structures for deposition in the MemProtMD database (162, 163).

Umbrella sampling and potential of mean force calculations

The free energy differences between two states can be calculated using the umbrella sampling technique (164). In this thesis the free energy of binding of MACPF/CDC proteins to membranes is calculated to determine membrane affinity under different conditions. To efficiently calculate the free energy difference between membrane bound and unbound states, a pathway, or reaction coordinate, between these two states must be defined. This parameter can take various forms, including the distance or angle between the two states, here being defined as the distance between the centre of mass of the protein and membrane. Sampling can involve one or more reaction coordinates, however here only one reaction coordinate is defined, resulting in a 1-dimensional potential of mean force (PMF) being calculated for each binding event.

Efficient sampling of phase space must be carried out along this reaction coordinate to ensure that both low energy events as well as high energy, rare events are sampled evenly. Various sampling methods can be used (165), however here we used Umbrella sampling, an accelerated sampling method based on equilibrium properties of the system, in which an additional energy term is applied to ensure efficient sampling. The reaction coordinate path is split into windows, each of which is sampled in a separate simulation. A bias potential is applied, usually in the form of harmonic restraints, to maintain the system near to the reaction coordinate reference point for that window (166). Windows are defined along the reaction coordinate to allow a degree of overlap between each window.

Once all simulations are produced, probability distributions are calculated for each of these windows, combined, and the bias potential removed to calculate the overall potential of mean force of the system. The PMF represents the negative logarithm of the probability of being at any point along the reaction coordinate and is more or less considered to be equivalent to the free energy of binding. To combine the Umbrella sampling windows, the weighted histogram analysis method (WHAM) was used, implemented in GROMACS as g_wham, to compute final PMF values (167, 168) .

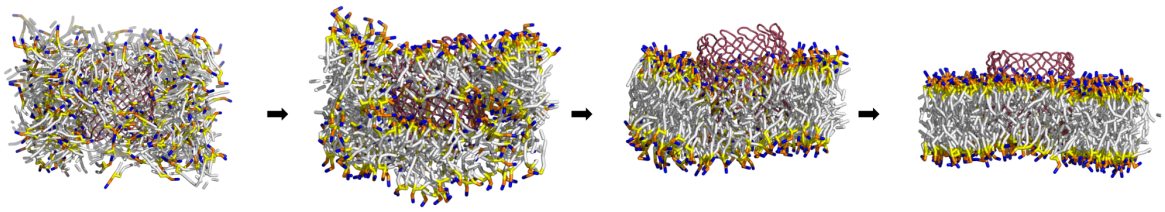


Figure 2.4 Membrane self-assembly

Lipids are placed randomly around a protein structure (left panel) and allowed to self-assembly into a bilayer during the course of a 100 ns simulation. A cartoon representation of a beta-barrel protein is shown, with lipids shown as a stick representation. Lipid tails are coloured in grey, while glycerol is coloured in yellow, phosphate in orange and PC headgroups in blue.

All Umbrella sampling was performed using Martini CG simulations and produced using GROMACS. To set up the system, a steered MD simulation was used to pull the protein from a membrane bound conformational state along the membrane normal. A series of overlapping windows along this reaction coordinate were then defined, typically every 0.1 nm, and each window was simulated for 1000 ns. Lipids were fixed along the Z-component of the simulation using a force constant of 10 kJ mol^{-1} nm^{-2}. The centre of mass of the protein relative to the membrane was restrained in each simulation with a force constant of 1000 kJ mol^{-1} nm^{-2}. The PMF was calculated using GROMACS g_wham, with error estimates obtained using bootstrap analysis. All values were shifted so that the PMF value was 0 in bulk water. Convergence was judged based on the change in PMF value calculated across consecutive 100 ns time intervals.

Data analysis and visualisation software

Analysis of simulation data was performed using a combination of GROMACS software (169), the MDAnalysis library (170) , and custom Python scripts. Visualisation of protein structure was carried out using the Pymol molecular graphics suite (171). Data is presented primarily using the Matplotlib Python plotting software (172) and the Seaborn Python data visualization library (173).

2.2 Protein Production and X-ray crystallography

Protein expression and purification of PvPLP C terminal domains

Expression and production of plasmodium PLP proteins was carried out using transient expression in a mammalian expression system, followed by purification and crystallization. Full length PvPLP1, 2, 3 and 5 were PCR amplified from *Plasmodium vivax* genomic DNA. The PvPLP1 gene contains 8 exons so a synthetic gene containing only exons and optimised for human expression was ordered (Thermofisher, GeneArt).

The gene regions of interest were PCR amplified and cloned using the restriction cloning method into the pHLSec plasmid (174), which appends an N terminal signal peptide to the recombinant protein, and a hexahistidine tag to the C terminal. The recombinant genes were verified by DNA sequencing. PvPLP1 and PvPLP2 constructs were expressed by transiently transfecting HEK 293T cells using polyethylenimine as described in by Aricescu *et al*. Kifunensine was added in order to modify the glycans into a homogeneous form to simplify structural studies.

Five days post transfection media containing the secreted recombinant protein was removed from the cell culture and centrifuged at 5000 rpm for 30 minutes. The media were filtered using 0.22 μm Amicron filters and dialysed into 10 mM sodium phosphate dibasic, 1.5 mM sodium phosphate monobasic, and 300 mM NaCl buffer using bucket dialysis at 4 °C overnight.

The media were filtered again, and 5 mM Imidazole was added. The protein was loaded onto a nickel affinity column (Histrap HP 5ml. GE healthcare) equilibrated with 20 mM Tris, 500 mM NaCl, 5 mM Imidazole pH 7.5. The column was washed with 100 ml of the same buffer after media were loaded. PLP was eluted from the column using a gradient of elution buffer, 20 mM Tris, 500 mM NaCl, 500 mM Imidazole pH 7.5, from 0% to 100% over 60ml.

The NetNGly 1.0 server (175) revealed two potential glycosylation sites. A deglycosylation test was performed to experimentally confirm results. 10 µl of 1 mg ml^{-1} of endoglycosidase F1 was added to 50 µl of protein from the nickel column. After 20 minutes this sample was analysed using SDS-PAGE to detect a shift in molecular weight. As deglycosylation was required, the proteins were deglycosylated overnight by addition of 300 µl 1 mg ml^{-1} endoglycosidase F1 to the protein and incubated at 4 °C. The proteins were further purified using size exclusion chromatography. A HiLoad 26/60 Superdex 75 prep grade column (GE Healthcare) was used for purification of PvPLP2 CTD, while a HiLoad 16/60 superdex 75 prep grade SEC column (GE Healthcare) was used for purification of PvPLP1 CTD. The columns were equilibrated, and protein eluted, using 20 mM HEPES pH 7.5, 150 mM NaCl buffer. Collected protein was analysed using SDS- PAGE and concentrated to around 20 mg ml^{-1} using 10 kDa molecular weight cut-off filters (Amicron) in preparation for crystallisation.

PvPLP1 APC-β crystallisation, data collection and structure determination

Crystals were grown using the sitting-drop vapour diffusion method of crystallisation. 100 nl of protein (20 mg ml^{-1} in 20 mM HEPES pH 7.5, 150 mM NaCl) was mixed with 100 nl reservoir solution into sitting drops and incubated to equilibrate at room temperature. Crystallisation trials of both PvPLP1 and PvPLP2 CTD were set up.

Crystals of PvPLP1 CTD grew after about 20 days incubation in 20 % PEG 3350, 0.2 M Tri-potassium citrate. This condition was used to perform optimisation screens, again using the sitting-drop vapour diffusion method. A dilution series of this condition was created using 100-78% made in 2 % increments across the 12 columns of a 96-well crystallisation tray, and from 100-56 % made in 4% increments per well across another tray. Three duplicate rows of each were set up in this way, and a different ratio of protein to reservoir liquid was used to set up drops in each row. Ratios used were 1:1, 2:1 and 1:2, with 1:1 being 100 nl protein followed by 100 nl reservoir liquid.

Resulting crystals were cryo-cooled in liquid nitrogen using 20% (v/v) glycerol as a cryoprotectant. The crystal diffraction data was collected on beamline I24 at Diamond Light Source (Didcot, UK). Molecular replacement was used to determine the structure using TgPLP1 APC-ß as a model, using phenix.phaser (176) followed by phenix.refine (177) for refining the model and manual reconstruction in Coot (178).

2.3 Biophysical techniques

Thermofluor assays

Thermofluor experiments were conducted in which 4μg of the protein sample was mixed with SYPRO orange stain (Life Technologies S-6650) with a final concentration of 3x. This sample was distributed across a range of pH from 4.0 to 9.5 with increments of 0.5, in 200 mM NaCl, and in serial concentrations of NaCl (0 mM, 100 mM, 200 mM and 500 mM) at pH 7.5. Fluorescence was determined using a real-time PCR machine (179), and was measured every 1 °C while temperature was ramped from 25 °C to 98 °C.

Circular dichroism

Circular dichroism (CD) was used to determine secondary structure composition of proteins. CD spectra were calculated with the protein sample at 100 μg ml^{-1} concentration in 10 mM potassium phosphate buffer at pH 7.5. A cuvette with a 1 mm path length was used.

Thermal stability was measured by measuring the spectra while temperature was increased from 20 to 80°C with steps of 2°C. The wavelength was maintained at 220 nm. Secondary structure composition was also calculated using CD. Temperature was maintained at 20°C while the wavelength was increased from 195 to 250 nm with an incremental increase of 0.5 nm. All CD experiments were conducted using a Jasco J-815 Spectropolarimeter in the Molecular Biophysics Suite, Department of Biochemistry, University of Oxford.

2.3.1 Lipid binding assay

Lipid dot blots were used to assess lipid binding of PPLP APCβs. Membrane lipid strips from Echelon Biosciences (P-6002), which contain 15 important membrane lipids, were used for this assay. Membrane were first incubated in PBST (PBS plus 0.1% Tween20), then blocked using PBST with 3% Non-fat dry milk for 1 hour at room temperature. The strips were then washed and incubated overnight with 0.5 μg/ml protein in PBST + milk buffer at 4°C. Membranes were washed 5 times with PBST for 10 minutes each time, then incubated for 1 hour with rabbit anti-penta-Histidine polyclonal antibody. Membranes were washed again as before, then incubated for a further hour with anti-rabbit-HRP conjugate. A final washing step was performed, and imaging was performed to detect lipid-bound protein.

Chapter 3

Toxoplasma gondii perforin-like protein 1

3.1 Introduction

3.1.1 *Toxoplasma gondii* is the causative agent of toxoplasmosis

Toxoplasma gondii is a global parasite with over 6 billion people infected worldwide. Although infection in healthy individuals is often asymptomatic, Toxoplasmosis can develop in the immunocompromised and through congenital transmission resulting in severe, acute symptoms including neurological and visual impairment and encephalitis, sometimes resulting in death (180). Human infection is contracted through ingestion of parasite oocysts or cyst contaminated raw meat and is characterised by an acute active phase of parasite replication and expansion followed by a chronic, almost unidentifiable phase of very slow replication within tissue cysts. Replication is accomplished by a lytic cycle of cell invasion, replication and escape, or egress. Once inside the cell, parasites replicate within a Parasitophorous vacuole (PV), forming one of two bounding membranes that they must disrupt to escape from cells following replication. Studies into cell egress have fallen behind those of cell invasion as they utilise similar cellular processes of microneme secretion and motility, however appreciation of the differences between the two has recently spurred investigation into egress specific mechanisms.

3.1.2 TgPLP1 facilitates membrane disruption during cell egress

A range of effector proteins are secreted by apicomplexans to facilitate membrane rupture and cell egress, which are thought to be further complemented by activation of cell motility to allow the parasite to rupture both the PV membrane and host cell plasma membrane to allow escape (94, 181). The perforin-like protein TgPLP1 has been identified as a key effector molecule in facilitating egress. Recombinant TgPLP1 was shown to lyse red blood cells and host cells (182), while gene disruption caused defective egress due to significant reduction in both PV and plasma membrane disruption (98). No egress was observed when knockout of TgPLP1 was combined with parasite immobilisation, suggesting TgPLP1 and motility are primary factors in facilitating egress through membrane rupture (98).

3.1.3 The MACPF and C-terminal domain structures of TgPLP1

TgPLP1 is a member of the MACPF/CDC family, containing a central MACPF domain, an N-terminal domain (NTD) and C-terminal membrane binding domain. It is thus thought to facilitate membrane disruption through the characteristic oligomerisation and pore formation of the MACPF protein superfamily. Similar to other members of the MACPF family, the C-terminal domain (CTD) of TgPLP1 is responsible for initial membrane binding (182). This domain is referred to as the ApiPLP C-terminal β-pleated sheet (APCβ) domain (43). The N-terminal domain (NTD) is not conserved across the phylum, with no identifiable domain architectures (43). While the function of the NTD is not known, evidence shows that it is capable of binding to membranes, although at a lower affinity than the CTD (182), and

also that it may have a role in regulation of TgPLP1 pore forming activity through pH dependent proteolysis (104).

The structures of the TgPLP1 MACPF and APCβ domains have recently been solved by our group (91). The MACPF domain was solved in two oligomeric structures alongside a monomeric form, with oligomers forming either in a hexameric closed ring or a continuous helix of six subunits per turn (**Figure 3.1**). The domain consists of the canonical four-stranded twisted β-sheet. Transmembrane Hairpin 1 (TMH1) is resolved in the crystal structures, and packs into the domain between the β-sheet and alpha helical arrangement at the top of the structure, however TMH2 is not resolved. These oligomeric forms are considerably smaller than pores formed by other MACPF domains. The structure is bulkier than other solved MACPF structures, in part due to two extra helices, partly explaining its capacity to form smaller oligomers. Alongside the two oligomeric crystal forms, TgPLP1's capacity to form transmembrane pores is supported by observation of detergent induced oligomerisation of the MACPF-APCβ, and MACPF domain alone, as well as conductance measurements of oligomers in membranes showing pores of variable size (91).

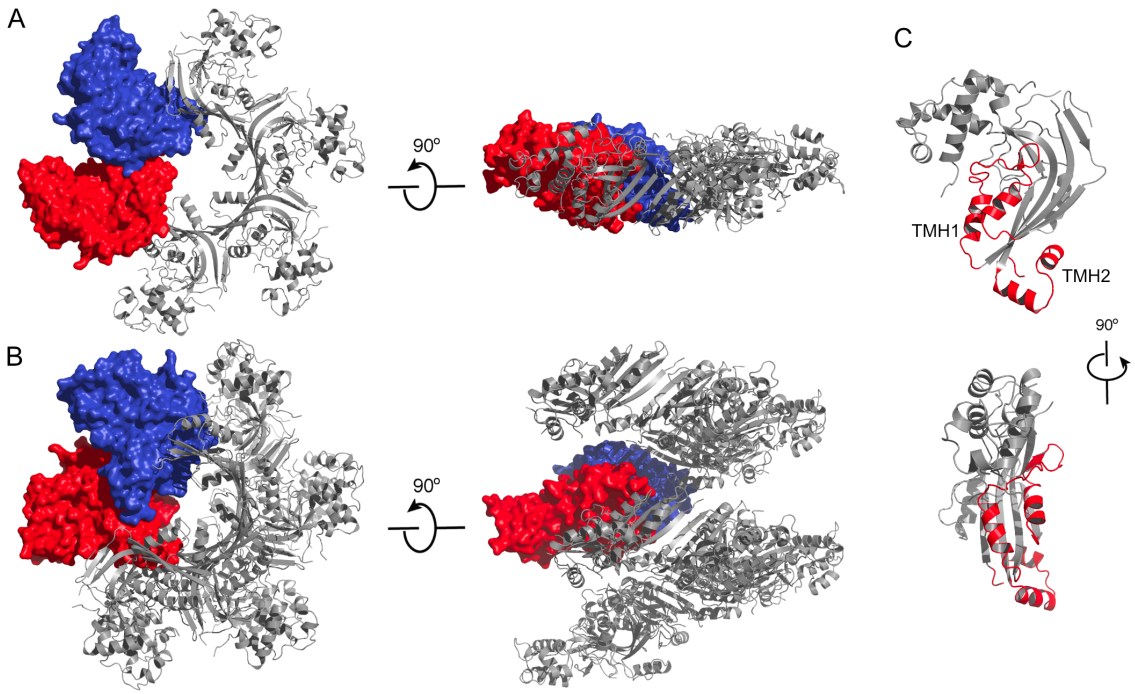


Figure 3.1 Crystal structures of TgPLP1 MACPF domain.

Oligomeric ring (A) (PDB ID: 5OUQ) and helix (B) (PDB ID: 5OUP) crystal structures show 6 MACPF domains in each oligomer. The helical crystal structure consists of a continuous MACPF helix, with only 6 represented here. The oligomeric ring structure is a closed ring containing 6 subunits. Two subunits are shown as surface representations coloured in red and blue in each oligomer. (C) MACPF domain oligomer (PDB ID: 5OWN) represented as a ribbon structure. TMH1 and TMH2 are shown coloured in red.

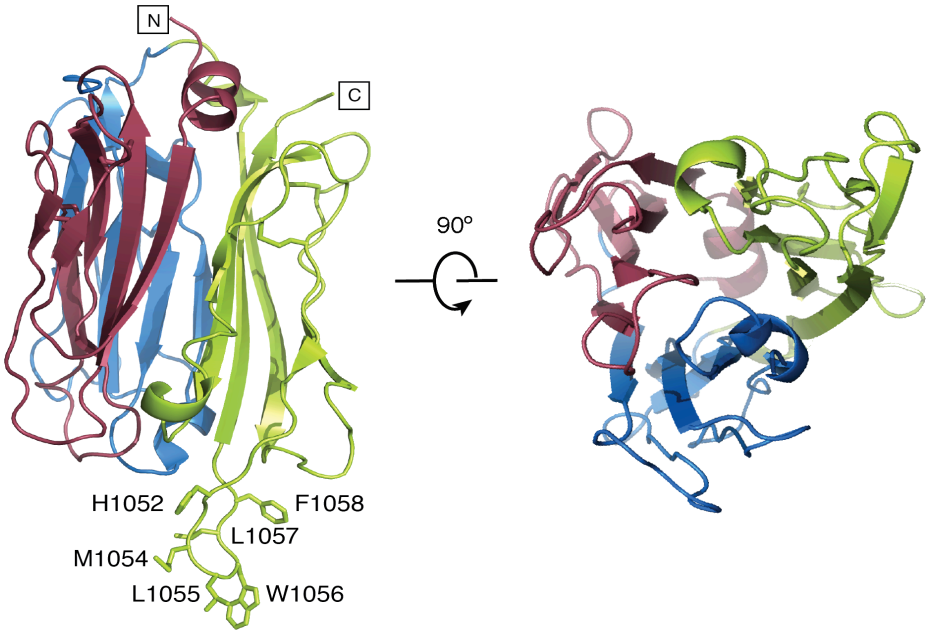


Figure 3.2 The TgPLP1 APCβ domain crystal structure.

Figure from (91). (PDB ID: 5OUO) Each homologous repeat is coloured separately, and disulphide bonds are shown. Residues in the hydrophobic extended loop at the base of the protein are highlighted.

The C-terminal APCβ domain crystal structure was also solved in this study, revealing a structure consisting of three repeats of a β-sandwich pseudo-domain arranged in a prism (**Figure 3.2**) (91). Four inner and two outer β-strands form each repeat, locked together by two disulphide bonds which are formed by cysteines conserved across the ApiPLP family. The prism core is hydrophobic, with a high proportion of aromatic hydrophobic residues holding the prism structure in place. A connecting loop at the base of the third β-sandwich is extended compared to others. It is hydrophobic in nature and tipped with a tryptophan residue, similar in character to membrane anchoring loops in other CTDs of the MACPF/CDC family (**Figure 3.2**).

3.1.4 Regulation of TgPLP1 activity

Both the egress and invasion processes of *T. gondii* involve induction of cell motility and microneme secretion, meaning TgPLP1 is secreted both during invasion and egress (182). It is therefore vital that the parasites use mechanisms to carefully regulate and orchestrate cell egress events to ensure accidental cell damage does not occur during cell invasion and replication. Acidification of the PV occurs late in replication, and beyond a certain threshold acts as a signal to induce egress, while cell invasion occurs in neutral pH conditions (104). Roiko *et al.* showed that membrane permeabilization is sensitive to pH, and that this sensitivity is due at least in part to sensitivity of TgPLP1 to pH (104). Membrane binding of the full-length TgPLP1 and CTD alone increases at low pH, indicating that activity is regulated at

the membrane binding stage of pore formation by pH. It has been proposed that pH sensitive degradation of the NTD may also play a role in regulation of PLP1 activity. The N-terminal domain is proteolytically processed by Subtilisin 1 (TgSUB1) (182, 183). This degradation does not occur at low pH, while it is active at neutral and high pH, suggesting that degradation at neutral pH leads to inhibition (104).

3.1.5 Proposed mechanisms of TgPLP1 mediated egress

The precise mechanism by which TgPLP1 allows parasite escape is unknown. Several lines of evidence support that TgPLP1 forms transmembrane pores to facilitate escape. While TgPLP1 pores have not been directly observed on cells, PLP1 was found in tightly associated oligomers on PV and host plasma membranes from naturally egressed cultures (182), suggesting a transition to the oligomeric membrane inserted form. We have also observed oligomerisation induced by detergent treatment, as well as conductance measurements of oligomers in membranes showing pores of variable size (91). The size of oligomeric MACPF crystal structures suggests that the pores formed by TgPLP1 may be considerably smaller than other known MACPF proteins, however Kafsack *et al.* showed that TgPLP1 forms pores that allow other effector proteins through the PVM (98).

The primary mechanism by which TgPLP1 acts may be to allow release of small molecules such as calcium (184) and proteases (185). A further suggestion is that TgPLP1 may act in concert with egress induced parasite motility to weaken membranes to help parasites break through (98).

With the crystal structures of TgPLP1 APCβ and MACPF domains recently solved, here we aimed to use MD approaches to study the molecular mechanisms of TgPLP1 oligomerisation, membrane binding, and pore formation. Analysis of MACPF domain oligomerisation was performed to investigate each of the two oligomeric crystal structures, to further elucidate their roles potential role in facilitating egress. In addition, the membrane binding APCβ domain forms an important first step in pore formation, in addition to conferring membrane specificity and regulation of pore formation. Therefore, we aimed to use MD to probe lipid binding and membrane dynamics at high resolution. Finally, we aimed to predict and model the transmembrane β-barrel pore of TgPLP1, in order to produce a model of a fully inserted membrane pore.

3.2 Results and Discussion

3.2.1 Oligomerisation of TgPLP1 MACPF domain

The structure of the TgPLP1 MACPF domain was previously solved in our group through crystallisation as a monomer and in two distinct oligomeric forms. The oligomeric forms were achieved using in-situ trypsinisation under different crystallisation conditions, one of which resulted in a 6-subunit flat ring oligomer of MACPF domains (PDB ID: 5OUQ*)*, while the other formed a continuous helical structure (PDB ID: 5OUP*)* (**Figure 3.1**). The helical crystal structure had good symmetry and was resolved to

2.03 Å, while the ring oligomer had a significantly lower resolution at 5.0 Å. The monomer alone was solved at 3.11 Å (PDB ID: 5OWN).

Crystallisation is typically achieved under extreme non-physiological conditions, and static structures do not necessarily reflect the oligomerisation capacity and conformational variability of a protein. We therefore assessed these two oligomers looking primarily at hydrogen bonding patterns and structural flexibility using molecular dynamics aiming to understand which oligomeric form may be more physiologically relevant in pore formation.

3.2.2 Stability of MACPF oligomers in atomistic simulations

Each structure was firstly simulated to assess overall stability of the MACPF domain and overall oligomers. Crystal structures lacked density in 3 loops of the MACPF domain, between residues 512-524, 612-615 and in the TMH2 residues 695-737. Modeller (186) was first used to connect free termini of missing loops 1 and 2, where the missing sequence was too long to accurately model. As loop 3 was only missing 4 residues, these were modelled back into the structure. Atomistic MD simulations were then performed for TgPLP1 MACPF monomer, hexameric closed ring crystal structure and a hexamer of the helical crystal structure. Each structure was simulated for 100 ns in triplicate.

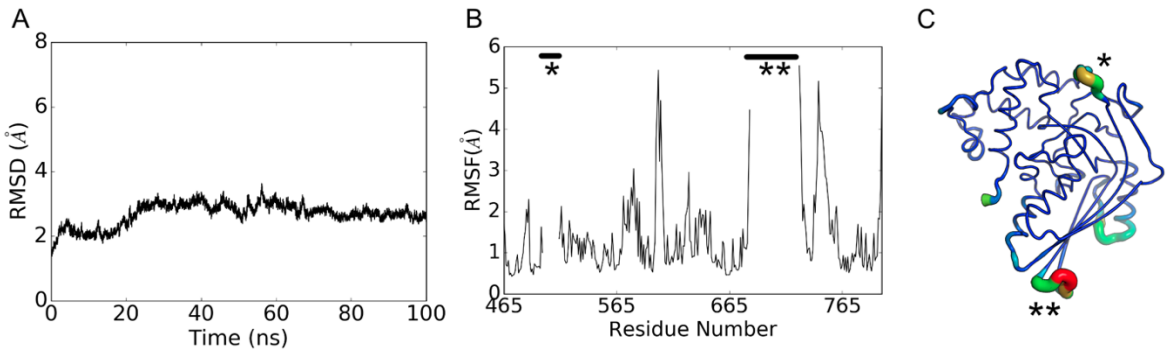


Figure 3.3 Analysis of structural stability of TgPLP1 MACPF monomeric crystal structure during MD simulations.

RMSD (A) and RMSF (B) plots of TgPLP1 MACPF monomeric crystal structure simulated in a 100 ns simulation. Two regions of missing structure are highlighted using black bars and labelled with asterisks (*) (C) Cartoon representation of TgPLP1 MACPF subunit, indicating regions of structural stability during simulations as Cα β-factors on a blue to red scale. Segments of missing structure are labelled with asterisks corresponding to labelling in (B). The structure is shown with free ends connected as described.

Overall the MACPF monomer had a highly stable core structure **(Figure 3.3)** with a Cα RMSD averaging 3.6 Å in the monomer. Both oligomeric structures were quite stable in simulations (**Figure 3.4**). The MACPF domains remained stable across the simulations, while the RMSD of the entire oligomers showed greater fluctuation. The helical oligomer overall had slightly greater stability of both the whole oligomer and for each individual monomer. These simulations indicate that the MACPF domain is a compact domain that is stable in solution as a monomer and oligomer.

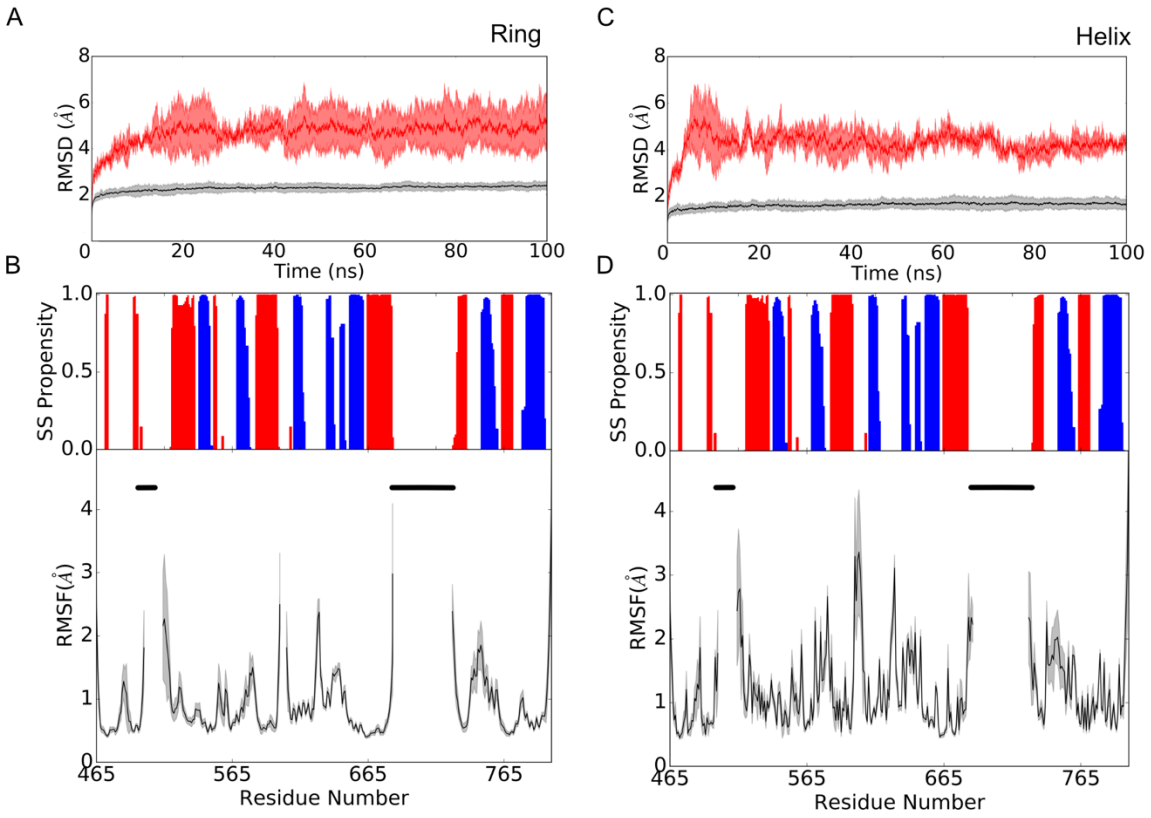


Figure 3.4 Structural stability of TgPLP1 MACPF hexameric helix and ring structures in simulations

RMSD (A&C) and RMSF (B&D) plots of TgPLP1 MACPF hexameric ring (left panels) and helix (right panels) showing structural stability from 3 replicates of 100 ns atomistic simulations. RMSD was calculated for each MACPF subunit (black lines) and for the whole molecule (red lines). Secondary structure propensity across the structure was calculated and represented with blue as helix and red as β strand. Missing segments of the structure are highlighted with black bars in RMSF plots.

3.2.3 Analysis of hydrogen bonding at MACPF oligomer interfaces

In order to form oligomeric assemblies, the MACPF domains oligomerise along each free edge of their central β-strands to form a continuous β-sheet across each domain. The hydrogen bonding pattern and shear between each interface must vary to form different oligomeric structures. In crystal structures, the helical interfaces had 6 main-chain hydrogen bonds between them. Each interface was identical in terms of number of bonds and residues involved. The ring interfaces varied however, both in number of bonds and residues involved **(Table 3.1)**. To further assess the stability of each interface, all main-chain hydrogen bonding pairs across adjacent MACPF domains were identified in the helix and ring simulations. The existence of each pair across the simulation was calculated to give a normalised frequency for each bonding pair (**Figure 3.5**). In the helix simulation, hydrogen bonding patterns were similar between each interface. Six bonding pairs were conserved between all interfaces and were present for at least 80% of the simulation time. Two bonding pairs were not conserved across each interface and formed intermittently through the simulations (**Figure 3.5 A**).

INTERFACE	CRYSTAL STRUCTURE	OLIGOMERIC MD	DIMER MD
Helix	6	6	6
Ring A-B	3	4	2
Ring B-C	0	1	N/A
Ring C-D	0	0	5
Ring D-E	3	4	N/A
Ring E-F	0	1	N/A
Ring F-A	1	4	5

Table 3.1 Counts of main chain hydrogen bonds between each adjacent pair of MACPF chains in the ring and helical crystal structures.

Main-chain hydrogen bonding pairs for each interface in the oligomeric MD and dimer MD were counted if they were present for over 80% of the 100 ns simulation. Only one interface of the helix is shown here as all interfaces in this structure were highly similar.

In comparison, ring interfaces were highly variable. There were fewer hydrogen bonding pairs across the ring overall, many of which existed at a lower frequency than those in the helix. Furthermore, that simulation does not recover a compacted stabilised ring structure: three distinct interfaces were observed in the ring, characterised by identical bonding patterns. Interfaces between monomers A-B, D-E and F-A were highly similar in each simulation, with 4 hydrogen bonds present between each subunit. Interfaces B-C and E-F were also similar, with only one hydrogen bonding pair well conserved across the simulations. The interface between C-D did not form any main-chain hydrogen bonds of significant frequency during the simulations, suggesting this interface did not form a continuous β-sheet (**Figure 3.5 B**).

These results suggest that the formation of hexameric ring species restrain interfaces in sub-optimal positions to maintain their planar structure. In order to investigate this, we isolated a series of dimers representing each interface type. Isolation away from the larger structure removed oligomeric restraints put on each interface, potentially allowing interfaces to settle into more favourable interactions. These dimers were then simulated, and hydrogen bonding conservation was again calculated (**Figure 3.5 C and Figure 3.6**)

While the number and identity of bonding pairs in the helix were maintained when oligomerisation restraints were removed, ring interfaces were altered. Removal of restraints allowed monomers of ring interfaces C-D, and F-A to settle closer together and form identical interfaces (**Figure 3.6**). Interestingly, interface A-B lost several hydrogen bonds, and formed only 2 of the 5 bonding pairs identified for ring dimers. This suggests that the formation of an almost closed MACPF ring forces subunits to adopt unfavourable interfaces, which are able to recover upon removal of structural restraints. It also suggests two more favourable primary β-strand interfaces between the domains. The first interface being the helical interface, in which 6 main-chain hydrogen bonds are formed, and the domains are less tilted in relation to the pore axis allowing greater surface contact and a more stable interaction. The second interface involves 5 hydrogen bonds, with a smaller contact area between the subunits due to greater tilt, as shown in dimers C-D and F-A.

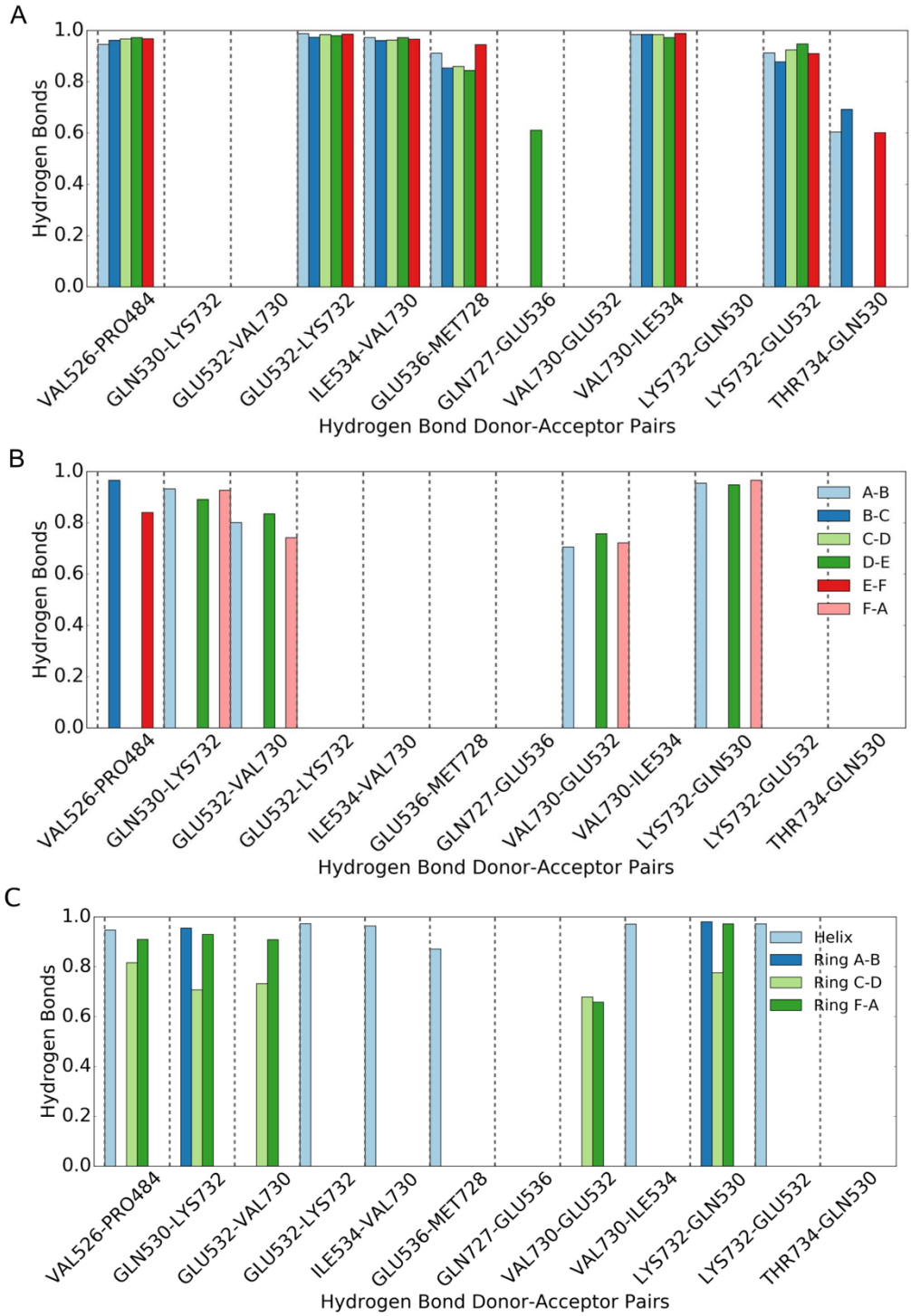


Figure 3.5 Hydrogen bonding in TgPLP1 MACPF oligomers.

Normalised bar charts showing frequency of inter-subunit main chain hydrogen bonds present in over 50% of simulation time during TgPLP1 MACPF helix (A), ring (B) and representative dimers (C) simulations. The hydrogen bond donor is the first residue listed for each bonding pair. Frequency of bonding is calculated as an average of 3 replicates.

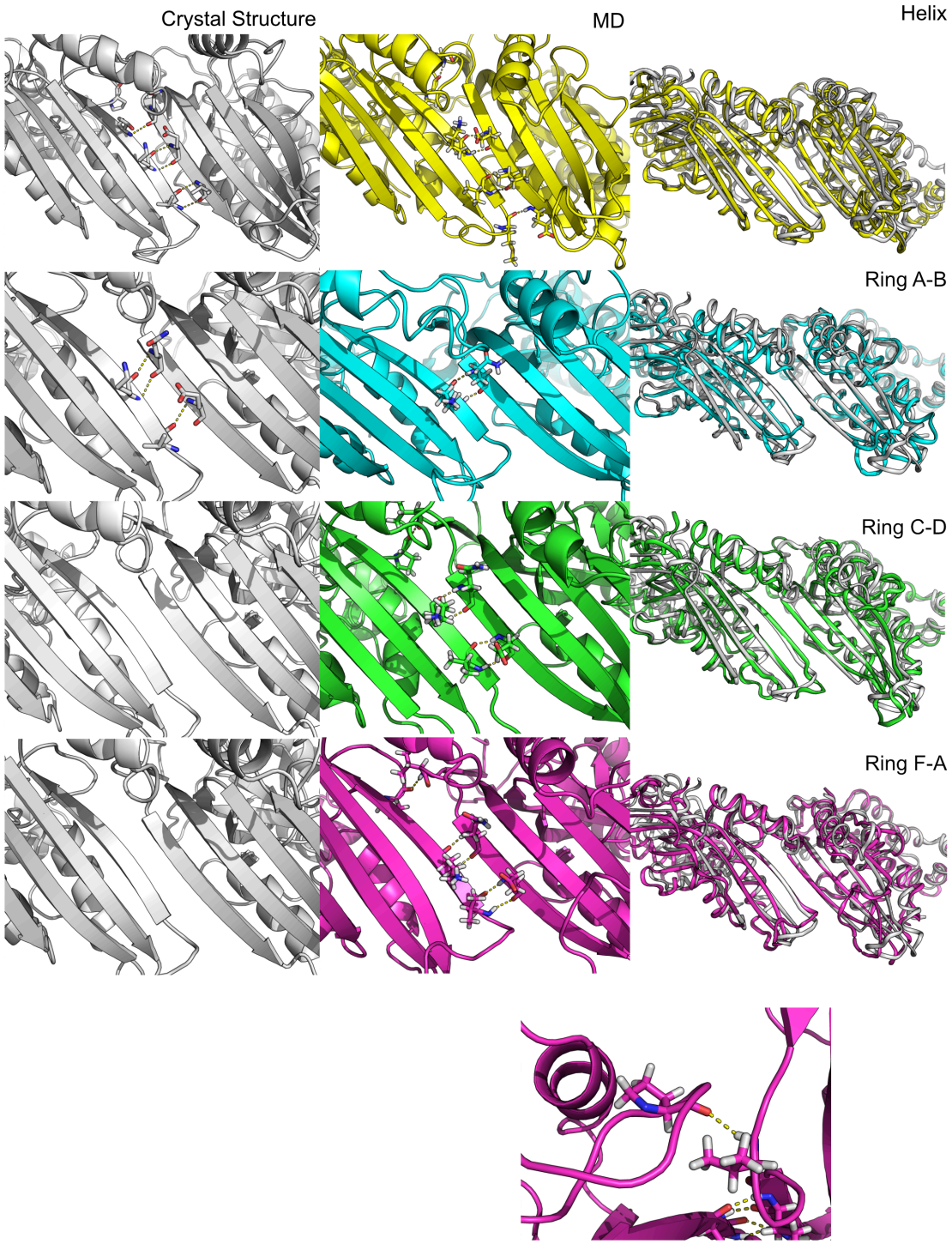


Figure 3.6 Comparison of hydrogen bonding at MACPF oligomer interfaces.

Crystal structures (grey) and frames from atomistic simulations (coloured) are shown. Representative dimers of the four different interface categories are shown here. Interfaces following MD simulations show high frequency main chain hydrogen bonds identified following frequency calculations.

Altogether, these results suggest that the helical oligomeric structure is significantly more stable than the hexameric ring crystal structure. Hydrogen bonds in the helix were both greater in number and better conserved. Fewer bonds, and greater variation, between subunits of the ring suggest that these interfaces would be less favoured during oligomerisation in physiological conditions. In order to form a ring oligomer, each monomer must tilt to a greater angle against the pore axis, meaning fewer bonds can be formed and ultimately the ring is unable to completely close. These results are supported by, and provide an explanation for, the differences in resolution of each crystal structure.

3.2.4 Large scale fluctuations of the two oligomeric forms

Hexameric ring structure

Following on from the detailed study of oligomer stability, we wanted to assess larger scale conformational fluctuations of the two crystal forms. Firstly, the atomistic MACPF ring simulations were assessed to determine oligomer dynamics. Visual observation during simulations showed a slight deformation to form a more oval structure (**Figure 3.7 A**). Distances were calculated between central residues of each pair of opposite subunits to quantify deformation during the simulation and showed slight variation in distance across the simulation (**Figure 3.7 B**). The 100 ns timescale of the atomistic simulations is unlikely to be long enough to observe larger scale fluctuations, so the system was coarse grained, and 10 simulations were produced each 1 μs in length.

The MACPF ring frequently underwent deformation, as can be seen by the variance in distances between each opposite pair of subunits (**Figure 3.7 E**). In many cases, the ring split into several smaller oligomers, which stayed in a loose ring structure but were not close enough to form a continuous central β-sheet (**Figure 3.7 D**). In these cases, distance measurements indicated a more oval shaped oligomer, and visual observation showed that the ring tended to split into two trimers, in which the curvature of each trimer increased. In some cases, the ring split apart and tightened to form a flat spiral, again leading to increased curvature at interfaces. While reduced resolution in CG simulations cannot fully reflect complex protein-protein interactions, they can give a picture of the oligomeric stability at longer timescales. In both the AT simulations and CG simulations the ring structure did not remain stable. Breaking apart of subunits and deformation reflected a preference for increased curvature of the oligomer.

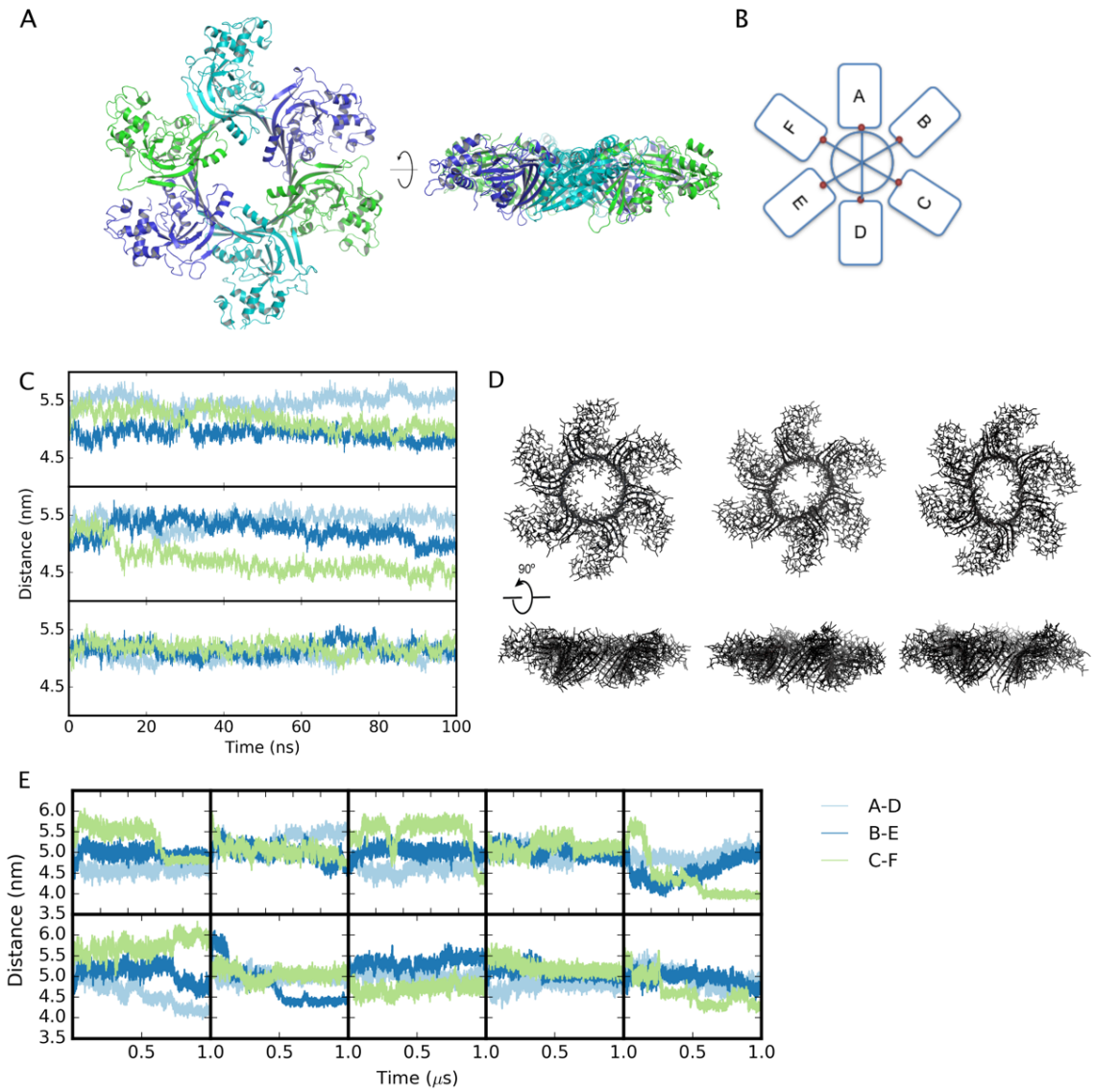


Figure 3.7 Assessing large scale stability of TgPLP1 MACPF ring.

(A) final snapshot from AT simulation showing deformation of ring. (B) Distance measurements were taken between opposite subunits in the ring. The Cα atom of residue 588 in each MACPF, representing the centre of the central MACPF 4-stranded β-sheet was used. (C) Distance measurements between each opposite subunit from three 100 ns AT simulations. (D-E) Deformation of TgPLP1 MACPF ring crystal structure following 1 μs martini coarse-grained simulations. (D) Cartoon representation of snapshots showing changes in structure observed during the simulation, showing the initial starting structure (left), and two representations of typical structural changes (middle and right). (E) Distance measurements between each pair of opposite subunits in the hexamer from each of 10 replicate simulations.

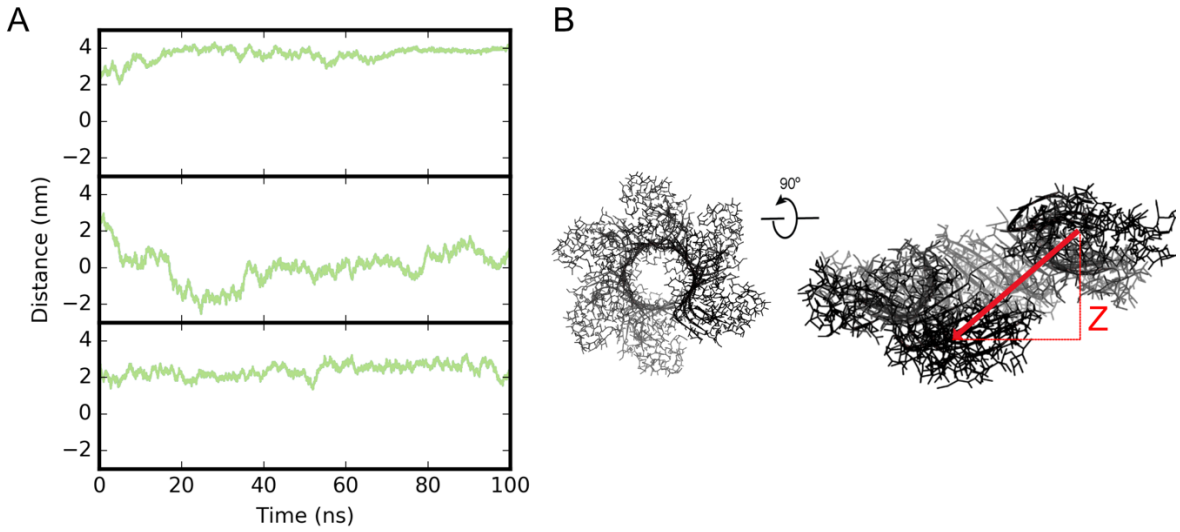


Figure 3.8 Vertical motions in MACPF helix simulations.

(A) Z component of distance between centre of mass of first and last subunit of hexamer of TgPLP1 MACPF helical crystal structure is shown over three 100 ns atomistic simulations. (B) The distance between first and last subunit of the helix hexamer is measured, and the vertical component along the Z axis is used to calculate distance in (A).

The Helical crystal structure

The atomistic helix simulation was then assessed for large scale fluctuations. The Z distance between first and last subunit was measured to determine if the tilt angle between monomers is able to alter leading to stretching or expansion of the helix or remains more static. The atomistic simulations showed little fluctuation in the course of the 100 ns simulation (**Figure 3.8**). Again, 10 CG simulations were set up to assess the flexibility at the longer time-scale of 1 μs. While interfaces remained intact, coarse grained simulations showed significant variation in helical tilt angles, with the helix both relaxing to form a looser helix with reduced distances between first and last subunit and twisting to increase tilt angle and reduce the internal channel diameter (**Figure 3.9**). A histogram plot of all values of distances between first and last subunit from the 10 repeats show that two distances, ~4 nm and ~7 nm were favoured within simulations, compared to the crystal structure distance of 3.1 nm (**Figure 3.9**).

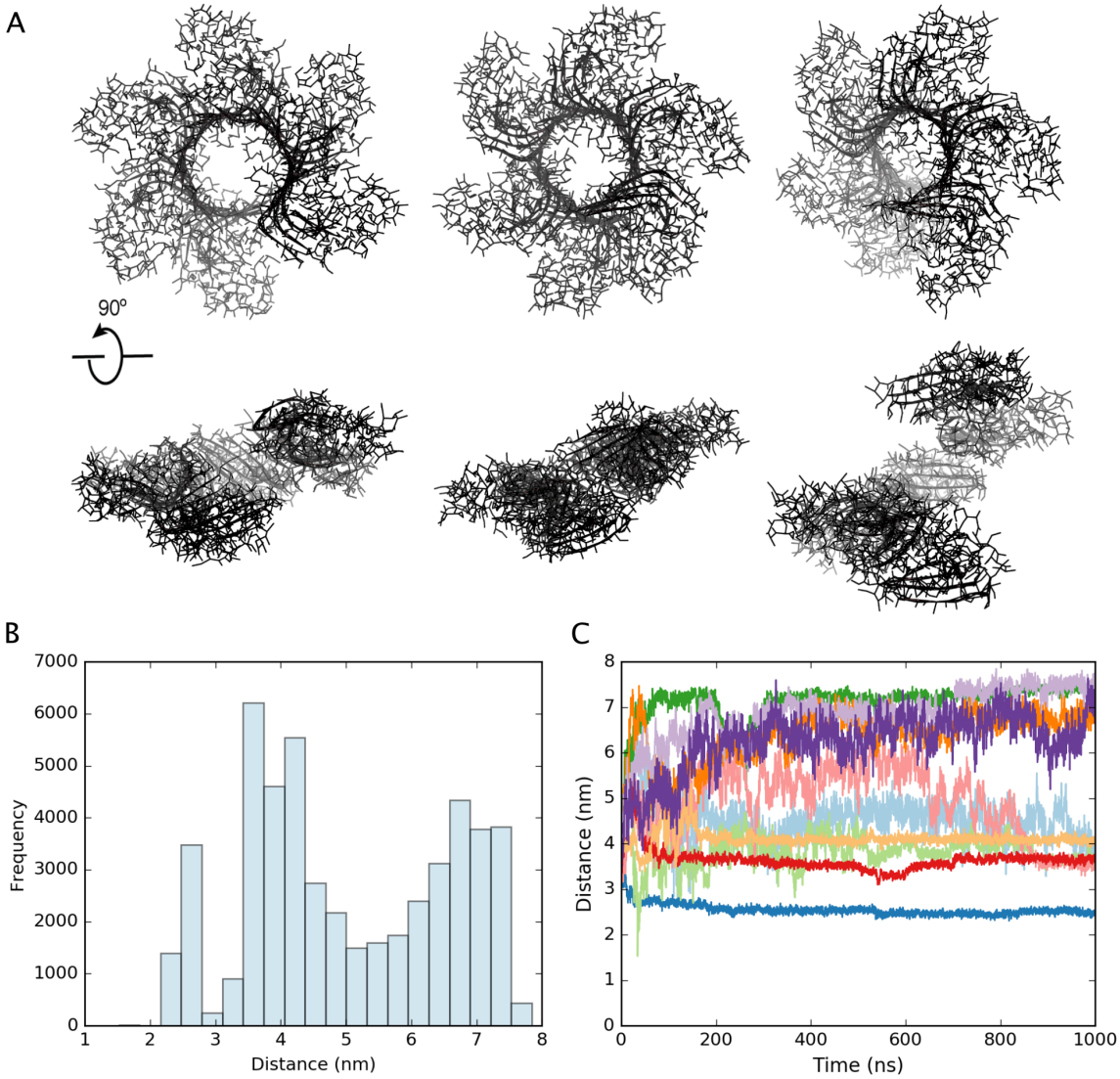


Figure 3.9 Helix motions in CG simulations.

(A) Cartoon representation of snapshots from Martini coarse-grained simulations of the TgPLP1 MACPF helix representing points of significant changes in pitch. (B) Frequency of Z-distances between the centre of mass of the first and last subunit of TgPLP1 helix from all 10 1 μs repeat simulations. (C) Change in Z-distance between first and last subunit of helix through each 1 μs simulation. Each simulation is represented by a different colour.

3.2.5 Discussion of TgPLP1 MACPF oligomerisation interface preferences

Larger scale observations of both oligomeric structures further support results from hydrogen bonding analysis. The ring was unstable in both atomistic and coarse-grained simulations, tending to either deform or split into several smaller oligomers. The oligomers also appeared to favour an increase in curvature. Helical interfaces on the other hand remained stable. In coarse grained simulations the helix

was able to change in tilt while maintaining its overall conformation. These motions involved both tightening and a relaxation of the helix, leading to changes in pitch and pore size.

MACPF/CDC proteins form pores by oligomerising on the surface of membranes, which has been observed using electron microscopy, showing both complete rings and arcs formed on membranes (82). As such, the functional oligomeric form of MACPFs is considered to be a flat planar oligomeric ring or arc. Previous cryo-electron microscopy studies of bacterial CDCs found that the CDC pneumolysin was able to form of helical assemblies in solution (74), and that these structures maintained the subunit interfaces found in the planar ring-shaped pore (79). The recent high-resolution structure of the MAC complex resolved a split-washer type assembly, with a twist in the pore preventing the ring from completely closing (87, 88). In addition, our recent work on solving a high-resolution structure of the perforin-2 pore [Manuscript under review, see Chapter 5], has also revealed a subset of results in the split washer conformation.

TgPLP1 may function using a similar mechanism, with the higher stability helical interface favouring formation of a split ring on the membrane. The ability of the helix to change significantly in pitch suggests that this structure could accommodate a wide range of helical conformations. The current structures contain the MACPF domain in isolation. It is likely that presence of the APCβ domain and N-terminal domain in the full-length protein will restrict the range of conformations available during oligomerisation resulting in a helix with reduced pitch, similar to the split-washer conformation. The APCβ domain also has affinity for membranes, so may exert a force on the oligomer upon binding resulting in a flat ring structure. In this case, the propensity of the oligomer to form helical assemblies may exert a shearing force on the membrane that could contribute to membrane disruption and pore formation.

LIPIDS	**COMPOSITION (%)**	**TOTAL MEMBRANE BINDING**	**TOTAL 'UPRIGHT' BINDING**	**TOTAL 'UPSIDE DOWN' BINDING**
PC	100	1	1	0
PE	100	15	15	0
PC:PS	80:20	5	5	0
PC:PE:PS	45:45:10	3	3	0
PC:PE:PS:PI(3,4,5)P_3	45:42:10:3	16	7	9
PC:PE:PS:PI(4,5)P_2	45:42:10:3	20	9	11

Table 3.2 Table indicating the total number of simulations out of 20 in which TgPLP1 APCβ bound to the membrane for each bilayer composition.

These results suggest that MACPF oligomer interfaces may be capable of flexing to adopt a wider range of tilt angles. This puts into question the range of functional oligomeric states that the MACPF/CDC family of proteins can adopt. It is possible that the twisted, helical assembly may have functional relevance, or that interconversion between different forms occurs as part of the pore formation mechanics.

3.2.6 Membrane association of TgPLP1 APCβ

Following investigation of TgPLP1 MACPF oligomerisation, we next turned our attention to the APCβ domain. Given structural similarities to other MACPF/CDC C-terminal domains, we hypothesised that this domain would bind to membranes using the extended tryptophan tipped loop at its base. We firstly performed a series of coarse-grained membrane association simulations to investigate the mode of binding of the APCβ domain. Simulations were performed with a series of bilayers of different lipid composition to investigate lipid specificity.

The protein was placed 10 nm away from a series of pre-formed bilayers with lipid compositions listed in **Table 3.2**. Ensembles of 20 unbiased simulations of 1 μs in length were performed per bilayer. The APCβ domain was rotated consecutively along the X, Y and Z-axis in each simulation to reduce bias, and the protein was able to diffuse freely towards the membrane.

Protein-membrane separation distance along with rotation of the protein along the y-axis against an initial set orientation was used to generate 2-dimensional histograms to identify binding orientation (**Figure 3.10**). The APCβ domain bound to the membrane in a single upright orientation on membranes containing PC, PE and PS alone and in combination. As predicted, the protein bound to the membrane using the tryptophan tipped extended loop to anchor into the lipid bilayer (**Figure 3.10 C**). The N-terminal tail sits at the top of the domain, providing space for the MACPF domain to stack on top of the domain.

Upon addition of PIP lipid species, the protein only found this orientation in around 45% of binding events (**Table 3.2**). Further binding occurred in an upside-down orientation, with the top of the domain loosely bound to the membrane, sitting at a greater protein-lipid separation than in the upright orientation. This upside-down orientation is considered to be unlikely due to the location of the MACPF domain which is connected to the APCβ domain a linker of only 11 residues. Visual observation of the simulations showed that the APCβ domain is primarily in contact with PIPs in the upright, as can be seen in the snapshot in **Figure 3.10 C**. This was also further validated later using lipid contact analysis (**Figure 3.13**).

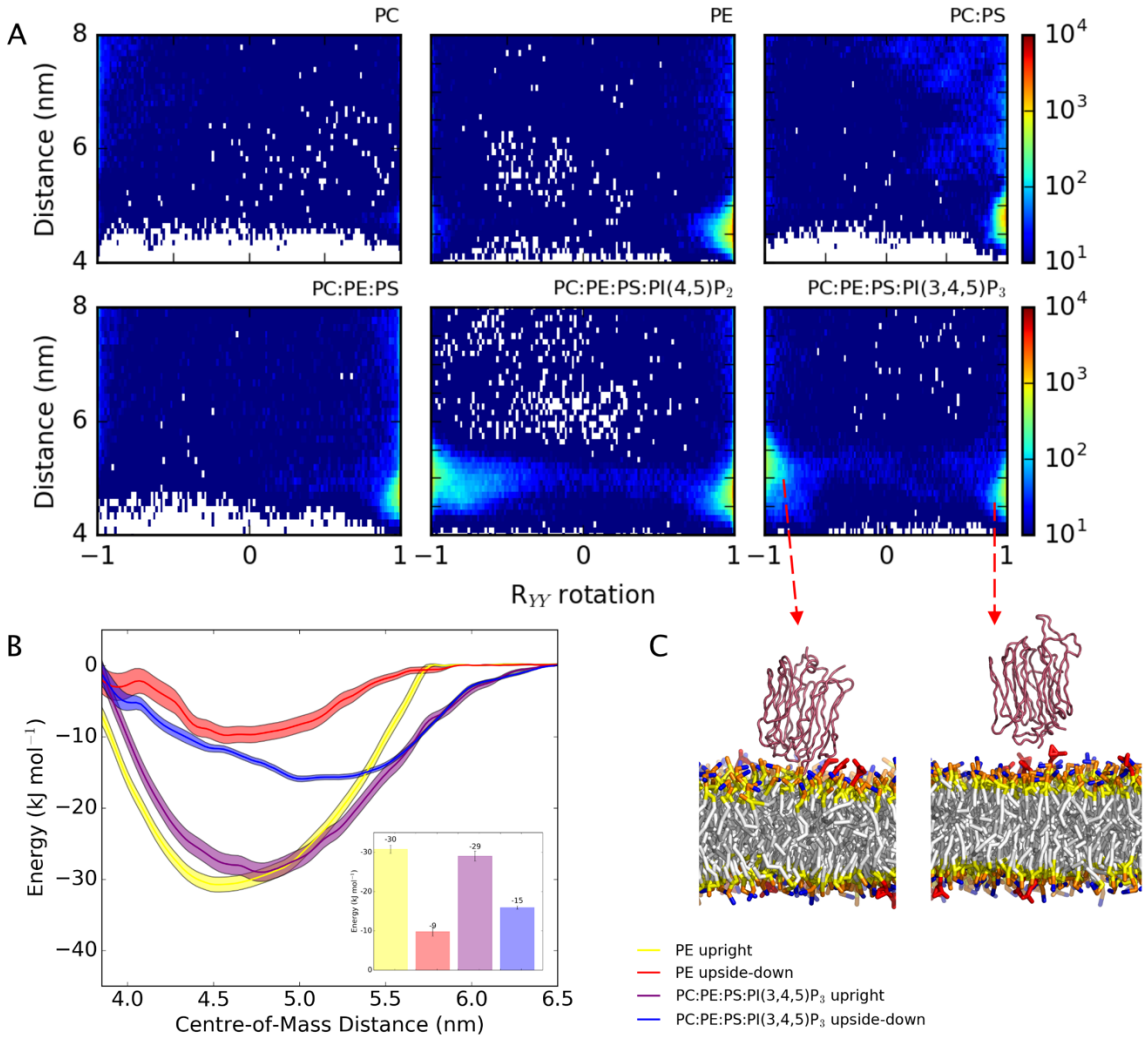


Figure 3.10 Membrane binding of TgPLP1 APCβ domain.

(A) Favoured binding orientations are shown from membrane association simulations of TgPLP1 APCβ domain on a series of bilayers of different lipid composition. 2D histogram plots show distance between centre of mass of the membrane and protein against the Y component of the rotational matrix of the APCβ domain. (B) PMF profiles showing free energy of binding of the APCβ domain in upright and upside-down orientations on PE and PC:PE:PS:PI(3,4,5)P_3 membranes. Bootstrap analysis was used to generate error estimates. (C) Graphical representations of the two upright (left) and upside down (right) binding orientations.

Umbrella sampling and PMF calculations were performed to investigate the energetics of these binding orientations (**Figure 3.10 B**). The centre of mass of the APCβ domain was used to pull the protein along a reaction coordinate perpendicular to the bilayer, and umbrella sampling was used to construct a PMF profile of binding energy. Strong binding to PE and PC:PE:PS:PI(3,4,5,)P_3 bilayers was observed in the upright orientation, while the same calculations for the protein in the upside-down orientation showed a reduction in binding energy of 21 and 14 kJ mol^{-1} for PE and PI(3,4,5)P_3 containing membranes respectively. The upside-down binding to the PC:PE:PS:PI(3,4,5)$_3$ membrane had a greater affinity for the membrane than for PE alone, supporting the role of PIPs in this orientation. We believe this secondary orientation is an artefact due to the extra charge density of the PIP species interacting with residues at the top of the domain which would not be accessible to the membrane in full-length TgPLP1 due to the presence of the MACPF domain at this point in space.

3.2.7 PMF calculations to investigate lipid preferences

The number of binding events in each ensemble gave an initial indication of lipid binding preferences for TgPLP1 (**Table 3.2**). TgPLP1 had little affinity for bilayers containing PC alone, with only one binding event occurring for 100 % PC membranes. Membranes containing a high proportion of PE led to increased binding, as well as those containing phosphatidylinositols PI(3,4,5)P_3 and PI(4,5)P_2 especially.

PMF calculations were used to further investigate the energetic contribution of specific lipids to binding. Umbrella sampling was performed using the same reaction coordinate as previously described. Starting orientations for each PMF were taken from a final frame representing binding in each set of membrane association simulations. All simulations were performed using the upright binding orientation.

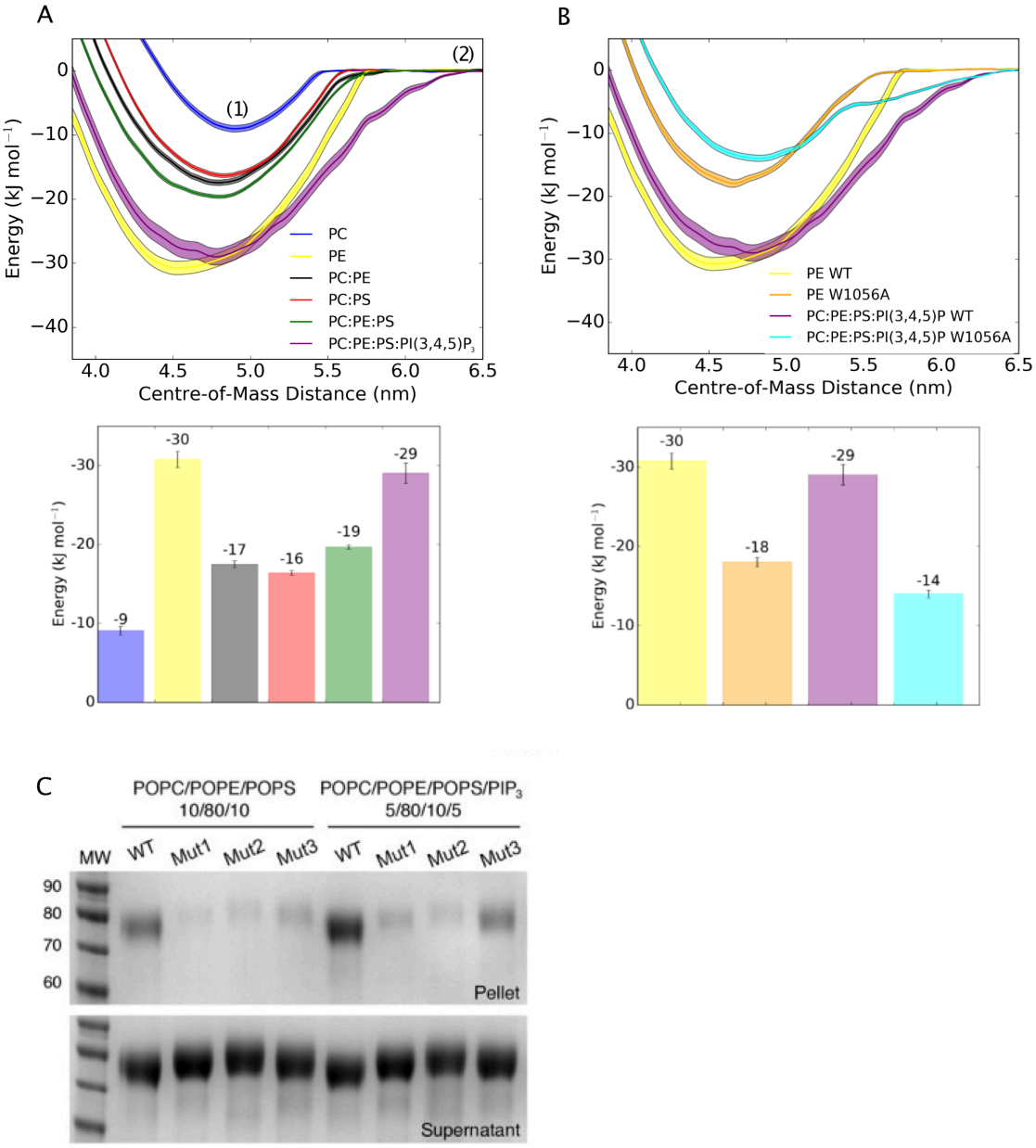


Figure 3.11 Lipid binding preferences and mutational studies of TgPLP1 APCβ domain.

PMF profiles of the APCβ domain on a series of membranes of different lipid compositions (A), and W1056A mutation on PE and PC:PE:PS:PI(3,4,5)P_3 membranes. Free energy at the minima of each energy curve is plotted below. Bootstrap analysis was used to generate error estimates. (C) Liposome sedimentation assays of mutated TgPLP1 APCβ domain performed by Dr Tao Ni. Mutant 1 (mut1) affects the tryptophan loop tip (^{1054}MLWL^{1057}-AAAA); mut2 gives charge neutralization (R851A K995A R996A R1030A K1048A); mut3 affects an exposed histidine (H1052A).

The APCβ had a low binding affinity of -9 kJ mol^{-1} on PC only bilayers. Binding energy was increased by 6-10 kJ mol^{-1} following addition of PE and PS alone or in combination, while addition of $PI(3,4,5)P_3$ led to a significant increase in binding energy down to -29 kJ mol^{-1} (**Figure 3.11**). Complementary liposome assays were also performed by Dr Tao Ni showing increased binding to $PI(3,4,5)P_3$ containing membranes over PC:PE:PS alone (**Figure 3.11 C**).

The APCβ had high affinity for 100% PE membranes, with binding energy of -30 kJ mol^{-1}. While not physiological, this suggests the protein has a strong preference for PE lipids, which has also been observed by Guerra *et al*. (187). A shift in the energy well towards the membrane was seen in PMFs of binding to 100% PE membranes compared to other bilayers, with the energy minima of PE at a protein-lipid separation of 4.5nm compared to 4.9 nm in the PC PMF.

To investigate the strong preference for PE, we analysed binding in the umbrella sampling windows corresponding to the energy minima in PC and PE. While similar contact was made with the tryptophan tipped loop in each, greater contact was made with PE across the base of the protein than with PC (**Figure 3.12**). We hypothesised that preference for this greater contact would in part be due to the hydrogen bonding potential of PE, as greater contact was made with polar residues including asparagine, serine and arginine. The role of hydrogen bonding in increased PE affinity was explored further in atomistic simulations below. In addition to increased hydrogen bonding potential, PE is a structural lipid with a conical geometry due to the relatively small size of its head group which confers a curvature to the membrane. Inclusion of PE, especially in high concentrations, could cause curvature and a reduced density of lipid packing, with greater space between adjacent headgroups. This could in turn, allow deeper insertion of membrane anchoring residues, such as tryptophan, which prefer to sit at the interface between lipid head groups and the hydrophobic core. To investigate this, bilayer densities were taken, both from the umbrella windows described above, as well as from membrane association simulations in which the APCβ domain was bound.

This allowed for comparison between freely associated protein and the steered umbrella sampling. The density of W1056 was calculated to show depth of anchoring into the bilayer. In the umbrella sampling windows, W1056 sat at 1.89 nm away from the centre of the PC bilayer compared to 1.54 nm in PE. In membrane association simulations the separation between W1056 and the PC membrane was unchanged, while the separation in PE simulations increased by 0.43 nm to sit at 1.97 nm away from the bilayer centre. This disparity suggests that while PMF calculations indicate that the reduced separation between protein and membrane is favourable, this position is not accessible to the protein without addition of an external force.

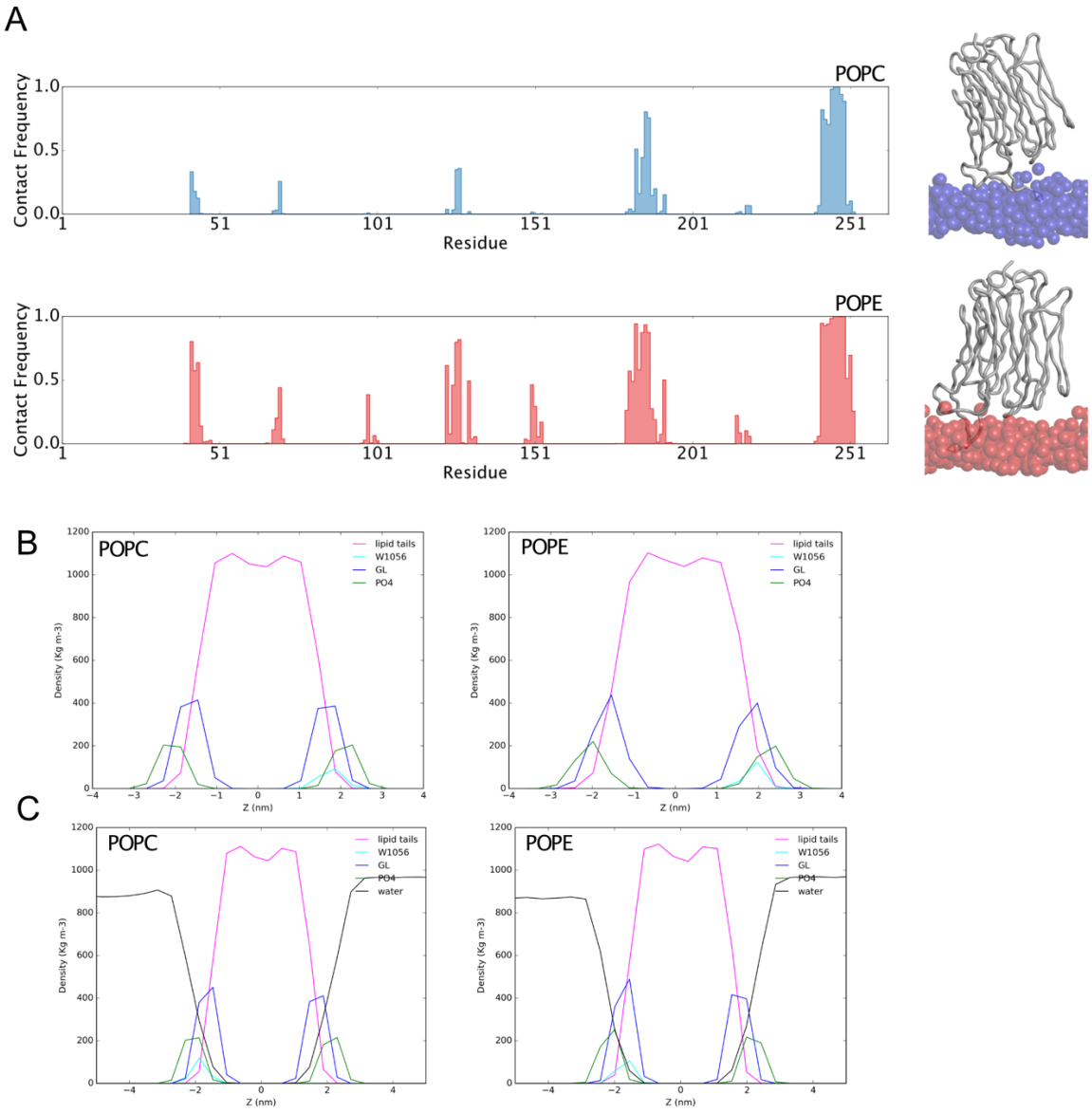


Figure 3.12 Comparison of TgPLP1 APCβ domain binding to PC compared to PE bilayers.

(A) Lipid contact frequencies of the APCβ domain calculated from PMF window at energy minima for the 100% PC and 100% PE bilayers, and cartoon representatives showing the final frame of each respective simulation. (B-C) Bilayer density profiles from membrane association simulations of each membrane (B) and umbrella sampling windows corresponding to energy minima for each bilayer (C). The phosphate (PO4) is shown in green, glycerol (GL) shown in blue, and lipid tails in magenta. Location of the tryptophan (W1054) is calculated from the Trp side chain.

Mutation of the tip tryptophan to alanine (W1056A) was produced and umbrella sampling was performed again to test the significance of this residue in membrane binding. PMFs were calculated for the W1056A mutant on a PE and PC:PE:PS:PI(3,4,5)P_3 bilayer, leading to a reduction in binding of 12 kJ mol^{-1} and 15 kJ mol^{-1} respectively compared to wild type protein (**Figure 3.11 B**). This supports the key role of W1056 in binding to the membrane, while bilayer affinity was maintained in its absence confirming that other residues across the base of the protein are also involved in binding. This was confirmed in liposome sedimentation assays showing that mutation of the hydrophobic loop to alanines (^{1054}MLWL^{1057} → AAAA) prevented membrane binding (**Figure 3.11 C**).

Lipid contact analyses for coarse grained membrane association simulations

Lipid contact analysis was performed to identify residues responsible for lipid binding preferences. Contacts were calculated for PC:PE:PS, PC:PE:PS:PI(4,5)P_2 and PC:PE:PS:PI(3,4,5)P_3 membrane association simulations, and averaged for each lipid species (**Figure 3.13**). Contacts in simulations in which the protein bound in an upright versus an upside-down orientation were calculated separately to identify differences in binding, while simulations in which the protein did not bind were excluded.

In all cases, the pattern of contact to PC and PE were similar across the protein. Contact with PE and PC to PC:PE:PS membranes was dispersed across the loops at the base of the protein, indicating these residues maintain a broad, non-specific interaction with the membrane. The tryptophan tipped loop made significant contact with the membrane, as did a second positively charged loop ^{992}NANKR^{996}, which made greater contact with PE. This loop sits next to and just above the extended tryptophan tipped loop, forming the first loop of pseudo-repeat 3.

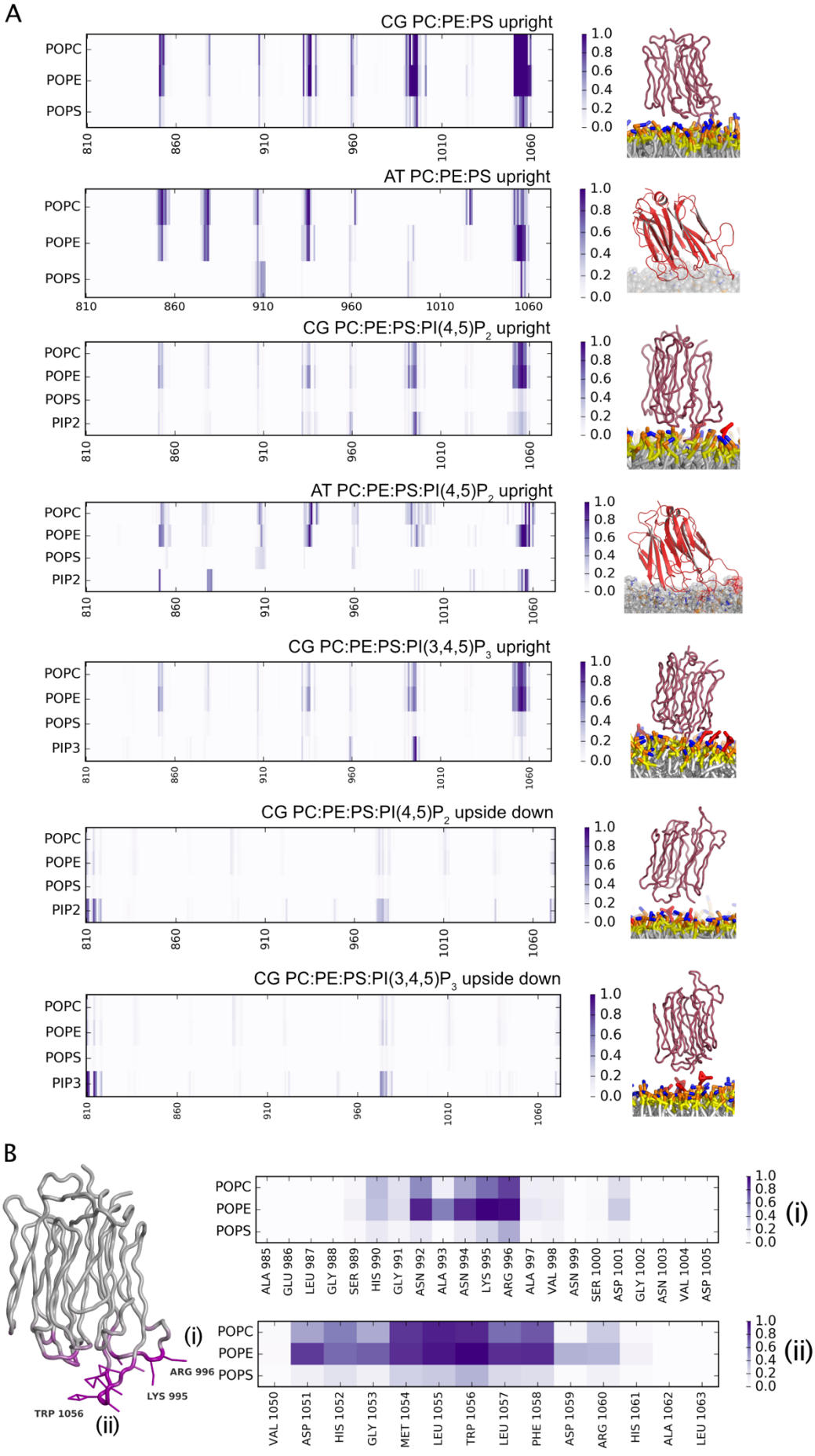

A
CG PC:PE:PS upright
AT PC:PE:PS upright
CG PC:PE:PS:PI(4,5)P2 upright
AT PC:PE:PS:PI(4,5)P2 upright
CG PC:PE:PS:PI(3,4,5)P3 upright
CG PC:PE:PS:PI(4,5)P2 upside down
CG PC:PE:PS:PI(3,4,5)P3 upside down
POPC
POPE
POPS
PIP2
PIP3
810
860
910
960
1010
1060
1.0
0.8
0.6
0.4
0.2
0.0
B
(i)
(ii)
ARG 996
LYS 995
TRP 1056
ALA 985
GLU 986
LEU 987
GLY 988
SER 989
HIS 990
GLY 991
ASN 992
ALA 993
ASN 994
LYS 995
ARG 996
ALA 997
VAL 998
ASN 999
SER 1000
ASP 1001
GLY 1002
ASN 1003
VAL 1004
ASP 1005
VAL 1050
ASP 1051
HIS 1052
GLY 1053
MET 1054
LEU 1055
TRP 1056
LEU 1057
PHE 1058
ASP 1059
ARG 1060
HIS 1061
ALA 1062
LEU 1063

Figure 3.13 CG and AT APCβ lipid contacts.

(A) APCβ lipid contact fingerprint plots are shown on the left. Normalised frequency of contact was plotted for each lipid from residue contacts averaged across all membrane association simulations for each bilayer in which the protein binds in an upright orientation for CG contacts, and across three repeats for AT simulations. The left panel shows cartoon representations of representative binding orientations. PIP lipid headgroups are coloured in red throughout. (B) Averaged contact for all lipids in the upright binding orientation from PC:PE:PS membrane associations is plotted onto a cartoon representation of the APCβ domain. The secondary positively charged binding loop (i) and tryptophan tipped hydrophobic loop (ii) are indicated, and normalised frequency of contacts are shown for each.

Lipid contact with PS was reduced compared to PC and PE in PC:PE:PS bilayers, while specific residues K851 and R932, histidines H907 and H1052, and the positively charged loop residues K995 and R996 had greater contact (**Figure 3.14**). Upon addition of $PI(4,5)P_2$ and $PI(3,4,5)P_3$, contact with PC and PE maintained the same pattern across the base of the protein compared to the PC:PE:PS bilayer and maintained the same orientation on the membrane. PS contacts were reduced following PIP addition to the membrane, apparently displaced by contact with PIP molecules.

Closer analysis of PIP contacts showed that both $PI(4,5)P_2$ and $PI(3,4,5)P_3$ Made strong contact with residues R959, K995 in the positively charged ^{992}NANKR^{996}binding loop already identified, and to a lesser degree with R959 in pseudo-repeat 2 (**Figure 3.14**). Observation of simulations shows PIP molecules bound almost consistently to the positively charged loop, although tracking of lipids shows exchange throughout the simulations, rather than tight contact with one molecule. The radial distribution function (RDF) for each lipid around the backbone bead of R996 was calculated to test this observation across all upright bound simulations, and showed clustering of $PI(4,5)P_2$, while no clustering of other lipid species was observed (**Figure 3.14 B**). The same observations were made for $PI(3,4,5)P_3$, though $PI(3,4,5)P_3$ clustering was less pronounced than for $PI(4,5)P_2$.

Lipid contact analysis of the upside-down orientation of TgPLP1 in PIP containing membranes showed fewer interactions between lipids and the protein. Most contacts were mediated by the PIP molecules, which bound specifically to residues R810, N811 and K814 and the extreme N-terminus (**Figure 3.13**).

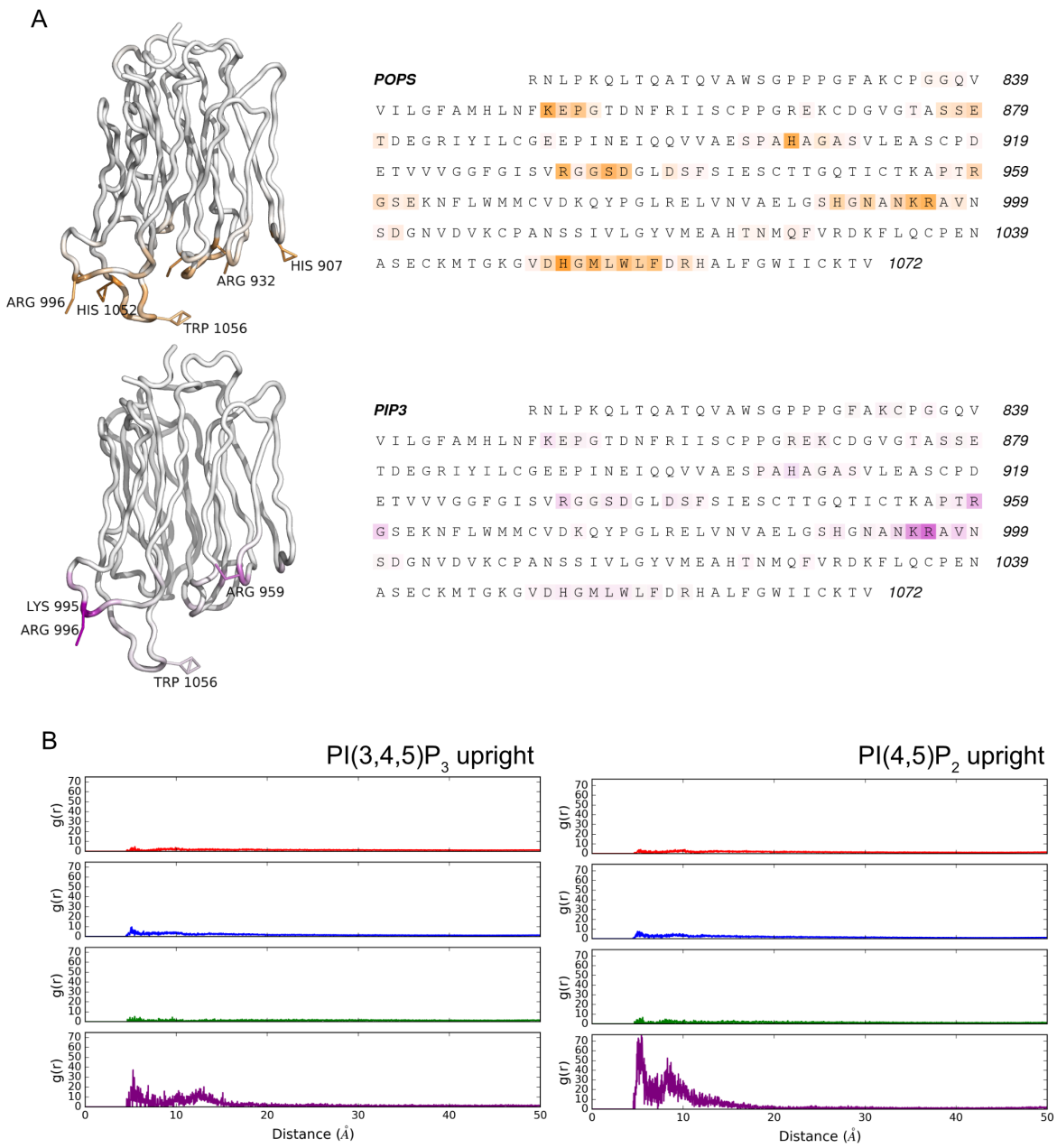


Figure 3.14 Electrostatic interactions and lipid clustering of TgPLP1 APCβ

(A) Normalised lipid contact frequency for PS and PI(3,4,5)P_3 plotted onto cartoon representation and sequence of the APCβ domain. (B) Representative radial distribution functions of each lipid calculated from residue K996. Contacts and RDF calculated from PC:PE:PS:PI(3,4,5)P_3/PI(4,5)P_2 membrane association simulations. Lipid contacts were averaged across all incidences of upright binding. RDF calculations are representative of upright binding in respective membrane binding association simulations.

3.2.8 Detailed atomistic investigation of lipid binding characteristics of APCβ

Binding poses representative of the upright membrane binding orientations on PE, PC:PE:PS and PC:PE:PS:PI(3,4)P_2 bilayers were converted into atomistic structures (158), and 100 ns atomistic simulations were run in triplicate to probe lipid binding at greater detail. Stability of each protein in simulations was similar to the monomer when comparing Cα RMSD values (**Figure 3.15 A**). Comparison of RMSF values across the APCβ domain bound to each membrane shows that the core of the domain remains very stable, with most fluctuations at the base of the protein. In particular, the positively charged ^{992}NANKR^{996} binding loop was stabilised on the PC:PE:PS:PIP(4,5)P_2 compared to the PC:PE:PS and PE membranes (**Figure 3.15 B**).

In comparison to CG simulations, the AT simulations allow the APCβ domains to settle onto the membrane in optimised orientations through refining inter-molecular interactions. In both PC:PE:PS simulations and PC:PE:PS:PI(4,5)P_2 simulations the domains maintained a similar orientation and pattern of lipid contacts, however while the tryptophan loop remained anchored into the membrane, they were able to tilt on the membrane increasing contact across the base of the protein (**Figure 3.13**). The positively charged loop lost contact with the membrane on the PC:PE:PS bilayer, which is in contrast to CG simulations where it binds stably. This provides an explanation for the high RMSF in this loop, where it is able to move freely in the solvent. Presumably the lower charge density of only PS in the membrane is not strong enough to maintain binding of this loop. In the PI(4,5)P_2 containing membrane this loop remains bound to the membrane, being stabilised by PIP contact.

In addition to hydrophobic and charge-charge interactions, the contribution of hydrogen bonding to membrane binding was investigated by calculating the frequency of hydrogen bonds to lipids in PC:PE:PS and 100 % PE bilayers (**Figure 3.16**). The tryptophan tipped loop made multiple contacts with PE, while across the rest of the base of the domain, S878 and S935 hydrogen bonding was significant. PS bonding was identified around His 907, matching lipid contact analysis. Frequency of hydrogen bonding to the 100% PE bilayer was calculated to continue investigations into the cause of high PE affinity. Overall frequencies of hydrogen bonds were significantly higher than in the PC:PE:PS bilayer, indicating that hydrogen bonding plays a greater role in overall protein binding in this membrane, while S878, S935 and W1056 were also identified as important PE bonding partners.

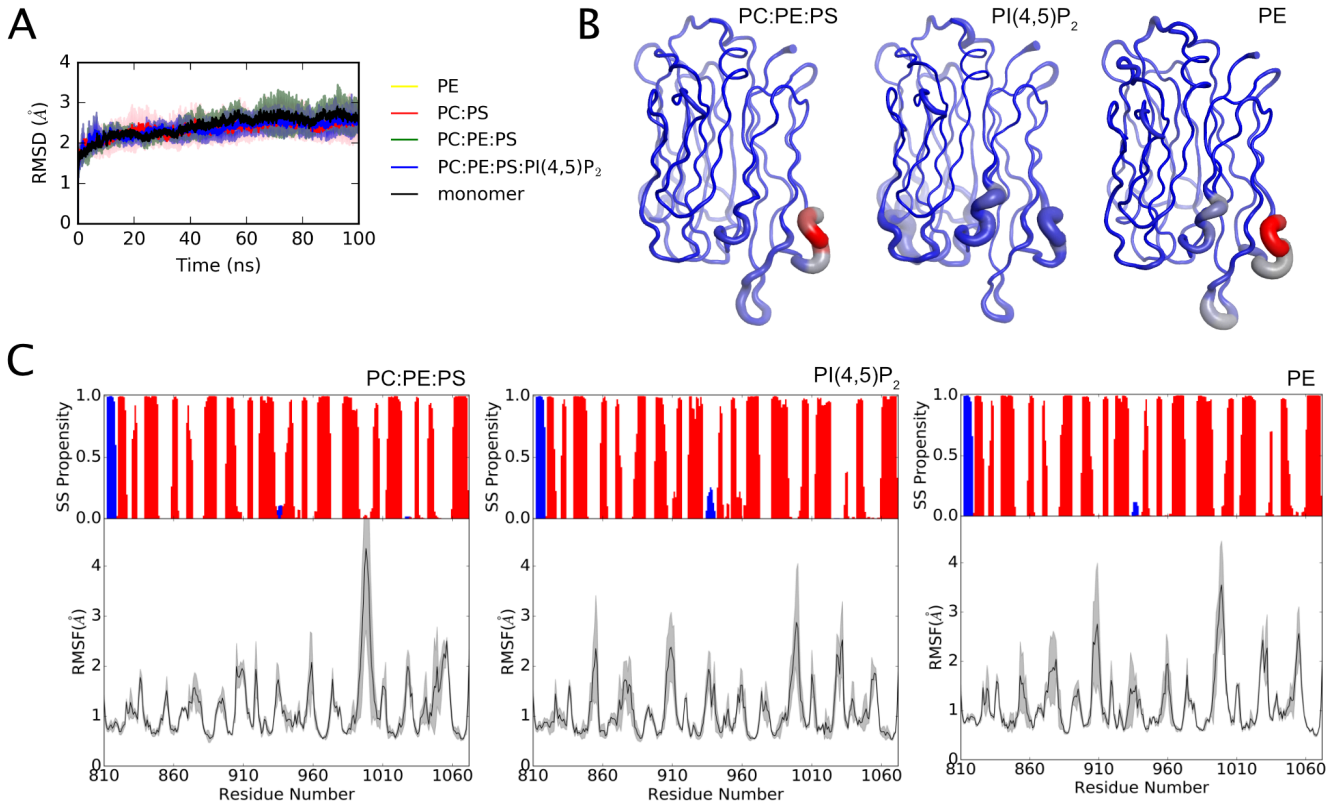


Figure 3.15 Stability of TgPLP1 membrane bound and soluble APCβ domain.

CG2AT simulations of the APCβ domain on PE, PC:PE:PS and PC:PE:PS:PI(4,5)P3 membranes, and in water, were run for 100 ns. (A) RMSD for each simulation. (B) Cartoon representations RMSF values on each bilayer plotted onto the structure. (C) Secondary structure propensity and RMSF calculations for the protein on each membrane. Mean values and standard deviation for RMSD and RMSF were calculated from 3 repeat simulations.

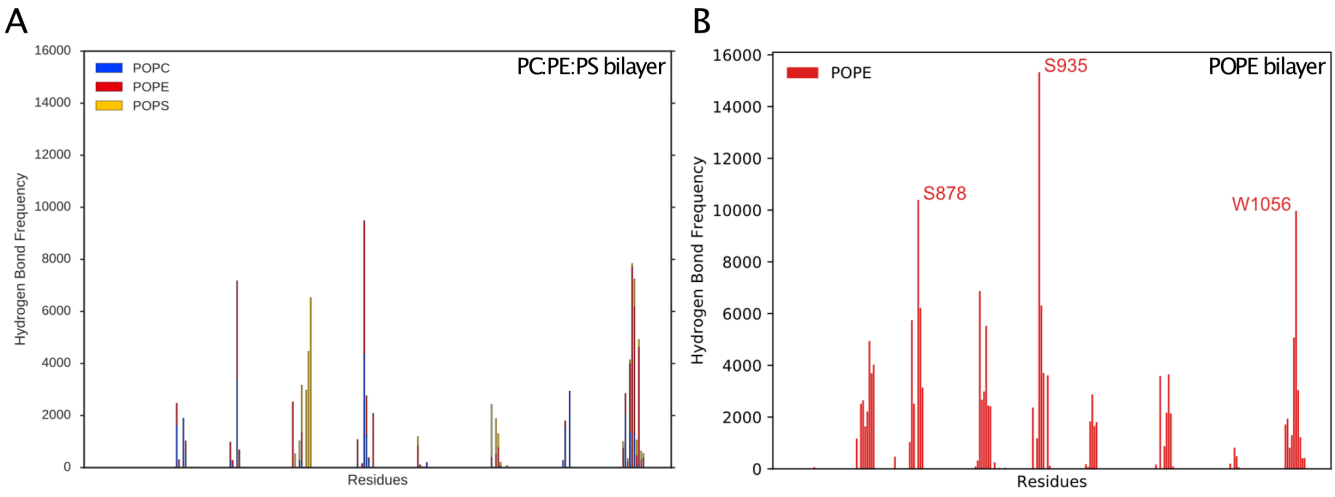


Figure 3.16 TgPLP1 APCβ domain protein-lipid hydrogen bonds

Mean frequency of hydrogen bonds were calculated for each residue in PC:PE:PS and PE simulations from three repeat simulations.

3.2.9 APCβ membrane binding discussion

The family of MACPF/CDC proteins typically use a C-terminal domain to bind to protein or lipid receptors on cell membrane before oligomerisation and pore formation. Previous studies have shown that both the C-terminal domain and N-terminal domain of TgPLP1 can bind to membranes, but only the APCβ domain is critical for pore formation (182). Following the structural determination of the TgPLP1 APCβ domain, we hypothesised that the domain would bind to membranes using the extended hydrophobic tryptophan tipped loop, a mechanism similar to that used by other MACPF/CDC membrane binding domains. We have shown through membrane association simulations that the domain binds to the membrane using this loop. Lipid contacts showed consistent binding of this loop throughout simulations, and mutation of the tip tryptophan alone to alanine significantly reduced binding in free energy calculations, which was confirmed in liposome sedimentation assays showing that mutation of the hydrophobic loop significantly reduced binding. This was consistent with later studies that used CRISPR/CAS9 gene editing to generate mutants affecting the ^{1055}LWL^{1057} tip motif, showing defects in parasite egress equivalent to knockout of PLP1 entirely (187). Intrinsic tryptophan fluorescence was also used to show that tryptophan on the surface of the protein inserts into liposomes upon binding. The use of tryptophan and other aromatic residues by peripheral membrane proteins to anchor into the membrane has been seen for MACPF/CDC C-terminal domains and other peripheral membrane proteins (188-190).

While an upside-down orientation of binding was also observed upon addition of PIPs in membrane association simulations, analysis of these contacts showed they were caused by binding of PIPs to several charged residues at the N-terminus of the domain, which would likely be shielded from the membrane through steric effects of the MACPF domain in the full-length protein, and PMF calculations show a reduced affinity for membrane in comparison to wild type protein.

A preference for binding to PE, PS and PIPs was identified in membrane association simulations, and further confirmed using free energy of binding calculations. While the protein has an affinity for PE and PS individually, complementation of both led to greater binding, while addition of PI(4,5)P_2 or PI(3,4,5)P_3 significantly increased binding, likely due to the increased charge density of PIPs. These results were consistent with the finding of a preference for PE, PS and PI(3,4,5)P_3 containing membranes in liposome sedimentation assays in our own work (91) and the recently published work by Guerra *et al*. (187).

Basic and polar residues across the base of the protein were in contact with the membrane and found to contribute to lipid specificity. A basic loop in the third pseudo-repeat of the structure was identified with motif ^{992}NANKR^{996}. K^{995} and R^{996} had strong binding to PS and PIPs in coarse grained simulations indicating that electrostatic interactions were important for binding, while atomistic simulations showed that this loop particularly favours PIP contact. The greater detail and flexibility available to the protein

in atomistic simulations highlighted that this loop was only able to bind to the membrane with PIPs present, and therefore PIPs stabilised binding of the protein as a whole on the membrane. Whether the APCβ has a specific affinity for PIPs, or if interaction is solely charge-based, is unknown as no other charged lipids were tested. Surface charge neutralisation through alanine substitution of all surface arginines and lysines, that included K^{995} and R^{996}, showed a reduction in binding to liposomes (**Figure 3.11**), however this basic loop was not targeted specifically. Further mutations, both in PMF calculations and liposome sedimentation assays, could be performed to assess the specific contribution of this loop to binding.

Several further charged and polar residues across the base of the protein were also found to maintain consistent lipid contact throughout simulations, including R^{932}, H^{907}, H^{1052} and serines S^{878} and S^{935}. The serines in particular maintained frequent hydrogen bonds with PE lipids, suggesting they may be partly responsible for the strong affinity for PE, along with the general increased capacity of PE to form hydrogen bonds across the protein binding surface. Mutation of H1052 to alanine reduced binding to liposomes in sedimentation assays, however the effect was not as significant as the hydrophobic or charge neutralisation mutations (91).

These analyses suggest that a combination of hydrophobic, electrostatic interactions and hydrogen bonding are responsible for membrane binding of TgPLP1. While the hydrophobic extended loop is critical for binding, the preference for PE, PS and PIP lipids, and increased binding in their presence, suggests that additional electrostatic interactions may be important for targeting of the protein to specific membranes. We found only weak binding of the APCβ domain to PC membranes. This finding was supported by liposome assays with full-length TgPLP1 and APCβ domain individually that showed an affinity for PE and PS, but not PC, sphingomyelin or cholesterol (187). Eukaryotic membrane inner leaflets contain higher concentrations of PE, PS and PI lipids, while the outer leaflet is enriched in sphingomyelin and cholesterol, contains little PE, PS or PI (1). TgPLP1 is localised to micronemes within the protists, secretion of which is triggered both during cellular invasion and cell egress after the replication cycle (98). This suggests that it may be exposed to membranes during invasion as well as egress and would require specific regulatory mechanisms to control pore forming action and prevent off-target damage. The preference for inner leaflet lipids such as PE and PS, and lack of binding to outer-leaflet lipids, may play a role in regulation, preventing cellular damage during invasion. This theory was proposed by Guerra *et al.* (187) for TgPLP1, however the role of the membrane binding domain of MACPF/CDC proteins is well established for other family members. CDCs require presence of cholesterol as a receptor in the membrane, binding using a cholesterol recognition motif. In addition, pH was shown to increase binding of the CTD (104). As invasion occurs at neutral pH, while PV acidification occurs during egress, this was proposed as a mechanism of regulation preventing activity during invasion. The role of pH in membrane binding has not been investigated here, however the

identification of several lipid binding histidines (H^{907} & H^{1052}) at the base of the protein provides clues to the mechanism of pH binding regulation and may warrant further study.

Multi-scale approaches to analysing membrane binding in peripheral membrane protein

Membrane binding of the APCβ domain was explored using a multi-scale approach, with CG membrane association simulations used to identify binding orientations followed by conversion to atomistic for more detailed analysis (133, 158). These were then used to identify lipid binding residues, complemented by the use of umbrella sampling and free energy calculations. These computational approaches have previously been used successfully to study membrane binding of peripheral membrane proteins, including PH domains (132, 191). These studies, along with results in our lab, have shown good agreement between computational and experimental results, thus supporting this approach for the study of peripheral membrane binding proteins, including domains such as APCβ, which has a broader affinity for membranes and less of an established foundation of research upon which to validate computational results.

One notable result was the observation of an additional upside-down orientation in the presence of $PI(4,5)P_2$ and $PI(3,4,5)P_3$ in the membrane. As discussed above, these results are thought to be artefactual, due to the isolation of the APCβ domain alone exposing normally buried residues at the MACPF-APCβ domain interface. The role of these residues has not been supported in mutational membrane binding studies, and this orientation did not have as strong an affinity for membranes as the upright orientation, as shown by PMF calculations. Additionally, the protein surface in this location does not have hydrophobic characteristics associated with PMPs. These results only occurred in membranes with increased charge density, and care should therefore be taken when adding PIPs into the membrane to identify any unexpected and aberrant results.

Lipid contact analysis was performed on both coarse grained and atomistic simulations to assess membrane binding at different scales. These approaches in general gave similar results, with the reduced resolution in CG simulations allowing exploration of membrane association and lipid binding on longer times-scales (192), followed by conversion to atomic detail allowing optimisation of lipid contacts and analysis of hydrogen bonding.

A trade-off between atomistic detail and sampling time must be taken into account when analysing the finer details of lipid-residue contacts. In some cases, residues that appeared important for lipid binding in coarse grained simulations did not appear significant in atomistic simulations and vice versa. For example, atomistic simulations containing $PI(4,5)P_2$ identified residues K851 & $^{878}SET^{880}$ had high $PI(4,5)P_2$ contact, while these residues did not have high PIP contact in coarse grained simulations (**Figure 3.13**). While these residues appear to be involved in $PI(4,5)P_2$ binding, they were likely overemphasised in atomistic lipid contact analysis as the PIPs were not able to diffuse and explore the protein surface during the simulation time scale.

This highlights a draw-back of atomistic simulations in that significant computational power is required to simulate large systems for time scales long enough to allow diffusion and equilibration of lipids. These simulations were run for 100 ns, which was deemed enough time to allow the protein to settle into an optimised orientation on the membrane, and to analyse atomistic details such as hydrogen bonds. However, when analysing contacts with lipids, such as $PI(4,5)P_2$, which are rarer in the membrane, the time frame is not long enough to allow significant diffusion and sampling of contacts. This also emphasises the importance of carefully choosing starting simulations in which important lipids are already bound in an optimal orientation identified through CG simulations. While attempts were made to choose starting structures that reflected these interactions, the results still appeared to overemphasise some interactions while missing others. On the other hand, AT simulations proved useful to identify important PE and PS binding residues, including S878 and S935 which were not identified in CG simulations.

The combination of reduced detail and use of elastic networks in Martini approaches mean that detailed analysis of protein stability is not accessible. To this end, it was interesting to observe the role of $PI(4,5)P_2$ in stabilising the positively charged loop adjacent to the membrane anchoring loop, which was only picked up in atomistic simulations.

3.2.10 Modelling the TgPLP1 transmembrane pore

Currently there is no structure of the TgPLP1 transmembrane pore. As we have now assessed the oligomerization state of the MACPF domain, we aimed to use this knowledge to predict the structure of the TgPLP1 pore transmembrane β-barrel, with the aim of then combining this with the MACPF and APCβ domain structures to build a full model of the TgPLP1 pore. Upon oligomerization into a pre-pore structure, MACPF/CDC proteins undergo a transition involving structural rearrangement of two TMH segments either side of the centre twisted β-sheet. Each TMH transitions from a pair of helices into a β-hairpin then inserts into the membrane. Each monomer contributes two hairpins that bond together to form a transmembrane β-barrel.

Bioinformatic analysis of TMH sequences

As we have not observed pore formation of TgPLP1 directly, we initially performed bioinformatics analysis of the TMH sequences. Firstly, we wanted to confirm that TgPLP1 does indeed form transmembrane pores, as would be predicted based on the presence of the MACPF domain. Secondly, we wanted to predict which residues within the TMH regions would contribute to the transmembrane section of the β-barrel and identify where the hairpin turn is located.

Transmembrane segments of β-barrel sequences typically consist of alternating hydrophobic and hydrophilic residues that alternately face the membrane core and pore lumen respectively. β-strands are typically at least 10 residues in length, although this can vary depending on the shear value of the barrel

(193). Various methods exist to predict transmembrane β-barrels based on protein sequence. We used three prediction softwares, BWARE (194), BOCTOPUS 2 (195), and PRED-TMBB 2 (196), to predict transmembrane segments in TgPLP1. The majority of β-barrels are bacterial outer membrane proteins. As such, most β-barrel prediction software has been solely trained on outer membrane proteins, meaning they may not perform well on other sequences (197). In particular, β-pore forming toxins in which transmembrane segments are not contiguous may be challenging to predict.

We therefore tested several sequences including TMH1 and TMH2 independently, TgPLP1 MAPCF domain, and a concatenated sequence consisting of 6 copies each of alternating TMH1 and TMH2 segments, which would mimic the β-barrel segment of the transmembrane pore. All but one software/sequence combination yielded negative results (**Table 3.3**). PRED-TMBB 2 predicted 10 transmembrane strands in the full TgPLP1 MACPF sequence, four of which spanned the TMH sequences, two within TMH1 and another pair within TMH2.

SOFTWARE		SEQUENCE	TM?	PREDICTED STRANDS	STRAND SEQUENCE
PRED-TMBB (196)	2	TgPLP1 MACPF	Yes	10	506:513, 558:566, 573:581, 668:682, 684:691, 705:715, 718:725, 744:750, 796:803, 805:811
		B-barrel	No		
		TMH1	No		
		TMH2	No		
BWARE (194)		TgPLP1 MACPF	No		
		B-barrel	No		
		TMH1	No		
		TMH2	No		
BOCTOPUS 2 (195)		TgPLP1 MACPF	No		
		B-barrel	No		
		TMH1	No		
		TMH2	No		

Table 3.3 Results of transmembrane β-barrel prediction. Residues highlighted in red are within TMH regions.

PSI-PRED was also used to predict secondary structure of the TMH regions, which did not provide clear predictions for β-stranded regions.

As the TMHs transition from alpha-helix to β-hairpin, and from soluble to transmembrane, upon pore formation, these sequences are unlikely to conform to standard parameters used for both secondary structure and β-barrel prediction. As prediction methods were only moderately successful, we also used manual methods to predict regions of transmembrane β-sheet and identify residues facing the pore lumen and membrane. Assuming that the TMH segments form a continuous extension of β-sheet from the MACPF core, without forming β-bulges which would offset the sequence, we calculated the location of the hairpin turn based on sequence length extending from the MACPF domain (**Figure 3.17**). Two glycine residues exist at the predicted turn in TMH1, residues that would be associated with a tight β-turn. Residues either side of the predicted turn displayed a somewhat alternating pattern of hydrophobic

and polar residues, although not perfect. These predictions also match up well with PRED-TMBB 2 predictions within TMH1. TMH2 also did not contain a strong pattern of alternating hydrophobicity, however the same location of the β-turn was predicted using both manual prediction and PRED-TMBB 2, also consisting of two glycine residues.

Despite the lack of a strong alternating amphipathic sequence, both predicted transmembrane β-strands were high in Gly, Thr, Ala, Leu and Val residues which are known to be higher in concentration in transmembrane β-barrels (198), and regions were predicted to best optimise partitioning between polar and hydrophobic residues on the favourable surface of the barrel.

Predicting β-barrel parameters

We hypothesised that we could model the transmembrane pore by using interfaces identified in the MACPF oligomeric crystal structures to predict characteristics of the β-barrel structure. B-barrel structures can be classified using two parameters: the number of strands (n), and the shear number (S) (199). Using the hexameric oligomeric structures, we therefore know that the transmembrane barrel should contain 24 strands. In general, structures of MACPF/CDC pores show a continuous β-sheet formed by the TMH hairpins, extending bonding from the core MACPF twisted sheet down through the barrel. Therefore, as we have already analysed hydrogen bonding patterns in the oligomeric crystal structures of the MACPF, we can use these structures to predict the shear of the barrel. We took the TMH sequences and calculated predicted hydrogen bonding patterns between TMH strands extending from the bonding present in the oligomers. Within the MACPF domain, the hydrogen bonding patterns will be consistent in all structure, however the interface bonds vary depending on whether we use the helical or ring interfaces.

We then used this to calculate the shear of the several barrels each with a different interface bonding pattern. Using bonding patterns in the helix oligomer, we calculated a shear value of 36.

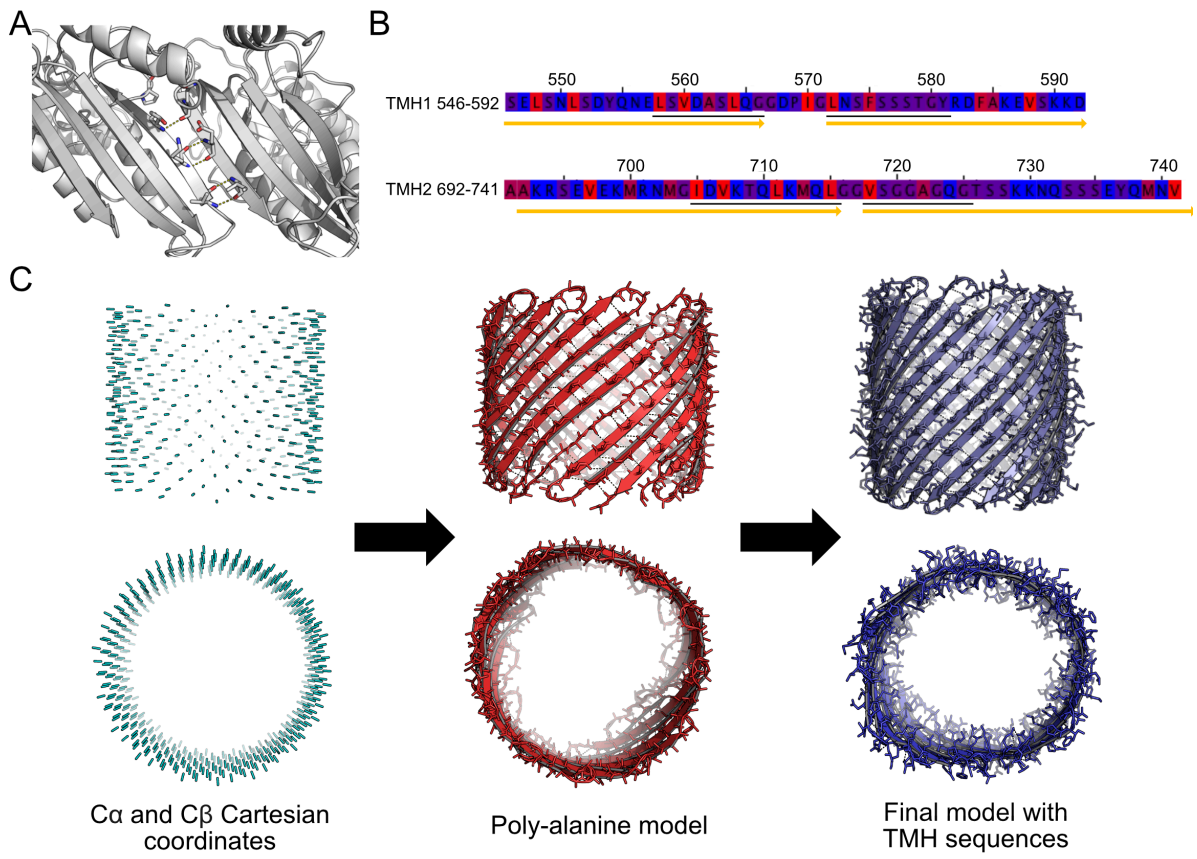


Figure 3.17 TgPLP1 β-barrel modelling procedure.

(A) A ribbon representation of the TgPLP1 MACPF helical interface showing main-chain β-sheet hydrogen bonds used to calculate bonding pattern for the TMH strands. (B) Sequences of TMH1 and TMH2 coloured by hydrophobicity show an alternating pattern of hydrophobic and polar residues. Yellow arrows represent predicted β-strands. Black bars represent predicted transmembrane regions of the β-strands. (C) The β-barrel modelling procedure. Cartesian coordinates of Cα and Cβ atoms placed according to strand number, length and shear. Next, a polyalanine model was built using the cartesian coordinates, and hydrogen bonding and secondary structure was applied. The alanine model was then replaced with the predicted TMH sequences, with additional alanine turns at the top of the barrel to close adjacent free strands to produce a final model.

Model building procedure

Ab initio modelling of the transmembrane β-barrel was started following calculation of the β-barrel parameters (**Figure 3.17**). Firstly, an in-house script was used to generate a PDB file containing coordinates for the Cα and Cβ atoms of each reside. Coordinates were automatically calculated based on the input strand number (24), strand length and shear value (36). The script available could model a maximum length of 20 residues per strand, meaning that in the final procedure additional length was added to the barrel later in the modelling procedure. Modeller (186) was then used to replace each atom in the PDB with alanine residues, and manually apply hydrogen bonds and secondary structure restraints based on previous calculations.

Using the prediction of hydrogen bonding and location of β-turns calculated previously, alanine residues were then replaced with the predicted strand sequence. Several iterations of the model were produced using slightly different hydrogen bonding patterns, based on either interfaces from the ring or from the helix oligomeric structures. The free termini at the top of the β-barrel, which would extend from the MACPF domain, were connected to hold the barrel shape. Two alanines were added to each strand and connected together to create poly-alanine turns.

Firstly, β-barrels only 20 residues in length were modelled, however to create a more complete model that could be directly connected to the current oligomeric MACPF structures, a further series of barrels were created to extend the length of the barrel to the full TMH length extending from the MACPF domain core twisted β-sheet. In this case, the starting barrel was modelled as above, however the top of the TMH strands directly connecting to the MACPF core were modelled first, and polyalanine turns were used to connect both free ends of the barrel. Once this model was refined, the polyalanine turns at the base of the barrel were removed, and the additional 6 residues completing the full TMH hairpins were modelled in. Hydrogen bonding and secondary structure restraints were also applied to these additional residues as predicted above. This also enabled modelling of tight 3-residue turns at the base of the β-barrel.

Before simulating, several prediction servers were used to predict the membrane positioning of the final structure. The OPM database PPM server (200) and the MEMEMBED server (201) were both tested and provided differing results. The PPM server predicted full membrane insertion of the segment of β-barrel predicted to be transmembrane, with the turn of each TMH positioned at the membrane-solvent interface. On the other hand, MEMEMBED only predicted insertion of the turns of the β-barrel into the membrane, indicating the barrel exists as more of a peripheral membrane protein.

The final β-barrel model

The β-barrel was modelled with the transmembrane TMH regions located at the base of the barrel, with the entire portion of β-hairpin extending from the transmembrane region up to the central MACPF domain modelled in above. Hydrogen bonding between strands of each MACPF ‘monomer’ were taken

from the MACPF domain, while interface strand bonding was taken from the helix. The turns of TMH hairpins were modelled tightly as three-residue β turns, enabled by glycine residues at the turns. An average of 10 additional residues per strand extended up from the transmembrane regions, followed by 4 alanine residues (2 per strand) to accomplish the turn at the top of the barrel.

Residues facing into the barrel lumen in the transmembrane region were strongly polar, with strand 1 of TMH2 also contributing several charged residues. The predominant characteristic of residues in the predicted transmembrane region facing outwards from the barrel had small hydrophobic side chains, with the notable inclusion of 3 threonine residues. Both internal and external facing residues in this region are predominantly polar and charged, in contrast to the alternating nature of the transmembrane region.

Membrane insertion of the β-barrel

Membrane self-assembly simulations were first used to evaluate the β-barrel model's propensity to insert into a membrane. Using an in-house script, MEMEMBED was used to predict placement of the β-barrel into a membrane. As MEMEMBED did not predict the barrel to be inserted into the membrane, the 'force' option was used to ensure the lipids were placed around the barrel correctly. PC lipids were then placed randomly around the barrel in the plane predicted by MEMEMBED and allowed to self-assembly into a membrane (**Figure 3.18**).

Lysenin, a nonameric pore forming toxin of the aerolysin family was used as a positive control, for which both a 3.1 Å cryo-EM structure, and an x-ray crystal structure of the membrane inserted protein has been solved (202, 203). Full membrane insertion of lysenin was observed during membrane assembly (**Figure 3.18**). In preliminary simulations, the TgPLP1 β-barrel was rejected from the membrane during assembly, finally resting along the surface of the membrane (**Figure 3.18**).

The final β-barrel model, which was used within simulations discussed from this point, was fully inserted following membrane self-assembly and formed a central solvent filled pore.

The results of 5 repeats are summarized in **Table 3.4**, showing that all but one repeat resulted in the insertion into the membrane. In one repeat, the barrel did insert into the membrane however insertion was only partial, with the barrel sitting halfway into the membrane and tilted at an angle. The lower leaflet of the membrane formed a partial toroidal structure to compensate for this semi-insertion.

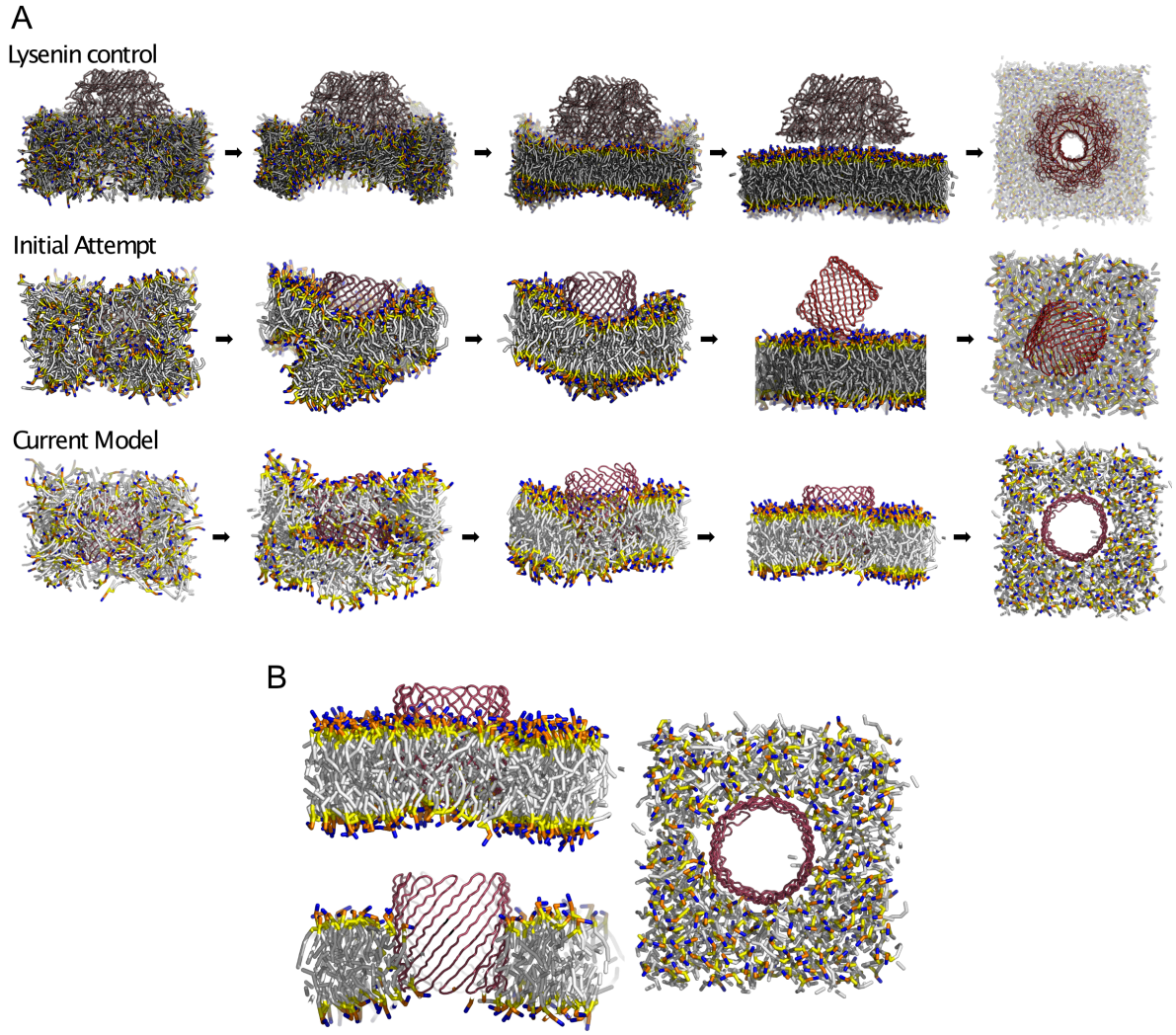


Figure 3.18 Membrane self-assembly simulations of the TgPLP1 ß-barrel model.

(A)Snapshots from the membrane self-assembly procedure showing initial lipid placement around the centre of the barrel, followed by lipid rearrangement into a complete bilayer during simulations. Results from the lysenin control, an attempted β-barrel that is shorter in length than the final barrel, and the final barrel are shown. The final central lumen of the barrel can be seen in simulations with lysenin and the final model, while the earlier barrel model did not insert into the membrane. (B) The final β-barrel model inserted into the membrane showing barrel positioning in the membrane.

The RMSD and RMSF of the barrel model were calculated for a representative self-assembly simulation, showing that the overall core of the barrel was stable in coarse grained simulations, with greatest fluctuation at the poly-alanine turns at the top of the barrel. This would be expected as these turns were modelled in to connect termini that would extend up to the core MACPF domain and are non-physiological (**Figure 3.19 A**). CG2AT conversion was then performed using the final frame of this simulation and simulated for 100 ns. Basic structural analysis showed that the barrel was stable in the simulation, with an average

RMSD of ~ 0.42 nm (**Figure 3.19 B**). In agreement with the CG analysis, the β-strands of the barrel remained stable, with more fluctuation at the turns, particularly at the poly-alanine turns at the top of the barrel. Calculation of secondary structure and maintenance of hydrogen bonds throughout the simulation time complemented this analysis in showing that both factors also remained consistent throughout the simulation (**Figure 3.19 C&D**). The barrel also remained stably inserted into the membrane throughout the simulation.

Repeat	*Inserted?*
1	Yes
2	No
3	Partial Insertion
4	Yes
5	Yes

Table 3.4 Results of 5 repeats of PC membrane self-assembly around the final β-barrel model.

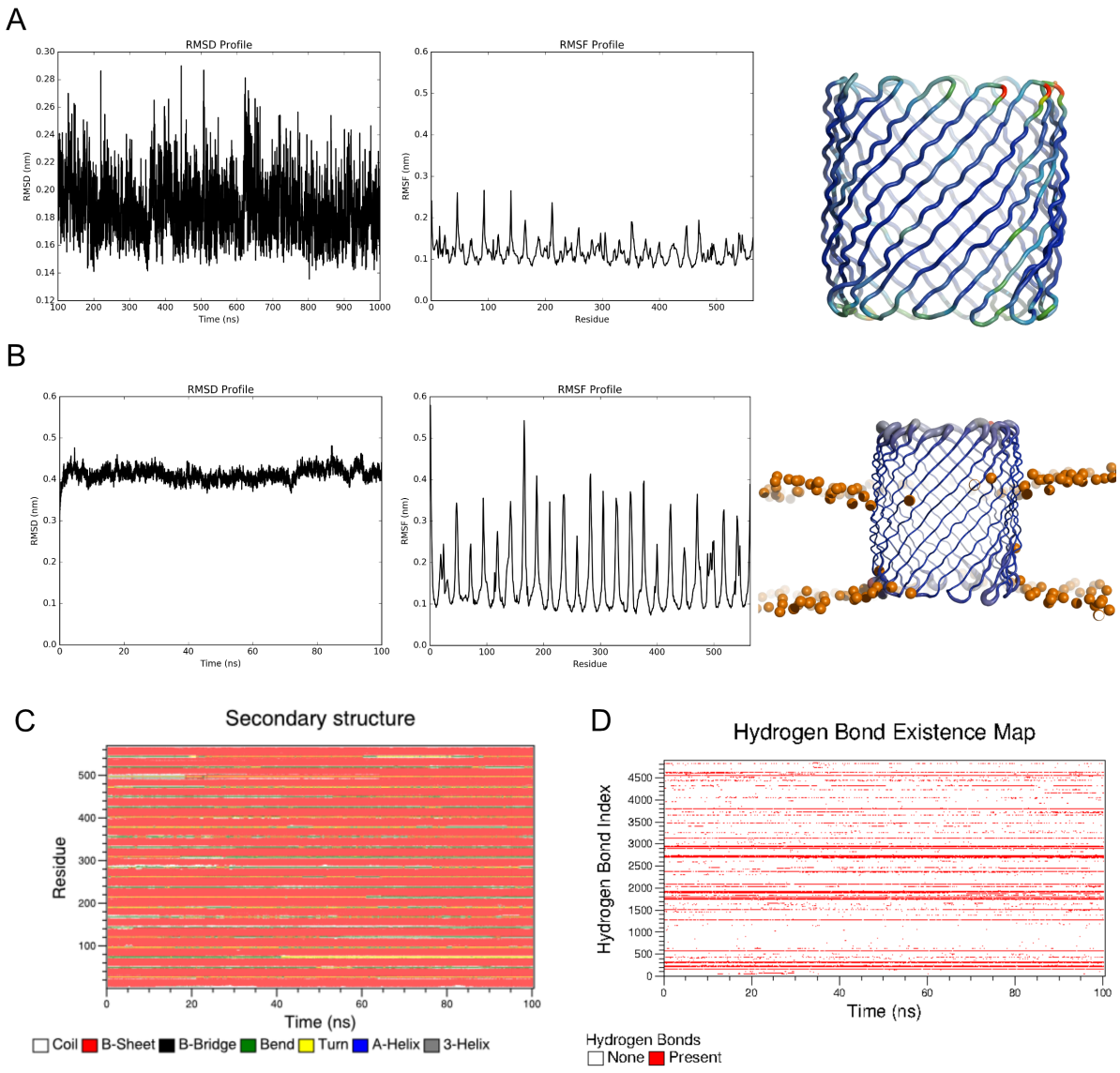


Figure 3.19 Stability of the TgPLP1 β-barrel in CG and AT simulations

(A) RMSD and RMSF calculations of the β-barrel model in a PC membrane following membrane self-assembly. RMSF is plotted onto the structure of the barrel to indicate areas of flexibility. (B) RMSD and RMSF profile of the same β-barrel model following CG2AT conversion of the self-assembled system and simulation for 100 ns. Lipid head group phosphates are shown as orange beads to indicate bilayer location. (C) Secondary structure and (D) hydrogen bond existence calculated across the 100 ns AT simulation.

β-barrel membrane interactions

We first observed the orientation of the β-barrel in the membrane once inserted. In all cases where the barrel was fully inserted into the membrane, the β-turn at the end of the TMHs sat in the plane of the lipid headgroups rather than extending out of the bilayer. The transmembrane predicted β-strand regions displaying alternating hydrophobicity were situated within the bilayer, while the strands extending between the MACPF and transmembrane segments sat above the bilayer as expected. A thinning of the bilayer around the β-barrel was also observed in both CG and AT simulations. This indicated a hydrophobic mismatch between the thickness of the membrane and that of the hydrophobic segment of the barrel.

Protein-lipid contacts were calculated for each simulation, firstly to validate consistency of contacts across the simulation (**Figure 3.20**). Contacts were calculated separately for lipid tails and headgroups and normalized for each simulation. Contacts for all self-assembly simulations show that repeats 1, 4 and 5 were inserted into the membrane in the same way, while repeat two did not insert into the membrane and repeat 3 was tilted, with some but not all contacts matching the fully inserted models. These favoured lipid contacts were consistent across CG and AT simulations, as seen in **Figure 3.20 B**.

Through zooming into lipid contacts of 4 strands of the β-barrel, representing one MACPF monomer, we can observe residue contacts in detail (**Figure 3.20 C**). The alternating patterns of contact with lipid tails can be seen within the transmembrane β-strands, with headgroup contacts being stronger at either end of the strand as expected. Some residues within the transmembrane β-strands are hydrophobic and would be expected to interact with the membrane core, such as leucines, valines, isoleucines, methionine and phenylalanine. However, some unexpected residues contact the lipid core, predominantly the polar residue threonine, three of which are found facing the lipid tails across TMH1 and TMH2. The residue side chain orientations in this model represent the best fit in optimizing placement of hydrophobic and polar residues in a membrane context. While it is not optimal having threonine facing the lipid tails, polar residues faced the lipid core due to the nature of the sequence. While threonine residues were found to be favoured in transmembrane β-structures (198), they would typically be orientation towards the pore lumen of the barrel to satisfy their hydrogen bonding capacity. Polar and charged residues lysine, arginine and asparagine tended to exist at the edges of the β-strands and were in contact with lipid head groups.

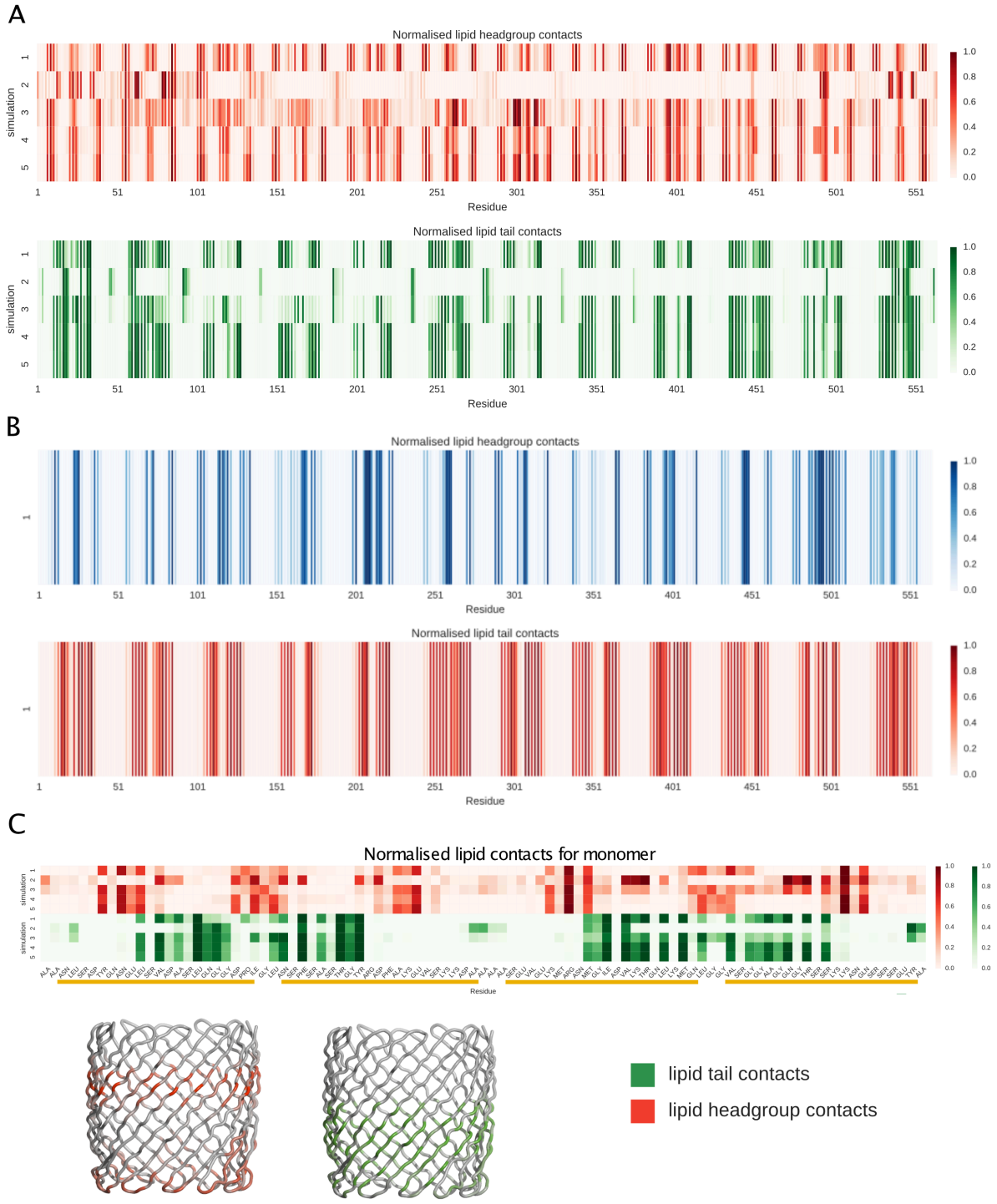


Figure 3.20 Protein-lipid contacts for the TgPLP1 β-barrel model.

Normalised protein-lipid contact fingerprint plots for (A) CG self-assembly simulations and (B) the CG2AT simulation. Contacts were calculated independent for the lipid head group (red/CG and blue/AT), and tails (Green/CG and Red/AT). (C) Normalised lipid contacts for the equivalent of one MACPF monomer from CG self-assembly simulations. Orange arrows show the location of β-strands. Contacts are plotted onto a representation of the β-barrel structure.

β-barrel discussion

Taken together these results suggest that this β-barrel model is stable in structure. The barrel model maintained low RMSD and RMSF values, and secondary structure and H-bonds in atomistic simulations, indicating the residue placement and location of turns is favourable.

In our experimental studies, we were unable to directly observe TgPLP1 pore formation. However, the computation studies described here suggest that the TgPLP1 TMH strands are able to favourably interact with the membrane to form a characteristic MACPF transmembrane pore. Prediction software does not work well for this class of proteins, consistently failing to identify transmembrane regions, except for PRED-TMBB2 which predicted roughly the correct TMH regions, but also a large number of false positives. The failure of these β-barrel predictive software is likely due to the fact that they are trained solely on bacterial outer membrane β-barrel proteins. In addition, the nature of the TMH segments in transitioning from α-helical to β-sheet structures is likely to impose limitations on residue composition at these regions. However, while the predicted β-strand region did not have strong amphipathic qualities, the resulting structure was effectively inserted into membranes in self-assembly simulations and was stably maintained in both coarse-grained and atomistic simulations. The membrane maintained its bilayer properties, rather than forming toroidal edges, indicating that the strands are hydrophobic enough to interact favourably with the core of the lipid bilayer.

Membrane thinning was seen around the barrel, induced by hydrophobic mismatch between the barrel and membrane in all simulations. Only PC bilayers were tested here, so it would be interesting to observe any differences with different lipid types, however the mismatch was significant and induced thinning of both leaflets of the bilayer. Whether this induced thinning of the bilayer could have any functional role during or after pore formation is unknown and would be interesting to investigate further. The precise role of pore formation during *Toxoplasma* egress is unknown, but pore formation occurs prior to membrane disruption by the parasite after induction of motility (98). Along with allowing efflux of effector molecules form the PV, formation of clusters of TgPLP1 pores and subsequent thinning of the PV and PM membrane, is likely to weaken the membrane aiding in parasite escape.

Discussion on β-barrel modelling procedure

As cryo-EM methods are improving, solving structures of large transmembrane pore forming proteins is becoming easier, however significant time and resources are needed to determine new structures. Additionally, achieving high resolution of large oligomeric potentially heterogeneous structures, especially of transmembrane proteins, poses further challenges.

To address these questions, we aimed to develop a procedure to model the transmembrane β-barrel structure of MACPF proteins at atomic detail. Further, MD simulations would then allow investigation into the membrane interactions of the pore at details inaccessible using standard experimental

techniques. One question in particular is the structural role of lipids in pore formation, especially in MACPF/CDC arcs and split ring structures which include a free lipid edge (204).

Overall the procedure of calculating hydrogen bonding patterns, shear and secondary structure characteristics based on TMH sequence and the core MACPF domain was successful based on validation of the final model produced. It is clear that transmembrane characteristics of MACPF proteins are different from typical β-barrel OMPs, suggesting the need for improved predictive software to evaluate the different characteristics of β-pore forming proteins.

As the model is complete and can be stably inserted into a membrane, next steps in investigating the β-barrel characteristics and pore formation mechanism could involve using steered MD to pull the barrel into a pre-formed membrane, observing reorganisation of lipids to accommodate the barrel. Strands of the barrel could then be removed to form arc assemblies which could then be investigated using the same mechanism. The existence and stability of MACPF/CDC arcs has long been a subject of discussion, in particular the structural stability and characteristics of such a proteo-lipidic pores (204). MD would provide an ideal method to investigate such phenomenon. Umbrella sampling and free energy calculations could be used to complement these studies to investigate the energetics of membrane insertion.

In addition to observing structural characteristics of the barrel, the hypothesis that these pores are used to facilitate release of further effector molecules could be investigated by performing computational electrophysiology on the pore structure to investigate conductance and flux of small molecules across the pore (205).

Finally, it would be interesting to extend this protocol to other MACPF/CDC proteins to investigate membrane interactions across the family of proteins. In order to do this, work would have to be carried out to automate this protocol. Firstly, modification of the β-barrel building software to extend the length and strand number of barrels it is capable of producing would be needed, followed by automation of modelling residues and hydrogen bonds onto the structure based on set hydrogen-bonding and sequence parameters.

3.3 Conclusions

We firstly looked at TgPLP1 MACPF domain, aiming to assess the stability and functional relevance of two oligomeric crystal structures. Results from AT simulations, hydrogen bonding analysis, and CG simulations observing large scale structural fluctuations, all pointed towards the conclusion that the TgPLP1 MACPF domain favours a helical oligomeric state.

In combination with recent work on the MAC complex and perforin-2, these results suggest that the interfaces between MACPF proteins may be flexible and able to accommodate tilting of the structures.

Whether or not the structure maintains a helical conformation on the membrane, or if the strength of membrane binding provides enough force to hold the oligomer in a flat planar conformation is unknown. Full interpretation of these results is difficult without further structural work to observe pores in-situ at high resolution; however, these results are in contrast with the accepted theory that MACPF/CDC proteins form circular pores or arc on the surface of membranes.

It will be interesting to see if future studies of MACPF/CDCs reflect the recent observations of split-washer pores and tendency towards helical formations. Technical advances in cryo-EM technology have led to improvements in resolution and particle image processing, meaning we are better able to identify heterogeneity in sample conformations which previously would have been averaged resulting in blurring (206). It is therefore possible that the split, helical conformation is more prevalent among MACPFs which was previously obscured by imaging techniques.

If this propensity towards a helical conformation is indeed favoured, this could provide a further mechanism by which TgPLP1 pores could weaken and tear membranes and facilitate parasite egress.

We next moved onto investigating membrane binding of the APCβ domain. After solving the structure we proposed that this domain would bind to membranes using the hydrophobic tryptophan tipped loop at its base, which is similar in character to the conserved undecapeptide motif seen in CDC membrane binding domains (188). Using a combination of MD simulations and membrane binding assays, we were able to show that the APCβ domain does indeed bind to membranes using this extended loop. Binding is further complemented by electrostatic interactions across the base of the protein which lead to a preference for binding to membranes containing PE, PS and PIP lipids.

These results strongly suggest that TgPLP1 has a preference for binding the cytosolic inner leaflet of plasma membranes and identified key residues across the structure that contribute to this lipid specificity. As membrane binding of MACPF proteins is a critical first step in pore formation, understanding the biochemical basis of membrane binding could contribute to the development of further treatments and vaccinations against toxoplasmosis in the future. It has been suggested that pH has a role in regulating and increasing membrane binding of TgPLP1 (104), therefore this identifies a further avenue of study of TgPLP1 membrane binding that could be pursued.

Once we had established a favoured interface for oligomerisation, and membrane binding orientation of the APCβ domain, we finally moved on to modelling the transmembrane β-barrel pore. This was done with the aim of putting together a final model of TgPLP1 transmembrane pore structure and being able to further investigate the biochemical and electrophysical properties of the pore.

Through development of a novel modelling procedure, we were able to produce a model of the transmembrane β-barrel based on shear calculated from the helical interface. This model was successfully inserted into a membrane and had properties expected of a transmembrane β-barrel.

Hydrophobic mismatch leading to membrane thinning was observed in simulations, which could contribute to weakening of the membrane for egress, especially through formation of clusters of pores.

Further work could also be done to characterise the transmembrane pore. We were able to see from the β-barrel model that residues facing the lumen were polar in character which in combination with the pore diameter would suggest that the pore is able to facilitate escape of small molecules across the PV and plasma membrane. Computational electrophysiology could complement these observations to understand the biophysical properties of the pore. Further, steered MD to pull the barrel, incomplete arc structures and single TMH segments into preformed bilayers, along with PMF calculations, could be used to investigate lipid rearrangement and the energetics of the membrane insertion step of pore formation to determine whether lipids are able to flow out from the pore into surrounding membrane during formation, or are expelled as vesicles from the membrane as has been observed using MD simulations (207). Finally, this procedure could be extended to study further MACPF/CDC pores including medically important PFPs such as perforin and MAC, as well as for biotechnological applications such as biosensing and sequencing where refinements in the transmembrane pore structure could lead to functional improvements.

Chapter 4

Membrane binding characteristics of *Plasmodium* PLPs

4.1 Introduction

4.1.1 Apicomplexan Perforin-like proteins

The plasmodium lifecycle is complex and dynamic, punctuated by critical periods of cell motility and traversal across host tissues, followed by cell invasion, intracellular replication and egress of daughter parasites. These processes require parasite modification and interaction with cell membranes from both the mammalian and arthropod host, and each is vital for successful disease transmission. Apicomplexan perforin-like proteins were first identified by Kaiser *et al.* in 2004 (90), who identified five MACPF-domain containing proteins, *Plasmodium* perforin-like proteins (PPLPs) 1-5. They posited that as *Plasmodium* has many varied and complex interactions with host membranes, it is likely that the PPLPs have a role in pore formation similar to other MACPF pore forming proteins used by pathogenic microbes for host cell invasion, destruction and escape. Reports since this date have shown that PPLPs function throughout the *Plasmodium* lifecycle in the processes of cell traversal and cell egress, however to date, no detailed structural or biochemical studies have been performed on the *Plasmodium* PLP family, and therefore many questions remain regarding the precise mechanism of action of these proteins. Here we have aimed to address the membrane binding characteristics of several members of the PPLP family. These interactions mediate initial cell recognition and binding. They are therefore critical to understanding PPLP function as a whole. Recognition of the varied environments these proteins are functioning within will also be important to identify functional diversity across the family.

4.1.2 *Plasmodium* PLPs function at different points in its lifecycle

Following injection of sporozoites into the skin by infected female *Anopheles* mosquitos, sporozoites traverse across both the dermis and liver hepatocytes before replicating in the liver. PPLP1 is critical for traversal through both dermal cells (110, 111), and across cells of the liver sinusoidal membranes and hepatocytes (112, 113). Disruption of PPLP1 leads to reduced infectivity in rodent models (112), while *in vitro* cell wounding assays showed PPLP1 is required for wounding cells, further supporting a role in membrane disruption (110, 112).

Risco-Castillo *et al.* (111) demonstrated cell traversal is achieved by formation of transient vacuoles (TV) derived from host plasma membrane, and showed that PPLP1 acts during egress from this TV. While the parasitophorous vacuole (PV) membrane is filtered and remodelled by parasites, the TV membrane still contained host proteins, indicating it is not re-modelled and remains similar to the host PM. Initial parasite entry is achieved without membrane damage and occurs in a PPLP1 independent manner. Fusion of host lysosomes within the TV is then proposed to induce acidification of the TV, and activation of *Plasmodium* egress. Release of PPLP1 at this stage facilitates escape from the TV and cell egress (111).

PPLP1 expression during asexual blood stage replication has been identified in *Plasmodium falciparum*, where it was found to localise to the RBC and PV membranes during egress and disrupt RBC membranes in a calcium dependent manner (114).

Evidence therefore suggests that PPLP1 acts to disrupt the TV membrane, and possibly host cell plasma membrane following entry into the cell during dermis and liver cell traversal. The mechanism by which this disruption is achieved, and how this facilitates parasite egress, is not known.

In contrast to PPLP1, PPLP2 is used by *Plasmodium* gametocytes to escape host erythrocyte cells following ingestion by the *Anopheles* mosquito (115, 116). Egress is activated by conditions in the midgut, leading to release of secretion of PPLP2 and other motility and egress effectors, which first rupture and escape from the PV membrane followed by the RBC membrane (117, 118).

Rupture of the PV membrane occurs within the first 2 minutes of activation and can occur in a PPLP2 independent manner (115, 116). Disruption of the RBC membrane occurs about 10 minutes post activation (117) and is facilitated by PPLP2, with gene knockout generating egress deficient mutants (115). Male gametocytes exit the RBC in a process termed exflagellation. While PV disruption occurred rapidly in PPLP2 knockouts (115, 116), exflagellation was reduced by 98.9% in *P. falciparum* (116). These were unable to disrupt the RBC membrane and remained enveloped in the RBC membrane, leading to impaired transmission (115, 116).

Some questions remain as to the differences in gender specific roles of PPLP2 across the *Plasmodium* species. PPLP2 expression and phenotypic effect was only identified in male gametocytes of *P. berghei* (115). On the other hand, PPLP2 has a role in egress in both male and female gametocytes in *P. falciparum* (116). In addition, PPLP2 expression has been identified in the asexual blood stage (114, 116), however replication occurs normally in PPLP2 mutants indicating that PPLP2 may not have a critical role in the asexual blood stage infection.

4.1.3 PPLPs 3, 4 and 5 act in cell traversal across the mosquito midgut

While the previously discussed PPLPs function in mammalian host tissue, PPLPs 3,4 and 5 act in the mosquito midgut to facilitate traversal across the midgut epithelia. Following escape from blood cells in the midgut, gametocytes form a zygote which develops into the ookinete stage which, once formed, rapidly traverses the mosquito midgut. On the basal side, ookinetes then form oocysts ultimately dividing into new sporozoites for another round of mammalian infection. Unlike PPLPs 1 and 2, PPLPs 3, 4 and 5 are all expressed during the ookinete phase (208-210). Disruption of PLP4 in *P. falciparum* (211) and *P. berghei* (212) showed mutant ookinetes accumulate in the mosquito midgut and are unable to traverse the midgut epithelium. PPLP4 gene disruption led to an accumulation of ookinetes in the blood bolus in mosquito, which rendered infected mosquitos unable to transmit infection to mice (212). A similar phenotype was observed for PPLP3 and 5 mutants, which both are unable to traverse midgut

epithelia (209, 210). However, in contrast to PPLP4, PPLP 3 and 5 mutants were able to escape the blood bolus and were found associated to the inner face of the midgut epithelia (209, 210).

Loss of any one of PPLP3-5 results in disruption of cell traversal, indicating that each of these proteins has a non-redundant role. The requirement for multiple PPLPs at this life cycle stage is unknown. It has been suggested that these proteins may form hetero-oligomeric pores, in a similar fashion to the formation of the MAC (211, 212). Loss of one PLP would then either prevent formation of pores or result in loss of pore stability. However, it was noted that PPLP4 was retained in the blood bolus, suggesting that it may have a role in disruption of the peritrophic membrane surrounding the blood bolus, while PPLPs 3 and 5 were able to cross this membrane and were retained at the midgut epithelial wall instead. Further direct evidence is required to assess the role of PPLP4 in crossing the peritrophic membrane, but this does not exclude the possibility of formation of heterooligomeric pores of either PPLP3 and 5, or PPLP3, 4 and 5 together.

4.1.4 Evidence for membrane disruption and pore formation in PLPs

Many questions remain as to the functional mechanisms with which PFPs disrupt membranes and facilitate parasite transport across the membrane. Most PPLP studies have focused on observing phenotypic changes in mutants which have indirectly suggested that PLPs function in membrane disruption during cell traversal and egress. That all PLPs contain a MACPF domain, alongside a putative N-terminal signal peptide, for trafficking to the cell secretory system, and C-terminal membrane binding domain, supports the hypothesis that this membrane disruption is facilitated by pore formation.

Direct evidence for membrane disruption by PLPs has been observed in cell wounding assays showing that PLP1 is capable of damaging cell membranes (110), while recombinant PLP2 facilitated RBC leakage in lysis assays (116). A study of PvPLP2 has suggested a mechanism by which this protein aids membrane disruption.

Live cell microscopy of gametocyte egress was used to demonstrate that erythrocyte membrane (EM) permeabilisation occurs several minutes before membrane rupture (116). This permeabilization was blocked in PPLP2- cells. Equinatoxin II, a pore forming toxin, was able to rescue the PPLP2 deficient phenotype (115, 116), while application of the pore sealant Tetronic 90R4 to wild-type gametocytes prevented EM disruption (116). These results indicate that PPLP2 disrupts the EM in a pore forming manner.

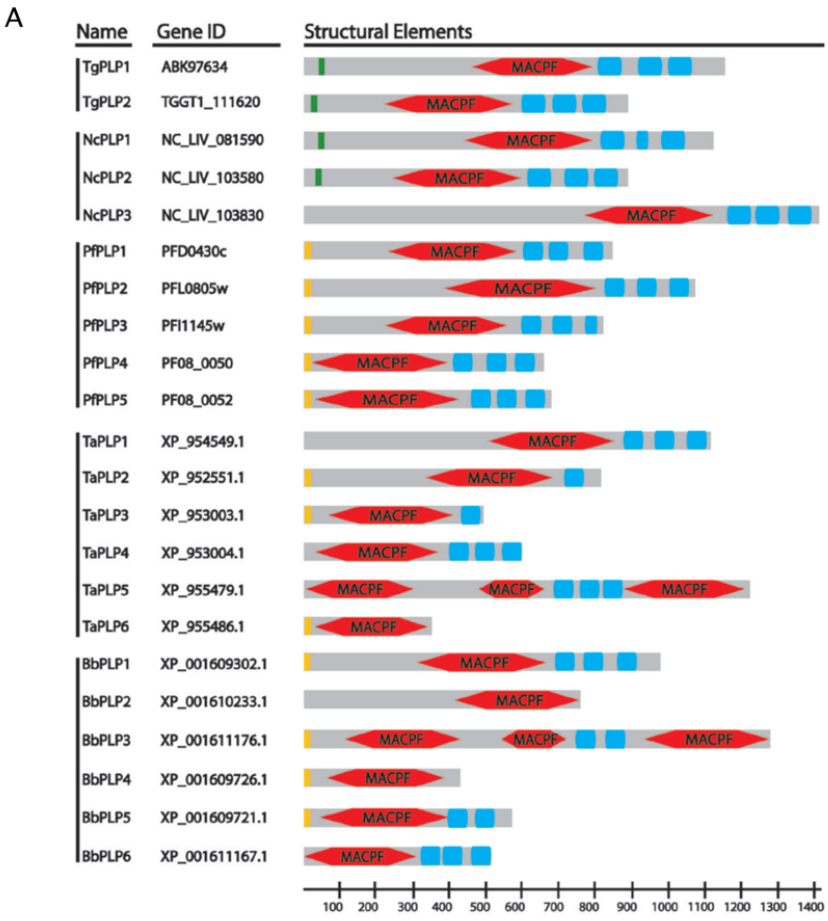
Further, Sologub *et al.* (117) showed that rupture of the EM occurs at a single point, indicating mechanical disruption of the membrane, possibly by rounding up of the gametocyte during activation. It is also known that proteases have critical roles in egress and may aid escape by degrading host cell cytoskeleton and are critical for membrane disruption and gametocyte escape (117, 118). These findings suggest a two-step mechanism of EM disruption, in which PPLP2 first permeabilises membranes,

allowing release of effector molecules such as proteases. These then weaken cell structures, allowing final rupture of the EM and escape of the gametocyte.

Our work on TgPLP1 (91) (**Chapter 3**) has provided a foundation of study of the ApiPLPs, suggesting these proteins are able to oligomerise into relatively small pores on membranes. Direct study of *P*PLPs is required to further investigate species differences in pore formation and mechanism of action. How large these pores are, and how they aid the parasite in crossing the membrane, are both critical questions.

4.1.5 Phylogenetic analysis of PLPs

Plasmodium PLPs contain N-terminal domains, a central MACPF domain and a C-terminal domain. Previous bioinformatic analysis showed that the C-terminal domains are β-pleated sheet rich, and unique to Apicomplexa, leading to their designation as ApiPLP C-terminal β-pleated sheet (APCβ) domains (43). Originally, pattern searches suggested these domains are ~55 amino acids long, and usually occur in triplicate at the C-terminus of the Apicomplexan PLPs (ApiPLPs) **(Figure 4.1)**. However, our recent structural studies of TgPLP1 (91) showed that each of the three beta-sheet rich repeats are in fact folded into a single stable domain consisting of the three pseudo-repeats arranged into a prism structure, which is responsible for membrane binding (see Chapter 3). The N-terminal domains are of variable length and have no identifiable domain structure or homology hinting at their functional role. In contrast to Plasmodium, not all Apicomplexan species have five predicted ApiPLPs. Toxoplasma gondii has only two predicted ApiPLPs, while *Babesia bovis* and *Theileria annulata* both have six predicted ApiPLPs **(Figure 4.1)**. While all are predicted to have at least one MACPF domain, *Theileria* and *Babesia* each have one PLP with three predicted MACPF domains. The number of APCβ pseudo-repeats predicted in several *Theileria* and *Babesia* PLPs vary from 0-3 repeats.


A
Name
Gene ID
Structural Elements
TgPLP1 ABK97634
TgPLP2 TGGT1_111620
NcPLP1 NC_LIV_081590
NcPLP2 NC_LIV_103580
NcPLP3 NC_LIV_103830
PfPLP1 PFD0430c
PfPLP2 PFL0805w
PfPLP3 PFI1145w
PfPLP4 PF08_0050
PfPLP5 PF08_0052
TaPLP1 XP_954549.1
TaPLP2 XP_952551.1
TaPLP3 XP_953003.1
TaPLP4 XP_953004.1
TaPLP5 XP_955479.1
TaPLP6 XP_955486.1
BbPLP1 XP_001609302.1
BbPLP2 XP_001610233.1
BbPLP3 XP_001611176.1
BbPLP4 XP_001609726.1
BbPLP5 XP_001609721.1
BbPLP6 XP_001611167.1
MACPF
100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400

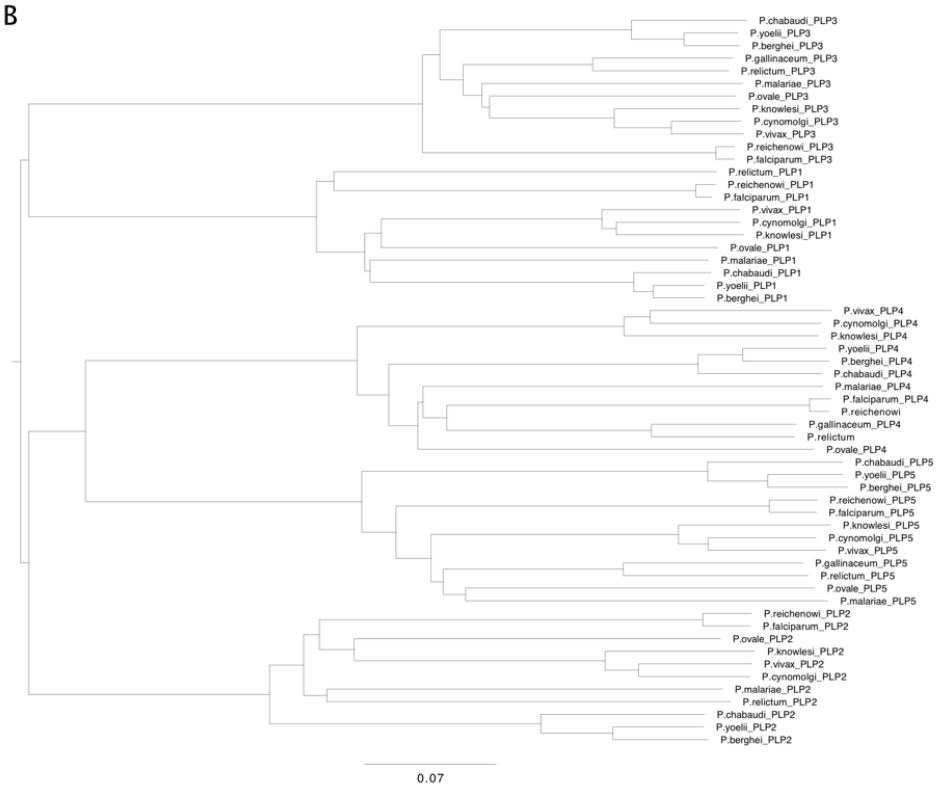

B
P.chabaudi_PLP3
P.yoelii_PLP3
P.berghei_PLP3
P.gallinaceum_PLP3
P.relictum_PLP3
P.malariae_PLP3
P.ovale_PLP3
P.knowlesi_PLP3
P.cynomolgi_PLP3
P.vivax_PLP3
P.reichenowi_PLP3
P.falciparum_PLP3
P.relictum_PLP1
P.reichenowi_PLP1
P.falciparum_PLP1
P.vivax_PLP1
P.cynomolgi_PLP1
P.knowlesi_PLP1
P.ovale_PLP1
P.malariae_PLP1
P.chabaudi_PLP1
P.yoelii_PLP1
P.berghei_PLP1
P.vivax_PLP4
P.cynomolgi_PLP4
P.knowlesi_PLP4
P.yoelii_PLP4
P.berghei_PLP4
P.chabaudi_PLP4
P.malariae_PLP4
P.falciparum_PLP4
P.reichenowi
P.gallinaceum_PLP4
P.relictum
P.ovale_PLP4
P.chabaudi_PLP5
P.yoelii_PLP5
P.berghei_PLP5
P.reichenowi_PLP5
P.falciparum_PLP5
P.knowlesi_PLP5
P.cynomolgi_PLP5
P.vivax_PLP5
P.gallinaceum_PLP5
P.relictum_PLP5
P.ovale_PLP5
P.malariae_PLP5
P.reichenowi_PLP2
P.falciparum_PLP2
P.ovale_PLP2
P.knowlesi_PLP2
P.vivax_PLP2
P.cynomolgi_PLP2
P.malariae_PLP2
P.relictum_PLP2
P.chabaudi_PLP2
P.yoelii_PLP2
P.berghei_PLP2
0.07

Figure 4.1 PLP domain architecture and phylogenetic tree.

(A) Domain architecture of Apicomplexan PLPs. Predicted signal peptides and signal anchors are indicated with orange and green. MACPF domains are shown as red and each APCβ pseudo-repeat is indicated in blue. Figure reproduced from from Kafsack et al. 2010. (B) Sequence based Phylogenetic tree of Plasmodium PLPs. Sequences were aligned using MUSCLE multiple sequence alignment software, and the tree was generated using Figtree (213).

A phylogenetic tree of *Plasmodium* PLPs was produced **(Figure 4.1)**. Analysis of the tree suggests that the duplication of PLPs occurred early in their evolutionary history, before speciation of the main apicomplexan species. PPLPs 1 and 3 are closer together in evolutionary history, occurring from a later duplication event, as are PPLPs 4 and 5. Sequence alignment of *P. vivax* PLP APCβ domains shows that few residues are conserved between the PLPs, mainly consisting of structural cysteines **(**Error! Reference source not found.**)**. A Hidden Markov Model produced by Jackhmmer from alignment of ApiPLP CTDs identified a 90-residue long consensus motif of primarily conserved cysteines, which exists in triplicate in ApiPLPs. Each repeat contains four conserved cysteines, as well as a well-conserved glycine, proline, and a hydrophobic patch before the fourth cysteine residue **(**Error! Reference source not found.**)**.

The expansion of PLPs within *Plasmodium* indicates that gene duplication was required to produce multiple PLPs adapted to different conditions within the lifecycle. As an ancient mechanism for cellular attack, the MACPF domain itself appears to use a fundamentally mechanical mechanism to punch membrane pores, meaning it has little inherent specificity. Therefore, the MACPF/CDC family primarily depends on diversity in their C-terminal membrane binding domain to provide a method of regulation (65). As a result, membrane binding is likely to be key to understand the differences in function between each *Plasmodium* PLP and why multiple PLP homologues are required within the lifecycle. Further, membrane association of MACPFs is an important regulatory bottleneck for pore formation and therefore an attractive target for therapeutic disruption. For these reasons, we have focused primarily on the membrane binding characteristics of PPLPs. We have first focused on solving the structure of PvPLP1 APCβ domain, and then used methods developed in Chapter 3 to extend our work to other *Plasmodium* PLPs.

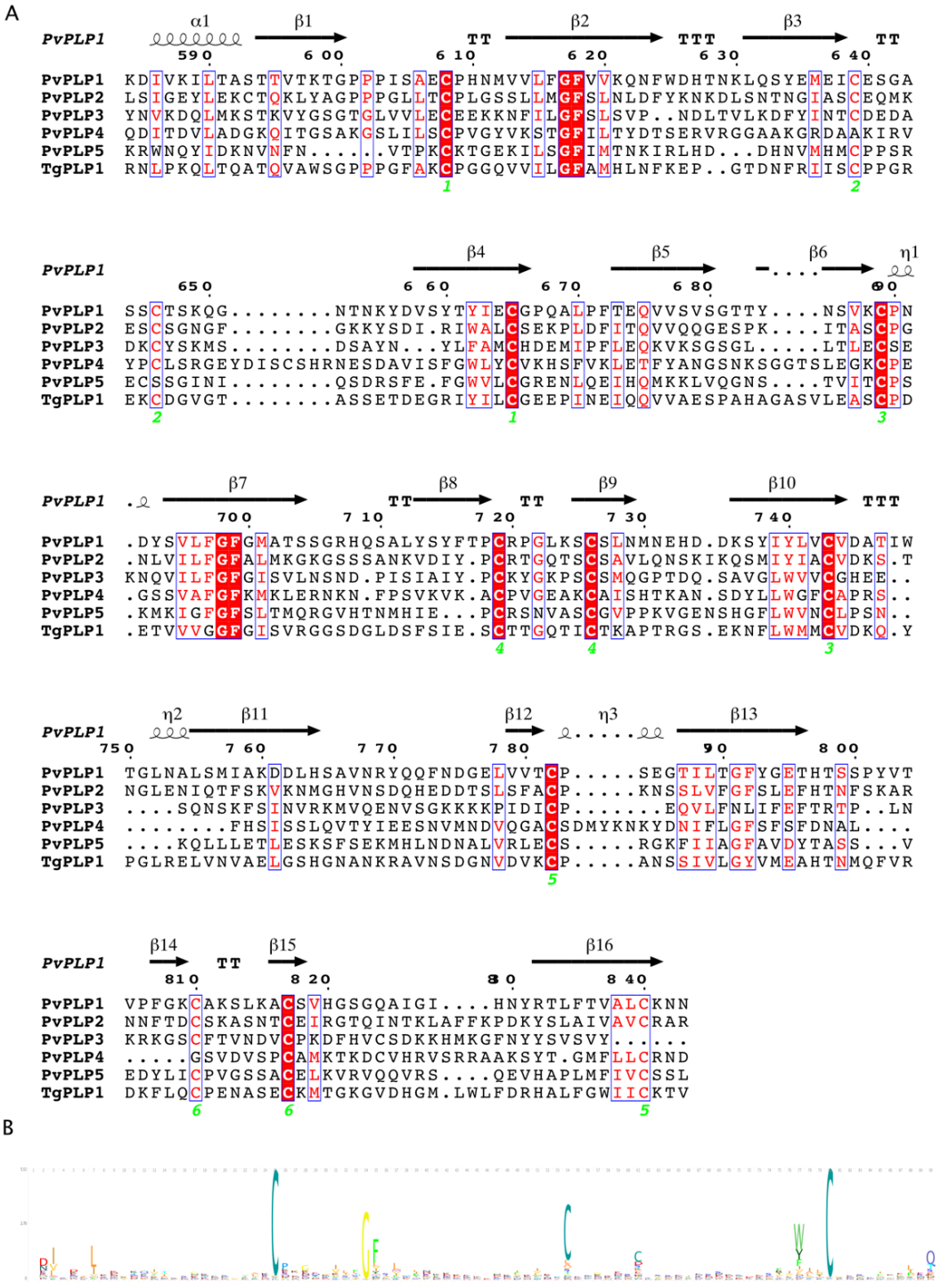


Figure 4.2 PPLP sequence alignment.

(A) Plasmodium vivax APCβ sequence alignment generated using ESPript 3.0. (B) Hidden Markov Model (HMM) showing conserved residues in the PLP APCβ domain based on sequence alignment of PvPLP1 APCβ homologous sequences. The first 90 residues of the HMM are shown here, forming one of three pseudo-repeats in the full-length sequence.

4.2 Results and Discussion

4.2.1 Structural Determination of the PvPLP1 APCβ domain

Expression screens of PvPLPs

Expression screening was carried out for PvPLPs 1, 2, 3, and 5 APCβ domains. Previous work in optimising PPLP expression found that *E. coli* and yeast expression systems do not yield well-folded protein, most likely because PPLPs are secreted proteins which require maturation through the protein secretory pathway and formation of disulphide bonds (214). A protocol using a mammalian secretion expression system was used to successfully solve the TgPLP1 crystal MACPF and APCβ crystal structures (91), and so was replicated here.

Previous expression screening found success in producing *P. Vivax* PLPs 1 and 2 (214). The *Plasmodium* genus has high genomic AT content, up to 80% in coding DNA, making structural studies challenging (215). *P. vivax* causes malaria in humans and has a more balanced AT content than *P. falciparum*, possibly resulting in the better expression observed for this species and was therefore chosen for study here.

A codon-optimised synthetic PvPLP1 gene was used for this study. The PPLP1 gene consists of 8 exons, so was difficult to obtain as a full-length wild type protein. Codon optimisation for expression in a mammalian system probably also led to higher protein yields. Wild-type, non-optimised genes were used for all other PLPs.

PvPLP 1, 2, 3 and 5 APCβ constructs were cloned into pHLsec expression vectors with a C-terminal Histidine tag. Small scale-expression tests were performed using transient expression in the mammalian HEK 293T cell expression system **(Figure 4.3)** (216). All constructs were detected in cell lysate, indicating successful expression in the cells. Only PvPLP1 and PvPLP2 APCβs were detected in cell media, indicating they were the only constructs that were successfully folded and targeted towards the secretion system.

Small scale expression tests were also carried out for a series of PvPLP1 MACPF and N-terminal domain constructs. Alignment of the PvPLP1 sequence with sequences of the solved TgPLP1 MACPF and APCβ structures was used to predict domain boundaries. Several constructs each of the MACPF alone, and in tandem with the NTD, were generated around these domain boundaries aiming to optimise protein folding for successful secretion. Constructs were cloned into the PhLSec vector and expressed transiently in HEK 293T cells. In all cases protein was detected in cell lysate but not secreted, indicating poor folding or degradation within the cell **(Figure 4.4)**. Expression of the MACPF and NTD was not pursued further here.

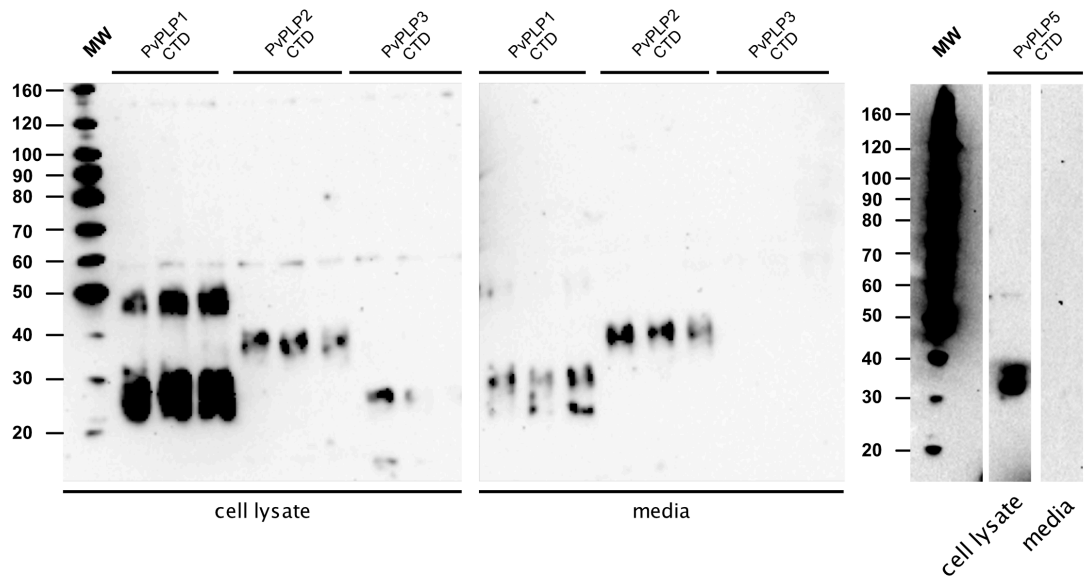


Figure 4.3 PvPLP1 APCβ small scale expression tests.

Western blot of PvPLP APCβ small scale mammalian expression tests. PvPLP1 and PvPLP2 were the only samples detected in secretion media. Each sample is run in triplicate.

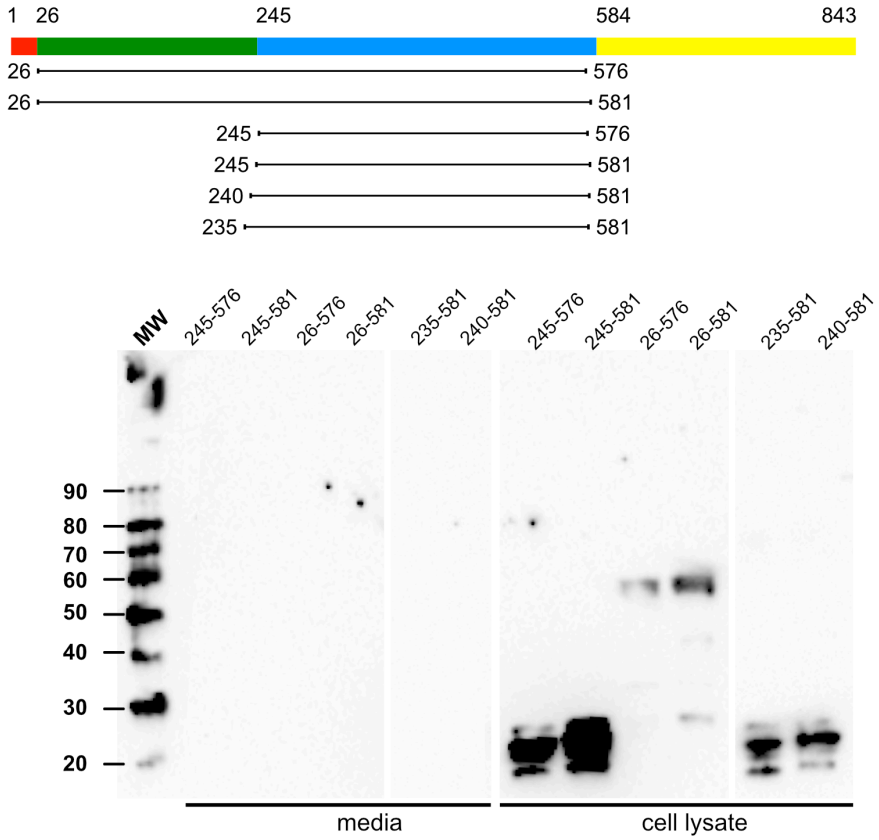


Figure 4.4 Small scale mammalian expression tests of PvPLP1 MACPF and NTD constructs.

(A) Construct boundaries are represented against the full PvPLP1 gene. Red represents the signal peptide, the NTD is green, MACPF is blue and APCβ is yellow. (B) Western blot of expression tests showing no expression detected in the media.

Large scale expression and purification of PvPLP1 APCβ was carried out for biophysical and structural studies. Ni-NTA affinity chromatography was used to purify large scale expression from HEK 293T cells. Final size exclusion chromatography (SEC) purification was performed using a Superdex 75 16/60 column **(Figure 4.5)**. A laddering effect was observed following SEC of PvPLP1, which was also observed in small scale expression tests. Two protein peaks were detected during SEC corresponding to monomer and dimer of PvPLP1, with further higher order oligomers visible in SDS-PAGE gels. The appearance of these oligomers appears to be a concentration dependent effect, with greater order oligomers visible in the higher concentration second peak from size exclusion. Denaturing SDS-PAGE gels did not disrupt the oligomers, indicating high stability.

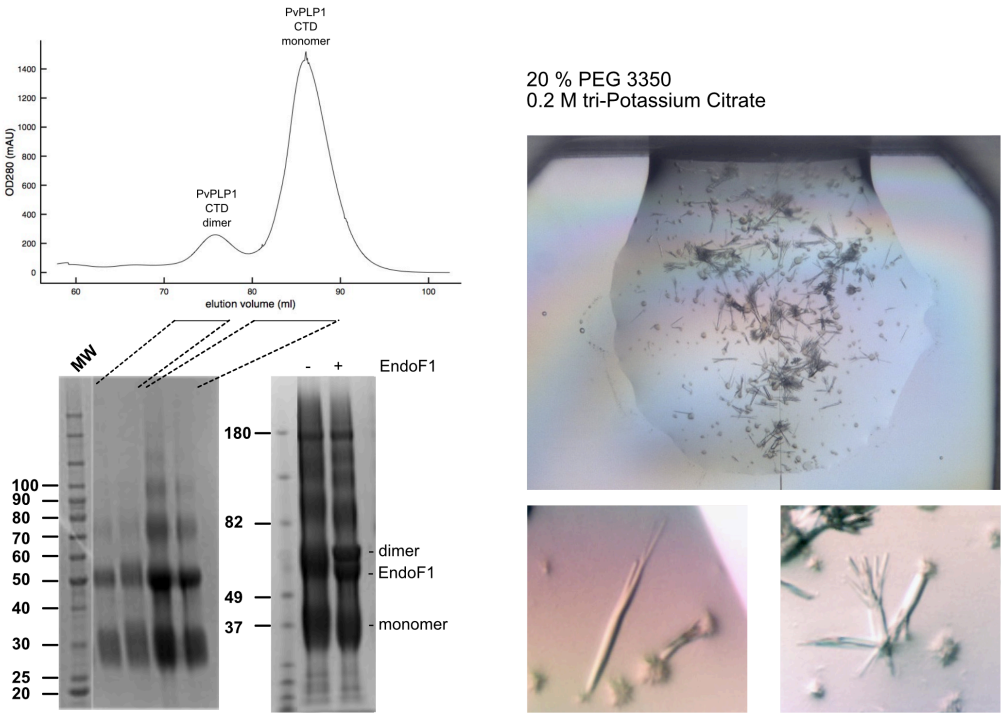


Figure 4.5 Large Scale purification and crystallisation of PvPLP1 APCβ.

(A) Size exclusion chromatography trace of APCβ purification. The peaks correspond to PvPLP1 monomer and dimer, which oligomerises in solution shown in western blots. A de-glycosylation test confirmed PvPLP1 APCβ was not glycosylated. (B) Optimised crystals grown in conditions shown.

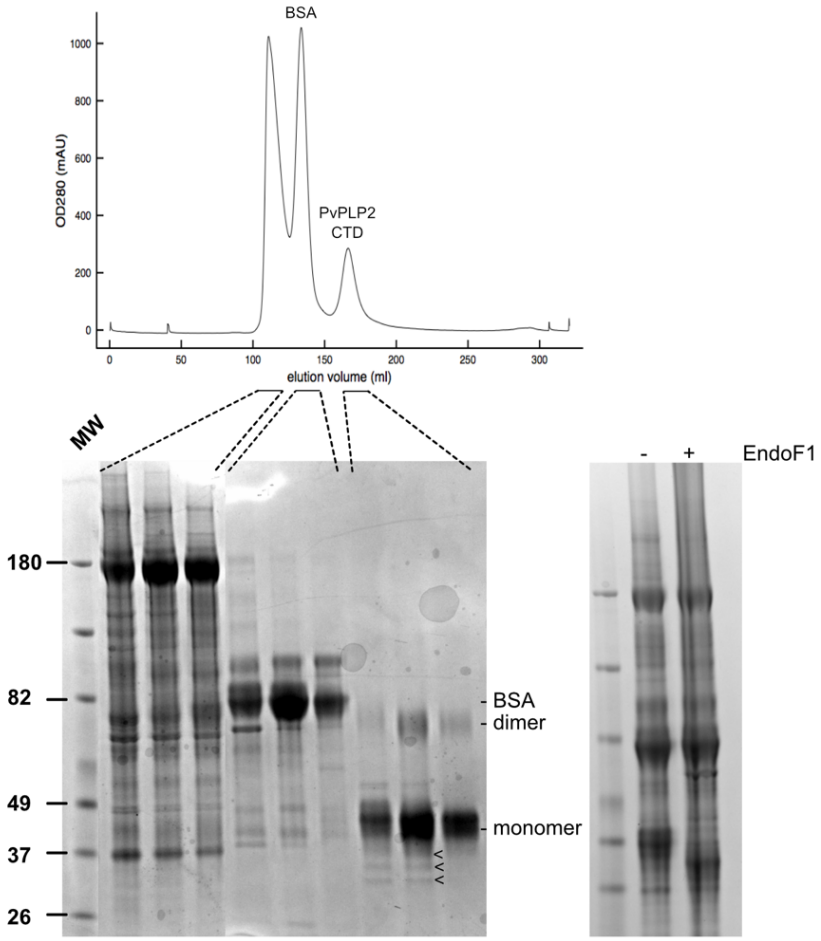


Figure 4.6 Large Scale purification of PvPLP2 APCβ.

Two peaks of monomeric and dimeric BSA, and a smaller peak of PvPLP2 are present in the SEC trace. Possible degradation bands are indicated with arrows (<). PvPLP2 was glycosylated, with deglycosylation using EndoF1 treatment showing a shift in size of the monomeric band.

The same protocol was used to produce PvPLP2 APCβ **(Figure 4.6)**. Two glycosylation sites were predicted in PvPLP2 using NetNGly 1.0 server (217), which were confirmed following overnight deglycosylation treatment with EndoF1. Oligomeric species of PvPLP2 were also observed, however presumably due to low protein concentration only lower order oligomers were seen. Several smaller protein bands appeared in purified PvPLP2 samples, which may be caused by partial proteolysis.

Biophysical analysis of structural stability

The thermal stability of PvPLP1 and PvPLP2 APCβs was assessed using thermofluor assays under a range of pH and ionic conditions. TgPLP1 APCβ (protein provided by Dr Tao Ni) was also assessed for comparison. PvPLP2 and TgPLP1 maintained high stability across a wide range of temperatures, pH and salt conditions **(**Error! Reference source not found. **A)**. Melting temperatures were similar, in

the range of 64-72 °C for PvPLP2 compared to 52-72 °C for TgPLP1. PvPLP2 displayed greater overall stability than TgPLP1, especially under low pH conditions. The denaturation curve produced by PvPLP1 APCβ exhibited a very high initial background signal, making interpretation difficult. This is often observed with membrane proteins, which display hydrophobic residues on their surface, or for poorly folded proteins.

Circular dichroism (CD) was used to confirm the folded state of PvPLP1 APCβ and was compared with CD spectra for TgPLP1 to give an indication of secondary structure similarity. Thermal stability of PvPLP1 APCβ calculated by CD was in a similar range to PvPLP2 and TgPLP1 thermofluor results **(**Error! Reference source not found. **B)**. PvPLP1 and TgPLP1 had very similar CD spectra, indicating that PvPLP1 is well folded and has a similar secondary structure to TgPLP1.

Crystallisation and structural determination

Following large-scale expression and purification, PvPLP1 APCβ was concentrated to 20 mg ml^{-1} for crystallisation screening. Crystals appeared after about 20 days in 20% PEG 3350, 0.2 M Tri-potassium citrate. As initial crystals were small and highly branched, further optimisation screening using a dilution series yielded more pin-like crystals **(Figure 4.5)**. Following several crystallisation trials at differing concentrations, PvPLP2 did not crystallise. In-situ proteolysis of crystallisation drops was also performed to test if trimming of flexible regions of protein may result in crystal formation. This protocol yielded success for crystallisation of the TgPLP1 MACPF domain (91). Crystallisation trials of this this PvPLP2 APCβ construct did not successfully yield crystals, however more recent work in the lab produced PvPLP2 crystals.

Diffraction data of PvPLP1 crystals was collected at I24 at the Diamond Light Source (Didcot, UK). Data were collected to a resolution of 3.2Å in space group $P2_1$, with two molecules per asymmetric unit **(Table 4.1)**. The structure was solved using molecular replacement with the similar TgPLP1 APCβ crystal structure (PDB: 5OUO, (91)), using phenix.phaser (176), followed by phenix.refine (177), and manual reconstruction in Coot (178).

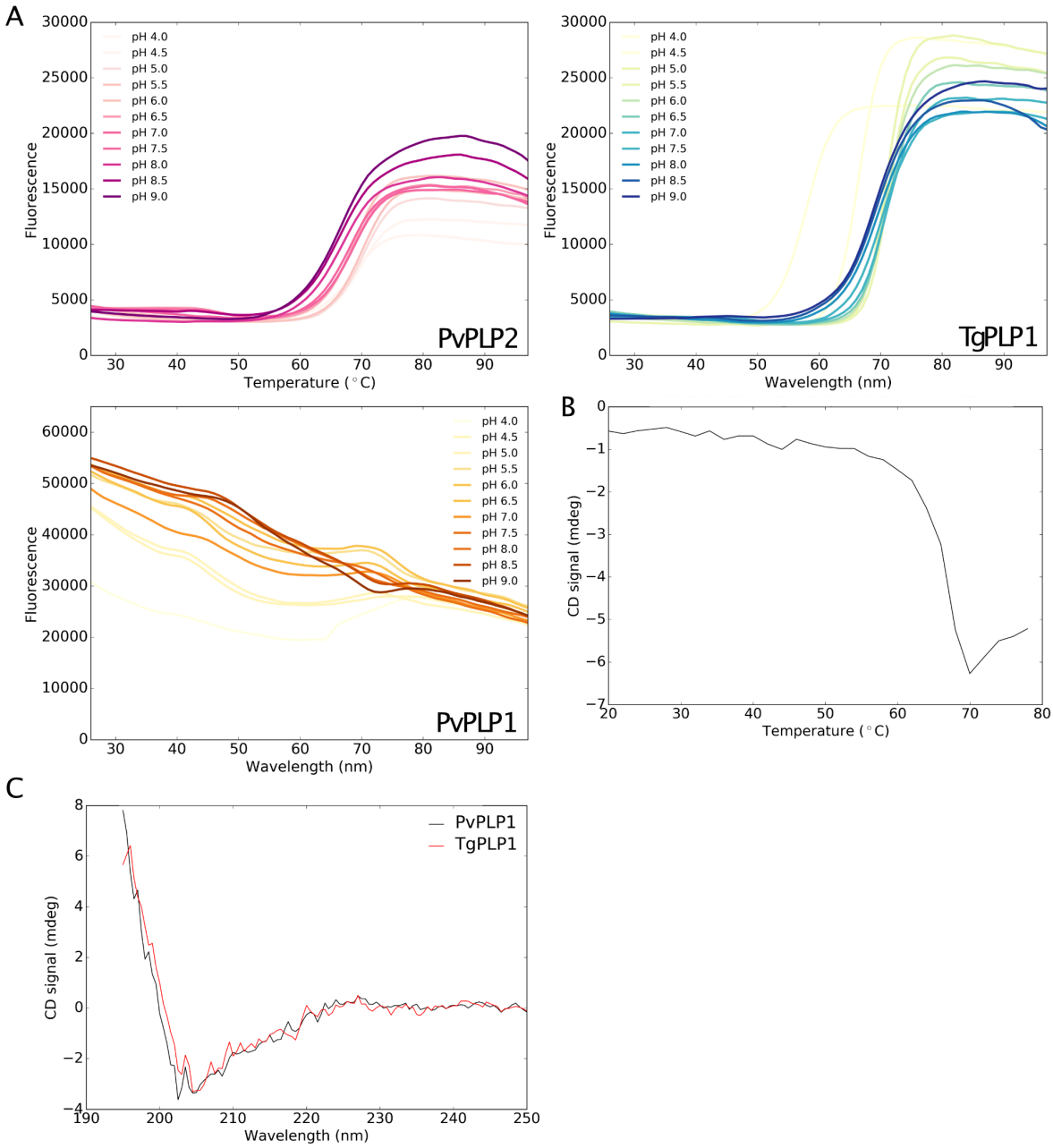


Figure 4.7 Stability of PvPLP1, PvPLP2 and TgPLP1

(A) Melting curves comparing stability of PvPLP1, 2 and TgPLP1 APCβ using thermofluor assays under a pH range from 4.0 – 9.0 in 200 mM salt. (B) Thermostability of PvPLP1 APCβ measured with circular dichroism. (C) Comparison of CD spectra of PvPLP1 and TgPLP1.

Resolution range	63.01 - 3.15 (3.263 - 3.15)
Space group	P 1 21 1
Unit cell	73.74 42.02 87.22 90 104.95 90
Total reflections	122538 (2207)
Unique reflections	9099 (790)
Multiplicity	13.5 (2.8)
Completeness (%)	0.98 (0.85)
Mean I/sigma(I)	10.54 (1.56)
Wilson B-factor	85.91
R-merge	0.1942 (0.5668)
R-meas	0.201 (0.6824)
CC1/2	0.996 (0.697)
CC*	0.999 (0.906)
Reflections used in refinement	9098 (790)
Reflections used for R-free	431 (37)
R-work	0.2740 (0.3730)
R-free	0.2793 (0.3500)
CC(work)	0.887 (0.651)
CC(free)	0.906 (0.327)
Number of non-hydrogen atoms	3697
macromolecules	3697
Protein residues	484
RMS(bonds)	0.015
RMS(angles)	1.29

Table 4.1 PvPLP1 APCβ data collection and refinement statistics.

4.2.2 The PvPLP1 APCβ domain crystal structure

PvPLP1 APCβ structure

The PvPLP1 APCβ crystal structure consists of three repeated β pseudo-domains folded into a prism structure **(Figure 4.8)**. Each pseudo-repeat contains 6 β-strands, with four antiparallel β-strands on an inner face and two on the outer face forming a β-sandwich structure. Each inner face is highly hydrophobic with several aromatic residues facing into the core of the prism to maintain packing and stability. Each sandwich is locked together by two disulphide bonds connecting the inner and outer face of the structures.

The surface of the APCβ is mostly polar and charged, with few surface hydrophobic residues **(Figure 4.8)**. A positively charged groove sits along the base of the protein between pseudo-repeats 1 and 2, with another positively charged patch at the base of repeat 3, which could interact with negatively charged lipids in the bilayer. The protein lacks any significant hydrophobic or aromatic residues on extended loops at its base. This, along with the primarily polar surface, indicates it may have different membrane interactions than those identified in TgPLP1 APCβ (see Chapter 3). The loop analogous to the TgPLP1 tryptophan tipped membrane binding loop contains the motif ^{825}AIGIHN^{830}, with the histidine at its tip **(Figure 4.11)**. In addition, six histidines are found on the surface of this protein, four of which are located at the base of the protein, indicating a possible role of pH in protein function **(Figure 4.8 B)**.

Loops between strands β11-12 and β15-16, connecting the inner and outer face of repeat 3 were not resolved in the crystal structure due to lack of density. This is presumably due to flexibility in these loops, which we assessed further using MD simulations.

The missing loops were modelled into the structure, which was then simulated for 100ns in triplicate to assess overall dynamics of the protein **(Figure 4.9)**. The structure was highly stable, with an average Cα RMSD of 2.5 Å. Secondary structure was well maintained throughout the simulation, and RMSF calculations indicated that the core of the structure, including packing of the three repeats and locked beta-sandwiches, was highly stable. Loops at the base of the protein in the first two repeats were slightly more flexible, while the loops in the third repeat were significantly more mobile. While this could be due to modelling of the loops, the fact that density was missing in this region of the crystal structure suggests that these loops are indeed flexible.

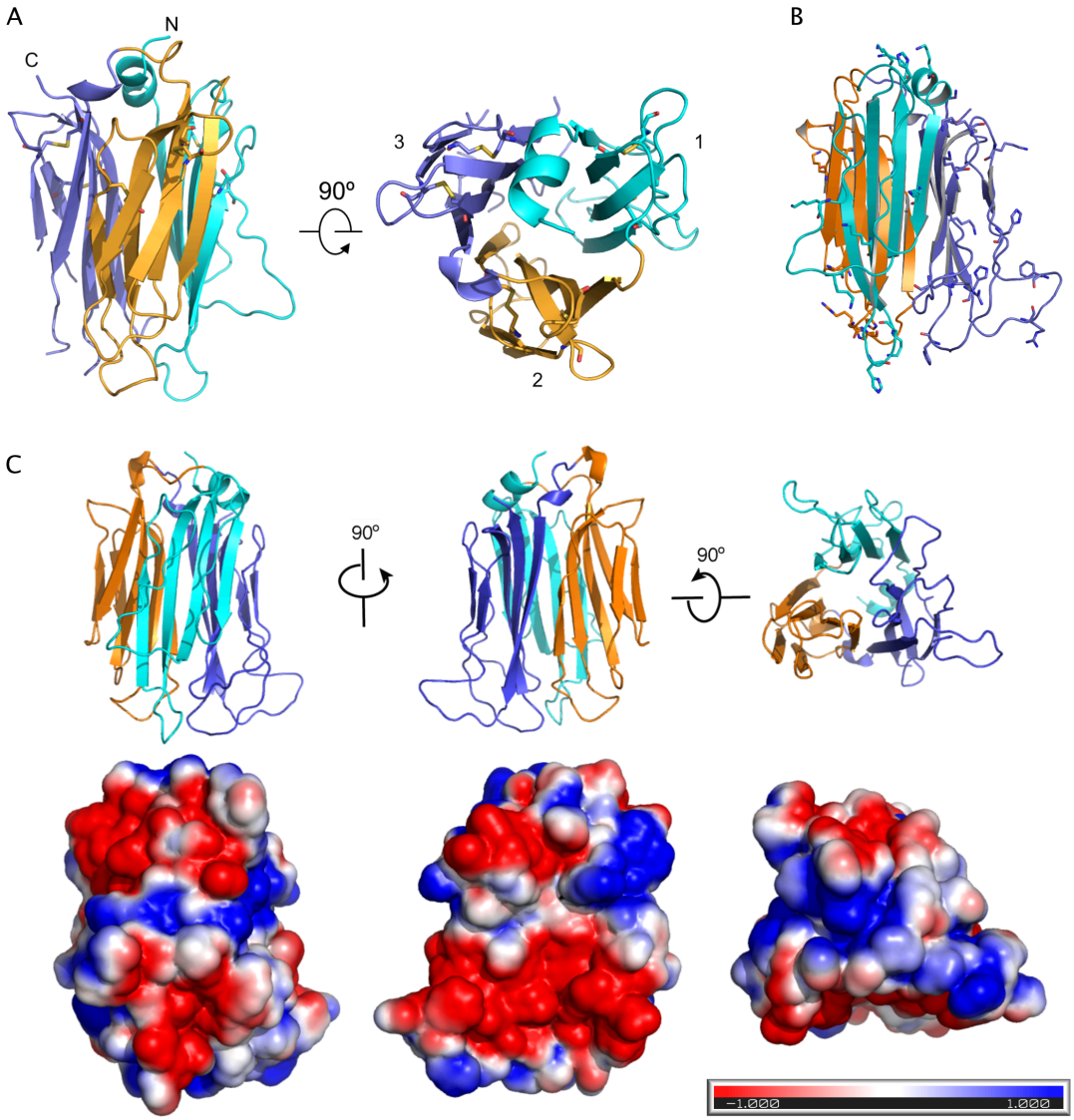


Figure 4.8 PvPLP1 APCβ crystal structure.

(A) Front and top view of the 3.1Å crystal structure of PvPLP1 APCβ. The structure is composed of three repeats of a beta-sandwich folds, coloured in cyan, orange and purple. Density for the turns at the base of repeat 3 was missing. Cysteine residues involved in disulphide bonds are shown as sticks. (B) Cartoon view of PvPLP1 APCβ crystal structure with missing loops modelled in. Positively charged residues are shown in sticks. (C) View of the modelled PvPLP1 APCβ structure showing electrostatic properties of surface in the bottom panel.

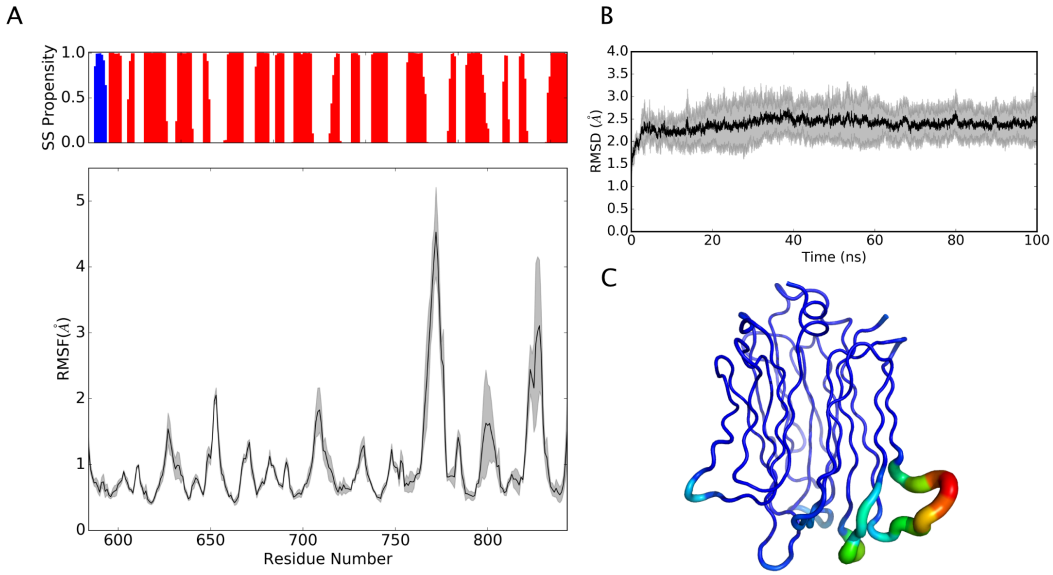


Figure 4.9 Structural stability of PvPLP1 crystal structure in MD simulations.

(A) Alpha helix (blue) and beta-strand (red) secondary structure propensity (top) and Cα RMSF (bottom) and (B) Cα RMSD, averaged from three 100 ns simulations. Errors were calculated as standard deviations. (C) Cartoon putty representation of PvPLP1 crystal structure. Colouring and width of putty represent RMSF calculated for Cα atoms of each residue averaged across the three simulations.

Bioinformatic analysis of TgPLP1 and PLPs

Sequence alignment of *P. vivax* PLP APCβs showed good conservation of cysteines and hydrophobic core residues, indicating conservation of core domain architecture (Error! Reference source not found.). The remaining sequence has low conservation. Conservation of apicomplexan APCβ domains mapped onto the PvPLP1 structure highlights that conserved residues are clustered around the core of the domain and related to structural integrity of the β-sandwich repeat structure **(Figure 4.10 A)**. Surface residues on the other hand were poorly conserved, especially around loops at the base of the protein, including the loop homologous to the TgPLP1 membrane anchoring loop.

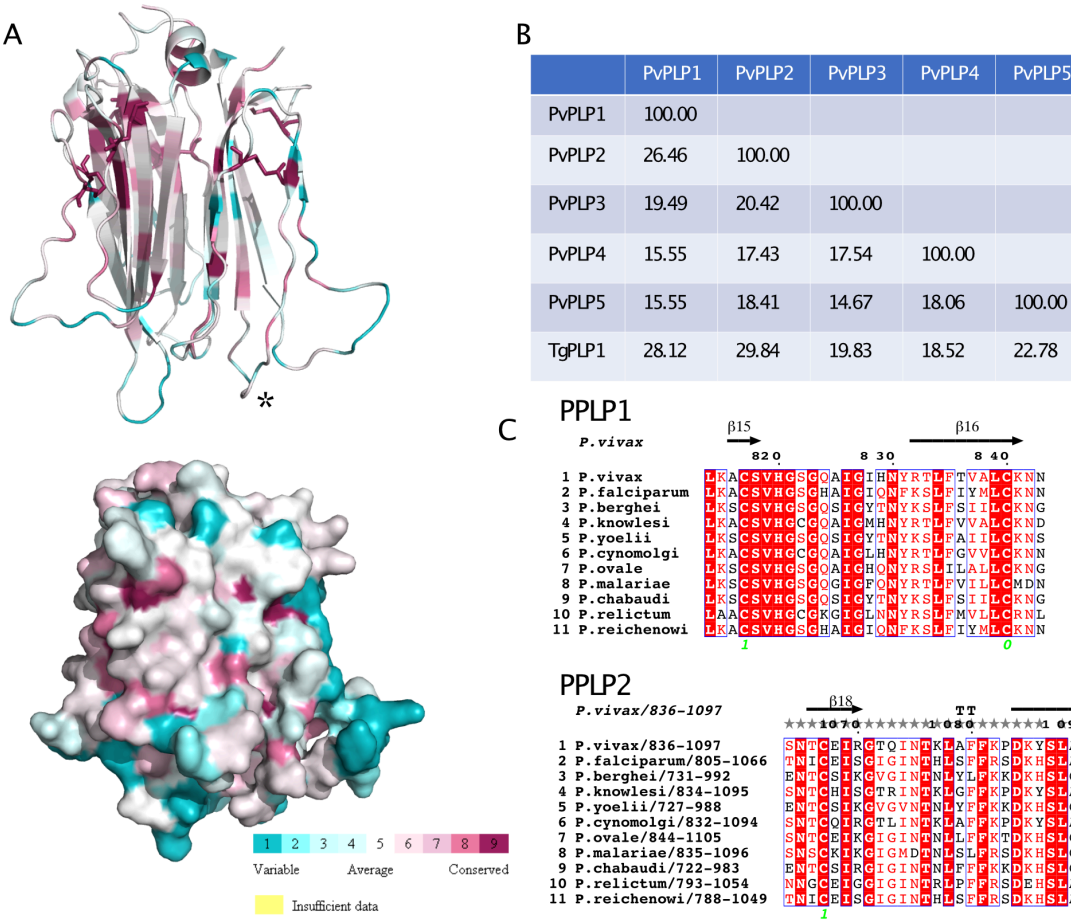
A percentage identity matrix (PIM) was calculated for the PvPLP and TgPLP1 APCβs, showing sequence identity of between 15-28% **(Figure 4.10 B)**. In contrast to the low levels of homology between PLP paralogs, sequence alignment of orthologous PLP1 and PLP2 APCβs across multiple Plasmodium species showed much higher conservation of residues within the membrane binding loop **(Figure 4.10 C)**. This indicates a functional importance of these residues, suggesting that each PLP APCβ functions in a different, but conserved manner. Sequence variation across functionally important

regions of these domains highlights the need to assess membrane binding characteristics of each PLP individually.

Homology models of PvPLP2, 4, and 5

High conservation of cysteine residues and hydrophobic core residues meant we were able to use the PvPLP1 and TgPLP1 crystal structures as templates to produce homology models of PvPLP2, 4 and 5 APCβ domains **(Figure 4.11)**. Calculation of sequence identity of each beta-sandwich repeat separately showed that the third repeat is the most variable. Due to poor conservation of the final cysteine and shorter sequence of this repeat **(Figure 4.11 C)**, we were not able to confidently generate a model of PvPLP3 APCβ.

While all PLPs have a conserved architecture, loops at the base of the protein showed some variation in length **(Figure 4.11 B)**. PvPLP2 and TgPLP1 are most similar in structure, both containing the same extended aromatic residue tipped loop at the base of the third pseudo-repeat. In TgPLP1 this residue is a tryptophan, while the homologous loop in PvPLP2 is tipped with two phenylalanines, one of which is conserved across all PLP2 species analysed. The equivalent loops in other PLPs are shorter, and do not contain aromatic residues, indicating a different molecular mechanism of membrane binding **(Figure 4.11 B)**. While at least one aromatic residue was identified at the base of each PvPLP, they were not located on a protruding loop and therefore did not appear to be optimally positioned for membrane anchoring in PvPLP1, 4 and 5. Positively charged residues dominate the base of PvPLP4 and 5, while PvPLP1 has fewer positively charged residues on its surface compared to other PPLPs.



	PvPLP1	PvPLP2	PvPLP3	PvPLP4	PvPLP5
PvPLP1	100.00				
PvPLP2	26.46	100.00			
PvPLP3	19.49	20.42	100.00		
PvPLP4	15.55	17.43	17.54	100.00	
PvPLP5	15.55	18.41	14.67	18.06	100.00
TgPLP1	28.12	29.84	19.83	18.52	22.78

Figure 4.10 Sequence conservation across PPLPs.

(A) Conservation of PLP residues mapped onto PvPLP1 APCβ crystal structure with a cartoon and surface representation. Homologous protein to PvPLP1 were identified and a multiple sequence alignment produced using HMMER. Conservation scores were mapped onto the structure using the Consurf server (218). The loop homologous to the tryptophan tipped membrane anchoring loop of TgPLP1 is marked with an asterisk (*). (B) Sequence identity matrix of PvPLPs and TgPLP1. (C) Multiple sequence alignment of Plasmodium PLP1 (top) and PLP2 (bottom) from key Plasmodium species membrane binding loops shows high sequence conservation within each PLP across the species. Generated using ESPript 3.0 (219).

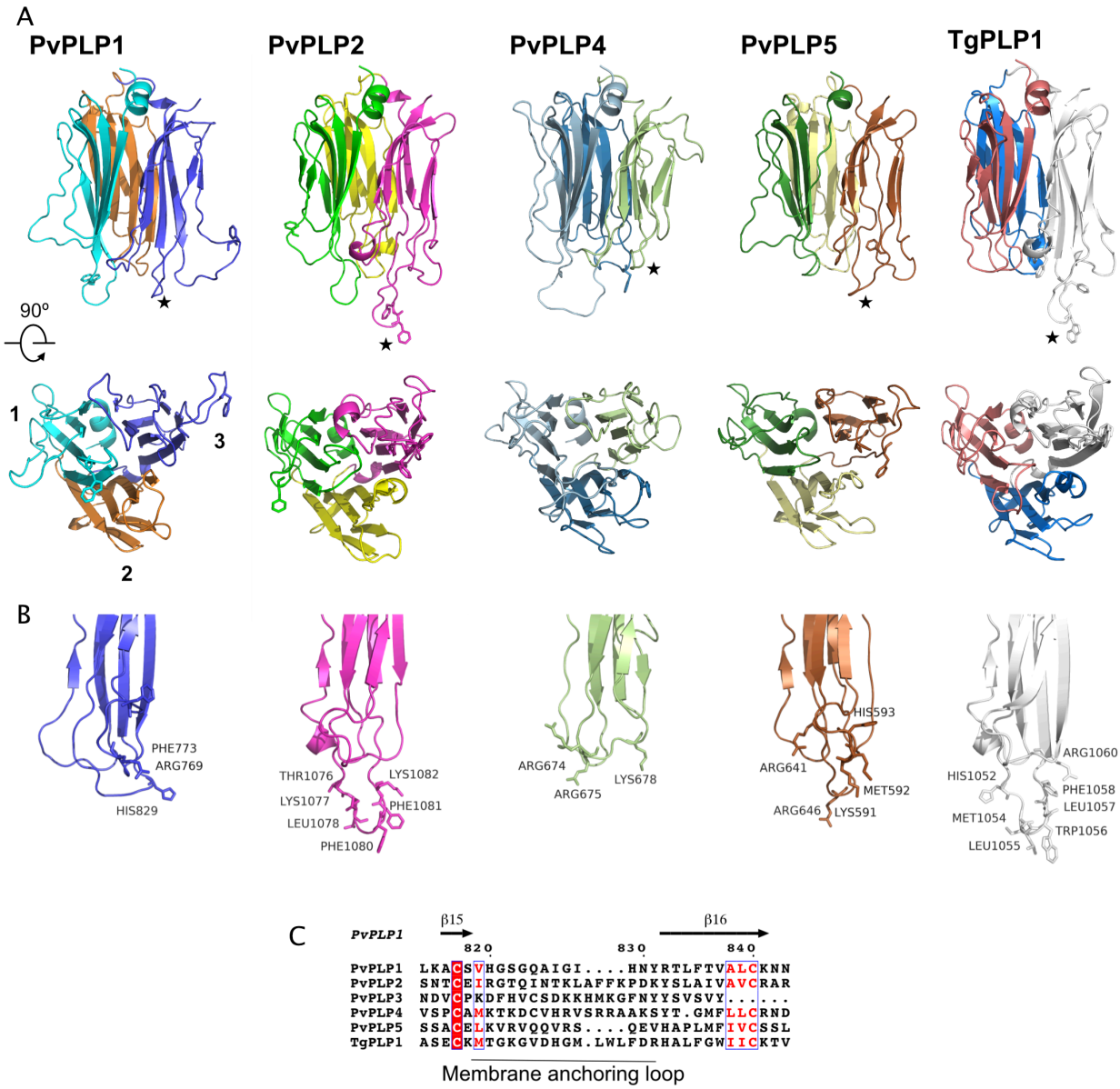


Figure 4.11 Structural comparison of PvPLP1, 2, 4, 5 and TgPLP1 APCβ structures.

Homology models of PvPLP2, 4 and 5 are shown in cartoon representation alongside PvPLP1 and TgPLP1 crystal structures. Each beta-sandwich repeat is coloured separately, with the loop homologous to TgPLP1 membrane anchoring loop labelled with a star (★). Each pseudo-repeat is labelled in PvPLP1, with all other PLPs following the same pattern. (B) Close up view of the base of pseudo-repeat 3, highlighting residues predicted to have a role in membrane binding. (C) Sequence alignment of predicted membrane anchoring loops for each PLP show poor sequence conservation in this loop.

4.2.3 Membrane binding characteristics of PvPLP1 and PvPLP2 APCβ domains

Membrane binding of PvPLP1 and PvPLP2 APCβ was then tested, firstly using lipid dot blot experiments with a range of eukaryotic membrane lipids. PvPLP2 bound to anionic lipids PS, Cardiolipin (CL) and PIPs **(Figure 4.12)**, with particular affinity for PI(4)P. PvPLP1 on the other hand

did not bind to any lipids tested. Liposome sedimentation assays with PvPLP1 and PvPLP2 APCβ were performed by Dr Xiulian Yu to test binding to liposomes composed of 40% PC, 30% PE and 30% of varied lipids. In agreement with lipid dot blots, PvPLP1 did not bind to any liposomes. PvPLP2 showed no binding to cholesterol, sphingomyelin, or brain or liver lipid extracts. Intermediate binding was observed for PS and L-α-lysophosphatidylserine liposomes, while all protein remained in the pellet in CL and L-α-phosphatidic acid (LPA) liposomes, indicating strong binding to negatively charged lipids **(Figure 4.12 B-C)**.

These results suggest PvPLP2 may prefer binding to the inner leaflet of the bilayer which contains more anionic lipids such as PS and PIPs. Unexpectedly, PvPLP1 had no affinity for any lipids tested. This agrees with the observations that PvPLP1 APCβ contains comparatively fewer basic residues on its surface and lacks a hydrophobic loop homologous to the TgPLP1 tryptophan tipped loop.

A series of membrane association simulations were then performed with PvPLP1 and PvPLP2 APCβ domains to assess membrane binding in further detail. Simulations were first performed by allowing the protein to freely associate with a bilayer containing PC, PE and PS in the composition shown in **Table 4.2**, representing a simple model of the eukaryotic membrane inner leaflet. Simulations were run for 2 μs each, and 20 repeats were performed for each system. The protein was sequentially rotated around the x, y and z axis in each repeat, to reduce bias.

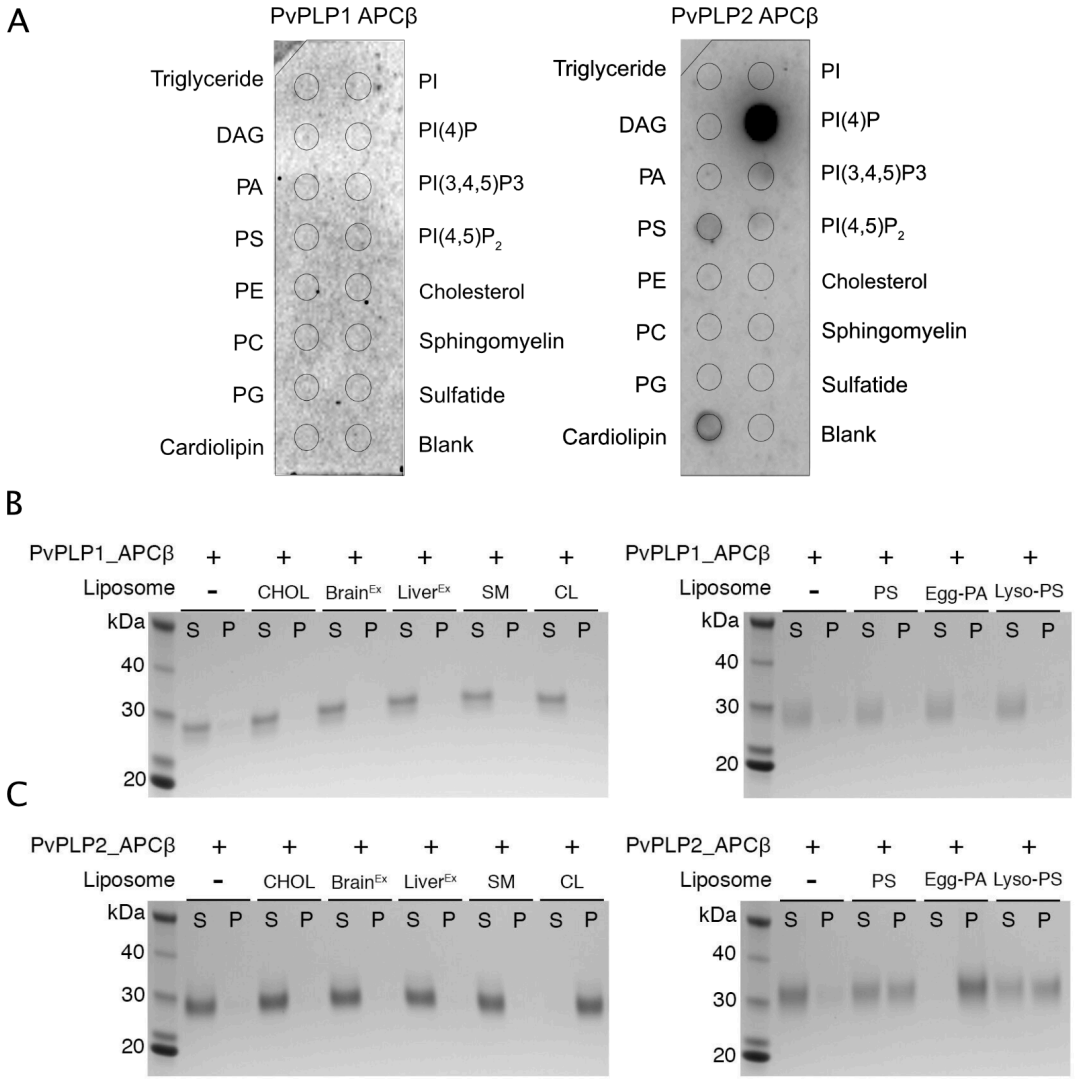


Figure 4.12 Lipid binding of PvPLP1 and PvPLP2 APCβs

(A) *Lipid dot-blots of PvPLP1 and PvPLP2 APCβs on nitrocellulose membrane dotted with common mammalian membrane lipid species.* ***(B-C)*** *Liposome sedimentation assays of APCβ domains of PvPLP1* ***(B)*** *and PvPLP2* ***(C)****. Liposomes are composed of 40% PC, 30% PE and 30% of varied lipids. CHOL, Cholesterol; Brain^{Ex}, Brain Total lipid Extract; Liver^{Ex}, Liver Total Lipid Extract; SM, Sphingomyelin; CL, cardiolipin; PS, phosphatidylserine; Egg-PA, L-α-phosphatidic acid; Lyso-PS, L-α-lysophosphatidylserine. Liposome assay results kindly provided by Dr Yu.*

Protein	Lipids	Composition (%)	pH	Simulations/length (ns)	Type (CG/AT)
PvPLP1	PC:PE:PS	45:45:10		20/2000	Membrane association (CG)
PvPLP1	PC:PE:PS:PI(3,4,5)P_3	45:42:10:3		20/2000	Membrane association (CG)
PvPLP1	PC:PE:PS	45:45:10	5.5	5/2000	Unbiased (CG)
PvPLP2	PC:PE:PS	45:45:10		20/2000	Membrane association (CG)
PvPLP2	PC:PE:LPA	45:45:10		20/2000	Membrane association (CG)
PvPLP2	PC:PE:PS:PI(4)P	45:42:10:3		3/2000	Unbiased (CG)
PvPLP2	PC:PE:PS	45:45:10		3/100	Unbiased, upright orientation (AT)
PvPLP2	PC:PE:PS	45:45:10		3/100	Unbiased, side orientation (AT)
PvPLP1	PC:PE:PS	45:45:10		25/1000	Umbrella sampling (CG)
PvPLP1	PC:PE:PS	45:45:10	5.5	25/1000	Umbrella sampling (CG)
PvPLP2	PC	100		25/1000	Umbrella sampling (CG)
PvPLP2	PC:PE	80:20		25/1000	Umbrella sampling (CG)
PvPLP2	PC:PS	80:20		25/1000	Umbrella sampling (CG)
PvPLP2	PC:PE:PS	45:45:10		25/1000	Umbrella sampling (CG)
PvPLP2	PC:PE:LPA	45:45:10		25/1000	Umbrella sampling (CG)
PvPLP2	PC:PE:PS:PI(4)P	45:42:10:3		25/1000	Umbrella sampling (CG)
PvPLP2 mutants	PC:PE:PS	45:45:10		25/1000	Umbrella sampling (CG)

Table 4.2 Summary of Simulations

Summary of simulations run in this chapter, including type of simulations and lipid compositions. See ***Table 4.3*** *for details on PvPLP2 mutations.*

PvPLP1 membrane binding

In agreement with lipid binding experiments, PvPLP1 did not associate strongly with the membrane, binding in only 2 out of 20 repeats **(Figure 4.13 A & B)**. Both binding events were followed by subsequent dissociation from the bilayer. In one case PvPLP1 bound to the membrane in an upside-down orientation, with the N- and C- termini sitting on the membrane surface **(Figure 4.13 C i)**. Similar to the results found in TgPLP1 membrane association simulations, these simulations were performed with the domain in isolation and this orientation was deemed non-physiological in the full-length protein containing the MACPF domain. In the second binding event, PvPLP1 was bound to the membrane in a more upright orientation **(Figure 4.13 C ii)**, with residue F^{773} on the first loop of pseudo-repeat 3 (β11-12) associating with the bilayer.

As TgPLP1 and PvPLP2 both have increased affinity for PIP molecules, further membrane association simulations were performed incorporating $PI(3,4,5)P_3$ into the bilayer to test the possible impact of increased electrostatic interactions with the membrane **(Figure 4.14)**. Membrane association was increased to around 40% of simulation time, while a more complex energy landscape revealed multiple binding orientations. Protein-lipid contact analysis showed most binding events involved interactions between N- and C- termini, and visual inspection confirms that these involve the protein oriented in mostly 'upside-down' orientations with the top of the domain in contact with the bilayer. As with binding observed in the PC:PE:PS bilayer, these are unlikely to be physiologically relevant due to steric clashes with the MACPF domain in full-length PvPLP1. An upright orientation was adopted in two simulations, in which the first loop of the third pseudo-repeat anchors the protein to the membrane, as seen in orientation (ii) on the PC:PE:PS bilayer, however addition of PIP_3 in these simulations appears to stabilise this orientation by binding to R^{769} in the same loop.

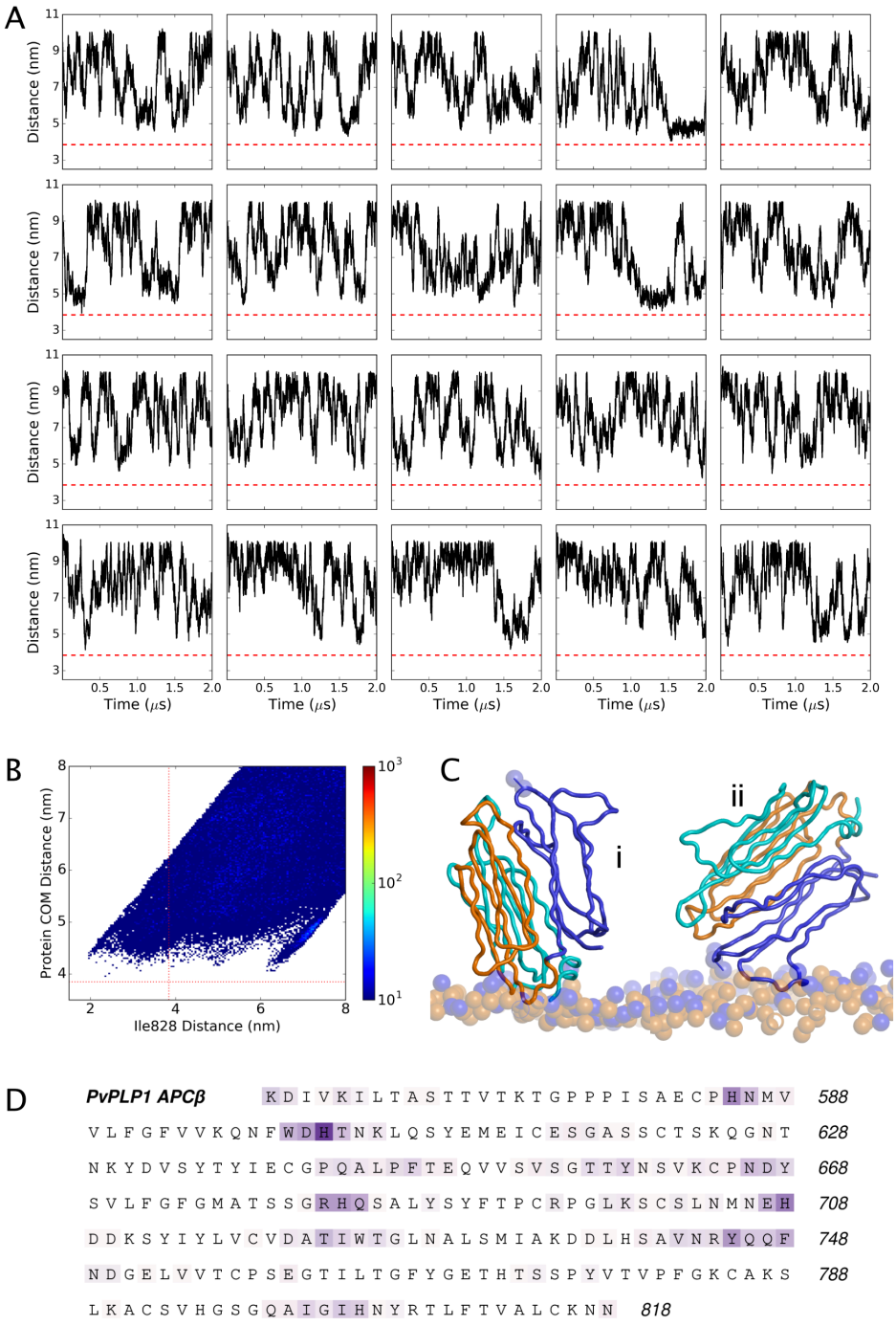


Figure 4.13 Membrane association simulations of PvPLP1 APCβ on PC:PE:PS bilayer.

(A) Grid of protein-membrane COM distance plots for all simulation repeats. Red dotted lines represent location of lipid head group phosphates. (B) Two-dimensional histogram using protein-membrane distances to identify binding frequency and orientations. Very little binding was observed. (C) Cartoon representation of the two binding poses identified for PvPLP1, with each pseudo-repeat coloured as in Figure 11, with Ile828 shown with sticks and spheres. Lipid head groups are shown as beads with phosphates coloured orange. (D) Normalised averaged lipid contacts plotted onto PvPLP1 sequence.

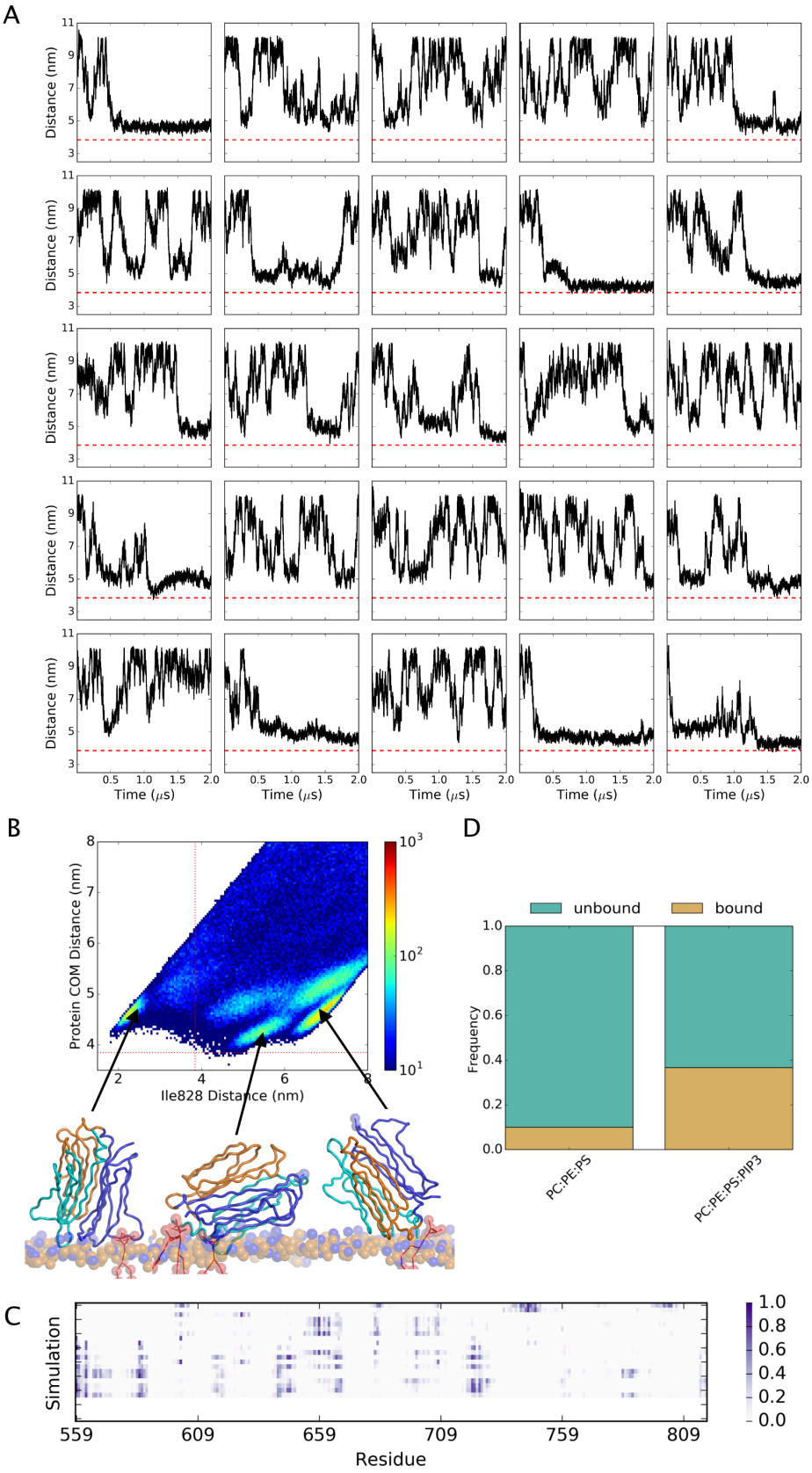

A
Distance (nm)
Time (μs)
B
Protein COM Distance (nm)
Ile828 Distance (nm)
D
unbound
bound
Frequency
PC:PE:PS
PC:PE:PS:PIP3
C
Simulation
Residue
559
609
659
709
759
809

Figure 4.14 Membrane association simulations of PvPLP1 APCβ on PC:PE:PS:PI(3,4,5)P_3 bilayer.

(A) Distance between COM of protein and membrane over time show progress of binding in each simulation. Phosphate bead of lipid headgroups is represented by the dotted red line. (B) Two-dimensional histogram plot of centre-of-mass distance (COM) distance between protein and membrane against distance between COM of membrane and Ile828. Snapshots below show cartoon representations of significant binding orientations. Each pseudo-repeat is coloured separately. Lipid headgroups are shown with spheres, and PI(3,4,5)P_3 is shown in red spheres with sticks. (C) Protein residue-lipid contact plot showing consistency of membrane binding patterns across all simulations. (D) Normalised frequency of binding between PC:PE:PS and PC:PE:PS:PI(3,4,5)P_3 bilayer.

Free Energy of Binding of PvPLP1 to membranes

Umbrella sampling was performed to calculate free energy of binding of the upright orientation of PvPLP1 on the PC:PE:PS membrane. The starting binding orientation for PMF calculations was taken from the final frame of the bound membrane association simulation. The membrane normal was used as the reaction coordinate along which to pull the protein. PvPLP1 was found to have no affinity of binding in PMF calculations, indicating unfavourable energetics between protein and membrane **(Figure 4.15)**. A small energy well is observed in the PMF profile at a distance of 5.0nm between the centre of mass of the protein and bilayer, which corresponds to orientation (ii) identified in membrane association simulations. The energy of association at this orientation is however still unfavourable, indicating this is not likely a stable interaction.

Overall, while some binding events occurred between PvPLP1 and the membrane, which were increased by presence of PI(3,4,5)P_3, most binding orientations were considered non-physiological. Additionally, free energy of binding calculations showed unfavourable energetics with the bilayer.

The effect of pH on PvPLP1 membrane binding

Analysis of average protein-lipid contacts during membrane association indicated that residues with the greatest lipid interaction were clustered around surface histidines **(Figure 4.13 D)**. The PVPLP1 APCβ contains four surface exposed histidine residues at its base, indicating a possible role of pH. Previous studies have shown that PvPLP1 mediated membrane disruption is triggered by acidification (111). Several MACPF/CDCs are known to be activated by pH, which can regulate activity at multiple stages of pore formation. We investigated the possible role of acidification in promoting membrane binding by calculating the free energy of binding of PvPLP1 APCβ at pH5.5 in a one-dimensional PMF. pKa calculations of PvPLP1 APCβ at pH5.5 were produced, and residue charges were set to reflect these pH conditions. PMF calculations were then performed as previously described. The deepest energy well in the PMF profile gave a free energy of binding of -2 kJ/mol and corresponded to the same F^{773} anchored orientation as was identified in membrane associations but anchored further into the

bilayer **(Figure 4.15)**. Additional CG simulations of PvPLP1 on PC:PE:PS bilayers at pH5.5 were produced, with 5 replicates of 2 μs each **(Figure 4.16)**. PvPLP1 dissociated almost immediately in all but one of the simulations, while in one repeat it remained bound for around 1 μs before dissociating. These results suggest that low pH makes no significant difference in increasing membrane binding of PvPLP1 to PC:PE:PS bilayers.

Small scale expression tests were also carried out for a series of PvPLP1 MACPF and N-terminal domain constructs. Alignment of the PvPLP1 sequence with sequences of the solved TgPLP1 MACPF and APCβ structures was used to predict domain boundaries. Several constructs each of the MACPF alone, and in tandem with the NTD, were generated around these domain boundaries aiming to optimise protein folding for successful secretion. Constructs were cloned into the PhLSec vector and expressed transiently in HEK 293T cells. In all cases protein was detected in cell lysate but not secreted, indicating poor folding, degradation or toxicity of the protein within the cell **(Figure 4.4)**. Expression of the MACPF and NTD was not pursued further here.

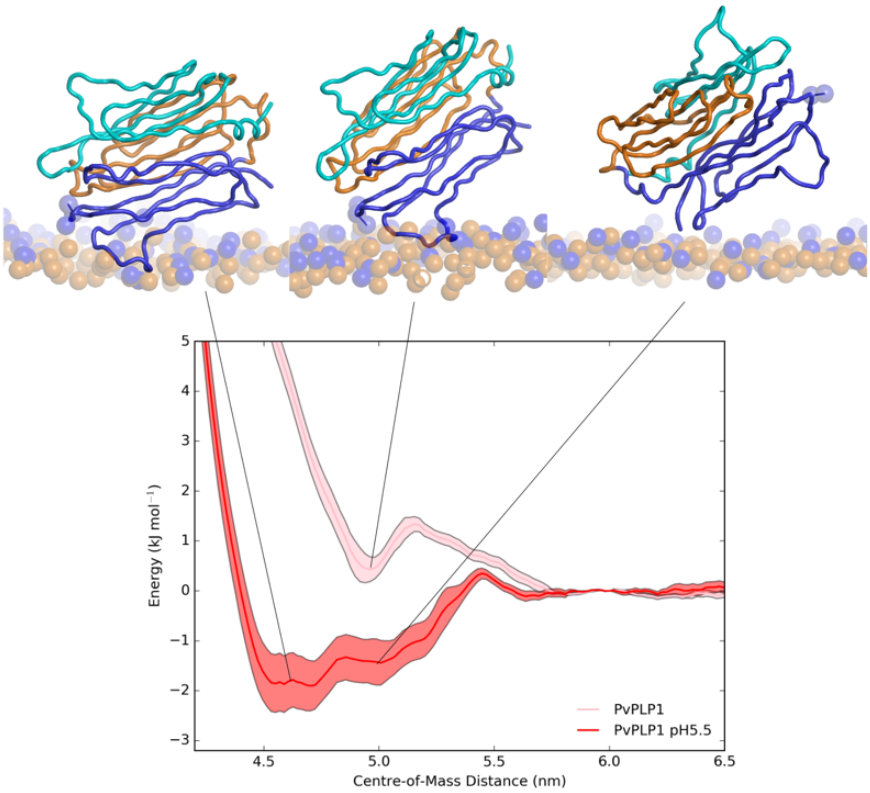


Figure 4.15 Free energy of binding of PvPLP1 APCβ at pH 7 and pH 5.5

PMF profile of PvPLP1 APCβ at neutral and pH5.5 pulled along a reaction coordinate between centre of mass of protein and membrane, perpendicular to the membrane. The membrane composition was PC:PE:PS in both simulations. Binding orientation at energy minima are shown with cartoon representations. Lipid head groups are shown by spheres and PvPLP1 Ile828 is shown with spheres and sticks.

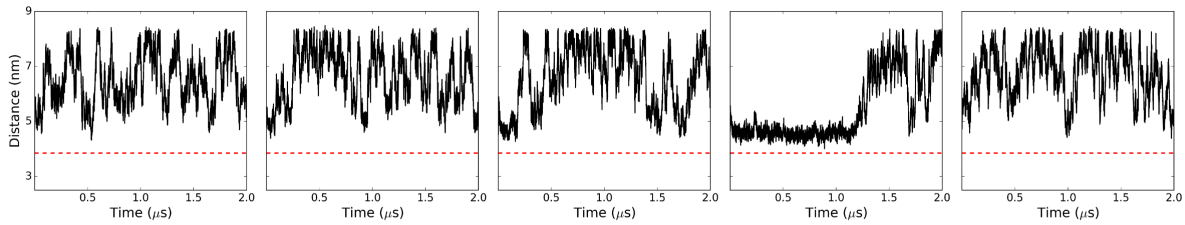


Figure 4.16 Simulations of PvPLP1 APCβ protonated to pH5.5 on PC:PE:PS bilayers.

5 repeats were produced where protein was started in membrane binding orientation of PvPLP2, and freely allowed to settle on the membrane during the 2 μs simulation.

PPLP1 mechanisms of membrane binding

The lack of membrane binding observed here has several possible explanations. It is possible that PvPLP1 binds to specific lipids or proteins in the membrane that were not tested here. Several members of both the MACPF and CDC families require further protein partners to bind the membrane. A subclass of CDCs requires human CD59 to bind to membranes (55), while pleurotolysin appears to be unique in consisting of two components, a MACPF component which first requires binding of a second, membrane binding component, to form pores (62). Within the MACPF family, the C9 component of MAC requires the prior formation of the C5b8 complex on the membrane before polymerisation on the membrane and form a complete pore (220).

It is possible that PPLP1 requires a separate membrane binding partner, either originating from the host membrane, or secreted by Plasmodium alongside PPLP1. If this were the case however, it is interesting to note that the C-terminal domain has been retained by PPLP1 and is well conserved across species, suggesting a further functional role. In this case, a secondary protein binding partner which would associate with the APCβ domain directly is more likely, and a proteomics-based approach could be used to screen for interacting proteins within the PVM and PM.

The requirement of a specific lipid ligand is also a possibility, although most common eukaryotic lipids have been tested here. While we tested the influence of pH here, these investigations could have been taken further to test a greater range of pH. It has been reported that PfPLP1 acts in a calcium dependent manner during merozoite egress from red blood cells (114), however subsequent work suggests PPLP1 is not expressed during blood stage infection, putting this into question.

As experiments have shown that disruption of PPLP1 significantly reduces, or entirely abolishes, the infectivity of Plasmodium (110-113), further work to establish the mechanism of membrane binding and pore formation of PPLP1 may lead to a better understanding of the Plasmodium infectious cycle, and elucidate potential therapeutic targets.

4.2.4 PvPLP2 membrane binding

PvPLP2 binds to the membrane in two orientations

Based on lipid dot blot experiments, and the amphipathic tryptophan tipped loop at the base of PvPLP2, we expected the APCβ domain to have membrane binding characteristics similar to those of TgPLP1. PvPLP2 bound to PC:PE:PS membranes in 55% of membrane association simulations **(Figure 4.17)**. 2-dimensional histograms of protein-membrane distances show that two distinct orientations were adopted on the membrane. This was confirmed by assessing protein-lipid contacts, which identified two separate binding patterns, and showed that these patterns were consistent in their membrane contacts across multiple simulations **(Figure 4.17)**. The first orientation adopted an upright position, with the phenylalanine tipped loop sitting at the interface between lipid headgroups and tails and was found in

73% of simulations **(Figure 4.17 B i)**. The secondary orientation, in which the protein was oriented sideways with its axis lying along the membrane, was adopted 27% of the time **(Figure 4.17 B ii)**. PMF calculations of both orientations on a PC:PE:PS bilayer gave a free energy of binding of -15 and -14 kJ mol^{-1} for the primary and secondary orientations respectively, indicating that on the PC:PE:PS membranes these two orientations are both favourable and thus likely of equal significance **(Figure 4.19 A)**.

As liposome binding assays showed strong binding of PvPLP2 to liposomes containing LPA, we performed a further set of membrane association simulations containing LPA **(Table 4.2)**. Binding occurred in 55% of these simulations, an increase compared to PC:PE:PS simulations **(Figure 4.17)**. In contrast to simulations with PC:PE:PS, these simulations identified only a single binding orientation on the membrane that matched the upright orientation on PC:PE:PS.

These results confirmed that PvPLP2 APCβ has an affinity for membranes in simulations and suggested the presence of two binding modes. While the orientations have similar affinities for PC:PE:PS membranes, the upright binding mode appears to be favourable in the presence of LPA species in the membrane. This may be due to an increased affinity of the protein to LPA and other charged lipids in the membrane which may drive binding to this orientation.

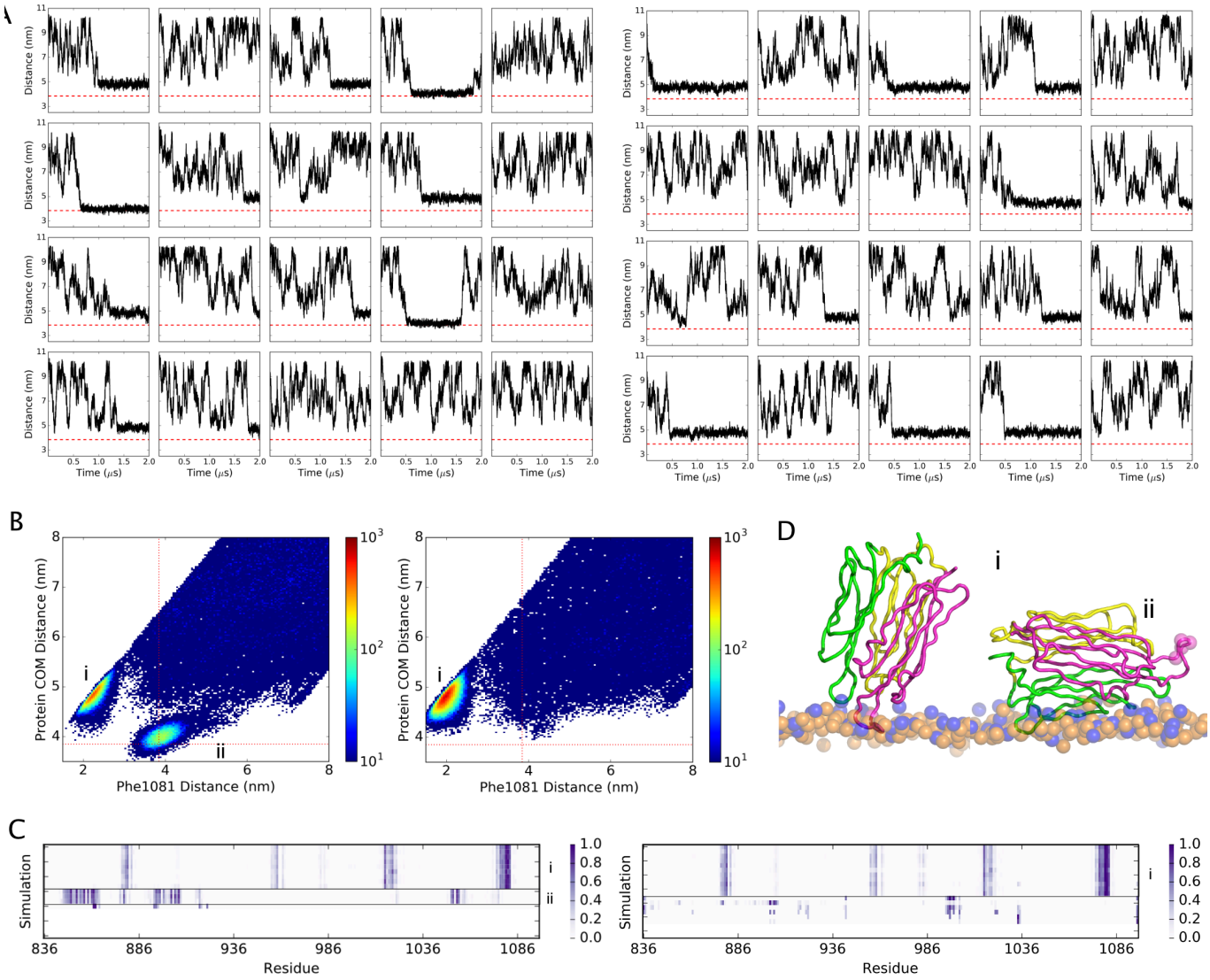


Figure 4.17 PvPLP2 APCβ membrane association simulations.

Simulations of PvPLP2 APCβ on PC:PE:PS (left) and PC:PE:LPA (right) bilayers. (A) COM distances between protein and membrane for each of 20 repeat simulations of 2 μs in length. Red dotted line indicates level of lipid head group phosphate beads. (B) 2D histogram of protein to membrane distance showing two binding orientations. (C) Cartoon representations of each of the two binding orientations labelled in the 2D histogram plots. PvPLP2 is coloured separately for each pseudo-repeat, with Phe 1081 shown with sticks and spheres at the base of the protein. Lipid head groups are shown with blue and orange spheres. (D) Normalised protein-lipid contact plot for all 20 simulations. Each binding orientation is shown surrounded by a box.

Atomistic simulations reveal details of lipid binding

A representative system for each orientation was converted from CG to atomistic detail and simulated again for 100 ns with three repeats for each orientation. The APCβ domain was stable on the membrane in both sets of simulations, with average RMSDs of 2.56 ± 0.21 Å and 2.65 ± 0.25 Å for the primary and secondary orientation respectively **(Figure 4.18)**. The protein remained bound to the membrane throughout all simulations. In one instance the upright orientation began transitioning towards the secondary binding pose throughout the simulation by tilting on its side, however all other simulations maintained their original orientations.

In the primary orientation, PvPLP2 APCβ sits upright along the membrane normal. Protein-lipid contact analysis was performed by identifying all residues within 4 Å of lipid headgroups. These were then counted across each simulation, normalised, and then averaged across the three repeats. Contact analysis showed that binding is facilitated by two loops at the base of the protein, as shown by mapping frequency of lipid contacts onto the structure **(Figure 4.18 C-D)**. F1080 has greatest lipid contacts, and acts as a hydrophobic dagger inserting into the membrane, while L1078 also has significant contact with the membrane. K1077 and K1082 sit higher up on the loop, positioned at the surface of the membrane to interact electrostatically with charged lipid head groups. A secondary lysine rich loop was also identified with the motif ^{878}KNK^{880} at its tip. This loop also sits at the surface of the bilayer and may contribute both hydrogen bonding and electrostatic interactions to stabilise membrane interaction.

The secondary binding orientation bound to the membrane via the first pseudo-repeat. Key hydrophobic residues L862, L857, F903 and Y907 are dispersed across the face of this repeat and identified with high frequency of contact with lipids. Several polar and basic residues along the surface also contribute, with N879, K878 and K905 forming particularly high contact.

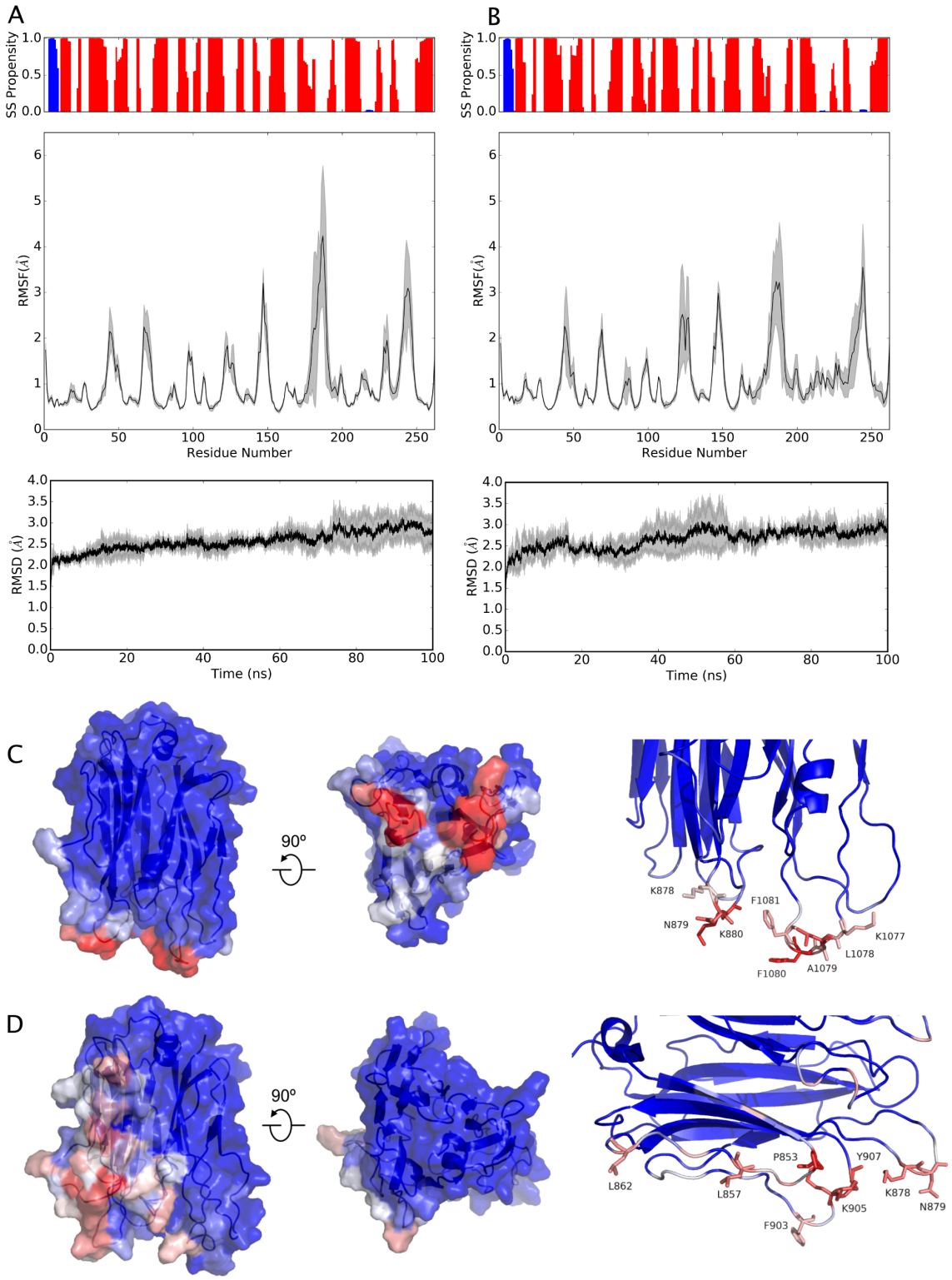


Figure 4.18 Stability of PvPLP2 APCβ CG2AT conversion of membrane binding in upright and sideways orientation.

Starting systems were taken from representative simulations of the upright binding orientation (A) and sideways binding orientation (B) from CG membrane associations. Top panel shows secondary structure propensity with alpha helical structure in blue and beta-sheet in red. The middle panel

indicates Cα RMSF, and the bottom panel shows Cα RMSD for each orientation. Each orientation was simulated in triplicate, for 100ns. All plots show mean value across three repeats and standard deviation. Protein-lipid contacts for PvPLP2 APCβ bound the PC:PE:PS bilayers in upright orientation (C) and sideways orientation (D). Normalised protein-lipid contacts are calculated for each residue across 100ns simulations and plotted onto the structure. High contact areas are coloured in red and low contact areas are coloured in blue. Representation of protein surface (left) shows surface area in contact with the membrane. Details are shown with cartoon representations, with significant residues shown in sticks (right). Protein-lipid contacts are calculated whenever lipid is within 4 Å of a residue. Simulations were performed in triplicate, and contacts were calculated for each simulation then averaged.

These atomistic simulations, along with PMF calculations and coarse-grained simulations, indicate that this secondary sideways orientation of binding is stable. However, as these simulations are conducted with the APCβ in isolation, the orientation of the MACPF domain in relation to the APCβ domain is unknown. There are approximately 9-residues linking the two domains, which is similar in length to the linker between TgPLP1 MACPF and APCβ. SAXS data of TgPLP1 tandem MACPF-APCβ constructs (91) indicate that these two domains, containing a linker of similar length, can rotate in relation to each other, and display two well-defined conformations, in one of which the APCβ domain lies below the MACPF domain and in the other of which it sits behind it. Whether the sideways orientation of the APCβ domain found for PvPLP2 in MD simulations would be possible within the full protein on a membrane, and the functional role of this orientation, is unknown. However, the two orientations could relate to a pre-pore and a pore-forming state of PvPLP2.

Mutations of key PvPLP2 APCβ binding residues

Having used a combination of coarse grained and atomistic simulations to identify binding orientation and key lipid binding residues, we produced a series of mutations to test these observations. A series of *in silico* alanine substitution mutants were produced targeting the primary and secondary membrane anchoring loops, as shown in **Table 4.3**. The full ^{1077}KLAFFK^{1082} motif of the primary loop was mutated to alanines, in addition to mutations of tip phenylalanines and positive charges in the loop to test the critical binding function of this loop. As lipid contact analysis indicated that F1080 had the greatest frequency of contact with lipids, a single point mutation F1080A was also produced. This also provided a direct comparison between the roles of aromatic residue in the extended membrane binding loops of PvPLP2 and TgPLP1.

Region	Mutant
^{1077}KLAFFK^{1082}	KLAAAK
^{1077}KLAFFK^{1082}	ALAFFA
^{1077}KLAFFK^{1082}	ALAAAA
^{1077}KLAFFK^{1082}	AAAAAA
F^{1080}	A
^{878}KNK^{880} and ^{903}FGKK^{906}	^{878}AAA^{880} and ^{903}AAAA^{906}
K^{880}	A

Table 4.3 Summary of PvPLP2 mutations.

As further lipid interactions were identified in secondary loops at the base of PvPLP2, alanine substitution mutations of positively charged regions in two loops ^{878}KNK^{880} and ^{903}FGKK^{906} were also produced to assess the contributions of these loops.

Umbrella sampling was then used to calculate the binding energy of each mutant. Simulations were set up using the mutated APCβ domains in the upright orientation on a PC:PE:PS bilayer to match the wild-type simulations performed previously **(Figure 4.19)**. PMFs confirmed that the phenylalanine tipped loop is critical for binding to the membrane, with the full mutant entirely abolishing binding affinity. The double Phe mutant also had no membrane affinity, showing the critical role of the hydrophobic residues in anchoring the protein to the membrane. The F1080 phenylalanine contributes most of this binding energy, with a reduction in binding energy of 11 kJ mol^{-1} following mutation. On the other hand, mutation of the two positively charged lysines (ALAFFA) only reduced free energy of binding by 2 kJ mol^{-1}. Binding energy was calculated for one further single point mutation of K880A in the secondary binding loop, resulting in a reduction in binding energy by 3 kJ mol^{-1}.

Mutations in the two secondary binding loops ^{878}KNK^{880} and ^{903}FGKK^{906} surprisingly showed an increase in binding over the wild type protein, with a binding energy of -17 kJ mol^{-1} accompanied by a shift in the energy well towards the membrane. A possible explanation for this observation could be that removal of positively charged residues combined with the force pulling PvPLP2 into the membrane improved favourable hydrophobic interactions. As shown in **Figure 4.19 E-G**, this is not experimentally validated. This potentially highlights a shortcoming of umbrella sampling of peripheral membrane proteins. The unphysiological act of exerting force on the protein to pull it into the bilayer is expected to yield unfavourable interaction energies, however in this case it appears to allow the protein access to an orientation that is energetically more favourable than the starting orientation, but nonetheless may not be accessible in unbiased simulations or in biology.

Overall, mutant PMFs showed that the upright binding mode of PvPLP2 is similar in character to that seen with TgPLP1 APCβ. These results indicate that a combination of the two hydrophobic F1080 and F1081 are responsible for a substantial proportion of membrane binding, while positively charged residues in the primary and secondary binding loops contribute to binding to negatively charged lipids through multiple weaker electrostatic interactions. These are likely to contribute proportionally more to the energetics of binding when further charged lipids such as PI(4)P and LPA are added to the membrane.

Site-directed mutagenesis of the hydrophobic tip (^{1077}KLAFFK^{1082} → AAAAAA), and the two secondary charged binding loops (^{878}KNK^{880} → AAA and ^{903}FGKK^{906} → AAAA) of full-length recombinant PvPLP2 was performed by Dr Yu that match those described for *in silico* mutants. Liposome sedimentation assays were then performed, also by Dr Yu, to test binding of mutants to liposomes containing 40% PC, 30% PE and 30% of either CL, LPA or PS **(Figure 4.19 E-G)**. Mutation of the primary binding loop reduced binding to all liposomes, however some binding was retained with both CL and LPA. Mutations of loops 2 and 3 reduced binding to a similar degree as mutation of the hydrophobic tipped loop.

These results complement the PMF calculations in showing the important role for the primary binding loop. While the mutant PMFs were performed with PC:PE:PS bilayers, the liposome assays were also performed using CL and LPA, and importantly indicate that while the hydrophobic tipped loop is important, charged residues play their own important role in binding to negatively-charged bilayers, and can compensate for binding when the hydrophobic tip is lost.

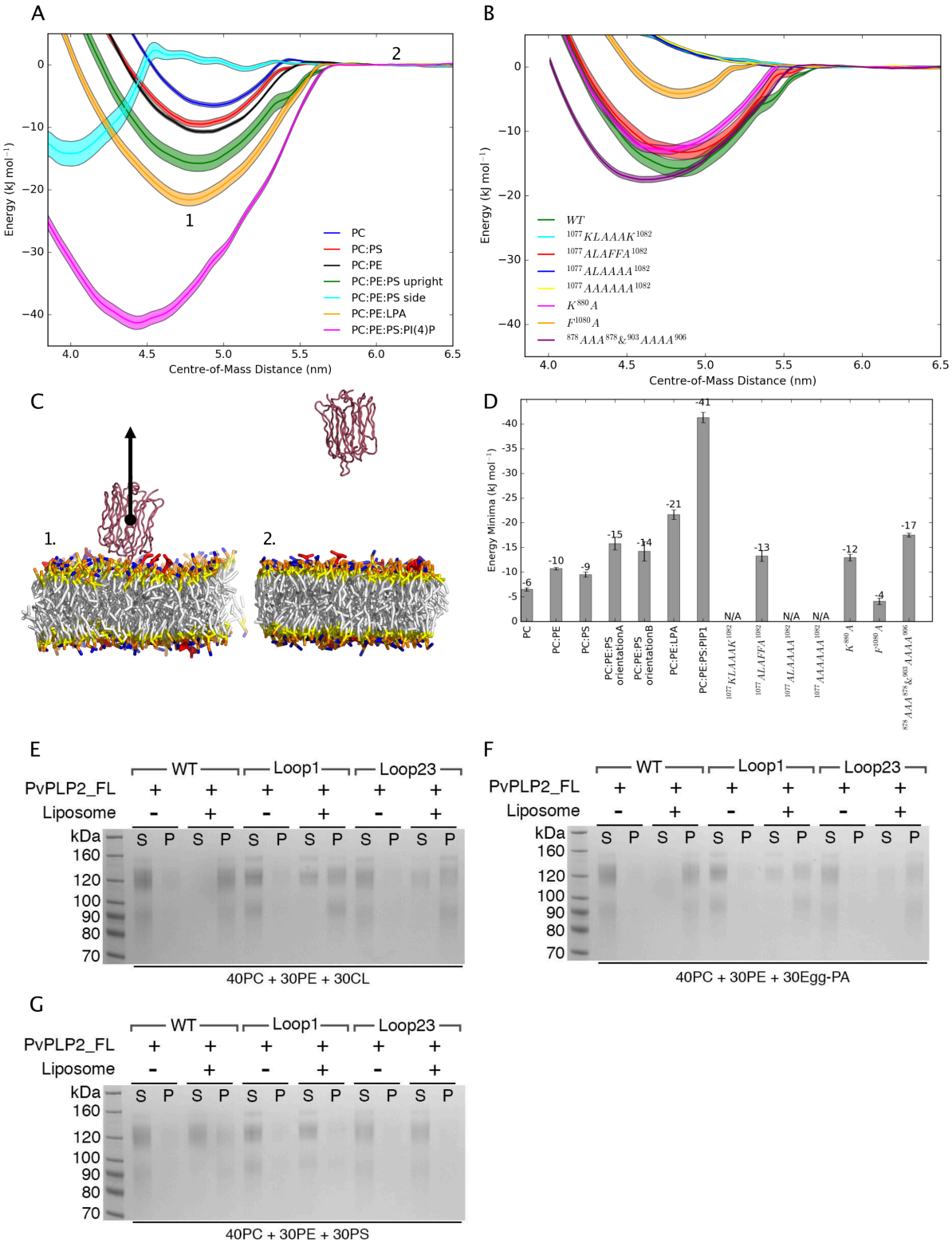


Figure 4.19 Free energy of binding of PvPLP2 APCβ on membranes.

Unless otherwise specified, PvPLP2 is pulled along a reaction coordinate perpendicularly from the membrane starting in the upright binding orientation. Errors are calculated from bootstrap analysis. ***(A)*** *PMF profiles for PvPLP2 APCβ of membranes of different lipid compositions, and in both upright*

and sideways orientation. ***(B)*** *PMF profiles for PvPLP2 alanine substitution mutations.* ***(C)*** *Cartoon representation of PvPLP2 on membrane in the starting orientation (1) and pulled away from the membrane along the reaction coordinate indicated by the arrow to generate umbrella sampling windows (2).* ***(D)*** *Energy minima calculated from PMF profiles give free energy of binding for PvPLP2 on membranes of different lipid composition and with several mutations on the membrane anchoring loops.* ***(E-G)*** *Ultracentrifugation-based liposome-binding assays showing that mutations in the hydrophobic tip and charged residues within the APCβ domain (Loop 1 and Loop 2 3 mutants) reduce the binding of PvPLP2 full-length protein to Cardiolipin* ***(E)****; L-α-phosphatidic acid* ***(F)*** *and phosphatidylserine* ***(G)****. "Loop1" represents the mutant* ^{1077}*KLAFFK*^{1082}*-AAAAAA; "Loop23" indicates the combined mutations* ^{878}*KNK*^{880}*-AAA &* ^{903}*FGKK*^{906}*-AAAA giving charge neutralisation. Liposome sedimentation assay results kindly provided by Dr Yu.*

4.2.5 Lipid specificity of PvPLP2 APCβ

Umbrella sampling investigating the effects of different membrane composition on binding

As shown previously, lipid dot blot experiments and liposome binding assays identified that PvPLP2 has an affinity for negatively charged lipids, in particular PI(4)P, LPA, PS and CL **(Figure 4.12)**. To probe the energetics of PvPLP2 preferences for different lipid species, further umbrella sampling was performed testing free energy of binding to a series of bilayers. PvPLP2 was positioned in the upright orientation, and umbrella sampling simulations on a series of membranes detailed in **Table 4.2** were performed.

A low free energy of binding was observed in a pure PC bilayer, while membrane affinity increased a similar proportion when either PE or PS was added **(Figure 4.19)**. Complementation of PC with PE and PS increased binding energy to -15 kJ mol^{-1}, while addition of PE and LPA resulted in a greater increase in energy to -21 kJ mol^{-1}. Addition of PI(4)P to the bilayer significantly increased binding energy to -41 kJ mol^{-1}.

These results indicate that while hydrophobic interactions with PC result in binding, electrostatics and hydrogen bonding also play important roles in protein-membrane interactions, with the negatively charged PI(4)P and LPA significantly increasing the strength of binding.

Comparison of these results with similar PMFs with TgPLP1 **(Figure 3.10)** show that addition of strong negative charge in the membrane increases binding of PvPLP2 to a greater degree than does PI(3,4,5)P_3 addition to TgPLP1 systems. Overall this indicates that the role of charge is more significant in PvPLP2 than TgPLP1, which is likely due to the greater number of basic residues on the surface of PvPLP2. This increased role of electrostatic interactions for PvPLP2 may reflect functional specificity. While the action of PvPLP2 must be carefully controlled both spatially and temporally to activate at the inner

leaflet of red blood cells in the mosquito midgut, TgPLP1 functions in a broader range of cell types during egress events so perhaps requires less regulation at the stage of membrane attack.

PI(4)P clustering around PvPLP2

We have shown that PvPLP2 has a high affinity for PI(4)P. To investigate the nature of this affinity, we produced three unbiased simulations of PvPLP2 APCβ on a PC:PE:PS:PI(4)P bilayer, each running for 2 μs. The protein was positioned on the membrane in the upright orientation. Radial distribution function (RDF) of lipids around F1080 of the primary binding loop was calculated to identify lipid clustering around the protein **(Figure 4.20)**. Clustering of PI(4)P around the base of PvPLP2 was significant, while clustering of PS and PE was also identified, but significantly less than identified for PI(4)P. Diffusion of PI(4)P, calculated from the mean square of displacement (MSD), was reduced compared to other lipid species indicating clustering of lipids around the protein rather than diffusing freely through the bilayer.

Protein-lipid contacts averaged for all lipids, and specifically for the PI(4)P lipid headgroup were performed. Comparison contacts between the PC:PE:PS membrane association simulations and PI(4)P containing simulations showed an increase in total lipid contact upon addition of PI(4)P, primarily in pseudo-repeat 1 and 2 **(Figure 4.20)**. PI(4)P headgroup contacts were strong in these areas, while PI(4)P made minimal contact with the third pseudo-repeat containing the primary membrane binding loops.

While contact analysis with the PC:PE:PS bilayer showed two key binding loops **(Figure 4.18)**, lipid contacts are increased in three additional loops in pseudo-repeats 1 and 2. Each of these loops contain two lysine residues. Both loops in pseudo-repeat 1 also contain a tyrosine residue, while the additional loop in repeat 2 contains polar residues serine and asparagine.

Observation of PIP diffusion in these simulations shows that the lipids cluster around positively charged residues, but do not appear to associate and remain in specific binding pockets. This is also confirmed when averaged POP1 headgroup contacts are plotted on to the PvPLP2 structure, showing a dispersed pattern of contact concentrated around the base of pseudo-repeats 1 and 2.

4.3 Conclusion

Here we have described the crystal structure of PvPLP1 APCβ domain, the first crystal structure of a *Plasmodium* perforin-like protein to date. This structure has shown that the PvPLP1 APCβ shares a common domain structure with the *Toxoplasma* PLP1 APCβ domain (91) which, along with bioinformatic analysis, supports the presence of a conserved membrane binding domain unique to Apicomplexans. While the overall architecture of these domains has been conserved, we have shown that the sequence, particularly in loops at the base of the protein, is poorly conserved between different PLPs. Molecular dynamics simulations and lipid binding assays have shown that PvPLP1, PvPLP2 and TgPLP1 membrane binding characteristics are not conserved across the family.

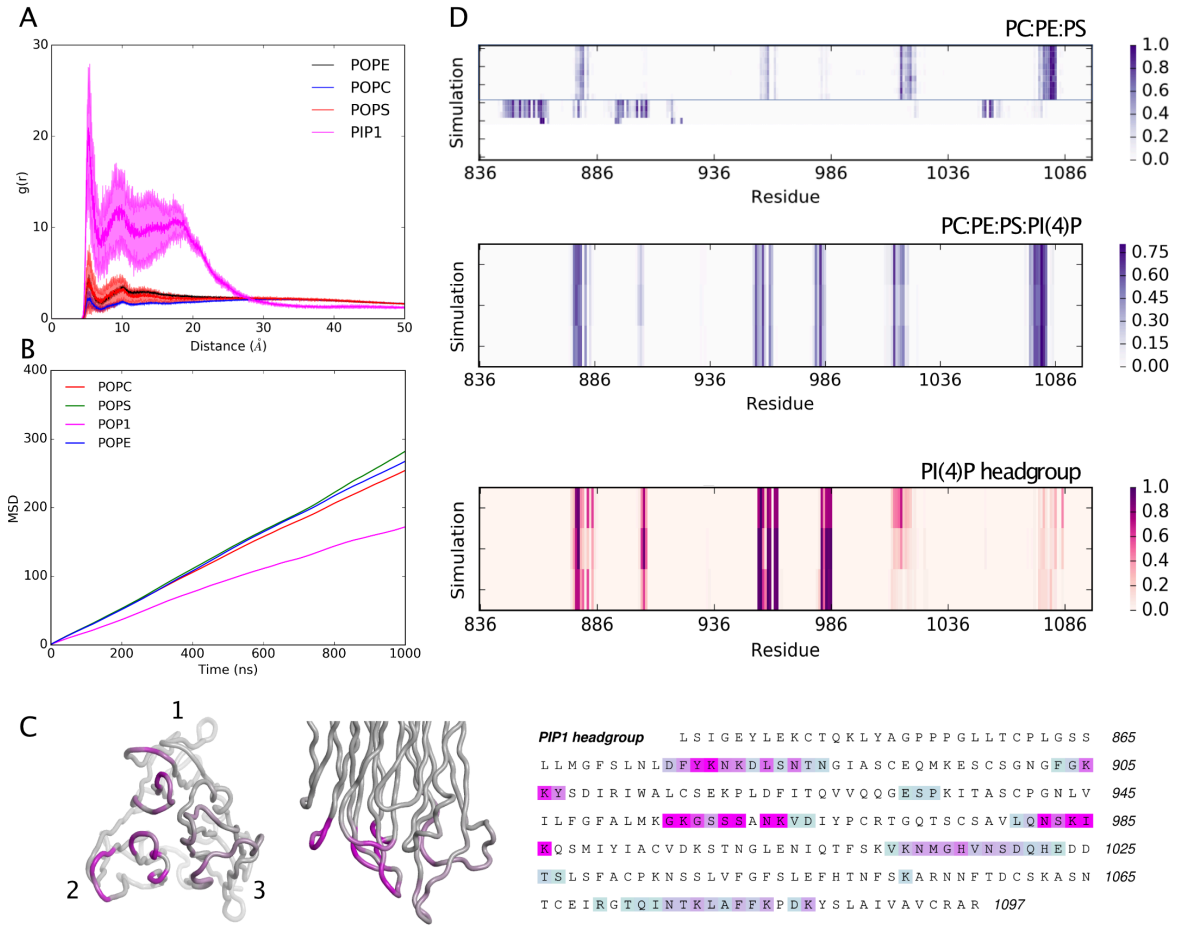


Figure 4.20 CG simulations show clustering of PI(4)P around PvPLP2 APCβ.

PvPLP2 was simulated in the upright orientation on a PC:PE:PS:PI(4)P bilayer for 2 μs with 3 repeats. (A) Radial distribution function was calculated for each lipid headgroup around Phe1080. (B) Representative MSD for each lipid species. (C) Comparison between lipid-contact fingerprint plot in membrane association simulation on PC:PE:PS bilayer, with box around simulations in upright orientation (top), averaged lipid-contacts for PC:PE:PS:PI(4)P simulations (middle), and normalised contacts with PI(4)P lipid headgroups alone (bottom). (D) Normalised frequency of PI(4)P headgroup contact plotted onto a cartoon representation of PvPLP2 APCβ, with each pseudo-repeat labelled, and onto the protein sequence to identify key binding residues.

We have shown that PvPLP2 shares a similar membrane binding mechanism to TgPLP1. Both proteins utilise an extended loop at their base, tipped with an aromatic residue, to anchor into the membrane and adopt similar binding orientations. In addition to the hydrophobic anchoring into the bilayer, binding of PvPLP2 appears to be significantly strengthened by weaker electrostatic interactions across the base of the protein, leading to a strong but non-specific affinity for negatively charged membranes. In comparison to TgPLP1, we posit that this increased sensitivity to negatively charged lipids in the membrane has evolved to adapt to the specific functional requirements of PvPLP2. Specifically, the recognition of erythrocyte inner membrane during gametocyte egress.

In contrast, PvPLP1 did not bind to membranes in any of our experiments. It is possible that PvPLP1 binds to specific lipids or proteins in the membrane that were not tested here or uses an entirely different mechanism to bind to the membrane.

Overall, these differences support the theory that each PLP has evolved to function in specific cellular environments, suggesting that further characterisation of each PPLP individually is warranted. The function of each PLP may be tailored to its lifecycle stage, requiring presence of specific lipids, membrane characteristics, or other ligands, to bind to the membrane and disrupt pores. These domains act as the first step in pore formation by PLPs, and therefore understanding the molecular basis behind membrane binding could provide a basis for developing a therapeutic target for *Plasmodium* PLPs.

During study of PvPLP2 membrane binding, we also identified a second binding orientation which has not been previously observed for any MACPF/CDC family protein. While this may be a result of simulation of the membrane-binding domain alone, if valid, it may have implications for the mechanism of pore formation across the ApiPLP family as a whole and could represent a transitionary state during pore formation. Recent structural studies of perforin-2 in our lab revealed an unexpected orientation of the MACPF domain and membrane binding P2 domain, in which the MACPF and P2 domains exist in an antiparallel orientation in monomers. They then undergo large-scale rotation to orient the MACPF towards the membrane for pore formation **(see Chapter 5)**. These unexpected results highlight the possibility of previously unobserved larger scale structural rearrangements during pore formation, indicating that the relevance of this orientation should not be dismissed without further study.

Chapter 5

Perforin-2

5.1 Introduction

Perforin-2 is a pore forming effector protein in the innate immune system. Originally identified in macrophages and named macrophage expressed gene 1 (MPEG1) (221), perforin-2 (PFN2) contains a MACPF domain and is found in a wide range of animals, both invertebrates including sponges, the oldest metazoans still in existence, and many vertebrates including humans (122). Due to its evolutionary conservation, PFN2 is thought to be part of an ancient innate immune defence system against bacterial pathogens and therefore an important anti-bacterial agent (122). While its role in the human immune system is still being established, several mutations have been identified that lead to increased susceptibility to infections with intracellular bacteria (127). Several *in vivo* studies have also shown its importance in clearance of intracellular bacterial infections in both mice and zebrafish (124, 222).

5.1.1 The role of perforin-2 in the immune system

While originally identified in Macrophages, PFN2 is expressed in many parenchymal cells, including fibroblasts and epithelial cells (124-126). PFN2 has a high basal expression level in macrophages (125), while in other cells expression is induced during bacterial infection or following or interferon exposure (124-126). In both professional phagocytes and non-immune cells, PFN2 aids the cell in killing both Gram-positive and Gram-negative intracellular bacteria (124). Studies have shown that knockout of PFN2 activity in both macrophages and other cell lines results in great reduction or complete inability to kill intracellular bacteria, while *in vivo* studies showed that mice lacking PFN2 are susceptible to both MRSA and *S. Typhimurium* infection (124). Several mutations have been identified in humans that have been linked to pulmonary non-tuberculosis mycobacterial infection along with susceptibility to further infections with intracellular bacteria (127). These gene variants were shown to reduce the efficiency with which mycobacteria could be killed as well as several other intracellular bacteria including *Salmonella typhimurium* and *Staphylococcus aureus* (127).

5.1.2 Molecular insights into perforin-2 activity

Unusually for pore forming proteins, PFN2 contains a transmembrane region. With its MACPF domain at the N-terminus, responsible for pore formation, and an additional conserved domain labelled the perforin-2 domain (P2), it also has a C-terminal transmembrane α-helix. The MACPF and P2 domains are thought to be localised to the extracellular side of the membrane (within cytoplasmic endosome-like vesicles), while a short cytoplasmic domain projects into the cytoplasm (122).

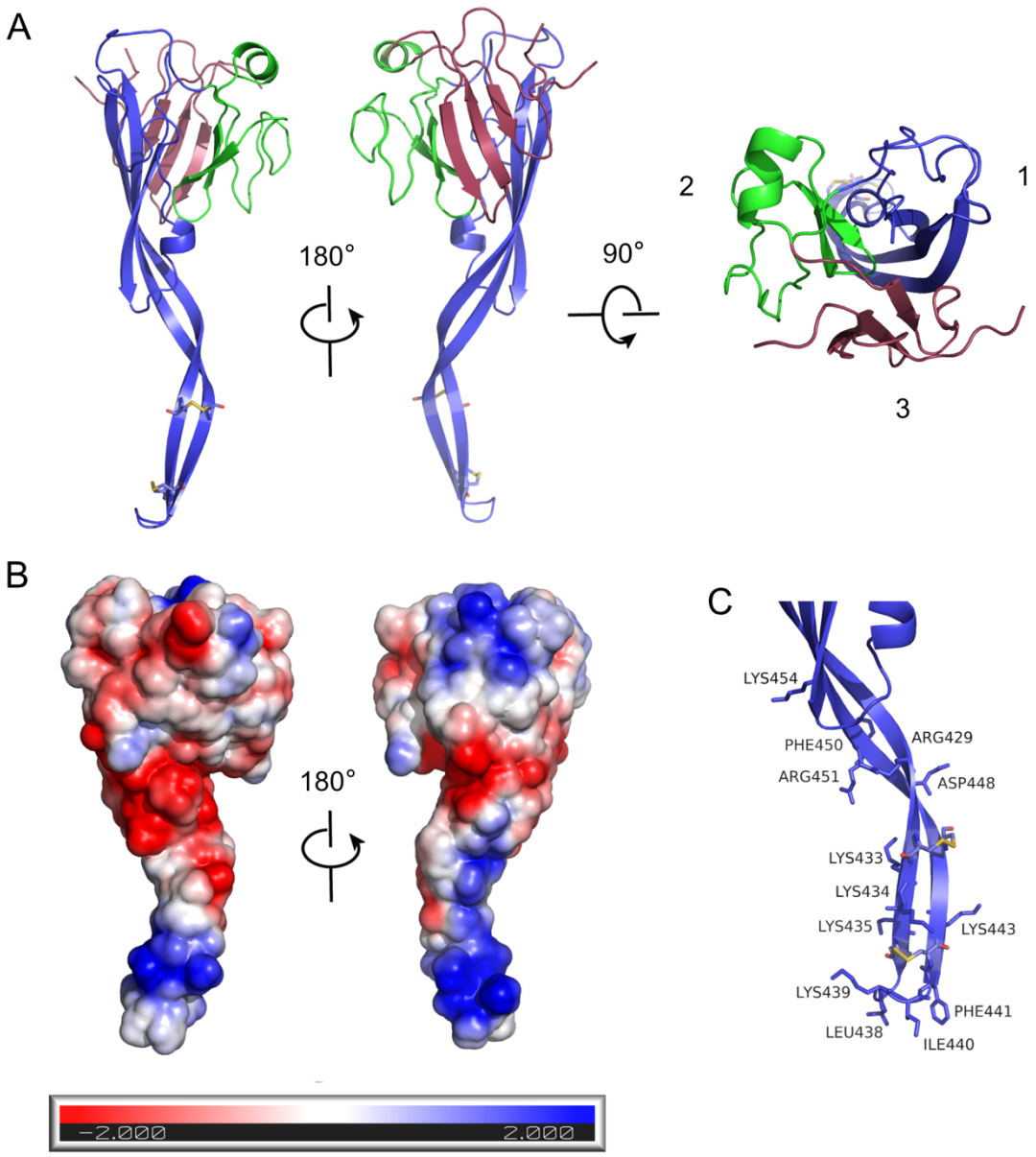


Figure 5.1 The Perforin-2 P2 domain crystal structure

A cartoon representation (A) and electrostatic surface representation (C) of the PFN2 P2 domain are shown. Each pseudo-repeat in the structure is coloured separately and labelled from 1 to 3 in the right-hand panel of (A). (C) Zoomed in cartoon representation of the β-hairpin of P2, showing two disulphide bonds and critical hydrophobic and positively charged residues. Structure solved by Lucy Ginger and Dr Tao Ni.

PFN2-containing vesicles fuse with endosomes containing bacteria following infection (124, 223). Identification of a second isoform of PFN2 lacking the transmembrane segment and cytoplasmic domain suggests it may also be exist as a soluble protein within endosome-like vesicles (223).

Following intracellular infection, PFN2 forms pores on bacterial outer membranes, resulting in clusters of pores around ~100 Å each in diameter (124). Formation of pores on the bacterial cell wall has been proposed to facilitate access of effector molecules. In professional phagocytes, PFN2 is required for susceptibility of phagocytosed bacteria to killing by reactive oxygen species (ROS) and reactive nitrogen species (RNS) (124), while the Gram-positive bacterium *Mycobacteria smegmatis* cannot be killed by lysozyme attack unless perforin-2 is present (125).

5.1.3 The perforin-2 structure

The structure of the murine PFN2 pre-pore has recently been solved in our group using X-ray crystallography and cryo-electron microscopy (cryo-EM).

The crystal structure of the PFN2 P2 domain in isolation was first solved using X-ray crystallography by Lucy Ginger and Dr Tao Ni. Solved at 2.05 Å resolution (manuscript under review), the structure is β-stranded in nature, consisting of three pseudo-repeats arranged in a prism **(Figure 5.1)**. Although similar in nature to the Apicomplexan PLP APCβ domains, each repeat contains only three β-strands locked together by disulphide bonds. The first repeat contains one extended β-hairpin that is 45 Å long and locked together by two further disulphide bonds for stability. Due to its location next to the MACPF domain, and structural similarities with other MACPF/CDC membrane binding domains, we have hypothesised that this domain is responsible for membrane binding.

Next, the full PFN2 structure was solved using cryo-EM as an oligomeric pre-pore containing 16 subunits formed in solution at endosomal pH 5.5. The PFN2 monomer consists of a MACPF domain, P2 domain and a C-terminal tail region (CTT). The MACPF domain is similar to the canonical MACPF domain structure and is linked to the P2 domain by a short linker, however the globular core of P2 forms close contacts with the MACPF domain. The P2 domain solved in these structures is similar to the crystal structure, while the CTT extends from the core of the P2 domain and folds around the interface between the P2 and MAPCF domains, ending along the top of the MACPF domain with a small β-sheet followed by a short α-helix (**Figure 5.2**).

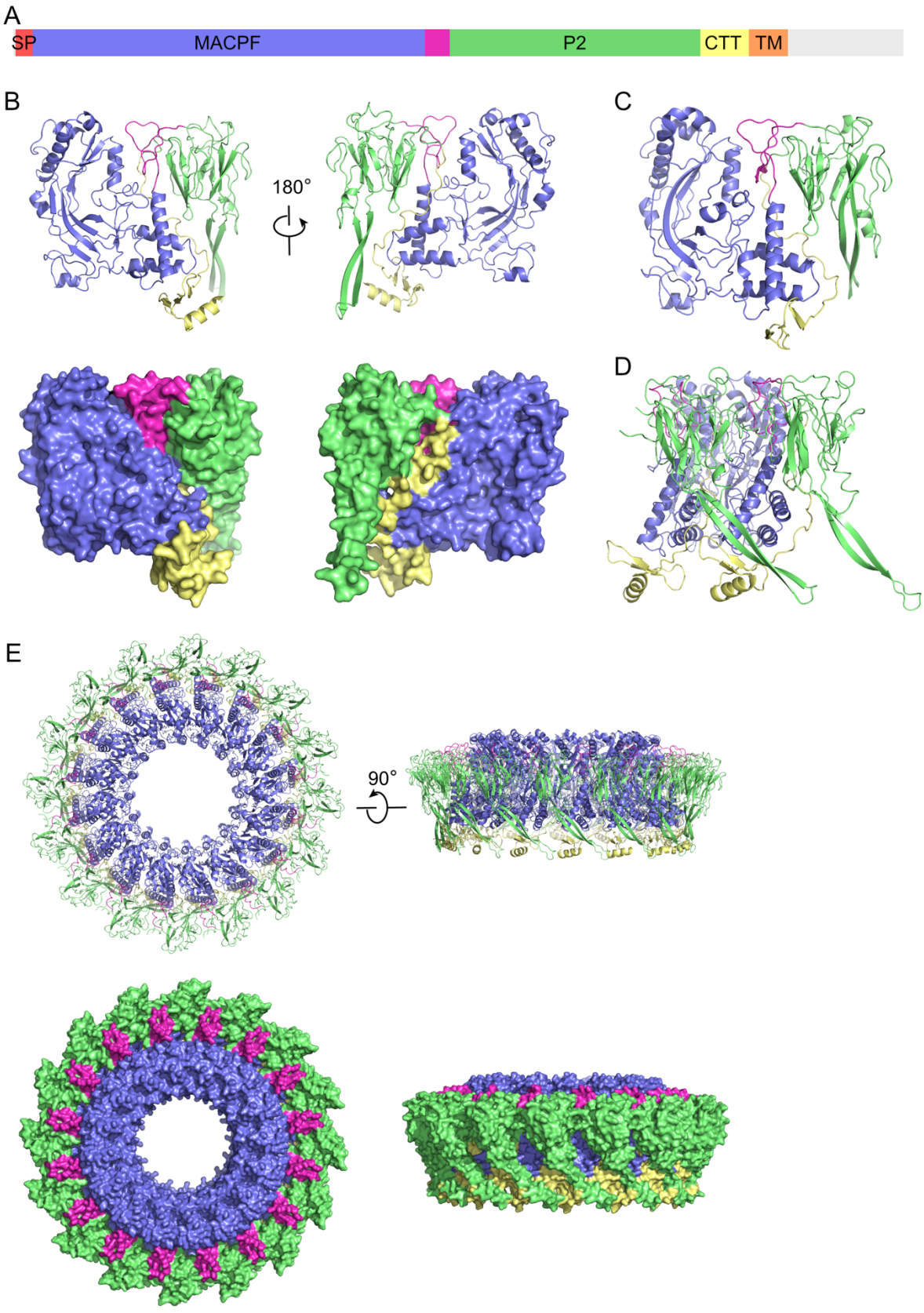


Figure 5.2 The perforin-2 pre-pore structure

(A) Schematic of the domain architecture of PFN2. SP = Signal peptide. CTT = C-terminal tail. TM = transmembrane segment. In both the schematic and following structural diagrams, the MACPF domain is coloured in blue, P2 domain is green, the linker between the MACPF and P2 domains is coloured in magenta and the CTT is coloured in yellow. (B) Surface and cartoon representations of a single copy of PFN2. (C) Initial structure of PFN2 missing the tip of the P2 β-hairpin, and CTT terminal helix. (D) Two subunits of the PFN2 pre-pore showing the oligomerisation surface on the MACPF domain and P2-CTT β-sheet. (E) Cartoon and surface representations of the 16-mer PFN2 pre-pore.

Two pre-pore structures were solved in our group, the second of which was solved part way through this study. The two structures were highly similar with a few exceptions making the second structure more complete. In the first structure the P2 domain was missing density at the tip of the extended β-hairpin (**Figure 5.2 C**) , which was complete in the second structure. The first structure was also missing the extreme C-terminus of the CTT, which consists of 12 residues forming a short α-helix (**Figure 5.2 B**).

Oligomerisation of PFN2 in the pre-pore shows inter-subunit interfaces formed by PFN2 initially before pore formation. PFN2 monomers remain in an almost vertical orientation following oligomerisation, with the flat face of the MAPCF domains forming the oligomerisation surface, the narrow edge of the MACPF domain facing into the pre-pore lumen and the P2 domain oriented outwards from the pore. The CTT β-sheet forms a continuous β-sheet with the P2 of the adjacent subunit, forming a secondary set of contacts between each subunit alongside the main MACPF domain core contacts (**Figure 5.2 D**).

Unusually, the MACPF domain and P2 domain are found in an anti-parallel orientation. The extended P2 β-hairpin, which we hypothesise is responsible for membrane binding, is found on the opposite edge of the protein from the MACPF TMH helices, leading us to question how the monomers bind to the membrane, how the pre-pore remains associated with the membrane, and how pore formation occurs.

In this chapter we aimed to investigate the initial steps of PFN2 pore formation. We first investigated membrane binding of the P2 domain, PFN2 monomers and full pre-pore structures, in order to evaluate the binding orientation of this unusual structure as well as identify lipid binding preferences. We then further investigated the dynamics of the PFN2 monomer to further understand the structural transition from pre-pore to pore.

5.2 Results and Discussion

As several structures of perforin-2 were solved in the Gilbert group, we aimed to use MD simulations to build up a picture of PFN2 membrane interactions both before and during pore formation. We first aimed to understand the molecular basis with which the protein can bind to membranes, the first stage in pore formation. As most C-terminal domains of MACPF/CDC proteins are responsible for membrane

binding, we initially assessed the P2 domain of PFN2 in isolation to determine if it is responsible for membrane targeting.

5.2.1 Atomistic simulation of P2 domain monomer in aqueous solution

The P2 domain consists of three β-stranded pseudo-repeats in a prism arrangement. The first repeat contains a very long extended β-hairpin. This hairpin has a hydrophobic tip, with a phenylalanine residue at its extreme end, while the rest of the hairpin is basic in nature, containing a high proportion of lysine residues **(Figure 5.1)**. The extended hairpin is locked together by two disulphide bonds, which are conserved across species. These suggest that, while unusual, this structure is conserved and stable. Density was missing in the crystal structure between residues 540-548, which were first connected together using Modeller (224). We then performed MD simulations to assess this structure in water, in particular the flexibility of the extended hairpin.

The P2 domain was simulated for 100 ns in triplicate. RMSD and RMSF calculations show that the domain is very stable, with an average RMSF of 3 Å (**Figure 5.3**). The core of the domain remained rigid, while significantly more motion was observed in the extended β-hairpin, with greatest motion at its tip. Despite an increased flexibility, the secondary structure of the hairpin, and domain as a whole, was well conserved. The extended and flexible nature of this hairpin, along with its dual hydrophobic tip and basic nature, suggests it has adapted a specific role in membrane binding.

5.2.2 P2 domain simulations

We next aimed to assess the membrane binding capacity of the P2 domain. Using CG membrane association simulations, we assessed the binding of the P2 domain to a series of both eukaryotic and prokaryotic membranes, whose lipid compositions are listed in **Table 5.1**. Structural analysis of the domain suggested that the extended β-hairpin loop, tipped with a phenylalanine residue, may be responsible for membrane binding in a similar mechanism to that identified for many MACPF membrane binding domains.

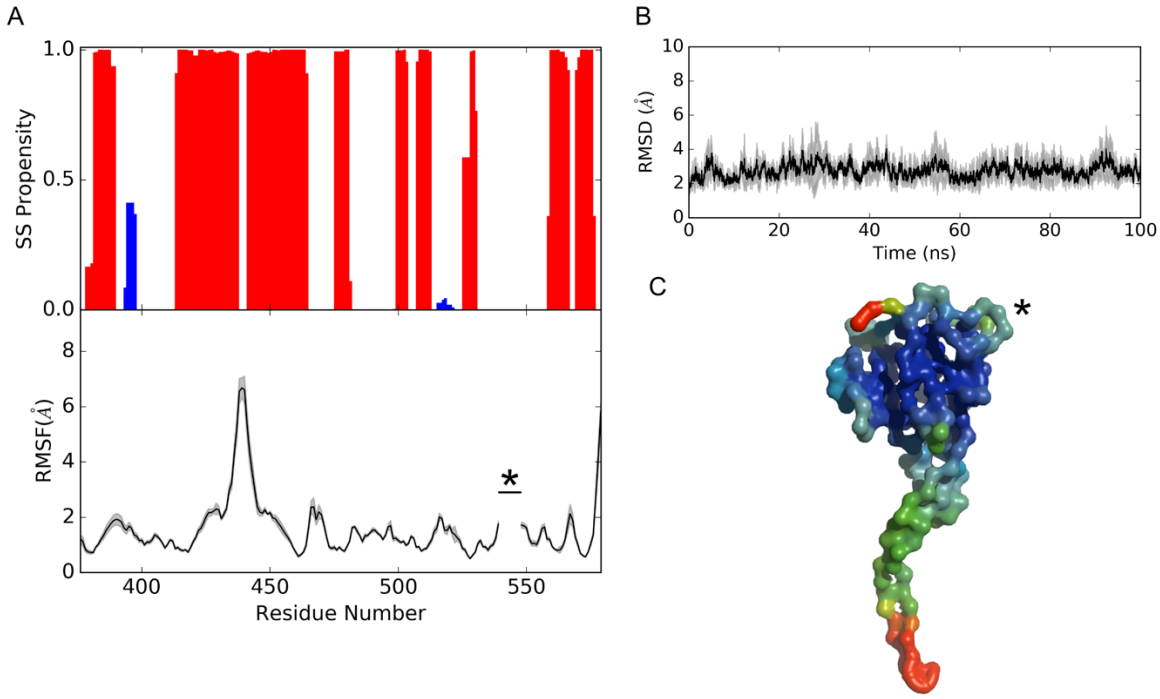


Figure 5.3 Stability of P2 domain in solution

RMSF (A), secondary structure propensity (A) and RMSD (B) of the PFN2 P2 domain simulated for 100 ns. (C) RMSF values were plotted onto a surface representation of the P2 structure and coloured from blue to red. Mean values were calculated from three repeat simulations and standard deviations are shown. Missing residues in the crystal structure are highlighted with an asterisk.

Twenty-five simulations of 2 μs in length were run per bilayer, with the protein starting orientation sequentially rotated around the X, Y and Z axis as previously described. As perforin-2 is known to attack bacterial membranes, binding was assessed with PG:PE bilayers, as well as two lipopolysaccharide (LPS) containing membranes representing rough (RER/Ra-LPS) and smooth (SER/Re-LPS) Gram-negative LPS outer membrane. Binding to several eukaryotic membranes was also tested. The protein associated in all simulations with all eukaryotic and prokaryotic bilayers tested, except in a single simulation with the Ra-LPS membrane (**Table 5.1**).

On average, time to membrane binding was reduced upon addition of charged lipids in phospholipid bilayers. The P2 domain took longer to bind to the PC:PE bilayer, while it quickly found PG:PE and PC:PE:PS:PI(3,4,5)P_3 bilayers, indicating a preference for charged bilayers. Binding to LPS membranes also occurred quickly, in particular to Re-LPS.

Membrane	Lipid compositions	Simulations run/duration (μs)	Number of membrane binding events
PC:PE	55:45	25/2	25
PC:PE:PS	45:42:10	25/2	25
PC:PE:PS:PIP3	45:42:10:3	25/2	25
PG:PE	25:75	25/2	25
Re-LPS	Upper leaflet – 100 % LPS Lower leaflet – PG:PE 25:75	25/2	25
Ra-LPS	100% LPS	25/2	24

Table 5.1 *Lipid composition of membrane association simulations performed with PFN2 P2 domain*

Observation of the simulations suggested that the P2 domain bound to the membrane in several orientations. In all eukaryotic phospholipid bilayers, two binding orientations were identified (**Figure 5.4**). In both orientations the P2 domain was bound to the membrane by F441 at the tip of the extended β-hairpin, however in pose A, the domain remained vertically oriented along the membrane normal, while in pose B a greater surface area of the hairpin, including F450, bound to the membrane, pulling the P2 domain axis more parallel with the bilayer plane (**Figure 5.4 C**). Binding to the PG:PE bilayer also occurred in two orientations. Both orientations resemble pose B, whereby the tip Phe residue is anchored into the membrane and the P2 domain is tilted towards the membrane. However, in one orientation Phe450 is directly bound to the membrane, as shown in pose B, while in the other the protein is rotated 180° so the Phe450 face is directed away from the membrane. In comparison to the phospholipid bilayers, binding to Re-LPS and Ra-LPS membranes was more complex, with multiple binding orientations adopted. In almost all cases however, the P2 domain bound to the membrane with the extended β-hairpin, details of which will be explored further by lipid-contact analysis (see below).

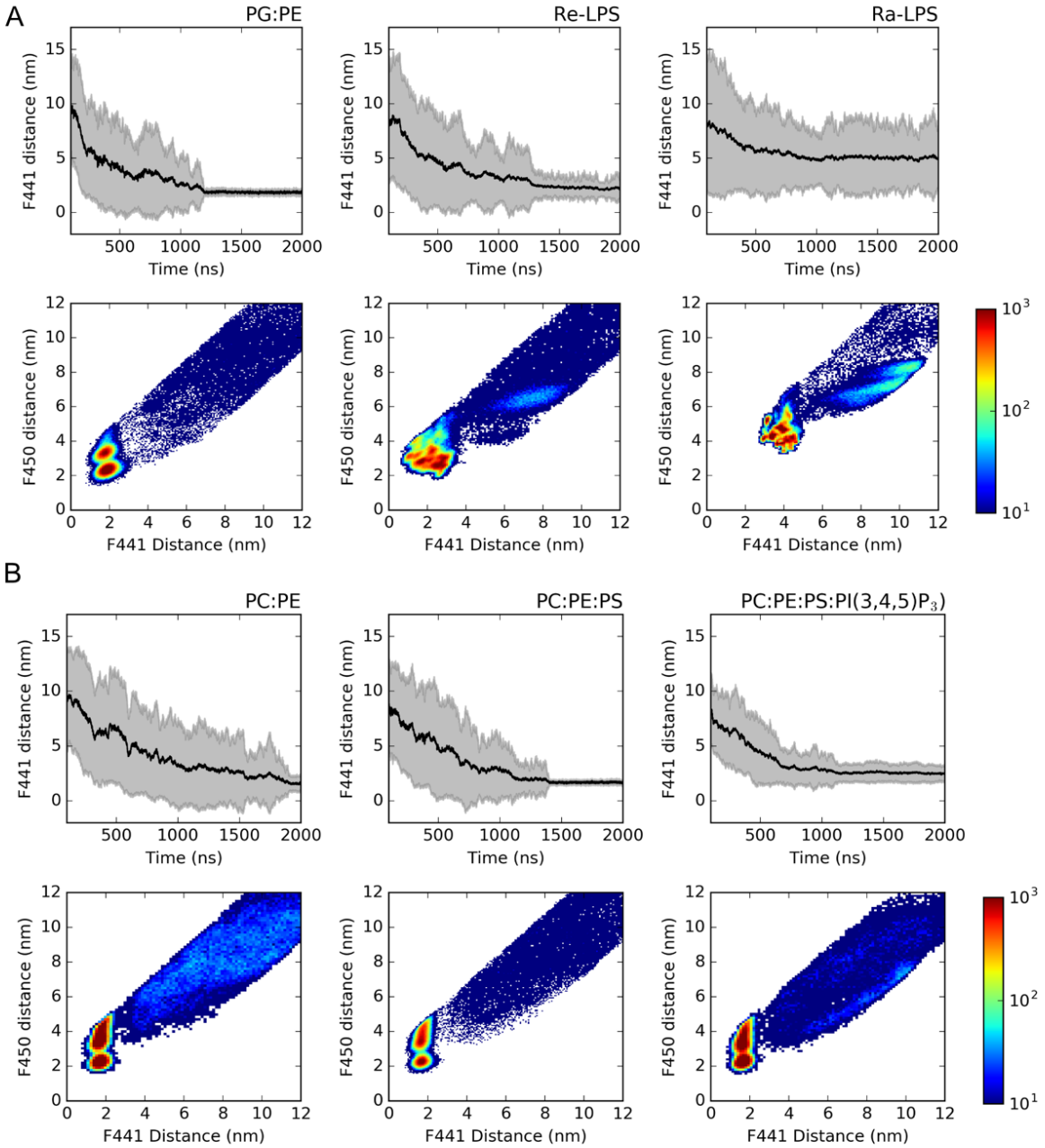

A
PG:PE
Re-LPS
Ra-LPS
F441 distance (nm)
Time (ns)
F450 distance (nm)
F441 Distance (nm)
B
PC:PE
PC:PE:PS
PC:PE:PS:PI(3,4,5)P3
C
Phe450
Phe441
Pose A
Pose B

Figure 5.4 Perforin-2 P2 domain membrane association simulations on prokaryotic and eukaryotic membranes

25 membrane association simulations were performed on each representative prokaryotic (A) and eukaryotic (B) membranes. Mean distance between the P2 domain F450 and COM of the bilayer is shown against simulation time, with errors shown as standard deviation. 2-dimensional histogram plots represent the binding orientations adopted by the domain by measuring distance between the COM of the bilayer and F450 at the tip of the extended β-hairpin, and F441 near the start of the extended β-hairpin. (C) Cartoon representation of the P2 domain showing location of F441 and F450 (left) and the two binding orientations pose A and pose B on the membrane (right). Lipid headgroups are represented by spheres.

These results suggest that the P2 domain has a non-specific membrane binding capacity. The protein primarily anchors into the membrane using its phenylalanine-tipped extended β-hairpin, in a hydrophobic interaction that initially guides the domain to the bilayer. The P2 domain is then able to remain in this upright orientation, or tilt towards the membrane into a secondary pose, in which the length of the β-hairpin is in contact with the membrane. Speed of membrane association appears to depend on charge density of the bilayer, with a greater proportion of charged lipids leading to faster membrane association. This is likely explained by the positively charged residues found along the β-hairpin strands, which also associate with the membrane when tilted towards the bilayer. The more complex binding map of Re-LPS and Ra-LPS is likely due to the hydrophilic nature of the core polysaccharides, allowing greater bonding between the P2 domain and the membrane. In these cases, the β-hairpin remains important for contact with the membrane. Binding characteristics are further explored next using protein-lipid contact analysis.

Lipid binding of the P2 domain

Following identification of membrane binding patterns of the PFN2 P2 domain, binding was assessed in greater detail to identify residues responsible for binding. As PFN2 is a bactericidal agent, we specifically focused on binding to bacterial membranes including the PG:PE bilayer, representative of the bacterial inner membrane, and LPS bilayers Re-LPS and Ra-LPS of Gram-negative bacteria.

Lipid-contact analysis of the P2 domain bound to PG:PE bilayers clearly identified two distinct binding patterns as described earlier. Fingerprint plots show that in both orientations the β-hairpin makes primary contact with the bilayer, while the core of the domain contacts the membrane in one of two orientations (**Figure 5.5**). In the first orientation, Phe450 binds to the bilayer, and pseudo-repeats 1 and 3 are mostly in contact with the bilayer (**Figure 5.5, Orientation 1**). In contrast, in the second orientation the opposite side of the domain binds, with pseudo-repeat 2 mostly in contact with the membrane (**Figure 5.5, Orientation 2**). The first orientation described occurred more frequently, in

roughly 68% of the simulations. Contacts with PE and PG lipids were plotted individually to determine any lipid specificity. Patterns of contact were very similar for both lipids, with slightly higher contact with PE overall. Residues at the hydrophobic motif ^{437}TLKIFC^{442} at the tip of the β-hairpin all made strong contact with the membrane, and residue Phe441 inserted into the membrane and the interface between the lipid head groups and tails.

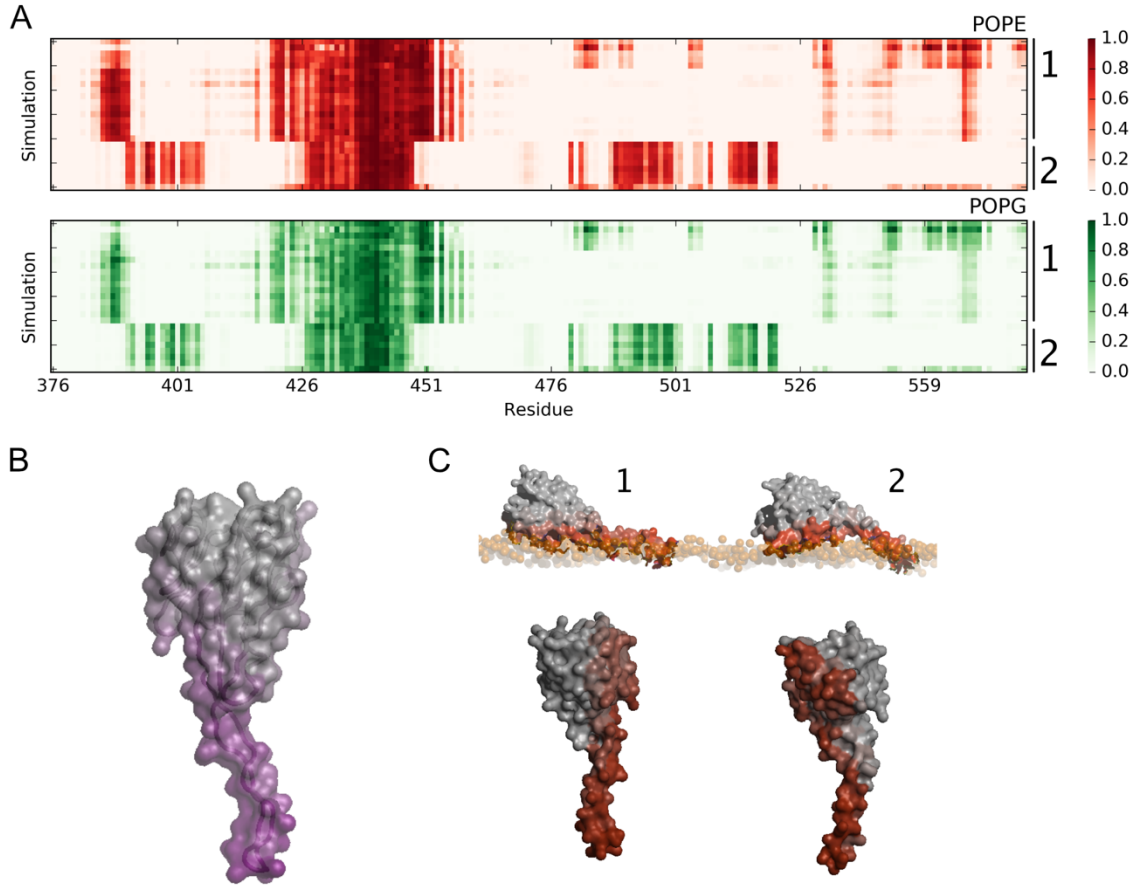


Figure 5.5 PG:PE lipid contact analysis of PFN2 P2 domain

(A) Normalised protein-lipid contact fingerprints showing P2 binding to PG:PE membranes in each of 25 membrane association simulations. The two identified binding poses are identified as orientations 1 & 2. (B) Averaged contacts for both PG and PE lipids across all membrane association simulations plotted onto the P2 structure, coloured from white to purple. (C) Averaged lipid contacts calculated separately for each of the two orientations, plotted onto the P2 structure both binding to the membrane (top panel) and onto an aligned structure (bottom panel), showing binding on opposite faces on the domain.

Contacts with Ra-LPS and Re-LPS were also analysed. As was seen with protein orientation, contacts with LPS were more variable across simulations **(Figure 5.4)**. The extended β-hairpin was bound to the membrane in most simulations. Through plotting averaged lipid contacts onto the P2 domain structure, we can see that the β-hairpin makes greatest lipid contact across the structure, consistent with a role in membrane anchoring. Based on contact fingerprints, four different binding orientations were identified on Re-LPS. These contacts varied primarily in the region of the core domain that was in contact with the membrane, showing a rotation around the protein axis. Flexibility in the β-hairpin facilitates this rotation of the binding surface while maintaining Phe441 contact. Ra-LPS contacts had a similar pattern of variability as Re-LPS, however contact with the globular core of the domain was greater than with Re-LPS. In addition, a subset of four simulations bound to the membrane by the top of the core of the domain near the N- and C- terminus rather than the β-hairpin.

As Gram-negative bacteria outer membranes contain LPS, the phospholipid bilayer is protected from attack by the long polysaccharide O-antigen chains. Therefore, we proposed that the function of the extended-β-hairpin may be to insert into the polysaccharide chains to anchor PFN2 and facilitate membrane attack. Thus, contacts between the protein and lipid A tails, headgroup and LPS core sugars were calculated separately **(Figure 5.6)**. As expected, contact was greatest with the LPS core on the surface of the membrane for both Re-LPS and Ra-LPS. In Re-LPS, which contains only two sugar molecules from the LPS core, contact with lipid A headgroups was primarily made by the β-hairpin, and a few residues dispersed across the core of the domain. Contact with Lipid A tails was made in 48% of simulations and was restricted to residues at the tip of the β-hairpin, showing burrowing into the membrane. As Ra-LPS contains a longer polysaccharide chain, no residues were able to contact the lipid A tails, however the β-hairpin inserted into the LPS and contacted the lipid A headgroups in 3 simulations **(Figure 5.6 D**).

Atomistic detail of P2 domain binding to PG:PE membranes

A representative frame from each membrane binding orientation on PG:PE bilayers was converted to atomistic resolution and simulated for 100 ns to further evaluate binding. In contrast to CG simulations, AT simulations maintain no secondary structure restraints or elastic networks on the protein structure, and therefore in combination with the greater resolution, allow the protein to settle onto the membrane and optimise lipid interactions and structure.

In comparison to simulation in solution alone, membrane binding stabilised the domain overall, as well as at the β-hairpin (**Figure 5.7**). In pose 1, the extended β-hairpin maintains similar membrane interactions throughout the simulation, while the core of the domain flattens onto the membrane and pseudo-repeat 3 buries into the membrane core (**Figure 5.8 B**). In contrast, this region is facing away from the bilayer and exposed to water in pose 2. This buried segment is found at the top corner of pseudo-repeat 3 around residue 550, and is destabilised in pose 2, while burying into the core stabilises

it (**Figure 5.7**). In pose 2 the extended β-hairpin buries further into the bilayer throughout the simulation, while the second pseudo-repeat remains peripherally bound to the membrane.

Discussion of membrane binding

In this section we have assessed membrane binding of the P2 domain of PFN2 in isolation. We predicted that this domain would bind to membranes using the extended β-hairpin located in the first pseudo-repeat of the structure. This prediction was based on similarities to other MACPF/CDC membrane binding domains including the ApiPLP APCβ domain (see Chapter 4), specifically the characteristic extension from a core β-sheet rich domain and hydrophobic tip motif including an aromatic phenylalanine residue.

Using membrane association simulations, we have confirmed that this domain does indeed bind to membranes using the phenylalanine tipped extended β-hairpin. The protein had high affinity for all membranes tested, binding in over 99% of simulations and showed a preference for negatively charged lipids. Binding appears to be non-specific, due to the hydrophobic characteristics of the Phe tip of the extended β-hairpin which directed association with the membrane. Binding was further strengthened by interaction between charged lipids, such as PS, PIPs and LPS, and positively charged residues along the length of the extended hairpin. These results were also confirmed in our group through pull-down experiments with liposomes of various lipid concentrations, that showed a preference for positively charged lipids. Truncation of the P2 β-hairpin was also performed which entirely blocked membrane binding in pull-down experiments, while mutation of the β-hairpin in fibroblasts reduced their ability to kill intracellular bacteria (manuscript under review).

This suggests that PFN2 may bind to both eukaryotic and prokaryotic membranes, highlighting the need for tight regulation of expression and activation of pore formation, including rapid degradation by the proteasome (125), and pH-dependent activation (104, 225). While some MACPF/CDC membrane binding domains recognise a specific lipid, such as cholesterol in CDCs (51, 54), protein, such as CD59 in some CDCs (226), or environmental conditions such as the presence of calcium for perforin-1 C2 domain (57), these results suggest that perforin-2 may have little specificity at this stage and use a different regulatory mechanism. Free energy of binding of the P2 domain to membranes of different lipid concentrations could be calculated using potential of mean force (PMF) calculations or metadynamics to provide further details of membrane binding.

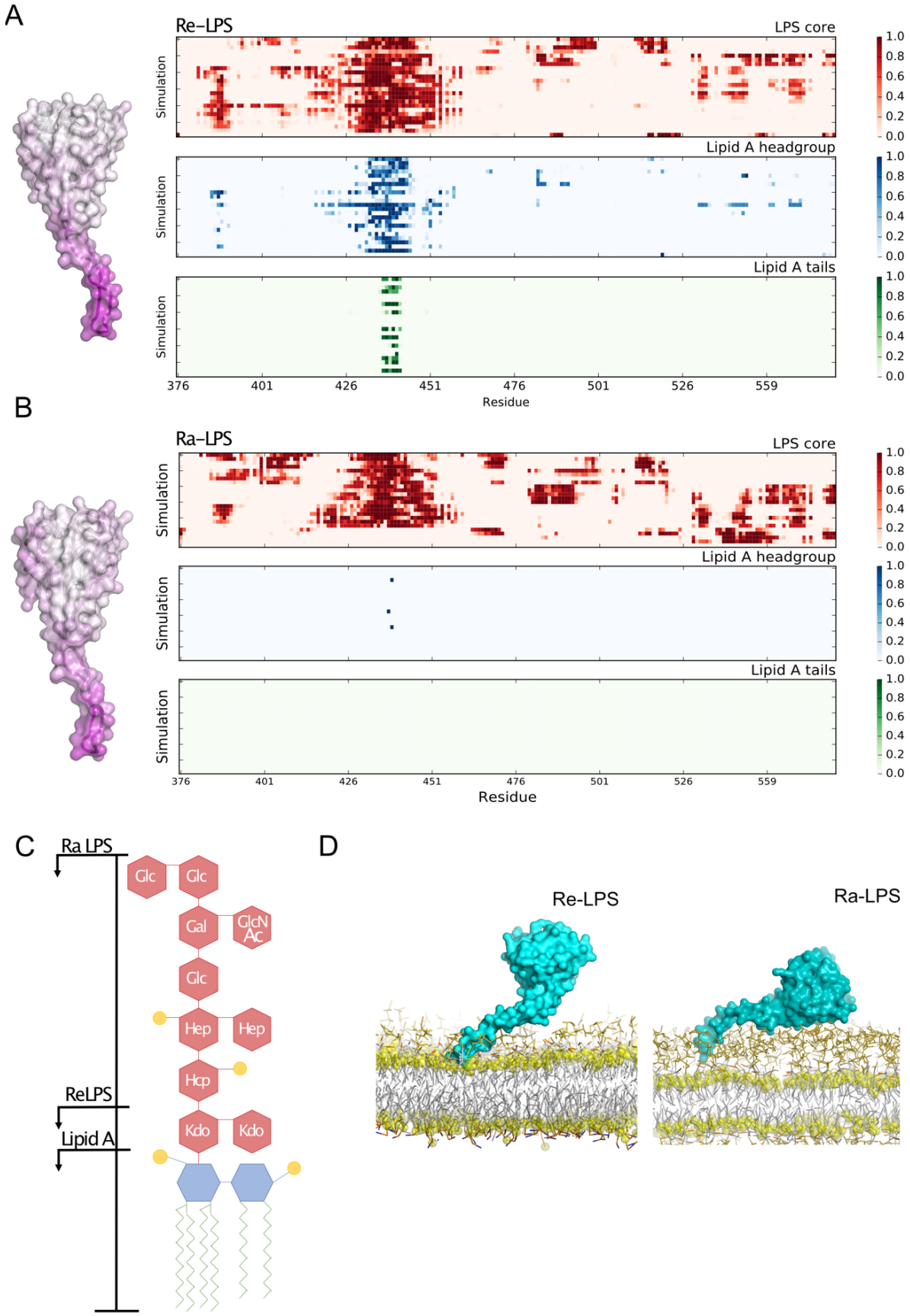

A
Re–LPS
LPS core
Lipid A headgroup
Lipid A tails
Simulation
376
401
426
451
476
501
526
559
Residue
1.0
0.8
0.6
0.4
0.2
0.0
B
Ra–LPS
LPS core
Lipid A headgroup
Lipid A tails
Simulation
Residue
C
Ra LPS
ReLPS
Lipid A
Glc
Glc
Gal
GlcN Ac
Glc
Hep
Hep
Hcp
Kdo
Kdo
D
Re-LPS
Ra-LPS

Figure 5.6 P2 domain contact with LPS

Lipid contact analysis of P2 binding to Re-LPS (A) and Ra-LPS (B) membranes. For each, averaged lipid contacts are plotted onto a surface representation of the P2 structure, coloured from white to purple to show frequency of contact. Normalised lipid contacts were calculated separately for the LPS core (red), Lipid A headgroup (blue) and lipid A tails (green) to observe burrowing of the β-hairpin into the membrane. (C) Schematic of LPS showing Ra-LPS and Re-LPS structures coloured consistently with lipid contact diagrams. (D) Representative frames from simulations showing the P2 domain β-hairpin burrowing between the LPS polysaccharide chains of Re-LPS and Ra-LPS. P2 is shown as a surface representation, lipids are shown as sticks, with lipid tails coloured grey.

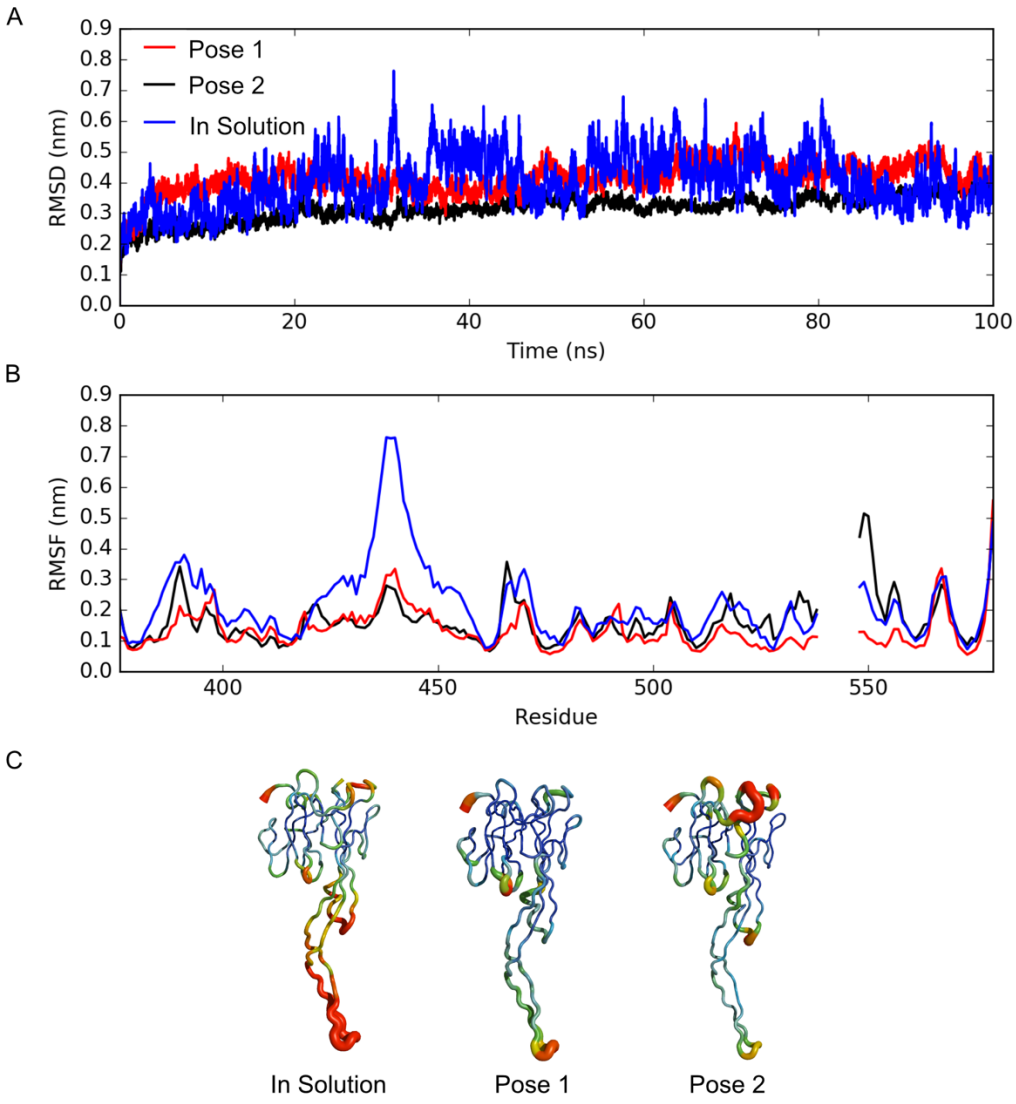


Figure 5.7 AT simulation of the P2 domain on PG:PE bilayers

RMSD (A) and RMSF (B) of P2 simulated in solution, and on PG:PE membranes in pose 1 and pose 2 for 100 ns each. (C) RMSF plotted onto the P2 structure coloured from blue to red.

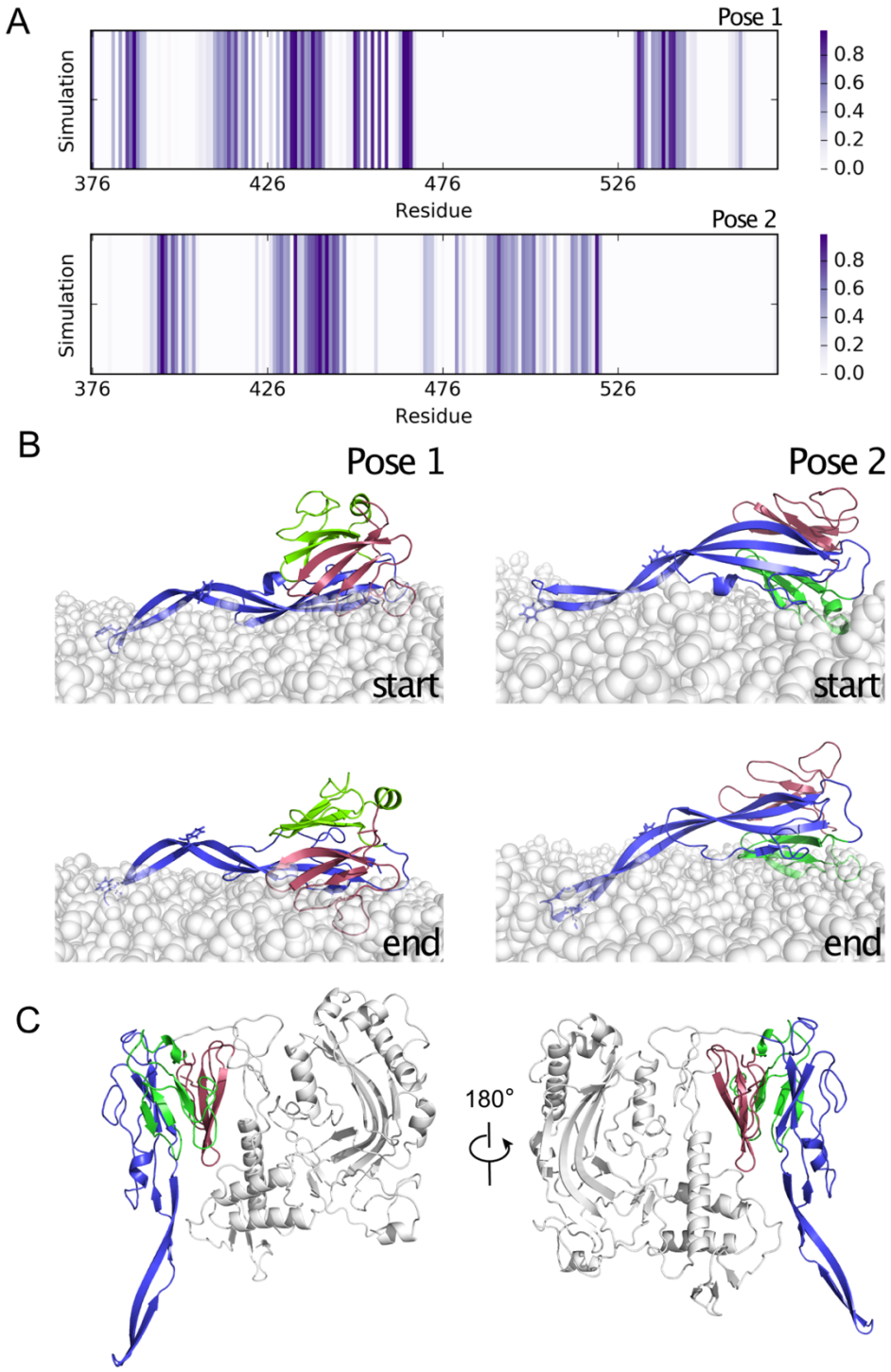


Figure 5.8 Membrane contact of P2 domain with PG:PE bilayers in AT simulations

Normalised protein-lipid contacts for the P2 domain simulated on PG:PE bilayers in pose 1 and pose 2. Three repeats are shown as three rows. (B) Start and end positions of the P2 domain PG:PE membranes after 100 ns simulation. (C) Ribbon representation of the PFN2 structure showing the MACPF-P2 domain interface. Each pseudo-repeat is coloured from blue (repeat 1), green (repeat 2) and maroon (repeat 3).

In many simulations, the core of the P2 domain also bound to membranes. The flexibility of the hairpin allowed binding of both the tip of the β-hairpin as well as the core of the domain in multiple orientations. While binding was more variable on LPS membranes, binding of the core of the domain could generally be grouped into two categories involving either pseudo-repeat 2 or 3 of the core structure.

These simulations are of the domain in isolation. This allows assessment of membrane binding of the domain while maintaining a smaller simulation system size, therefore reducing computational costs. As we also have the structure of the full PFN2 monomer, we can compare the surface accessibility of the P2 domain alone with that of the full monomer. While all pseudo-repeats of the P2 domain interact with the membrane in association simulations, observation of the full PFN2 structure shows that the third pseudo-repeat forms an interface with the MACPF domain and CTT (**Figure 5.8 C**). This suggests that, at least in the pre-pore and monomer, this surface of the P2 domain will not be accessible to the membrane, making binding orientations involving this surface, primarily pose 1 on PG:PE membranes, unlikely to occur.

Similar to previous simulations of the ApiPLP APCβ domains (**Chapters 3 & 4**) which have identified both the correct membrane binding site as well as a secondary orientation that is not thought to be accessible to the membrane in the full PLP protein, these simulations have identified membrane binding orientations that are unlikely to be physiological due to the presence of additional protein domains. Protein-protein and protein-lipid interfaces share some similarities in surface properties, namely surface exposure of hydrophobic residues with ionic and hydrogen bonding contributions (227-229). This region in P2 consists of hydrophobic and polar residues dispersed across the surface. These results suggest that both CG and atomistic simulations may not able to differentiate well between protein-lipid and protein-protein interfaces. This highlights a drawback of the membrane association method if simulating domains separately, especially those that form close contacts with other domains to form a reasonably compact protein such as PFN2. As the MACPF domain is oriented in the opposite direction to the P2 domain, further study was also required to identify if the sideways orientations observed would put the MACPF domain in an optimal position for membrane attack. In **section 5.2.3**, we address this further by investigating membrane binding of the PFN2 protein as a whole.

In addition, we saw that flexibility of the β-hairpin was required to allow simultaneous binding of the hairpin and domain core. While the β-hairpin was free to move in the P2 domain alone, in the pre-pore structure the β-hairpin forms a continuous β-sheet with the adjacent monomer, therefore reducing flexibility and thus accessibility to some of the orientations identified here.

Finally, we questioned the purpose of the unusually long extended β-hairpin. While the membrane binding motif of MACPF/CDCs typically forms a loop projecting from the base of the membrane binding domain (91, 188), the extended β-hairpin forms an unusual extension that is locked by two

disulphide bonds. In simulations on LPS, we saw that the tip of the hairpin was able to tunnel through the LPS polysaccharide layer and contact the lipid A at the bilayer. The LPS models used contain only the sugars from the core polysaccharide chain of LPS, and therefore the P2 domain would have a greater thickness of sugars from the O-antigen to tunnel into the Gram-negative bacterial surface. The basic nature of the hairpin is likely to aid in these interactions.

Despite this observation, the structure of the PFN2 pre-pore shows the tip of the extended β-hairpin is almost level with the CTT and MACPF domain. This position, along with orientation alongside the MACPF domain rather than underneath it which is typical for many MACPF/CDC proteins, suggests that at least in the pre-pore this extension may remain locked in place at the membrane surface, with only the hydrophobic tip of the hairpin able to anchor into the membrane. The opposite orientations of the MACPF domain and P2 domain may explain the evolution of this unusually extended β-hairpin, which must span the height of the MACPF domain in order to bind to the membrane. Transition to a transmembrane pore structure may free up these restraints resulting in a greater membrane accessibility and flexibility of the P2 domain, which would be interesting to investigate further.

McCormack *et al* found that Gram-positive bacteria could not be killed by lysozyme, which degrades the peptidoglycan cell wall, without PFN2 present (125). A similar mechanism of action has been suggested for MAC, which acts only on Gram-negative bacteria (230, 231). This suggests that PFN2 is also able to bind to the outer impermeable proteoglycan layer of the Gram-positive cell wall, causing perforations which then allow the passage of lysozyme and small molecules across the cell wall to rapidly kill the bacteria. This interaction has not been investigated here, however it further supports the requirement for a broad specificity of membrane binding and pore forming activity of PFN2.

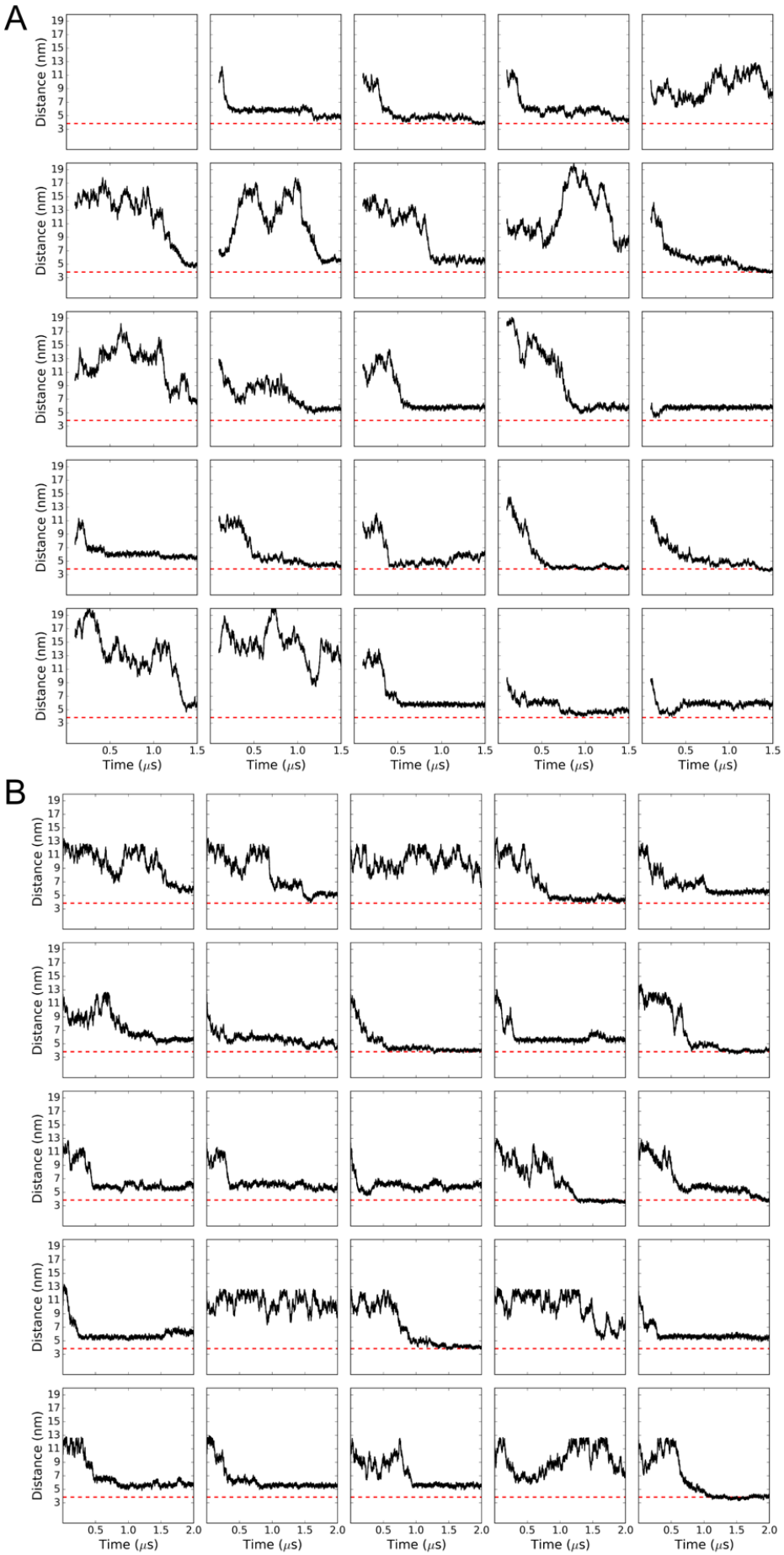

A
B
Distance (nm)
Time (μs)

Figure 5.9 Perforin-2 monomer binding to PG:PE bilayers

Membrane association simulations of PFN2 monomer original structure (A) and newer more complete structure (B). Distances are plotted for each simulation between protein and membrane centre of mass.

5.2.3 Monomer membrane binding

After assessing the membrane binding orientation of the P2 domain alone, we next looked at membrane binding of the full PFN2 monomer. Given the opposite orientations of the MACPF domain and P2 extended β-hairpin, we aimed to confirm that the P2 domain is responsible for membrane binding and assess monomer membrane association before the oligomerisation step.

Two structures of the monomer were solved, firstly a structure containing the MACPF domain, P2 domain missing the tip of the extended β-hairpin, and a partial C-terminal tail. The second structure was solved part way through this study, containing the full P2 β-hairpin and a short α-helix at the C-terminus of the CTT. As the more complete structure was only available part way through this study, we primarily simulated the first PFN2 monomer structure. Modeller (186) was used to combine the P2 domain crystal structure with the PFN2 monomer to complete the β-hairpin, and as with the P2 domain alone, connect the free ends of the missing segment between residues 539 and 548 in P2.

Membrane association simulations were first performed with this initial PFN2 monomer, allowing PFN2 to bind to PG:PE membranes mimicking the bacterial inner membrane. 25 simulations were produced for 1.5 μs each, rotating the protein sequentially around its X, Y and Z axis as before. Out of 25 simulations, 22 bound to the membrane (**Figure 5.9 A**). Of those that bound, 80% bound to the membrane with the extended P2 β-hairpin. In those that did not bind via the P2 domain, PFN2 bound to the membrane using the MACPF domain.

Lipid contact analysis was performed to further assess PFN2 binding (**Figure 5.10)**. As was visually observed, PFN2 bound to the membrane via the P2 domain β-hairpin in most simulations. In some simulations, the C-terminal tail binds along with the β-hairpin, together forming a flat platform upon which the MACPF domain sits and is able to retain an upright orientation (**Figure 5.10 C**). As the β-hairpin is flexible, it is able to curve to extend the surface of binding while maintaining the MACPF domain's vertical orientation. In a few cases, binding was also observed in which the P2 domain bound to the membrane using the β-hairpin, and the entire PFN2 structure tilted onto its side with the core of the P2 domain also in contact with the membrane. This pattern of binding matches that seen in pose 2 of the P2 domain on PG:PE bilayers (**Figure 5.5**).

Once it was solved, the membrane association simulations were then repeated using the new PFN2 structure to understand the role of the additional C-terminal helix. Simulations were run for 2 μs in this

case to allow extra time for membrane binding. After the 2 μs, PFN2 remained unbound in three simulations (**Figure 5.9 B**). Similar to without the helix, PFN2 bound to the membrane by the P2 β-hairpin in 82% of bound simulations. Protein-lipid contact analysis shows a more variable pattern of binding (**Figure 5.10**). The CTT bound to the membrane more frequently in the new structure than the previous structure, suggesting that the additional helix contributes to membrane binding. The helix contains two arginine residues (R628 & R629), along with a threonine (T632) and several hydrophobic residues L625, L627 and M631 directed towards the bilayer. In 12 simulations, the P2 β-hairpin and CTT helix bind to the membrane forming a stable foundation of contact with the membrane (**Figure 5.10 D**), similar to without the extra CTT helix. However, the extended length of the CTT in the new structure maintains the P2 domain in a straighter position so only the tip of the extended β-hairpin is in contact with the membrane. This also rotates the MACPF domain around so that the N-terminus at the top of the domain also contacts the membrane.

While the MACPF domain retains a vertical profile as would be expected during oligomerisation, in some simulations the protein turns on its side with the flat edge of the MACPF domain sitting on the membrane. These observations suggest that while the P2 domain and CTT often form a stable platform upon which PFN2 initially associates with the membrane, oligomerisation is required to maintain a functional upright orientation. As will be further discussed later in this chapter, the MACPF must be in an orientation in which it can rotate around the P2 domain into its final pore forming orientation.

5.2.4 Pre-pore membrane binding

We next turned our attention to membrane binding of the pre-pore structure as a whole. While simulations and structural characteristics of both the P2 domain alone and PFN2 monomer strongly suggested that PFN2 binds to membranes using the extended β-hairpin of the P2 domain, the antiparallel orientation of the P2 and MACPF domains still posed questions as to how the protein orients itself on the membrane to allow the MACPF TMH segments to deploy into the membrane for pore formation. Thus, we tested membrane binding of both the MACPF/TMH exposed face and the P2 β-hairpin exposed face of the first pre-pore structure.

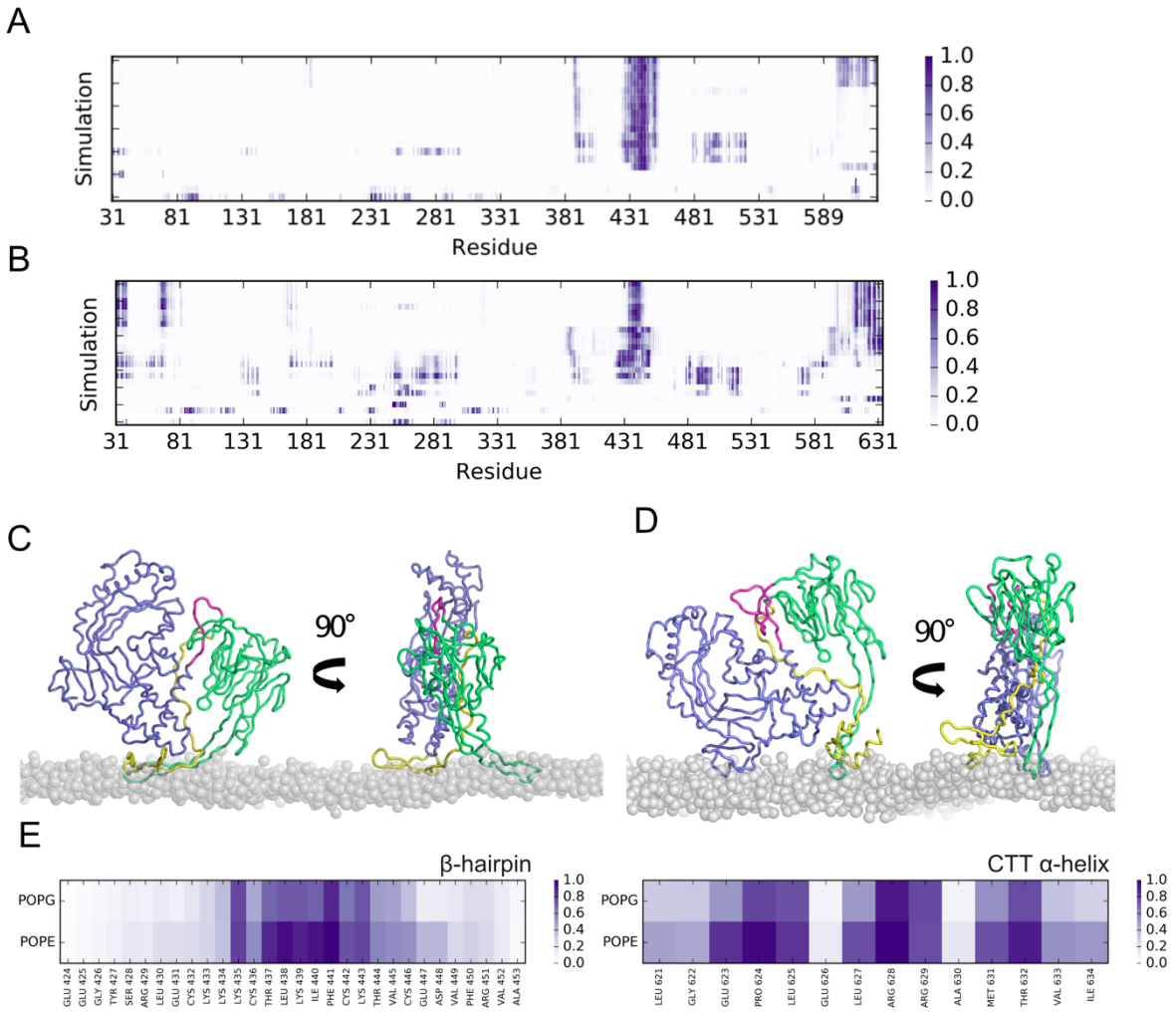


Figure 5.10 Lipid contacts and membrane binding of PFN2 monomers

Normalised protein-lipid contacts for original PFN2 structure (A) and more complete structure (B) from membrane association simulations on PG:PE bilayers. PFN2 binding orientation with P2 hairpin bound to the membrane in original structure (C) and more complete structure (D). Lipid head groups are shown as white spheres. The MACPF domain is coloured blue, linker is shown in magenta, P2 domain in green and the CTT is coloured yellow. Lipid contacts for the P2 β-hairpin tip and CTT α-helix are shown averaged across all 25 simulations showing detailed residue contacts.

Simulations were performed with the PFN2 pre-pore structure that was solved first, and Modeller was used to model in the missing tip of the β-hairpin and connect residues at either side of the missing segment (540-547) of the P2 domain. The crystal structure of the P2 domain was used as a template to model in the extended hairpin of the P2 structure. As the density of this region was missing in the structure and the extended hairpin is flexible, position of the hairpin had to be estimated based on best predicted fit into the pre-pore oligomer.

The model was then placed 10 nm away from a PG:PE bilayer in either a MACPF-biased or P2-biased orientation (**Figure 5.11**). Ten repeat simulations were run for 1 μs for each orientation, and the resulting binding was assessed visually and using lipid-contact analysis to compare to monomer binding.

In the MACPF-biased simulations, the pre-pore bound to the membrane in 7 out of the 10 simulations (**Figure 5.11**). Additionally, in one simulation the structure crossed the periodic boundary of the simulation box and began to bind via the P2 domain. While the top of the MACPF domain did bind to the membrane, the surface area of contact was small, and binding did not occur consistently across the simulations. In contrast, when oriented with the P2 domain facing the bilayer, the pre-pore quickly docked onto the membrane in every simulation, and remained in a consistent position on the bilayer (**Figure 5.11 C&D**).

Protein-lipid contacts were also calculated, averaging the contacts of each subunit in the pre-pore to show the contacts for each simulation plotted onto one PFN2 monomer sequence (**Figure 5.11 E & F**). Consistent with distance calculations, the lipid contacts in the P2 biased simulations were almost identical across the 10 repeats, with the extended β-hairpin of the P2 domain and the CTT making contact with the membrane. In comparison with simulations of the monomer alone, binding was restricted to these regions with only a few residues in the MACPF domain in contact with the membrane, as oligomerisation restrained the position of the protein. The MACPF-biased simulations showed some variation in binding, however binding primarily occurred via the TMH1 helices of the MACPF.

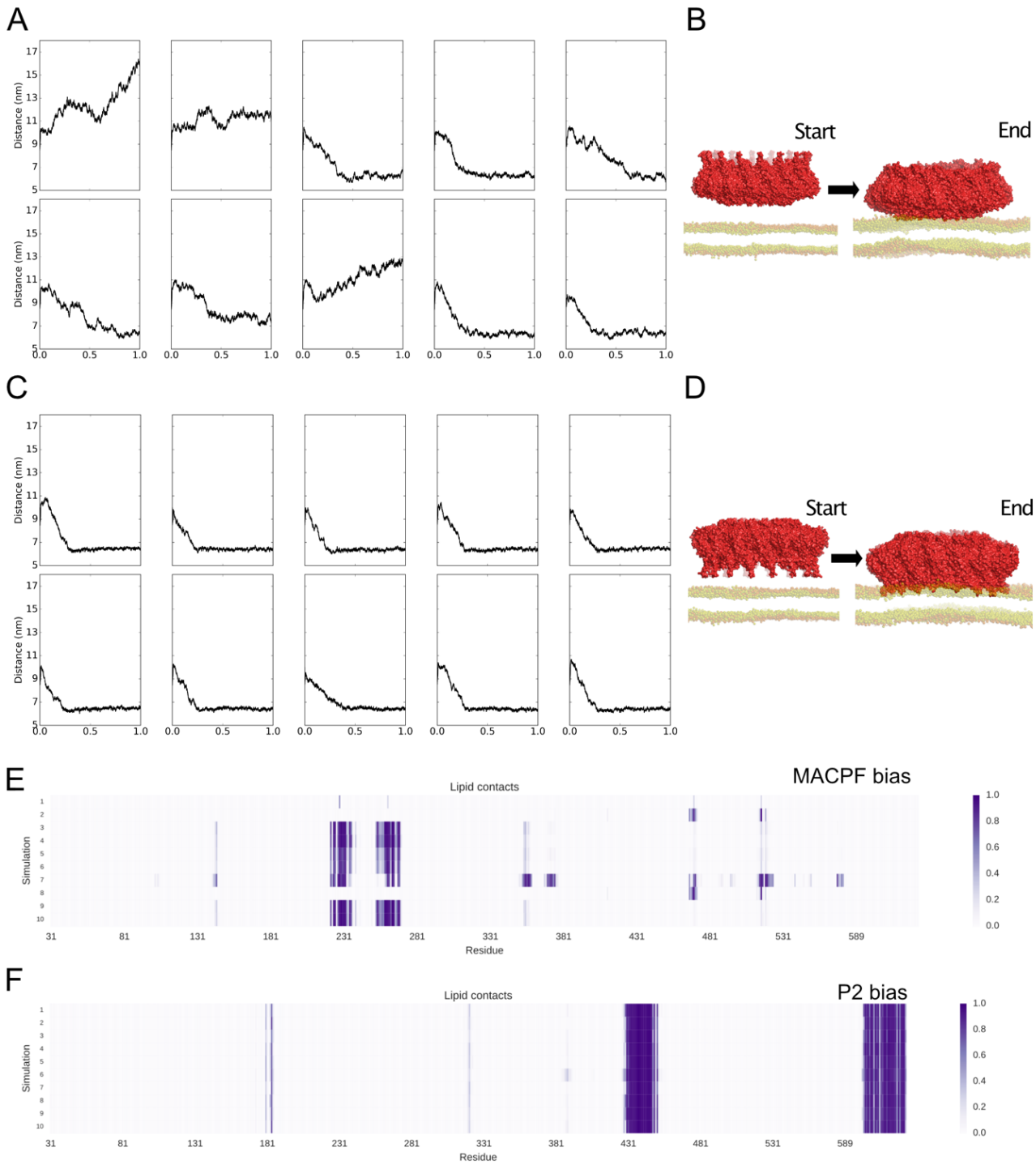


Figure 5.11 Pre-pore membrane binding

Protein to membrane distances measured for each of 10 repeat pre-pore membrane association simulations from both MACPF biased (A) and P2 biased (B) starting positions. Starting positions along with representative bound positions are shown for the MACPF biased (B) and P2 biased (D) simulations. Normalised lipid contacts were averaged across all 16 subunits for each simulation for the MACPF biased (E) and P2 biased (F) simulations.

PFN2 pre-pore arcs

As our studies revealed incomplete PFN2 arc pre-pores/pores, we also produced an arc structure by removing 8 of the 16 subunits in the pre-pore. Association simulations were set up as described above for the P2 biased pre-pore membrane binding, resulting in a model of a PFN2 arc bound to the membrane in a pre-pore state, which could be further simulated to understand the dynamics of PFN2 on the membrane, and investigate flexibility in the structure that could lead to the transition to a pore state. Importantly, the binding pose found for the arc was the same as that observed for the full pre-pore ring.

5.2.5 PFN2 membrane binding discussion

Through performing membrane association simulations of PFN2 monomer, we were able to investigate how the subunits may associate with membranes prior to oligomerisation. In almost all cases, PFN2 made contact with the membrane using the P2 β-hairpin. Binding was complemented with the CTT terminal helix, inclusion of which increased incidences of stable binding. In the stably bound pose, the axis of the MACPF domain was perpendicular to the membrane, with TMH segments oriented away from the membrane. The oligomerisation surface of the MACPF domain remained exposed upon binding in many simulations, putting it in an optimal position for oligomerisation. Liposome-binding assays performed in the lab in which a CTT truncation was produced showed similar frequency of binding with and without the CTT (manuscript under review). Combining these data with simulation data, this suggests that the CTT may have more of a role in binding stabilisation for efficient pore formation than membrane binding itself.

Binding of the P2 β-hairpin was similar to that seen with the P2 domain alone, with greatest lipid contact at the hydrophobic phenylalanine tipped β-hairpin. Compared to simulations with the P2 domain alone, the MACPF domain restricted contact of the P2 core which had significantly less contact with the membrane.

As very little binding was seen ‘upside down’, with the MACPF TMH regions facing towards the membrane, these simulations support the theory that PFN2 binds to membrane using P2 hairpin, and therefore the MACPF domain must rotate itself around the P2 domain to orient the TMH hairpins towards the bilayer. In light of this large reorientation of the protein, we next performed MD simulations to observe the dynamics between the CTT, P2 domain, MACPF domain.

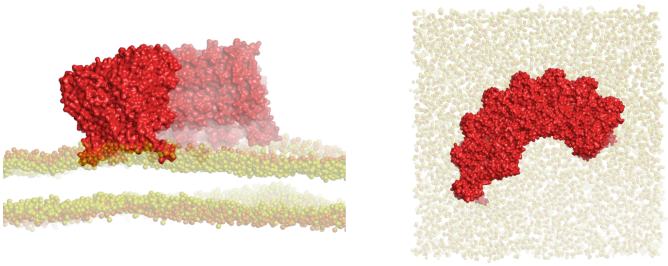


Figure 5.12 A model of the PFN2 pre-pore bound to the membrane

Representative final bound position of the PFN2 pre-pore 8-mer arc.

Association of the pre-pore continued to confirm conclusions drawn from monomeric simulations, with the speed and consistency of pre-pore binding in the P2 biased orientation suggesting that this is the correct orientation for binding. Lipid contacts matched those seen for the PFN2 monomer but restricted the contact to just the base of the protein as each subunit is constrained in the pre-pore and not able to tilt sideways on the membrane. These simulations also provided a membrane bound structure for further simulation and analysis.

Finally, a model of an incomplete pre-pore arc was also produced and docked with the membrane using membrane association simulations. This bound as expected based on previous simulations, and along with the full closed pre-pore, provided structures of the PFN2 pre-pore bound to the membrane for further study.

5.2.6 Investigating the mechanism of pre-pore to pore transition

Due to the opposite orientations of the P2 and MACPF domains, the MACPF domain is initially facing into solution upon initial membrane docking, and unable to reach the membrane for pore formation. The pre-pore and pore assembly was investigated by our collaborators, who found that while PFN2 pre-pores often formed complete rings on membranes, pores mostly were formed by incomplete arcs which increased in height in comparison to the pre-pores (manuscript under review). This height increase was due to a domain re-organisation, with the MACPF domain rotating 180° around the P2 domain, resulting in orientation of the TMH helices towards the membrane for pore formation. The breaking up of complete pre-pores into arcs is presumably required to provide space for the MACPF domain rotation.

To understand the molecular mechanisms that facilitate this large-scale rotation, we analysed the interfaces between the P2 domain, MACPF domain and CTT. The CTT and the MACPF domain appear to have a more stable interaction with hydrogen bonds found along the length of the CTT folded around the MACPF domain. Several hydrophobic pockets were identified at the domain interfaces, one between the P2 core, CTT and MACPF domain, and between the

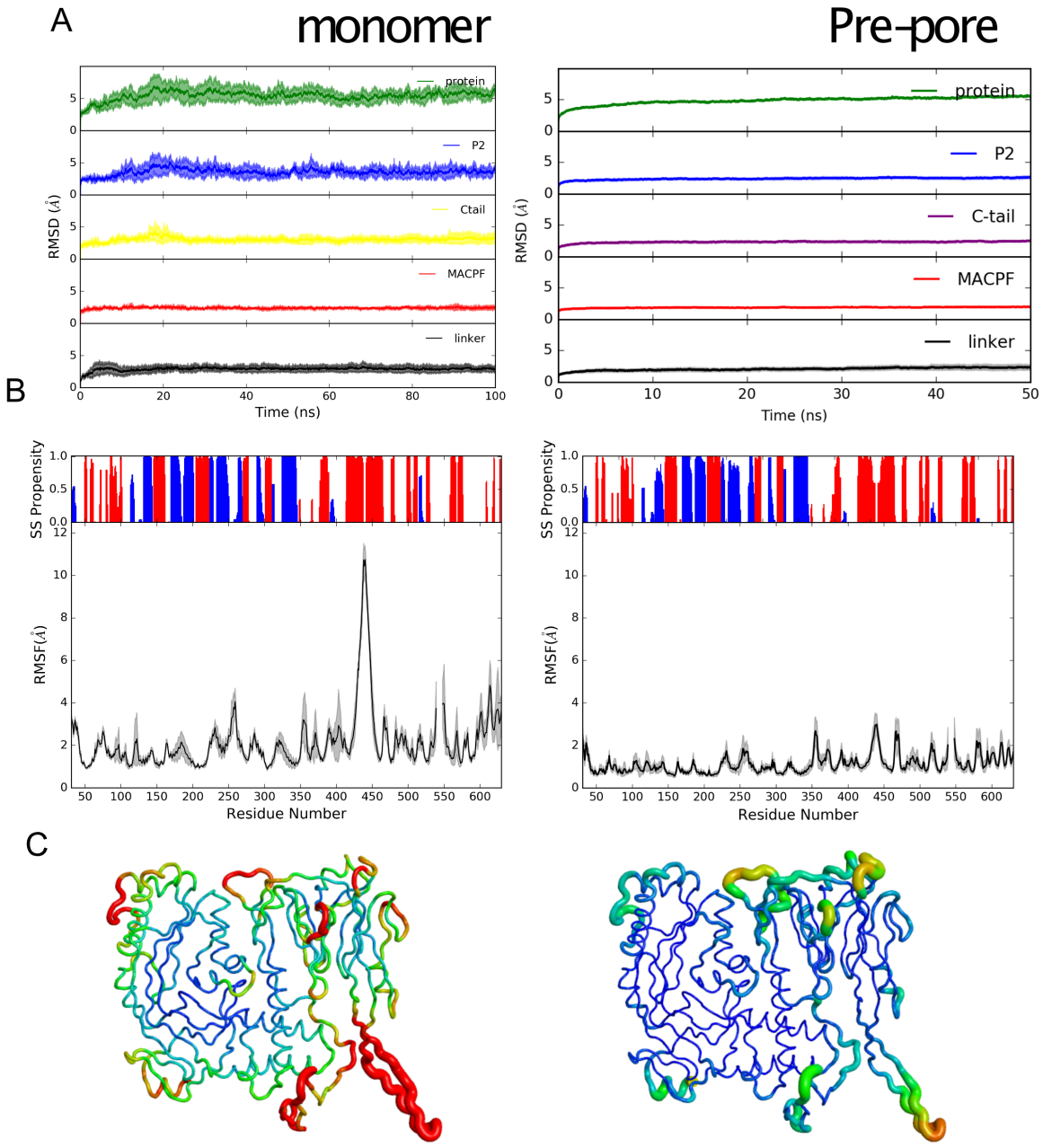


Figure 5.13 Stability of PFN2 monomer and pre-pore

The PFN2 monomer (left panel) and pre-pore (right panel) were simulated atomistically. RMSD (A), RMSF and secondary structure propensity (B) were calculated for each. The mean value was averaged from 3 repeats of the monomer, while each subunit was averaged together for the pre-pore. Standard deviations are shown. (C) RMSF values are plotted onto cartoon representations of the structure, coloured from blue to red to show regions of greater fluctuation across simulations.

MACPF-P2 linker, MACPF domain and CTT. Only one hydrogen bond was found between the P2 domain and MACPF domain.

Atomistic simulations of the original monomer and full pre-pore structures were performed to assess dynamics of the PFN2 domain interfaces. The monomer was simulated in water for 100 ns, with three replicates. The pre-pore was also simulated in water for only one repeat due to the large system size, however results could be averaged across the 16 subunits in the ring. Overall stability of the systems was observed using RMSD and RMSF calculations. Average RMSD values for each domain of PFN2 were calculated individually, and were similar in both the monomer and pre-pore, while standard deviations in the pre-pore were lower, presumably due to oligomerisation conferring greater stability to the structure (**Figure 5.13 A**). RMSF values in the pre-pore were also lower than those for the monomer alone, again showing that oligomerisation stabilises the structure. RMSF at the P2 extended β-hairpin and CTT were significantly reduced in particular, due to formation of a stable β-sheet between the P2 domain and CTT of the adjacent subunit (**Figure 5.13 B&C**).

Next, distances between the P2 domain and MACPF domain, and between the CTT β-sheet and MACPF domain were calculated across each simulation (**Figure 5.14**). Distance between K188 of the MACPF domain and D567 of the P2 domain reflected the motion of the P2 domain core away from the MACPF domain, while F450-N185 reflected the motion of the P2 β-hairpin away from the MACPF domain. The F450-N185 distance showed greatest fluctuation is simulation of the monomer, due to the flexible nature of the hairpin. Distance between K188 and D567 remained more stable, however in repeat one of the monomeric simulations we observed motion of P2 away from the MACPF domain reflected in both distance measurements. The contact between the CTT β-sheet and the top of the MACPF domain is also destabilised by this motion, suggesting that the CTT may also contribute to this domain interface (**Figure 5.14 B right panel**). The number of hydrogen bonds between each domain was also reduced during this time period (**Figure 5.14 C**). One significant hydrogen bond was identified between the P2 and MACPF domains, between residues K188 and D567, which broke when the domains rotated away from each other in MD1. In contrast to the monomer, the pre-pore structure remained stable throughout the simulation showing little motion between the domain (**Figure 5.14**). This highlights the way in which the pre-pore is assembled to maintain stability until appropriately triggered for pore formation.

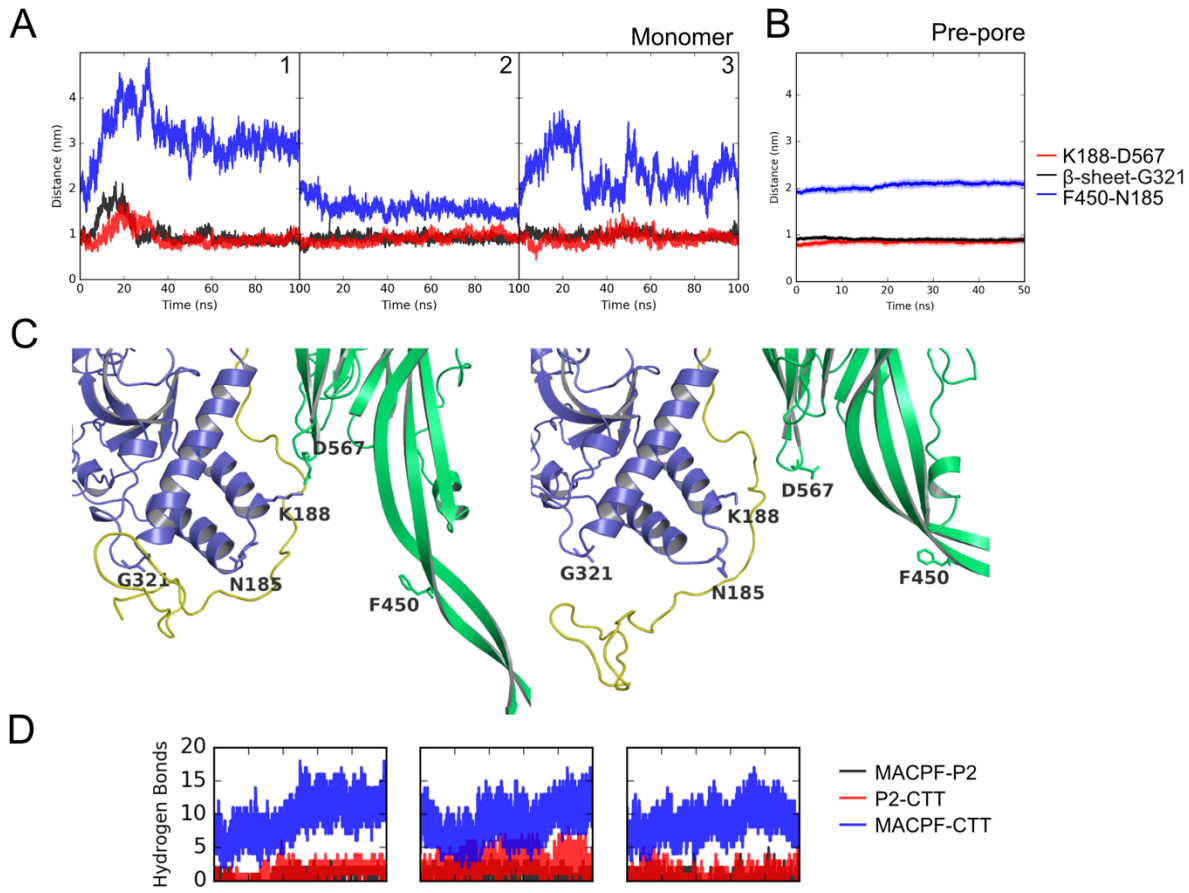


Figure 5.14 PFN2 domain interfaces

(A) Inter-domain distances calculated for each of three repeated PFN2 monomer simulations. (B) The same inter-domain distances calculated for each subunit of the PFN2 pre-pore structure. Mean distances were calculated from all 16 subunits of the pre-pore and standard deviations are shown. (C) Residues used for inter-domain distance measurements are labelled on the PFN2 structure at the start of the simulation, and at 20 ns into the first repeat of the monomer simulation, showing motion of the P2 domain away from the MACPF domain. (D) Number of hydrogen bonds between each domain calculated for each PFN2 monomer simulation repeat.

5.2.7 Discussion and further work to investigate the pre-pore to pore transition

Through simulating the monomeric PFN2 structure in solution, one destabilisation event occurred between the MACPF and P2 domains which may reflect a less stable interface between these domains that may break apart during MACPF domain rotation and pore formation. We found greater numbers of hydrogen bonds between the CTT and MACPF domain, however the C-terminal β-sheet was able to move away from the MACPF domain during destabilisation. In contrast, interaction between the P2 domain and MACPF is limited to a single significant hydrogen bond, and a hydrophobic pocket shared between the core of the P2 domain, CTT and MACPF domain. Observation of the structure shows that the P2 domain also has a lower surface area of contact with the MACPF and CTT. Together, these

results indicate that the CTT and MACPF domain form a more stable connection and may rotate together around the P2 domain which remains bound to the membrane.

These simulations represent the first steps in investigating PFN2 structural reorganisation during pore formation. Several factors govern PFN2 activation, including low pH, membrane binding and oligomerisation. These simulations were performed at neutral pH, at which perforin-2 is not active. PFN2 functions at an acidic pH in the phagosome (232), so a further set of simulations could be performed at a lower pH. Histidines are known to be a factor controlling conformational changes, particularly in viral proteins, by either disrupting hydrophobic interactions or forming charged interactions (233, 234). Several histidines, H340 H347 and H561, were identified in the structure at the hydrophobic interface between the domains. These could possibly play a role in triggering structural rearrangement at low pH through ionisation which could trigger disruption of hydrophobic interactions.

We simulated PFN2 as a monomer and full pre-pore. We have shown with atomic force microscopy (AFM) that on transition to a pore structure PFN2 rings break up into incomplete arcs of subunits, due to the need to provide space for the MACPF rotation (manuscript under review). Therefore, simulations of PFN2 as both a dimer and incomplete arc would reveal the impact of oligomerisation on domain stability. The membrane association simulations of PFN2 monomers and pre-pore arcs could provide starting structures which could be used for these analyses.

5.3 Conclusions

Through taking a multi-scaled approach to studying membrane binding of perforin-2, we have identified that PFN2 binds to membranes with its unique P2 domain. Separate experimental data have confirmed this is the case (manuscript under review). The hydrophobic tip of the P2 extended β-hairpin, containing a phenylalanine residue at its extreme end, binds to membranes in a non-specific manner. Further along from the tip, the surface of the β-hairpin is positively charged, leading perforin-2 to preferentially bind negatively charged lipids. While the PFN2 had a preference for negatively charged bilayers, it did not appear to have significant lipid specificity, and bound to LPS bilayers as well as bacterial inner membrane and eukaryotic membranes. Simulations of PFN2 monomer membrane binding showed that the α-helix of the CTT also contributed to stable membrane binding.

In contrast to the Complement system, which kills only Gram-negative bacteria, PFN2 has broad spectrum antibacterial properties. We have shown that it is able to bind to LPS present on the surface of Gram-negative bacteria, where the long β-hairpin may function to anchor the protein into the polysaccharide chains of LPS to enable pore formation. Questions remain as to how PFN2 may act to kill Gram-positive bacteria, which are protected by a thick proteoglycan layer in the cell wall.

The discovery that the membrane binding and MACPF domains of perforin-2 are located in opposite orientations posed a challenge in understanding how pore formation is achieved. Through experimental evidence, our hypothesis is that the MACPF domain rotates 180° around the P2 domain to orient towards the membranes, a motion that requires pores to form from arcs rather than closed circles. We used molecular dynamics to predict the molecular mechanisms by which this rotation may occur. We have observed that the interaction between the MACPF and P2 domain is relatively weak, with only a loose association between the two domains, whereas the CTT is more tightly folded around the MACPF domain. Further study into the effects of both membrane binding, oligomerisation and pH on the interface stability would be interesting to understand this transition further. It would also be interesting to observe how membrane interaction of both the P2 domain and CTT helix change throughout this transition.

Overall however, these studies have helped to confirm that the perforin-2 pore forming mechanism is unique among the MACPF/CDC family in that the MACPF domain is oriented away from the membrane upon binding and must undergo significant conformational change to reorient and form membrane inserted arcs.

Chapter 6

Conclusions

6.1 Summary

The work presented in this thesis aimed to use molecular dynamics and structural biology to investigate the membrane binding and mechanisms of action of two groups of MACPF/CDC proteins; the apicomplexan PLPs and perforin-2. A serial multi-scaled approach to studying the membrane interactions of these MACPF proteins membrane binding domains proved effective at investigating several novel membrane binding domains from initial mode of membrane interaction to the molecular detail of membrane recognition and lipid affinities. These methods combined well with structural and biophysical approaches to develop an understanding of the mechanisms of action of TgPLP1 and perforin-2, as well as beginning to investigate the molecular basis of PPLP function.

In chapter 3, MD was used to study TgPLP1 at multiple stages of the pore forming process. Firstly, membrane interaction studies of the APCβ domain revealed its membrane binding orientation and identified that this interaction is primarily mediated by a hydrophobic tryptophan-tipped loop that anchors the protein in the bilayer. Simulations of several crystal structures of the oligomeric MACPF domain found that the interfaces in the helical structure were more stable and are likely to be the interface present in transmembrane pores, while a novel method for modelling MACPF transmembrane beta-barrels was developed to predict the transmembrane pore β-barrel structure of TgPLP1 based on its oligomerisation state and TMH sequences. Together these studies provide an overall picture of ways in which TgPLP1 interacts with the membrane from initial docking through to pore formation.

In chapter 4, the study of ApiPLPs was extended into the *Plasmodium* PLPs, with the aim of investigating the membrane binding domains of this family. We solved the first crystal structure of a *Plasmodium* PLP membrane binding domain, and the high structural similarity to TgPLP1 APCβ supported the theory that the APCβ domain is unique to ApiPLPs. Simulation studies were then performed to investigate membrane binding interactions and lipid affinities, showing that while the overall structure is similar across the ApiPLP family, PvPLP1 and PvPLP2 APCβ domains behaved in distinct ways and no membrane binding for seen for PvPLP1. This is a key observation, showing that PPLPs have different membrane recognition mechanisms across the family and are therefore likely to have distinct differences in function depending on the point in the *Plasmodium* lifecycle at which they are expressed.

In chapter 5 these investigative methods were extended to study perforin-2, a recently discovered MACPF/CDC protein with important bactericidal properties in the human immune system. Another unique protein fold was discovered in the crystal structure of the PFN2 P2 domain, containing a highly extended beta-hairpin, while the monomer as a whole contains the MACPF and P2 domains in an antiparallel arrangement facing in opposite directions. Membrane association simulation studies of this domain, along with experimental evidence, showed that the P2 beta-hairpin is responsible for membrane

binding of the domain, while simulations with a variety of bilayers showed that this domain has low lipid specificity and can associate with both bacterial and eukaryotic membranes.

Further simulations of the PFN2 monomer and pre-pore confirmed this method of membrane interactions, suggesting that the MACPF domain faces away from the membrane upon initial docking. We proposed that the MACPF domain undergoes a rotation around the P2 domain during pore formation to orient itself towards the membrane before puncturing the membrane. This represents a novel mechanism of pore formation which has not been seen in other members of the MACPF/CDC family.

6.2 Future Directions

With the work on TgPLP1 in this thesis, we were able to identify the membrane binding interactions of the APCβ domain, and oligomerisation states of the MACPF domain. The model of the transmembrane beta-barrel showed promising lipid interactions when inserted into the membrane, suggesting that the residue orientations and hydrogen bonding are favourable. Together these results provide a basis for producing a model of the complete TgPLP1 pore structure in a membrane.

A key question remaining around TgPLP1 is its precise roles in facilitating parasite escape. One proposed mechanism involves the formation of large patches of destabilised membrane that, along with the action of other effector molecules, would be sufficiently weakened to allow mechanical disruption of the membrane and escape. Another theory suggests that the pores provide a conduit for which further secreted effector molecules including lipases and proteases could be released from the PV and degrade cell structures to facilitate escape.

The hexameric crystal structures of the TgPLP1 MACPF domain suggest that TgPLP1 forms small, discrete pores that would allow passage of small molecules and ions through. Once a model of the TgPLP1 pore is produced, this could be investigated further using computational electrophysiology to investigate the flow of ions through the pore to determine what size and characteristics of molecules are able to pass through the pore. While the crystal structures of TgPLP1 reported here were hexameric in nature, it is possible that, similar to many other MACPF/CDC proteins, TgPLP1 pores may exist in a range of oligomeric forms. Further cryo-EM and AFM studies of this protein could be used to investigate the range of pore sizes formed by TgPLP1, which would help inform the mechanism by which it may act.

Alternatively, the role of TgPLP1 could be to weaken the membrane, allowing further disruption by parasite motility. The observed preference for formation of a helical assembly may support this hypothesis. CG simulations of a helical TgPLP1 model in the membrane over long timescales could be used to determine if the protein induces membrane distortion, thinning or curvature both during the process of membrane insertion and once fully inserted, and therefore whether or not this mechanism of egress is likely.

Stepping back to investigate the fundamental biophysics of MACPF/CDC membrane insertion, there has been little investigation into the energetic costs involved in insertion of beta-hairpins into the membrane, as well as the molecular details of insertion of pores into the membranes from a lipid point of view. As a technique, molecular dynamics is well placed to investigate some of these questions that are difficult to unpick using other experimental methods.

It would be interesting to use the TgPLP1 beta-barrel model and pre-pore structures to investigate this. Steered MD and PMF calculations could be used to investigate the energetics of membrane insertion of both full, closed, pores and β-barrels as well as partial structures representing incomplete arcs and single monomers. This would also pose a mechanism of investigating the rearrangement of lipids during membrane insertion as well as how the membrane may respond to insertion of incomplete arcs, both questions that remain in the field (72, 204). A recently published paper uses similar methods in MD simulations to investigate pneumolysin pore formation (207), in which insertion of a transmembrane β-sheet formed from several pneumolysin oligomers into a membrane caused lipids to move away from the β-sheet and form a toroidal membrane edge, with water filling the space in-between. They also tested insertion of a full penumolysin pore, at several stages of pre-pore to membrane insertion, and found that lipid was expelled as a vesicle from the centre of the pore. Whether or not MACPF/CDC proteins are able to form incomplete arcs as well as full pores has been a contentious issue in the field, and while there is now general agreement that arcs are able to form functional pores in addition to full rings of subunits (72). The study of the TgPLP1 membrane insertion could continue to contribute to this debate by simulating both complete pores and incomplete arcs of the structure in a membrane. Pulling of the structure into the membrane could be performed to investigate how lipids may rearrange during membrane insertion, and determine if arc structures are stable in a membrane.

This work has investigated the APCβ membrane binding domain of *Plasmodium* PLPs, showing that there are important differences in membrane binding between members of the family, presumably due to differences in location and function. This provides a strong argument for further exploration of PPLP function, in particular in understanding the differences between each PPLP at the point of membrane binding and activation of pore formation. Knockout studies have shown that PPLPs are important for malarial transmission meaning that these proteins could prove an effective target for vaccination trials or therapy (110, 116, 210, 211).

As with TgPLP1, the mechanism by which these proteins facilitate *Plasmodium* egress and traversal is unknown. PPLP1 did not bind to membranes here, but i*n vitro* work suggests that PLP1 is responsible for membrane disruption (98, 111). Lack of lipid binding suggests that PLP1 may recognise a specific protein receptor in the membrane. A proteomics approach would be useful to identify proteins that interact with PvPLP1. After exposing PvPLP1 to PV membranes, a pull-down assay could be used,

followed by analysis of binding partners using mass spectrometry. Once a binding partner is identified, further investigation into the mechanism of pore formation by PvPLP1 could be carried out.

Expression of the PPLP1 MACPF domain and N-terminal domain along with the already solved APCβ domain would provide a foundation for further study, which could then be extended to other PPLPs. It has been proposed that PLP2 perforates membranes, allowing proteases and other effector molecules to pass out of the PV into the cell cytoplasm, further damaging the cell and promoting egress (116). This suggests that these proteins may form discrete pores that allow small effector molecules to flow through. Electron microscopy could be used to visualise pores formed by PvPLP1, to determine whether the size of the pores would be sufficient to allow traversal of molecules, as well as determine if PvPLP1 forms a helical asymmetric assembly or canonical circular pore shape. These structures could then be used for further simulation studies into membrane binding of monomers, pre-pores and pores.

It would also be interesting to investigate PPLPs 3, 4 and 5. These are expressed at the same point as each other in the *Plasmodium* lifecycle and it has been proposed that they may act cooperatively to form heterooligomeric pores. The use of lysis assays with each protein alone and in combination could be used to determine if these proteins are indeed able to form hetero-oligomers that may function cooperatively, or if each protein functions independently.

Additionally, the functions of the N-terminal domain of the PLPs remains unknown. One study showed that the TgPLP1 NTD has membrane binding properties, however removal of the domain does not appear to prevent membrane attack or binding of the MACPF-APCβ (182). As the lengths and sequence of the NTD vary considerably across the PLP family, it is possible that these domains perform diverse functions across the family.

Finally, the perforin-2 work described here describes the initial steps of membrane binding of the protein. Since this work was done, a cryo-EM structure of the full PFN2 pore was solved in the Gilbert group, providing the final pore structure. To investigate how the pore structure interacts with surrounding lipids, a series of simulations could be performed to build a lipid bilayer back around this pore structure to visualise the pore in a native membrane-inserted configuration. Simulations could then be performed to calculate membrane thickness, density and disorder around the pore to determine how insertion of perforin-2 into the membrane affects the stability of the surrounding membrane, and whether this may contribute energetically to pore formation. These simulations could also be used to further investigate lipid preferences of the protein. Our work thus far suggests that perforin-2 has a broad preference for negatively charged lipids, and simulations spanning multiple microseconds could be performed to investigate lipids interactions and clustering around the oligomer to determine if any further lipid preferences can be identified.

Our work showed that perforin-2 pore formation is triggered by low pH. This may trigger a destabilisation of the interface between the P2 domain and MACPF which would then allow for rearrangement of the domains. Work to continue investigating the mechanism of pre-pore to pore transition could be done, in which simulations of membrane bound monomers and oligomers at neutral and low pH could be produced to identify any residues responsible for this transition at greater resolution. A series of simulations could be produced at low pH, both of the perforin-2 monomer, and oligomers of varying sizes on a membrane to identify interactions and structural transition required to facilitate domain rearrangement.

A further perforin-2 oligomeric structure has been recently published, solving a 16 subunit pre-pore structure at 2.4 Å using cryo-electron microscopy (235). This structure formed in solution and was highly similar to the pre-pore structure reported here. A further 3.6 Å pre-pore structure was solved on a PC:PS liposome, showing that the P2 extended beta-hairpin did indeed interact with the membrane in a similar way to that proposed here. This research also found that perforin-2 is activated by acidification, however they were unable to observe the pre-pore to pore transition that we believe is accompanied by a rotation of the P2 domain around the MACPF to bring the two domains in alignment with the membrane for pore formation. Instead, a mechanism in which perforin-2 remains bound to the endosomal membrane while the MACPF domain forms pores in a closely apposed bacterial membrane is proposed, supported by the observation of pores bridging between liposomes in cryo-EM. This may function an alternative mechanism by which perforin-2 acts, which may be effective in preventing damage to the host membranes.

The identification of a fixed PFN2 pre-pore is in contrast to other MACPF/CDCs. For example, the kinetics of perforin-1 assembly have been investigated using AFM, showing that small oligomers are able to bind and punch into the membrane, forming a free lipid edge to which more subunits may subsequently join to form a larger arc or ring pore (86). While the unique structural transition of perforin-2 may require formation of a complete pre-pore before membrane insertion, it is also possible that the observed pre-pore structure could form a stable inactive form of the protein, which acts separately to the pore-forming pathway, and could be required for host protection due to its widespread expression in many non-immune cells.

The pore forming proteins investigated in this thesis are all recently discovered proteins that have important roles in health and disease but of which relatively little is known about their function. MD has been successfully used to probe mechanisms such as membrane binding which are not easily accessible for study using other methods, and therefore has proven a useful tool in their study. A combination of structural, computational and biophysical methods will be useful in the future to further investigate the pore formation of ApiPLPs and PFN2 and to come to a greater understanding of their respective roles in the apicomplexan lifecycle and immune system defence against intracellular bacteria.

References

1. van Meer G, Voelker DR, Feigenson GW. Membrane lipids: where they are and how they behave. Nature Rev Mol Cell Biol. 2008;9:112-24.

2. Cho W, Stahelin RV. Membrane-protein interactions in cell signaling and membrane trafficking. Ann Rev Biophys Biomol Str. 2005;34:119-51.

3. Shevchenko A, Simons K. Lipidomics: coming to grips with lipid diversity. Nature reviews Mol Cell Biol. 2010;11(8):593-8.

4. Leventis PA, Grinstein S. The Distribution and Function of Phosphatidylserine in Cellular Membranes. Annu Rev Biophys. 2010;39:407-27.

5. Vance JE, Tasseva G. Formation and function of phosphatidylserine and phosphatidylethanolamine in mammalian cells. Biochim Biophys Acta. 2013;1831(3):543-54.

6. Vance JE, Steenbergen R. Metabolism and functions of phosphatidylserine. Prog Lipid Res. 2005;44(4):207-34.

7. Aktas M, Danne L, Möller P, Narberhaus F. Membrane lipids in Agrobacterium tumefaciens: biosynthetic pathways and importance for pathogenesis. Front Plant Sci. 2014;5(109).

8. Meer Gv, Kroon AIPMd. Lipid map of the mammalian cell. J Cell Sci. 2011;124(1):5-8.

9. Ohvo-Rekila H, Ramstedt B, Leppimaki P, Slotte JP. Cholesterol interactions with phospholipids in membranes. Prog Lipid Res. 2002;41(1):66-97.

10. Brown DA, London E. Functions of lipid rafts in biological membranes. Annu Rev Cell Dev Biol. 1998;14:111-36.

11. Lingwood D, Simons K. Lipid rafts as a membrane-organizing principle. Science. 2010;327:46-50.

12. Monje-Galvan V, Klauda JB. Peripheral membrane proteins: Tying the knot between experiment and computation. BBA - Biomembranes. 2016;1858(7, Part B):1584-93.

13. Whited AM, Johs A. The interactions of peripheral membrane proteins with biological membranes. Chem Phys Lipids. 2015;192:51-9.

14. Wimley CW, White SH. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nature Struct Biol. 1996;3(10):842-8.

15. Murray D, Arbuzova A, Honig B, McLaughlint S. The role of electrostatic and nonpolar interactions in the association of peripheral proteins with membranes. Current topics in membranes. 52: Academic Press; 2002. p. 277-307.

16. Lemmon MA. Pleckstrin homology (PH) domains and phosphoinositides. Biochem Soc Symp. 2007;74:81-93.

17. Stahelin RV, Long F, Diraviyam K, Bruzik KS, Murray D, Cho W. Phosphatidylinositol 3-phosphate induces the membrane penetration of the FYVE domains of Vps27p and Hrs. J Biol Chem. 2002;277(29):26379-88.

18. White SH, Wimley WC. Hydrophobic interactions of peptides with membrane interfaces. Biochim Biophys Acta. 1998;1376:339-52.

19. Kachel K, Asuncion-Punzalan E, London E. Anchoring of Tryptophan and Tyrosine Analogs at the Hydrocarbon-Polar Boundary in Model Membrane Vesicles. Biochemistry. 1995;34(47):15475-9.

20. Yau W-M, Wimley WC, Gawrisch K, White SH. The Preference of Tryptophan for Membrane Interfaces. Biochemistry. 1998;37(42):14713-8.

21. Drin G, Antonny B. Amphipathic helices and membrane curvature. FEBS Lett. 2010;584(9):1840-7.

22. Zasloff M. Antimicrobial peptides of multicellular organisms. Nature. 2002;415(6870):389-95.

23. Bernstein LS, Grillo AA, Loranger SS, Linder ME. RGS4 binds to membranes through an amphipathic alpha -helix. J Biol Chem. 2000;275(24):18520-6.

24. Szeto TH, Rowland SL, Rothfield LI, King GF. Membrane localization of MinD is mediated by a C-terminal motif that is conserved across eubacteria, archaea, and chloroplasts. PNAS. 2002;99(24):15693-8.

25. Bischofberger M, Gonzalez MR, van der Goot FG. Membrane injury by pore-forming proteins. Curr Opin Cell Biol. 2009;21(4):589 - 95.

26. The Comprehensive Sourcebook of Bacterial Protein Toxins. Popoff JEAMR, editor: Academic Press; 2006.

27. Gouaux E. Channel-forming toxins: tales of transformation. Curr Opin Struct Biol. 1997;7(4):566-73.

28. Lakey JH, Slatin SL. Pore-forming colicins and their relatives. Curr Top Microbiol Immunol. 2001;257:131-61.

29. Kristan K, Podlesek Z, Hojnik V, Gutiérrez-Aguirre I, Gunčar G, Turk D, et al. Pore Formation by Equinatoxin, a Eukaryotic Pore-forming Toxin, Requires a Flexible N-terminal Region and a Stable β-Sandwich. J Biol Chem. 2004;279(45):46509-17.

30. Choe S, Bennett MJ, Fujii G, Curmi PM, Kantardjieff KA, Collier RJ, et al. The crystal structure of diphtheria toxin. Nature. 1992;357(6375):216-22

31. Parker MW, Feil SC. Pore-forming protein toxins: from structure to function. Prog Biophys Mol Biol. 2005;88(1):91-142.

32. Parker MW, Buckley JT, Postma JPM, Tucker AD, Leonard K, Pattus F, et al. Structure of the Aeromonas toxin proaerolysin in its water soluble and membrane channel states. . Nature 1994;367:292-5.

33. Petosa C, Collier RJ, Klimpel KR, Leppla SH, Liddington RC. Crystal structure of the anthrax toxin protective antigen. Nature. 1997;385(6619):833-8.

34. Rosado CJ, Kondos S, Bull TE, Kuiper MJ, Law RHP, Buckle AM, et al. The MACPF/CDC family of pore-forming toxins. Cell Microbiol. 2008;10(9):1765-74.

35. Seveau S. Multifaceted activity of listeriolysin O, the cholesterol-dependent cytolysin of Listeria monocytogenes. Subcell Biochem. 2014;80:161-95.

36. Rossjohn J, Feil SC, McKinstry WJ, Tweten RK, Parker MW. Structure of a Cholesterol-Binding, Thiol-Activated Cytolysin and a Model of Its Membrane Form. Cell. 1997;89(5):685-92.

37. Anderluh G, Kisovec M, Krasevec N, Gilbert RJ. Distribution of MACPF/CDC proteins. Subcell Biochem. 2014;80:7-30.

38. Tschopp J, Masson D, Stanley KK. Structural/functional similarity between proteins involved in complement- and cytotoxic T-lymphocyte-mediated cytolysis. Nature. 1986;322(6082):831-4.

39. Peitsch MC, Tschopp J. Assembly of macromolecular pores by immune defense systems. Curr Opin Cell Biol. 1991;3(4):710-6.

40. Voskoboinik I, Smyth MJ, Trapani JA. Perforin-mediated target-cell death and immune homeostasis. Nat Rev Immunol. 2006;6(12):940-52.

41. Anderluh G, Kisovec M, Kraševec N, Gilbert RJC. Distribution of MACPF/CDC Proteins. In: Anderluh G, Gilbert R, editors. MACPF/CDC Proteins - Agents of Defence, Attack and Invasion. Dordrecht: Springer Netherlands; 2014. p. 7-30.

42. Ni T, Gilbert RJC. Repurposing a pore: highly conserved perforin-like proteins with alternative mechanisms. Philos Trans R Soc Lond B Biol Sci. 2017;372(1726).

43. Kafsack BFC, Carruthers VB. Apicomplexan perforin-like proteins. Commun Integr Biol. 2010;3(1):18-23.

44. Rosado CJ, Buckle AM, Law RHP, Butcher RE, Kan W-T, Bird CH, et al. A Common Fold Mediates Vertebrate Defense and Bacterial Attack. Science. 2007;317(5844):1548-51.

45. Hadders MA, Beringer DX, Gros P. Structure of C8alpha-MACPF reveals mechanism of membrane attack in complement immune defense. Science. 2007;317(5844):1552-4.

46. Lukoyanova N, Saibil HR. Friend or foe: the same fold for attack and defense. Trends Immunol. 2008;29(2):51-3.

47. Heuck AP, Savva CG, Holzenburg A, Johnson AE. Conformational changes that effect oligomerization and initiate pore formation are triggered throughout perfringolysin O upon binding to cholesterol. J Biol Chem. 2007;282(31):22629-37.

48. Polekhina G, Giddings KS, Tweten RK, Parker MW. Insights into the action of the superfamily of cholesterol-dependent cytolysins from studies of intermedilysin. Proc Natl Acad Sci USA. 2005;102(3):600-5.

49. Gilbert R. Structural Features of Cholesterol Dependent Cytolysins and Comparison to Other MACPF-Domain Containing Proteins. In: Anderluh G, Gilbert R, editors. MACPF/CDC Proteins - Agents of Defence, Attack and Invasion. Dordrecht: Springer Netherlands; 2014. p. 47-62.

50. Heuck AP, Hotze EM, Tweten RK, Johnson AE. Mechanism of membrane insertion of a multimeric beta-barrel protein: perfringolysin O creates a pore using ordered and coupled conformational changes. Mol Cell. 2000;6(5):1233-42.

51. Ramachandran R, Heuck AP, Tweten RK, Johnson AE. Structural insights into the membrane-anchoring mechanism of a cholesterol-dependent cytolysin. Nat Struct Biol. 2002;9(11):823-7.

52. Giddings KS, Johnson AE, Tweten RK. Redefining cholesterol's role in the mechanism of the cholesterol-dependent cytolysins. Proc Natl Acad Sci USA. 2003;100(20):11315-20.

53. Farrand AJ, LaChapelle S, Hotze EM, Johnson AE, Tweten RK. Only two amino acids are essential for cytolytic toxin recognition of cholesterol at the membrane surface. Proc Natl Acad Sci USA. 2010;107(9):4341-6.

54. Dowd KJ, Farrand AJ, Tweten RK. The cholesterol-dependent cytolysin signature motif: a critical element in the allosteric pathway that couples membrane binding to pore assembly. PLoS Pathog. 2012;8(7):e1002787.

55. Giddings KS, Zhao J, Sims PJ, Tweten RK. Human CD59 is a receptor for the cholesterol-dependent cytolysin intermedilysin. Nat Struct Mol Biol. 2004;11(12):1173-8.

56. Law RHP, Lukoyanova N, Voskoboinik I, Caradoc-Davies TT, Baran K, Dunstone MA, et al. The structural basis for membrane binding and pore formation by lymphocyte perforin. Nature. 2010;468:447-53.

57. Voskoboinik I, Thia M-C, Fletcher J, Ciccone A, Browne K, Smyth MJ, et al. Calcium-dependent plasma membrane binding and cell lysis by perforin are mediated through its C2 domain: A critical role for aspartate residues 429, 435, 483, and 485 but not 491. J Biol Chem. 2005;280(9):8426-34.

58. Antia R, Schlegel RA, Williamson P. Binding of perforin to membranes is sensitive to lipid spacing and not headgroup. Immunology letters. 1992;32(2):153-7.

59. Traore DA, Brennan A. J. Law R. H. Dogovski C. Perugini M. A. Lukoyanova N. Leung E. W. Norton R. S. Lopez J. A. Browne K. A. Yagita H. Lloyd G. J. Ciccone A. Verschoor S. Trapani J. A. Whiss- tock J. C., Voskoboinik I. Defining the interaction of perforin with calcium and the phospholipid membrane. Biochem J. 2013;456:323-35.

60. Yagi H, Conroy PJ, Leung EWW, Law RHP, Trapani JA, Voskoboinik I, et al. Structural Basis for Ca2+-mediated Interaction of the Perforin C2 Domain with Lipid Membranes. J Biol Chem. 2015;290(42):25213-26.

61. Sakurai N, Kaneko J, Kamio Y, Tomita T. Cloning, expression, and pore-forming properties of mature and precursor forms of pleurotolysin, a sphingomyelin-specific two-component cytolysin from the edible mushroom Pleurotus ostreatus. Biochim Biophys Acta. 2004;1679(1):65-73.

62. Tomita T, Noguchi K, Mimuro H, Ukaji F, Ito K, Sugawara-Tomita N, et al. Pleurotolysin, a novel sphingomyelin-specific two-component cytolysin from the edible mushroom Pleurotus ostreatus, assembles into a transmembrane pore complex. J Biol Chem. 2004;279(26):26975-82.

63. Lukoyanova N, Kondos SC, Farabella I, Law RHP, Reboul CF, Caradoc-Davies TT, et al. Conformational changes during pore formation by the perforin-related protein pleurotolysin. PLoS Biol. 2015;13(2):e1002049.

64. Preissner KT, Podack ER, Muller-Eberhard HJ. The membrane attack complex of complement: relation of C7 to the metastable membrane binding site of the intermediate complex C5b-7. J Immunol. 1985;135(1):445-51.

65. Dunstone MA, Tweten RK. Packing a punch: the mechanism of pore formation by cholesterol dependent cytolysins and membrane attack complex/perforin-like proteins. Curr Opin Struct Biol. 2012;22(3):342--9.

66. Sharp TH, Koster AJ, Gros P. Heterogeneous MAC Initiator and Pore Structures in a Lipid Bilayer by Phase-Plate Cryo-electron Tomography. Cell Reports. 2016;15(1):1-8.

67. Morgan BP, Boyd C, Bubeck D. Molecular cell biology of complement membrane attack. Seminars in cell & developmental biology. 2017;72:124-32.

68. Reboul CF, Whisstock JC, Dunstone MA. Giant MACPF/CDC pore forming toxins: A class of their own. BBA - Biomembranes. 2016;1858(3):475-86.

69. Ramachandran R, Tweten RK, Johnson AE. Membrane-dependent conformational changes initiate cholesterol-dependent cytolysin oligomerization and intersubunit beta-strand alignment. Nat Struct Mol Biol. 2004;11(8):697-705.

70. Ramachandran R, Tweten RK, Johnson AE. The domains of a cholesterol-dependent cytolysin undergo a major FRET-detected rearrangement during pore formation. Proc Natl Acad Sci USA. 2005;102(20):7139-44.

71. Tweten RK. Cholesterol-dependent cytolysins, a family of versatile pore-forming toxins. Infect Immun. 2005;73(10):6199-209.

72. Gilbert RJC. Protein-lipid interactions and non-lamellar lipidic structures in membrane pore formation and membrane fusion. Biochim Biophys Acta. 2016;1858(3):487-99.

73. Shepard LA, Shatursky O, Johnson AE, Tweten aRK. The Mechanism of Pore Assembly for a Cholesterol-Dependent Cytolysin: Formation of a Large Prepore Complex Precedes the Insertion of the Transmembrane β-Hairpins. Biochemistry. 2000;39(33):10284-93.

74. Gilbert RJ, Jimenez JL, Chen S, Tickle IJ, Rossjohn J, Parker M, et al. Two structural transitions in membrane pore formation by pneumolysin, the pore-forming toxin of Streptococcus pneumoniae. Cell. 1999;97(5):647-55.

75. Shatursky O, Heuck AP, Shepard LA, Rossjohn J, Parker MW, Johnson AE, et al. The mechanism of membrane insertion for a cholesterol-dependent cytolysin: a novel paradigm for pore-forming toxins. Cell. 1999;99(3):293--9.

76. Shepard LA, Heuck AP, Hamman BD, Rossjohn J, Parker MW, Ryan KR, et al. Identification of a membrane-spanning domain of the thiol-activated pore-forming toxin Clostridium perfringens perfringolysin O: an alpha-helical to beta-sheet transition identified by fluorescence spectroscopy. Biochemistry. 1998;37(41):14563-74.

77. Hotze EM, Heuck AP, Czajkowsky DM, Shao Z, Johnson AE, Tweten RK. Monomer-monomer interactions drive the prepore to pore conversion of a beta-barrel-forming cholesterol-dependent cytolysin. J Biol Chem. 2002;277(13):11597-605.

78. Czajkowsky DM, Hotze EM, Shao Z, Tweten RK. Vertical collapse of a cytolysin prepore moves its transmembrane β-hairpins to the membrane. The EMBO journal. 2004;23(16):3206-15.

79. Tilley SJ, Orlova EV, Gilbert RJC, Andrew PW, Saibil HR. Structural basis of pore formation by the bacterial toxin pneumolysin. Cell. 2005;121(2):247-56.

80. Pee Kv, Neuhaus A, D'Imprima E, Mills DJ, Kühlbrandt W, Yildiz Ö. CryoEM structures of membrane pore and prepore complex reveal cytolytic mechanism of Pneumolysin. eLife. 2017;6:e23644.

81. Tschopp J. Ultrastructure of the membrane attack complex of complement. Heterogeneity of the complex caused by different degree of C9 polymerization. J Biol Chem. 1984;259(12):7857-63.

82. Gilbert RJC, Mikelj M, Dalla Serra M, Froelich CJ, Anderluh G. Effects of MACPF/CDC proteins on lipid membranes. Cell Mol Life Sci. 2013;70(12):2083-98.

83. Podobnik M, Marchioretto M, Zanetti M, Bavdek A, Kisovec M, Cajnko MM, et al. Plasticity of Listeriolysin O Pores and its Regulation by pH and Unique Histidine. Sci Rep. 2015;5:9623.

84. Gilbert RJC, Sonnen AF-P. Measuring kinetic drivers of pneumolysin pore structure. Eur Biophys J. 2016;45(4):365-76.

85. Lovelace LL, Cooper CL, Sodetz JM, Lebioda L. Structure of human C8 protein provides mechanistic insight into membrane pore formation by complement. J Biol Chem. 2011;286(20):17585-92.

86. Leung C, Hodel AW, Brennan AJ, Lukoyanova N, Tran S, House CM, et al. Real-time visualization of perforin nanopore assembly. Nat Nanotechnol. 2017;12(5):467-73.

87. Serna M, Giles JL, Morgan BP, Bubeck D. Structural basis of complement membrane attack complex formation. Nat Commun. 2016;7:10587.

88. Menny A, Serna M, Boyd CM, Gardner S, Joseph AP, Morgan BP, et al. CryoEM reveals how the complement membrane attack complex ruptures lipid bilayers. Nat Commun. 2018;9(1):5316.

89. Parsons ES, Stanley GJ, Pyne ALB, Hodel AW, Nievergelt AP, Menny A, et al. Single-molecule kinetics of pore assembly by the membrane attack complex. Nat Commun. 2019;10(1):2066.

90. Kaiser K, Camargo N, Coppens I, Morrisey JM, Vaidya AB, Kappe SHI. A member of a conserved Plasmodium protein family with membrane-attack complex/perforin (MACPF)-like domains localizes to the micronemes of sporozoites. Mol Biochem Parasitol. 2004;133(1):15-26.

91. Ni T, Williams SI, Rezelj S, Anderluh G, Harlos K, Stansfeld PJ, et al. Structures of monomeric and oligomeric forms of the Toxoplasma gondii perforin-like protein 1. Sci Adv. 2018;4(3).

92. Chitnis CE, Blackman MJ. Host Cell Invasion by Malaria Parasites. Parasitol Today. 2000;16(10):411-5.

93. Saliba KJ, Kirk K. Nutrient acquisition by intracellular apicomplexan parasites: staying in for dinner. Int J Parasitol. 2001;31(12):1321-30.

94. Blackman MJ, Carruthers VB. Recent insights into apicomplexan parasite egress provide new views to a kill. Curr Opin Microbiol. 2013;16(4):459-64.

95. Baum J, Papenfuss AT, Baum B, Speed TP, Cowman AF. Regulation of apicomplexan actin-based motility. Nat Rev Microbiol. 2006;4(8):621-8.

96. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson M-A, Roy SL, et al. Foodborne illness acquired in the United States--major pathogens. Emerg Infect Dis. 2011;17(1):7-15.

97. Aguirre AA, Longcore T, Barbieri M, Dabritz H, Hill D, Klein PN, et al. The One Health Approach to Toxoplasmosis: Epidemiology, Control, and Prevention Strategies. EcoHealth. 2019.

98. Kafsack BFC, Pena JDO, Coppens I, Ravindran S, Boothroyd JC, Carruthers VB. Rapid membrane disruption by a perforin-like protein facilitates parasite exit from host cells. Science. 2009;323(5913):530-3.

99. Kim K, Weiss LM. Toxoplasma gondii: the model apicomplexan. Int J Parasitol. 2004;34(3):423-32.

100. Carruthers VB. Host cell invasion by the opportunistic pathogen Toxoplasma gondii. Acta Tropica. 2002;81(2):111 - 22.

101. Robert-Gangneux F, Darde ML. Epidemiology of and diagnostic strategies for toxoplasmosis. Clin Microbiol Rev. 2012;25(2):264-96.

102. Blader IJ, Coleman BI, Chen CT, Gubbels MJ. Lytic Cycle of Toxoplasma gondii: 15 Years Later. Annu Rev Microbiol. 2015;69:463-85.

103. Black MW, Boothroyd JC. Lytic cycle of Toxoplasma gondii. Microbiol Mol Biol Rev. 2000;64(3):607–23.

104. Roiko MS, Svezhova N, Carruthers VB. Acidification Activates Toxoplasma gondii Motility and Egress by Enhancing Protein Secretion and Cytolytic Activity. PLoS Pathog. 2014;10(11):e1004488.

105. Frénal K, Dubremetz J-F, Lebrun M, Soldati-Favre D. Gliding motility powers invasion and egress in Apicomplexa. Nat Rev Microbiol. 2017;15:645.

106. Egarter S, Andenmatten N, Jackson AJ, Whitelaw JA, Pall G, Black JA, et al. The toxoplasma Acto-MyoA motor complex is important but not essential for gliding motility and host cell invasion. PLOS ONE. 2014;9(3):e91819.

107. Mehta S, Sibley LD. Actin depolymerizing factor controls actin turnover and gliding motility in Toxoplasma gondii. Mol Biol Cell. 2011;22(8):1290-9.

108. Phillips MA, Burrows JN, Manyando C, van Huijsduijnen RH, Van Voorhis WC, Wells TNC. Malaria. Nat Rev Dis Primers. 2017;3:17050.

109. Mota MM, Rodriguez A. Invasion of mammalian host cells by Plasmodium sporozoites. BioEssays. 2002;24(2):149-56.

110. Amino R, Giovannini D, Thiberge S, Gueirard P, Bertrand B, Dubremetz J-F, et al. Host cell traversal is important for progression of the malaria parasite through the dermis to the liver. Cell Host Microbe. 2008;3:88-96.

111. Risco-Castillo V, Topcu S, Marinach C, Manzoni G, Bigorgne AE, Briquet S, et al. Malaria Sporozoites Traverse Host Cells within Transient Vacuoles. Cell Host Microbe. 2015;18(5):593-603.

112. Ishino T, Chinzei Y, Yuda M. A Plasmodium sporozoite protein with a membrane attack complex domain is required for breaching the liver sinusoidal cell layer prior to hepatocyte infection. Cell Microbiol. 2005;7(2):199-208.

113. Yang ASP, O'Neill MT, Jennison C, Lopaticki S, Allison CC, Armistead JS, et al. Cell Traversal Activity Is Important for Plasmodium falciparum Liver Infection in Humanized Mice. Cell Reports. 2017;18(13):3105-16.

114. Garg S, Agarwal S, Kumar S, Yazdani SS, Chitnis CE, Singh S. Calcium-dependent permeabilization of erythrocytes by a perforin-like protein during egress of malaria parasites. Nat Commun. 2013;4:1736.

115. Deligianni E, Morgan RN, Bertuccini L, Wirth CC, Silmon de Monerri NC, Spanos L, et al. A perforin-like protein mediates disruption of the erythrocyte membrane during egress of Plasmodium berghei male gametocytes. Cell Microbiol. 2013;15(8):1438-55.

116. Wirth CC, Glushakova S, Scheuermayer M, Repnik U, Garg S, Schaack D, et al. Perforin-like protein PPLP2 permeabilizes the red blood cell membrane during egress of Plasmodium falciparum gametocytes. Cell Microbiol. 2014;16(5):709-33.

117. Sologub L, Kuehn A, Kern S, Przyborski J, Schillig R, Pradel G. Malaria proteases mediate inside-out egress of gametocytes from red blood cells following parasite transmission to the mosquito. Cell Microbiol. 2011;13(6):897-912.

118. Wirth CC, Pradel G. Molecular mechanisms of host cell egress by malaria parasites. Int J Med Microbiol. 2012;302(4-5):172-8.

119. Voskoboinik I, Whisstock JC, Trapani JA. Perforin and granzymes: function, dysfunction and human pathology. Nat Rev Immunol. 2015;15(6):388-400.

120. Osińska I, Popko K, Demkow U. Perforin: an important player in immune response. Centr Eur J Immunol. 2014;39(1):109-15.

121. McCormack R, Armas Ld, Shiratsuchi M, Podack ER. Killing machines: three pore-forming proteins of the immune system. Immunol Res. 2013;57(1-3):268-78.

122. McCormack R, Podack ER. Perforin-2/Mpeg1 and other pore-forming proteins throughout evolution. J Leukoc Biol. 2015;98(5):761-8.

123. D'Angelo ME, Dunstone MA, Whisstock JC, Trapani JA, Bird PI. Perforin evolved from a gene duplication of MPEG1, followed by a complex pattern of gene gain and loss within Euteleostomi. BMC Evolutionary Biology. 2012;12:59.

124. McCormack RM, Armas LRd, Shiratsuchi M, Fiorentino DG, Olsson ML, Lichtenheld MG, et al. Perforin-2 is essential for intracellular defense of parenchymal cells and phagocytes against pathogenic bacteria. eLife. 2015;4:e06508.

125. McCormack R, de Armas LR, Shiratsuchi M, Ramos J, Podack ER. Inhibition of intracellular bacterial replication in fibroblasts is dependent on the perforin-like protein (Perforin-2) encoded by macrophage expressed gene 1. J Innate Immun. 2013;5(2):185-94.

126. Fields KA, McCormack R, de Armas LR, Podack ER. Perforin-2 Restricts Growth of Chlamydia trachomatis in Macrophages. Infect Immun. 2013;81(8):3045-54.

127. McCormack RM, Szymanski EP, Hsu AP, Perez E, Olivier KN, Fisher E, et al. MPEG1/perforin-2 mutations in human pulmonary nontuberculous mycobacterial infections. JCI Insight. 2017(8).

128. Dowhan W, Bogdanov M, Mileykovskaya E. Chapter 1 - Functional Roles of Lipids in Membranes. In: Ridgway ND, McLeod RS, editors. Biochemistry of Lipids, Lipoproteins and Membranes (Sixth Edition). Boston: Elsevier; 2016. p. 1-40.

129. May KL, Silhavy TJ. Making a membrane on the other side of the wall. BBA - Molecular and Cell Biology of Lipids. 2017;1862(11):1386 - 93.

130. Brown L, Wolf JM, Prados-Rosales R, Casadevall A. Through the wall: extracellular vesicles in Gram-positive bacteria, mycobacteria and fungi. Nat Rev Microbiol. 2015;13:620.

131. Bagheri M, Keller S, Dathe M. Interaction of W-Substituted Analogs of Cyclo-RRRWFW with Bacterial Lipopolysaccharides: the Role of the Aromatic Cluster in Antimicrobial Activity. Antimicrob Agents Chemother. 2011;55:788-97.

132. Naughton FB, Kalli AC, Sansom MSP. Association of Peripheral Membrane Proteins with Membranes: Free Energy of Binding of GRP1 PH Domain with Phosphatidylinositol Phosphate-Containing Model Bilayers. J Phys Chem. 2016;7(7):1219-24.

133. Yamamoto E, Kalli AC, Yasuoka K, Sansom MSP. Interactions of Pleckstrin Homology Domains with Membranes: Adding Back the Bilayer via High-Throughput Molecular Dynamics. Structure. 2016;24(8):1421-31.

134. Lumb CN, He J, Xue Y, Stansfeld PJ, Stahelin RV, Kutateladze TG, et al. Biophysical and computational studies of membrane penetration by the GRP1 pleckstrin homology domain. Structure. 2011;19(9):1338-46.

135. Jaud S, Tobias DJ, Falke JJ, White SH. Self-induced docking site of a deeply embedded peripheral membrane protein. Biophys J. 2007;92:517–24.

136. Manna D, Bhardwaj N, Vora MS, Stahelin RV, Lu H, Cho W. Differential roles of phosphatidylserine, PtdIns(4,5)P2, and PtdIns(3,4,5)P3 in plasma membrane targeting of C2 domains. Molecular dynamics simulation, membrane binding, and cell translocation studies of the PKCalpha C2 domain. J Biol Chem. 2008;283(38):26047-58.

137. Lumb CN, Sansom MSP. Finding a needle in a haystack: the role of electrostatics in target lipid recognition by PH domains. PLoS Comput Biol. 2012;8(7):e1002617.

138. Naughton FB, Kalli AC, Sansom MSP. Modes of Interaction of Pleckstrin Homology Domains with Membranes: Toward a Computational Biochemistry of Membrane Recognition. J Mol Biol. 2018;430(3):372-88.

139. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput. 2015;11(8):3696-713.

140. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF. 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. 2004;25(13):1656-76.

141. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc. 1996;118:11225-36.

142. Brooks BR, Brooks CL, 3rd, Mackerell AD, Jr., Nilsson L, Petrella RJ, Roux B, et al. CHARMM: the biomolecular simulation program. J Comput Chem. 2009;30(10):1545-614.

143. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins. 2010;78(8):1950-8.

144. Notman R, Anwar J. Breaching the skin barrier — Insights from molecular simulation of model membranes. Adv Drug Deliv Rev. 2013;65(2):237-50.

145. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. Interaction models for water in relation to protein hydration. In: B.Pullman, editor. Intermolecular Forces. Dordrecht: Reidel; 1981. p. 331-42.

146. Berweger CD, van Gunsteren WF, Mullerplathe F. Force-field parametrization by weak-coupling - reengineering SPC water. Chem Phys Lett. 1995;232:429-36.

147. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. J Comp Chem. 1997;18:1463-72.

148. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1-2:19-25.

149. van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. J Comput Chem. 2005;26(16):1701-18.

150. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys. 1995;103(19):8577-93.

151. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys. 2007;126(1):014101.

152. Parrinello M, Rahman A. Polymorphic transitions in single-crystals - a new molecular-dynamics method. J Appl Phys. 1981;52:7182-90.

153. Periole X, Marrink S. The Martini Coarse-Grained Force Field. Methods Mol Biol. 2013;924:533-65.

154. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH. The MARTINI force field: coarse grained model for biomolecular simulations. J Phys Chem B. 2007;111(27):7812-24.

155. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink S-J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. J Chem Theory Comput. 2008;4(5):819-34.

156. de Jong DH, Singh G, Bennett WF, Arnarez C, Wassenaar TA, Schafer LV, et al. Improved Parameters for the Martini Coarse-Grained Protein Force Field. J Chem Theory Comput. 2013;9(1):687-97.

157. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984;81:3684-90.

158. Stansfeld P, Sansom M. From Coarse Grained to Atomistic: A Serial Multiscale Approach to Membrane Protein Simulations. J Chem Theory Comput. 2011;7(4):1157-66.

159. Stansfeld PJ, Sansom MS. Molecular simulation approaches to membrane proteins. Structure. 2011;19(11):1562-72.

160. Lomize AL, Pogozheva ID, Lomize MA, Mosberg HI. Positioning of proteins in membranes: A computational approach. Protein Sci. 2006;15(6):1318-33.

161. Scott KA, Bond PJ, Ivetac A, Chetwynd AP, Khalid S, Sansom MSP. Coarse-grained MD simulations of membrane protein-bilayer self-assembly Structure. 2008;16:621-30.

162. Stansfeld PJ, Goose JE, Caffrey M, Carpenter EP, Parker JL, Newstead S, et al. MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes. Structure. 2015;23(7):1350-61.

163. Newport TD, Sansom MSP, Stansfeld PJ. The MemProtMD database: a resource for membrane-embedded protein structures and their lipid interactions. Nucleic Acid Res. 2019;47(D1):D390–D7.

164. Souaille M, Roux Bt. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. Comput Phys Commun. 2001;135(1):40-57.

165. Kästner J. Umbrella sampling. WIREs Comput Mol Sci. 2011;1(6):932-42.

166. Torrie GM, Valleau JP. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. Chem Phys Lett. 1974;28(4):578-81.

167. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. The weighted histogram analysis method for free-energy calculations on biomolecules .1. The method. J Comp Chem. 1992;13:1011-21.

168. Hub JS, de Groot BL, van der Spoel D. g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. J Chem Theory Comput. 2010;6(12):3713-20.

169. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. Comp Phys Comm. 1995;95:43-56.

170. Gowers RJ, Linke M, Barnoud J, Reddy TJE, Melo MN, Seyler SL, et al., editors. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. Proceedings of the 15th Python in Science Conference; 2016.

171. Schrodinger L. The PyMOL Molecular Graphics System, Version 1.8. 2015.

172. Hunter JD. Matplotlib: A 2D graphics environment. Comput Sci Eng. 2007;9(3):90-5.

173. Waskom M, Botvinnik O, O'Kane D, Hobson; P, Ostblom; J, Lukauskas; S, et al. mwaskom/seaborn: v0.9.0 (July 2018). 2018.

174. Aricescu AR, Lu W, Jones EY. A time- and cost-efficient system for high-level protein production in mammalian cells. Acta Crystallogr D Biol Crystallogr. 2006;62(10):1243-50.

175. Gupta R, Jung E, Brunak S. Prediction of N-glycosylation sites in human proteins. 2004;46:203-6.

176. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007;40(4):658-74.

177. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, et al. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D. 2012;68(4):352-67.

178. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D. 2004;60(12):2126-32.

179. Nettleship JE, Brown J, Groves MR, Geerlof A. Methods for protein characterization by mass spectrometry, thermal shift (ThermoFluor) assay, and multiangle or static light scattering. Methods Mol Biol. 2008;426:299-318.

180. Furtado JM, Smith JR, Belfort R, Jr., Gattey D, Winthrop KL. Toxoplasmosis: a global threat. J Glob Infect Dis. 2011;3(3):281-4.

181. Caldas LA, de Souza W. A Window to Toxoplasma gondii Egress. Pathogens. 2018;7(3):13.

182. Roiko MS, Carruthers VB. Functional Dissection of Toxoplasma gondii Perforin-like Protein 1 Reveals a Dual Domain Mode of Membrane Binding for Cytolysis and Parasite Egress. J Biol Chem. 2013(288):8712-25.

183. Lagal V, Binder EM, Huynh M-H, Kafsack BFC, Harris PK, Diez R, et al. Toxoplasma gondii protease TgSUB1 is required for cell surface processing of micronemal adhesive complexes and efficient adhesion of tachyzoites. Cell Microbiol. 2010;12(12):1792-808.

184. Chandramohanadas R, Davis PH, Beiting DP, Harbut MB, Darling C, Velmourougane G, et al. Apicomplexan parasites co-opt host calpains to facilitate their escape from infected cells. Science. 2009;324(5928):794-7.

185. Roiko MS, Carruthers VB. New roles for perforins and proteases in apicomplexan egress. Cell Microbiol. 2009;11(10):1444-52.

186. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Protein Sci. 2016;54: 5.6.1–5.6.37.

187. Guerra AJ, Zhang O, Bahr CME, Huynh M-H, DelProposto J, Brown WC, et al. Structural basis of Toxoplasma gondii perforin-like protein 1 membrane interaction and activity during egress. PLOS Pathogens. 2018;14(12):1-21.

188. Tavares J, Amino R, Ménard R. MACPF/CDC Proteins - Agents of Defence, Attack and Invasion. In: Anderluh G, Gilbert R, editors. MACPF/CDC Proteins - Agents of Defence, Attack and Invasion. Subcellular Biochemistry. 80: Springer Netherlands; 2014. p. 241-53.

189. Feng JW, Wehbi H, Roberts MF. Role of tryptophan residues in interfacial binding of phosphatidylinositol-specific phospholipase C. J Biol Chem. 2002;277(22):19867-75.

190. Fuglebakk E, Reuter N. A model for hydrophobic protrusions on peripheral membrane proteins. PLoS Comput Biol. 2018;14(7):26.

191. Ni T, Kalli A, Naughton F, Yates L, Naneh O, Kozorog M, et al. Structure and lipid binding properties of the kindlin-3 pleckstrin homology domain. Biochem J. 2016;474(4):539–56.

192. Stansfeld PJ, Sansom MSP. Molecular Simulation Approaches to Membrane Proteins. Structure. 2011;19(11):1562-72.

193. Punta M, Forrest LR, Bigelow H, Kernytsky A, Liu J, Rost B. Membrane protein prediction methods. Methods. 2007;41(4):460-74.

194. Savojardo C, Fariselli P, Casadio R. BETAWARE: a machine-learning tool to detect and predict transmembrane beta-barrel proteins in prokaryotes. Bioinformatics. 2013;29(4):504-5.

195. Hayat S, Peters C, Shu N, Tsirigos KD, Elofsson A. Inclusion of dyad-repeat pattern improves topology prediction of transmembrane beta-barrel proteins. Bioinformatics. 2016;32(10):1571-3.

196. Tsirigos KD, Elofsson A, Bagos PG. PRED-TMBB2: improved topology prediction and detection of beta-barrel outer membrane proteins. Bioinformatics. 2016;32(17):665-71.

197. Koehler Leman J, Ulmschneider MB, Gray JJ. Computational modeling of membrane proteins. Proteins. 2015;83(1):1-24.

198. Ulmschneider MB, Sansom MSP. Amino acid distributions in integral membrane protein structures. Biochim Biophys Acta. 2001;1512:1-14.

199. Murzin AG, Lesk AM, Chothia C. Principles determining the structure of β-sheet barrels in proteins I. A theoretical analysis. J Mol Biol. 1994;236(5):1369-81.

200. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: resources for positioning of proteins in membranes. Nucl Acids Res 2012;40:370-6.

201. Nugent T, Jones DT. Membrane protein orientation and refinement using a knowledge-based statistical potential. BMC Bioinformatics. 2013;14:276.

202. Bokori-Brown M, Martin TG, Naylor CE, Basak AK, Titball RW, Savva CG. Cryo-EM structure of lysenin pore elucidates membrane insertion by an aerolysin family protein. Nat Commun. 2016;7:11293.

203. Podobnik M, Savory P, Rojko N, Kisovec M, Wood N, Hambley R, et al. Crystal structure of an invertebrate cytolysin pore reveals unique properties and mechanism of assembly. Nat Commun. 2016;7:11598.

204. Gilbert RJC, Serra MD, Froelich CJ, Wallace MI, Anderluh G. Membrane pore formation at protein--lipid interfaces. Trends Biochem Sci. 2014;39(11):510-6.

205. Kutzner C, Grubmüller H, de Groot Bert L, Zachariae U. Computational Electrophysiology: The Molecular Dynamics of Ion Channel Permeation and Selectivity in Atomistic Detail. Biophys J. 2011;101(4):809-17.

206. Serna M. Hands on Methods for High Resolution Cryo-Electron Microscopy Structures of Heterogeneous Macromolecular Complexes. Front Mol Biosci. 2019;6(33).

207. Vögele M, Bhaskara RM, Mulvihill E, van Pee K, Yildiz Ö, Kühlbrandt W, et al. Membrane perforation by the pore-forming toxin pneumolysin. PNAS. 2019;116(27):13352-7.

208. Andreadaki M, Hanssen E, Deligianni E, Claudet C, Wengelnik K, Mollard V, et al. Sequential Membrane Rupture and Vesiculation during Plasmodium berghei Gametocyte Egress from the Red Blood Cell. Sci Rep. 2018;8(1):3543.

209. Kadota K, Ishino T, Matsuyama T, Chinzei Y, Yuda M. Essential role of membrane-attack protein in malarial transmission to mosquito host. Proc Natl Acad Sci USA. 2004;101(46):16310-5.

210. Ecker A, Pinto SB, Baker KW, Kafatos FC, Sinden RE. Plasmodium berghei: plasmodium perforin-like protein 5 is required for mosquito midgut invasion in Anopheles stephensi. Exp Parasitol. 2007;116(4):504-8.

211. Wirth CC, Bennink S, Scheuermayer M, Fischer R, Pradel G. Perforin-like protein PPLP4 is crucial for mosquito midgut infection by Plasmodium falciparum. Mol Biochem Parasitol. 2015;201(2):90-9.

212. Deligianni E, Silmon de Monerri NC, McMillan PJ, Bertuccini L, Superti F, Manola M, et al. Essential role of Plasmodium perforin-like protein 4 in ookinete midgut passage. PLOS ONE. 2018;13(8):1-20.

213. Rambaut A. FigTree. GitHub2018.

214. Ni T. Structural and Functional Study of MACPF/CDC proteins: University of Oxford; 2017.

215. Vedadi M, Lew J, Artz J, Amani M, Zhao Y, Dong A, et al. Genome-scale protein expression and structural biology of Plasmodium falciparum and related Apicomplexan organisms. Mol Biochem Parasitol. 2007;151(1):100-10.

216. Aricescu AR, Assenberg R, Bill RM, Busso D, Chang VT, Davis SJ, et al. Eukaryotic expression: developments for structural proteomics. Acta Crystallogr D. 2006;62(10):1114-24.

217. Gupta K, Selinsky BS, Kaub CJ, Katz AK, Loll PJ. The 2.0A resolution crystal structure of prostaglandin H(2) synthase-1: structural insights into an unusual peroxidase. J Mol Biol. 2004;335:503-18

218. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 2016;44:344-50.

219. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 2014;42(W1):W320-W4.

220. Podack ER, Tschoop J, Muller-Eberhard HJ. Molecular organization of C9 within the membrane attack complex of complement. Induction of circular C9 polymerization by the C5b-8 assembly. J Exp Med. 1982;156(1):268-82.

221. Spilsbury K, O'Mara MA, Wu WM, Rowe PB, Symonds G, Takayama Y. Isolation of a novel macrophage-specific gene by differential cDNA analysis. Blood. 1995;85(6):1620-9.

222. Benard EL, Racz PI, Rougeot J, Nezhinsky AE, Verbeek FJ, Spaink HP, et al. Macrophage-Expressed Perforins Mpeg1 and Mpeg1.2 Have an Anti-Bacterial Function in Zebrafish. J Innate Immun. 2015;7(2):136-52.

223. Xiong P, Shiratsuchi M, Matsushima T, Liao J, Tanaka E, Nakashima Y, et al. Regulation of expression and trafficking of perforin-2 by LPS and TNF-alpha. Cell Immunol. 2017;320:1-10.

224. Nugent T, Jones DT. Accurate de novo structure prediction of large transmembrane protein domains using fragment-assembly and correlated mutation analysis. Proc Natl Acad Sci USA. 2012;109(24):E1540-7.

225. Praper T, Besenicar Mp Fau - Istinic H, Istinic H Fau - Podlesek Z, Podlesek Z Fau - Metkar SS, Metkar Ss Fau - Froelich CJ, Froelich Cj Fau - Anderluh G, et al. Human perforin permeabilizing activity, but not binding to lipid membranes, is affected by pH. Mol Immunol. 2010;47(15):2492-504.

226. Lawrence SL, Gorman MA, Feil SC, Mulhern TD, Kuiper MJ, Ratner AJ, et al. Structural Basis for Receptor Recognition by the Human CD59-Responsive Cholesterol-Dependent Cytolysins. Structure. 2016;24(9):1488-98.

227. Keskin O, Gursoy A, Ma B, Nussinov R. Principles of protein-protein interactions: what are the preferred ways for proteins to interact? Chem Rev. 2008;108(4):1225-44.

228. Lomize AL, Pogozheva ID, Lomize MA, Mosberg HI. The role of hydrophobic interactions in positioning of peripheral proteins in membranes. BMC Struct Biol. 2007;7(44).

229. Lemmon MA. Membrane recognition by phospholipid-binding domains. Nat Rev Mol Cell Biol. 2008;9(2):99-111.

230. Heesterbeek DAC, Angelier ML, Harrison RA, Rooijakkers SHM. Complement and Bacterial Infections: From Molecular Mechanisms to Therapeutic Applications. J Innate Immun. 2018;10(5-6):455-64.

231. Hadders MA, Bubeck D, Roversi P, Hakobyan S, Forneris F, Morgan BP, et al. Assembly and regulation of the membrane attack complex based on structures of C5b6 and sC5b9. Cell Rep. 2012;1(3):200-7.

232. Bai FF, McCormack RM, Hower S, Plano GV, Lichtenheld MG, Munson GP. Perforin-2 Breaches the Envelope of Phagocytosed Bacteria Allowing Antimicrobial Effectors Access to Intracellular Targets. J Immunol. 2018;201(9):2710-20.

233. Kalani MR, Moradi A, Moradi M, Tajkhorshid E. Characterizing a histidine switch controlling pH-dependent conformational changes of the influenza virus hemagglutinin. Biophys J. 2013;105(4):993-1003.

234. Valéry C, Deville-Foillard S, Lefebvre C, Taberner N, Legrand P, Meneau F, et al. Atomic view of the histidine environment stabilizing higher-pH conformations of pH-dependent proteins. Nat Commun. 2015;6:7771.

235. Pang SS, Bayly-Jones C, Radjainia M, Spicer BA, Law RHP, Hodel AW, et al. The cryo-EM structure of the acid activatable pore-forming immune effector Macrophage-expressed gene 1. Nat Commun. 2019;10(1):4288.