The Influence of Environment: Simulations of the Influenza BM2 Proton Channel

Sarah L. Rouse



Structural Bioinformatics & Computational Biochemistry,

Department of Biochemistry and Keble College, Oxford

TRINITY 2011

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy at the University of Oxford

The Influence of Environment: Simulations

of the Influenza BM2 Proton Channel

Thesis Abstract, submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford in Trinity Term 2011.

Sarah L. Rouse

Department of Biochemistry and Keble College, University of Oxford

The *influenza* BM2 proton channel is a small, tetrameric α-helical membrane protein, which is vital in multiple stages of the life cycle of the virus. The transmembrane (TM) domain is responsible for modulating pH to facilitate release of viral genetic material whilst the cytoplasmic domain is purported to play a role in virus budding *via* interactions with another protein, BM1.

Coarse-grained molecular dynamics (CGMD) simulations have been employed in several studies of α-helical membrane protein folding with success. This methodology is applied to the BM2 TM domain to generate a structure of the tetramer based on sequence alone. A converged tetrameric bundle with left-handed packing was generated and stable following conversion to atomistic representation and conventional atomistic MD simulation. Sequential protonation of His19 residues within the putative HxxxW gating motif led to subtle rearrangements in the packing of the helices, allowing formation of a water wire once three of the His19 residues were protonated.

The influence of solubilising detergents on membrane protein structure is a topic of debate in the literature. The recent solution NMR structure of the BM2 TM domain was used as the basis of a study on the influence of two types of detergent micelle environment compared to a dipalmitoylphosphatidylcholine (DPPC) bilayer. The detergents chosen were zwitterionic dihexanoylphosphatidylcholine (DHPC) and non-ionic β-dodecylmaltoside (DDM). The results of these simulations suggest that BM2 is able to modulate its helix packing to compensate for hydrophobic mismatch in the DPPC bilayer. The protein-detergent complexes appeared to retain some of the features of detergent-only micelles. DDM detergent molecules replicated the binding mode of DPPC lipids (parallel to the protein bundle axis) whereas the DHPC detergent molecules adopted a binding mode perpendicular to the protein bundle axis. This appeared to influence the binding on/off rates of each detergent.

Recent developments in the field of mass spectrometry have allowed intact membrane protein complexes in detergent micelles to be stabilised in the gas phase. Questions remain over the structure of these gas phase complexes compared to solution phase. A combination of non-equilibrium and equilibrium MD simulations was used to study the effects of transfer from solution to vacuum of protein-detergent complexes. The DHPC detergent micelle formed destabilising interactions with BM2 upon dehydration whereas the presence of the DDM micelle stabilised the solution phase protein conformation.

In the final Chapter the multiscale simulation approach is used to study the interactions between BM2 and BM1. Protein-protein interactions formed *via* CGMD simulation were conserved upon atomistic simulation. The results of this thesis highlight the worth of multiscale simulation approaches in observing a greater range of timescales and processes.

Acknowledgements

Firstly I would like to thank Prof MSP Sansom for being an excellent supervisor and for proving that nice people (I haven't heard any threats from Mark about running over grad students for at least a year now) can be successful.

I would also like to thank:

Timmay and Robbie (copper nanotube) for showing me the ropes when I first joined the group. Carol V. Robinson and Julien Marcoux for helpful discussions on the mass spec work. Mark Howarth for supervising my first rotation project. Ranjit, Ben, Phil F, Oli and Caroline for their help and general enthusiasm.

Phillip J Stansfeld has my undying gratitude for responding to 1062 separate CG2AT incidents over the years and still being chirpy most of the time. Thanks Stanners ☺

Goseph Joose also gets some credit for being a role model even though he tries to interfere with my methods (yes, I do need 100 terminals open at any given time).

The rest of SBCB (current and ex) for making the lab fun, including: Alan, Crag Rat, DP, Shabs the Barn Owl, Jussi, J-rome, Mitch (the Uni Club will never be the same), Auntie Janye, DP's beard, and of course Khalid, S (youtube sensation).

Outside of work I would like to thank my family for their support and weekly interrogations about tax-dodging. I am grateful to my parents for getting me to where I am.

Thank you to Liz of Rouch Ltd. for still being here despite writing up doom. You are my compass etc.

Finally, I would like to thank my friends in advance for hopefully relocating me into the wild. Oxford is filled with some Good Eggs. Especial thanks to the people who have kept the final year fun, against the odds – Blubo, Shoulder Angel, Danger (49 %), big r, big k and all the others in Manny Road and OUWAFC.

List of Abbreviations

AT	Atomistic
AT-X	Atomistic simulation of the CG-generated model of the BM2TM domain, where X is the number of doubly-protonated His19 residues
BM2TM	The 27-mer putative transmembrane region of the influenza BM2 protein
BOG	*β*-octylglucoside
CG	Coarse Grain
CG2AT	Coarse-Grained to Atomistic Conversion
CMC	Critical Micelle Concentration
DDM	*β*-dodecylmaltoside
DHPC	Dihexanoylphosphatidylcholine
DPC	Dodecylphosphocholine
DPPC	Dipalmitoylphosphatidylcholine
EM	Energy Minimisation
ESI	Electrospray Ionisation
GpA	Glycophorin A
GNM	Gaussian Network Model
LDAO	Lauryldimethylamine-*N*-oxide
MD	Molecular Dynamics
MS	Mass Spectrometry
N_{agg}	Aggregation Number
NMR	Nuclear Magnetic Resonance
OmpA	Outer Membrane Protein A
PBC	Periodic Boundary Conditions
PDB	Protein Data Bank

PDC	Protein-Detergent Complex
PME	Particle Mesh Ewald
R_g	Radius of Gyration
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
SAXS	Small Angle X-ray Scattering
SMD	Steered Molecular Dynamics
TM	Transmembrane

List of Publications Arising From this Thesis

1) Rouse, SL, Carpenter, TS, Sansom, MSP. Coarse-Grained Molecular Dynamics Simulations of Membrane Proteins. Molecular Simulations and Biomembranes: From Biophysics to Function. *RSC Biomolecular Sciences*. **2010.**

2) Rouse, SL, Carpenter, TS, Stansfeld, PJ, Sansom, MSP. Simulations of the BM2 Transmembrane Domain from Influenza Virus B. *Biochemistry*. **2009**.

Table of Contents

Abstract ii

Acknowledgements iii

List of Abbreviations iv

List of Publications vi

Chapter 1 ..... **1**

Introduction **1**

1. Cell Membranes and Membrane Proteins *1*

1.1. The Hydrophobic Effect 1

1.2. The Fluid Mosaic Model, Lipid-Protein Interactions and Hydrophobic Mismatch, and Lipid Rafts 2

1.3. Membrane Protein Structures and the Two Stage Folding Model 4

2. Structural Studies of Membrane Proteins and the Influence of Non-Native Environments *6*

2.1. Solution NMR Spectroscopy and X-ray Crystallisation: Detergent Micelles 7

2.2. Electrospray Ionisation Mass Spectrometry: "Gas Phase Detergent Micelles" 8

2.3. Two Classes of Membrane Proteins? 8

3. The Influenza M2 Proton Channels *9*

3.1. Structure and Function of the M2 Channels 9

3.2. Drug Inhibition and The HxxxW Motif 11

3.3. Structures of M2 Channels and Conformational Plasticity of A/M2 12

4. Molecular Dynamics Simulations of Membrane Proteins *13*

4.1. Simulations of Membrane Protein Folding 14

4.2. Simulations of Membrane Proteins in Non-Bilayer Environments 15

5. Work Presented in this Thesis *16*

6. References *18*

Chapter 2 ..... **23**

Methods **23**

1. Classical Molecular Dynamics Simulations *23*

1.1. Bonded Interactions 24

1.2. Non-Bonded Interactions 26
1.3. Pair Exclusions 27
1.4. The Particle-Mesh Ewald method 27
1.5. United Atoms 28
1.6. Force Field Parameterisation 28
2. Simulation Techniques 28
2.1. Equations of Motion 29
2.2. Choice of timestep 30
2.3. Periodic Boundary Conditions 30
2.4. Non-Bonded Neighbour lists 31
2.5. Charge Groups 31
2.6. Bond constraints 31
2.7. Positional Restraints 31
2.8. Temperature and pressure control 32
3. Energy Minimisation 33
4. Coarse-Graining 34
4.1. The MARTINI Lipid Model 34
4.2. Extension to Proteins 35
4.3. Advantages and Disadvantages of Coarse-Graining 36
4.4. Multiscale simulations and "CG2AT" 37
4.5. Sidekick 38
5. Non-equilibrium Methods (Steered Molecular Dynamics Simulations) 39
6. Modelling Loops and Homology Modelling 40
7. Detergent Parameterisation 40
8. Software and Analysis 41
9. References 42

Chapter 3 46

The Influenza B Virus Proton Channel 46
1. Introduction 46
1.1. Self-Assembly Simulations 46
1.2. Tetramer Oligomerisation Simulation 48
1.3. Control simulations: Randomized Sequence 53
2. Atomistic Simulations of CG Generated Model 56
2.1. Conversion to Atomistic Representation 56
2.2. Stability of the CG2AT Channel 57

2.3. Potential Gating Mechanism 58
3. Discussion *66*
3.1. CGMD to Model Membrane Protein Folding 66
3.2. Conversion to Atomistic Resolution (CG2AT) 66
3.3. The HxxxW Motif and its Role in Channel Gating 67
4. Further Work *68*
4.1. CGMD to Predict Oligomer State: Pentamer Control Study 68
5. Summary *70*
6. References *71*

Chapter 4 ..... **73**

Solution NMR structure **73**
1. Unusual coiled-coil structure *73*
1.1. Generation of Bilayer System for the NMR Structure 74
1.2. ATMD Simulations of the BM2 Solution NMR Structure 75
2. Discussion *85*
2.1. Structural Comparison of CGMD-Generated Model and Solution NMR Model 85
2.2. Pore Diameter 87
2.3. Orientation of His27 Sidechains in a Bilayer Environment 90
2.4. Solvation of the N-terminal Region 91
2.5. Hydrophobic Mismatch, Lipid Binding Sites, and the Effect of Lipid Ordering 92
2.6. The HxxxW Motif and its Role in Channel Gating 93
3. Further Work *93*
3.1. Generation of a CGMD Model of Residues 1 to 33 93
4. Summary *94*
5. References *96*

Chapter 5 ..... **98**

Protein-Detergent Complexes: Formation **98**
1. Introduction *98*
1.1. Properties of Detergent Micelles 98
2. Methods: Generation of Starting Configurations for Simulations *101*
2.1. Protein-Bilayer systems 101
2.2. Protein-Detergent Micelle Complexes 101
3. Simulation of Detergent-Only Micelles *101*
3.1. Dihexanoylphosphatidylcholine (DHPC) Micelles 101
3.2. β-Dodecylmaltoside (DDM) Micelles 105

4. Protein-Detergent Complexes *109*
4.1. BM2-DHPC Complex Formation 109
4.2. BM2-DDM Complex Formation 113
5. Discussion *118*
5.1. Generation of Protein-Detergent Complexes 118
5.2. Comparison to Experimental Data 119
5.3. Architecture of DHPC and DDM Protein-Detergent Complexes 119
6. Further Work *120*
6.1. Conformational Transitions in DDM Micelles 120
7. Summary *121*
8. References *122*

Chapter 6 123

Protein-Detergent Complexes: Protein Stability and Conformation, Protein-Detergent Interactions, and Detergent Dynamics 123
1.1. The Solution NMR Structure in DHPC Micelles 123
1.2. Stability of the Equilibrated NMR Structure in DHPC and DDM Micelles 125
2. Architecture of Protein-Detergent Complexes *129*
3. Protein-Lipid Interactions & Detergent Binding Modes *133*
4. Lipid Dynamics *137*
5. Water Penetration Into the Hydrophobic Core and Solvation of Serine-rich N-terminal Region *139*
6. Discussion *141*
6.1. Architecture of DHPC and DDM Protein-Detergent Complexes 142
6.2. Protein Conformation, Detergent Packing, and Dynamics 143
6.3. Water Penetration and Solvation of Serine-lined Pore Region 143
7. Further Work *144*
7.1. Gaussian Network Modelling of the Solution NMR Structure of the BM2 TM Domain 144
7.2. Extension to Native Lipid Environment 144
8. Summary *145*
9. References *146*

Chapter 7 147

MD for MS: "Micelles Protect Membrane Protein Complexes from Solution to Vacuum" 147
1. Introduction *147*

2. Methods *150*
2.1. Transfer of protein-detergent complex from solution to vacuum using steered MD 150
2.2. Simulation in "true" vacuum 151
2.3. Simulation at High Temperatures 152
3. Results *152*
3.1. Transfer of Protein-Detergent Complex from Solution to Vacuum 152
4. Extended Simulation Under Dehydrating Conditions *162*
4.1. Dehydration of the BM2-DHPC Complex 163
4.2. Increasing (Collision) Energy 167
5. Discussion *171*
5.1. Methods used 171
5.2. Transition from Solution to Vacuum 171
5.3. Water evaporation 172
5.4. Increasing the "Collision" Energy 173
6. Further Work *174*
7. Summary *176*
8. References *177*

Chapter 8 ..... **180**

Conclusions and Future Directions **180**
1. Introduction *180*
2. Summary *180*
3. Future Directions *182*
3.1. Cross-Validation of Experiment and MD Simulation 182
3.2. Mechanism of Detergent Solubilisation 184
3.3. Modelling Full-Length BM2: Interactions of the Cytoplasmic Domain with Matrix Protein 1 186
4. References *189*

Appendix A ..... **I**
Transfer of Protein-Detergent Complex from Solution to Vacuum ..... *I*
Choice of Parameters for Steered MD I

Chapter 1

Introduction

1. Cell Membranes and Membrane Proteins

For living cells to contain and regulate their contents, it is necessary for a barrier to exist between their external environment and their interior. This barrier is the cell membrane, shown schematically in Figure 1:1A. Biological membranes comprise a complex arrangement of lipids, proteins and carbohydrates. The basis of the membrane is a phospholipid bilayer, formed by two layers of phospholipids aligning such that their polar headgroups form the external surfaces and their acyl chains form a non-polar region in between them (Figure 1:1B and C). The bilayer contains integral (membrane-spanning) membrane proteins as well as peripheral (membrane-associated) and lipid-anchored (membrane-anchored) membrane proteins. These membrane proteins can be broadly grouped into *i)* enzymes, *ii)* receptors, and *iii)* channels and transporters.

1.1. The Hydrophobic Effect

The driving force for bilayer formation is known as the "hydrophobic effect" and is a product of the extensive hydrogen bonding networks formed in liquid water. Addition of a nonpolar molecule to water causes a reordering of the water molecules nearby to give a 'cage' around the solute. This ordering of waters leads to an unfavourable loss in entropy [1]. To minimise this entropic cost, hydrophobic groups are sequestered into aggregates, reducing the numbers of water molecules that are ordered on the surface (as solvent exposed hydrophobic surface is decreased). In the case of bilayer-forming lipids the most favourable type of aggregate formed is a bilayer, however other types of aggregate are possible such as micelles (approximately spherical aggregates of limited

size) and bicelles (bilayered disc-like mixed micelles). The propensity to form discrete aggregates rather than a bilayer is related to the shape of the amphiphilic molecule in question [2]. Generally, a cylindrical shape amphiphile will form bilayers whereas conical or wedge-shaped amphiphiles will form micelles or inverse micelles. The inclusion of non-cylindrical lipids into a membrane can induce curvature. This variation in lateral pressure has been shown to modulate the properties of proteins within the membrane (see *e.g.* [3]).

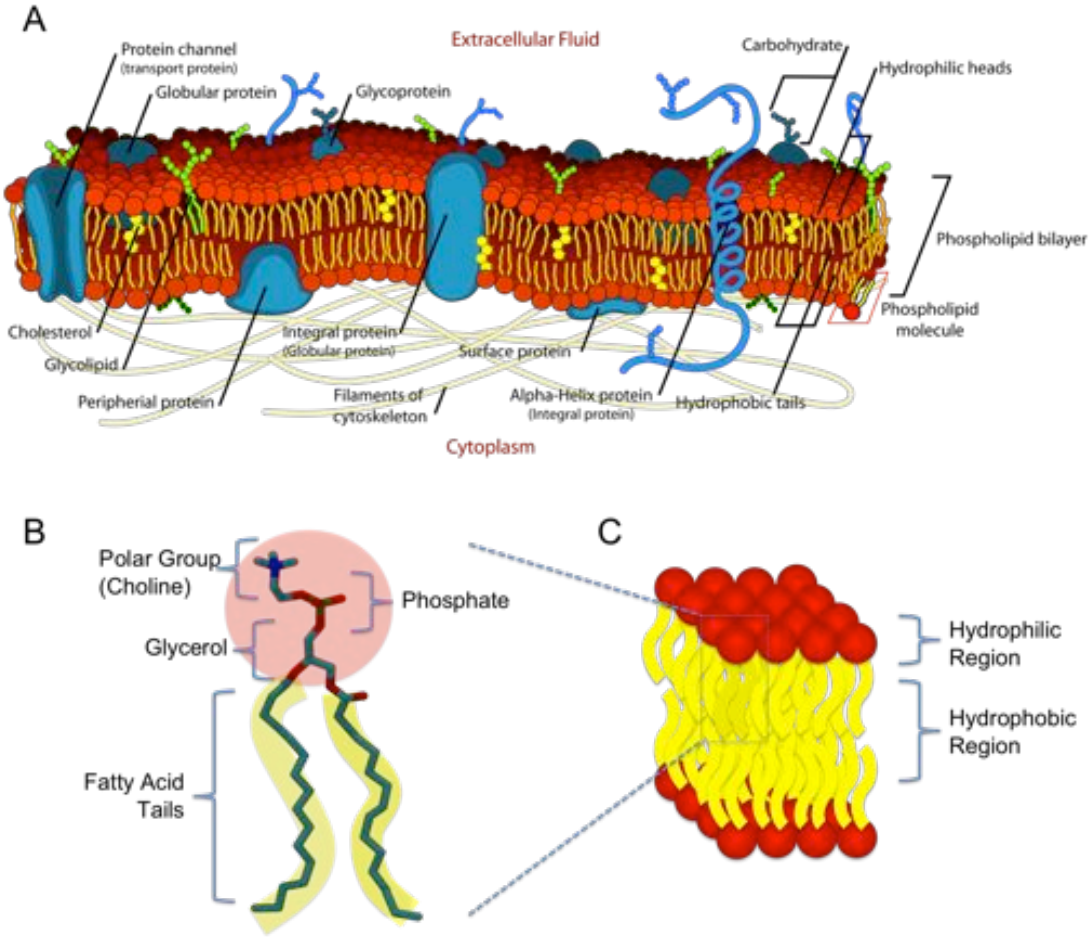


Figure 1:1. The cell membrane and the phospholipid bilayer. A: The Singer and Nicolson Fluid Mosaic Model. Image from Wikimedia Commons. B: Composition of a phospholipid. Phospholipids are built around a central glycerol moiety that is phosphorylated. The phosphate group is then further modified by a polar group, in this case choline. Two fatty acid chains are attached to the glycerol unit. C: Packing of phospholipid molecules into a bilayer system, Such that the polar head groups face outwards while the hydrophobic tails are sequestered into the centre of the bilayer.

1.2. The Fluid Mosaic Model, Lipid-Protein Interactions and Hydrophobic Mismatch, and Lipid Rafts

The understanding of the cell membrane has changed much since it was first described as rigid lipid rods [4]. The so-called fluid mosaic model is the textbook description of the

membrane adopted since Singer and Nicolson published their model in 1972 [5]. Within this model the integral membrane proteins are free to laterally diffuse amongst a fluid lipid bilayer, which acts as a solvent (although within this model it is recognised that some lipids may interact specifically with proteins).

A typical biological membrane contains over a hundred types of lipids, therefore it is improbable that lipids exist simply as a solvent for proteins. The necessity of this vast range in lipid content remains something of an enigma: the importance of specific protein-lipid interactions in several signalling pathways is known; the properties of a lipid within a bilayer have been shown to modulate the thermodynamics of *e.g.* channel gating [6]; and quantitative data on aspects of native membranes drastically affecting the dimer affinity of glycophorin A (a dimeric α-helical membrane protein) were recently published [7]. However, despite the importance of specific interactions in several cases, there is some evidence to suggest that many membrane proteins have evolved to be capable of withstanding changes in the cell membrane [8]. For example, the aquaporin-0 membrane proteins isolated from three domains of life (bacteria, eukaryote and archaea) have been found to have similar structures despite dissimilar lipid compositions of their native membranes [8].

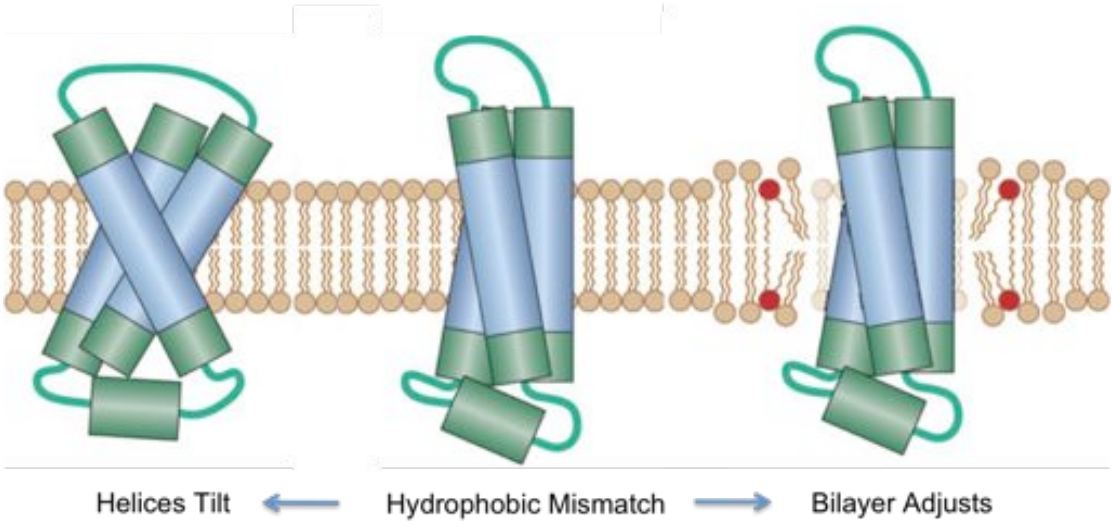


Figure 1:2. Relieving Hydrophobic Mismatch. In the central case the hydrophobic length of the protein (shown in blue) is longer than that of the bilayer hydrophobic core. To alleviate this unfavourable mismatch either the protein may tilt (left most figure) or the bilayer may distort to increase its hydrophobic width close to the protein (right). An alternative reduction in unfavourable mismatch can be achieved by protein aggregation. Figure adapted from [9]

When the hydrophobic region of a membrane protein is not the same thickness as the hydrophobic core of the membrane this is referred to as "hydrophobic mismatch" and is energetically unfavourable. To minimise the unfavourable energetics associated with

this mismatch either the bilayer must extend or compress or the protein must alter its conformation by *e.g.* tilting with respect to the bilayer normal (Figure 1:2).

A recent addition to the picture of cell membranes is the discovery of lipid raft domains. These are nanometre-scale sphingolipid- and cholesterol-rich regions ("rafts") in the membrane [10]. These raft domains are thought to be important in lateral ordering of proteins (*via* hydrophobic matching and/or protein glycosylation) and to have roles in signalling [11]. Some controversy remains over the size, lifetimes and composition of these raft domains *in vivo* [12].

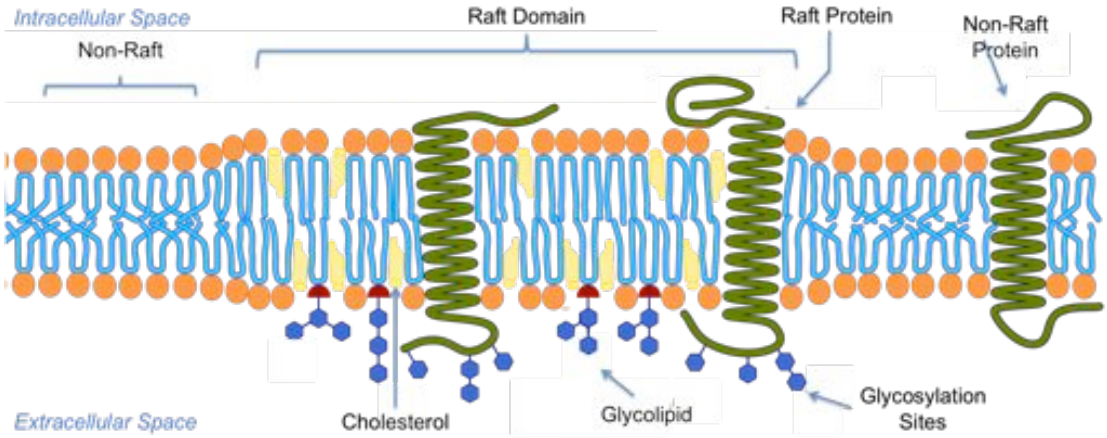


Figure 1:3. Lipid raft domains. The increased cholesterol concentration and bilayer thickness is indicated as well as sorting of membrane proteins into raft and non-raft regions. Figure adapted from Wikimedia Commons.

The depiction of the biological membrane is therefore complex and continues to evolve with the emergence of new data.

1.3. Membrane Protein Structures and the Two Stage Folding Model

Membrane proteins encode ~25 % of the human genome and correspond to ~50 % of drug targets [13–15]. Despite their importance, well-documented difficulties in structure determination have resulted in a disproportionately low number of 3D structures being obtained; at the time of writing, only 862 of the ~70,000 protein structures deposited in the Protein Data Bank [16] correspond to membrane proteins (http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html) [17].

To date, all of the determined structures of membrane proteins fall into two classes: α-helical and β-barrels, shown in Figure 1:4 (although it cannot be ruled out that alternative structures will be discovered as more structures are solved). The β-barrel

proteins are found in the outer membranes of bacteria, mitochondria and chloroplasts [18]. The *α*-helical membrane proteins are generally confined to all membranes apart from the bacterial outer membrane, with a few exceptions (*e.g*. the *E.coli* transporter Wza [19]).

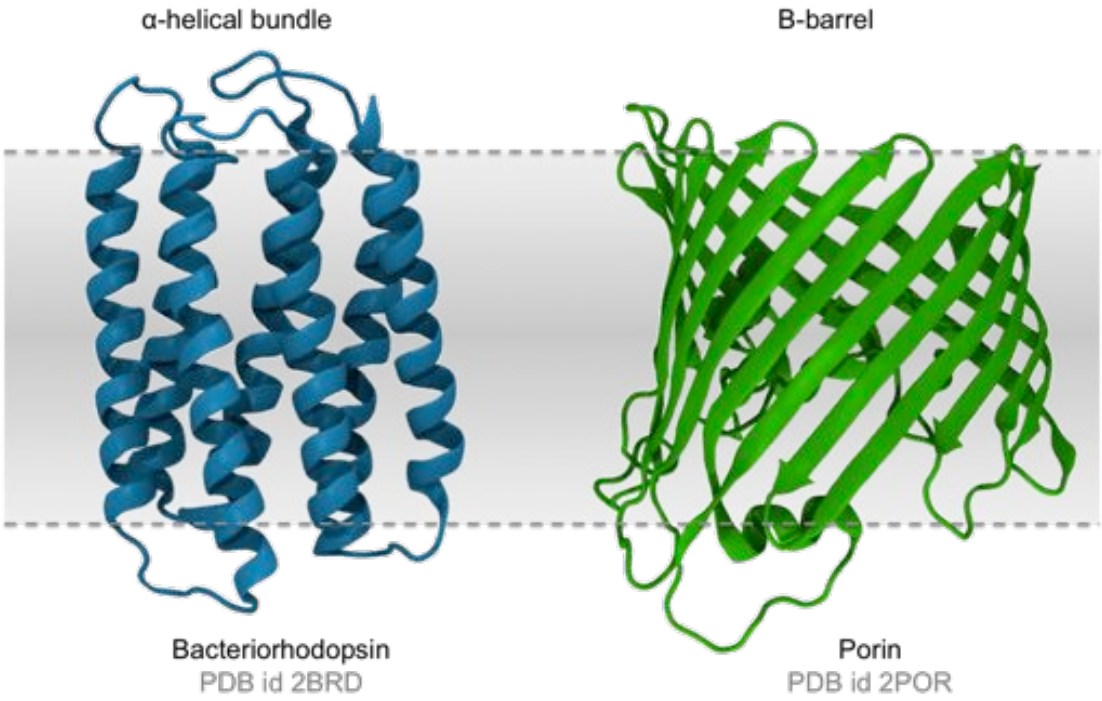


Figure 1:4. The two major architectures of membrane proteins. An *α*-helical bundle is shown in cartoon format in cyan. A monomer of the *β*-barrel porin (trimeric *in vivo*) is shown in green.

The two classes of membrane proteins are thought to fold *via* different pathways. The folding of *α*-helical membrane proteins may be described by the two-stage folding hypothesis [20], whereas folding of *β*-barrels is thought to form *via* a single-stage concerted folding and insertion mechanism [21]. In the two-stage folding model individual helices are inserted into the bilayer *via* transolocon-mediated insertion (Stage I). These helices are stable in their own right, and are able to then rearrange within the bilayer to pack together and give the final tertiary structure (Stage II). This process is shown schematically in Figure 1:5. Support for this model came in the form of studies on bacteriorhodopsin, (Figure 1:4) in which each of its seven TM helices were synthesized, and five of these were shown to be independently stable [22].

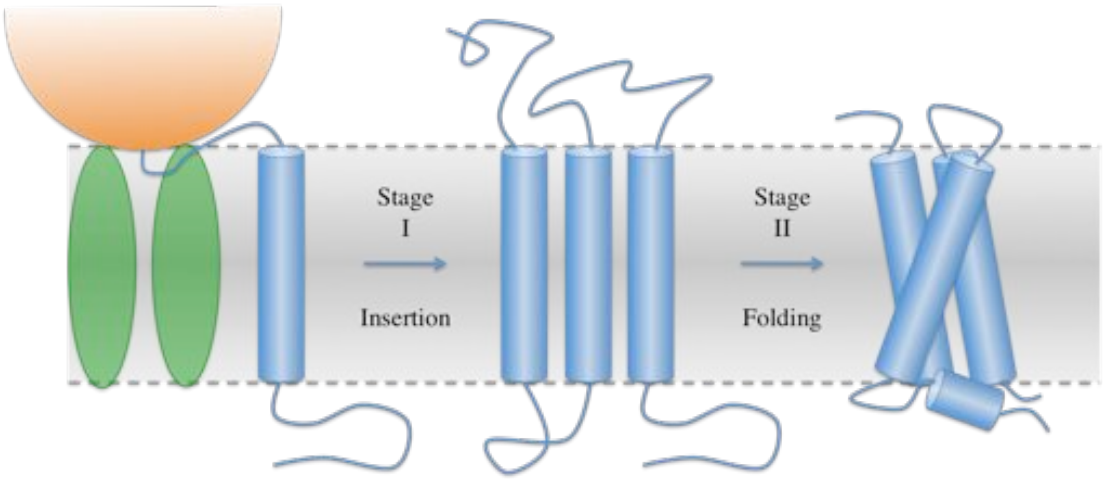


Figure 1:5. The Two Stage Model of α-helical membrane protein folding. In the first stage, individual helices are inserted into the bilayer via interactions with the translocon (shown here in green). In the second stage these helices rearrange to find the global energy minimum – the native structure. Figure based on [21]

2. Structural Studies of Membrane Proteins and the Influence of Non-Native Environments

Anfinsen states in his protein folding hypothesis that "the native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, **in a given environment**" [23]. The final clause may be expected to be especially important for membrane proteins where relatively weak protein-protein interactions are modulated by interactions with lipids [24]. Indeed, the conformational plasticity of many membrane proteins is thought to be necessary for them to perform their roles (*e.g.* protein kinases [25]).

It is often necessary to study membrane proteins in membrane-mimetic environments such as detergents (routinely used in membrane protein purification) or model membranes. In structural studies, detergents are used in solution NMR experiments and crystallisation, while model membranes are required for solid state NMR studies (where the heterogeneous nature of a native membrane is not compatible with the technique) [26]. The question then becomes, how are these environments expected to impact on the conformation and dynamics of a membrane protein compared to the native environment? This is a topic of debate in the literature, with two camps of view at opposite ends of the spectrum. In the following sections two types of environment used in membrane protein structural characterisation are discussed.

2.1. Solution NMR Spectroscopy and X-ray Crystallisation: Detergent Micelles

Detergent micelles form in solutions of detergent over a certain concentration, known as the critical micelle concentration (cmc). Below this concentration the detergent exists as monomers. The concentration of monomers above the cmc remains approximately equal to the cmc; for a typical detergent this monomer concentration is orders of magnitude greater than that of a bilayer-forming lipid [27–29]. A wide range of detergents are commercially available, and the choice of which is dependent on the down-stream applications. For a crystallization study it is important to maintain protein-protein contacts and 'soft' detergents (which tend to form large micelles, e.g. *n*-dodecyl-*β*-maltoside (DDM)) have been found to be most successful [30]. In the case of solution NMR studies, the protein-detergent complex needs to tumble on the timescale of the experiment, requiring the use of small-micelle forming detergents such as the zwitterionic detergents dodecylphosphocholine (DPC) and dihexanoylphosphatidylcholine (DHPC) [31].

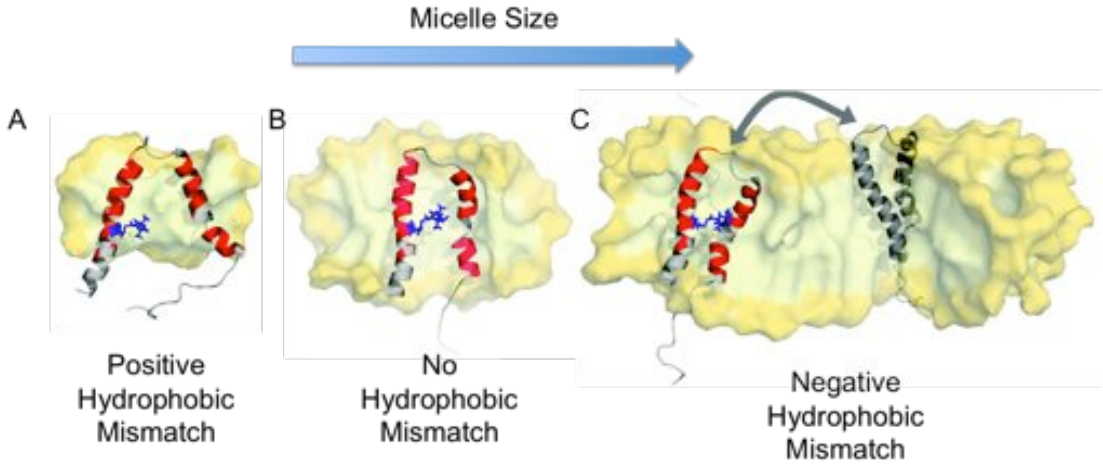


Figure 1:6. Effect of Detergent Micelle Size on TM0026. In A the hydrophobic regions of the helices are longer than that of the small FC micelle and the helices separate in order to bury their hydrophobic regions within the micelle. In B the hydrophobic dimensions of the protein are well matched to those of the detergent micelle (e.g. DM). In C the protein is able to adopt more than one conformation within the large DDM micelle whilst maintaining hydrophobic interactions. Figure adapted from reference [32].

The effect of hydrophobic mismatch on a membrane protein conformation was investigated by a combination of NMR, EPR and SAXS studies [32]. The protein in question was TM0026, a small bacterial membrane protein of unknown function, consisting of two helices joined by a short linker. The conclusion from this work was that to avoid conformational switching (between protein conformation and/or position of the protein within a large detergent micelle) that leads to poor resolution of NMR

data the hydrophobic core of the detergent micelle should be matched to that of the protein (Figure 1:6). This work suggests that the protein arranges to accommodate the hydrophobic length of the micelles. It seems probable that the degree of protein conformational change would be dampened for a less flexible protein and that the micelles would instead vary their hydrophobic thickness.

2.2. Electrospray Ionisation Mass Spectrometry: "Gas Phase Detergent Micelles"

X-ray crystallography and solution NMR spectroscopy are perhaps the most established methods employed in protein structure determination with the majority of structures deposited in the Protein Data Bank being obtained by these techniques (88 % X-ray crystallisation, 11 % solution NMR). A relatively recent technique in mass spectrometry is that of electrospray ionisation (ESI) mass spectrometry. In mass spectrometry, particles are detected as a function of their charge and molecular weight. Recent developments in ESI mass spectrometry have allowed membrane proteins in complex with detergent micelles to be studied in the gas phase (Figure 1:7) – previously thought only possible for soluble proteins [33]. This development paves the way for future studies of membrane protein stability, structure and oligomerisation.

The gas phase environment is very different to that of a cell membrane, and questions have been raised over the conformation of proteins in the gas phase [34]. Evidence is emerging to suggest that aspects of the native solution structure are stabilised in the gas phase [35] (*e.g*. entire virus particles have been shown to retain activity following the electrospray ionisation process [36]). The nature of mass spectroscopic experiment is such that computational methods are one of the only techniques available to attempt to determine and quantify the changes occurring upon transfer from solution to vacuum [37], [38].

2.3. Two Classes of Membrane Proteins?

The two structural classes of membrane proteins (α-helical bundles and β-barrels) were discussed above. It appears that, as well as differing in folding mechanisms and distribution amongst the domains of life, these structural motifs lead to differing sensitivities to their environments. The degree to which each class is influenced remains a matter of debate.

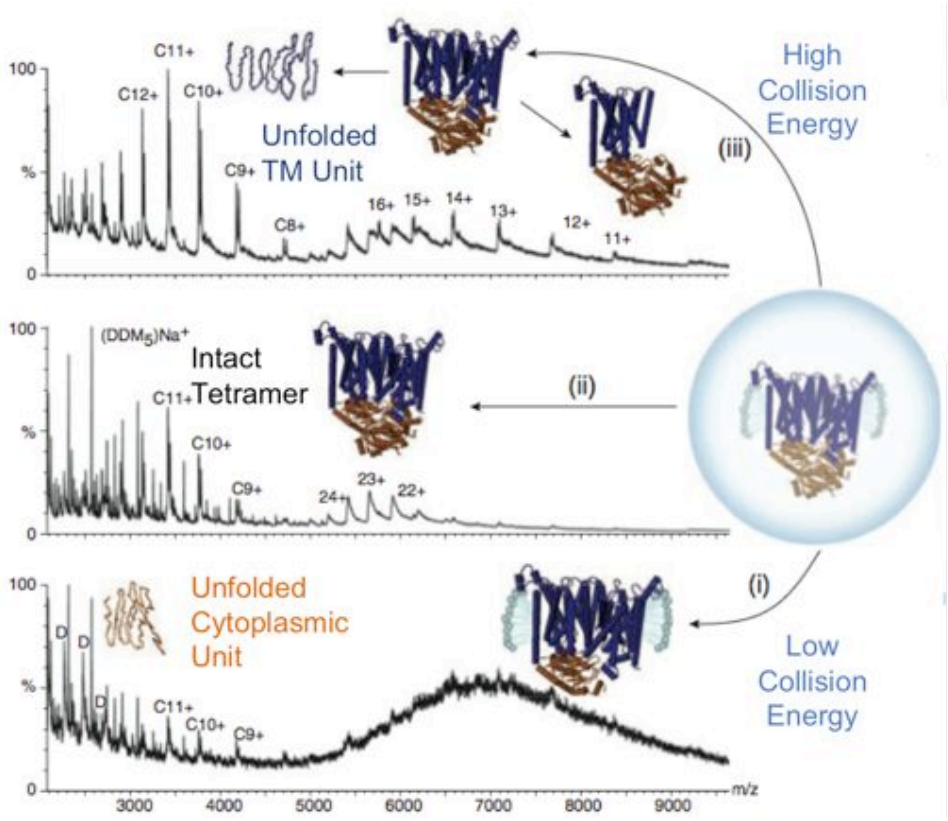


Figure 1:7. Electrospray Ionisation Mass Spectrometry of the bacterial ABC-transporter BtuCD. This protein consists of two cytoplasmic units (shown in orange) and two transmembrane units (shown in blue). The initial BtuCD-detergent complex in a charged nanodroplet of water is shown on the right. Three unfolding pathways under varying activation conditions are shown. Careful manipulation of voltages and collision energies allows detection of the complete tetramer to be determined (pathway ii). When the detergent is removed (at high collision energies, pathway iii) the transmembrane unit is observed to dissociate. However, at low collision energies the detergent remains in complex with the transporter, and the cytoplasmic domain instead unfolds. It appears the presence of detergent therefore stabilizes the transmembrane units in some manner. Figure adapted from reference [33].

3. The Influenza M2 Proton Channels

3.1. Structure and Function of the M2 Channels

The *influenza* A and B viruses contain three types of membrane proteins in their viral envelopes that are essential for viral replication (Figure 1:8). These are hemagglutinin (HA), neuraminidase (NA) and the matrix 2 proton channels (A/M2 in *influenza* A and BM2 in *influenza* B). The *influenza* B virus also contains a non-essential NB membrane protein. The M2 proton channels are small, (97 residues for A/M2, 109 residues for BM2) tetrameric proteins with a small N-terminal ectodomain, a ~20 amino acid transmembrane domain, and a large C-terminal cytoplasmic domain [39]. The BM2 channel is shown in Figure 1:8.

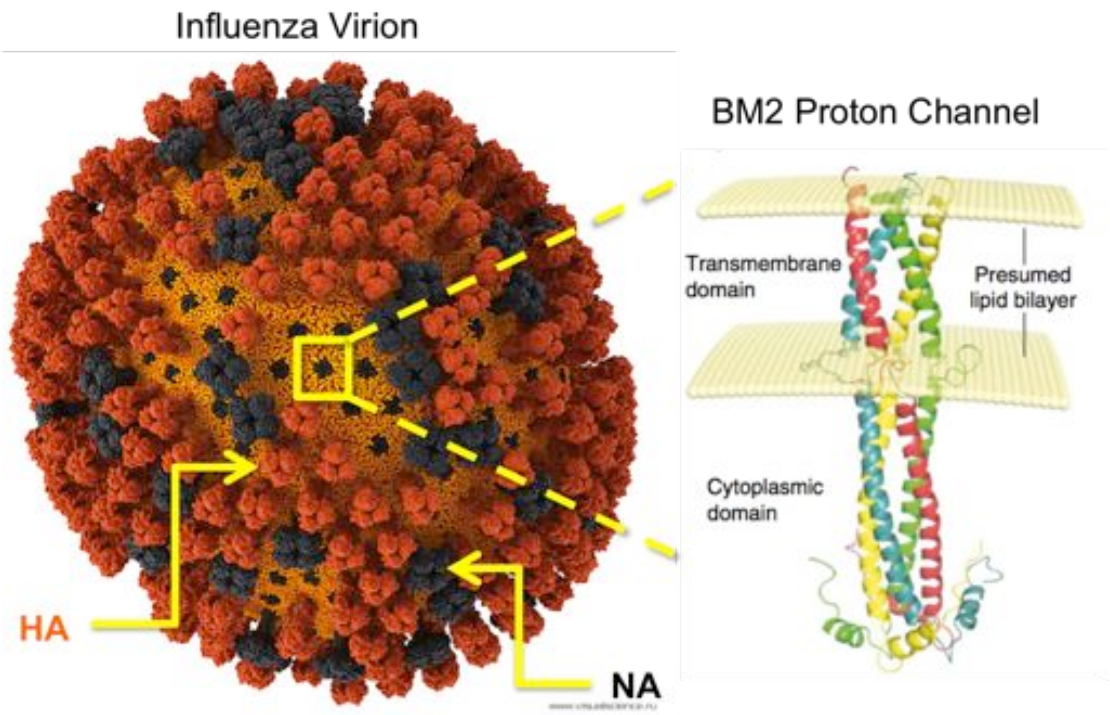


Figure 1:8. The Influenza Virion (left). The membrane lipids are shown in yellow, Hemagglutinnin proteins are shown in dark orange (labelled HA), neuraminidase proteins are shown in black and labelled HA. The M2 proton channels are shown in black and indicated by a box. Figure modified from http://visualscience.eu. Right: the full BM2 channel structure based on two constructs of the transmembrane (PDB id 2KIX) and cytoplasmic domains (PDB id 2KJ1), as solved by NMR spectroscopy. Figure adapted from reference [40].

The transmembrane region is known to function as a proton-specific ion channel that plays a role in two stages of the influenza life cycle (shown in Figure 1:9). The M2 channel is required to allow the interior of the virion to be acidified when the external pH is low. This stage is necessary for the viral genetic material to be released. Without a functioning M2 channel, the *influenza* A and B viruses are unable to replicate [41]. The M2 channels are also necessary in modulating pH of the trans-Golgi apparatus in a later stage of the cycle, preventing premature HA conformational changes .

The cytoplasmic domain of M2 is thought to be important in a later stage of the virus life cycle, where it becomes relocated to the region of viral budding *via* interactions with the M1 matrix protein.[42], [43] The presence of M2 in this region is thought to alter bilayer curvature, allowing budding to occur [44].

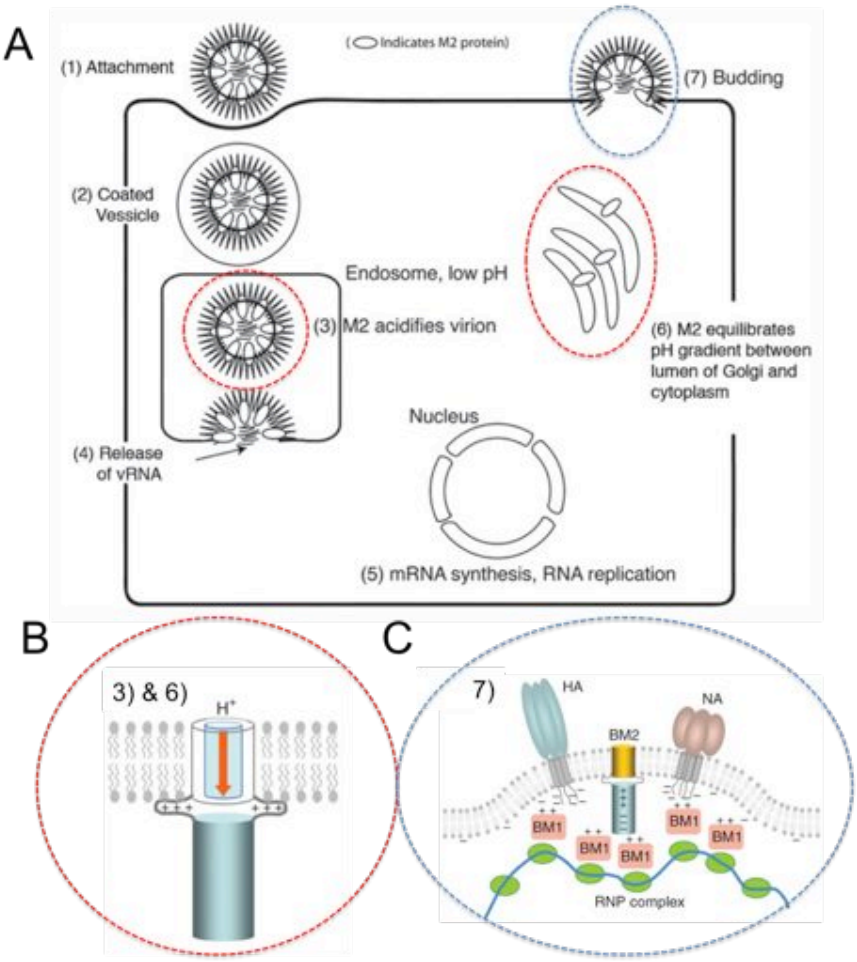


Figure 1:9. The Roles of M2 in the Influenza Virus Life Cycle. A: The life cycle of the influenza virus, adapted from reference [41] .The stages at which the M2 channel has been proposed to be vital for are indicated by dashed circles. B: M2 functioning as a proton channel: M2 is necessary in acidifying the interior of the virion in stage 3, allowing release of genetic material. In a later stage (stage 6) the M2 channel is required to modulate pH and prevent premature hemagglutinnin conformational change. C: The role of the cytoplasmic domain of M2: It has been proposed that interactions of the M2 cytoplasmic domains with the matrix protein M1 are required for correct virus budding. Figures B and C are adapted from reference [40].

3.2. Drug Inhibition and The HxxxW Motif

The A/M2 and BM2 channel are homologous in terms of function, however their only sequence conservation is contained within a single turn of the TM helix – a 5-residue HxxxW motif (Figure 1:10). Mutational studies suggest that the histidine tetrad, containing the only ionisable sidechains conserved between the two channels, is key for gating. Several studies of A/M2 have focussed on elucidating this gating mechanism, with proposals including a transporter-like mechanism [45], a helix-rotation mechanism [46] and histidine/tryptophan side chain 'flipping' [46].

The BM2 channel contains several pore-lining serine residues in the N-terminal region of the TM domain [47], [48]. This has been suggested to be responsible for the differing drug sensitivities of A/M2 and BM2; the A/M2 channel was until recently blocked by the

antiviral drugs amantadine and rimantidine, however the BM2 channel is resistant to both ([49]). The A/M2 channel has become resistant to adamantane-based drugs in recent years thus understanding the differing sensitivities is of interest [50].

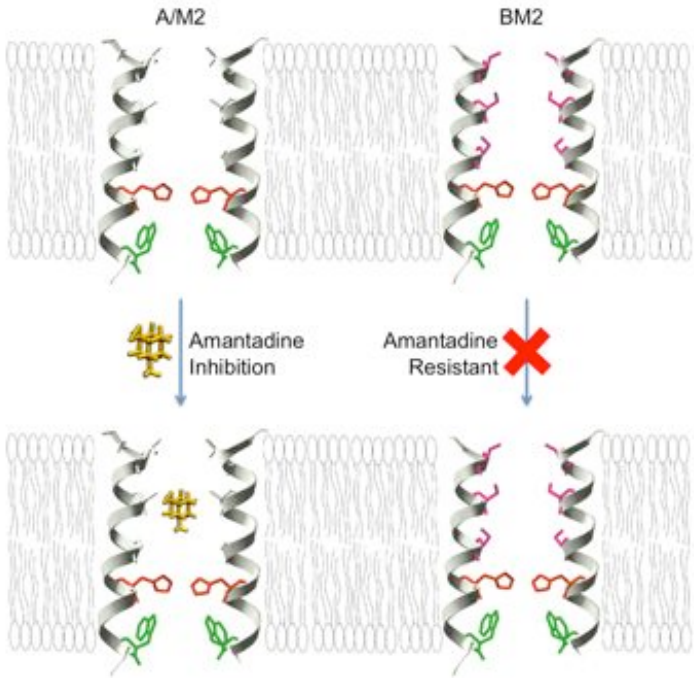


Figure 1:10. The HxxxW gating motif and differing amantadine sensitivities of the influenza A/M2 (left) and BM2 (right) channels. Only two of the four helices are shown for clarity. Key residues are shown explicitly (red = histidine, green = tryptophan, magenta = serine). An amantadine molecule is shown in yellow. Figure adapted from [41]

3.3. Structures of M2 Channels and Conformational Plasticity of A/M2

The A/M2 TM domain is the only protein structure to have been solved using three techniques in three different environments [41]. Some of the structures of the TM domain are shown in [24] along with the single structure of the BM2 channel. There is some variation between the structures solved in each environment. The solution NMR structure (PDB id 2RLF) appears to have the lowest tilt angles, while one of the X-ray crystal structures (PDB id 3BKD) adopts a 'splayed' conformation in which the C-termini are further apart than the N-termini. There are also differences in the crossing point of the helices. It remains to be seen if all of these structures correspond to functioning conformations of the proton channel.

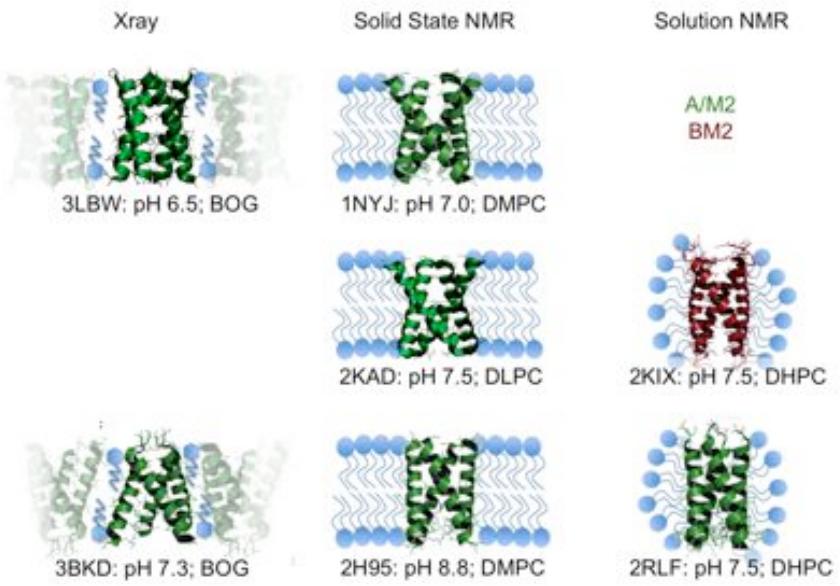


Figure 1:11. Influence of environment on conformation of the M2 proton channel(s). The X-ray structures show detergents packing between adjacent proteins. The solid state and solution NMR show the presence of the bilayer and detergent micelles in cartoon format. The BM2 solution NMR structure is shown in red. This figure is based on [51]

There are some differences in the BM2 and A/M2 solution NMR structures. Both display left-handed packing but while the A/M2 TM domain is a bundle of α-helices, the TM domain of the BM2 channel is proposed to adopt a coiled-coil motif (an alternative explanation being a bend in three of the four helices) [51]. The cytoplasmic domain of BM2 also adopts a coiled-coil configuration, whereas to date no structure is available of the full cytoplasmic domain of A/M2.

4. Molecular Dynamics Simulations of Membrane Proteins

Due in part to the difficulty in obtaining experimental structures of membrane proteins, Molecular Dynamics (MD) simulations have become a powerful tool in understanding their conformational changes and mechanisms. For a recent review of the field of membrane protein simulations see *e.g.* [52]. Here the discussion is limited to two of the areas that are of particular relevance to the work presented in this thesis: *i)* membrane protein folding and *ii)* simulation of proteins in non-native, but experimentally relevant, environments.

4.1. Simulations of Membrane Protein Folding

The two-stage folding model has important implications for computational studies of membrane protein folding as it allows the two stages to be simulated separately. The process of helix-helix association within a bilayer requires timescales that are currently not readily accessible using conventional atomistic MD simulations. However success has been had using coarse-grained (CG) simulations, in which the number of particles is reduced, to achieve longer timescales and larger system sizes. The MARTINI-based CG force field, in which four non-hydrogen atoms are mapped to a single CG particle, has been used to study the folding of glycophorin A (GpA) [53], a paradigm for helix-helix interactions, amongst other proteins [54].

The first stage of membrane protein folding (insertion) may be simulated by the process of a lipid bilayer assembling around a helix to ensure that a given helix is independently stable within a bilayer. The energetics of helix insertion into bilayers may also be studied using potential of mean force (PMF) calculations (*e.g.* [55]). Multiple helices inserted in a TM orientation in a bilayer may then be simulated and are observed to pack together to give the oligomeric protein structure (Figure 1:12). The structures of GpA and A/M2 obtained from this technique matched solution NMR data, to within ~1 Å RMSD of the Cα atoms [53], [56].

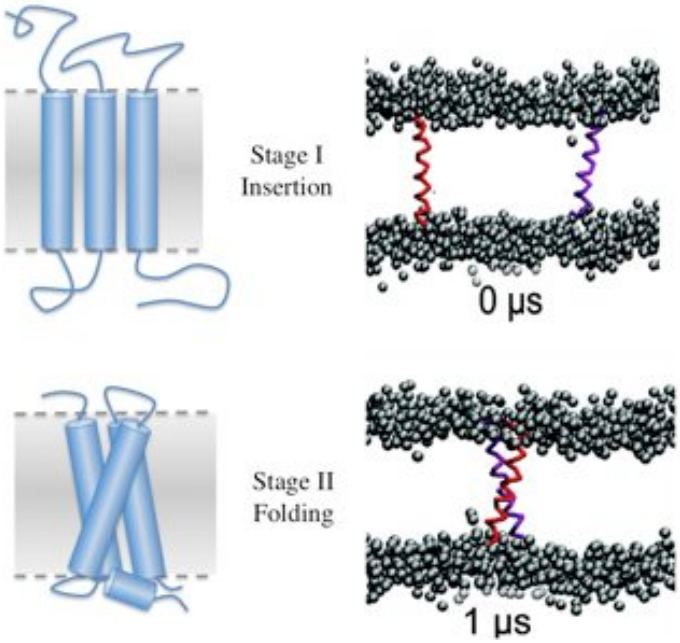


Figure 1:12. Simulations of the Two-Stage Folding Model Molecular Dynamics simulations of glycophorin A folding. RHS snapshots are adapted from [53] The helix backbones are shown as coloured traces. Headgroup particles of the bilayer are shown as grey spheres.

4.2. Simulations of Membrane Proteins in Non-Bilayer Environments

A unique feature of MD simulations is that it is possible for absolute control to be had over the environment - membrane proteins can be transferred from *e.g.* a lipid bilayer to a detergent micelle complex and the conformation and dynamics may be directly compared. However, the vast majority of computational studies of membrane proteins performed in recent years have focussed on simulation in a bilayer environment. There have been a few studies of membrane protein-detergent complexes, with one of the main limiting factors being the large numbers of water particles required to simulate detergent solutions at experimentally relevant concentrations [57–61] .

The discussion in this section is limited to a single protein. Outer membrane protein A (OmpA) is one of the most extensively studied (both experimentally and computationally) membrane proteins, and is one of the few cases where its behaviour in a range of environments has been studied *via* MD simulation (Figure 1:13) [62]. The X-ray structure (PDB id 1BXW) was solved in *n*-tetraoxyethylene detergent [63], while the solution NMR structure (PDB id 2GE4) was solved in DPC detergent micelles [64].

The X-ray structure has been simulated in a DPPC bilayer (model membrane environment) and a DPC micelle (to match the solution NMR structure obtained) [65]. The crystal lattice unit cell (PDB id 1QJP) containing four copies of the protein and BOG detergent was also simulated in a separate study [66]. Finally, the solution phase OmpA-DPC micelle complex was simulated in water shells of varying thickness (from 6 Å to 0 Å) to elucidate structural changes occurring during the process of dehydration in *e.g.* electrospray ionisation [67].

There was some dependence of OmpA dynamics on environment, with the protein found to be more flexible in the micelle environment than the bilayer environment. The simulations of the crystal lattice gave comparable *B* factors to the experimental data. An intriguing observation in the crystal lattice simulation was that of the detergent mobility – in the absence of a bilayer, transverse motion of these detergent molecules was possible. In the simulations of the OmpA-DPC under vacuum conditions the major structural changes were limited to the detergent micelle, which rearranged to give a partially inverted micelle. In all simulations the protein loops were the most flexible region of the protein, whilst the β-barrel residues were relatively independent of environment.

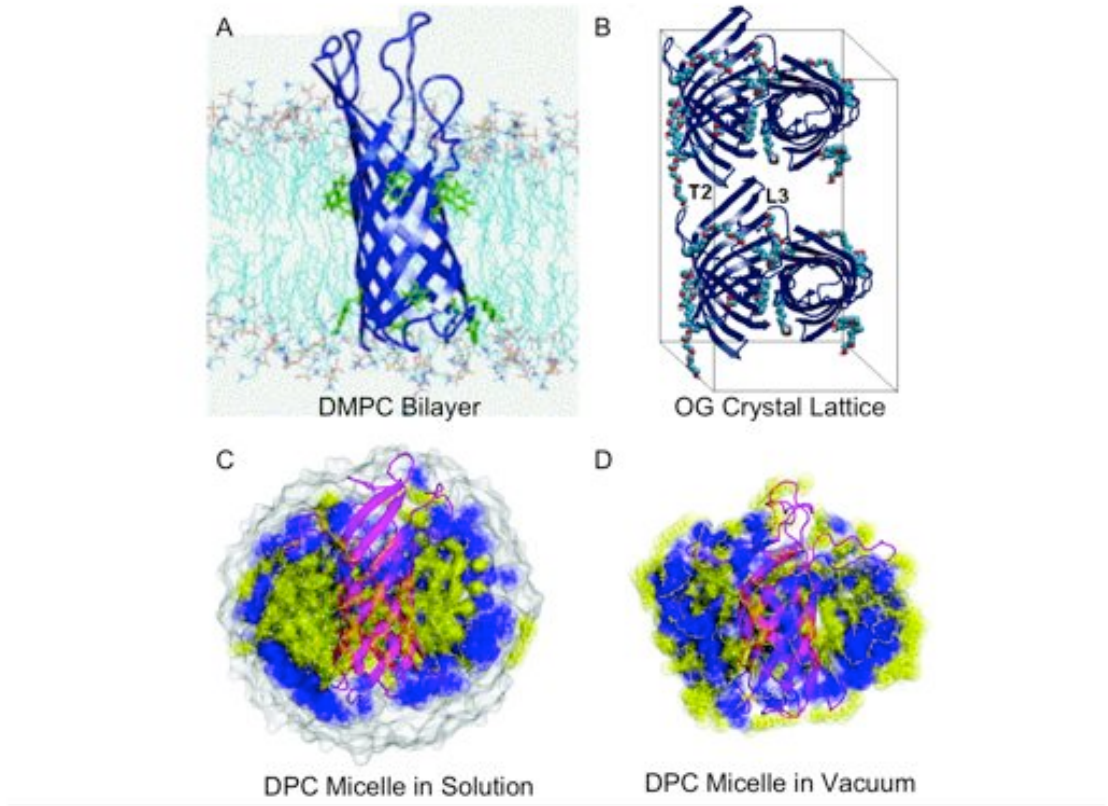


Figure 1:13. Molecular Dynamics studies of OmpA in various environments. A: ATMD simulation in a bilayer. Figure modified from [65] B: Simulation of the OmpA crystal lattice [66]. C: Simulation of OmpA in a DPC micelle. Detergent headgroups are shown in blue, acyl tails in light green. Figure adapted from [67]. D: Simulation of an OmpA-DPC micelle in vacuum (water molecules removed). Coloured as in C. Figure adapted from [67]. In each case the protein is displayed in cartoon format.

5. Work Presented in this Thesis

The broad theme of the work presented in this thesis may be described as simulations of membrane proteins under experimental conditions, that is, **membrane proteins as experimentalists see them**. As discussed previously, MD simulation is a powerful tool in the study of membrane proteins in particular, and the simulations presented in this thesis are a further step towards the use of simulations to inform experiments. The BM2 proton channel has been chosen as the focus of these studies; as a small, α-helical bundle it is an ideal protein for investigating the influence of the surrounding environment. Most previous molecular dynamics studies of the membrane proteins in non-bilayer environments have focussed on β-barrel proteins where the surrounding environment may have less impact.

A brief description of the work presented in each Chapter follows.

Chapter 3: The majority of the thesis is based on simulations of the solution NMR BM2 TM structure. However no structure was available when this work was begun, therefore a model of the BM2 TM domain was generated using the CGMD method described in [67]. This model was then converted to atomistic resolution to explore a potential gating mechanism of the channel. This work was published before release of the solution NMR structure and serves as a further test case for the CGMD folding method. The release of a solution NMR structure by no means makes this work redundant – indeed, comparison with A/M2 would suggest that BM2 is likely to exist in a range of conformations, and the one generated by CGMD in a bilayer is expected to correspond to one of the conformations adopted *in vivo*. The remainder of the thesis is based on simulations of the BM2 solution NMR structure.

Chapter 4: The BM2 solution NMR structure, originally solved in DHPC detergent micelles, was simulated in a DPPC bilayer to assess if this conformation is likely to be adopted in a bilayer environment, and to allow comparison to the CGMD-generated model.

Chapters 5 & 6: Protein-detergent complexes of BM2 with DHPC (a zwitterionic, small-micelle forming detergent commonly used in solution NMR spectroscopy experiments) and DDM (a non-ionic, large-micelle forming detergent commonly used in X-ray and Mass Spectrometry experiments) are generated using a multiscale simulation approach. The simulations were designed to match experimental conditions where possible.

Chapter 7: As discussed above, MD is a useful tool for elucidating the processes that occur during mass spectrometry experiments. Here, MD simulations are used to uncover the changes to protein-detergent complexes upon transfer from bulk solution to vacuum.

Chapter 8: Conclusions and future directions are discussed, including Multiscale simulations of the cytoplasmic domain of BM2 and its interactions with Matrix Protein 1 (BM1).

6. References

[1] E. Fisicaro, C. Compari, and A. Braibanti, "Entropy/enthalpy compensation: hydrophobic effect, micelles and protein complexes," *Physical Chemistry Chemical Physics*, vol. 6, no. 16, p. 4156, Jun. 2004.

[2] C. Tanford, "Micelle shape and size," *The Journal of Physical Chemistry*, vol. 76, no. 21, pp. 3020-3024, Oct. 1972.

[3] R. Phillips, T. Ursell, P. Wiggins, and P. Sens, "Emerging roles for lipids in shaping membrane-protein function.," *Nature*, vol. 459, no. 7245, pp. 379-85, May 2009.

[4] J. D. Robertson, "Membrane structure.," *The Journal of cell biology*, vol. 91, no. 3 Pt 2, p. 189s-204s, Dec. 1981.

[5] S. J. Singer and G. L. Nicolson, "The Fluid Mosaic Model of the Structure of Cell Membranes," *Science*, vol. 175, no. 4023, pp. 720-731, Feb. 1972.

[6] J. Gullingsrud and K. Schulten, "Lipid bilayer pressure profiles and mechanosensitive channel gating.," *Biophysical journal*, vol. 86, no. 6, pp. 3496-509, Jun. 2004.

[7] H. Hong and J. U. Bowie, "Dramatic destabilization of transmembrane helix interactions by features of natural membrane environments.," *Journal of the American Chemical Society*, p. null, Jun. 2011.

[8] C. R. Sanders and K. F. Mittendorf, "Tolerance to Changes in Membrane Lipid Composition as a Selected Trait of Membrane Proteins," *Biochemistry*, no. iii, 2011.

[9] J. U. Bowie, "Solving the membrane protein folding problem.," *Nature*, vol. 438, no. 7068, pp. 581-9, Dec. 2005.

[10] D. Lingwood and K. Simons, "Lipid rafts as a membrane-organizing principle.," *Science (New York, N.Y.)*, vol. 327, no. 5961, pp. 46-50, Jan. 2010.

[11] K. Simons and D. Toomre, "Lipid rafts and signal transduction.," *Nature reviews. Molecular cell biology*, vol. 1, no. 1, pp. 31-9, Oct. 2000.

[12] S. Munro, "Lipid RaftsElusive or Illusive?," *Cell*, vol. 115, no. 4, pp. 377-388, Nov. 2003.

[13] A. L. Hopkins and C. R. Groom, "The druggable genome.," *Nature reviews. Drug discovery*, vol. 1, no. 9, pp. 727-30, Sep. 2002.

[14] A. P. Russ and S. Lampel, "The druggable genome: an update.," *Drug discovery today*, vol. 10, no. 23-24, pp. 1607-10, Dec. 2005.

[15] J. P. Overington, B. Al-Lazikani, and A. L. Hopkins, "How many drug targets are there?," *Nature reviews. Drug discovery*, vol. 5, no. 12, pp. 993-6, Dec. 2006.

[16] H. M Berman et al., "The Protein Data Bank.," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235-242, 2000.

[17] S. H. White, "The progress of membrane protein structure determination.," *Protein science: a publication of the Protein Society*, vol. 13, no. 7, pp. 1948-9, Jul. 2004.

[18] W. C. Wimley, “The versatile beta-barrel membrane protein.,” *Current opinion in structural biology*, vol. 13, no. 4, pp. 404-11, Aug. 2003.

[19] C. Dong et al., “Wza the translocon for E. coli capsular polysaccharides defines a new class of membrane protein.,” *Nature*, vol. 444, no. 7116, pp. 226-9, Nov. 2006.

[20] J. L. Popot and D. M. Engelman, “Membrane protein folding and oligomerization: the two-stage model,” *Biochemistry*, vol. 29, no. 17, pp. 4031-4037, May 1990.

[21] S. H. White and W. C. Wimley, “MEMBRANE PROTEIN FOLDING AND STABILITY: Physical Principles,” *Physiology*, 1999.

[22] J. F. Hunt et al., “A biophysical study of integral membrane protein folding.,” *Biochemistry*, vol. 36, no. 49, pp. 15156-76, Dec. 1997.

[23] C. B. Anfinsen, “Studies on the Principles That Govern the Folding of Protein Chains.” Nobel Foundation, 1973.

[24] T. A. Cross, M. Sharma, M. Yi, and H.-X. Zhou, “Influence of solubilizing environments on membrane protein structures.,” *Trends in biochemical sciences*, vol. 36, no. 2, pp. 117-25, Feb. 2011.

[25] M. Huse and J. Kuriyan, “The Conformational Plasticity of Protein Kinases,” *Cell*, vol. 109, no. 3, pp. 275-282, May 2002.

[26] M. Luckey, *Membrane structural biology: with biochemical and biophysical foundations*. Cambridge University Press, 2008, p. 332.

[27] M. le Maire, P. Champeil, and J. V. Moller, “Interaction of membrane proteins and lipids with solubilizing detergents.,” *Biochimica et biophysica acta*, vol. 1508, no. 1-2, pp. 86-111, Nov. 2000.

[28] D. Linke, “Detergents: an overview.,” *Methods in enzymology*, vol. 463, pp. 603-17, Jan. 2009.

[29] T. Arnold and D. Linke, “The use of detergents to purify membrane proteins.,” *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]*, vol. 4, pp. Unit 4.8.1-4.8.30, Aug. 2008.

[30] S. Newstead, S. Ferrandon, and S. Iwata, “Rationalizing alpha-helical membrane protein crystallization.,” *Protein science: a publication of the Protein Society*, vol. 17, no. 3, pp. 466-72, Mar. 2008.

[31] C. R. Sanders and F. Sönnichsen, “Solution NMR of membrane proteins: practice and challenges.,” *Magnetic resonance in chemistry: MRC*, vol. 44, pp. S24-40, Jul. 2006.

[32] L. Columbus et al., “Mixing and matching detergents for membrane protein NMR structure determination.,” *Journal of the American Chemical Society*, vol. 131, no. 21, pp. 7320-6, Jun. 2009.

[33] N. P. Barrera, N. Di Bartolo, P. J. Booth, and C. V. Robinson, “Micelles protect membrane complexes from solution to vacuum.,” *Science (New York, N.Y.)*, vol. 321, no. 5886, pp. 243-6, Jul. 2008.

[34] J. A. Loo, “Studying noncovalent protein complexes by electrospray ionization mass spectrometry.,” *Mass spectrometry reviews*, vol. 16, no. 1, pp. 1-23.

[35] B. T. Ruotolo and C. V. Robinson, “Aspects of native proteins are retained in vacuum.,” *Current opinion in chemical biology*, vol. 10, no. 5, pp. 402-8, Oct. 2006.

[36] B. Bothner and G. Siuzdak, “Electrospray ionization of a whole virus: analyzing mass, structure, and viability.,” *Chembiochem: a European journal of chemical biology*, vol. 5, no. 3, pp. 258-60, Mar. 2004.

[37] N. P. Barrera and C. V. Robinson, “Advances in the mass spectrometry of membrane proteins: from individual proteins to intact complexes.,” *Annual review of biochemistry*, vol. 80, pp. 247-71, Jun. 2011.

[38] D. van der Spoel, E. G. Marklund, D. S. D. Larsson, and C. Caleman, “Proteins, lipids, and water in the gas phase.,” *Macromolecular bioscience*, vol. 11, no. 1, pp. 50-9, Jan. 2011.

[39] L. H. Pinto and R. A. Lamb, “The M2 proton channels of influenza A and B viruses.,” *The Journal of biological chemistry*, vol. 281, no. 14, pp. 8997-9000, Apr. 2006.

[40] J. Wang, R. M. Pielak, M. A. McClintock, and J. J. Chou, “Solution structure and functional analysis of the influenza B proton channel.,” *Nature structural & molecular biology*, vol. 16, no. 12, pp. 1267-71, Dec. 2009.

[41] L. H. Pinto and R. A. Lamb, “The M2 proton channels of influenza A and B viruses.,” *The Journal of biological chemistry*, vol. 281, no. 14, pp. 8997-9000, Apr. 2006.

[42] B. J. Chen, G. P. Leser, D. Jackson, and R. A. Lamb, “The influenza virus M2 protein cytoplasmic tail interacts with the M1 protein and influences virus assembly at the site of virus budding.,” *Journal of virology*, vol. 82, no. 20, pp. 10059-70, Oct. 2008.

[43] M. Imai, K. Kawasaki, and T. Odagiri, “Cytoplasmic Domain of Influenza B Virus BM2 Protein Plays Critical Roles in Production of Infectious Virus,” *Journal of Virology*, vol. 82, no. 2, pp. 728-739, 2008.

[44] J. S. Rossman and R. A. Lamb, “Influenza virus assembly and budding.,” *Virology*, vol. 411, no. 2, pp. 229-36, Mar. 2011.

[45] E. Khurana, M. Dal, R. Devane, S. Vemparala, W. F. Degrado, and M. L. Klein, “Molecular dynamics calculations suggest a conduction mechanism for the M2 proton channel from influenza A virus,” *Sciences-New York*, vol. 106, no. 4, pp. 1069-1074, 2009.

[46] H. Leonov and I. T. Arkin, “pH-driven helix rotations in the influenza M2 H+ channel: a potential gating mechanism.,” *European biophysics journal: EBJ*, vol. 39, no. 7, pp. 1043-9, Jun. 2010.

[47] M. Sharma et al., “Insight into the mechanism of the influenza A proton channel from a structure in a lipid bilayer.,” *Science (New York, N.Y.)*, vol. 330, no. 6003, pp. 509-12, Oct. 2010.

[48] J. R. Schnell and J. J. Chou, “Structure and mechanism of the M2 proton channel of influenza A virus.,” *Nature*, vol. 451, no. 7178, pp. 591-5, Jan. 2008.

[49] C. Ma et al., “Identification of the pore-lining residues of the BM2 ion channel protein of influenza B virus.,” *The Journal of biological chemistry*, vol. 283, no. 23, pp. 15921-31, Jun. 2008.

[50] R. A. Bright, D. K. Shay, B. Shu, N. J. Cox, and A. I. Klimov, "Adamantane resistance among influenza A viruses isolated early during the 2005-2006 influenza season in the United States.," *JAMA: the journal of the American Medical Association*, vol. 295, no. 8, pp. 891-4, Feb. 2006.

[51] T. A. Cross, "Flu BM2 structure and function.," *Nature structural & molecular biology*, vol. 16, no. 12, pp. 1207-9, Dec. 2009.

[52] M. S. P. Stansfeld, P.J., Sansom, "Simulation Approaches to Membrane Proteins," *Structure (London, England: 1993)*.

[53] E. Psachoulia, P. W. Fowler, P. J. Bond, and M. S. P. Sansom, "Helix-helix interactions in membrane proteins: coarse-grained simulations of glycophorin a helix dimerization.," *Biochemistry*, vol. 47, no. 40, pp. 10503-12, Oct. 2008.

[54] S. Rouse, T. Carpenter, and M. S. P. Sansom, "Coarse-grained Molecular Dynamics Simulations of Membrane Proteins," in *Molecular Simulations and Biomembranes: From Biophysics to Function*, M. Sansom and P. Biggin, Eds. ROYAL SOC CHEMISTRY, 2010, pp. 56-75 .

[55] A. Chetwynd, C. L. Wee, B. A. Hall, and M. S. P. Sansom, "The energetics of transmembrane helix insertion into a lipid bilayer.," *Biophysical journal*, vol. 99, no. 8, pp. 2534-40, Oct. 2010.

[56] T. S. Carpenter, "Simulation Studies of the Influenza M2 Channel Protein," University of Oxford, 2008.

[57] G. Patargias, P. J. Bond, S. S. Deol, and M. S. P. Sansom, "Molecular dynamics simulations of GlpF in a micelle vs in a bilayer: conformational dynamics of a membrane protein as a function of environment.," *The journal of physical chemistry. B*, vol. 109, no. 1, pp. 575-82, Jan. 2005.

[58] E. Psachoulia, P. J. Bond, and M. S. P. Sansom, "MD simulations of Mistic: conformational stability in detergent micelles and water.," *Biochemistry*, vol. 45, no. 30, pp. 9053-8, Aug. 2006.

[59] A. Choutko, A. Glättli, C. Fernández, C. Hilty, K. Wüthrich, and W. F. van Gunsteren, "Membrane protein dynamics in different environments: simulation study of the outer membrane protein X in a lipid bilayer and in a micelle.," *European biophysics journal: EBJ*, vol. 40, no. 1, pp. 39-58, Jan. 2011.

[60] P. J. Bond and M. S. P. Sansom, "Membrane protein dynamics versus environment: simulations of OmpA in a micelle and in a bilayer.," *Journal of molecular biology*, vol. 329, no. 5, pp. 1035-53, Jun. 2003.

[61] R. A. Böckmann and A. Caflisch, "Spontaneous formation of detergent micelles around the outer membrane protein OmpX.," *Biophysical journal*, vol. 88, no. 5, pp. 3191-204, May 2005.

[62] S. Khalid, P. J. Bond, T. Carpenter, and M. S. P. Sansom, "OmpA: gating and dynamics via molecular dynamics simulations.," *Biochimica et biophysica acta*, vol. 1778, no. 9, pp. 1871-80, Sep. 2008.

[63] A. Pautsch and G. E. Schulz, "Structure of the outer membrane protein A transmembrane domain.," *Nature structural biology*, vol. 5, no. 11, pp. 1013-7, Nov. 1998.

[64] T. Cierpicki, B. Liang, L. K. Tamm, and J. H. Bushweller, "Increasing the accuracy of solution NMR structures of membrane proteins by application of residual dipolar couplings. High-

resolution structure of outer membrane protein A.," *Journal of the American Chemical Society*, vol. 128, no. 21, pp. 6947-51, May 2006.

[65] P. J. Bond and M. S. . Sansom, "Membrane Protein Dynamics versus Environment: Simulations of OmpA in a Micelle and in a Bilayer," *Journal of Molecular Biology*, vol. 329, no. 5, pp. 1035-1053, Jun. 2003.

[66] P. J. Bond, J. D. Faraldo-Gómez, S. S. Deol, and M. S. P. Sansom, "Membrane protein dynamics and detergent interactions within a crystal: a simulation study of OmpA.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 25, pp. 9518-23, Jun. 2006.

[67] R. Friemann, D. S. D. Larsson, Y. Wang, and D. van der Spoel, "Molecular dynamics simulations of a membrane protein-micelle complex in vacuo.," *Journal of the American Chemical Society*, vol. 131, no. 46, pp. 16606-7, Nov. 2009.

Chapter 2

Methods

This thesis is based on the application of molecular dynamics (MD) simulations to the study of membrane proteins. In MD simulations the properties of a system as they evolve over time are calculated. The methods used in this work are described in the following sections. A good overview of the methods is found in Leach [1].

1. Classical Molecular Dynamics Simulations

In principle, the behaviour of any biological molecule can be fully described by consideration of the physical interactions between its nuclei and electrons. In practice, a complete, quantum mechanical description of atoms is computationally demanding, limiting its application to that of relatively small groups of atoms *e.g.* enzyme binding sites [2]. To study larger biomolecules with MD the electronic motions are ignored and the energy of the system is calculated as a function of nuclear positions alone [1]. This level of description of the system, in which quantum mechanical effects are left untreated, leads to "empirical force field methods", also known as classical MD simulations [1].

In classical MD simulations the "force field" is the description of interactions between atoms in the system [3]. The force field is a potential energy function $V(\mathbf{r}^N)$, where N is the number of particles and $\mathbf{r}$ is the position of the particles. $V(\mathbf{r}^N)$, is calculated from the sum of contributions from bonded (acting between covalently bound particles) and non-bonded interactions (acting between all atoms):

$$V(\mathbf{r}^N) = \sum V_{bonded} + \sum V_{non-bonded}$$

1.1. Bonded Interactions

The bonded interactions comprise the bond stretching (V_{bonds}), bond angle bending (V_{angles}), and bond torsion (rotation of dihedral angles, $V_{torsions}$,) components. Improper dihedral angles may be used to preserve stereochemistry ($V_{improper}$). The bonded interaction function may be written:

$$\sum V_{bonded} = \sum V_{bonds} + \sum V_{angles} + \sum V_{torsions} + \sum V_{impropers}$$

The bonded interactions are shown in Figure 2:1. The potential energy contributions from bond stretching and angle bending terms are frequently described using Hooke's law, in which the energy varies with the square of the displacement from a reference bond length (b_0) or angle (θ_0) where b_0 is a reference bond length (the bond length when all other terms in the force field are set to zero) [1]. The force constants k_b and k_θ are chosen (specific to a given atom set) to determine the magnitude of the penalty resulting from bond stretching and angle bending, respectively. The bond stretching and angle-bending terms are hard to deform significantly form their reference values, such that most of the conformational and energetic change during a simulation is due to the combination of torsional and non-bonded effects [1].

The torsional term is a four-body interaction and is periodic, usually based on a cosine function. For an alkane chain this potential corresponds to interconversion between the lower-energy "staggered" *anti* and *gauche* conformations via higher-energy "eclipsed" conformations, as shown in Figure 2:2 [4].

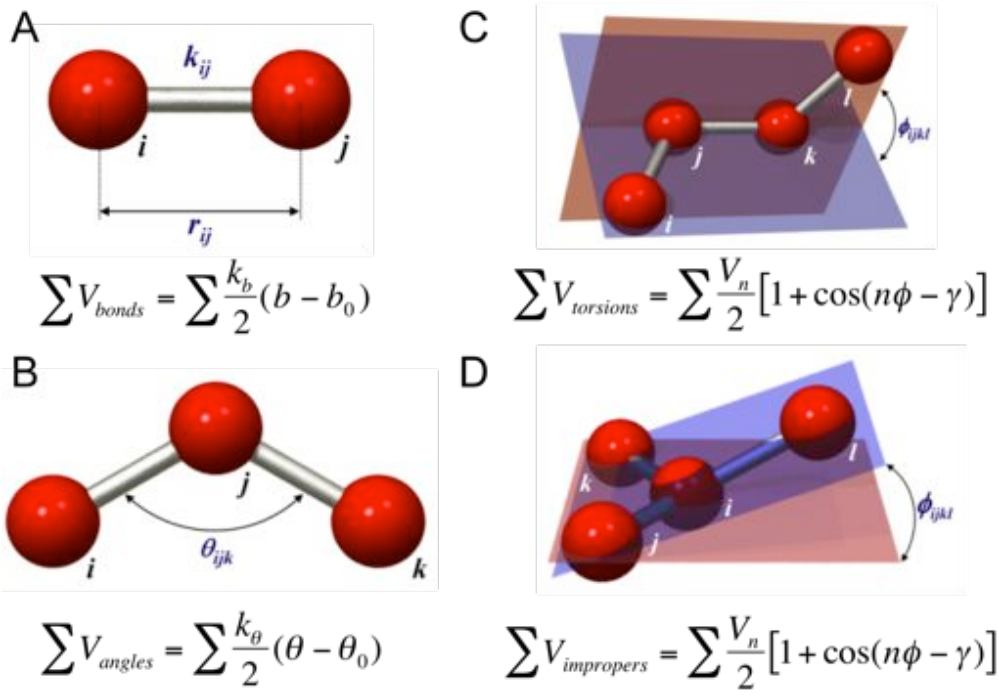


Figure 2:1. Bonded interactions in classical molecular dynamics. A: Bond stretching, generally approximated using a harmonic potential B: Bond angle bending, approximated by a harmonic potential C: Bond dihedral angles, or torsion, a periodic function. The value of *V* indicates qualitatively the relative barrier to rotation. n is the multiplicity (the number of minima when rotated through 360°). D: Improper dihedral angles, may take the same form as proper dihedral angles. Equation parameters are described in the text. Figure adapted from www.pumma.nl/index.php/theory/potentials.

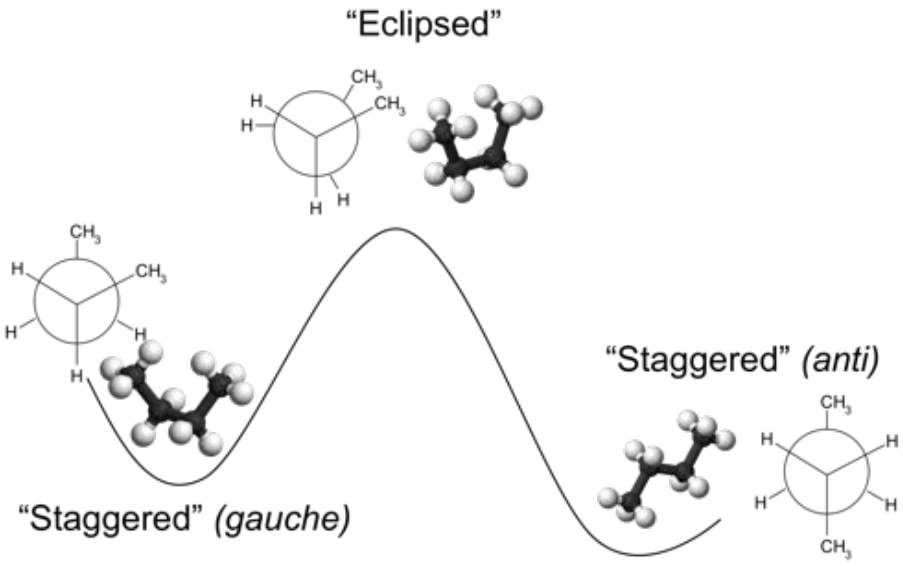


Figure 2:2. Conformations of butane. Fisher projections along with ball-and-stick representations of butane (carbon atoms shown in black, hydrogen atoms in light grey). The approximate potential energy path is shown as a line. The *anti* conformation is slightly lower in energy than the *gauche* conformation. Images adapted from Wikimedia Commons.

1.2. Non-Bonded Interactions

The non-bonded terms may be grouped into Van der Waals interactions and Coulombic interactions:

$$\sum V_{non-bonded} = \sum V_{VDW} + \sum V_{electrostatics}.$$

1.2.1. Van der Waals Interactions (Lennard-Jones Potential)

The Lennard-Jones (LJ) potential is used to describe the V_{VDW} term. This includes both the strong repulsion of atoms sufficiently close in space that their electron densities overlap, and the weak attractive forces (van der Waals attraction) occurring at larger distances, originating from short-range induced-dipole effects (sometimes referred to as the London force). The attractive part of the potential scales as the inverse sixth power of the interatomic distance (r), while the repulsive part scales as the inverse twelfth power of r [1]. The Lennard-Jones potential has the form:

$$\sum V_{VDW} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right]$$

where σ_{ij} is the distance at which the energy between the atoms is zero (*i.e.* at the collision diameter) and εi_j, the strength of interaction (potential energy well depth). These parameters are dependent on the identity of *i* and *j*.

The LJ potential decays with r^6, meaning that it is reasonable to use cut-offs to reduce the number of calculations that must be performed. In the simulations performed in this thesis the LJ potential was cut-off at 12 Å and the potential smoothed gradually from 9 Å to reduce artefacts from the use of straight cut-offs.

1.2.2. Coulombic Interactions

Due to electronegativity differences between non-identical atoms, the electron distribution in most molecules is not distributed evenly. In classical MD force fields this is usually dealt with by assigning fractional point charges (partial charges) to each atom [5]. The Coulombic interaction between (partially) charged particles (*i* and *j*) is calculated using Coulomb's Law:

$$\sum V_{electrostatics} = \sum \frac{q_i q_j}{4\pi\varepsilon r_{ij}}$$

where q_i and q_j are the partial charges of the atoms, r_{ij} is the distance between them and ε is the product of the permittivity of free space ε_0 and the relative permittivity of the medium ε_r. It is evident that the electrostatic effects decay to zero much slower (scaling with r^{-1}) than the Lennard-Jones potential. Therefore careful treatment of long-range electrostatic interactions is an important consideration in MD simulations [3]. An exact treatment of this would require evaluation over all pairs of atoms, such that the number of calculations scales with N^2 where N is the number of particles in the system.

1.3. Pair Exclusions

Non-bonded interactions only apply to pairs of atoms (*i* and *j*) that are either part of distinct molecules or are within the same molecule but separated by at least three covalent bonds. The $i \to i+1$ and $i \to i+2$ pairs are excluded completely, and the $i \to i+3$ pair typically use modified LJ parameters (*e.g.* scaling of the energy parameter by a factor of 0.5 in GROMOS and 0.83 in OPLS). These pair exclusions are necessary to avoid distortions between atoms that are too close in space for their interaction to be described by the Lennard-Jones and electrostatic potentials [5].

1.4. The Particle-Mesh Ewald method

To reduce the computational time necessary to calculate long-range electrostatic contributions, the particle-mesh-Ewald (PME) [6], [7] approach is currently the method of choice in MD [3]. This method is a lattice-based technique and makes use of the periodicity (see section on periodic boundary conditions later) of the simulation box (*i.e.* its use is restricted to systems where it is suitable to impose periodicity). All of the interactions within the infinite lattice are summed. PME is based on the Ewald method, in which the principle is to separate the calculation of electrostatic interactions into a short-range function that converges quickly in real space, and a long-range component that converges quickly in Fourier space. As both components converge quickly in their respective spaces it becomes possible to truncate them (using a cut-off distance for the real space part and a limited number of Fourier modes for the reciprocal part) without introducing large errors. In PME the particle charges are redistributed onto a finite grid,

allowing the use of fast Fourier transform to give faster conversion between real and Fourier space. This reduces the computational scaling to $N\log(N)$ rather than N^2 (where N is number of particles). All ATMD simulations in this thesis use the PME method for electrostatic calculation apart from those indicated in Chapter 7.

1.5. United Atoms

A common way to reduce the number of particles (and therefore the number of non-bonded interactions to calculate) in a system is to implicitly treat the non-polar hydrogen atoms of a molecule by combining them with their parent carbon atom into a single interaction site, or 'united atom'. For the lipid DPPC used in this thesis, this reduces the number of particles explicitly described from 130 to 50. An unwanted consequence of this is that chiral centres may be able to invert during a simulation, however improper dihedrals may be used to prevent this [1]. Three force fields are used within this thesis: GROMOS96 [8], [9], OPLS-UA [10] (both united-atom force fields) and OPLS-AA [11], [12] (an all-atom force field).

1.6. Force Field Parameterisation

The main force fields used in MD simulations of phospholipid membranes are CHARMM [13], GROMOS96 [9], and a united atom force field known as the Berger lipid set [14], based on a combination of OPLS and AMBER united atom versions with some terms taken from an older version of GROMOS. Force fields parameters are generally adjusted to match experimental data (density and thermodynamic parameters) and quantum mechanical calculations [1]. Each force field is developed based on its own data set and it is considered unwise to attempt to use parameters from one force field set with another [3]. Ideally simulations should only be performed under simulation conditions under which the original parameters were developed.

2. Simulation Techniques

In the previous sections the basis of the potential has been described. In this section describes the simulation stages (of a given force field). The general MD algorithm is shown in Figure 2:3. In principle the change in position of atoms over time can be calculated from the initial positions and velocities of atoms [5] .

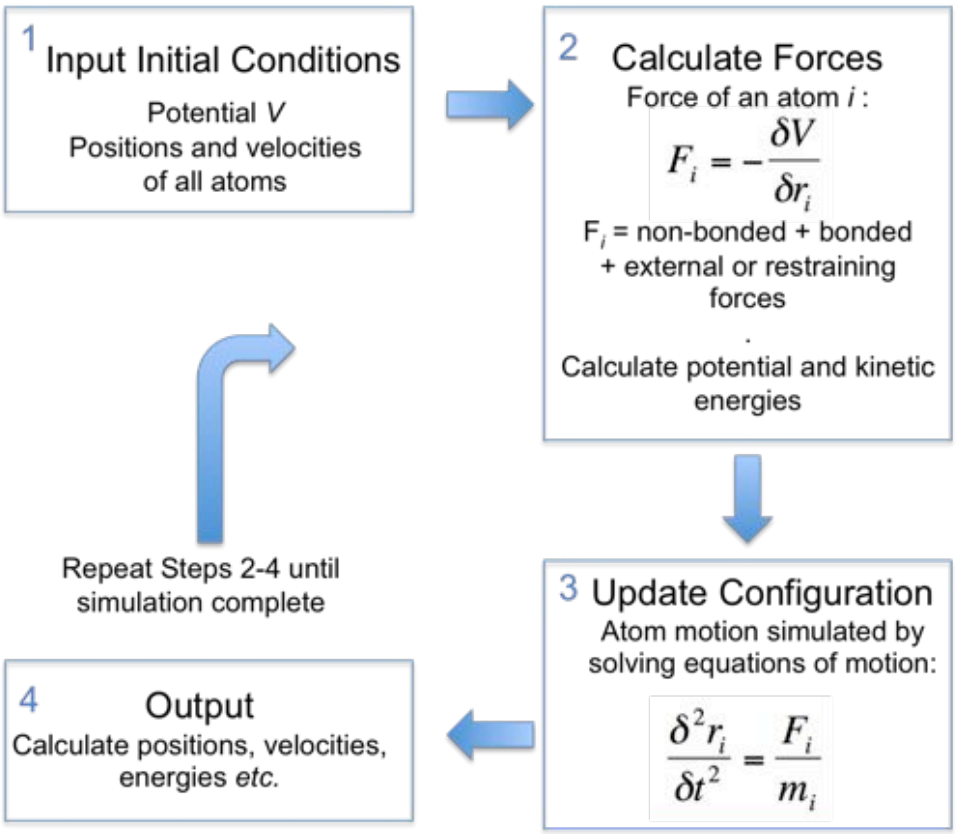


Figure 2:3. The MD algorithm. If no velocities are input they are randomly generated from a Maxwell-Boltzmann distribution at a chosen temperature. Figure based on [5]

2.1. Equations of Motion

In MD simulations, the trajectory is obtained by solving Newton's second law, that force is equal to the rate of change of momentum ($F=ma$):

$$\frac{d^2 x_i}{dt^2} = \frac{F_{x_i}}{m_i}$$

to give the motion of particle (mass m_i) along coordinate x_i where F_{xi} is the force on the particle in the x-direction [5]. In the many-body systems studied in MD it is not possible to solve the equations of motion analytically because the motion of all particles is coupled. To solve these equations a finite difference method is employed [1].

The finite difference method used in GROMACS is based on the Verlet algorithm and is called the leap-frog algorithm [15]. This uses the following terms:

$$\mathbf{v}(t + \frac{1}{2}\Delta t) = \mathbf{v}(t - \frac{1}{2}\Delta t) + \frac{\Delta t}{m}\mathbf{F}(t)$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \Delta t \mathbf{v}(t + \frac{1}{2}\Delta t)$$

2.2. Choice of timestep

The choice of timestep is important. As the timestep tends towards zero the integration error is reduced but the simulation time sampled is reduced. However, if the timestep is too large instabilities may occur, leading to the non-conservation of momentum. In general the timestep should be approximately a tenth of the time of the fastest motion (in an unconstrained system usually corresponding to the C-H bonds which vibrate with a repeat period of ~10 fs) [1]. In the ATMD simulations presented in this thesis a timestep of 2 fs was used, with bonds to hydrogen atoms constrained.

2.3. Periodic Boundary Conditions

Consider a system of a lipid bilayer in a simulation box of water. All those molecules at the boundaries of the simulation box are in contact with vacuum. This is unphysical and leads to unwanted edge effects. To eliminate this problem, periodic boundary conditions (PBCs) are imposed, in which infinite copies of the simulation box are extended in each direction. This means that none of the system is exposed to vacuum. The use of PBCs has been suggested to lead to other, unwanted artefacts such as increased ordering of a system [3]. There are other concerns particular to the simulation of membrane systems. The natural fluctuations of membranes are forced to fit within the simulation box, meaning that using MD simulations to study long-range motions is difficult [16]. A further concern is in asymmetric systems (*e.g.* with peptide insertion in one leaflet, or asymmetric lipid composition of the two leaflets as in bacterial membranes) where care must be taken to ensure that properties such as membrane curvature are not caused due to the periodic constraint of a continuous bilayer.

GROMACS uses periodic boundary conditions with the minimum image convention. This means that only particles in the nearest interacting box are considered in calculation of short-range non-bonded interactions. (This is accurate enough for short-range interactions but long-range interactions are generally treated with PME and include *all* images, as was described earlier.)

2.4. Non-Bonded Neighbour lists

The use of cut-offs in the calculation of the non-bonded Lennard-Jones interactions to reduce the number of calculations that must be performed has already been discussed. However, this approach in itself may not lead to a large decline in calculation time. This is because the positions of each particle relative to each other (to decide if two particles are within the cut-off) must still be calculated. The non-bonded neighbour list works by storing all those particles within the cut-off distance of each other (and those slightly further away). This allows the interactions between the nearest-neighbours to be calculated without needing to calculate the positions of every atom first. The neighbour list is usually updated every 10 or 20 molecular dynamics steps [5].

2.5. Charge Groups

Charge groups were originally developed to counter some of the effects of using plain cut-offs. With the development of PME these no longer necessarily lead to greater accuracy but are still useful in maximising the efficiency of simulations [5]. Charge groups are small groups of atoms (typically three or four) that are assigned a partial charge. When calculating non-bonded interactions, only those between charge groups that fall within a cut-off are calculated. In this case, the non-bonded interactions between all atoms of each charge group are calculated.

2.6. Bond constraints

As discussed previously, the timestep used in MD simulations should be less than that of the fastest motion in the system. To maximise the timestep possible it is common practice to constrain the bond lengths of the fastest vibrating bonds (bonds containing hydrogen). There are two well-known algorithms implemented in GROMACS to achieve this, LINCS [17] and SHAKE [18]. In the OPLS-AA simulations presented in this thesis, the faster and more stable LINCS algorithm was used to constrain hydrogen bonds. LINCS resets bonds to their correct lengths after an unconstrained update.

2.7. Positional Restraints

Restraints may also be used within a MD simulation. Restraints impose an energetic penalty upon deviation from the reference value (whereas a constraint does not allow

the value to change) [5]. Positional restraints (PR) are commonly used to preserve a protein structure while lipids and solvent are equilibrating around it. In all ATMD simulations presented in this thesis 2 ns pre-production MD runs were performed with positional restraints on the protein (unless stated otherwise).

2.8. Temperature and pressure control

A conventional MD simulation as described so far reproduces a constant-*NVE* (microcanonical) ensemble, where the number of particles (*N*), the volume (*V*) and the energy (*E*) of the system remain constant [1]. However, biological experiments are generally performed under constant temperatures and pressures and therefore an isothermal-isobaric system (*NPT*) is often preferred. This may be achieved by applying thermostat and barostat algorithms to reproduce constant-*NVT* and constant-*NVP* ensembles. The majority of the simulations presented in this thesis were performed under the constant-*NPT* regime and systems were coupled to a Berendsen [19] barostat and thermostat unless stated otherwise.

2.8.1. The Berendsen Thermostat

The system temperature is connected to the kinetic energies (*K*) of individual particles:

$$K_{NVT} = \frac{1}{2}\sum m_i v_i^2 = \frac{3}{2} N k_B T$$

The simplest way to control temperature is therefore to scale the velocities. In the Berendsen thermostat the system is coupled to an external heat bath that is fixed at the required temperature (T_{bath}). The velocities of each particle are then scaled at each step at a rate proportional to the difference between the temperature of the component (in this work the system is divided into the three components; protein, lipid/detergent and solvent (water including ions)) and T_{bath}:

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

where *T(t)* is the temperature of the system component (*e.g.* protein) at time *t*. The value of the coupling parameter τ_t determines how tightly the system is coupled to the bath.

The Berendsen thermostat suppresses thermal fluctuations, leading to incorrect sampling, an error which scales as $1/N$. For smaller systems the velocity rescaling thermostat can be more accurate. It is similar to the Berendsen thermostat but contains an additional stochastic term, ensuring the correct kinectic energy distribution. The velocity rescaling thermostat [20] was adopted for the vacuum simulations in Chapter 7.

2.8.2. The Berendsen Barostat

The simplest way to control the pressure of a system is to scale the volume. Analogously to the isothermal set-up, the pressure can be kept constant by coupling the system to a 'pressure bath'. The rate of change of the pressure is given by:

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

where $P(t)$ is the pressure of the component at time t, τ_p is the coupling constant, P_{bath} is the required pressure (the pressure of the pressure bath). The volume of the box is scaled, equivalent to scaling the atomic coordinates and consequently the inter-particle forces.

In bilayer-containing simulations it is preferable to maintain the surface area per lipid. Semi-isotropic pressure coupling along the plane of the bilayer (*xy*-plane) is used, such that the box length vectors *x* and *y* vary together, while *z* varies independently. A reference pressure of 1 bar in *xy* and *z* is used to correspond to realistic pressures [3]. In non-bilayer atomistic simulations the pressure coupling was isotropic (*x*, *y*, and *z* vary together) and again reference pressures of 1 bar were chosen.

3. Energy Minimisation

Energy minimization is an essential step prior to beginning a simulation [1]. It is used to reduce steric clashes in the system and give a stable starting configuration. The two EM methods employed in this thesis are the steepest descent and conjugate gradient methods. Both of these methods involve gradually changing the coordinates of atoms, reducing the energy of the system to the minimum *via* an iterative procedure. The method of steepest descent reaches the minimum most effectively when the system is far from the energy minimum. The conjugate gradient method bases the next step on the previous one and is more effective when the system is near the energy minimum [1].

For the CGMD simulations described steepest descent was used. For the ATMD simulations a minimum of 5000 steepest descent steps with a conjugate gradient step every five iterations was used.

4. Coarse-Graining

Many groups have developed coarse-grained force fields in which the system complexity is reduced to allow longer spatial and temporal scales to be accessed by simulation. These vary in the degree of coarse-graining, *e.g.* the MARTINI force field (~4 non-hydrogen atoms per CG particle) [21], residue-based coarse grain models (entire amino acid residues modelled by single CG particle) [22] and shape-based coarse-graining (~150 atoms per CG particle) [23]. The reduction in system complexity means that fewer calculations are required and longer timesteps may be used, leading to an overall speed-up of 10^2-10^3 compared to conventional ATMD simulation.

In the work presented in this thesis, two variants of the Martini-based force field are used – the MARTINI force field [21] and the Bond force field [24]. There are slight differences between them, which are discussed further below, although there is no compelling evidence to suggest that either is 'better'. Some evidence suggests that while the Bond force field is more accurate in prediction of insertion efficiencies of peptides, the MARTINI force field alone is more able to distinguish between subtle mutations (Hall, unpublished results). The Bond force field has been extensively used in membrane protein simulations with success – *e.g.* key lipid contacts based on experimental data were reproduced [25]. It has also been used to predict bilayer positions for all membrane proteins in the Protein Data Bank [26]. The majority of simulations were performed using the Bond force field. The simulations of DDM in Chapter 3 used the MARTINI force field to be consistent with the carbohydrate parameters the DDM parameters were based on. Figure 2:4 shows the 4-to-1 mapping of the Bond force field.

4.1. The MARTINI Lipid Model

The topology of lipids is the same in both Bond and MARTINI force fields and it is only in more recent additions to the MARTINI force field (proteins and carbohydrates) where they vary.

A general feature of both force fields is that only a few levels of Lennard-Jones interaction energies are used. In the original lipid model only four particle types are

used: *i*) **P** - polar (neutral groups), *ii*) **N**-non-polar (mixed polar-apolar groups), *iii*) **C**-apolar (hydrophobic) groups and *iv*) **Q**- charged (ionised groups). Particle types N and Q are further described by subgroups to indicate hydrogen bonding potential. The subgroups are *0* (no hydrogen bonding capability), *a* (hydrogen bond acceptor) *d* (hydrogen bond donor) *da* (could act as either a hydrogen bond acceptor or donor). The standard Lennard-Jones potential is used to describe non-bonded interactions, but only five levels of interaction energy (ε_{ij}) are defined: attractive (ε_{ij} = 5 kJ mol^{-1}, corresponding to polar water-water interactions); semi-attractive (ε_{ij} = 4.2 kJ mol^{-1}); intermediate (ε_{ij} = 3.4 kJ mol^{-1}, non-polar interactions in hydrophobic chains); semi-repulsive (ε_{ij} = 2.6 kJ mol^{-1}); and repulsive (ε_{ij} = 1.8 kJ mol^{-1}, to model hydrophobic repulsion between polar and non-polar phases). The values were calibrated using a water/alkane reference system. The more recent version of the MARTINI force field includes a greater number of particle types, allowing finer tuning of the coarse-grained particles to closer match the atomistic counterparts.

4.2. Extension to Proteins

The Bond and MARTINI force fields were developed from the original CG lipid model to include proteins independently, leading to some differences in i*)* Lennard-Jones energy levels *ii)* particle subtypes and *iii)* secondary structure description.

4.2.1. The Bond Force Field

The number of particles assigned to each amino acid side chain was based on optimising the volume of CG particles compared to its AT counterpart. The particle types assigned to protein residues were chosen by optimizing the free energies of transfer of the individual amino acids from water to cyclohexane to match experimental data. For proteins, the secondary structure is maintained by applying a harmonic restraint, or 'artificial hydrogen bond' between residues *i* and *i*+4 (except when *i*+4 is proline). The backbone particle type and secondary structure is based on the original atomistic structure input.

4.2.2. The MARTINI Force Field

In the Bond force field only one size of particle type is used whereas in MARTINI smaller particle types for ring particles (*e.g.* in histidine sidechains) have been included, leading to slight differences in the mapping of CG particles of side chains. The choice of particle

types for each amino acid were based on partitioning free energies between polar and apolar phases. More Lennard-Jones interaction energy levels are available although these cover approximately the same range of values (5.6-1.0 kJ mol^{-1} for MARTINI compared to 5.0 – 1.8 kJ mol^{-1} for Bond). For proteins, secondary structure restraints are again applied; however these take the form of dihedral angle restraints on the peptide backbone, as opposed to the *i*-(*i*+4) distance restraints used with the Bond force field. Unlike the Bond force field, there is no break in secondary structure due to the presence of proline residues.

4.3. Advantages and Disadvantages of Coarse-Graining

As stated above, the advantage of coarse graining is the large increase in system sizes and timescales that become accessible to simulation. Whichever force field is chosen out of Bond or MARTINI force field, a considerable increase in system size and timescales of simulation may be achieved.

Only short-range Lennard-Jones interactions are calculated (although it is possible to combine PME electrostatics with CG-MD this is not usually done) such that the number of calculations is reduced not only by the decrease in number of particles but also by the number of non-bonded interactions that must be calculated. The timescales attainable are further increased because the smoother energy potentials in the CG models allow for longer timesteps (typically 40 fs with Bond and 20 fs with MARTINI). There has been some discussion on the acceptable timestep limit but most evidence suggests that 40 fs is generally suitable for CG (but may be reduced when necessary).

The absence of hydrogen bonding potential in CG simulations requires the use of secondary structure restraints as described above. Coarse-grained simulations are (currently) limited therefore by the lack of secondary structure transitions, as well as the more obvious limitation of lowered resolution.

Due to the simplified interactions in the CGMD method, as well as the secondary structure restraints it is common to maintain the tertiary structure by applying a Gaussian Network Model (GNM) to the backbone particles [27]. This maintains the overall structure but allows some conformational change during a simulation.

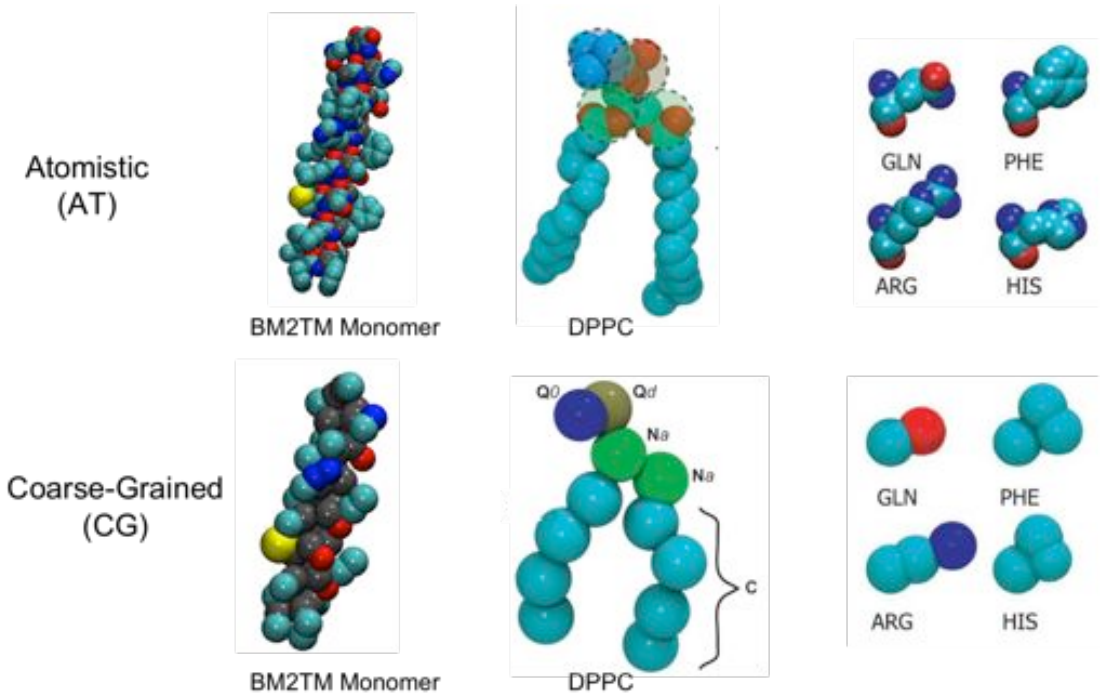


Figure 2:4. The 4-to-1 mapping of the Bond v2 CGMD force field used in this thesis. The BM2TM monomer is the transmembrane region of the BM2 channel used in Chapter 3. Simulations of bilayer systems in this thesis used DPPC. Four examples of amino acid residue mapping are shown.

4.4. Multiscale simulations and "CG2AT"

Of course, the use of CGMD leads to a loss in detail. To study properties that require atom-level description of the system (*e.g.* water penetration) it is useful to be able to convert the CG system to an atomistic representation. Originally this was achieved by mapping conformations of AT lipid molecules (from a library of thousands of lipids) to each CG lipid and choosing the one of closest fit [28]. Recently this somewhat laborious approach has been made redundant by a method that has been developed within the lab to substantially improve conversion of CG to AT ("CG2AT"). The method is fully described elsewhere [29] but involves a fragment-based fitting approach, leading to much faster conversion and a well-equilibrated ATMD system. Within this method the protein is fitted to the CG model using PULCHRA [30]. Some of the coarse-grain fragments and their atomistic counterparts are shown in Figure 2:5.

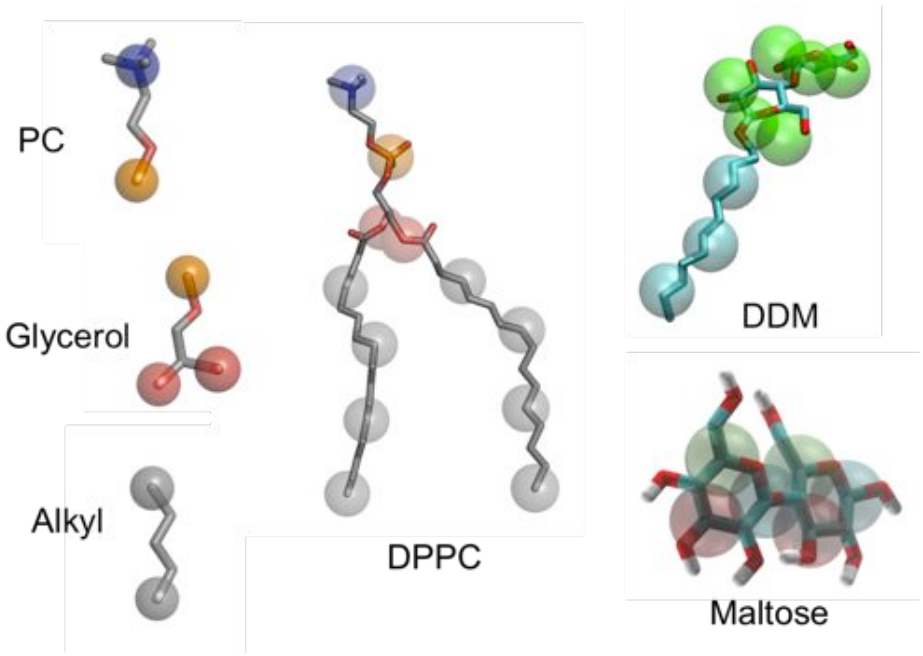


Figure 2:5. Fragment-based coarse-grain to atomistic conversion. The CG particles are shown as transparent spheres, superimposed on the corresponding atomistic structures. Figures for DPPC fragments modified from [29]. The maltose headgroup parameters for DDM were based on the MARTINI carbohydrate extension to the force field.[31]

This interconversion between AT and CG representations (and vice versa) leading to well-equilibrated systems is a powerful tool in studies of μs-timescale processes.

4.5. Sidekick

One of the benefits of using CGMD simulation is that the reduction in computational hours makes CGMD compatible with a high-throughput simulation approach, in which statistical data for the system of interest can be generated. The grid-based program Sidekick has been developed within the lab, automating the process of setting up and running self-assembly studies of helices partitioning into bilayers [32]. The only input required is a peptide sequence (Figure 2:6). An example application of this is found in Chapter 3. Sidekick is in the process of being extended to larger, more complex systems and processes such as helix dimerisation.

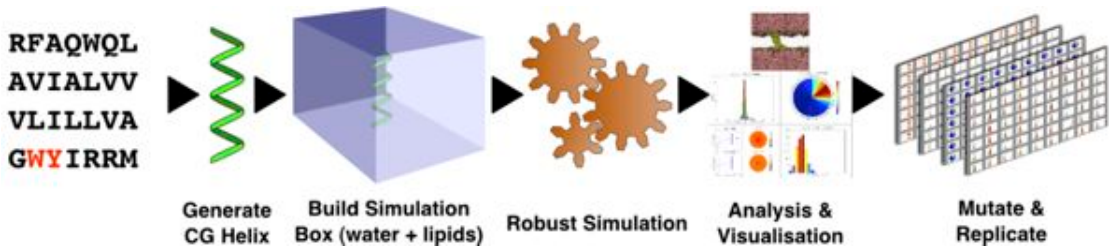


Figure 2:6. The Sidekick pipeline. Figure kindly provided by Benjamin Hall.

5. Non-equilibrium Methods (Steered Molecular Dynamics Simulations)

The methods discussed so far describe equilibrium MD simulations. In these 'brute force' simulations, as long as the system is sampled for long enough it should sample all states eventually. However, some processes are rare, *e.g.* solute transfer through a pore, and would only be observed using timescales that are currently inaccessible to simulation. In these cases, the use of non-equilibrium methods, such as steered molecular dynamics (SMD) simulations, may be used to simulate processes that would otherwise not be observed.

In SMD simulations an external force can be applied to drive a process in the required direction [33]. This external force may take a variety of forms, including a directional 'pulling' force or a rotational 'torque' force. In SMD, a virtual harmonic spring is attached to the centre of mass of the group that is to have the force applied to it (the 'pull group'). A reference group (or absolute reference point in space) is also defined. A force is then applied to the spring. This external force allows energy barriers to be overcome and is akin to micromanipulation of single molecules with optical tweezers, or scanning atomic force microscopy (AFM). SMD simulations have found many uses, *e.g.* in quantitatively and qualitatively predicting protein unfolding pathways, calculating dissociation constants of antibody/antigen complexes and studying ligand binding pathways.

In the work presented in this thesis the method of umbrella pulling was used to transfer a protein-detergent complex from bulk solution to vacuum, corresponding to the processes that occur in the electrospray ionisation process. In umbrella pulling a harmonic potential is applied between the two groups such that the pull force is a function of distance between the groups [5]. The timescales (tens of microseconds) required to simulate the solution-to-vacuum transfer of the protein-detergent complex are not currently feasible using equilibrium MD. By applying an external force to drive the process the simulation time may be reduced by two or three orders of magnitude. SMD is therefore a way to approach this problem, with the added advantage that others may readily reproduce the method using standard GROMACS code.

6. Modelling Loops and Homology Modelling

In the generation of the full BM2 structure (combining PDB id 2KIX and 2KJ1) a single helix was chosen and replicated to give complete four-fold symmetry. This was done so that the modelled loop regions connected the two domains would be identical. MODELLER [34] was used to build 100 models of the two domains connected by inter-domain loop regions (residues 34-44). Loop building within MODELLER is based on the CHARMM22 force field combined with potentials for dihedral angles of loop regions and non-bonded contacts. This function is then optimised by conjugate gradients minimization combined with simulated annealing. The lowest energy structure in which the loop region contained β-loop character was chosen as the final model. This was in order to match PSIPRED predictions. The unstructured C-terminal residues 104-109 were left in random coil configuration. The final model of the full BM2 channel was then converted to the Bond CG model using in-house scripts and simulated with GNM restraints.

The BM1 homology model was generated by sequence alignment with AM1 (PDB id 1AA7), as described elsewhere [35]. MODELLER was used to create the homology model. MODELLER works by converting the template structure (AM1) and the sequence alignment to give a set of spatial restraints for the target protein. Backbone distance restraints are used. The backbone structures are then optimized through conjugate gradients methods. Following backbone optimization, sidechain atoms are added and their positions optimized [34]. The resulting structures are scored and the one with the lowest (corresponding to the best model) score was chosen as the BM1 homology model.

7. Detergent Parameterisation

OPLS-AA parameters for dodecylphosphocholine (DPC), dihexanoylphosphatidylcholine (DHPC) and β-dodecylmaltoside (DDM) detergents were generated as follows:

1) The detergent in differing conformations (to improve sampling of conformational space and reduce chances of system becoming stuck in a local minimum) was edited using Gaussian to give ideal acyl tail dihedral angles. For DPC and DHPC these conformations were generated from simulation of a single

detergent molecule using the GROMOS96 force field in a solvated box. These conformations were then energy minimized. For DDM, conformations from crystal structures in the Protein Data Bank [36] were used.

2) The topology and parameters (GAFF) of the detergent were generated in AMBER format with ANTECHAMBER [37], [38] and converted to GROMACS format using the amb2gmx.pl script [39]. The detergents were optimized with Gaussian 03 [40] at the HF level with the 6-31G basis set. Restrained electrostatic potential (RESP) charges [41] were assigned as the atomic partial charges for the detergent by optimizing the charge fitting to QM electrostatic potential (ESP) maps.
3) The lowest energy parameters were chosen for use.

Coarse-grain parameters were developed for DDM based on the Martini carbohydrate parameters [31]. Coarse-grain parameters for lauryldimethylamine-*N*-oxide (LDAO) were based on the Martini force field. The parameters used in this thesis have been deposited in LipidBook [42].

8. Software and Analysis

All simulations presented in this thesis were performed using the GROMACS biomolecular simulation package [8]. Analysis was performed using in-built GROMACS tools or locally written code. Visualisation and molecular graphics used VMD [43] and were rendered with Tachyon. Pore radius profiles were calculated using HOLE [44]. Secondary structure calculations used the program DSSP [45]. Clustering analysis was performed using NMRClust [46].

9. References

[1] A. Leach, "Molecular Modelling: principles and applications," 2001.

[2] H. M. Senn and W. Thiel, "QM/MM methods for biomolecular systems.," *Angewandte Chemie (International ed. in English)*, vol. 48, no. 7, pp. 1198-229, Jan. 2009.

[3] D. P Tieleman, "Methods and Parameters for Membrane Simulations," *Most*, no. 20, pp. 1-25, 2010.

[4] Y. Mo, "A critical analysis on the rotation barriers in butane.," *The Journal of organic chemistry*, vol. 75, no. 8, pp. 2733-6, Apr. 2010.

[5] H. J. C. van der Spoel, D., Lindahl, E. , Hess, B. , van Buuren, A. R., Apol, E., Meulenhoff, P. J., Tieleman, D. P., Sijbers, A. L. T. M., Feenstra, K. A., van Drunen, R., Berendsen, "Gromacs User Manual version 4.5.4." 2010.

[6] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems," *The Journal of Chemical Physics*, vol. 98, no. 12, p. 10089, Jun. 1993.

[7] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," *The Journal of Chemical Physics*, vol. 103, no. 19, p. 8577, 1995.

[8] W. R. P. Scott et al., "The GROMOS Biomolecular Simulation Program Package," *The Journal of Physical Chemistry A*, vol. 103, no. 19, pp. 3596-3607, May 1999.

[9] L. D. Schuler, X. Daura, and W. F. V. Gunsteren, "An Improved GROMOS96 Force Field for Aliphatic Hydrocarbons in the Condensed Phase." 2001.

[10] W. L. Jorgensen, J. D. Madura, and C. J. Swenson, "Optimized intermolecular potential functions for liquid hydrocarbons," *Journal of the American Chemical Society*, vol. 106, no. 22, pp. 6638-6646, Oct. 1984.

[11] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids," *Journal of the American Chemical Society*, vol. 118, no. 45, pp. 11225-11236, Jan. 1996.

[12] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, "Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides †," *The Journal of Physical Chemistry B*, vol. 105, no. 28, pp. 6474-6487, Jul. 2001.

[13] A. D. MacKerell, et al., "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins †," *The Journal of Physical Chemistry B*, vol. 102, no. 18, pp. 3586-3616, Apr. 1998.

[14] O. BERGER, O. EDHOLM, and F. JAHNIG, "Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature," *Biophysical Journal*, vol. 72, no. 5, pp. 2002-2013, May 1997.

[15] R. HOCKNEY, S. GOEL, and J. EASTWOOD, "Quiet high-resolution computer models of a plasma," *Journal of Computational Physics*, vol. 14, no. 2, pp. 148-158, Feb. 1974.

[16] S. J. Marrink and A. E. Mark, "Effect of Undulations on Surface Tension in Simulated Bilayers," *The Journal of Physical Chemistry B*, vol. 105, no. 26, pp. 6122-6127, Jul. 2001.

[17] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, "LINCS: A linear constraint solver for molecular simulations," *Journal of Computational Chemistry*, vol. 18, no. 12, pp. 1463-1472, Sep. 1997.

[18] J. RYCKAERT, G. CICCOTTI, and H. BERENDSEN, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes," *Journal of Computational Physics*, vol. 23, no. 3, pp. 327-341, Mar. 1977.

[19] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of Chemical Physics*, vol. 81, no. 8, p. 3684, Oct. 1984.

[20] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling.," *The Journal of chemical physics*, vol. 126, no. 1, p. 014101, Jan. 2007.

[21] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, "The MARTINI force field: coarse grained model for biomolecular simulations.," *The journal of physical chemistry. B*, vol. 111, no. 27, pp. 7812-24, Jul. 2007.

[22] A. Y. Shih, A. Arkhipov, P. L. Freddolino, and K. Schulten, "Coarse grained protein-lipid model with application to lipoprotein particles.," *The journal of physical chemistry. B*, vol. 110, no. 8, pp. 3674-84, Mar. 2006.

[23] A. Arkhipov, P. L Freddolino, and K. Schulten, "Stability and dynamics of virus capsids described by coarse-grained modeling.," *Structure London England 1993*, vol. 14, no. 12, pp. 1767-1777, 2006.

[24] P. J. Bond and M. S. P. Sansom, "Insertion and assembly of membrane proteins via simulation.," *Journal of the American Chemical Society*, vol. 128, no. 8, pp. 2697-704, Mar. 2006.

[25] K. A. Scott, P. J. Bond, A. Ivetac, A. P. Chetwynd, S. Khalid, and M. S. P. Sansom, "Coarse-grained MD simulations of membrane protein-bilayer self-assembly.," *Structure (London, England: 1993)*, vol. 16, no. 4, pp. 621-30, Apr. 2008.

[26] A. P. Chetwynd, K. A. Scott, Y. Mokrab, and M. S. P. Sansom, "CGDB: a database of membrane protein/lipid interactions by coarse-grained molecular dynamics simulations.," *Molecular membrane biology*, vol. 25, no. 8, pp. 662-9, Dec. 2008.

[27] T. Haliloglu, I. Bahar, and B. Erman, "Gaussian Dynamics of Folded Proteins," *Physical Review Letters*, no. 1, pp. 2-5, 1997.

[28] T. Carpenter, P. J. Bond, S. Khalid, and M. S. P. Sansom, "Self-assembly of a simple membrane protein: coarse-grained molecular dynamics simulations of the influenza M2 channel," *Biophysical Journal*, vol. 95, no. 8, pp. 3790-3801, 2008.

[29] P. J. Stansfeld and M. S. P. Sansom, “From Coarse Grained to Atomistic: A Serial Multiscale Approach to Membrane Protein Simulations,” *Journal of Chemical Theory and Computation*, vol. 7, no. 4, pp. 1157-1166, Apr. 2011.

[30] P. Rotkiewicz and J. Skolnick, “Fast procedure for reconstruction of full-atom protein models from reduced representations.,” *Journal of computational chemistry*, vol. 29, no. 9, pp. 1460-5, Jul. 2008.

[31] C. A. Lo´pez, A. J. Rzepiela, A. H. de Vries, L. Dijkhuizen, P. H. Hu¨nenberger, and S. J. Marrink, “Martini Coarse-Grained Force Field: Extension to Carbohydrates,” *Journal of Chemical Theory and Computation*, vol. 5, no. 12, pp. 3195-3210, Dec. 2009.

[32] M. S. Hall, Benjamin A, Sansom, “Sidekick: A pipeline and framework for high throughput molecular dynamics simulations,” *In Preparation*.

[33] S. Izrailev et al., “Steered molecular dynamics.” Computational Molecular Dynamics: Challenges, Methods, Ideas, pp. 39-65, 1998.

[34] A. Sali and T. L. Blundell, “Comparative protein modelling by satisfaction of spatial restraints.,” *Journal of molecular biology*, vol. 234, no. 3, pp. 779-815, Dec. 1993.

[35] J. Wang, R. M. Pielak, M. A. McClintock, and J. J. Chou, “Solution structure and functional analysis of the influenza B proton channel.,” *Nature structural & molecular biology*, vol. 16, no. 12, pp. 1267-71, Dec. 2009.

[36] H. M Berman et al., “The Protein Data Bank.,” *Nucleic Acids Research*, vol. 28, no. 1, pp. 235-242, 2000.

[37] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, “Development and testing of a general amber force field.,” *Journal of computational chemistry*, vol. 25, no. 9, pp. 1157-74, Jul. 2004.

[38] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, “Automatic atom type and bond type perception in molecular mechanical calculations.,” *Journal of molecular graphics & modelling*, vol. 25, no. 2, pp. 247-60, Oct. 2006.

[39] D. L. Mobley, J. D. Chodera, and K. A. Dill, “On the use of orientational restraints and symmetry corrections in alchemical free energy calculations.,” *The Journal of chemical physics*, vol. 125, no. 8, p. 084902, Aug. 2006.

[40] M. Frisch et al., “Gaussian 03, Revision C.02,” 2003.

[41] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, “A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model,” *The Journal of Physical Chemistry*, vol. 97, no. 40, pp. 10269-10280, Oct. 1993.

[42] J. Domański, P. J. Stansfeld, M. S. P. Sansom, and O. Beckstein, “Lipidbook: a public repository for force-field parameters used in membrane simulations.,” *The Journal of membrane biology*, vol. 236, no. 3, pp. 255-8, Aug. 2010.

[43] W. Humphrey, A. Dalke, and K. Schulten, “VMD: visual molecular dynamics.,” *Journal of molecular graphics*, vol. 14, no. 1, pp. 33-8, 27-8, Feb. 1996.

[44] O. S. Smart, J. G. Neduvelil, X. Wang, B. A. Wallace, and M. S. Sansom, "HOLE: a program for the analysis of the pore dimensions of ion channel structural models.," *Journal of molecular graphics*, vol. 14, no. 6, pp. 354-60, 376, Dec. 1996.

[45] W. Kabsch and C. Sander, "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.," *Biopolymers*, vol. 22, no. 12, pp. 2577-637, Dec. 1983.

[46] L. A. Kelley, S. P. Gardner, and M. J. Sutcliffe, "An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies.," *Protein engineering*, vol. 9, no. 11, pp. 1063-5, Nov. 1996.

Chapter 3

The Influenza B Virus Proton Channel

1. Introduction

Coarse-grained molecular dynamics (CGMD) simulations have been used to model the two-stage folding model of α-helical membrane proteins. In this Chapter CGMD simulations are used to generate a model of the BM2 TM domain, with no experimental structure accessible at the time of generation. This model is then converted to an atomistic representation to allow a possible gating mechanism of the channel to be probed.

The process of membrane protein folding may be simulated in two steps:

1) The insertion of individual helices into a bilayer ("Self-Assembly Simulations")
2) Lateral diffusion of helices in the bilayer until oligomer is formed ("Tetramerisation Simulations")

These two stages are discussed below under the corresponding subsections.

1.1. Self-Assembly Simulations

To model the tetramerisation of the BM2 TM domain using this methodology, it is first necessary for the individual helices to be shown to be stable in a TM orientation in the bilayer. Therefore self-assembly simulations of a CG-DPPC bilayer were performed to assess the orientation of the BM2 TM peptide monomer (BM2TM).

A 27-mer peptide (P^{4}FQILSICSFILSALHFMAWTIGHLNQ^{30}) was used throughout this work to model the putative TM domain of BM2 plus several residues on either side (BM2TM, Figure 3:1). This region was chosen based on experimental evidence [1] as

well as the results from sequence-based TM prediction servers [2], [3]. An ideal atomistic helix was built in Modeller v9.2 [4] and then converted to the Bondv2 coarse-grain force field using in-house scripts, as has been described elsewhere [5]. The helix was constrained to be α-helical based on the PSIPRED secondary structure prediction server [6], with the terminal three residues at each end in a random coil configuration. The His27 residue was protonated throughout based on experimental data [7], however as this was on the boundary of the α-helical region it was found to make little difference to the CGMD structures (see Figure 3:5D).

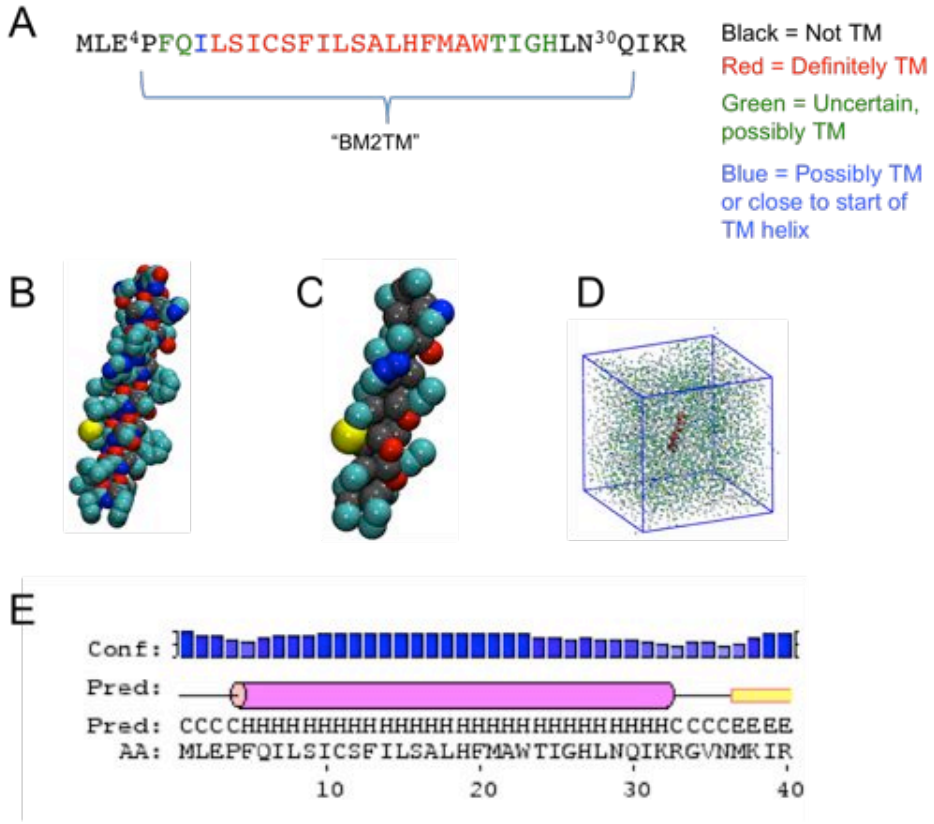


Figure 3:1. A: Residues 1 to 33 of the BM2 sequence. The peptide chosen for this study (4-30) is bracketed. The colour of the sequence corresponds to the consensus TM prediction server results. Red = definitely TM, black = non-TM residues, green = uncertain, possibly TM, blue = possibly TM or close to start of TM helix. B: The BM2TM atomistic α-helix built in Modeller. C: The CG BM2TM helix. Backbone particles are shown in grey. D: The starting point of monomer insertion simulations. The helix backbone particles are shown as a red trace in the centre of the simulation box that is filled with CG lipids and water particles in random configuration. E: Results of secondary structure prediction server for first 40 residues. The residues within BM2TM are predicted to be α-helical

The insertion simulations were set up as follows:

1) The CG helix was placed in a simulation box (100 Å^3)
2) ~300 CG-DPPC lipids were added in random orientation and position,
3) CG water particles were added using the program genbox.

4) Water particles were replaced by Na^+/Cl^- ions as required to ensure system neutrality.

This system was the starting point for 5 CGMD simulations of 200 ns duration (Figure 3:1D).

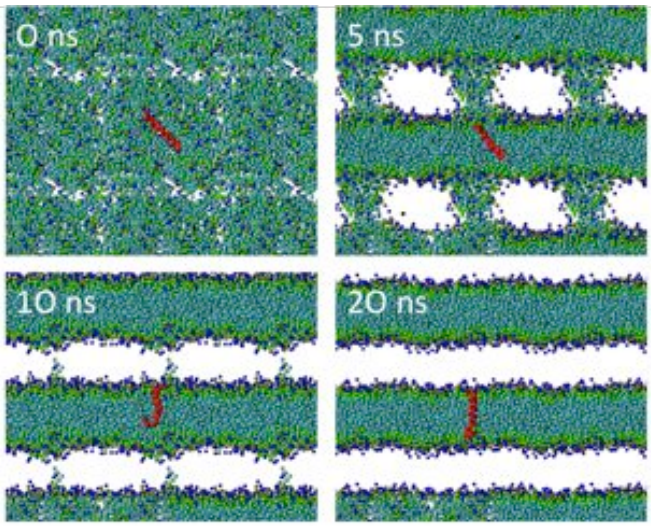


Figure 3:2. Stages of helix insertion during bilayer self-assembly simulation. The helix inserts as the lipids form a stable bilayer during the first ~ 20 ns. This bilayer remains intact (and the BM2-TM continues to span the bilayer) for the remainder of the 0.2 μs simulation. The backbone of the helix is portrayed as a red trace Water particles are omitted for clarity.

The system behaves as described previously for helix insertion simulations with GpA [8] and A/M2 [9]. The lipid molecules cluster together within a couple of nanoseconds and form a bilayer by ~ 20 ns (Figure 3:2). The outcome of each simulation was formation of a stable bilayer spanned by the BM2TM monomer. The BM2TM monomer remained in this transmembrane orientation for the duration of the 200 ns simulations. The average tilt angle was 14 ± 2°.

These monomer insertion simulations demonstrate that the BM2TM peptide adopts a stable, transmembrane orientation.

1.2. Tetramer Oligomerisation Simulation

To model the second stage of membrane protein folding, a system in which four helices were present in the bilayer patch was made. The final configuration of the helix monomer simulations was replicated to generate 3 more helices. These were inserted in the bilayer in parallel fashion at distances of ~45 Å from each other (Figure 3:3). Each of

the replicated helices was rotated randomly about its helix axis (bilayer normal) to prevent any bias due to initial orientation. Any overlapping lipids were removed.

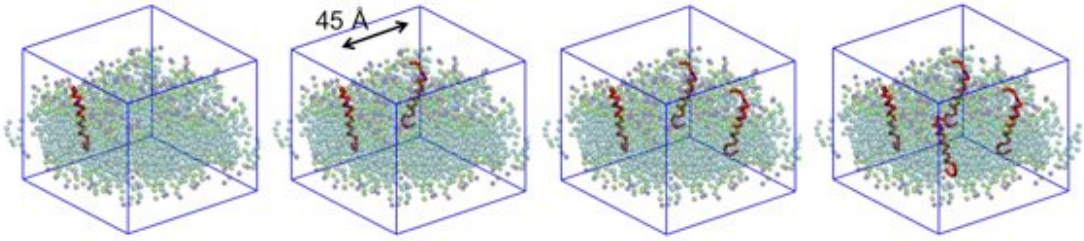


Figure 3:3. Starting configuration of tetramer simulations. Left-most panel shows the BM2TM monomer partitioned in the bilayer. This helix was replicated three times to give a total of 4 helices. These were placed in equidistantly in the bilayer and rotated randomly about the helix axis to give a non-biased starting configuration. The helices are displayed as a red backbone trace. Lipids are shown as transparent spheres. Water is omitted for clarity.

This system was used as the starting point for 5 independent 5 μs simulations, in which the helices were observed to laterally diffuse through the bilayer. The formation of helix bundles was monitored using the distance between the centres of mass of the backbone particles of pair-wise combinations of helices (see Figure 3:4). In each case the tetramer was seen to form *via* a dimer or trimer, with initial formation of a loose tetrameric assembly. This loose assembly subsequently equilibrated to form a tightly-packed left-handed bundle of approximate four-fold symmetry. Figure 3:4 shows these stages of tetramer formation. The average time to tetramer formation was 0.5 (± 0.2) μs.

The structures from these simulations were clustered based on their backbone CA particles (equivalent to Cα atoms) of the regions constrained to be helical using the program NMRClust [10]. A representative frame from each of the top 3 clusters in each trajectory was taken, to give a total of 15 frames. These 15 frames were clustered again to give the overall most representative frame. This clustering showed that 13 of the 15 structures belonged to a single cluster (average Cα RMSD = 2.1 Å) whereas 2 of the structures corresponded to a minor configuration. The superimposition of these 13 structures is shown in Figure 3:5 along with the relative RMSD for each backbone particle between each of the 13 frames. Further analysis showed that the minor conformation observed was due to a meta-stable 2-fold symmetric structure persisting for timescales of over 1 μs in one simulation (see Figure 3:7).

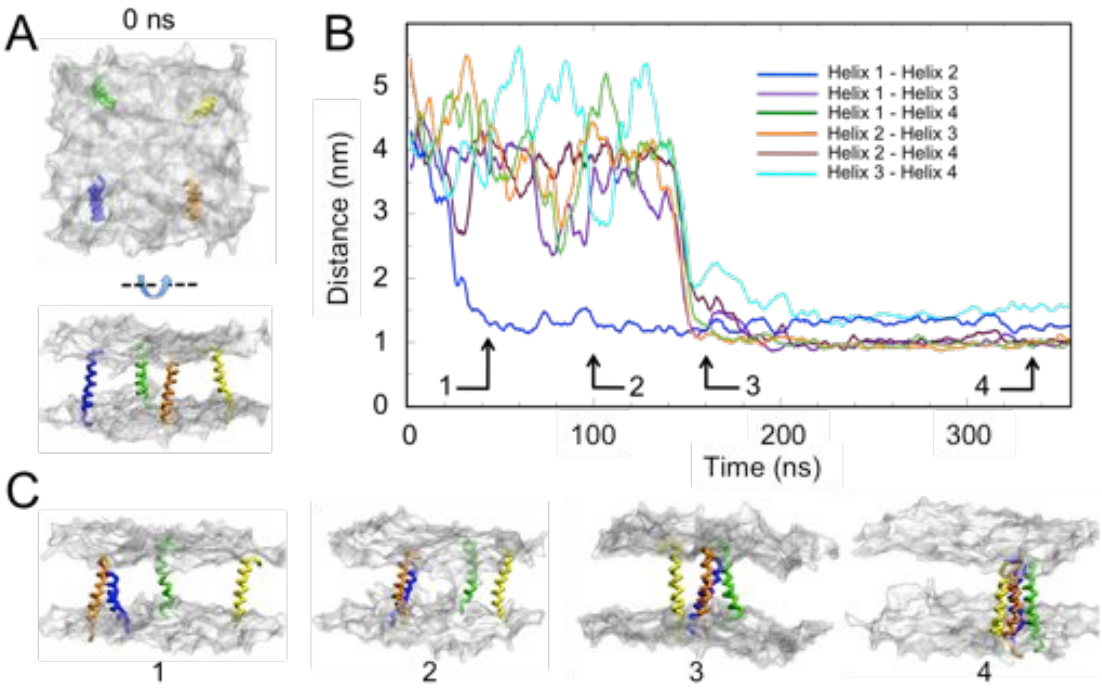


Figure 3:4. Tetramerisation of the BM2-TM helices. A: The system at 0 ns. B: Pair-wise helix-helix distances over time. C: Snapshots 1, 2, 3 and 4 correspond to the system at times marked by arrows on the inset. The lipid headgroups are shown as a grey surface. Helix backbones are shown as coloured traces.

The converged structure obtained from CGMD simulations had crossing angles of 13 ± 3° between opposite helices and 14 ± 4° between adjacent helices. This left-handed model is consistent with predicted pore-lining residues [11], namely Ser9, Ser12, Ser16 His19, and Trp23 (Figure 3:6). The helices are tightly packed at the N-terminus and slightly further apart in the C-terminal region where the bulky Trp sidechains must be accommodated.

Prior to the solution NMR structure of the BM2 channel being solved, the only structural model of the BM2 channel used computational methods and was based on alignment of predicted pore-lining residues. The RMSD between the backbone particles of the CG-generated model and this computational model from reference [11] was 1.6 Å. This is comparable to the deviation between backbone particles within individual clusters of the CGMD simulations (1.2 Å for the top cluster).

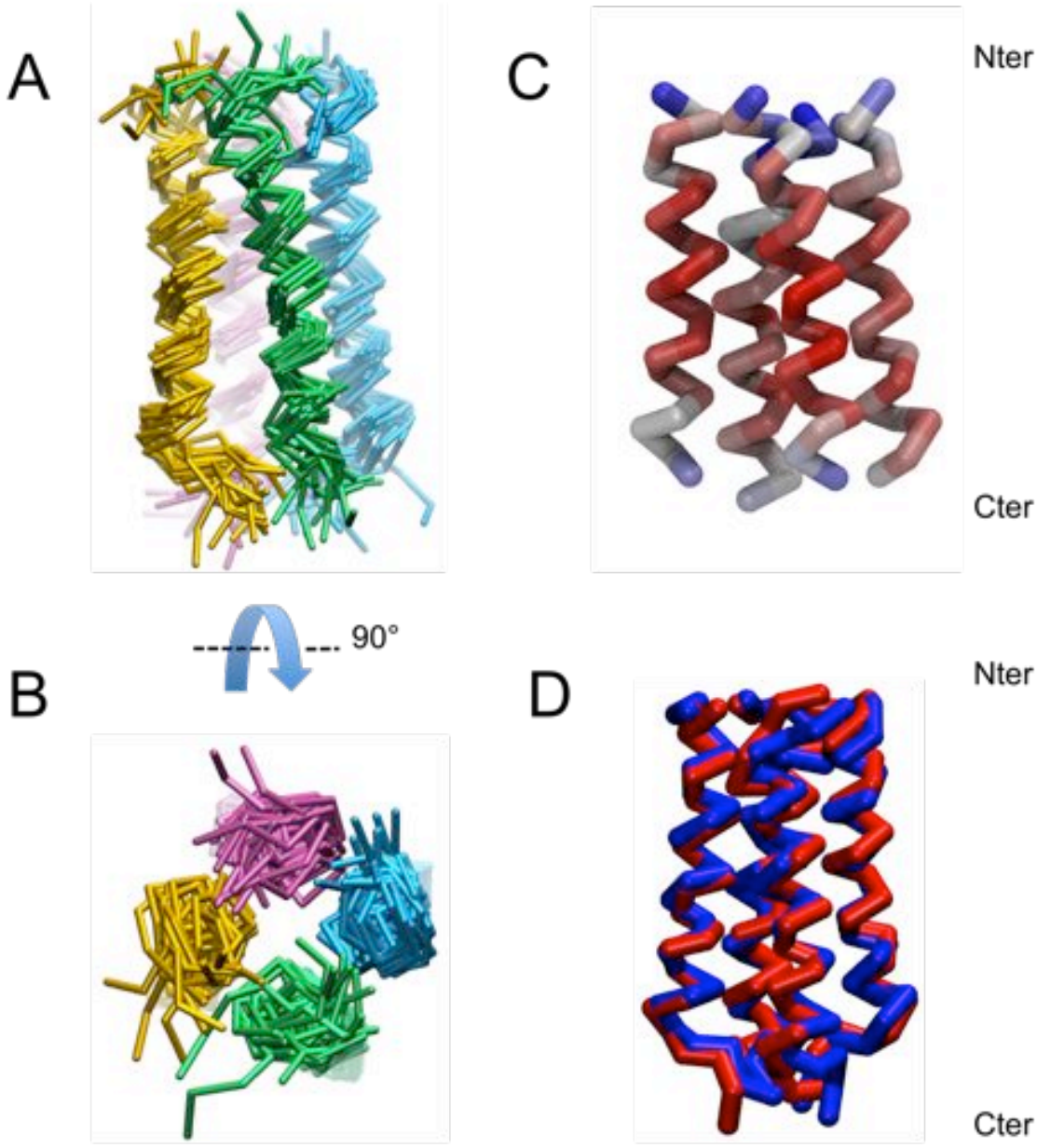


Figure 3:5. A: Superimposition of the backbones of top clustered frames from each simulation. B: View down bundle axis from the N-terminus. C: Tetramer coloured in terms of Cα RMSD between the top clusters (red = least deviation, blue = most deviation). D: The top model generated with His27 residues protonated (blue) superimposed on that generated with neutral His27 residues (red). His27 is on the boundary of the helical region therefore little difference was observed in the structure generated.

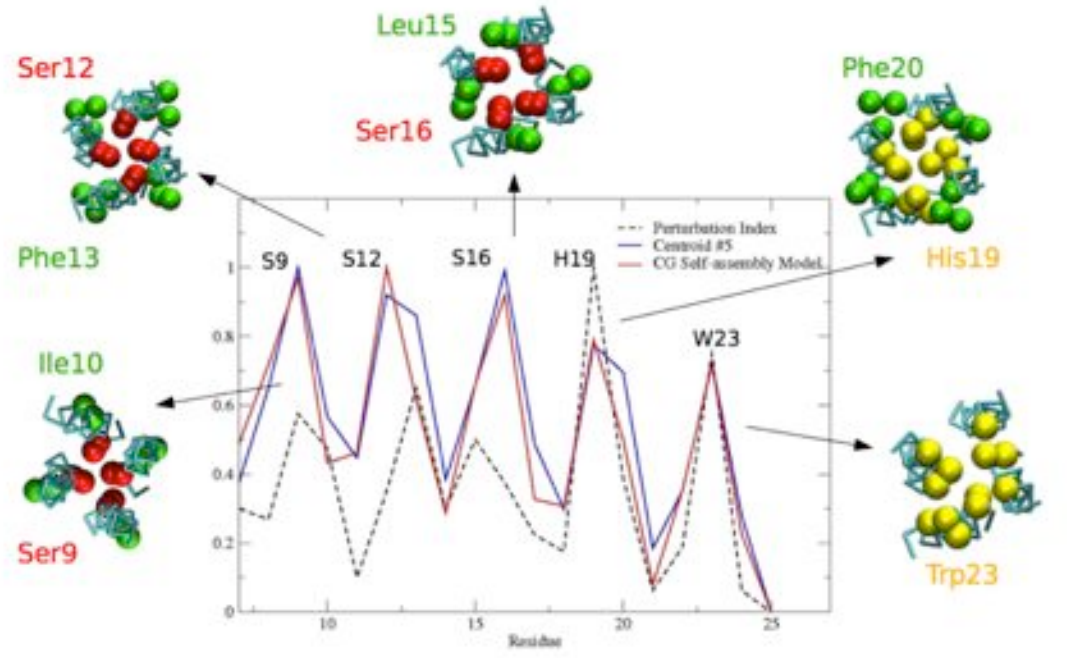


Figure 3:6 Pore-lining residues of the BM2 Channel as Predicted by Mutation Scanning and CGMD simulation. The normalized distances between residues of diagonal helices averaged over 4 μs of CGMD tetramer simulation time are plotted in red, along with the distances calculated from Centroid 5 from reference [11] in blue. Representations of the CGMD model are shown with sidechains coloured according to labels. The dashed black line shows the perturbation index. Residues with the highest perturbation index are predicted to be pore-lining.

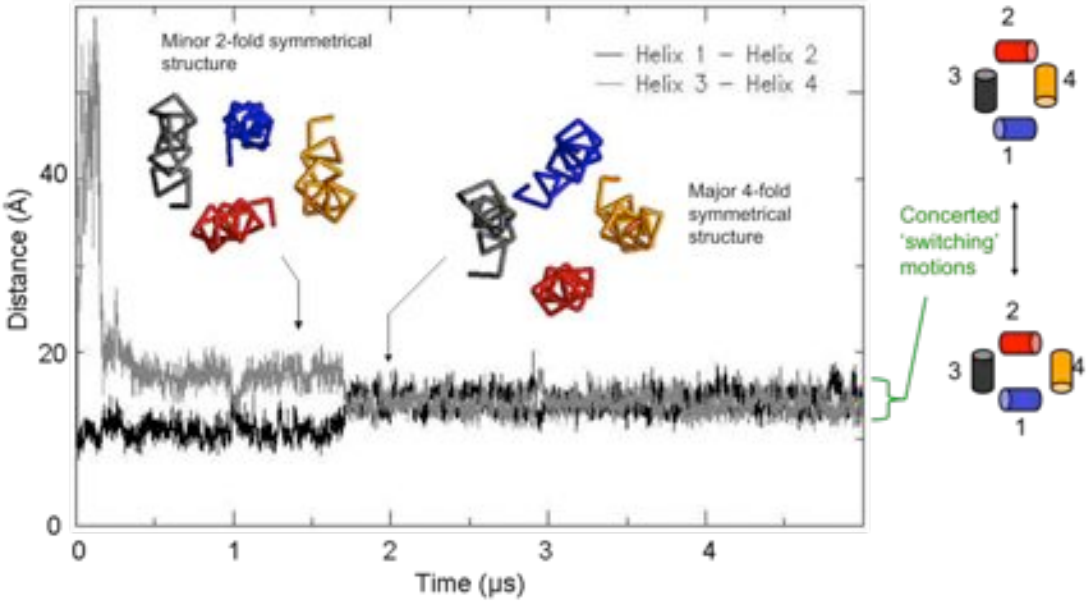


Figure 3:7. Dynamic behaviour of the BM2 TM tetramer. Pairwise inter-helix distances may be used to monitor the conformation and dynamics of the bundle during simulation. Here a meta-stable structure with 2-fold symmetry is seen transitioning to the major 4-fold symmetrical structure. Once the stable 4-fold symmetric structure is obtained, smaller concerted motions (in which one diagonal pair of helices move closer while simultaneously the opposite pair of helices move further apart) are observed on the timescale of tens of nanoseconds.

1.2.1. Tetramer Dynamics

Analysis of pairwise helix-helix distances demonstrates that the helix is highly dynamic on the timescale of CGMD. Concerted 'switching' motions are observed on the tens of nanoseconds timescale, in which one pair of diagonal helices move closer whilst the other pair simultaneously move further apart (Figure 3:7). This is similar to behaviour observed of a CGMD-generated structure of the A/M2 tetramer. However for BM2 the fluctuations are smaller with a 4-fold symmetry adopted, compared to the 2-fold symmetry observed for A/M2 [9]. Figure 3:7 shows that a minor structure with 2-fold symmetry persists in one simulation for almost 2 microseconds before converting to the more stable structure with a 4-fold axis of symmetry.

1.3. Control simulations: Randomized Sequence

The CGMD-generated models of GpA and A/M2 were were validated by comparison to experimental data and control simulations were performed [9], [12]:

1) Control simulations of A/M2 were performed using a randomised sequence of A/M2. This led to much lower frequency of helix partitioning into the bilayer, demonstrating that CGMD may be used to predict the propensity of a given helix to partition into a bilayer.
2) The dimer affinity of mutant GpA helices was simulated. The CGMD methodology was able to be used to distinguish between disruptive mutants and the wildtype dimers.

In the absence of an experimental structure to compare the CGMD-generated BM2 model to, one further test of the CGMD method was performed in order to ensure that the tetramer formed was as a result of sequence-specific helix-helix interactions. The BM2 sequence was taken and the hydrophobic region of the sequence randomized, retaining the 4 residues at each termini. This was to permit insertion of the helices to comparable levels as the wild type. Nine of these sequences were chosen along with the wildtype sequence, for the basis of a peptide insertion study. CGMD self-assembly simulations (in which the peptide is placed in a box with randomly orientated and

positioned CG-DPPC lipids) were performed. The system was set-up to maximize the chances of bilayer formation by having a layer of CG water particles in the xy-plane.

Sequence	Number of Simulations [a]	% Run Insertion [b] (% Time Insertion [c])
PFQILSICSFILSALHFMAWTIGHLNQ (WT)	300	64 (64)
PFQICAAHFFSLISWLSIIMTGLHLNQ	312	46 (47)
PFQIFIHLSFSILIMGLACWTASHLNQ	311	64 (64)
PFQIFLAMGFISLHTSICLSAWIHLNQ	345	48 (49)
PFQIFLLMGSLISIWICATHAFSHLNQ	312	75 (74)
PFQIGHLIFLSASMLSAWTSFIIHLNQ	327	69 (68)
PFQILWISLIAFLGSTAFSMIHSHLNQ	304	56 (56)
PFQISGTLLLCMSIASIWHAFFIHLNQ (S7)	353	72 (71)
PFQISILMLHSAWIASCTFFLIGHLNQ	308	49 (50)
PFQISWCAIMIASLSLGFFHLITHLNQ	312	59 (59)

Table 1. Insertion properties of BM2 monomer helices with randomized sequences compared to the WT. [a] The number of simulations out of 400 passing normality tests (including formation of a distinct bilayer). [b] Percentage of individual simulations in which the helix was inserted for over 50 % of simulation time. [c] Percentage of simulation time helix was inserted over all simulations.

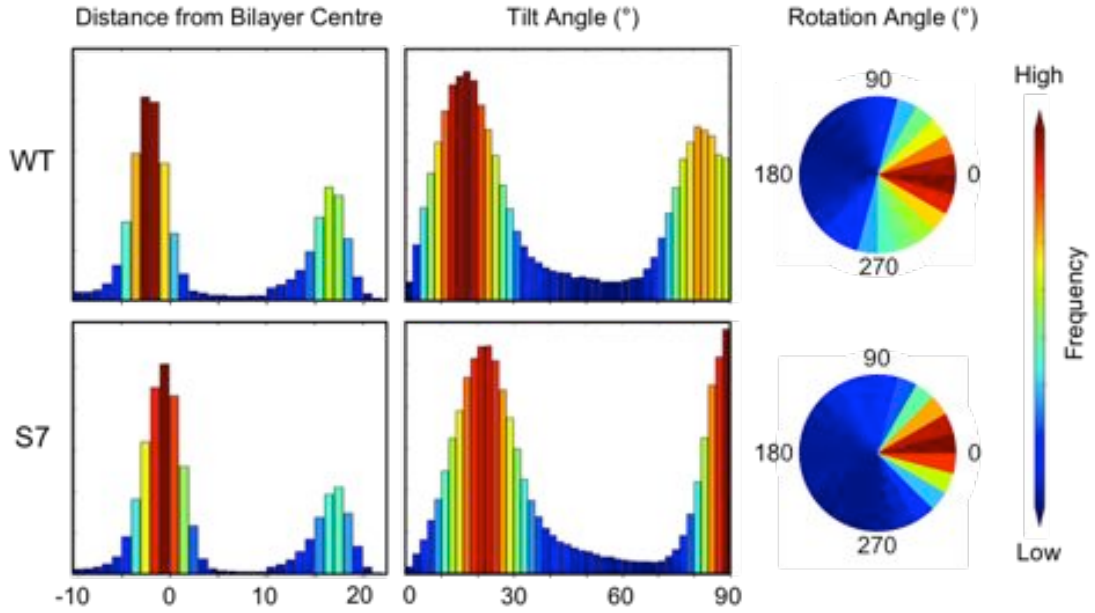


Figure 3:8 Sidekick analysis of WT simulations compared to randomized sequence S7. The distance of the protein from the bilayer centre is defined as the distance between the protein and bilayer centres of mass, along the bilayer normal. Tilt angle is calculated as the angle between the helical axis and the bilayer normal for each simulation frame. The rotation angle refers to the internal rotation angle of the helix.

400 simulations for each sequence were performed and analyzed in terms of helix position in the bilayer, tilt and internal rotation. The insertion properties of each sequence are displayed in Table 1. The program Sidekick [13] was used to perform these simulations (see Methods Chapter).

As expected, each sequence displays a comparable percentage of time inserted in a TM orientation. The insertion propensities, position in bilayer, tilt and internal rotation of each of the helices were compared to that of the wildtype, and the most similar sequence (labelled "S7" in Table 1) was chosen for further study.

The starting point for tetramerisation simulations of the scrambled sequence S7 was generated as described previously for the WT simulations: four copies of the TM-orientated helix were placed in parallel orientation at distance of ~ 45 Å from each other. This was used as the start point for 5 independent 5 μs CGMD simulations. As observed for the WT tetramer, the helices diffused laterally through the bilayer until aggregating *via* dimer or trimer formation.

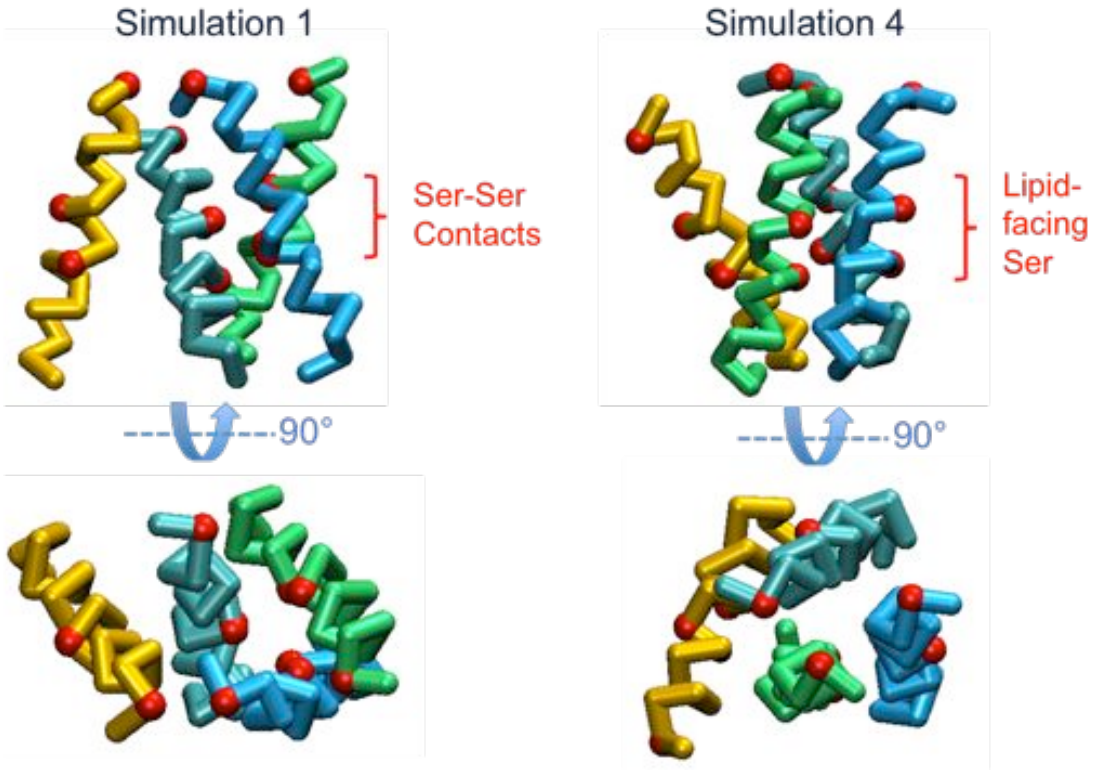


Figure 3:9. Representative frames from 2 simulations of the S7 oligomerisation simulations. In Simulation 1 the helices orientate such that the Ser residues of 3 helices interact. In simulation 4, these Ser residues do not interact but adopt an outwards, lipid-facing orientation.

In contrast to the WT tetramerisation simulation, cluster analysis of the bundles formed did not lead to a converged structure. In each simulation the helices packed such that three of the helices were tightly packed as a trimer, with the fourth helix loosely in

contact with one or more helices. Closer inspection of the clustered structures showed that the helices did not have common interaction interfaces in each simulation (Figure 3:9). In one simulation the pore-lining serine residues in the tetramer are pore-lining in the helix 'trimer' but face the lipids in the fourth helix, whereas in another simulation all of the serine residues face outwards to the lipids.

It is clear from these results that the tetrameric bundle generated from the simulations of spontaneous association of the WT BM2 TM peptides is not simply a case of the CGMD symmetrically packing 4 helices, but is a consequence of the peptide sequence used *i.e.* a sequence-specific structure.

2. Atomistic Simulations of CG Generated Model

2.1. Conversion to Atomistic Representation

To study the CGMD-generated tetramer in atomistic detail, the converged BM2TM structure in an equilibrated bilayer was converted to an atomistic representation, using the GROMOS force field. This CG2AT conversion was achieved by using a fragment-based approach (see Methods Chapter and [14]). Figure 3:10 shows an example of this conversion process.

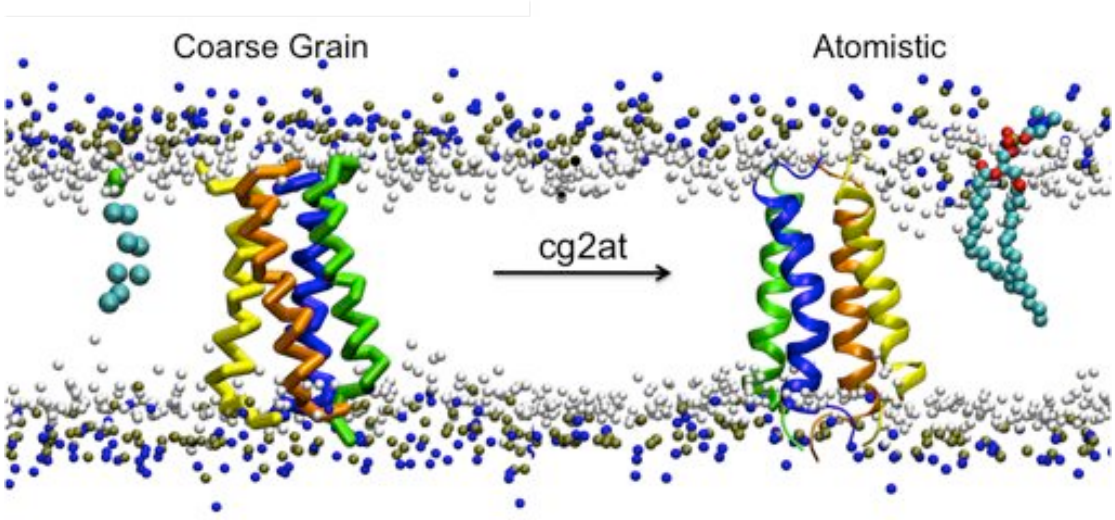


Figure 3:10. The Coarse-Grain (CG) to Atomistic (AT) Conversion. The left-most side shows the CG system used for the conversion. The right-hand side shows the converted AT system (in this case using the GROMOS96 forcefield). The helix Cα particles are shown in each case along with one DPPC lipid. The equivalent particles corresponding to the phosphates (tan) and nitrogens (blue) of the phosphocholine groups along with the glycerol oxygens (white) are shown as spheres.

2.2. Stability of the CG2AT Channel

A 30 ns ATMD production run of the CG2AT system was performed following equilibration steps described in the Methods Chapter. This ATMD simulation was analysed to assess the stability of the CG-generated BM2 structure. Figure 3:11 shows the superimposition of the tetramer structure at the end of the ATMD simulation and the structure generated by the original CG2AT conversion process. The overall Cα RMSD between the final structure and the initial structure was 2.2 Å.

Secondary structure analysis demonstrated that the helices fluctuated around ~90 % of the maximum α-helical content during the 30 ns of simulation time ((Figure 3:12C). Individual helix Cα RMSDs were between 1 and 2 Å. The RMSDs appear to plateau within the first ~5 ns (Figure 3:12A). The RMS fluctuations of 2 of the 4 helices were observed to be slightly higher in the C-terminal region than the N-terminal region. The presence of bulky Trp23 sidechains in this region cause the helices to be less closely-packed in this region and a higher RMSF is to be expected (Figure 3:12B).

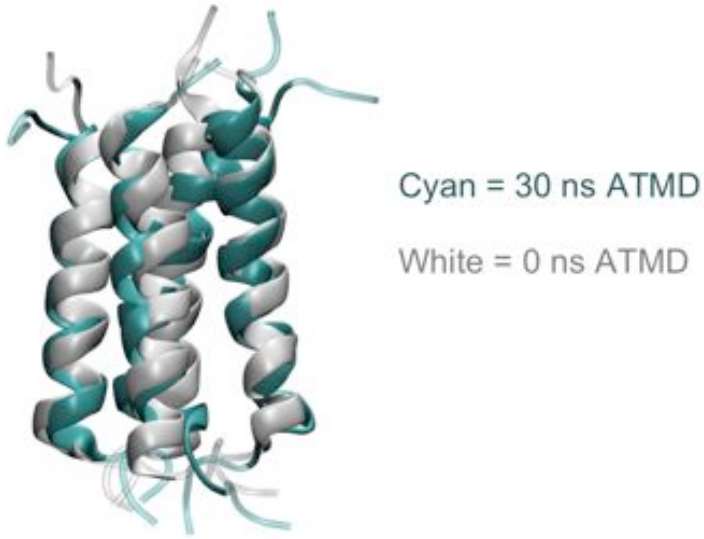
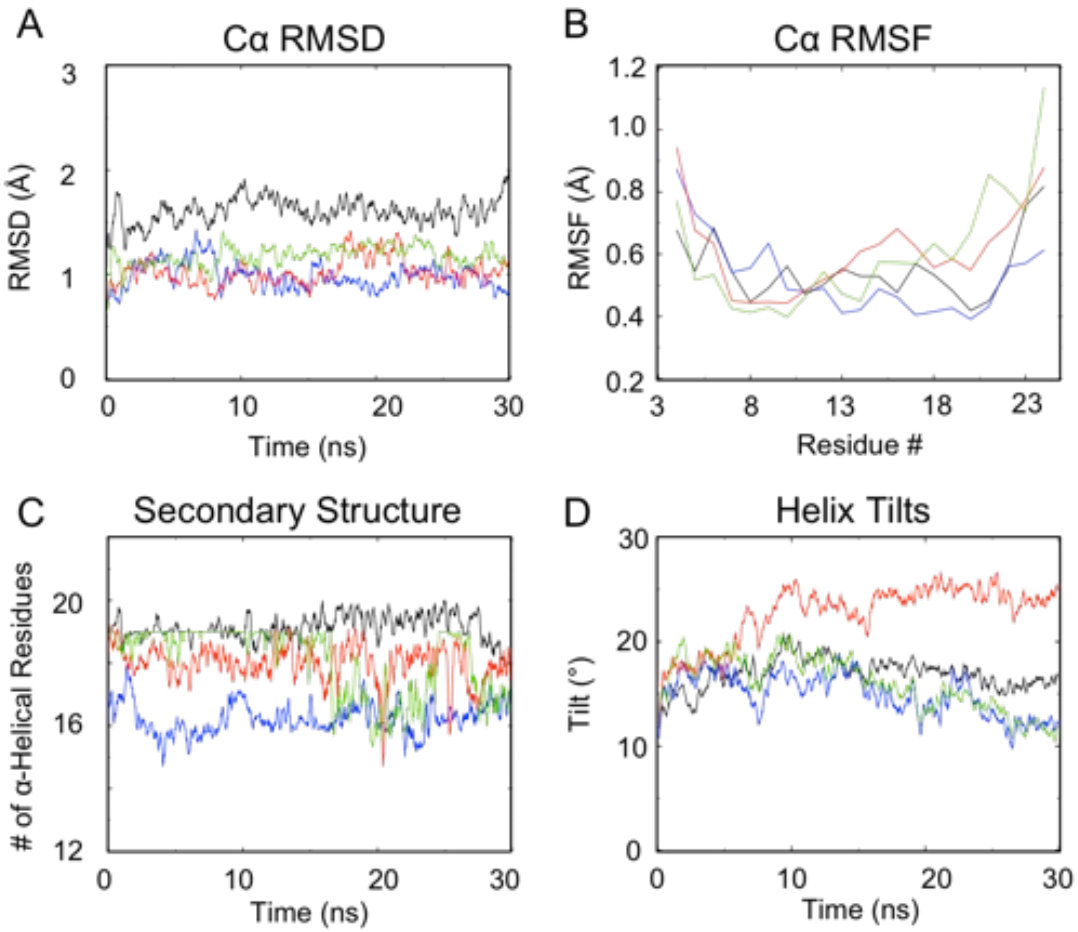


Figure 3:11. Superimposition of the initial CG2AT (grey) and final (cyan) structures of the BM2 channel. The tetramers are shown in NewCartoon representation.

The main change to the channel over the 30 ns simulation was a reorientation of the bundle such that the average tilt increased during the simulation ((Figure 3:12D). This is in order to maximize interactions of polar residues with the hydrated headgroup region ("hydrophobic mismatch"). The lower tilt in CGMD simulations may be attributed to the forced secondary structure not permitting any flexibility of the helices as well as the simplified interaction potentials.

The above analysis of the ATMD simulation confirms that CGMD can be used to generate a BM2TM tetramer structure that is stable on the atomistic timescales simulated.



(Figure 3:12. Stability of the cg2at converted BM2TM tetramer. RMSD (A) and RMSF (B) of the C*α* particles of the helical region of each helix during ATMD simulation. C: Helical content of each chain during simulation as calculated using the program DSSP. D: The tilt of each helix during simulation

2.3. Potential Gating Mechanism

Previous studies of A/M2 have linked channel opening at low pH to protonation of the His37 residue [15–18]. Similarly, the homologous residue in BM2, His19, is the only ionisable side chain within the TM domain and has been found to be vital for the channel to function; therefore, it is probable that channel opening coincides with protonation of this residue [11], [19], [20]. This possible gating mechanism was probed by successive increase in protonation states of the His19 residues. Initial pore profile analysis of CG2AT models of AT-0 and AT-2+ demonstrated that the pore was initially closed. It was therefore decided to focus on the AT-2+ to AT-3+ and AT-3+ to AT-4+ transitions.

The simulations performed are summarized in Figure 3:13 and Table 2. The first set of simulations involved a CG2AT model of the AT-2+ bundle used as the starting point for 30 ns production runs. The doubly-protonated His19 residues were either on adjacent helices or diagonal ('opposite') helices. Representative frames from these were then

used as the starting point for simulations with 3 of the 4 His19 residues doubly-protonated. For the AT-3+ simulations both possible choices of the 3rd histidine to be protonated were initially simulated and the most stable in terms of secondary structure were continued for full analysis.

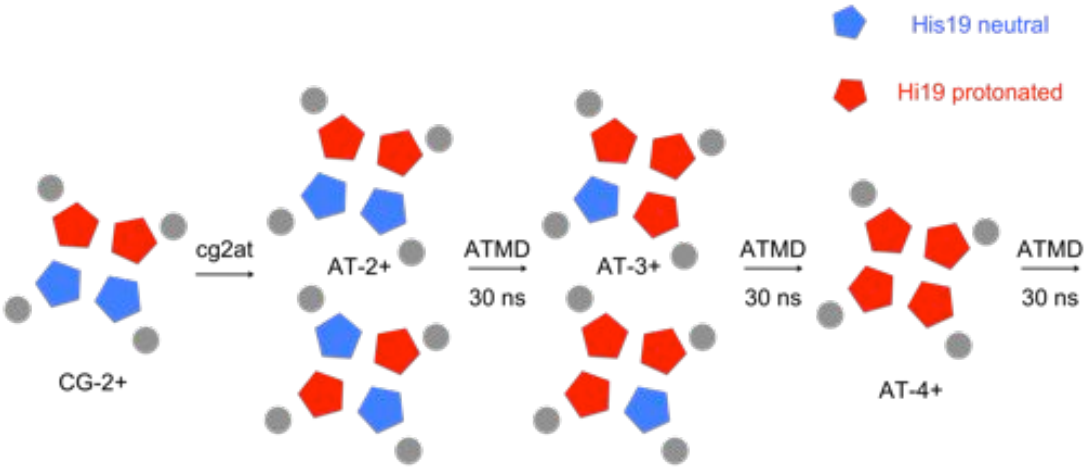


Figure 3:13. Simulation pathway for successive His19 protonation simulations. The initial AT-2+ is obtained by cg2at conversion of a doubly-protonated CG bundle. This is simulated for 30 ns and a representative frame taken and converted to a triply-protonated system. This is repeated for the AT-3+ to AT-4+ transition.

Name	# His19 +	Configuration	Start Point
AT-2+-A	2	Adjacent	CG2AT
AT-2+-B	2	Adjacent	CG2AT
AT-2+-opp	2	Opposite	CG2AT
AT-3+-A	3	-	AT-2+-A (20 ns)
AT-3+-A2	3	-	AT-2+-A (23 ns)
AT-3+-B	3	-	AT-2+-B
AT-3+-opp	3	-	AT-2+-opp
AT-4+-A	4	-	AT-3+-A
AT-4+-A2	4	-	AT-3+-A2
AT-4+-B	4	-	AT-3+-B
AT-4+-opp	4	-	AT-3+-opp

Table 2 His27 protonation state simulations performed. Column 2 shows the number of His19 residues that are charged. Column 3 shows the orientation of these for the AT-2+ systems. The rows are coloured in terms of the simulation series described in the text.

Each of these simulations was analyzed in terms of the channel stability, structure, dynamics, bilayer orientation and potential gating mechanism. Each of the AT-2+, AT-3+, AT-4+ simulation series are coloured in Table 2 in the same colours they are plotted in for subsequent analysis.

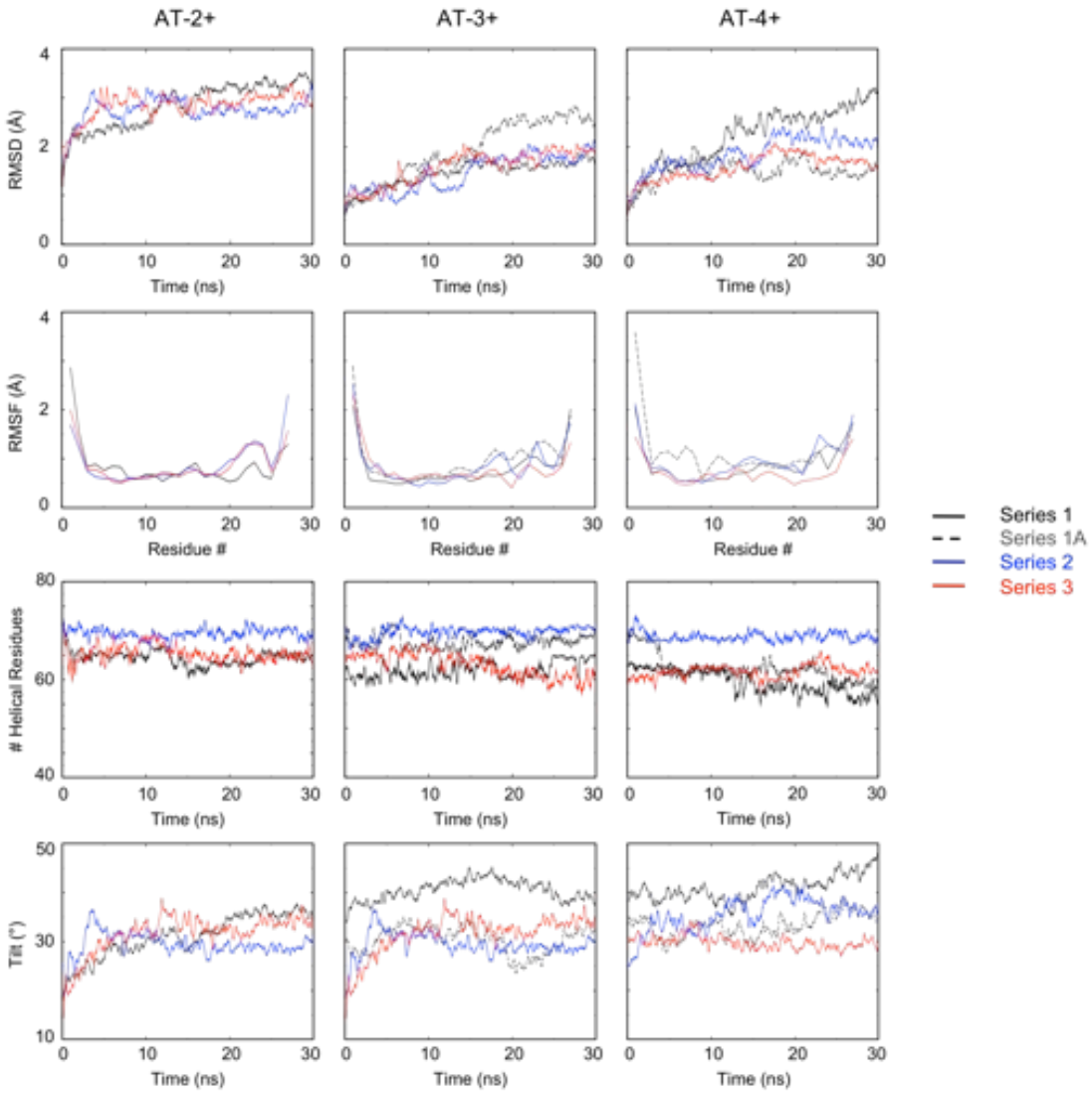


Figure 3:14. Analysis of each series of His27 simulations. Top panel: Backbone RMSD of helical regions during 30 ns ATMD. Second panel: RMS Fluctuation of all backbone residues during simulation. RMFSs of the residues in the first chain only are shown for clarity. Third panel: Helical content of protein during each simulation as calculated by DSSP. Bottom panel: The tilt of the bundle averaged over each helix for each simulation. The largest deviation is observed in the initial AT-2+ simulations that follow the cg2at conversion.

The largest Cα RMS deviations were observed for the AT-2+ simulations. This is unsurprising since these simulations are based on a CG model starting point and some relaxation of the system is expected. For the AT-3+ and AT-4+ simulations the deviation is due only to protonation of a single residue. The RMSDs of the AT-2+ systems ($\sim$ 3 Å) are slightly higher than that observed for the CG2AT simulation of the AT-0 system (2.2 Å), which may be expected due to the simplified description of charges in the CG

model. The RMSDs appear to converge in most simulations within the 30 ns timescale. A notable exception is simulation AT-4+-A; this simulation is discussed further below.

The largest fluctuations in each simulation are in the C-terminal region of the peptide – from His19 residue onwards. This was also observed in the simulation with no His19 residues protonated. The channel is more open at the C-terminal region that also contains the His19 residues that are additionally protonated. A combination of the two factors leads to the observed increase in fluctuation in this region.

Secondary structure analysis shows that, similarly to the AT-0 simulations, the α-helical content is largely maintained in the AT-2+ simulations. In two of the AT-3+ simulations (Series 1 & 3) a decrease in the number of α-helical residues is observed. Further loss in helicity is observed for the AT-4+ simulation in Series 1. Series 1A retains most secondary structure until all His19 residues are protonated.

As was observed for the AT-0 simulation, the tilt of the bundle increases during the AT-2+ simulations from $\sim 15°$ to $\sim 30°$ and remains close to this value upon increase of His19 charge. Although the tilt increases, the overall geometry of the bundle does not. This increase in tilt is likely to be a combination of the secondary structure restraints in the CG system preventing any flexibility of the helices, as well as the increase in system complexity causing the helices to relax into a slightly different conformation following the CG2AT conversion process.

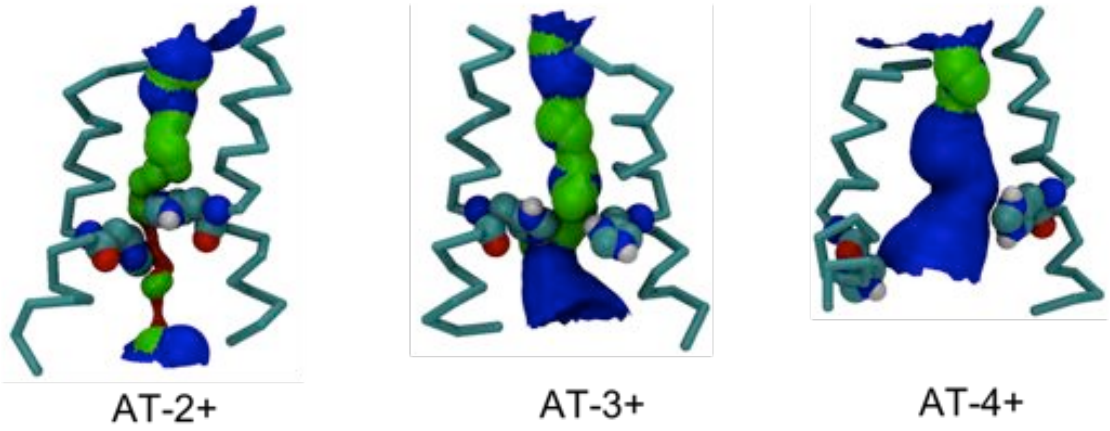


Figure 3:15. Pore surfaces of representative frames from the AT-2+, AT-3+ and AT-4+ simulations of Series 1. Cα particles of two helices are shown as a trace. His19 residues of these chains are shown in space-filling representation. Constrictions in the HxxxW region are observed in AT-2+ and removed in AT-3+ and AT-4+ simulations.

Pore profile analysis of each simulation was performed (Figure 3:15). For the AT-2+ systems a constriction was observed in the putative HxxxW gating region, preventing water wire formation. Upon protonation of a third His19 residue the helix packing was observed to subtly rearrange such that this constriction was removed. Protonation of

the final His19 led to fully open pores in each case. The most dramatic change was observed for the AT-4+ simulation of Series 1 as is evident from the large increase in Cα RMSD. In this simulation the repulsion between His19 sidechains led to a splaying of the helices, such that the only remaining contacts between helices were in the N-terminal region of the protein. The C-terminal region was held together only by an indirect hydrogen-bonding network and coordination of two His27 residues on adjacent chains to a chlorine ion. This splayed, conical structure appeared strikingly similar to that of crystal structure A/M2 solved at pH 6.5 in the presence of β-octylglucoside. To compare this, the CGMD structures were least squares superpositioned onto the A/M2 structures, with sequence alignment to the HxxxW motif. The superimposition is shown in Figure 3:16. The Cα RMSD for the overlapping regions is 3.6 Å.

Comparison of the Series 1 AT-2+ and AT-4+ structures and pore profiles to those of the A/M2 channels from X-ray (PDB id 3C9J) and solution NMR (PDB id 2RLF) showed several similarities (Figure 3:16). The pore profile and structure of the AT-2+ and solution NMR structures show constrictions in the HxxxW region and the Cα RMSD between overlapping regions was 2.3 Å. The pore profile of the AT-3+ representative structure shows that this constriction has been mostly removed, while the AT-4+ shows no constriction in this region.

As evidenced by the lower Cα RMSDs of the other AT-4+ simulations, the large-scale changes in bundle geometry were only observed for one simulation (Series 1). Inspection of the other AT-4+ simulations demonstrated that the reason no splaying was observed was the presence of one or more chloride ions in the pore (Figure 3:17). These were able to coordinate to the positively charged His19 residues and reduce the unfavourable electrostatic interactions between them, thus stabilizing the channel architecture.

The effect of His19 protonation was most clearly quantified by measuring changes in orientation of His19 sidechains. Figure 3:18 shows the variation of a property δ-His (see caption text) over time which corresponds to a measure of the distances of the His19 residues from the pore centre. The His19 sidechains repel each other once protonated and reorientate in order to maximise the distance between them. Observation of the variation of δ-His over time shows that the rearrangement of these sidechains occurs rapidly, within the first few ns of simulation. The distance remains approximately constant for the remainder of the simulation. This demonstrates that the protonation causes fast rearrangement of His19 sidechains, while the rest of the bundle takes between 5-30 ns of simulation to 'catch up' (as the Cα RMSDs continue to increase).

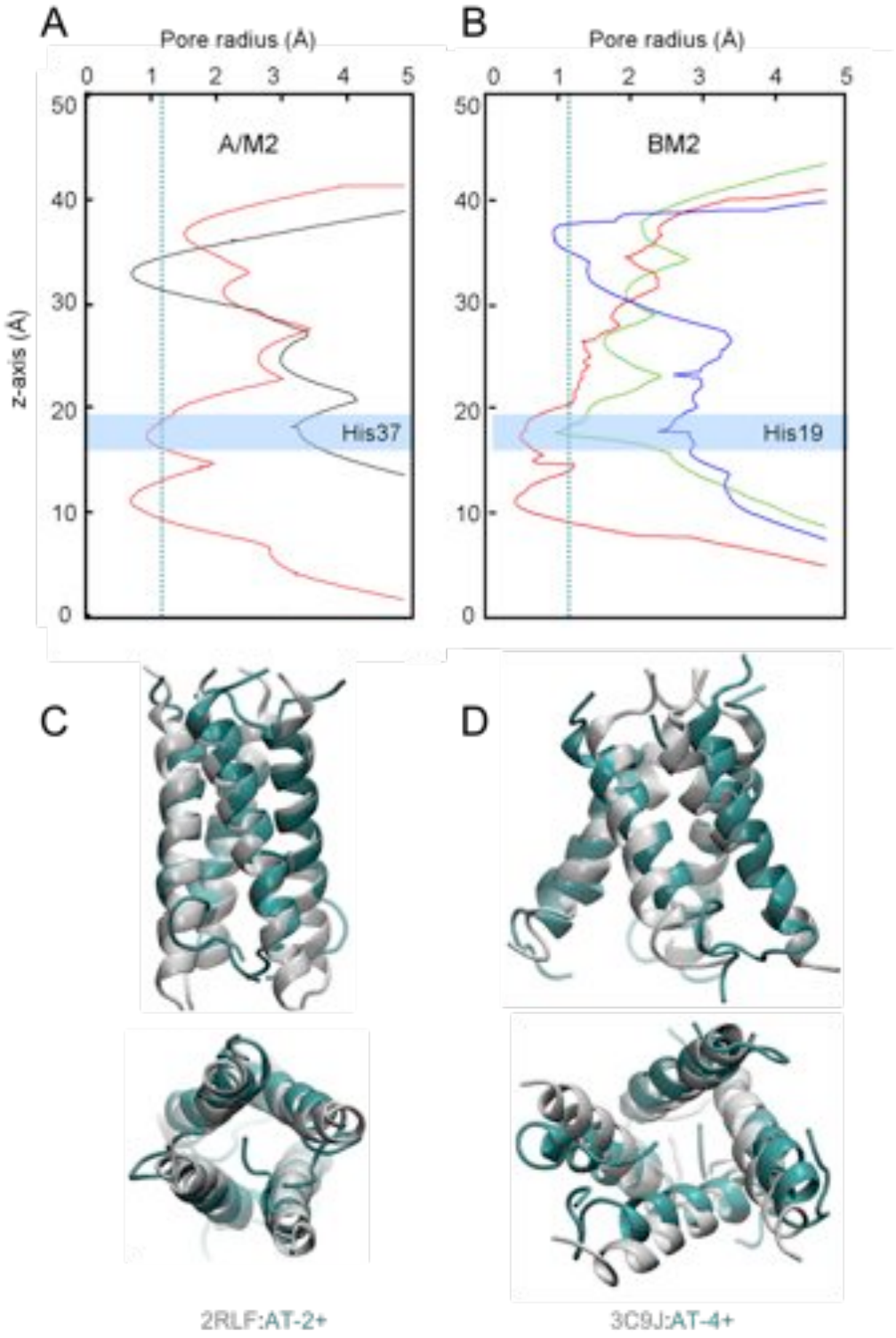


Figure 3:16. A: Pore profiles of the A/M2 solution NMR (red) and X-ray (black) structures. A constriction in the His37 region is observed in the NMR structure whereas the pore is open in this area in the X-ray structure. B: Pore profiles of representative frames from series 1 simulations. AT-2+ (red), AT-3+ (green), and AT-4+ (blue). The constriction in the His19 region is mostly removed in AT-3+ and completely removed in AT-4+. The pore profiles of AT-2+ and the A/M2 structures are similar in form as are those of AT-4+ and the X-ray structure. The dotted blue line shows the approximate van der Waals radius of a water molecule. C: Superimposition of the AT-2+ and the NMR structure. D: Superimposition of the AT-4+ and X-ray structures. The experimental structure is shown in grey in each case. The side view and the view down the pore from the C-terminus are shown.

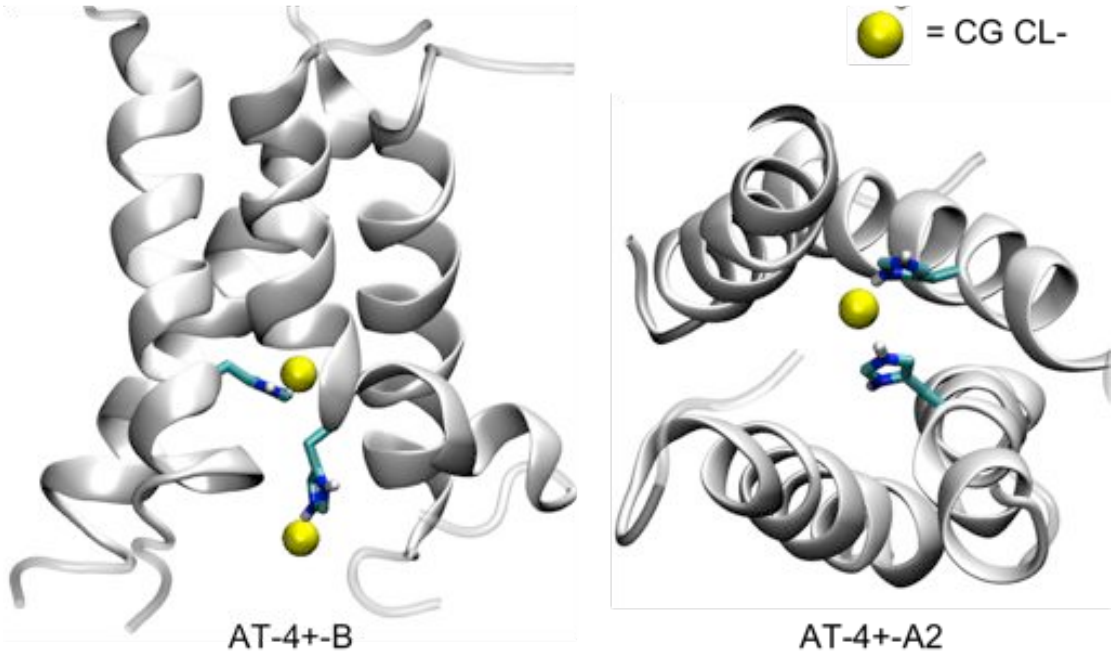


Figure 3:17 Coordination of charged His19 residues to chloride ions in AT-4+ simulations from Series 1A and Series 2. The presence of chloride in the pore reduced the Coulombic repulsion between charged His19 residues and stabilized the channel structure in these simulations, preventing the splayed structure observed in the Series 1 simulation form being observed. A large bend in one helix of AT-4+B is observed where the channel has rearranged to allow the presence of two chloride ions in the pore.

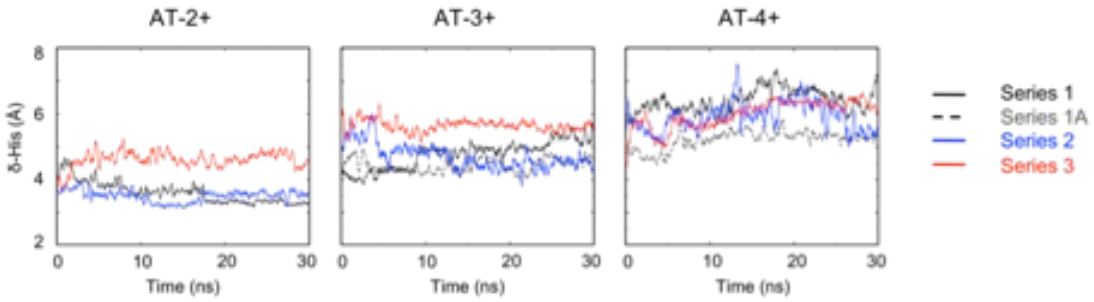


Figure 3:18. .Reorientation of His19 sidechains over time. δ-His is the average of the centre of mass (COM) of the nitrogen atoms of the His19 imidazole rings from the COM of the His19 Cα particles and corresponds to a measure of the sidechain distance from the centre of the pore.

The N-terminal region of each system remained hydrated throughout the 30 ns simulation for all systems. Z-coordinates (corresponding to the distances along the pore) of three water molecules in this region are plotted for the final 10 ns of AT-2+, AT-3+ and AT-4+ simulations in Figure 3:19.

In the AT-2+ system no waters were observed to cross the boundary of the His19 sidechains into the C-terminal region. In the AT-3+ simulation one of the three water molecules escapes from the protein pore. In the AT-4+ simulation there is no barrier to water molecules crossing this region, and no water molecules remained in the pore for more than 10 ns before exchanging with bulk water.

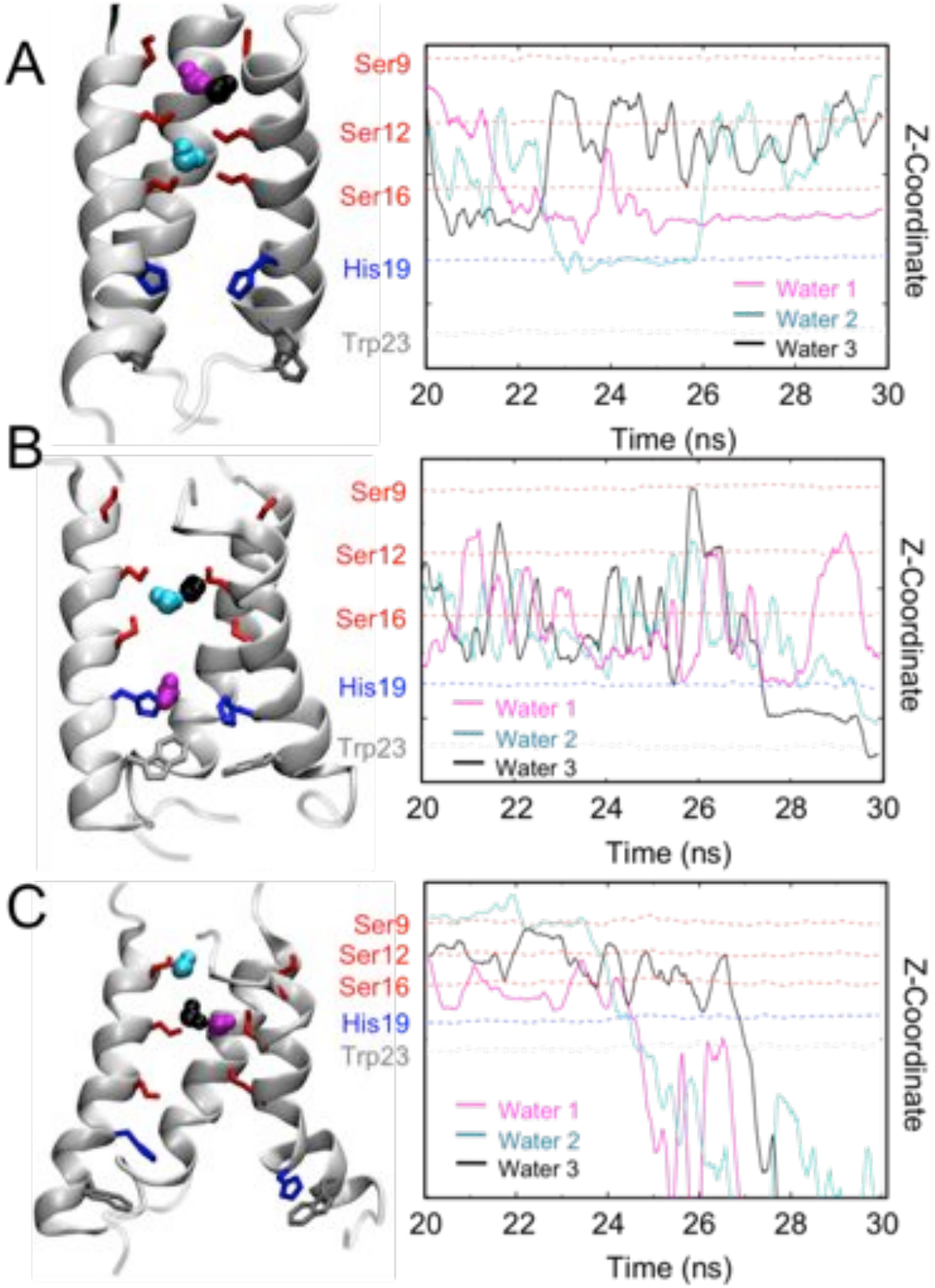


Figure 3:19 Position and dynamics of water molecules in the channel pore in the AT-2+ (A) AT-3+ (B) and AT-4+ (C) systems from simulation series 1. The z-coordinates of 3 molecules are plotted over the final 10 ns of each simulation. The initial positions of the water molecules within the pore are shown in the left hand side figures. Sidechains and water molecules are coloured corresponding to the key used in the plots. The average positions of the Cα atoms of serine residues are shown as dashed red line, His19 is shown as a dashed blue line, and Trp 23 is shown as a dashed grey line. The waters are seen to remain within the N-terminal region in the AT-2+ simulation. One water molecule crosses the His19 boundary in AT-3+, and all water molecules are free to move between bulk and N-terminal region in the AT-4+ system.

3. Discussion

3.1. CGMD to Model Membrane Protein Folding

The results of the CGMD two-stage folding simulations demonstrated that a helix monomer of the putative TM region is stable in a transmembrane orientation. When four helices were free to move within a bilayer, they were found to laterally diffuse until they contacted each other to form a loose oligomer. In each of the five cases, despite the initial association process being intrinsically stochastic (leading to different configurations of the initial oligomeric bundles), these loose tetrameric assemblies were found to rearrange to give a converged structure. This is consistent with the two-stage folding model in which individual components of an α-helical bundle are able to rearrange to give the global minimum.

The tetrameric bundle was found to have left-handed packing and corroborates an earlier predicted model of the TM domain. The CGMD-generated TM domain was found to be highly dynamic during CGMD simulations, with "switching" motions of the channel occurring on the timescale of tens of nanoseconds. Furthermore, in one simulation a meta-stable two-fold symmetric conformation persists for 2 μs. Taking into consideration the smoother energy landscape in CGMD simulations (leading to a faster rate of sampling) compared to ATMD, we might expect that a meta-stable structure would appear stable in an ATMD study under currently feasible simulation lengths. This highlights the usefulness of CGMD in sampling longer timescales that are currently not readily accessible to ATMD simulations.

3.2. Conversion to Atomistic Resolution (CG2AT)

The CGMD-generated structure was found to be stable upon conversion of the protein-bilayer system to atomistic resolution and ATMD simulation for 30 ns. The main change following CG2AT conversion was an increase in the overall tilt of the helix bundle. Despite this change in tilt, the overall geometry of the bundle remained similar. This variation in tilt appears to be related to the two levels of resolution compensating for hydrophobic mismatch in slightly different ways. The most probable explanation for the difference is that in the CGMD simulations the helices are constrained to be α-helical, whereas in ATMD simulations the helices are more flexible and able to alter their structure slightly by *e.g.* sidechain reorientation.

3.3. The HxxxW Motif and its Role in Channel Gating

A possible gating mechanism of the channel *via* His19 protonation was explored by a series of simulations in which additional His19 residues were charged in successive stages. As the number of positively-charged histidine residues of the HxxxW motif increased, the helix packing was shown to change to accommodate these charges, leading to the formation of a water wire in the AT-3+ simulation.

In one of the four simulations with all four histidine residues protonated the helices were shown to splay out at the C-terminus such that the tetramer resembled the A/M2 X-ray structure solved in BOG detergent (PDB id 3C9J). It has been proposed that the 3CJ9 structure does not correspond to a functioning proton-selective channel since there does not seem to be a barrier in the HxxxW gating region to water permeation [21]. There is also a *β*-octylglucoside (BOG) detergent molecule bound in the C-terminal end of the channel pore. The similarity of the simulated structure to the X-ray structure is remarkable and perhaps shows that the C-terminal splaying may not necessarily be an effect of a detergent molecule binding, but may in fact enable detergent binding.

The results of these simulations suggest that the protonation of the third His19 residue is responsible for channel opening. This is consistent with the findings of Hu et al, who used solid state NMR spectroscopy to determine the pKas of the histidine tetrad of the A/M2 channel [22]. The first two protons were proposed to bind in a cooperative manner, with protonation of the third histidine coinciding with channel opening. However, it is important to note that the timescales simulated are such that transient water wire formation in the AT-2+ state may not be observed, therefore it cannot be ruled out that the AT-2+ state is sufficiently open for proton gating. Furthermore, limitations in classical force fields mean that imidazole-imidazolium dimers (in which the extra proton is shared equally between two His19 residues) are unattainable in these simulations. Despite these simulations not leading to a firm conclusion of the exact mechanism of BM2 channel gating, it should be noted that in the case of the extensively studied A/M2 proton channel there is still debate in the literature over the mechanism of gating.

4. Further Work

4.1. CGMD to Predict Oligomer State: Pentamer Control Study

In the CGMD simulations of the BM2 TM tetramer formation, the total number of helices in the bilayer patch was four, limiting the maximum size of the bundle to a tetramer. A further question to be addressed was the possibility of using the CGMD method to predict the oligomeric state of TM bundles. To test the effectiveness of the CGMD method in predicting oligomer state, a system with 5 helices present in the bilayer was simulated. Figure 3:20 shows the initial configuration of the pentameric system.

Of 5 simulations, only one resulted in a pentamer in which all five helices were spanning the bilayer. In the other four simulations one or two of the helices adopted an interfacial orientation whilst the remaining TM helices associated to form a trimer or tetramer. In the simulation in which a pentameric TM bundle was observed, the cluster of five helices that results does not resemble a symmetric pore; rather, the bundle appears to maximize contacts between each of the helices, such that a tightly packed trimer that is associated with a tightly packed dimer results.

From visual inspection it is clear that the pentamer simulation does not lead to a well-defined pore. However, to fully test the ability of CGMD to predict oligomer states a much larger study would need to be undertaken, including positive controls in which a higher oligomeric structure is the expected outcome (*e.g.* pentameric phospholamban, and hexameric LS3). Despite the limitations of the test performed here, the results of these simulations provide positive signs that there may be some way to use high-throughput CGMD simulations in the prediction of oligomer state.

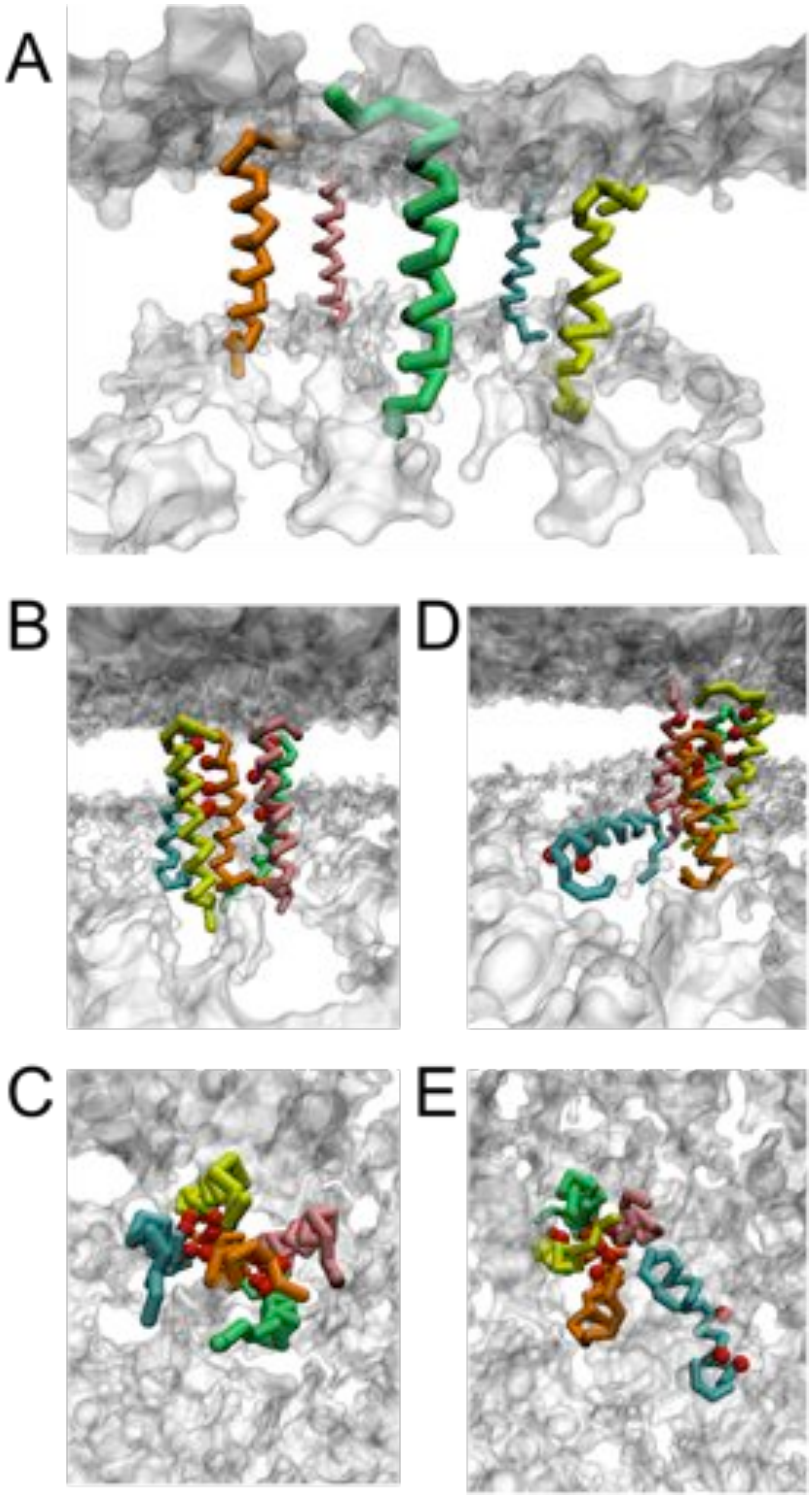


Figure 3:20. Control simulations of five BM2TM helices in a bilayer. A: Starting configuration with helices approximately equidistantly spaced in the bilayer. B: in 1 simulation out of 5 a pentameric bundle formed. C: View of B from C-terminus. 3 of the helices pack as a trimer whilst 2 others are associated as a dimer. D: A more common outcome of these simulations. One or more of the helices become interfacial in orientation. The remaining helices pack tightly as a trimer or tetramer. E. View from the C-terminus of D. Helix backbone particles are shown as a coloured trace. Serine sidechains are shown as red spheres.

5. Summary

Coarse-grained molecular dynamics simulations were used to generate a converged structure of the transmembrane domain of the BM2 proton channel, based on sequence and knowledge of oligomer state alone. The longer timescales accessible to CGMD simulation allowed the dynamics of the tetrameric bundle to be observed, with concerted motions of the helices occurring on the tens of nanoseconds timescale. This structure was then converted to an atomistic representation and was shown to be stable upon ATMD simulation. Simulation at this atomistic resolution allowed a potential gating mechanism *via* His19 protonation to be probed. When three of the four histidine residues in the putative HxxxW gating motif were protonated the channel was shown to be open enough to permit water wire formation, however it is also possible that transient water wire formation could occur in the AT-2+ channel on longer timescales not captured during these ATMD simulations.

Ideally, it would be possible to predict membrane protein structures from sequence alone (*i.e*. with no prior knowledge of oligomer state). Initial CGMD simulations of a system of five BM2 TM helices showed that no stable pentameric structure was formed. However there is some way to go before CGMD can be used with confidence in predicting the oligomeric state of a given membrane protein from sequence alone.

Following publication of this work a solution NMR structure of the full BM2 proton channel (from two overlapping constructs) was published [23]. The following Chapter is based on simulations of the solution NMR structure of the BM2 TM domain in a bilayer environment.

6. References

[1] R. G. Paterson, M. Takeda, Y. Ohigashi, L. H. Pinto, and R. A. Lamb, "Influenza B virus BM2 protein is an oligomeric integral membrane protein expressed at the cell surface.," *Virology*, vol. 306, no. 1, pp. 7-17, Feb. 2003.

[2] J. M. Cuthbertson, D. A. Doyle, and M. S. P. Sansom, "Transmembrane helix prediction: a comparative evaluation and analysis.," *Protein engineering, design & selection: PEDS*, vol. 18, no. 6, pp. 295-308, Jun. 2005.

[3] M. Amico et al., "PONGO: a web server for multiple predictions of all-alpha transmembrane proteins.," *Nucleic acids research*, vol. 34, no. Web Server issue, pp. W169-72, Jul. 2006.

[4] A. Sali and T. L. Blundell, "Comparative protein modelling by satisfaction of spatial restraints.," *Journal of molecular biology*, vol. 234, no. 3, pp. 779-815, Dec. 1993.

[5] P. J. Bond and M. S. P. Sansom, "Insertion and assembly of membrane proteins via simulation.," *Journal of the American Chemical Society*, vol. 128, no. 8, pp. 2697-704, Mar. 2006.

[6] K. Bryson, L. J. McGuffin, R. L. Marsden, J. J. Ward, J. S. Sodhi, and D. T. Jones, "Protein structure prediction servers at University College London.," *Nucleic acids research*, vol. 33, no. Web Server issue, pp. W36-8, Jul. 2005.

[7] K. Otomo, A. Toyama, T. Miura, and H. Takeuchi, "Interactions between histidine and tryptophan residues in the BM2 proton channel from influenza B virus.," *Journal of Biochemistry*, vol. 145, no. 4, pp. 543-554, 2009.

[8] P. J. Bond, J. M. Cuthbertson, S. S. Deol, and M. S. P. Sansom, "MD simulations of spontaneous membrane protein/detergent micelle formation.," *Journal of the American Chemical Society*, vol. 126, no. 49, pp. 15948-9, Dec. 2004.

[9] T. Carpenter, P. J. Bond, S. Khalid, and M. S. P. Sansom, "Self-assembly of a simple membrane protein: coarse-grained molecular dynamics simulations of the influenza M2 channel," *Biophysical Journal*, vol. 95, no. 8, pp. 3790-3801, 2008.

[10] L. A. Kelley, S. P. Gardner, and M. J. Sutcliffe, "An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies.," *Protein engineering*, vol. 9, no. 11, pp. 1063-5, Nov. 1996.

[11] C. Ma et al., "Identification of the pore-lining residues of the BM2 ion channel protein of influenza B virus.," *The Journal of biological chemistry*, vol. 283, no. 23, pp. 15921-31, Jun. 2008.

[12] E. Psachoulia, P. W. Fowler, P. J. Bond, and M. S. P. Sansom, "Helix-helix interactions in membrane proteins: coarse-grained simulations of glycophorin a helix dimerization.," *Biochemistry*, vol. 47, no. 40, pp. 10503-12, Oct. 2008.

[13] M. S. Hall, Benjamin A, Sansom, "Sidekick: A pipeline and framework for high throughput molecular dynamics simulations," *In Preparation*.

[14] P. J. Stansfeld and M. S. P. Sansom, "From Coarse Grained to Atomistic: A Serial Multiscale Approach to Membrane Protein Simulations," *Journal of Chemical Theory and Computation*, vol. 7, no. 4, pp. 1157-1166, Apr. 2011.

[15] V. Carnevale, G. Fiorin, B. G. Levine, W. F. Degrado, and M. L. Klein, "Multiple Proton Confinement in the M2 Channel from the Influenza A Virus.," *The journal of physical chemistry. C, Nanomaterials and interfaces*, vol. 114, no. 48, pp. 20856-20863, Oct. 2010.

[16] G. Fiorin, V. Carnevale, and W. F. DeGrado, "Structural biology. The flu's proton escort.," *Science (New York, N.Y.)*, vol. 330, no. 6003, pp. 456-8, Oct. 2010.

[17] A. L. Polishchuk, J. D. Lear, C. Ma, R. a Lamb, L. H. Pinto, and W. F. Degrado, "A pH-dependent conformational ensemble mediates proton transport through the influenza A/M2 protein.," *Biochemistry*, vol. 49, no. 47, pp. 10061-71, Nov. 2010.

[18] J. Wang, J. X. Qiu, C. Soto, and W. F. DeGrado, "Structural and dynamic mechanisms for the function and inhibition of the M2 proton channel from influenza A virus.," *Current opinion in structural biology*, vol. 21, no. 1, pp. 68-80, Feb. 2011.

[19] J. A. Mould et al., "Influenza B virus BM2 protein has ion channel activity that conducts protons across membranes.," *Developmental cell*, vol. 5, no. 1, pp. 175-84, Jul. 2003.

[20] L. H. Pinto and R. A. Lamb, "The M2 proton channels of influenza A and B viruses.," *The Journal of biological chemistry*, vol. 281, no. 14, pp. 8997-9000, Apr. 2006.

[21] T. A. Cross, M. Sharma, M. Yi, and H.-X. Zhou, "Influence of solubilizing environments on membrane protein structures.," *Trends in biochemical sciences*, vol. 36, no. 2, pp. 117-25, Feb. 2011.

[22] J. Hu et al., "Histidines, heart of the hydrogen ion channel from influenza A virus: toward an understanding of conductance and proton selectivity.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 18, pp. 6865-70, May 2006.

[23] J. Wang, R. M. Pielak, M. A. McClintock, and J. J. Chou, "Solution structure and functional analysis of the influenza B proton channel.," *Nature structural & molecular biology*, vol. 16, no. 12, pp. 1267-71, Dec. 2009.

Chapter 4

Solution NMR structure

1. Unusual coiled-coil structure

The structure of the TM domain (residues 1-33) of BM2 as solved by solution NMR spectroscopy is shown in Figure 4:1. Similarly to the CGMD-generated model, and structures of A/M2, a left-handed bundle is adopted. This structure appears to be somewhat 'loosely' packed - gaps between sidechains in the N-terminal region are clear. This structure was solved at pH 7.5 and is described in reference [1] as corresponding to the closed form of the channel. However, there is a pore in the channel (Figure 4:1C) that is large enough for a continuous water channel to form. Another surprising feature of this structure is the orientation of the His27 residues, which point outwards into an area that would normally be occupied by hydrophobic lipid tails. These aspects are discussed later.

As mentioned previously, this structure was obtained by solution NMR spectroscopy of the construct in dihexanoylphosphatidylcholine (DHPC) micelles. To determine if this conformation is likely to correspond to one which would be adopted in a bilayer environment, the structure was simulated in a dipalmitoylphosphatidylcholine (DPPC) lipid bilayer as described in the following sections.

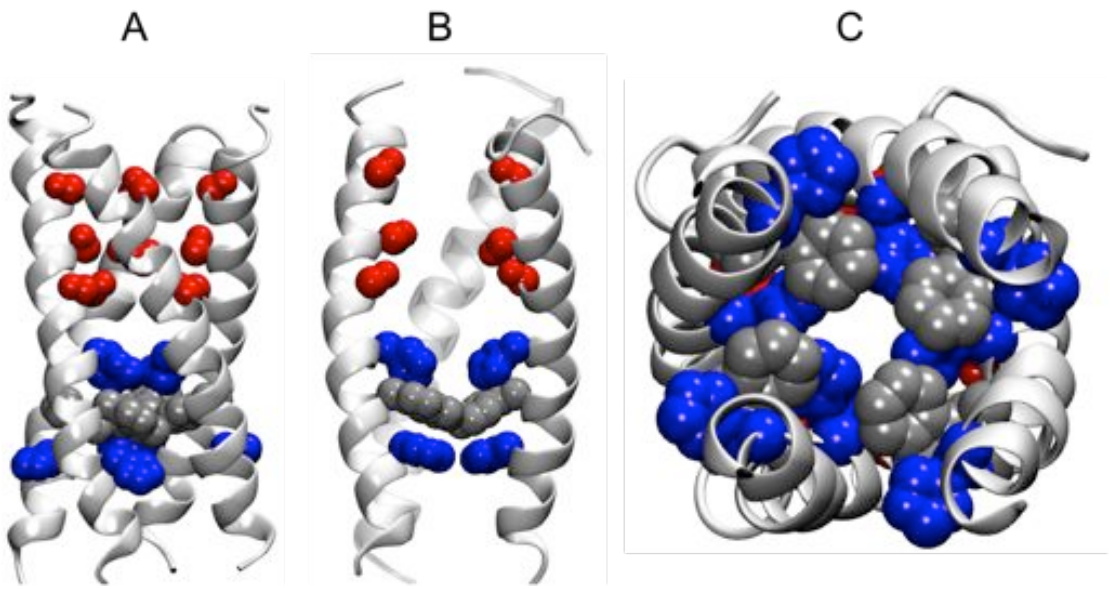


Figure 4:1. Predicted Pore-Lining Residues in the Solution NMR Structure. A and B show the relatively large gaps between helices in the N-terminal Ser-region. In B the front helix is removed for clarity. C shows the channel as viewed through the pore from the C-terminus. Key residues are displayed as space filling spheres: serine (red), histidine (blue) and tryptophan (grey).

1.1. Generation of Bilayer System for the NMR Structure

The NMR structure of the BM2 transmembrane domain was converted to a coarse grain (CG) model by fitting backbone particles and key sidechain atoms to the NMR structure. This process has been automated using in-house scripts (Figure 4:2). A bilayer self-assembly simulation was then performed as described earlier, by restraining the CG protein in a box full of randomly orientated and positioned CG-DPPC lipids. Upon CGMD simulation the lipids spontaneously self-assembled to form a bilayer within 20 ns of CGMD simulation.

Inter-helical restraints were necessary to maintain structural integrity of the channel, even once the bilayer had equilibrated. Removal of the restraints between helices and using GNM constraints for individual helices led to rapid collapse of the structure. This could be a consequence of the rigid angles defined in the CG model that do not allow for the slightly larger angles in a coiled-coil helix compared to an ideal α-helix. However the main change in structure appeared to be an 'inwards collapse' of the helices, in which the pore size decreased substantially. The BM2-bilayer system was simulated for 1 μs using CGMD then converted to an atomistic representation using the fragment-based approach, as described previously.

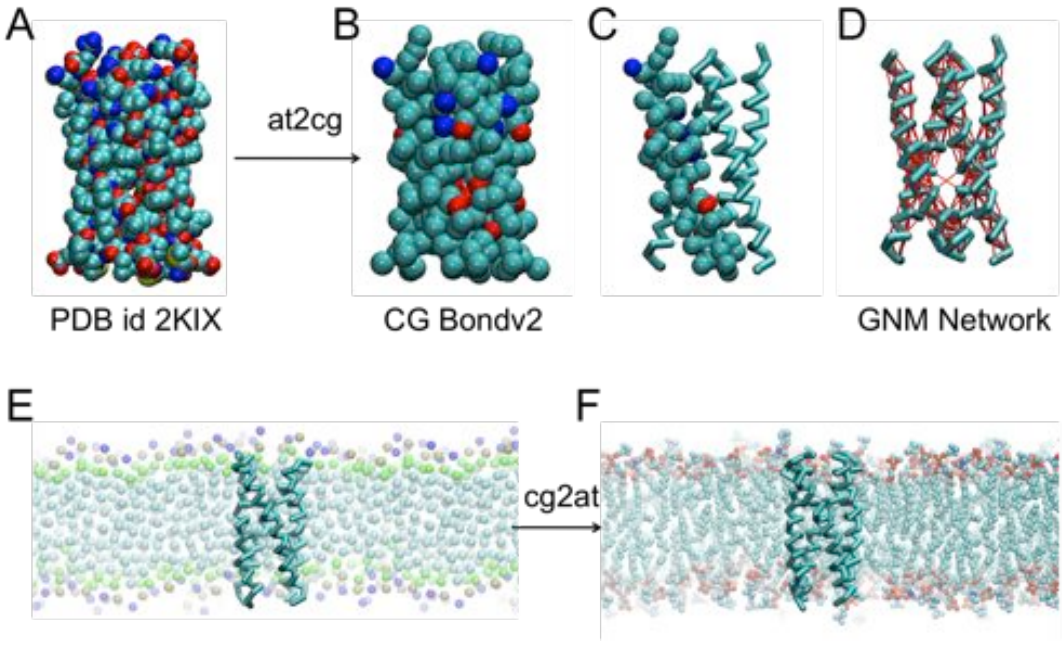


Figure 4:2. Generation of the NMR Structure-Bilayer Simulation System. A: The initial NMR structure. Each particle is shown as a coloured sphere. B: The coarse-grain structure. C: Coarse grain model showing 3 of the 4 helices as a Cα trace. D: The Gaussian Network Model (GNM) constraints (shown in red) applied to the chains. The GNM constraints were applied as described in [2]. Use of the GNM model allows some conformational change whilst maintaining the overall structure of a given protein. This has been employed in e.g. [3] E: Position of the CG model in the bilayer after 1 μs CGMD with positional restraints on the protein. F: The CG2AT converted system prior to simulation.

1.2. ATMD Simulations of the BM2 Solution NMR Structure

Table 3 shows the simulations performed from this assembly. Two force fields were used for this study, OPLS-UA [4], [5] and GROMOS96 [6]. Prior to each of these production runs, simulations were performed with positional restraints on the protein for 2 ns to allow the lipids to relax around the protein. Following this, the restraints were removed and production simulations of 100 ns performed. To investigate the effects of the phase of the lipid bilayer on the structure and dynamics of BM2 the OPLS simulation was repeated twice at 323 K and three times at 300 K. The gel-to-fluid phase transition temperature of DPPC is ~314 K, therefore the bilayer is expected to be in the gel phase in the simulations performed at 300 K, and the fluid phase in the simulations performed at 323 K. In one GROMOS simulation ("GROMOS1") the waters initially in the channel pore following the CG2AT conversion were removed. Waters were allowed to enter the pore during the 2 ns equilibration. In the second GROMOS simulation and all other simulations the waters added in the conversion were left during the positional restraints equilibration.

Name	Force field	Duration / ns	Temperature / K	Number of Simulations	Waters in pore
OPLS1, OPLS2, OPLS3	OPLS-AA (protein), OPLS-UA (lipid)	100	323	3	Yes
GROMOS1	GROMOS	100	323	1	No
GROMOS2	GROMOS	100	323	1	Yes
300K, 300K-1, 300K-2	OPLS-AA (protein), OPLS-UA (lipid)	100	300	3	Yes

Table 3. Simulations performed of the solution NMR BM2 structure in a bilayer. The simulation names are coloured as they appear in plots in the remainder of this Chapter. The number of simulations corresponds to the number of simulations performed from the same starting point with different initial velocities.

The simulations of the NMR structure were assessed in terms of stability and dynamics by monitoring certain properties during each simulation (Figure 4:3). Similar trends are observed for all 3 sets of simulations:

1) The Cα RMSDs generally plateau within the first 20 ns of simulation, however some large changes are observed, generally corresponding to a loss in secondary structure *i.e.* protein unfolding.
2) The OPLS simulations at both temperatures show the greatest Cα RMS fluctuations in the C-terminal region of the structure. The GROMOS simulations do not show this trend. The RMSFs in the simulation at 300 K are similar to the Cα RMSD between the 15 models within PDB id 2KIX.
3) There is a gradual loss in α-helicity over time that is most pronounced in the higher temperature OPLS simulations. The dotted lines (lower panel of Figure 4:3) correspond to the maximum α-helical count that was observed when the NMR structure was simulated with positional restraints. There is a clear distinction between this and the starting α-helical count, demonstrating that the NMR structure loses helicity as soon as the restraints are removed.

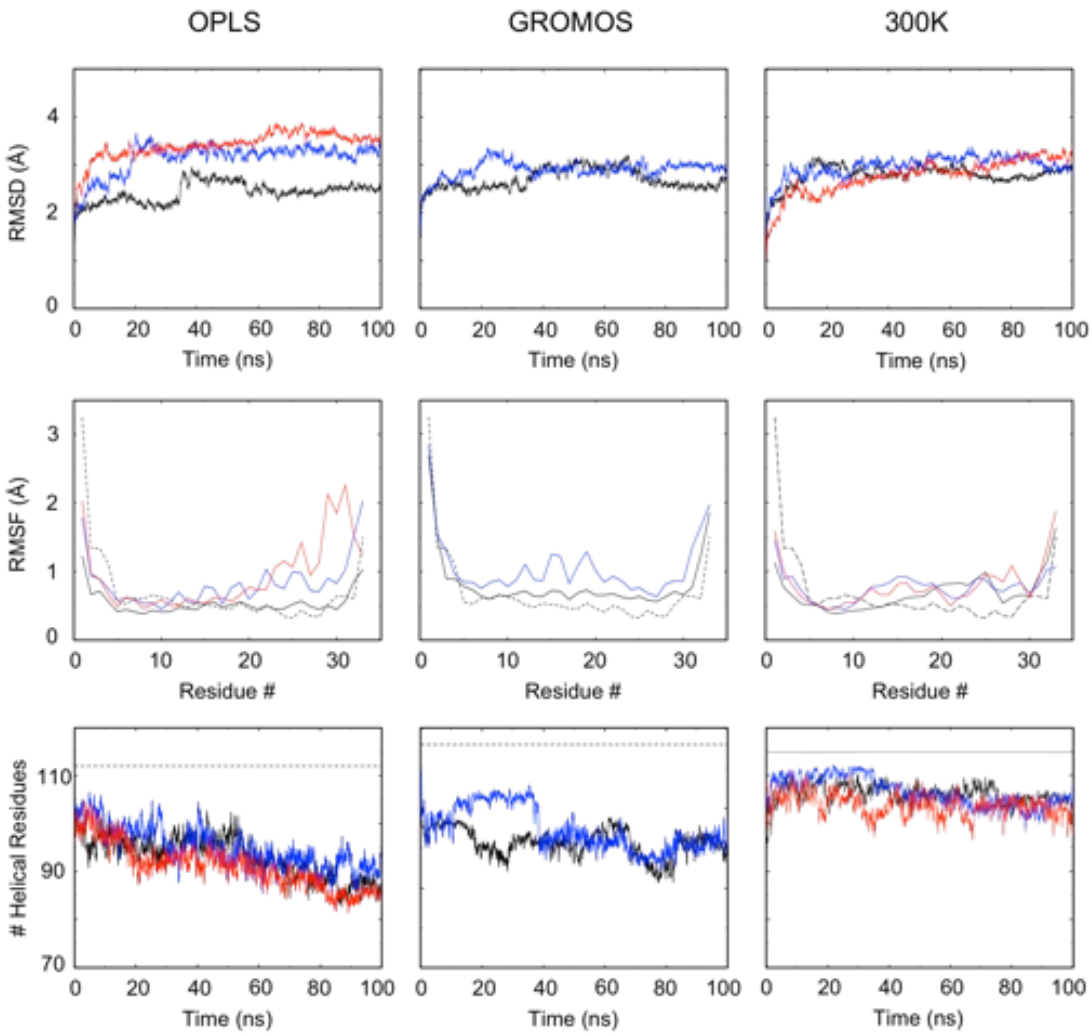


Figure 4:3. Cα RMSDs (top panel), Cα RMS fluctuations (middle panel) and α-helical residue count (lower panel) for each simulation. RMSDs were calculated for all residues not including the 2 end residues at C- and N-termini. RMSFs are shown for the first chain only for clarity. The dashed line in the RMSF plots shows the Cα RMSD between the 15 structures deposited in the protein data bank. The dotted line in the secondary structure plots corresponds to the maximum α-helical count observed for the NMR structure with positional restraints maintained. The traces of repeat simulations are coloured according to Table 3.

These initial observations suggest that the NMR structure is not particularly stable in a DPPC bilayer environment using conventional ATMD force fields.

Pore profile analysis demonstrates that the relatively large pore present in the NMR structure decreases in radius rapidly, within 1 ns of ATMD simulation (Figure 4:4). This was the case for simulations using both the GROMOS and OPLS force fields. Water that is initially present in the large cavities was forced out during this relaxation. Following this initial equilibration, constrictions are observed in the His19 region, as well as at the N-terminus of the pore.

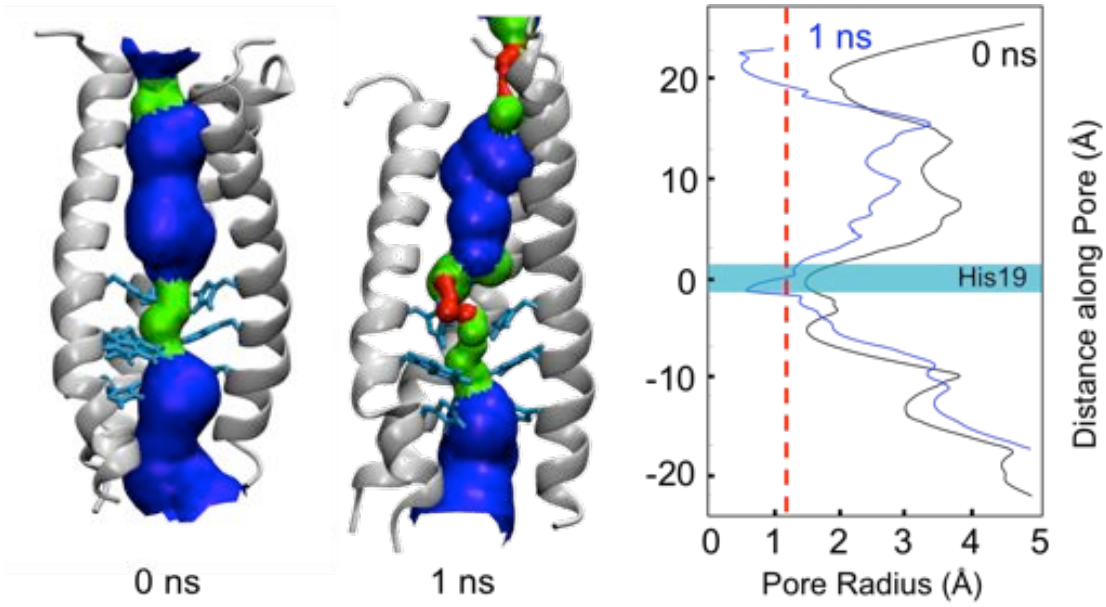


Figure 4:4. Left: 3D representations of the pore surfaces of the BM2 NMR structure at beginning of simulation and after 1 ns. These structures are taken from simulation OPLS1. Only three helices are displayed for clarity. The His19, Trp23 and His27 residues are shown in cyan in a ball-and-sticks representation. The colours of the 3D profiles are as follows: red, radius < 1.2 Å (no water can pass); green, 1.2 Å ≤ radius ≤ 2.3 Å (single-file water); blue, radius > 2.3 Å (multiple waters may pass). The plot on the right shows the average pore radius relative to the His19 region. The approximate van der Waals radius of a water molecule is plotted as a red dashed line.

A particularly unstable region appears to be around the His27 residue, where, after 1 ns of simulation, the backbone adopts a distorted, non-ideal geometry (Figure 4:4). The orientation of the His27 sidechains in the solution NMR structure was previously identified as an unusual structural feature [7] as the His27 residues are modelled as lipid-facing despite experimental evidence suggesting that these residues have a higher pKa than the His19 residues [8].

To quantify the changes in helicity of the tetramer at each residue, the VMD [9] plug-in Bendix [10] was used to calculate curvature of the initial structure and of the structures after 100 ns of simulation. Plots of the curvature averaged over the 4 helices are shown in Figure 4:5 and 3D representations are shown in Figure 4:7. It is clear that the main distortion is observed in the His27 region. Some bending of the helices is also observed near the N-terminus, particularly in the GROMOS simulations.

The structure that most resembles the initial NMR structure in terms of bending of the helices is from one of the simulations performed at 300 K. This shows the least bending overall, although the main distortion is again in the His27 region rather than approximately one turn earlier (central panel; Figure 4:7) as observed in the initial NMR structure.

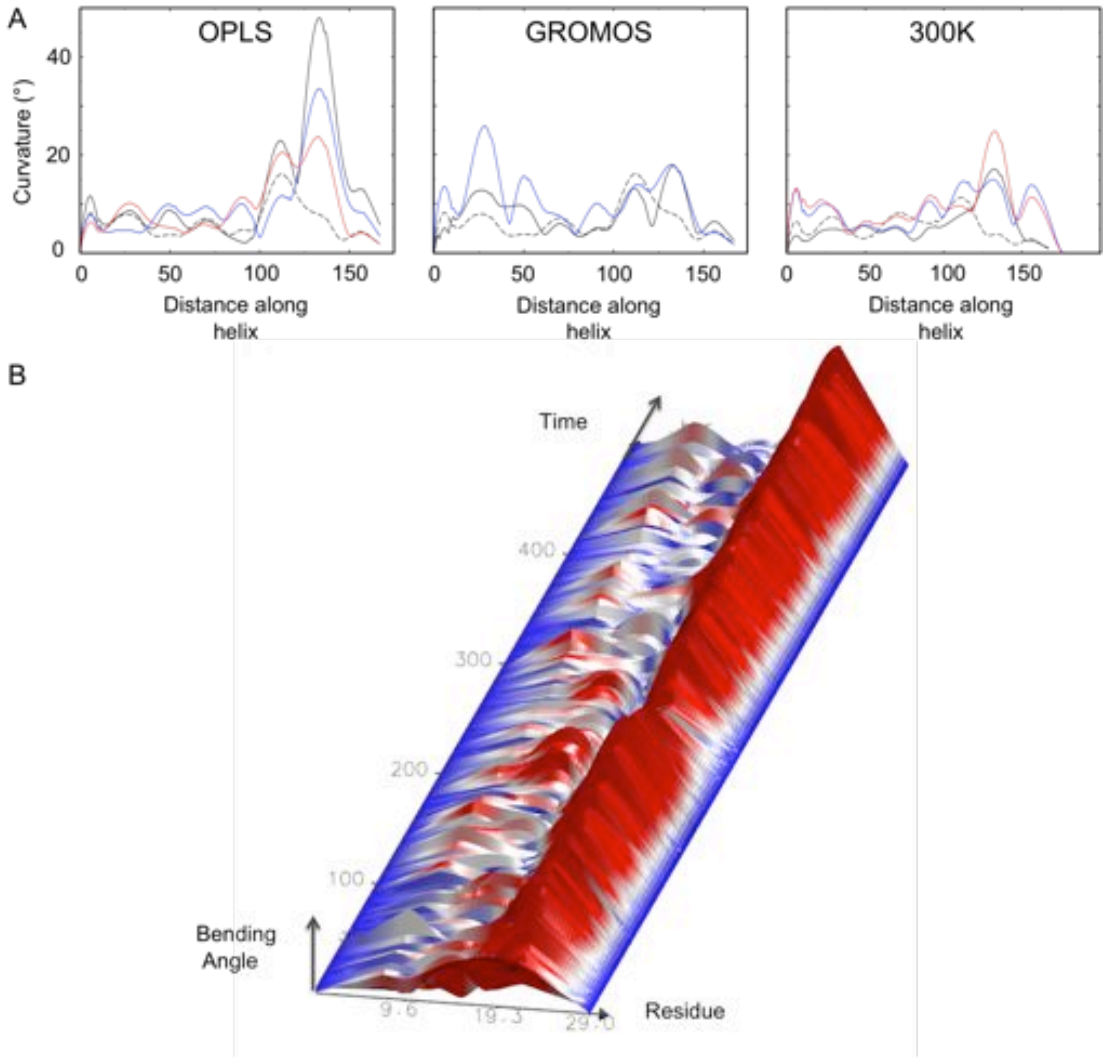


Figure 4:5. A: Curvature plots for each set of simulations. The curvature, or bend, has been averaged over the four helices and over 100 ns simulation time in each case. The plots are coloured according to Table 3. The dashed line corresponds to the calculated curvature of the NMR structure PDB id 2KIX. B: An example of the evolution of a single helix bending angle plot over 100 ns simulation.

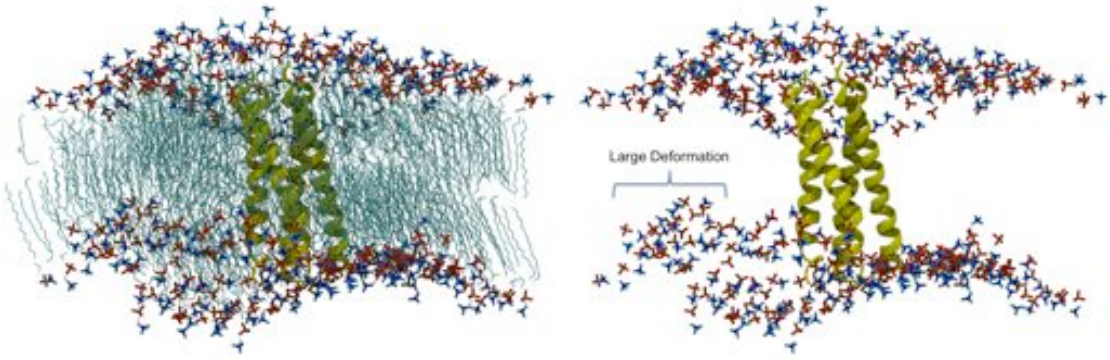


Figure 4:6. Deformation of DPPC Bilayer in the Gel Phase. The protein is shown in cartoon representation, coloured yellow. The lipid headgroups are shown coloured according to atom type (red = oxygen, blue = nitrogen, tan = phosphorus). In the left hand side figure the lipid acyl chains are shown in cyan. Waters are omitted for clarity. Large deformations are observed in the system at 300 K.

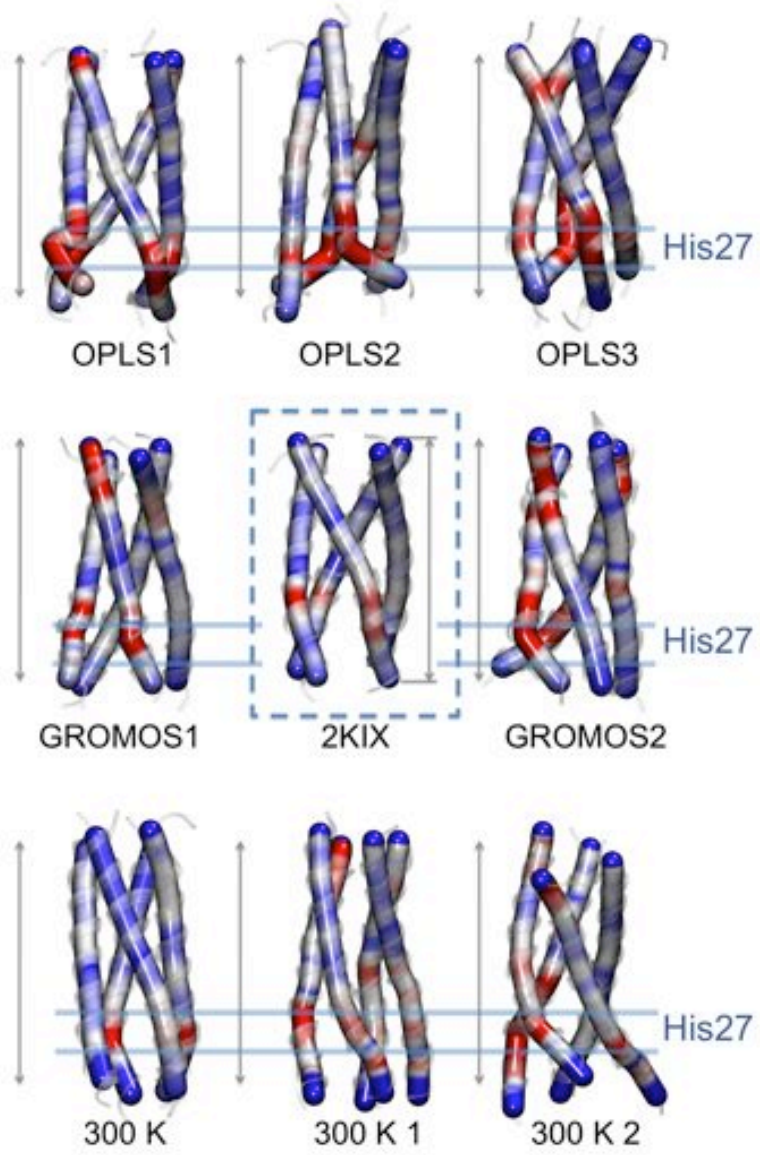


Figure 4:7. Visualization of helix curvature of the initial solution NMR structure (central panel; labeled 2KIX) and final structures of each 100 ns atomistic MD: simulations, indicated as labeled. The threshold angle used for colouring is 20°. The approximate location of the His27 residues is shown by blue lines as indicated. The helix backbones are superimposed on the Bendix helices in transparent grey, shown in cartoon format. The distance between N-terminal and C-terminal residues of the initial NMR structure is indicated by a grey arrow.

However, it should be noted that at this lower temperature the lipids become somewhat ordered. When the lipids are in fluid phase (323 K) the thickness of the bilayer gradually increases to a maximum around the protein. When the lipids are in the gel phase (300 K) different behaviour is observed. The deviation in bilayer thickness is much greater than in the fluid phase, although the average thickness is similar (Figure 4:6). The region around the protein is thicker than in the fluid phase simulations at 323 K, whereas the thinnest region of the bilayer is thinner than that at 323K (Figure 4:8). The plots of lipid density show a clear distinction between the system at the two temperatures. In the coarse-grained and atomistic simulations at 323 K there are specific binding sites

(regions of high density) around the protein but elsewhere the density becomes averaged. At 300 K the lipids all remain in a similar position relative to the protein, indicating the higher degree of ordering. A comparison of diffusion constants estimated from the MSD of each lipid relative to the protein indicate that the lipids move an order of magnitude less in the simulations at 300 K (data not shown).

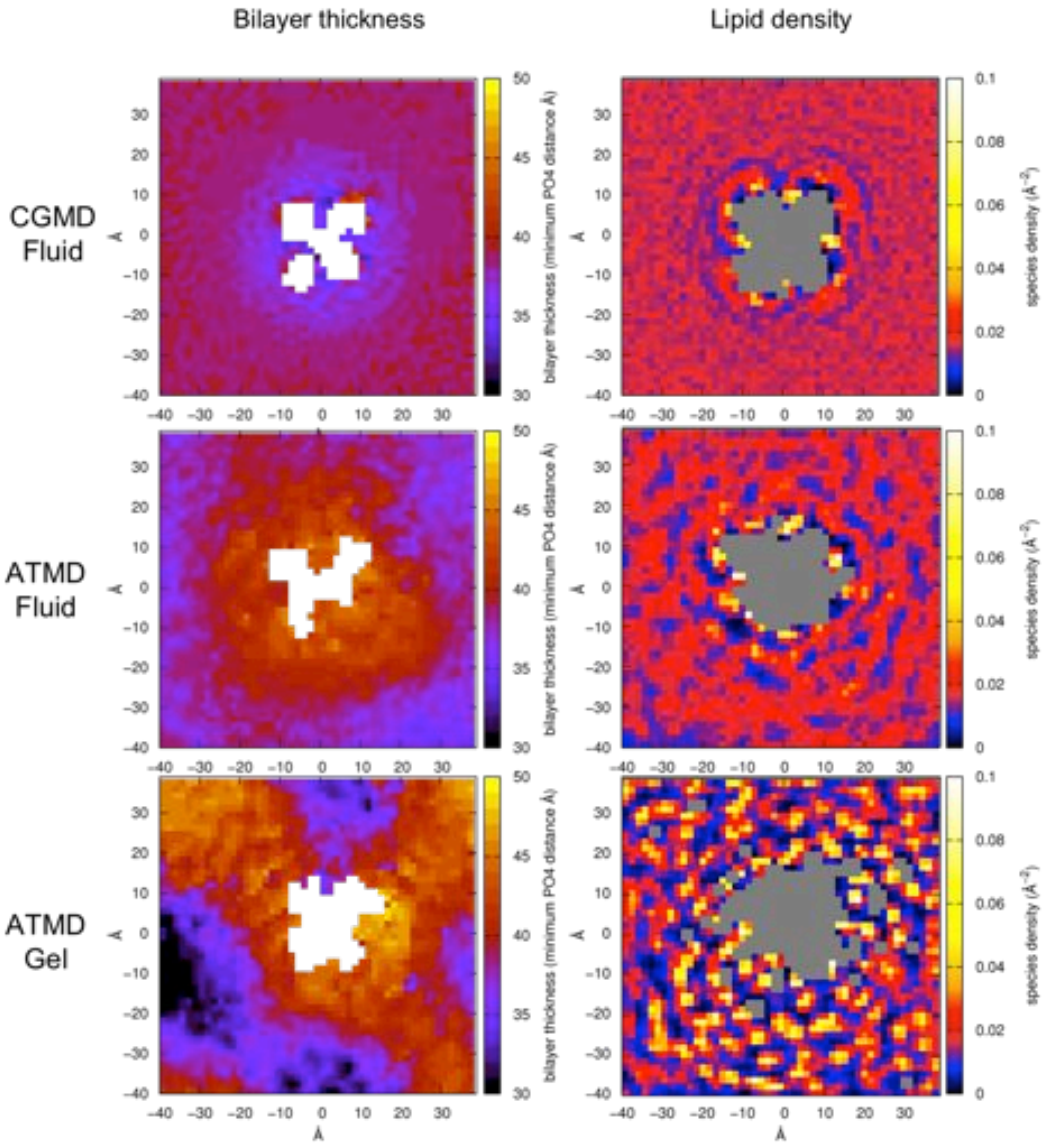


Figure 4:8. Bilayer thickness and lipid densities in coarse-grained (top) and atomistic simulations at 323 K (middle) and 300 K (bottom). Left panel: The minimum distance between phosphorus groups (calculated in terms of PO4 or P8 particles) is shown as an average over the 100 ns of simulation time. Right panel: The species density of the first carbon of a glycerol unit. Simulation CGMD Fluid = NMR structure with GNM restraints at 323 K. ATMD Fluid = simulation OPLS1 performed at 323K, ATMD Gel = simulation 300 K. Calculations performed in the reference frame of the protein.

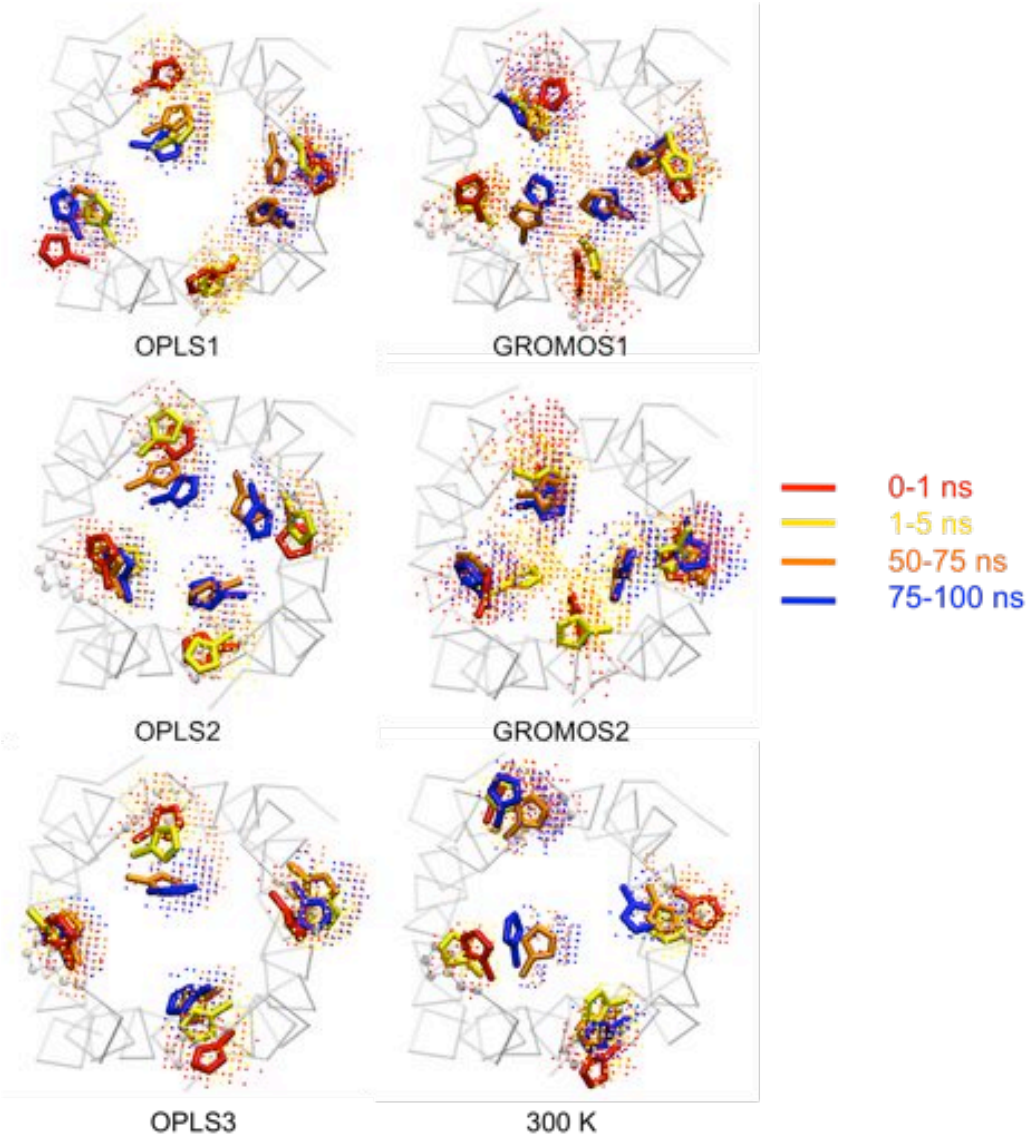


Figure 4:9. Spatial distribution of His27 sidechains during 100 ns simulation. The helix backbones are shown as a thin grey trace with the initial orientation of His27 residues displayed in stick-and-ball representation. The space sampled by the His27 sidechains during each block of time is shown as dots, coloured as indicated. The average position of the His27 sidechain for each time is shown in stick representation.

To investigate the conformations of the His27 residues further, the spatial distributions of the His27 sidechains (the space sampled by the His27 sidechains relative to the protein backbone) over each of the 100 ns simulations were calculated. In Figure 4:9 the spatial distributions of these sidechains are displayed as coloured dots. The average coordinates for the times indicated in the key are shown in the corresponding colours.

The average positions of the His27 sidechains over the final 25 ns of each simulation are shown in blue. It can be seen by the final position of these sidechains that at least one of the helices reorientates in each simulation to enable the His27 sidechain to point into the pore. The simulation at 300 K displays the least fluctuation in His27 sidechain position, nethertheless one of the His27 residues is pore-facing by the end of 100 ns

simulation. The GROMOS simulations show the most spatial sampling by the His27 sidechains (indicated by larger region occupied by dots). In each simulation the space sampled by the His27 sidechains in the first nanosecond of simulation is comparable to that sampled in the final 25 ns of simulation. This indicates that the His27 residues are most unstable at the start of the simulation, and most relaxation occurs within the first nanosecond.

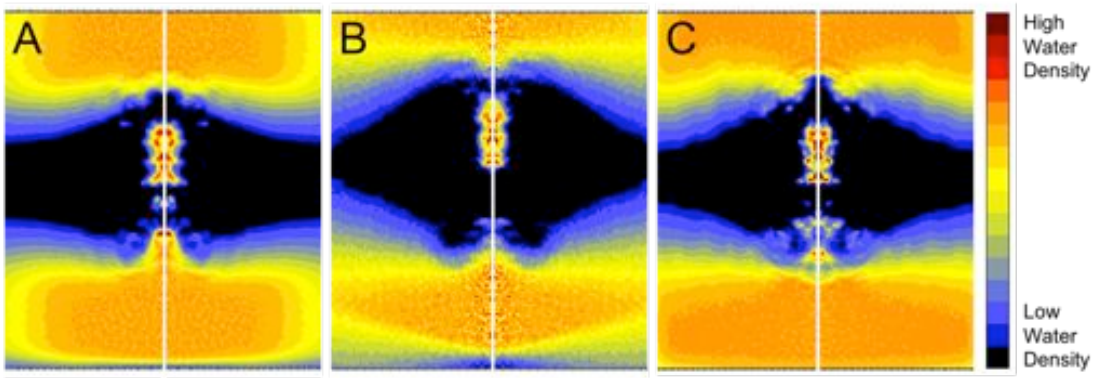


Figure 4:10. Radial density distributions of water averaged over simulation A: OPLS1, B: GROMOS1, C: 300 K. This data was generated using the program gridcount. The water densitiy is calculated in the reference frame of the protein (centred). The tetramer bundle axis is indicated by a white line.

In each of the simulations, the N-terminal serine-rich region was hydrated throughout. This was the case whether or not waters were included in the pore at the start of the 2 ns equilibration simulation with positional restraints on the protein. (GROMOS1 simulation, Figure 4:10B). The pore of the NMR structure is wide enough that waters were observed to enter the channel within the 2ns timescale. Most water molecules present in the pore at the start of each simulation were exchanged with bulk water during the simulation but some remained in the channel throughout. The z-coordinates (where the z-axis corresponds to the bundle axis) of two such water molecules in simulations at 323 K and 300 K are plotted in Figure 4:11. Waters that are initially in the HxxxW region are rapidly expelled from the channel within the first nanosecond of simulation (green trace in Figure 4:11). Following the formation of the constriction in the HxxxW region no waters were observed to pass through this region. The waters in the pore are highly dynamic and exchange with each other on the timescale of a few nanoseconds. This was the case even at lower temperature.

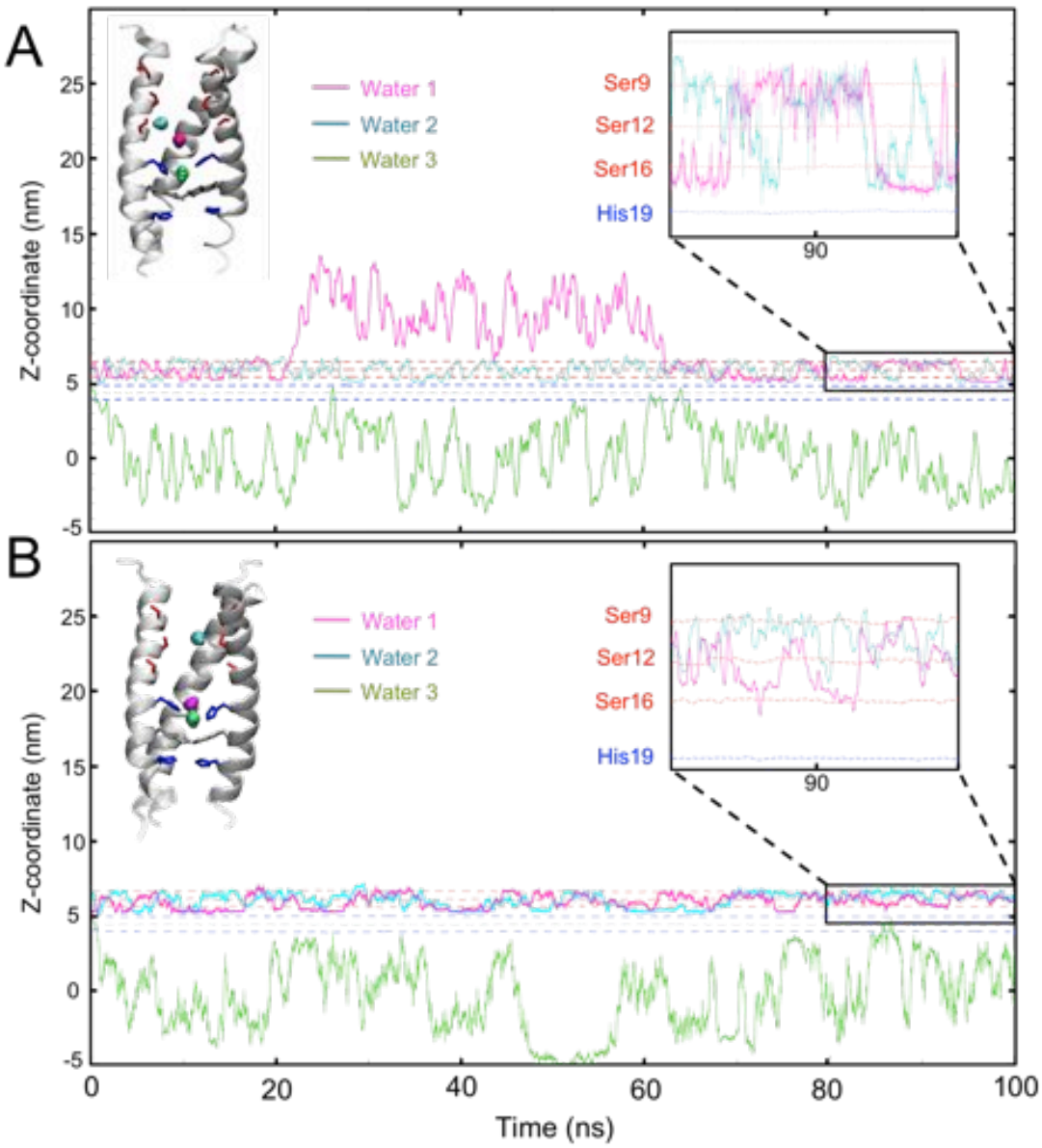


Figure 4:11. Position and dynamics of water in pore at 323 K (A) and 300 K (B). The waters in the pore are highly dynamic on the 10s of nanoseconds timescale. Hydrogen bond duration with residues typically lasts only a few ns before switching with another water molecule. Waters initially in the HxxxW region are expelled from the channel within the first nanosecond (green trace). Z-coordinates of the Ser9, Ser12 and Ser16 Cα atoms are plotted as a red dashed line. Z-coordinates of Cα particles of His19 and His27 are shown as blue dashed lines, and Trp23 in grey. The water molecules initially in the HxxxW region are rapidly (~1 ns) forced out of the channel (green trace in each plot).

2. Discussion

The simulations performed in this Chapter of the solution NMR structure and the previous Chapter of the CGMD-generated structure of the BM2 TM domain allow certain conclusions to be drawn. The two models show common features despite apparent structural differences. These are discussed in the following subsections.

2.1. Structural Comparison of CGMD-Generated Model and Solution NMR Model

Both the CGMD-generated model and the NMR structures have left-handed packing. The predicted pore lining residues Ser9, Ser12, Ser16, His19 are shown to be facing the pore in both structures. The Trp23 residue which was found to adopt either an interfacial or pore-lining conformation in the molecular models from reference [11] was pore-lining in the CGMD-generated model but seen to adopt an interfacial orientation during simulation of the NMR structure in a DPPC bilayer.

A comparison of the overlapping backbone residues of the following structures are compared:

1) The CGMD-generated model
2) Final frames of AT-2+, AT-3+ and AT-4+ simulations (from series 1)
3) The solution NMR structure 2KIX
4) Final frames of OPLS1 and GROMOS1

The largest Cα RMSD is between the 'splayed' AT-4+ structure with all other structures. The Cα RMSD between all other structures is below 5 Å. It can be seen that the GROMOS1 final structure is a closer fit to the CGMD model than to the initial solution NMR structure. The Cα RMSD fit between three published experimental A/M2 structures are in the range of 4.0- 4.6 Å (3BKD versus 1NYJ = 4.0 Å, 3BKD versus 2RLF = 4.0 Å, 1NYJ versus 2RLF = 4.6 Å) [12]. It is reassuring that the RMSDs between different simulation structures are comparable to those between experimental A/M2 TM structures.

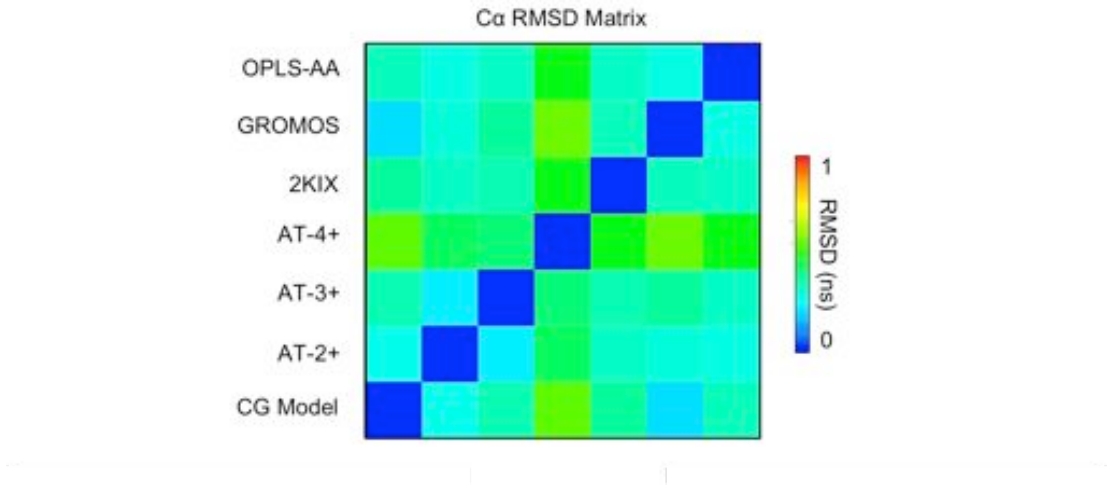


Figure 4:12. RMSD matrix of backbone particles of overlapping regions of i) CGMD-generated model ii) end point of simulation AT-2+, iii) AT-3+, iv) AT-4+, the solution NMR structure 2KIX, and the structures resulting from 100 ns simulation of 2KIX with the vi) GROMOS and vii) OPLS force fields.

2.1.1. Solution NMR Structure Refinement Compared to MD Simulation

The fact that the solution NMR structure of BM2 was determined within DHPC micelles rather than a bilayer environment has already been alluded to as a reason for instability of the structure in a DPPC bilayer. However, a further reason for inconsistencies between the NMR model in combination with conventional classical MD force fields is in the nature of the structure refinement methods used. In the case of the BM2 solution NMR structure, standard protocols were used, which may be summarised in the following stages [1]:

1) Intrasubunit NOEs involving backbone and side chain protons were assigned to completion using the 3D ^{15}N-edited and ^{13}C-edited NOESYs.
2) Remaining NOEs that could not be explained by intrasubunit distances were identified as intersubunit NOEs.
3) The secondary structure of the monomer was calculated using data from NOEs and chemical shifts, followed by a high-temperature simulated annealing protocol in which the bath temperature was cooled from 1,000 to 200 K.
4) The lowest energy monomer structure was used to generate an initial tetramer model. Simulated annealing was then performed in the presence of intersubunit NOEs and all other intrasubunit restraints. To maintain C4 rotational symmetry,

for all experimental intersubunit NOE between two adjacent subunits, four identical distance restraints were assigned throughout the tetramer.

There are a few potential sources of problems within this method. Firstly, the NMR experiments measure distances over timescales of hundreds of milliseconds. The restraints therefore are time-averaged, and it may not correspond to a realistic structure to use them in a non-time averaged sense during MD simulation. Consider for example the dynamic 'switching' behaviour observed of the BM2 tetramer during CGMD simulation in Chapter 3, where the structure alternated between slight variations of two-fold symmetric structures on a rapid timescale. The time-averaged structure of this corresponds to a four-fold symmetric structure, whereas during the simulation the majority of the time a two-fold symmetric structure is observed. It is possible that the time-averaged structure does not correspond to a major conformation (or indeed even an existing conformation). Following on from this, a likely consequence of step 3 of this procedure, the enforcement of four-fold symmetry by replicating each intersubunit distance throughout the tetramer, is that the structure is somewhat 'over-restrained'. Finally, in the simulated annealing step the structure is refined in the absence of the detergent micelle *i.e.* without accounting for the hydrophobic environment the measurements were made in. A combination of these factors may contribute to the observed instability of the NMR structure upon MD simulation. This remains a relatively unexplored problem and it may be expected that as the number of membrane protein structures solved by solution NMR spectroscopy (and indeed other structural techniques) continues to increase, the effort to more closely match the refinement protocols in solution NMR and MD simulation will increase also.

2.2. Pore Diameter

As discussed in Sections 1. and 1.2. , the NMR structure appears to have a surprisingly large pore diameter for a structure that has been described as corresponding to the "closed" state of a proton-specific channel. Representative pore surfaces of the AT-2+, AT-3+ and AT-4+ are shown in Figure 4:13 along with those of the A/M2 solution NMR and X-ray structures, and the solution NMR structure of BM2. The solution NMR BM2 pore surface appears to correspond most closely to AT-3+, which was found to be an open form of the channel.

In simulations of the NMR structure in a bilayer environment, a rapid decrease in the pore dimensions of the protein is observed. It was thought that this could be a consequence of the lateral pressure of the bilayer altering the helix packing compared to the micelle environment the structure was solved in. However, as will be discussed in Chapter 5, it was not possible to maintain the pore dimensions even in a micelle environment. Indeed, the helices are sufficiently loosely associated that upon CGMD simulation of self-assembly of a DHPC micelle in the presence of the NMR structure (with GNM restraints applied), two DHPC molecules were seen to insert between helices and interact with the polar pore-lining serine residues or water particles in the pore (Figure 4:14).

It seems probable that the structure solved in micelles is more loosely-packed than we would expect the tetramer to be in a native environment. Membrane proteins within a bilayer are subject to large local pressures [13]. This lateral pressure has been shown to influence protein properties in several cases, *e.g.* varying lateral pressure has been shown to modulate the dimerisation affinity of glycophorin A [14]. It may be expected, then, that this lateral pressure would lead to a more closely-packed helix bundle than a detergent micelle environment.

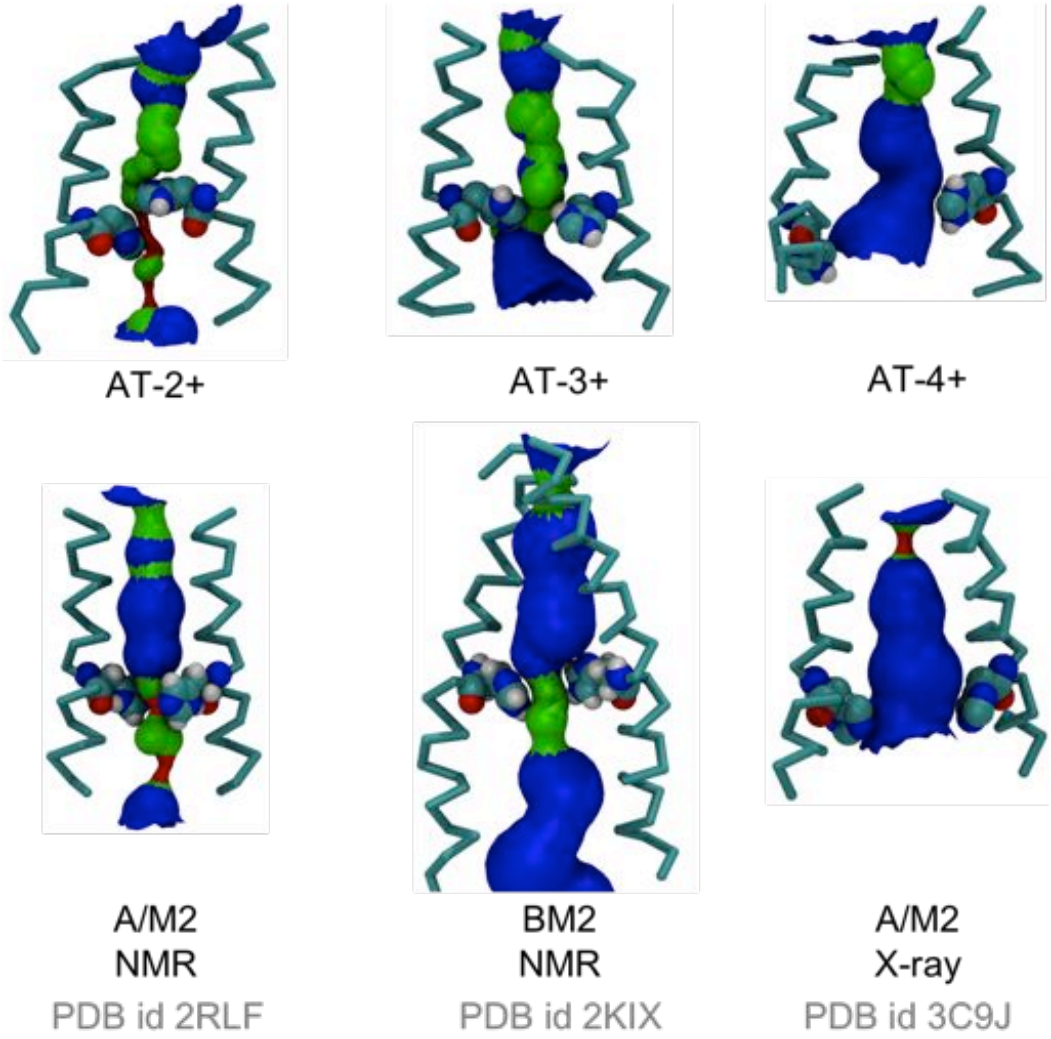


Figure 4:13. 3D representation of the pore profile surfaces of the CG2AT simulated models of the BM2 TM domain compared to solution NMR structures of the A/M2 and BM2 channels and the X-ray structure of the A/M2 channel.

However, it cannot be ruled out that other factors could be leading to this looser packing that are not captured by conventional force fields. It may be that structural waters stabilize a more 'open' structure, or that the classical mechanical nature of these force fields (*e.g*. modelling the His19 and His27 residues as 'neutral 'or 'protonated' rather than imidazole-imidazolium dimers) with no polarisation does not lead to a true stabilisation of the observed NMR conformation. Solving the structure of the BM2 TM domain using another technique such as solid state NMR would be of great use in determining the causes of these differences.

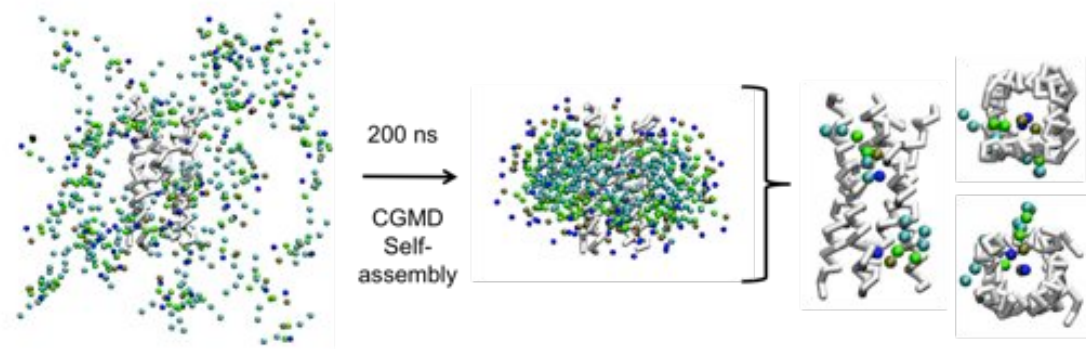


Figure 4:14. Self-assembly of a DHPC Micelle about the NMR BM2 Structure. Left: the initial system with ~80 DHPC molecules randomly orientated and positioned. Middle: The protein-detergent complex after 200 ns CGMD simulation. Right: Insertion of 2 DHPC molecules into the pore of the tetramer. One DHPC interacts with the polar serine residues, whereas the second DHPC is seen to interact with water molecules in the hydrated C-terminal region.

2.3. Orientation of His27 Sidechains in a Bilayer Environment

The simulations of the NMR structure in a bilayer with GROMOS and OPLS force fields demonstrate structural instability, particularly in the region of the His27 residue (Figure 4:15). These His27 residues tend to rotate such that they become orientated towards the pore, despite this leading to non-ideal geometry of the helix (Section 1.2.). The loss of ideal geometry is most pronounced in the OPLS simulations, but in all simulations at least one of the His27 sidechains have reorientated to point towards the channel pore.

When the simulations were performed at 300 K, it was observed in one simulation that the C-terminus retained most of the secondary structure. The His27 region was stabilized by the helices lengthening such that interactions with the hydrated headgroup region of the bilayer became possible. However, it should be noted that at this temperature the lipids were observed to enter an ordered, gel-like phase. The effect of lipid ordering on the protein structure and stability is unclear.

Membrane proteins adopting an α-helical bundle architecture are not strongly associated, indeed their dynamic nature is thought to be an important part of allowing them to perform their functions correctly [14]. Therefore it is entirely possible that the instability around the His27 region contributes to the proton channel being able to function effectively. BM2 has approximately 2-fold higher conductance than the A/M2 channel – this may be a consequence of this more variable structure in the C-terminal region of the TM domain.

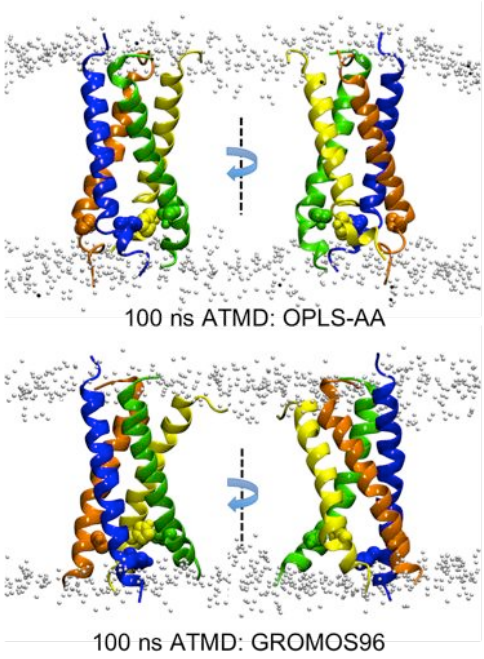


Figure 4:15: Instability of NMR Structure in DPPC Bilayer. Snapshots of the NMR TM domain following 100 ns of ATMD simulation. Upper panel: the OPLS force field. Lower panel: the GROMOS force field. Backbone particles are shown in cartoon representation and coloured by chain. The His27 sidechains are shown as spheres and coloured according to chain. Particles of the phosphatidylcholine groups are represented by grey spheres.

2.4. Solvation of the N-terminal Region

In simulations of the NMR structure and the CGMD-generated structure the serine-rich N-terminal region of the pore remains solvated on the timescales of the ATMD simulations. Figure 4:16 shows representative snapshots of these systems highlighting the presence of water in the pore. The solvation of the N-terminal region of the BM2 channel is thought to be important for channel function, although the relative importance of each serine residue for this is a matter for debate:

"Remarkably, replacement of the Ser OH of Ser9 with a Cys SH group almost entirely debilitates the channel's conductance […] only mutation at the absolutely conserved His19 is more disruptive." [11]

and *".. mutations S12A and S16A [,,,] had the largest effect, […] Replacing polar residues near the N- and C-terminal ends of the channel had less effect."* [1]

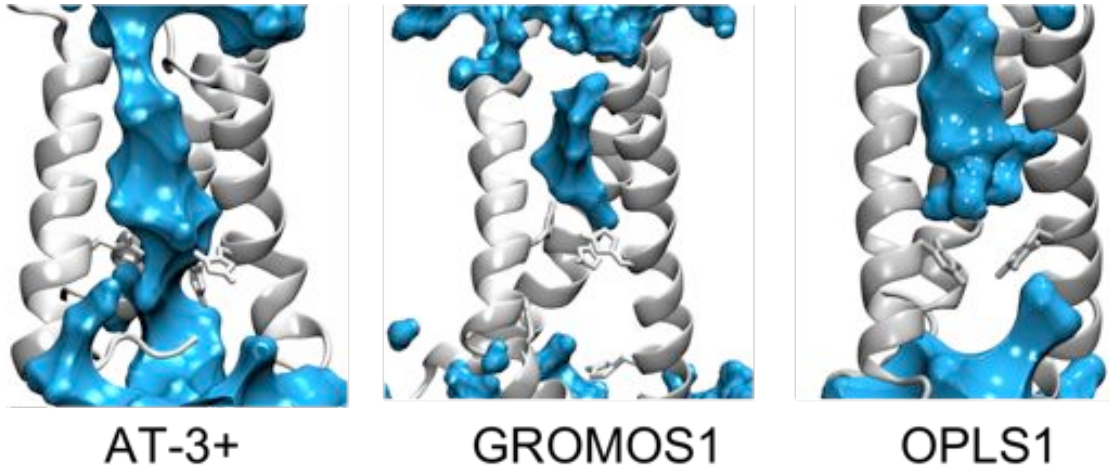


Figure 4:16. Hydration of the BM2 pore in AT-3+, GROMOS1 and OPLS1 simulations. The protein backbone is displayed in cartoon format and coloured grey. One helix is omitted for clarity. The His19 residues are shown explicitly, also in grey. Water is represented as a surface, coloured cyan.

Hydrogen bond analysis was performed to determine the number of interactions between each serine residue and water molecules during a 100 ns simulation (data not shown). No significant difference in the number of hydrogen bonds between the three serine residues (Ser9, Ser12 and Ser16) was observed. MD simulations could potentially be used to attempt to quantify the effects of mutating each of the three serine residues on gating. However, until a means of simulating the TM domain such that it is stable is available there would be significant doubt over the results of any such study.

2.5. Hydrophobic Mismatch, Lipid Binding Sites, and the Effect of Lipid Ordering

Comparison of the protein conformation in the fluid (323 K) and gel (300 K) phase bilayer phases showed contrasting behaviour. In the CGMD simulation and the fluid phase ATMD simulations the lipid packing was able to alter such that the lipid hydrophobic regions matched the local density around the protein hydrophobic core. However in simulations of the gel phase of the bilayer, it appeared to be the case that the protein altered its hydrophobic dimensions instead. Presumably the more ordered bilayer at lower temperature prevents the local distortion of lipid acyl chains necessary to minimise the hydrophobic mismatch, therefore the helices straighten to compensate. This structure is perhaps more representative of one that might be adopted if the BM2 TM domain was associated within the more ordered environment of a lipid raft.

From analysis of the lipid density about the protein, there appears to be conserved binding sites between CGMD and ATMD simulation, with the regions in the grooves between the helices being occupied by the lipid glycerol moieties throughout.

2.6. The HxxxW Motif and its Role in Channel Gating

It was not possible to study potential His19 gating mechanisms due to the instability of the His27 region of the protein. Initial simulations in which some or all of the His27 residues were charged did not lead to a more stable structure (data not shown). Therefore no further conclusions may be added on the mechanism from the previous Chapter.

3. Further Work

3.1. Generation of a CGMD Model of Residues 1 to 33

In order to assess the structure that would be obtained by the CGMD two-stage folding simulation method for residues 1-33 (the construct used in the solution NMR structure) a helical peptide corresponding to these residues was built as before, with all residues constrained as α-helical apart from terminal residues. This was found to adopt a transmembrane orientation. Clustering of 5 independent 5 μs simulations led to a wider range of structures than for the BM2 27-mer generated in Chapter 3. The top model was converted to an atomistic structure to compare to the NMR structure. The helices are much more tightly-packed in this CGMD-generated model compared to the solution NMR structure. There is also a difference in packing of the His27 residues, which face closer to the pore in the CGMD model. Simulation of this atomistic structure demonstrated that the main instability in this CG2AT model was in the serine-rich N-terminal region. These residues were too tightly packed for water to be placed in the pore upon the CG2AT conversion. This provides further evidence that it is necessary for channel stability and/or function for the N-terminal region to be solvated.

A limitation of the CGMD approach is that individual water particles correspond to four atomistic waters. This means that they are too large to fit in the pore of a self-assembled tetramer. Development of the CGMD force field to include solvent effects would be of interest.

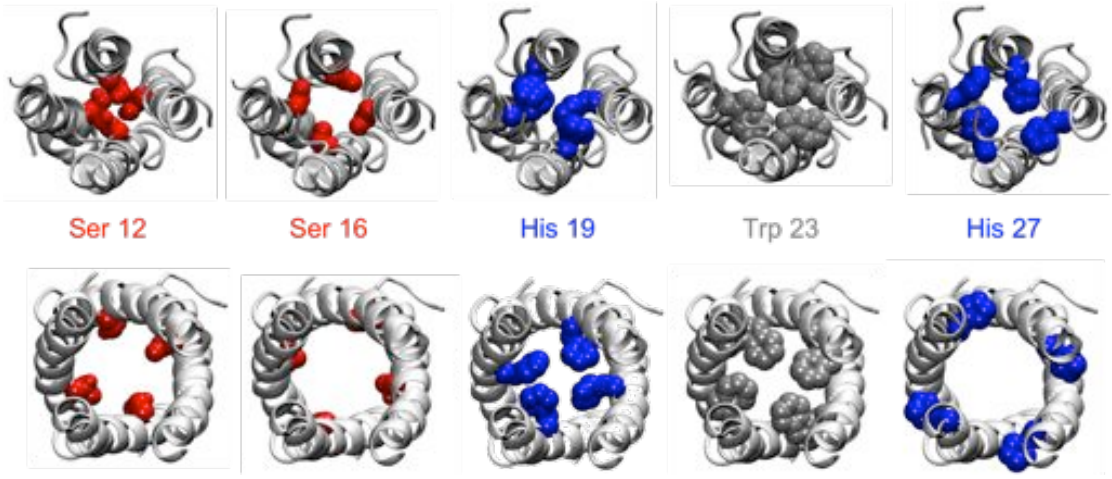


Figure 4:17. Pore lining residues of CGMD-generated model (upper panel) of residues 1-33 compared to those in the solution NMR structure (lower panel). The helix backbones are shown in cartoon format and coloured grey, with residues shown as van der Waals spheres and coloured as indicated. There are clear differences between the models. The helices are more closely-packed in the CGMD model and the orientation of His27 residues differ.

4. Summary

The NMR structure was found to be relatively unstable upon simulation in a bilayer environment, as might have been expected from some of its unusual features (large pore diameter and orientation of His27 sidechains). It is difficult to be certain if this is due to limitations in the NMR model or because the force fields used cannot correctly capture the true interactions between atoms. Studying potentially-stabilizing histidine imidazole-imidazolium dimers (that have been proposed to exist in the A/M2 channel [15]) using quantum mechanical methods is beyond the scope of this thesis. A crystal or solid-state NMR structure of the BM2 TM domain would be invaluable in understanding the reasons for the discrepancies.

Despite the differences between the CGMD-generated structure and the NMR structure, a key feature of experimental results was replicated; namely, that the N-terminal region of the channel was hydrated throughout. Indeed, a CGMD-generated model of residues 1-33 was found to be unstable in the N-terminal region where the helices were packed too tightly to include water in the pore.

The effect of the fluid *vs* gel phases of the DPPC bilayers appeared to impact upon the conformation of the TM domain. The more rigid acyl chains in the gel phase (300 K) appeared to lead to the protein increasing its hydrophobic dimensions (*via* straightening to helices) to reduce the hydrophobic mismatch. The effect of lipid

ordering on BM2 conformation is particularly of interest since it has been proposed that during virus budding it associates at the periphery of lipid-ordered and lipid-disordered phases and thus alters the curvature of the budding vesicle, permitting scission and completing the virus cycle. The possibility of BM2 being able to change its hydrophobic length by this bending of helices could play a role in this potential association.

The results of this Chapter provide some evidence to suggest that the BM2 TM domain may alter its structure to compensate for its environment. This property will be studied further in the remaining work of this thesis, where the TM domain is simulated in various environments.

5. References

[1] J. Wang, R. M. Pielak, M. A. McClintock, and J. J. Chou, "Solution structure and functional analysis of the influenza B proton channel.," *Nature structural & molecular biology*, vol. 16, no. 12, pp. 1267-71, Dec. 2009.

[2] P. J. Bond and M. S. P. Sansom, "Insertion and assembly of membrane proteins via simulation.," *Journal of the American Chemical Society*, vol. 128, no. 8, pp. 2697-704, Mar. 2006.

[3] M. S. P. Sansom, K. A. Scott, and P. J. Bond, "Coarse-grained simulation: a high-throughput computational approach to membrane proteins.," *Biochemical Society transactions*, vol. 36, no. 1, pp. 27-32, Feb. 2008.

[4] J. P. Ulmschneider and M. B. Ulmschneider, "United Atom Lipid Parameters for Combination with the Optimized Potentials for Liquid Simulations All-Atom Force Field," *Journal of Chemical Theory and Computation*, vol. 5, no. 7, pp. 1803-1813, Jul. 2009.

[5] D. Kony, W. Damm, S. Stoll, and W. F. Van Gunsteren, "An improved OPLS-AA force field for carbohydrates.," *Journal of computational chemistry*, vol. 23, no. 15, pp. 1416-29, Nov. 2002.

[6] L. D. Schuler, X. Daura, and W. F. V. Gunsteren, "An Improved GROMOS96 Force Field for Aliphatic Hydrocarbons in the Condensed Phase." 2001.

[7] T. A. Cross, "Flu BM2 structure and function.," *Nature structural & molecular biology*, vol. 16, no. 12, pp. 1207-9, Dec. 2009.

[8] K. Otomo, A. Toyama, T. Miura, and H. Takeuchi, "Interactions between histidine and tryptophan residues in the BM2 proton channel from influenza B virus.," *Journal of biochemistry*, vol. 145, no. 4, pp. 543-54, Apr. 2009.

[9] W. Humphrey, A. Dalke, and K. Schulten, "VMD: visual molecular dynamics.," *Journal of molecular graphics*, vol. 14, no. 1, pp. 33-8, 27-8, Feb. 1996.

[10] C. Dahl, "Bendix," *Manuscript in Press*, 2011.

[11] C. Ma et al., "Identification of the pore-lining residues of the BM2 ion channel protein of influenza B virus.," *The Journal of biological chemistry*, vol. 283, no. 23, pp. 15921-31, Jun. 2008.

[12] H. Leonov and I. T. Arkin, "Structure and dynamics of the influenza A M2 channel: a comparison of three structures.," *Journal of molecular modeling*, vol. 15, no. 11, pp. 1317-28, Nov. 2009.

[13] J. Gullingsrud and K. Schulten, "Lipid bilayer pressure profiles and mechanosensitive channel gating.," *Biophysical journal*, vol. 86, no. 6, pp. 3496-509, Jun. 2004.

[14] H. Hong and J. U. Bowie, "Dramatic destabilization of transmembrane helix interactions by features of natural membrane environments.," *Journal of the American Chemical Society*, p. null, Jun. 2011.

[15] J. Hu et al., “Histidines, heart of the hydrogen ion channel from influenza A virus: toward an understanding of conductance and proton selectivity.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 18, pp. 6865-70, May 2006.

Chapter 5

Protein-Detergent Complexes: Formation

1. Introduction

The solution NMR structure of the BM2 transmembrane domain showed regions of instability when simulated in a DPPC bilayer environment. However, the structure was solved in dihexanoylphosphatidylcholine (DHPC) micelles. In this chapter, protein–detergent complexes (PDCs) of BM2 with DHPC and β -dodecylmaltoside (DDM) are generated. Simulations of these systems are used to evaluate the stability and dynamics of the BM2 channel in each environment. The structures of DPPC, DHPC and DDM are shown in Figure 5:1.

1.1. Properties of Detergent Micelles

As discussed in the Introduction to this thesis, due to the hydrophobic effect it is favourable for amphiphilic species such as lipids and detergents to rearrange to sequester their hydrophobic moieties above a certain concentration [1]. For micelles this concentration is referred to as the critical micelle concentration (cmc) and is an intrinsic property of a given detergent. As the detergent concentration increases above this value the concentration of free monomers in solution remains at approximately the cmc.

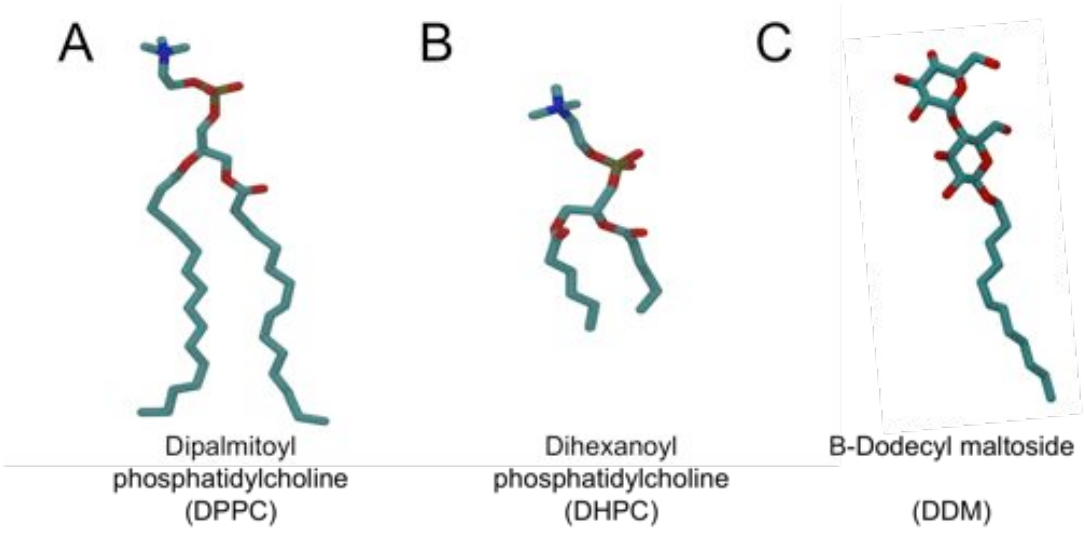


Figure 5:1. The three lipids/detergents compared in this chapter. A: dipalmitoylphosphatidylcholine, (DPPC) a bilayer-forming lipid. B: dihexanoylphosphatidylcholine (DHPC), a micelle forming detergent with the same zwitterionic headgroup as DPPC. C: β-dodecylmaltoside (DDM), a nonionic detergent that forms large micelles. Each is shown in stick representation, coloured according to atom type (cyan = carbon, red = oxygen, blue = nitrogen, tan = phosphorus). Hydrogen atoms are omitted for clarity.

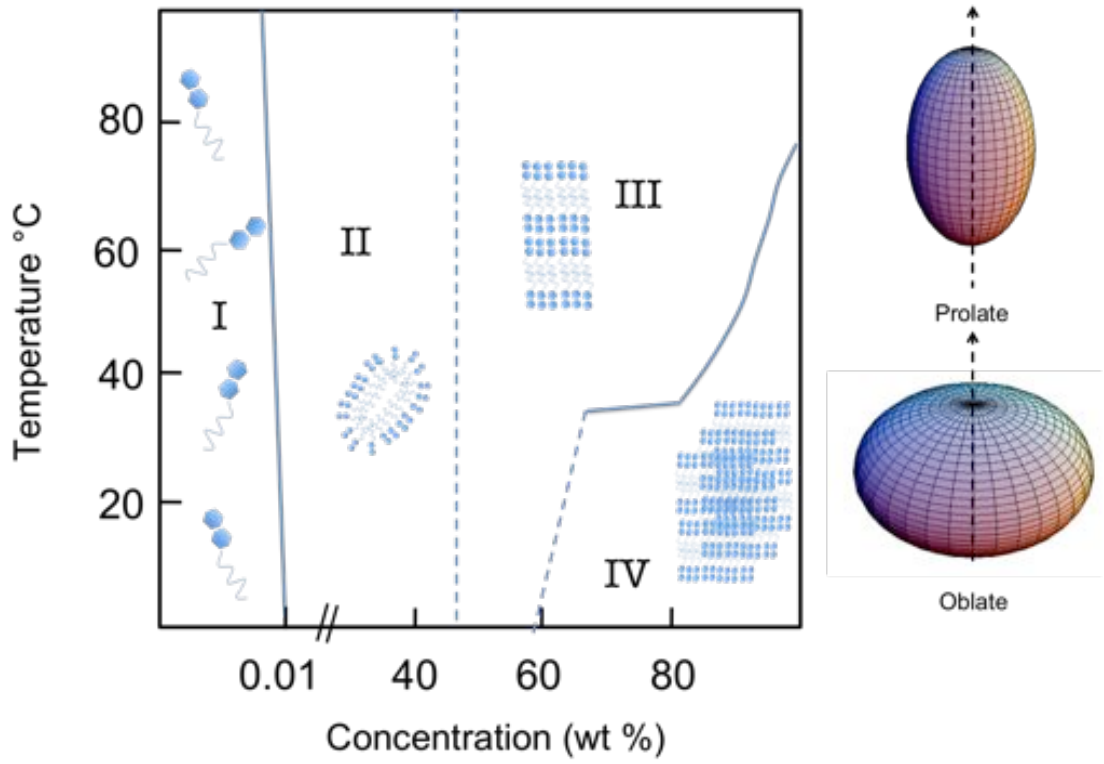


Figure 5:2. Approximate phase diagram of DDM detergent (Left). At concentrations below the cmc the detergent exists as monomers in solution (Phase I). At concentrations above the cmc various aggregated phases exist depending on detergent concentration and temperature. Phase II corresponds to the micellar phase, in which the detergent exists as a homogenous distribution of micelles (with a concentration of monomers present at the cmc). Above a certain boundary an extended-aggregate (lamellar) phase is formed (Phase III). The approximate boundary between Phase II and III is shown as a dotted line as the exact boundary was not well-defined. At high concentrations and low temperatures DDM exists as a solid (Phase IV). Schematic figures of the system in each phase are shown in cartoon form. Figure based on reference [2]. Right: Two shapes of micelles: prolate and oblate spheroids. The principle axis of symmetry is indicated in each by an arrow.

The reason for formation of discrete aggregates rather than phase separation is due to a balance of two factors i) favourable energy of sequestering the hydrophobic tails (the "hydrophobic effect" discussed in Chapter 1) and ii) unfavourable interactions between polar headgroups as the aggregate size increases. This means that there is a limit to number of detergent molecules that may be incorporated into a single micelle (the so-called aggregation number of a detergent (N_{agg}). Generally, detergents with small, charged or zwitterionic headgroups (*e.g.* DHPC) and short tails form smaller aggregates whereas detergents with large, nonionic headgroups and long tails (*e.g*. DDM) form larger aggregates.

The micellar phase only exists under certain conditions, and the phase diagrams of detergents can be complicated.

Figure 5:2 shows the phase diagram of DDM. In Phase I, below the cmc, the detergent molecules exist as free monomers. Above the cmc, it becomes favourable for the DDM molecules to aggregate. Phase II corresponds to the micellar phase, where the detergent forms a homogenous distribution of discrete micelles of size N_{agg}. As the concentration of detergent increases further, extended aggregates form (Phase III). The lamellar phase was detected for DDM but transitions to other phases (*e.g*. hexagonal or cubic) were not ruled out [2]. At higher concentrations still, DDM forms a solid (Phase IV) below temperatures of ~80 °C. In this work, Phases I-III are observed.

The shape of detergent micelles may generally be approximated by either a prolate or an oblate ellipsoid (Figure 5:2). Prolate ellipsoids are those in which the semiaxis length of the principle axis of symmetry (c) is greater than the semiaxis lengths of the two perpendicular axes (a and b). For an oblate ellipsoid the reverse is true ($c < a \sim b$).

The three properties of micelles discussed above (CMC, N_{agg} and shape) are all *measurable* properties. This means that micelle MD simulations may be directly compared to experimental data (*e.g.* Small Angle X-ray Scattering, SAXS). Table 4 shows these properties of DHPC and DDM detergents as found in the literature. There are several differences between DHPC and DDM:

Detergent	Type	CMC (mM)	Nagg	Shape
DHPC	Zwitterionic	14-15	25-35	Prolate
DDM	Non-ionic	0.18	135-145	Oblate

Table 4. Properties of DHPC and DDM micelles, as reported in the literature [3]

In this chapter the effect these intrinsic properties of the detergents may have on the structure and dynamics of protein-detergent complexes is studied.

2. Methods: Generation of Starting Configurations for Simulations

2.1. Protein-Bilayer systems

The BM2 DPPC bilayer simulations presented in this chapter are those that were previously described in Chapter 3. Briefly, these were generated by CGMD simulation of the protein NMR structure with GNM restraints in a bilayer and a representative snapshot was converted to the OPLS-UA force field [4].

2.2. Protein-Detergent Micelle Complexes

The generation of protein-detergent complexes (PDCs) is not a trivial problem. Several methods for PDC generation are possible and the methods chosen were based on the properties of the detergent in question. The generation of BM2-DHPC and BM2-DDM systems is described in the following sections.

3. Simulation of Detergent-Only Micelles

Prior to simulating protein-detergent complexes, the micelle properties of DHPC and DDM were studied in initial systems containing only the detergent and solvent.

3.1. Dihexanoylphosphatidylcholine (DHPC) Micelles

To assess the timescales required to observe micelle self-assembly from individual DHPC detergent molecules in random orientation, a control ATMD simulation of 200 DHPC molecules and ~ 75,000 water molecules was performed. The concentration of DHPC in this system is ~ 300 mM, intended to match the concentrations used in the NMR solution experiment. The lipid:protein ratio was ~200:1 therefore 200 DHPC monomers was chosen as a minimum system size.

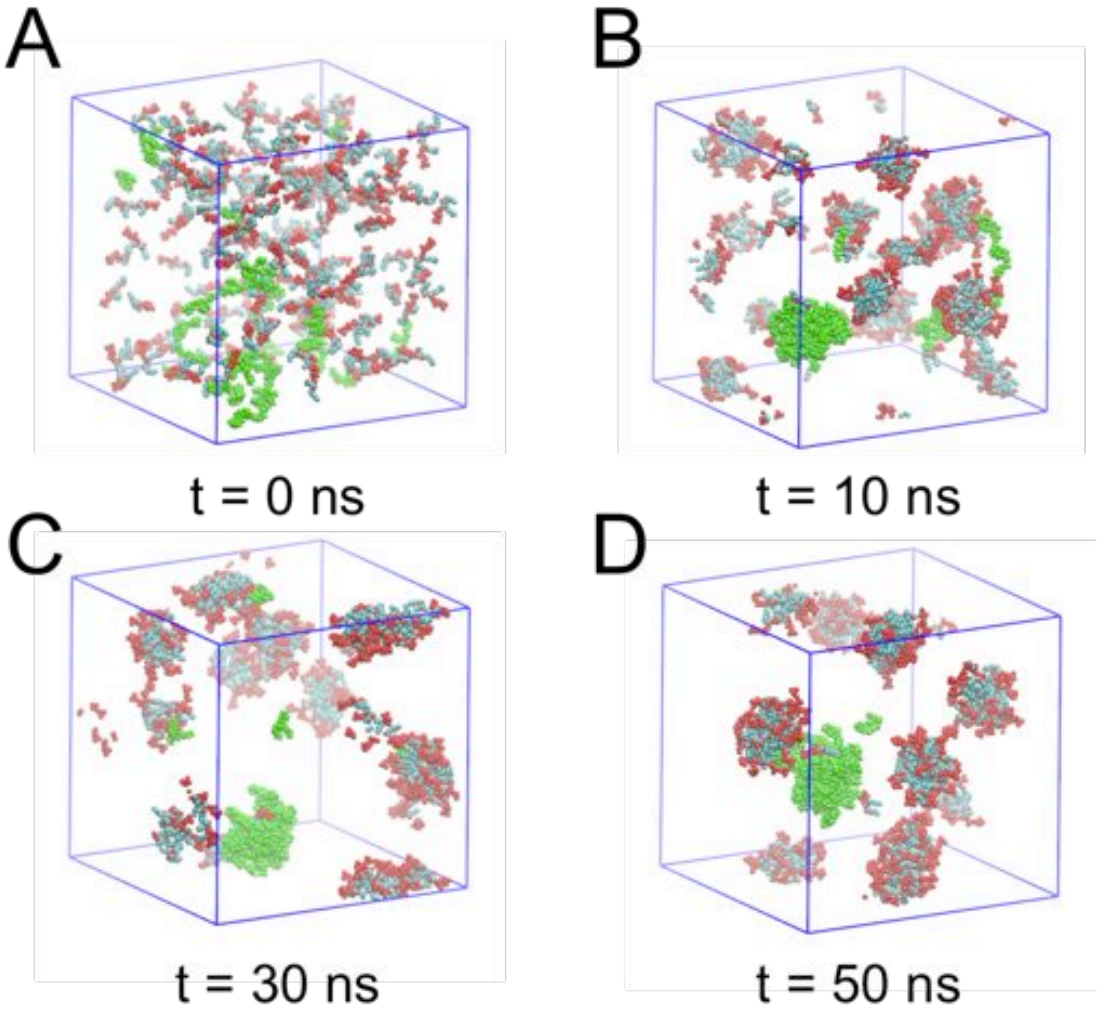


Figure 5:3. Snapshots showing self-assembly of DHPC monomers into micelles. A: The initial configuration of 200 randomly placed and orientated DHPC monomers. The system after B: 10ns, C: 30 ns and D: 50 ns. The phosphatidylcholine headgroups of DHPC are shown as red spheres, acyl tails are shown as cyan spheres. The detergent molecules coloured in green are those used in analysis of a single micelle (see text).

It is possible to follow the process of DHPC micellisation by monitoring the average cluster size along with the number of clusters (Figure 5:4 D & E). The number of clusters appears to plateau after ~ 50 ns of simulation time. The mean cluster size is ~ 28, in good agreement with experimental data (N_{agg} DHPC = 25-35). By the end of the 100 ns simulation there are 3 DHPC monomers present in solution, corresponding to a concentration of ~ 5 mM. This is slightly below the CMC of DHPC (14-15 mM) and may be a consequence of the limited system size or simulation time. Upon visual inspection of the final 10 ns of the trajectory, DHPC micelles were observed to interact with each other *via* headgroup interactions. However, minimal transfer of detergent monomers between micelles occurs on this timescale (three incidents of transfer during final 10 ns). The results of this self-assembly simulation demonstrate that modelling the initial fast process of DHPC micellisation (at experimentally relevant concentrations) is

feasible using ATMD (although the slower process of exchange between micelles would require longer timescales).

As the DHPC molecules aggregate, the number of DHPC hydrogen bonds with water is observed to decrease, whilst the number of water-water hydrogen bonds increases (Figure 5:4 A& B). This corresponds to release of water molecules from the ordered 'cage' of water molecules on the surfaces of DHPC molecules (the entropic effect discussed in the Introduction of this thesis).

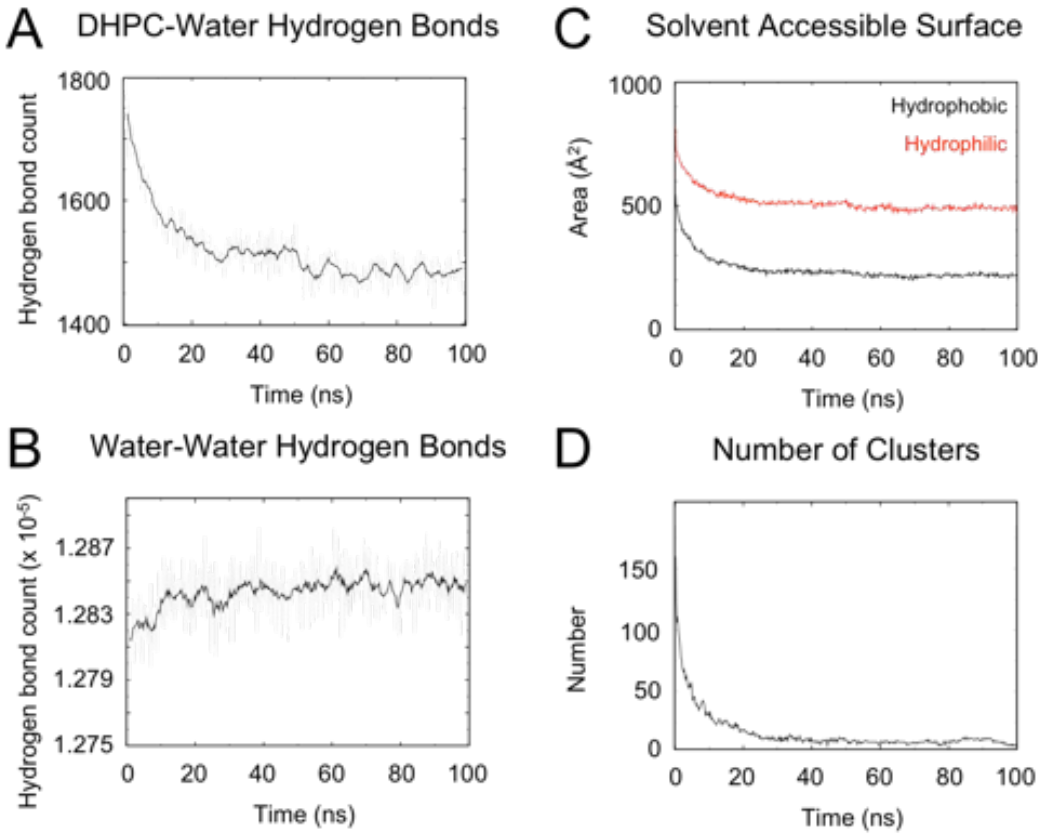


Figure 5:4.. Properties of the DHPC-water system during the micellisation process. The number of hydrogen bonds between A: DHPC and solvent and B: solvent molecules is shown as a function of time. C: The hydrophobic (black trace) and hydrophilic (red trace) SASA of the protein during simulation. D: Average (mean) size of clusters and E: Number of separate clusters versus simulation time.

Properties of an individual DHPC micelle that formed within 25 ns and remained intact for the duration of the simulation were calculated for comparison to experimental SAXS data. Two views of this micelle are shown in Figure 5:5. The prolate nature of the micelle ($c > a = b$) can be quantified by calculation of the semiaxis lengths. The semiaxis lengths (a, b and c) of micelles (mass M) may be calculated from the micelle moments of inertia ($I_1 > I_2 > I_3$) [5]. The polar length, corresponding to the thickness of the headgroup region, is found by calculating the semiaxis lengths of the acyl tails alone, and

subtracting these from the overall lengths. This, along with the micelle size and radius of gyration are shown in Table 5.

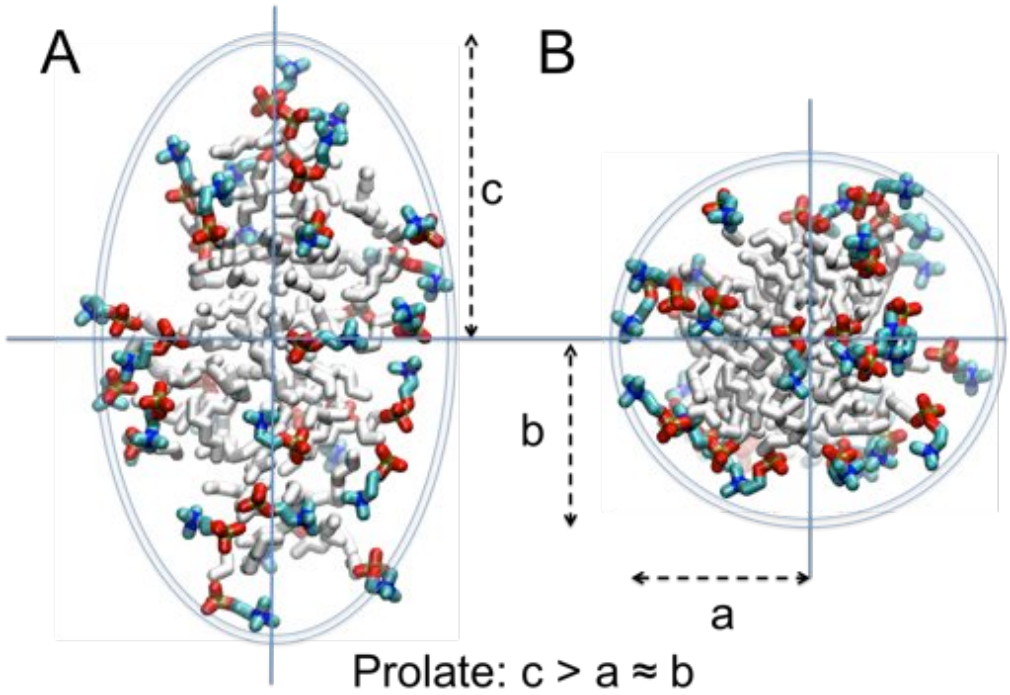


Figure 5:5. A: Side and B: top views of an individual DHPC micelle (N = 30). This micelle existed for ~ 75 ns of the 100 ns simulation. Blue lines indicate the axes of symmetry. Dashed black arrows indicate the semiaxis lengths.

	N_{agg}	R_g	a (Å)	c (Å)	pl (Å)	a/c
Experiment [3]	25-35	20	9.5-10.0	20.5-22.5	3.0-4.0	2.2
MD (self-assembly)	30	15.6 (± 0.6)	16.6 (± 1.6)	24.3 (± 3.2)	3.7 (± 0.4)	1.5 (± 0.7)

Table 5. Properties of DHPC micelles calculated from experimental data [3] and from MD simulation. N_{agg} corresponds to the calculated aggregation number in experiment or the size of the micelle analyzed in the MD experiment. The radius of gyration is calculated using the program g_gyrate in GROMACS. Semiaxis lengths (a and c) and the polar length (pl) are calculated indirectly from the instantaneous moments of inertia (see text). All values from MD are the averages over the final 10 ns of the trajectory, with twice the standard deviation shown in parentheses.

The prolate nature of the micelle (a/c ratio 1.5 ± 0.7) is in reasonable agreement with SAXS measurements (a/c ratio 2.2). There is good agreement between the thickness of the headgroup region calculated by SAXS (3.0-4.0) and using MD (3.7 ± 0.4). The radius of gyration (R_g) is lower than the experimental value, corresponding to a more compact species. The discrepancy between experimental R_g and axis lengths and the MD values may be related to hydrated regions in the headgroup region of the micelle contributing to the SAXS data [6]. However, the overall agreement between the experimental data and the MD data appears reasonable. The analysis of pure DHPC micelle formation

demonstrates that ATMD may in principle be used to monitor formation of protein-DHPC complexes.

3.2. β-Dodecylmaltoside (DDM) Micelles

The lower cmc of DDM compared to DHPC poses a challenge in simulations using this detergent. The concentration of DDM used in experiments is generally one to two orders of magnitude lower than that used for DHPC [7]. As discussed in the previous section, the concentration and system size used for DHPC was approaching the limit currently achievable for study by ATMD. To simulate micelle formation at experimentally relevant DDM concentrations would require a massive increase in system size due to increased number of water particles. Indeed, the only published MD study of DDM micelles of aggregate size consistent with literature values of N_{agg} was based on pre-formed micelles of DDM [6].

To assess the practicality of using ATMD to monitor DDM micelle formation, a similar system to the DHPC micelle self-assembly simulation was generated. This comprised a box of water containing 200 randomly positioned DDM molecules at a concentration of ~ 200 mM. However, very little rearrangement into well-defined aggregates was observed during 50 ns of simulation, (data not shown) presumably because the artificially high local concentration of DDM corresponds to the extended aggregate phase of DDM (Phase III in Figure 5:2).

Formation of DDM micelles is therefore not readily obtainable using conventional ATMD simulations. To determine the timescales necessary to simulate DDM formation at experimentally relevant concentrations, a CGMD system was generated of 200 CG-DDM monomers in a large box, at a concentration of ~ 2 mM. This required 11 million CG-water particles. A positionally-restrained BM2 was also present in the centre of the box. The system was simulated for 1.2 μs. Snapshots of the system over time are shown in Figure 5:7. Even after 1.2 μs simulation there are still a large number of detergent monomers present in solution. The much lower cmc of DDM means that in the system studied the number of free monomers in solution would be expected to be only one or two. The formation of micelles over time is shown in Figure 5:6. Unlike the DHPC simulation, where a clear plateau in the number of separate micelles levelled off within 50 ns of ATMD, there are no signs of the number of clusters levelling out on this timescale.

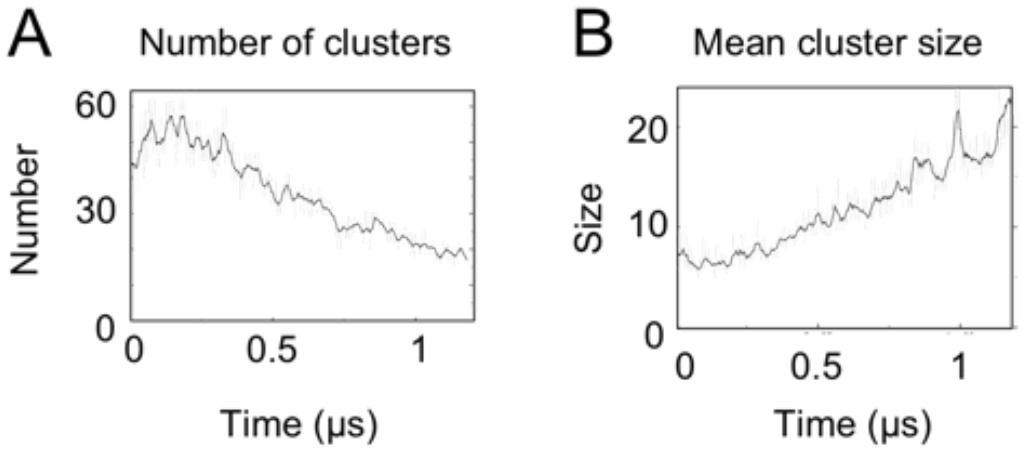


Figure 5:6. Cluster formation in CGMD simulation of DDM micelle formation. A; The number of separate clusters in the system over time and B: Mean cluster size during the simulation. There are no signs of convergence on this timescale.

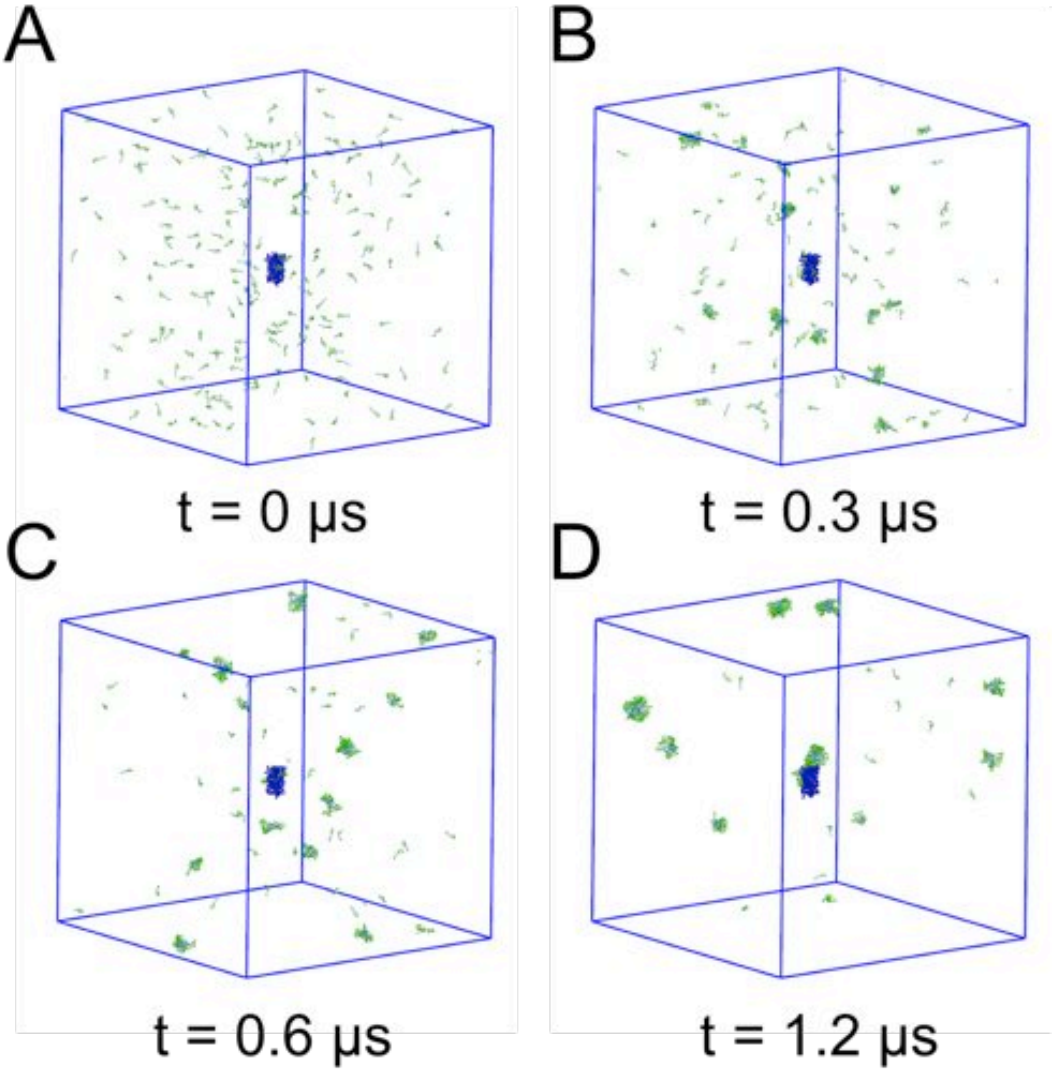


Figure 5:7. CGMD system of DDM and BM2 at experimentally-relevant concentrations. A: The starting configuration. Snapshots of the system at B: 300 ns, C: 600 ns and D: 1200 ns. The protein is shown as blue spheres. Detergent monomers are coloured green for the maltose headgroup and yellow for the acyl tails. Waters are omitted for clarity.

It therefore appears that timescales of several microseconds may be necessary to observe the process of DDM micellisation using CGMD. The large system size meant that even with the speed-up relative to conventional ATMD, this simulation took ~1 month to run on 32 processors. The generation of a protein-detergent DDM complex *via* condensation of the detergent onto the protein is therefore not currently feasible.

The monomer-aggregate micelle equilibrium lies much further to the right for DDM compared to DHPC (Table 4) such that the concentration of DDM monomers in bulk solution is approximately two orders of magnitude lower than for DHPC. It seems reasonable, therefore, to simulate the interaction of a DDM *aggregate* (*i.e.* a preformed micelle) with the membrane protein. Two sizes of micelles were built, N=132 and N=150. These are within the experimentally determined aggregation range and allow comparison to both experimental results and the recent MD study of DDM micelles.

The protocol used in the DDM simulations may be summarized in four steps:

1) Spherical CG-DDM aggregate built such that the end carbon of each of the acyl tails (furthest from headgroup) is at the centre of sphere.
2) CG-DDM micelle simulated for 1 μs to allow full equilibration
3) System following CGMD simulation converted to ATMD
4) ATMD simulations performed for comparison to experimental and MD data

The initial spherical aggregate for the N=150 system is shown in Figure 5:8. The program Packmol [8] was chosen to build the initial aggregate since it distributes the headgroups approximately equidistantly around the surface of the micelle. This gives an unbiased initial set up. It can be observed in Figure 5:8 that upon CGMD simulation the micelle 'flattens out' to become oblate in geometry. This is consistent with experimental predictions. There is also a tendency for the maltose headgroups to cluster together slightly on the surface, leaving some sections of the acyl tails exposed to solvent.

The final CG-DDM micelle was then converted to an atomistic representation (using the CG2AT approach described in Chapter 2), using both the OPLS-AA and GROMOS96 force fields. This was in order to test the stability of the CGMD-generated micelles. Upon ATMD simulation of the DDM micelle the dimensions are seen to decrease and the micelle ellipticity decreases (ratio of a/c axes tends towards 1) slightly, whilst maintaining an oblate shape (Figure 5:8). Some changes upon CG2AT conversion are expected due to the system relaxing. A further cause for changes to the detergent

packing is that waters are able to penetrate into the headgroup region of the AT micelle, whereas the larger water particles used in CGMD limit the degree of penetration. The clustering of maltose headgroups, leading to a slightly "rough" surface is maintained in ATMD simulations. The DDM micelle systems tested are shown in Table 6, with calculated parameters compared to those from MD and experiment.

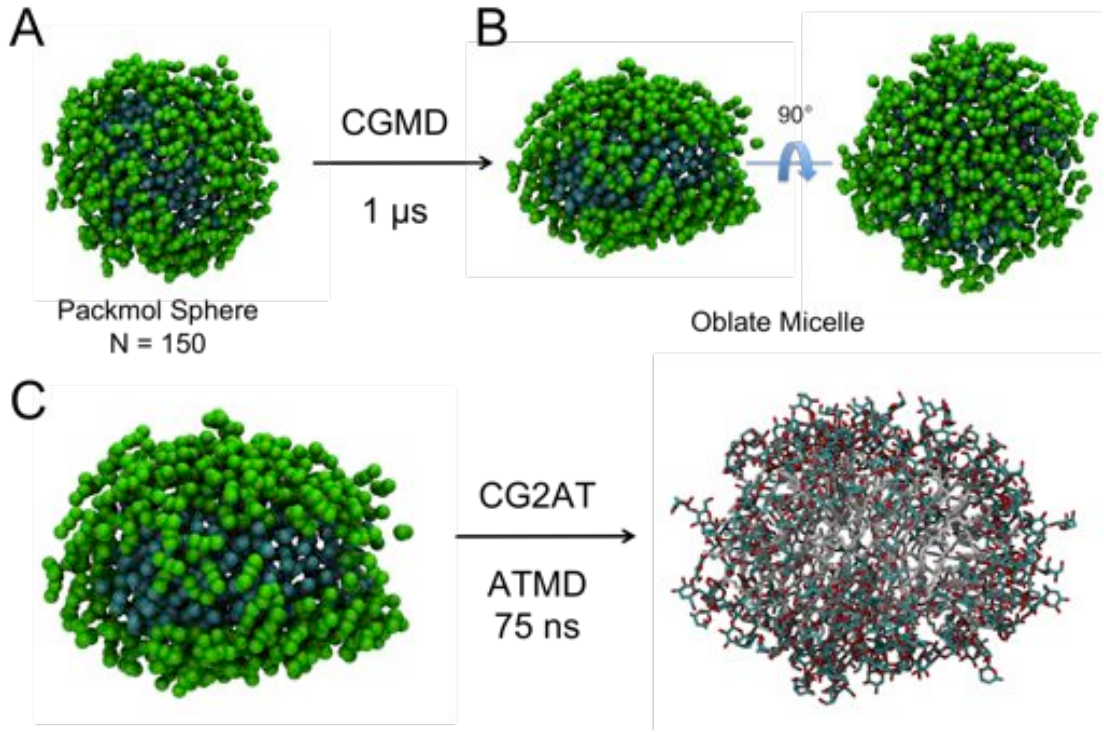


Figure 5:8. A: CGMD simulation of pre-formed DDM micelle. A: Initial spherical CG-DDM aggregate generated using the program Packmol. B: Side and top views of the DDM micelle following 1 μs of CGMD simulation. Maltose headgroups are shown as green spheres while the acyl tails are shown as cyan spheres. C: Conversion to atomistic (AT) representation and DDM micelle following a relatively long 75 ns ATMD simulation with the OPLS-AA force field. The micelle becomes more compact, while the oblate nature is retained Acyl tails are shown in grey, headgroups are coloured according to atom type (red = oxygen, cyan = carbon). Hydrogen atoms and water molecules are omitted for clarity.

As was observed for analysis of the DHPC micelle there is some difference between the calculated dimensions of the simulated DDM micelles (including those from reference [6]) and the experimental scattering data. The main disparity between experiment and simulation appears to be the degree of ellipticity (measured by the a/c ratio).

	N_{agg}	R_g (Å)	a (Å)	c (Å)	pl (Å)	a/c
Experiment [3]	135-145	33	28.0-29.5	13.8-14.3	6.0-6.3	0.48
MD [6]	132	25.2-26.4	37.2-38.8	26.5-30.1	7.2-7.7	0.71
OPLS-AA	132	24 (± 0.2)	33.1 (± 1.0)	28.4 (± 0.8)	7.1 (± 0.2)	0.86
GROMOS	150	24.9 (± 0.2)	34.0 (±1.0)	28.0 (±0.6)	5.1 (± 0.1)	0.82
OPLS-AA	150	24.4 (± 0.3)	34.2 (± 0.8)	27.2 (± 0.6)	7.1 (± 0.1)	0.79

Table 6. Properties of DDM micelles from experiment and molecular dynamics simulation. N_{agg} refers to the calculated aggregation number in the case of the experimental data, or the numbers of lipids chosen for the study in the case of MD work. R_g is the calculated radius of gyration of the detergent micelle. a and c correspond to the half-lengths of the secondary and principle axes of symmetry. pl corresponds to the length of the polar core (calculated as described previously in text). The ratio a/c gives and indication of the degree of ellipticity (tending towards one for a spherical aggregate).

4. Protein-Detergent Complexes

The simulations of the detergent-only systems provide a methodology for each detergent (atomistic self-assembly for DHPC, coarse-grained pre-formed aggregate for DDM) that may to be extended to simulate complex formation with BM2.

4.1. BM2-DHPC Complex Formation

A similar system was set up to the one described for the DHPC-only simulation. The NMR structure was restrained in the centre of the box, and 200 DHPC lipids were randomly orientated and positioned around it. To increase the rate of micelle formation the system set up was biased such that the DHPC monomers were contained within a smaller cube and the box dimensions were then extended prior to addition of solvent (Figure 5:9). This was the starting point for a 50 ns ATMD simulation during which the condensation of DHPC detergent molecules onto the protein surface was monitored. Similarly to the DHPC-only simulation, small DHPC clusters are observed to form within a few nanoseconds.

Following the initial 50 ns of ATMD simulation, 115 DHPC residues are found to be directly or indirectly in contact with the protein in a 'loose' aggregate. The majority of these are in contact with the hydrophobic region of the protein. A smaller micelle of N~25 is attached at the N-terminal extracellular region. Of these 115 detergent molecules the majority of them (92) bound to the protein-detergent complex as part of small detergent clusters. The remaining detergent molecules bound as monomers,

dimers or trimers. Figure 5:10 shows the protein surface following 50 ns simulation. It is observed that the majority of the hydrophobic surface residues are in contact with the DHPC tails, and water is excluded from this region. The charged residues near the protein termini remain largely accessible to water. The hydrophobic solvent accessible surface area (SASA) is reduced by ~50 % within the 50 ns simulation. The hydrophilic SASA is also reduced upon formation of protein-detergent contacts, but to a lesser extent (< 30 %).

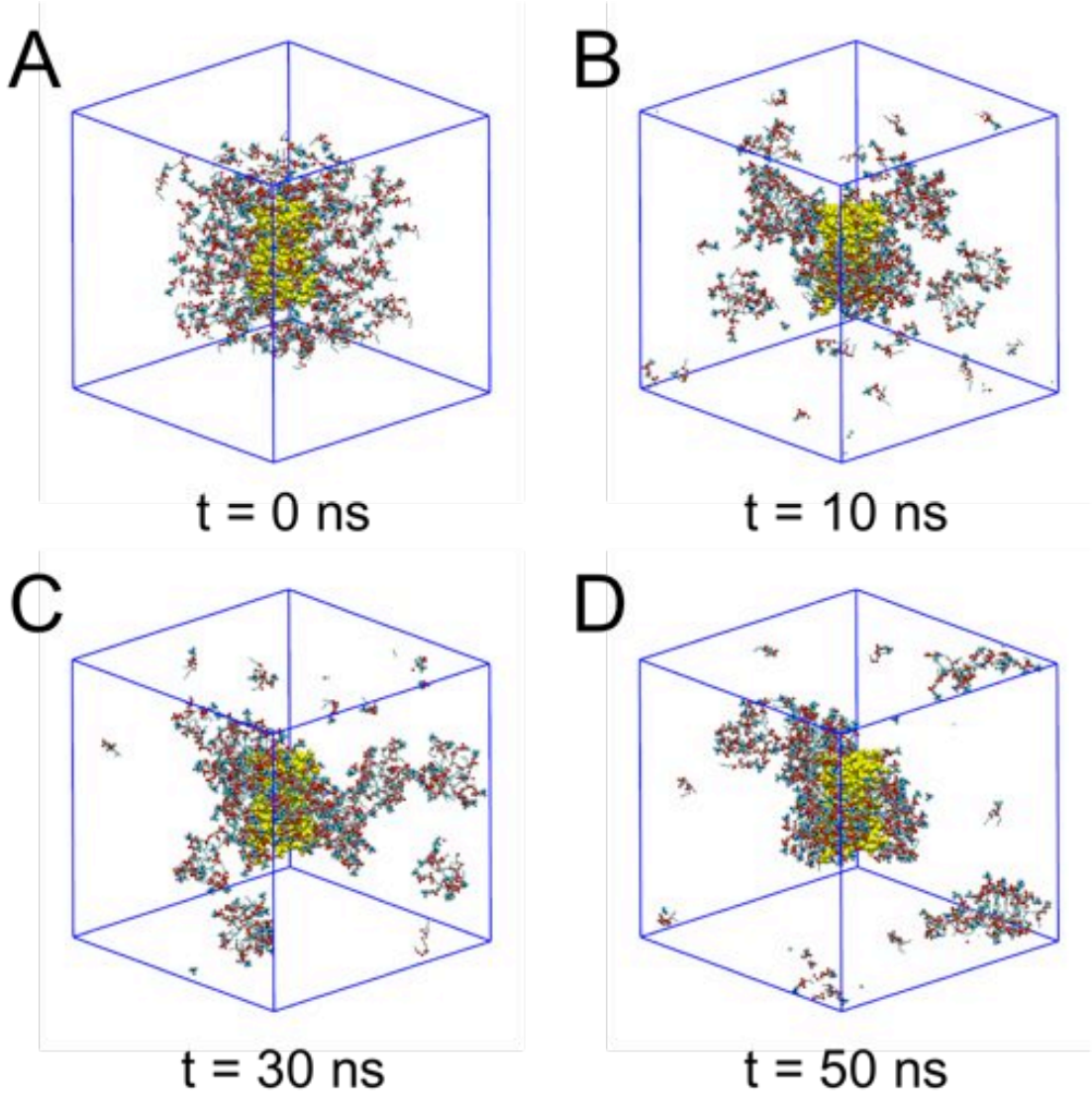


Figure 5:9. Self-assembly of BM2-DHPC complex. A: Starting point of simulation. Snapshots of the system at B: 10 ns, C: 30 ns and D: 50 ns. The phopshatidylcholine headgroups are shown as spheres (oxygen:red, carbon:cyan, nitrogen:blue, phosphorus:tan) and the carbon tails are shown as a cyan trace. The protein is shown as yellow spheres. Water is omitted for clarity.

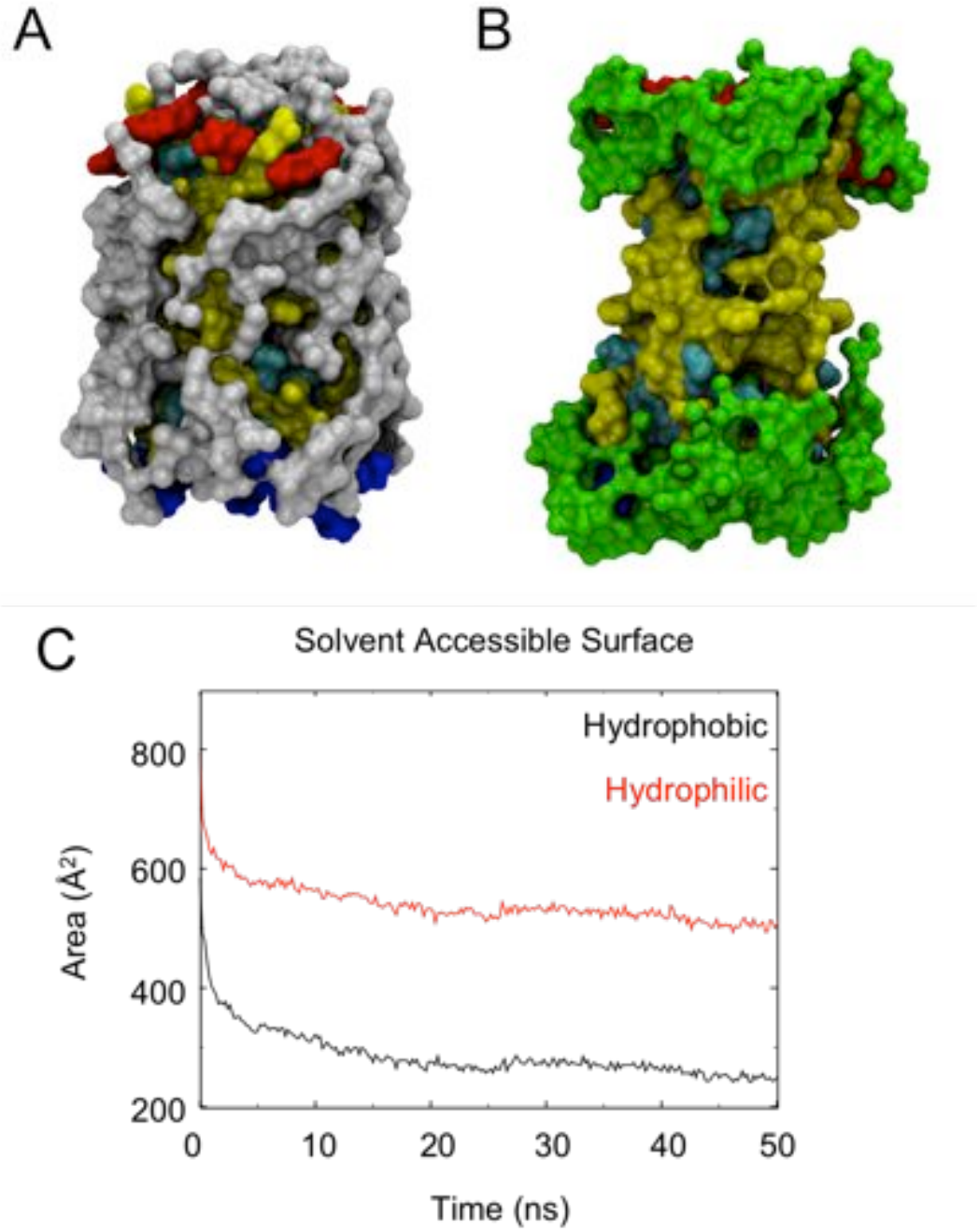


Figure 5:10. Solvent accessible surfaces of BM2. Top: protein surface after 50 ns coloured according to hydrophobicity of residues (red = negatively charged, blue = positively charged, cyan = polar, yellow = hydrophobic residues). A: DHPC atoms within 5 Å of the protein surface are shown as a grey surface. B: Water oxygen atoms within 5 Å of the protein surface are displayed as a green surface. C: Reduction in solvent accessible surface area of BM2 during initial 50 ns simulation. Surface areas of residues are classed as hydrophobic (black trace) and hydrophilic (red trace) based on partial charges (>0.2 corresponds to hydrophilic residue).

Once the initial 'loose' PDC complex formed (after 50 ns) the main PDC was removed and simulated further in a smaller box (whilst maintaining the same overall concentration of DHPC). This reduction in system size allowed a less computationally expensive simulation of the equilibration process. During 100 ns simulation the smaller DHPC micelle attached to the N-terminus was found to dissociate and the geometry of

the remaining protein-detergent complex appeared to stabilise within a 100 ns timescale as evidenced by the radius of gyration (Figure 5:11)

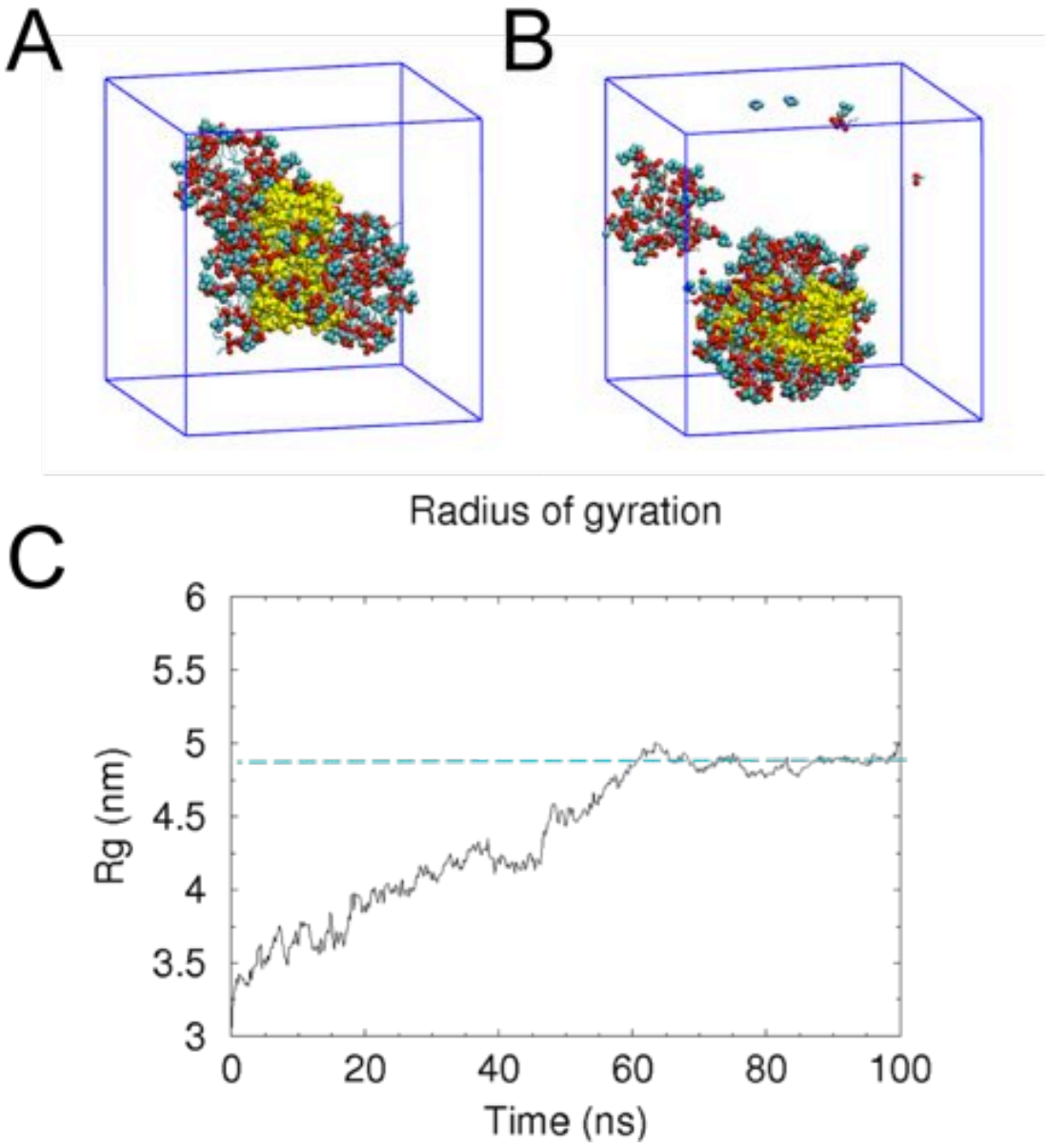


Figure 5:11. Snapshots of the aggregate following A: 50 ns self-assembly simulation and after B: 100 ns equilibration simulation. C: Calculated radius of gyration for those molecules in the final protein-detergent complex during 100 ns equilibration simulation. The dotted line in blue corresponds to the mean radius of gyration of the protein-detergent complex over the final 10 ns of the simulation.

4.2. BM2-DDM Complex Formation

The calculated values of the aggregation number of DDM micelles range from 98 up to 135-145 [9]. Previous simulations and calculations have shown that upon binding with a protein, the number of detergent molecules can increase above that of the aggregation number of pure detergent as the hydrophobic core of the protein provides extra surface area for the acyl tails to interact with [10].

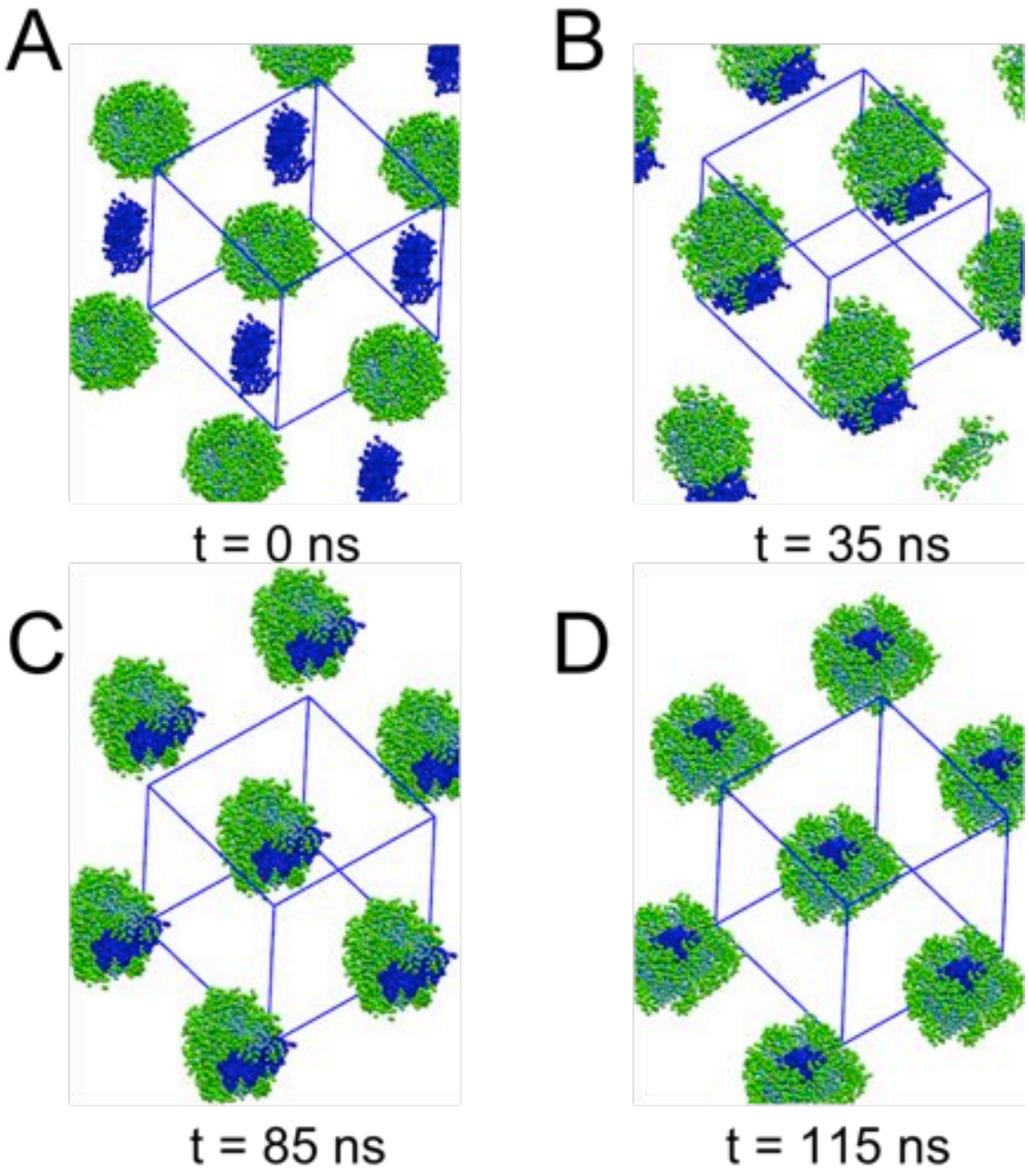


Figure 5:12. Snapshots of the BM2-DDM complex formation at A: 0 ns, B: 35 ns, C: 85 ns and D: 115 ns. The initial configuration has the BM2 and DDM micelle as far apart as possible. The protein rapidly contacts the DDM micelle and becomes incorporated into the centre of the micelle after 100 ns.

Based on this evidence, a micelle of DDM at the higher end of the predicted aggregation number, namely, 150 DDM molecules, was used as the starting point for the generation of a BM2-DDM complex. The system for formation of a BM2-DDM complex was generated by placing the BM2 protein and DDM micelle at sufficient distance apart in the box to ensure that there was no initial interaction between them (*i.e*. greater than the interaction cut-offs). Five simulations of 1 μs based on this starting configuration were performed. The stages of complex formation are displayed in Figure 5:12. Briefly, the DDM micelle and protein interact within tens of nanoseconds in four of the simulations and in one simulation ~200 ns (the average time to contact will be proportional to the box size chosen); the DDM maltose headgroups at the point of contact move away to interact with solvent, whilst hydrophobic interactions between the detergent carbon tails and the protein hydrophobic core are formed; the protein remains in this surface orientation for varying amounts of time before becoming incorporated into the centre of the DDM micelle. In one of the five simulations the protein remains on a surface orientation whereas in the other four the complex rearranges at 400 ns, 500 ns, 800 ns and 900 ns.

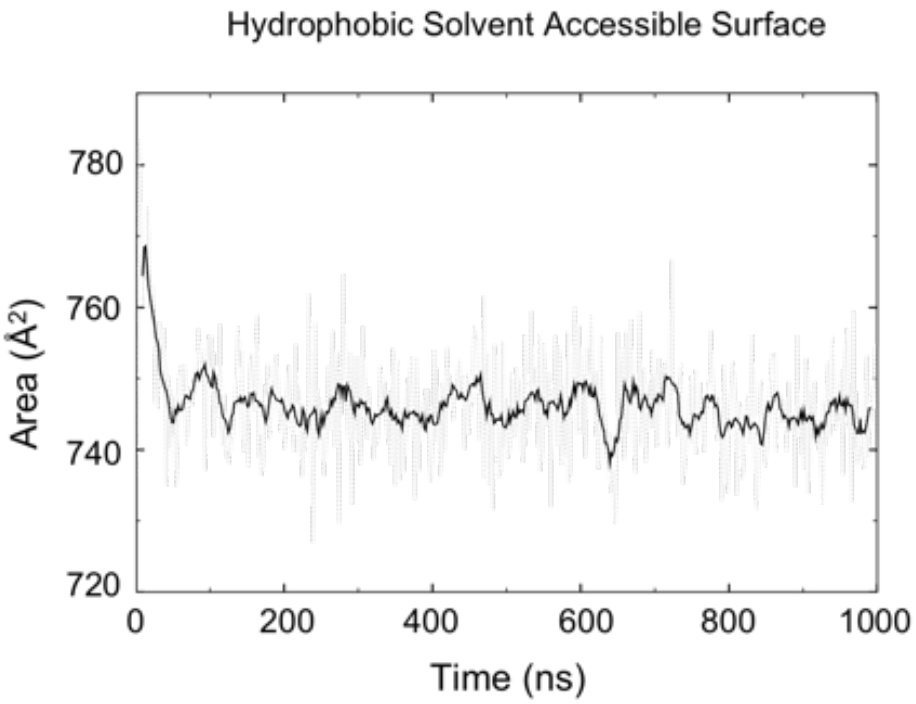


Figure 5:13. Hydrophobic solvent accessible surface area of the protein and detergent molecules during initial association of protein and DDM micelle (~20 ns) and during rearrangement from surface to buried protein orientation (~500ns). A running average over 10 data set points is shown as a black trace, with the full data shown as a dotted grey trace.

The main change to components of the potential energy of the system is during the initial interaction between the protein and DDM micelle. This allows the hydrophobic

surface area of the protein to be reduced significantly. The slow rearrangement from a surface to central orientation of the BM2 leads to less change in potential energy. The overall system hydrophobic SASA does not change significantly during this rearrangement because some of the DDM tails become exposed to the solvent during this process.

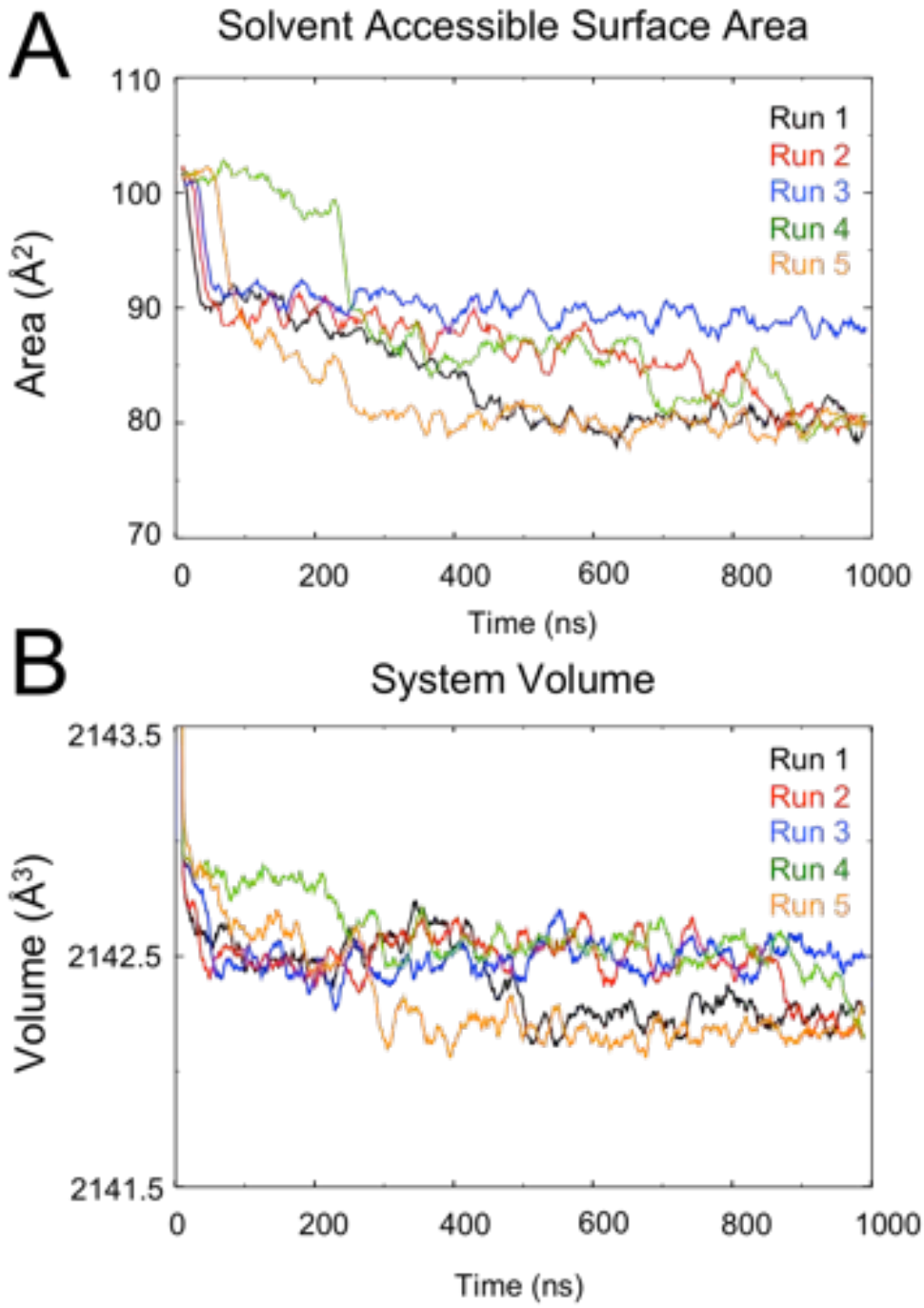


Figure 5:14. Properties of the system during BM2-DDM initial association and subsequent rearrangement. A: protein solvent accessible surface area (SASA) over time. B: system volume over time. The largest decrease in both SASA and volume is seen upon initial association with the DDM micelle. Data from each simulation is coloured according to plot legend. The data is shown as a running average over 100 data points.

The incorporation of the protein into the centre of the DDM micelle may be followed by monitoring the protein SASA and the overall volume of the system over time (Figure 5:14). Upon transfer to the micelle centre the SASA of the protein decreases. In the one simulation where the protein remains in a surface orientation there is little change to the SASA following the initial decrease. Incorporation of the protein into the centre of the micelle affects the system volume. The volume is reduced upon initial association due to the excluded volume of the combined BM2-DDM complex being lower than the sum of the individual components. This is due to a decrease in the number of waters interacting with the surfaces of solutes. A further reduction in box volume is observed upon rearrangement to central position of the BM2 as the overall surface area of the solutes is again reduced.

The distribution of DDM headgroups in each of the four simulations where BM2 becomes fully incorporated into the complex is shown in Figure 5:15. The average positions of the first glucose ring (furthest from the acyl tail) over the final 10 ns of each simulation are shown. The complex after 1 μs of simulation is seen to vary between the four repeats. However, the complex had < 200 ns and < 100 ns in the 2nd and 4th simulations (respectively) to equilibrate following the rearrangement. Comparison of the two most equilibrated systems (1st and 5th simulations) shows some variation in the position of the protein within the micelle. The PDC generated in the first simulation was chosen for further study.

The BM2-DDM complex was converted to an atomistic system (using the CG2AT method) and simulated for 20 ns at 323 K with positional restraints on the protein to enable the packing of the detergents to stabilise around the protein. Similarly to the DDM-only micelle control, the protein-micelle complex became more compact (measured by initial rapid decrease in R_g of the complex from 27.5 to 27.0) and the geometry stabilized (small fluctuations in R_g in the region of ± 0.2 Å) during the 20 ns simulation.

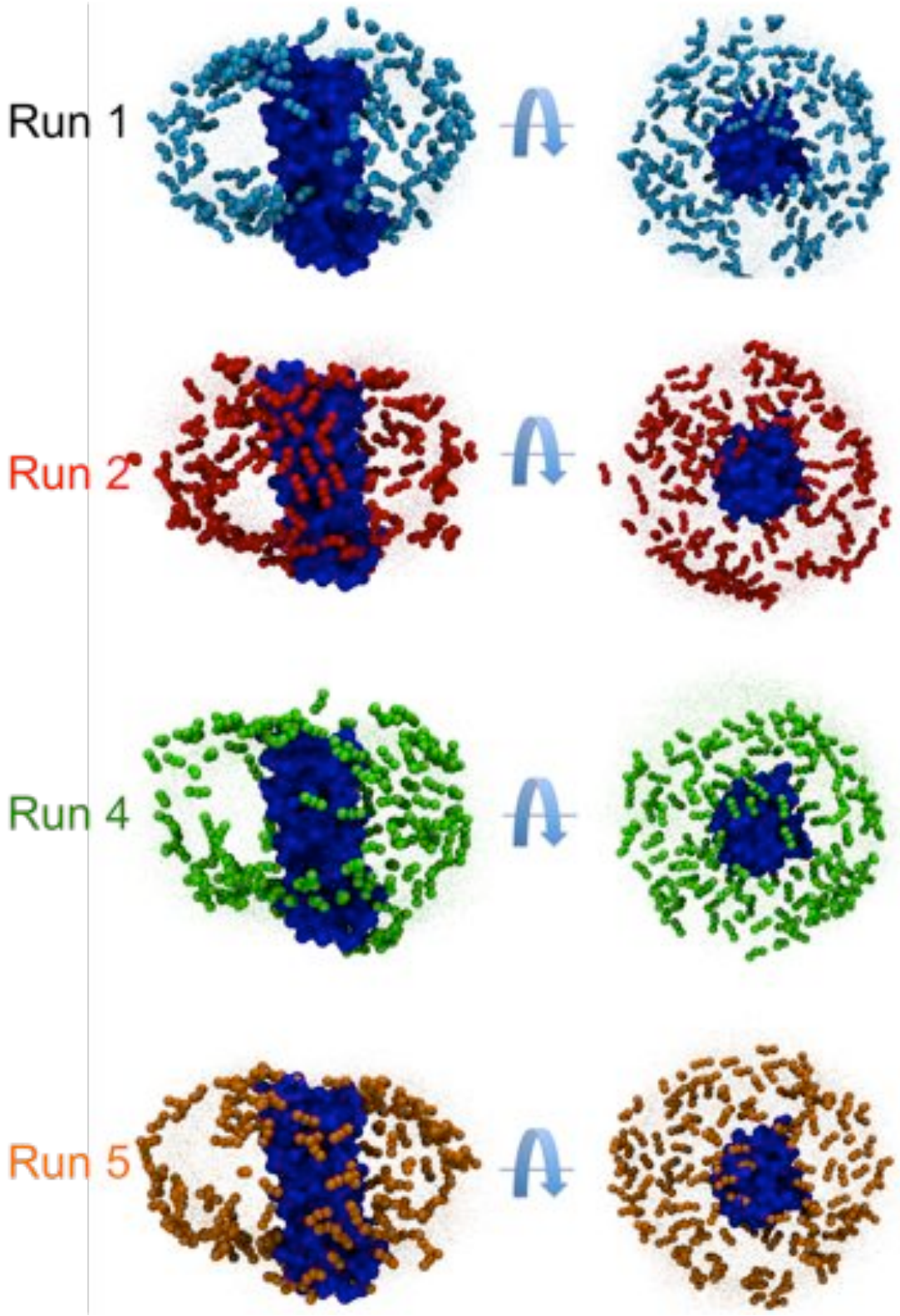


Figure 5:15. Final BM2-DDM complexes following 1μs CGMD simulation. The protein is shown as a blue surface. The most populated positions of the first ring of the maltose headgroups over the final 100 ns of CGMD simulation are shown spheres. The space sampled by the headgroups are shown by small dots. There is some evidence of headgroup clustering. Spatial distributions were calculated in the frame of reference of the protein.

5. Discussion

A multiscale simulation approach was found to be necessary to fully describe formation of the two types of detergent micelles: ATMD for zwitterionic DHPC detergent micelles and CGMD followed by ATMD for nonionic DDM detergent micelles. The results are discussed below.

5.1. Generation of Protein-Detergent Complexes

The analysis performed of DHPC micelle self-assembly simulations demonstrates that the timescales currently attainable in ATMD are sufficient to monitor micelle formation. To observe the slower process of detergent exchange between discrete micelles, longer timescales would be required. The use of CGMD allows much longer timescales to be simulated, however the lack of an upper limit to aggregate size using CGMD has been documented elsewhere [11]. Exploratory DHPC micelle self-assembly simulations demonstrated that CGMD could not replicate key features of DHPC micelle solutions. The simplified description of electrostatic interactions in the CG model prevented an upper limit to aggregation size, such that no monomers were present in solution. This is particularly significant for DHPC where the concentration of monomers in solution is ~15 mM and may be expected to have some bearing on the mechanism of micelle formation.

The two simulations performed to generate the protein-DHPC complex (self-assembly then equilibration) suggest a mechanism for formation of protein-DHPC detergent complexes. During the self-assembly simulation most of the binding events occurred *via* small clusters of DHPC. The overall mechanism of protein-DHPC complex formation may therefore involve binding of one or more DHPC micelles (of size N_{agg}) followed by rearrangement to the equilibrium geometry.

The self-assembly of lower concentration, nonionic micelle systems (corresponding to higher number of water particles) is not yet feasible using either ATMD or CGMD. Despite being unable to use a self-assembly system to determine the micelle size of the nonionic detergent DDM, the pre-formed CG micelle (the size of which was based on aggregation values from the literature) was stable upon conversion to ATMD, and the main features of the micelle as determined from experiment and ATMD simulation (oblate geometry and maltose headgroup clustering) were replicated in the CG model. Therefore the use of a CGMD pre-formed aggregate appears to be an acceptable method

in generating systems for ATMD simulation. Based on the time taken for the BM2-DDM initial aggregate to rearrange to incorporate BM2 into the centre of the micelle, an equivalent ATMD simulation of BM2-DDM complex formation is expected to require up to 4 μs simulation time.

5.2. Comparison to Experimental Data

The analysis of pure DHPC micelle formation *via* ATMD simulation showed reasonable agreement with experimental data from SAXS scattering. A key feature of the scattering data, the thickness of the 'polar length' (the part of the micelle not including the hydrophobic core) was matched well by the simulated micelle. Comparison of the pre-formed DDM micelle formed via CGMD simulation then conversion to atomistic resolution also showed good agreement with the experimental polar length. This observation suggests that the orientations of the headgroups of the DHPC and DDM molecules within the micelles are physical.

The a/c ratio of the micelles generated by MD simulation is consistently higher than that calculated from the SAXS data, regardless of force field used. This may be a consequence of the microscopic nature of the simulations compared to the macroscopic measurements obtained using SAXS. Another possible reason for this discrepancy, highlighted earlier, is that of ordered water molecules in the headgroup regions of the detergent micelles contributing to the scattering profiles of the micelles in SAXS measurements. Observation of water distribution on the surface of the DDM micelle showed an extended hydrogen bonded network between the maltose headgroups and surface waters. The disagreement between simulated micelle parameters compared to experimental data is conserved; there is good agreement between all simulation studies using five different parameter sets (three in reference [6] and two used in this thesis). Furthermore, comparison of the scattering profiles of the OPLS-AA system with 150 DDM molecules to recent SAXS data showed a good fit (Rouse and Durand, D., unpublished data).

5.3. Architecture of DHPC and DDM Protein-Detergent Complexes

As may be expected from the differing properties of DHPC and DDM detergents, the geometries of the BM2-DHPC and BM2-DDM complexes are correspondingly dissimilar.

The small-micelle forming nature of DHPC detergent leads to only 70 detergents (from a possible 200) packing on the surface of the protein, whereas the large-micelle forming DDM detergent incorporates the protein into the centre of a large oblate complex.

The DHPC headgroups appear to be distributed approximately evenly upon the surface of the complex whereas clear clustering of the maltose headgroups was apparent in both CG and AT simulations of the BM2-DDM complex. This clustering maximises the hydrogen bonding capacity of the maltose headgroups on the surface of the complex. This behaviour is consistent with that reported in a concurrent MD study of DDM micelles in solution, despite the pre-formed micelles being generated in different manners, suggesting that this is an important feature of DDM micelles rather than an artefact of the methodology used.

The results described in this chapter suggest that protein-detergent complexes (PDCs) appear to retain some of the features of the detergent-only complexes. This is investigated further in the following Chapter.

6. Further Work

6.1. Conformational Transitions in DDM Micelles

The simulations presented within this Chapter provide a methodology for generating protein-detergent complexes using the nonionic detergent DDM. This has been successfully applied to other, larger proteins *e.g*. LamB and BtuCD with DDM and β-octyl glucoside (Stansfeld and Rouse, unpublished data).

In the five simulations of BM2 becoming incorporated into a DDM micelle there was some evidence of differences in the packing of DDM detergent molecules around the protein. In [12], various techniques were used to study the small α-helical protein TM0026 in micelles of differing shape and sizes. They found that when the protein was incorporated into DDM micelles, NMR line broadening was observed, which they attributed to the protein adopting different conformations in the large DDM micelle rather than slower rotation of the overall complex due to the large size. Now that a protocol has been developed for forming protein-DDM complexes this could be used as the basis of an MD study on potential transitions between different conformations of a

small protein such as TM0026 or the extensively studied glycophorin A in DDM (or other nonionic detergent micelles such as β-octyl glucoside).

7. Summary

The results of this Chapter have shown that a multiscale simulation approach to detergent systems is necessary for a complete description of protein-detergent complex formation. The timescales required to observe the small membrane protein BM2 being engulfed by a DDM detergent micelle are in the order of microseconds, which is currently not feasible for an ATMD simulation.

The two types of detergent studied here formed micelles of differing properties that were qualitatively consistent with experimental data (and quantitatively in some cases). This is despite the fact that experimental measurements are made over a distribution of many micelles, rather than the single micelles analysed here. The comparison between experimental measurements and the micelles generated *via* MD simulations described within this Chapter allow reasonable confidence to be placed in the protein-detergent complexes that have been generated.

In the following Chapter the behaviour of the protein within these PDCs, along with differences in the detergent structural and dynamic properties, are investigated. The PDCs are also compared to the simulations within a DPPC bilayer previously described in Chapter 4.

8. References

[1] D. Linke, "Detergents: an overview.," *Methods in enzymology*, vol. 463, pp. 603-17, Jan. 2009.

[2] G. G. Warr, C. J. Drummond, F. Grieser, B. W. Ninham, and D. F. Evans, "Aqueous solution properties of nonionic n-dodecyl .beta.-D-maltoside micelles," *The Journal of Physical Chemistry*, vol. 90, no. 19, pp. 4581-4586, Sep. 1986.

[3] J. Lipfert, L. Columbus, V. B. Chu, S. A. Lesley, and S. Doniach, "Size and shape of detergent micelles determined by small-angle X-ray scattering.," *The journal of physical chemistry. B*, vol. 111, no. 43, pp. 12427-38, Nov. 2007.

[4] P. J. Stansfeld and M. S. P. Sansom, "From Coarse Grained to Atomistic: A Serial Multiscale Approach to Membrane Protein Simulations," *Journal of Chemical Theory and Computation*, vol. 7, no. 4, pp. 1157-1166, Apr. 2011.

[5] S. Abel, F. Sterpone, S. Bandyopadhyay, and M. Marchi, "Molecular Modeling and Simulations of AOT–Water Reverse Micelles in Isooctane: Structural and Dynamic Properties," *The Journal of Physical Chemistry B*, vol. 108, no. 50, pp. 19458-19466, Dec. 2004.

[6] S. Abel, F.-Y. Dupradeau, E. P. Raman, A. D. MacKerell, and M. Marchi, "Molecular simulations of dodecyl-β-maltoside micelles in water: influence of the headgroup conformation and force field parameters.," *The journal of physical chemistry. B*, vol. 115, no. 3, pp. 487-99, Jan. 2011.

[7] Z. E. R. Newby et al., "A general protocol for the crystallization of membrane proteins for X-ray structural investigation.," *Nature protocols*, vol. 4, no. 5, pp. 619-37, Jan. 2009.

[8] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez, "PACKMOL: a package for building initial configurations for molecular dynamics simulations.," *Journal of computational chemistry*, vol. 30, no. 13, pp. 2157-64, Oct. 2009.

[9] M. le Maire, P. Champeil, and J. V. Moller, "Interaction of membrane proteins and lipids with solubilizing detergents.," *Biochimica et biophysica acta*, vol. 1508, no. 1-2, pp. 86-111, Nov. 2000.

[10] O. Ivashyna, A. J. García-Sáez, J. Ries, E. T. Christenson, P. Schwille, and P. H. Schlesinger, "Detergent-activated BAX protein is a monomer.," *The Journal of biological chemistry*, vol. 284, no. 36, pp. 23935-46, Sep. 2009.

[11] S. A. Sanders and A. Z. Panagiotopoulos, "Micellization behavior of coarse grained surfactant models.," *The Journal of chemical physics*, vol. 132, no. 11, p. 114902, Mar. 2010.

[12] L. Columbus et al., "Mixing and matching detergents for membrane protein NMR structure determination.," *Journal of the American Chemical Society*, vol. 131, no. 21, pp. 7320-6, Jun. 2009.

Chapter 6

Protein-Detergent Complexes: Protein Stability and Conformation, Protein-Detergent Interactions, and Detergent Dynamics

1.1. The Solution NMR Structure in DHPC Micelles

The simulations of the solution NMR structure of the BM2 TM domain in a DPPC bilayer indicated there were two main destabilising features: *i)* the large pore dimensions and *ii)* the lipid-facing orientation of His27 sidechains. Simulations were performed of the solution NMR structure in a dihexanoylphosphatidylcholine (DHPC) micelle to assess whether this was a consequence of the protein adopting a different conformation in the micelle (the environment the structure was solved in) compared to that of a bilayer. Somewhat surprisingly, exploratory simulations found that the solution NMR structure was less stable in a DHPC micelle than it was in a bilayer. Two temperatures (300 K and 323 K) and three charge states of the His27 residues (zero, two and four protonated) were tested, with examples of the final structures shown in Figure 6:1. A large deformation in one helix was observed at 300 K, leading to a large water-filled region and a loss of protein interhelix contacts. Protonation of two or all of the His27 residues led to large distortions about these residues, although the rest of the structure retained more secondary structure compared to the neutral His27 simulations.

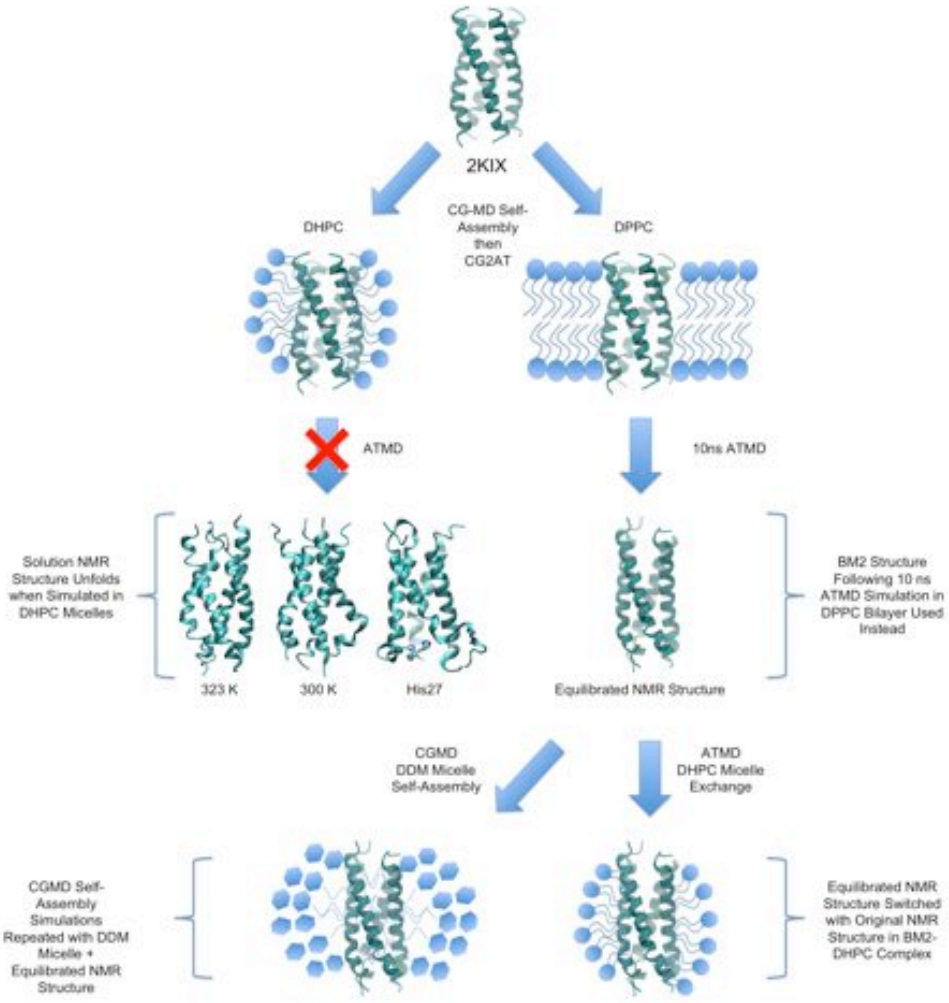


Figure 6:1. Equilibration of the solution NMR structure of the BM2 TM domain. Left: Snapshots of the solution NMR simulated in a DHPC micelle environment following i) 50 ns ATMD at 323 K ("323 K"), ii) 25 ns simulation at 300 K ("300 K "), C: 100 ns simulation with all 4 histidine residues doubly-protonated at 323 K ("His27"). Unfolding and deviation from ideal helicity is observed in each case. Once 25 % of α-helical structure was lost the simulations were terminated. This occurred within 50 ns and 10 ns for the systems at 323 K and 300 K, respectively.

Right: The solution NMR structure following simulation at 323 K for 10 ns in a DPPC bilayer was used to generate new BM2-detergent complexes. For BM2-DHPC the initial NMR structure was replaced with the equilibrated one using the program g_membed [1]. This system was then simulated for 20 ns with positional restraints on the protein to allow the micelle to relax. The CGMD formation of DDM and BM2 was repeated using the equilibrated structure (NB data described in Chapter 5 correspond to simulations using the equilibrated NMR structure).

1.1.1. The "Equilibrated NMR Structure"

The results of the simulations performed above indicate that the lateral pressure of the bilayer may have allowed the solution NMR structure to equilibrate and form interhelical contacts (coincident with decrease in pore radius). In the absence of this pressure, the helices appear to unfold before these contacts can be made. In order to simulate the NMR structure in a micelle environment (without significant helix unwinding and distortion), the structure of the TM domain following 10 ns equilibration in a DPPC bilayer (referred to as the "equilibrated NMR structure") was chosen as the starting point for further simulations. This is shown schematically in Figure 6:1. All following simulations of the "equilibrated NMR structure" refer to those using the NMR structure from 10 ns of ATMD simulation time in DPPC bilayer at 323 K as the starting structure. The initial configurations of the BM2-DHPC and BM2-DDM systems are shown in Figure 6:2.

1.2. Stability of the Equilibrated NMR Structure in DHPC and DDM Micelles

The structure and stability of the equilibrated NMR structure in each micelle environment was assessed in terms of the Cα RMSD and α-helical content during simulation at two temperatures. Figure 6:3A shows the Cα RMSD of the protein relative to the equilibrated NMR structure for simulations at 323 K for the DHPC and DDM micelle environments compared to a DPPC bilayer simulation. The structural drift during each simulation is comparable. Table 7 provides details of the structural stability of BM2 in repeat simulations. The least deviation from the equilibrated NMR structure is observed in the DHPC simulation at 300 K, closest to the conditions the solution NMR structure was solved in. The structure appears to equilibrate rapidly in the DDM environment as a large initial RMSD is observed which plateaus within 20 ns.

A decline in the number of residues in α-helical conformation is observed in all simulations. Similarly to the RMSD results, in DDM there appears to be a faster equilibration of the structure to the new environment, whereas in DHPC this process takes longer, with no clear plateau observed on the 100 ns timescale at 323 K.

The fluctuations of the protein show slight differences between the three environments. At 323 K the N-terminal region of the protein is more flexible in the bilayer environment

compared to the micelle systems. The C-terminal region, from His27 onwards, shows greater fluctuations than the N-terminal residues for DHPC and DPPC simulations, with the difference much more pronounced when in the DHPC micelle. In contrast, the protein in a DDM micelle has a fairly symmetrical increase in flexibility from the centre of the channel to its termini.

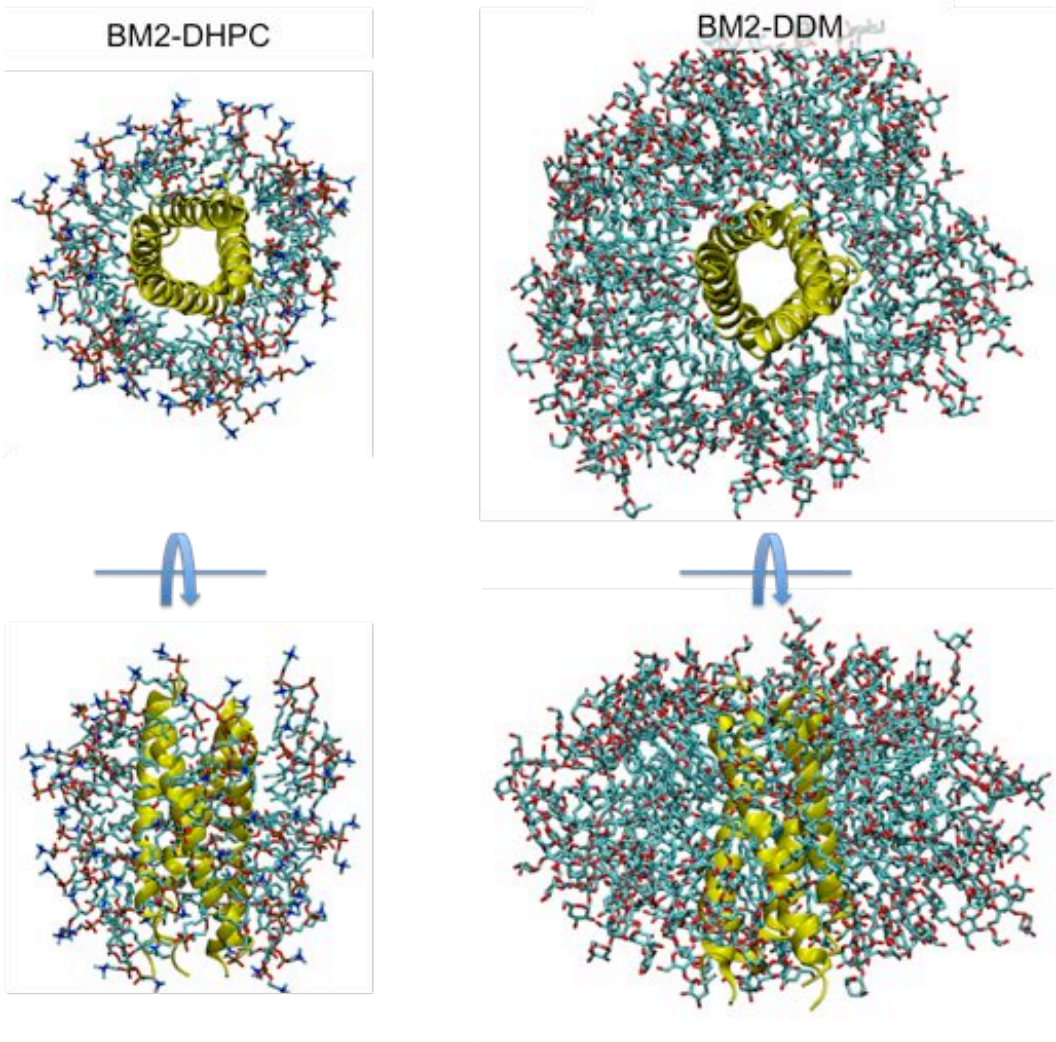


Figure 6:2. Starting configurations of micelle simulations. Side (A) and top (B) views of the equilibrated NMR structure in a DHPC micelle. 70 DHPC molecules are in the complex. Following self-assembly simulation. Side (C) and top (D) views of the equilibrated NMR structure within a DDM micelle, following encapsulation of the protein during CGMD simulation of the protein and a pure DDM micelle in solution. Here the top views correspond to the view down the principal axis of symmetry of the complex from the N-terminus of the protein.

The protein dynamics are dampened at lower temperature as may be expected. However, the fluctuations at the termini in the DHPC simulation are somewhat larger than that of DPPC or DDM. At this lower temperature the protein dynamics are very similar in both the DPPC bilayer and the DDM micelle, with only a slight increase in fluctuations of the terminal arginine residue in a DDM micelle compared to the DPPC simulations.

Detergent	Temperature (K)	Cα RMSD (Å)	α-helical Count (Number of Residues)	Maximum Bending Angle in Degrees
DHPC	323	1.8	90	32
DHPC	323	2.2	89	57
DHPC	323	1.7	100	51
DHPC	300	1.4	97	36
DDM	323	2.1	90	52
DDM	323	2.7	92	52
DDM	323	2.7	93	42
DDM	300	2.0	89	41

Table 7 Details of protein stability during simulations. The Cα RMSD corresponds to the deviation between structures at 0 and 100 ns of simulation. The number of residues in an α-helical conformation averaged over the final 10 ns is shown. The maximum bending angle in the structure at 100 ns is shown, calculated using Bendix [2].

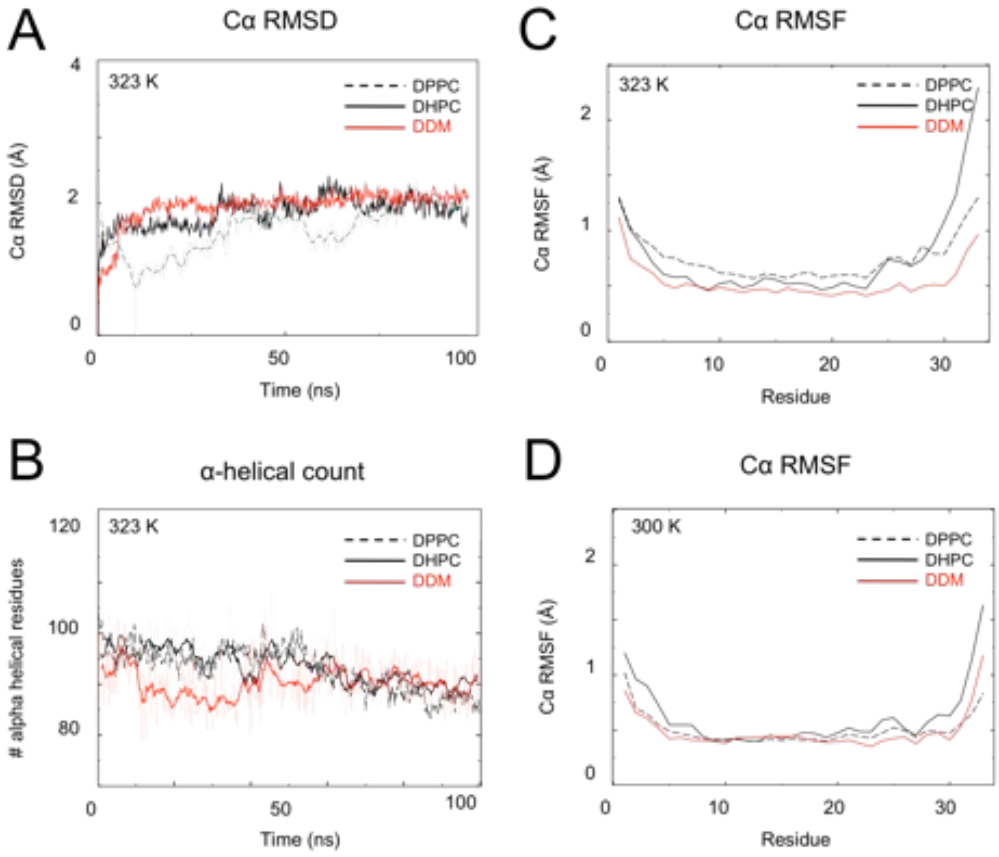


Figure 6:3. Stability and dynamics of the equilibrated NMR structure in a micelle environment. A: Cα RMSD from the equilibrated micelle structure, not including the terminal 2 residues at each end. B: α-helical count during 100 ns simulation. C: The RMS fluctuation of Cα particles during the final 10 ns of simulation at 323 K, averaged over the 4 chains. D: As for C, but for simulations at 300 K. The black dashed trace corresponds to DPPC bilayer simulations, the black solid trace corresponds to DHPC micelle simulations, and the red line corresponds to DDM simulations, as indicated in the figure legends.

As discussed above, a main destabilising feature of the solution NMR structure when incorporated into a bilayer appears to be the lipid-facing orientation of the His27 sidechains. To compare the preferred orientations adopted in micelle environments, spatial distribution plots showing the position of the His27 sidechains (in the reference frame of the protein) at various times over the 100 ns simulations were generated (Figure 6:4). The position of the His27 sidechains fluctuates by a similar order in the DHPC simulation as they did in the DPPC simulations, despite the initial structure being equilibrated in a bilayer. It appears that the His27 sidechains rearrange less upon transfer from a DPPC bilayer to a DDM micelle than they do upon transfer to a DHPC micelle.

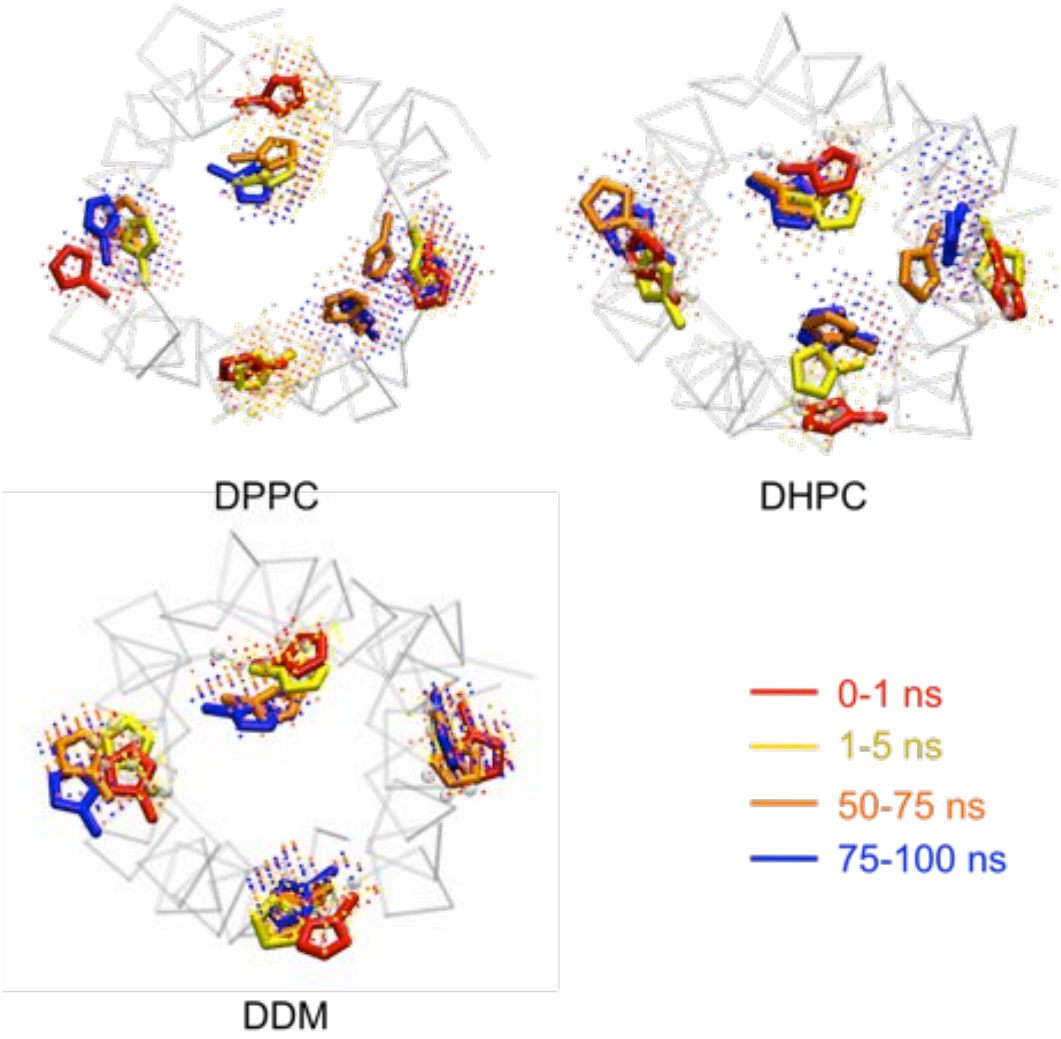


Figure 6:4. Spatial distribution plots of the His27 sidechains during 100 ns simulation in a DPPC bilayer, a DHPC micelle and a DDM micelle. The average coordinates during each time range are shown in stick representation and the area sampled by the sidechains during each time range is shown as dots. These are coloured according to simulation time as indicated in the key. The protein backbone is shown as a grey transparent trace. The initial His27 orientations are shown in grey (ball and sticks representation). Plots was generated using the tool g_spatial in GROMACS and visualising in VMD [3].

To assess if a conserved structure was obtained for the protein in each of the three environments, a Cα RMSD matrix was generated comparing the final structures from the three simulations at 323 K and one simulation at 300 K for each environment, as well as comparing the original solution NMR structure and the equilibrated NMR structure (Figure 6:5). The main structural conservation was observed for the four simulations of BM2 in a DDM micelle (average deviation < 2.5 Å). There was greater variation in the final structures between simulations in the DHPC micelle (average Cα RMSD > 3.5 Å).

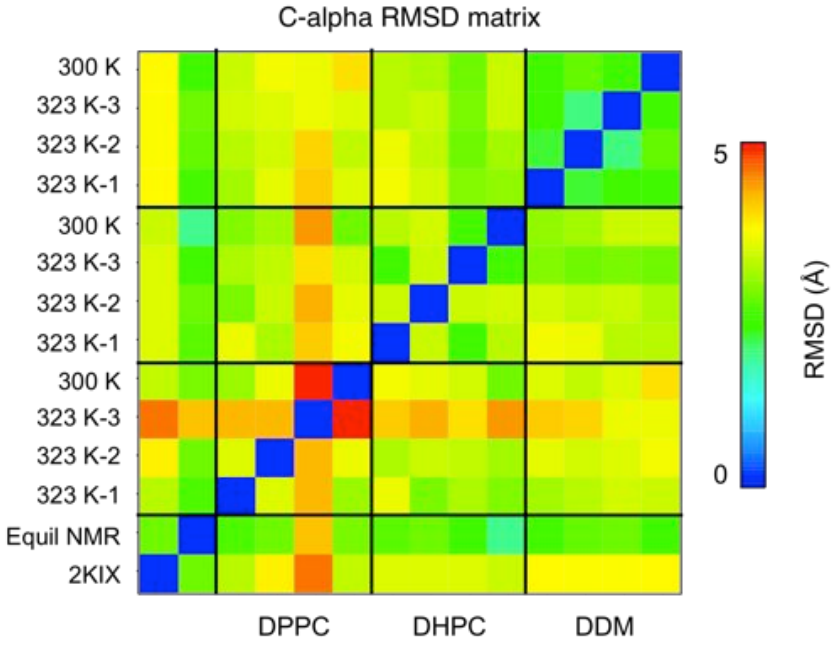


Figure 6:5 Cα RMSD matrix comparing the original solution NMR structure (2KIX) to the starting structure of the micelle simulations ("Equil NMR"), the final structures at 100 ns for three repeat simulations at 323 K and one at 300 K for bilayer simulations (DPPC) and micelle simulations (DHPC and DDM).

2. Architecture of Protein-Detergent Complexes

The shape of the protein-detergent complexes retain some of the features of the detergent-only micelles, as summarized in Table 8:

Protein	Detergent	a	b	c	a/c
No	DHPC	1.66	1.92	2.43	1.46
Yes	DHPC	2.48	2.69	3.10	1.25
No	DDM	3.42	3.04	2.72	0.79
Yes	DDM	3.91	3.29	2.84	0.73

Table 8 Semiaxis lengths (a and c) and ellipticity ratios of DHPC and DDM micelles with and without BM2 present. These were calculated from the instantaneous moments of inertia over the final 10 ns of each simulation (see text).

DHPC micelles are prolate both in the presence and absence of BM2. The architecture of the protein-DDM complex is similar to that of the DDM-only micelle. The overall shape of the complex is oblate, with the protein in the centre of the micelle. The detergent molecules close to the protein align such that the acyl tails are parallel to the helical bundle axis, while those further out from the complex centre of mass tend towards a perpendicular orientation. The maltose headgroups are again seen to cluster slightly on the surface as was observed in CGMD and ATMD simulations of the detergent-only micelle. To compare the DHPC and DDM micelle environments to that of the DPPC bilayer, density distribution plots of key residues and particles in each of the systems were calculated in intervals along the z-axis (principal axis of the protein) (Figure 6:6). Several key differences between each of the systems can be seen by comparison of the three density plots:

1) The protein is more compressed along its principal axis of symmetry in the DPPC bilayer compared to in DHPC or DDM micelles.
2) The protein sits asymmetrically in the DPPC and DDM systems. In DPPC the protein shifts upwards. This is presumably to allow interaction of the lysine and arginine residues with the DPPC headgroups. In DDM the protein sits below the centre of mass of the DDM micelle such that the lysine and arginine residues are exposed to the water. This is presumably because there are no possible saltbridge interactions between the non-ionic DDM detergent such that it becomes more favourable for these residues to interact with water instead.
3) A clear additional peak in the glycerol oxygen density is present in the DHPC micelle in the region of the lysine and arginine residues. There is no similar peak in the DPPC or DDM distributions. It appears the DHPC micelle is able to distort its geometry somehow such that lipid headgroups can interact with the whole protein.
4) There is density attributable to the acyl carbons of DHPC at the terminal regions of the protein. Similarly to 3), this is indicative of the DHPC micelle rearranging somehow to allow contacts of lipids with the protein termini that would not be possible in a bilayer environment.

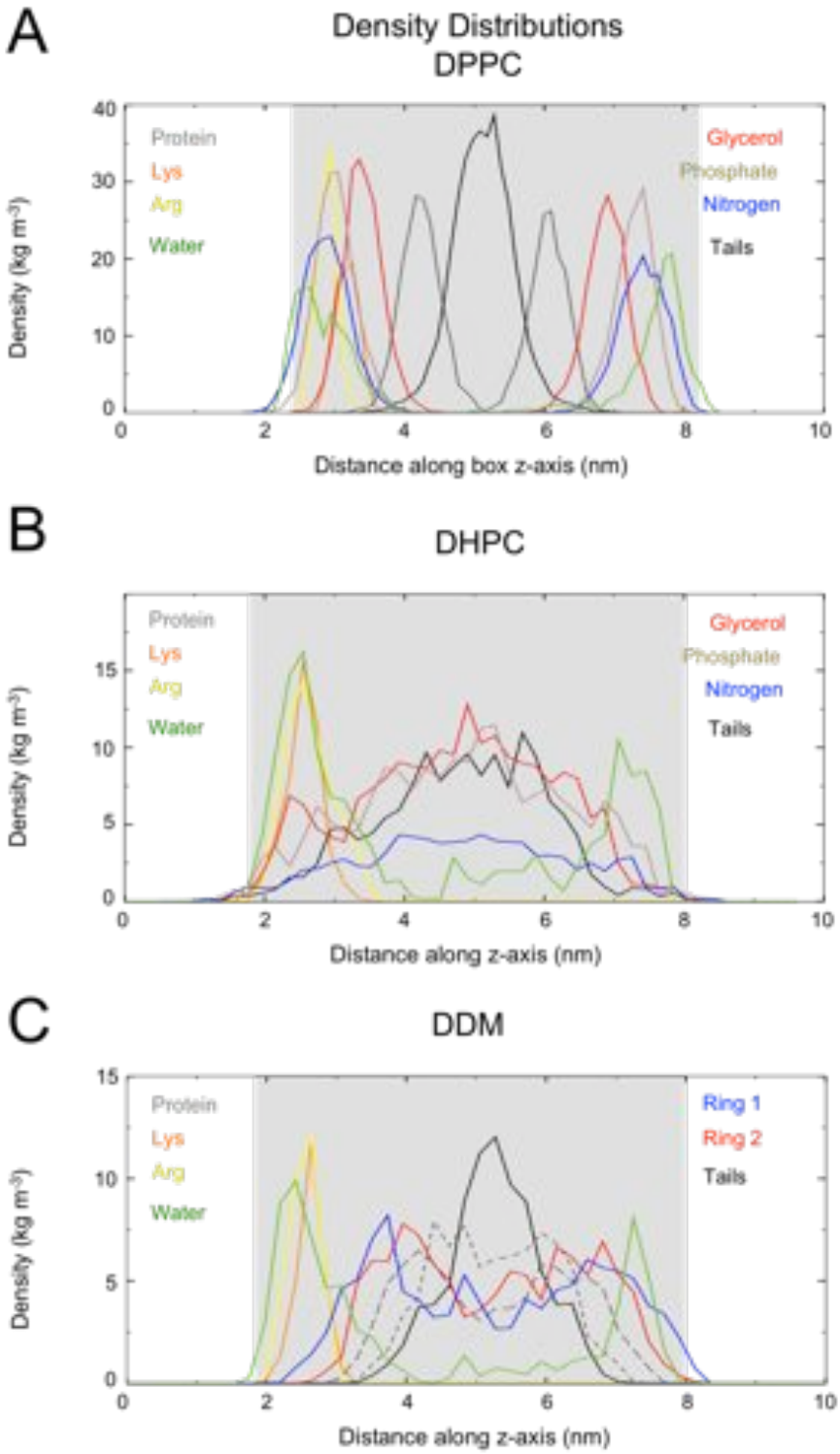


Figure 6:6. Density distribution plots of the A: DPPC B: DHPC and C: DDM systems. The density was calculated at 500 intervals along the z-axis (corresponding to the protein principle axis of symmetry). The area shaded in grey corresponds to the region containing protein. Lysine and Arginine residues are shown in orange and yellow, respectively. The nearest 100 water molecules to the protein are used in the calculation of water density. In A: only the closest 150 lipids to the protein are included in calculations and the density of the phosphorus atoms has been divided by two for clarity. In A and B only the first chain was used. The atoms used for calculating tails density are the terminal and middle carbons for DPPC, the terminal carbon for DHPC, and the carbons at each end of the acyl tail and the middle carbon for DDM. In C: the density Ring 1 corresponds to a single oxygen on the ring furthest from the acyl tail and the Ring 2 density corresponds to the equivalent oxygen of the other glucose ring.

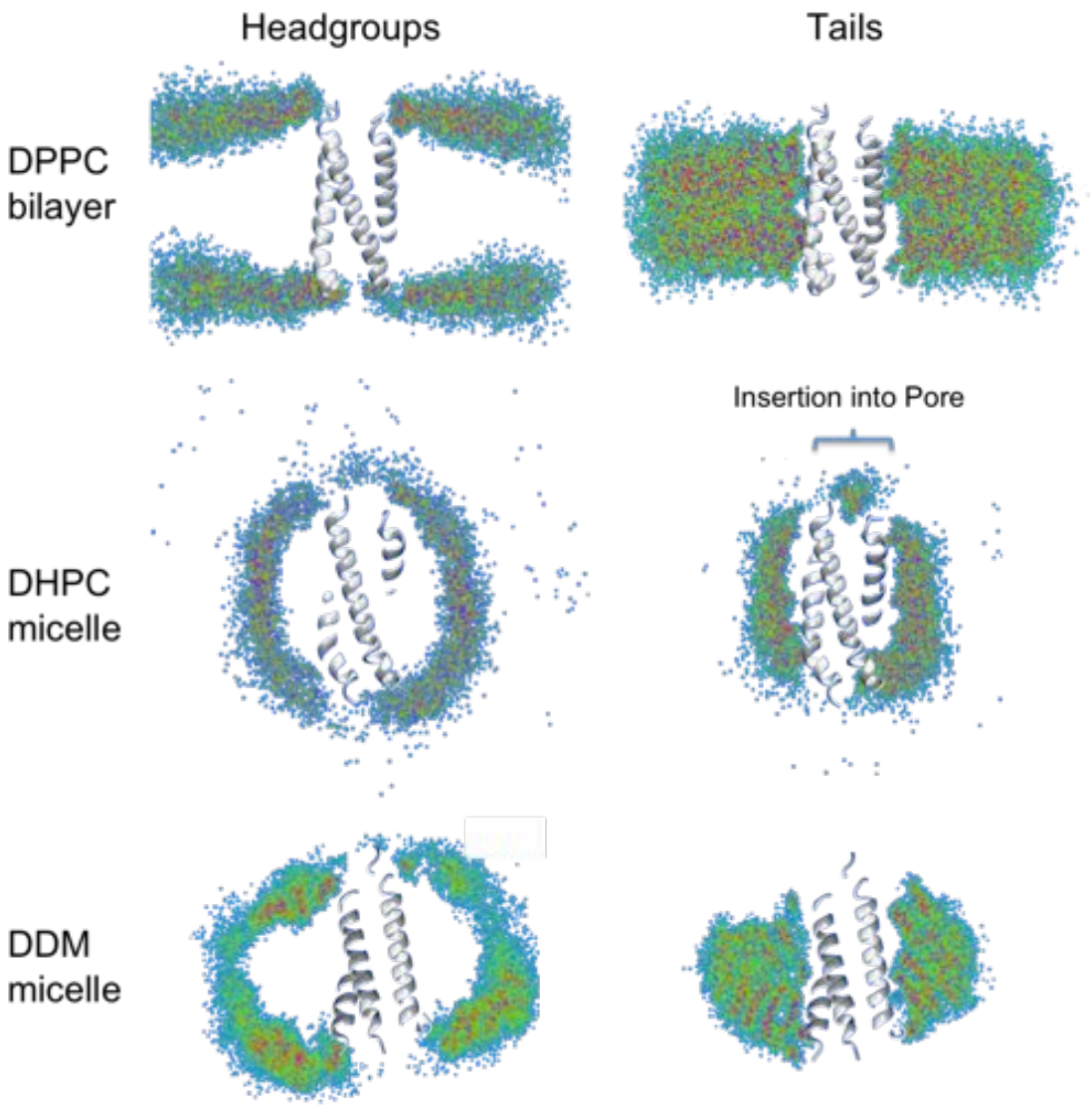


Figure 6:7. Volume slice through protein centre of mass of the spatial distributions of lipid/detergent headgroups and tails. These were calculated over the last 50 ns of simulation, in the reference frame of the protein. Red regions correspond to high density throughout simulation and blue regions correspond to lowest occupancy.

The differences between the distribution of the lipid/detergent headgroups and acyl tails were investigated further by generation of spatial distribution plots over the last 50 ns of simulation time for each system (Figure 6:7). In the DPPC bilayer it is observed that the lipid headgroups closest to the protein show the least movement during the simulation. This pattern is less pronounced for the DHPC and DDM systems. The density distribution of the DHPC micelle headgroups and tails is particularly illuminating. The presence of a detergent molecule forced into the pore at the N- and C-termini is indicated by regions of density. It is also apparent that one of more of the detergent molecules dissociates from the complex during the simulation. A more subtle difference between the DPPC and DDM distributions is in the dynamics of the acyl tails. Striations in the DDM micelle tail density indicate that some of the tails move very little with

respect to the protein during the simulation at 323 K. This suggests that the acyl chains may be more ordered in the protein-DDM complex than in the bilayer simulations.

Despite differences in the detergent distributions about the protein, in each environment the polar headgroup density is low around the hydrophobic core of the protein, matching the bilayer distribution and sequestering these residues from water.

3. Protein-Lipid Interactions & Detergent Binding Modes

The density distribution analysis gives information on the general architecture of the protein-detergent complexes. To understand more about how the lipids/detergents interact with the protein, lipid-protein contact analysis was performed (Figure 6:8). This demonstrated similarities in the patterns of contacts between the bilayer and micelle systems. A comparison of protein-headgroup interactions of the three systems shows that most of the contacts made in the DPPC system are replicated by the DHPC micelle. Some differences are observed in the DDM system, where the headgroups interact further up the protein from the C-terminus as the lysine and arginine residues are left to interact with solution. The protein-headgroup interactions at the N-terminal end of the protein are comparable in each of the three environments.

Comparison of the protein-tail contacts shows further differences between the systems. Firstly, it is evident that the micelle tails are able to pack more closely to the protein surface (increase in contact time compared to DPPC). The DHPC detergents forced into the C- and N- terminal regions of the channel are observed by a corresponding increase in contacts with tails in the regions whereas there are no equivalent contacts in DPPC. Again, the shifting of the DDM micelle hydrophobic core up towards the N-terminus compared to the DPPC hydrophobic core is observed (shown by increase in contact time of acyl tails in this region).

This contact analysis demonstrates that most of the contacts formed in a bilayer environment are fulfilled by the DHPC micelle environment (with some additional contacts) while for the DDM micelle the hydrophobic core is shifted slightly towards the N-terminus, allowing charged residues in the C-terminal region to interact with water.

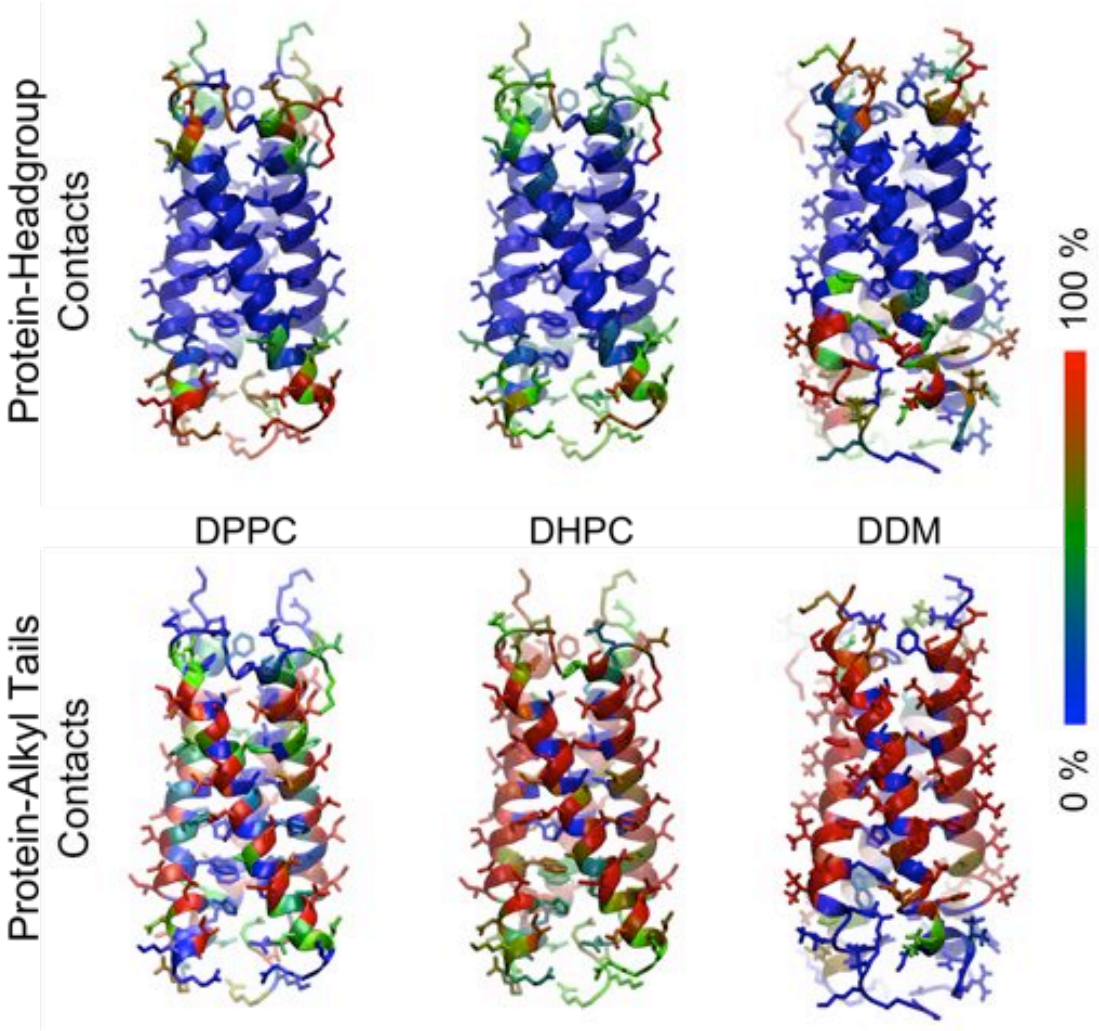


Figure 6:8. Lipid/Detergent interactions per residue. Residues are coloured according to time spent in contact with lipid/detergent headgroups or tails during simulation on a blue-green-red scale. Blue corresponds to 0 % of simulation, green = 50 % simulation time, and red = the residue is in contact with lipid headgroup or tail for 100 % of the simulation.

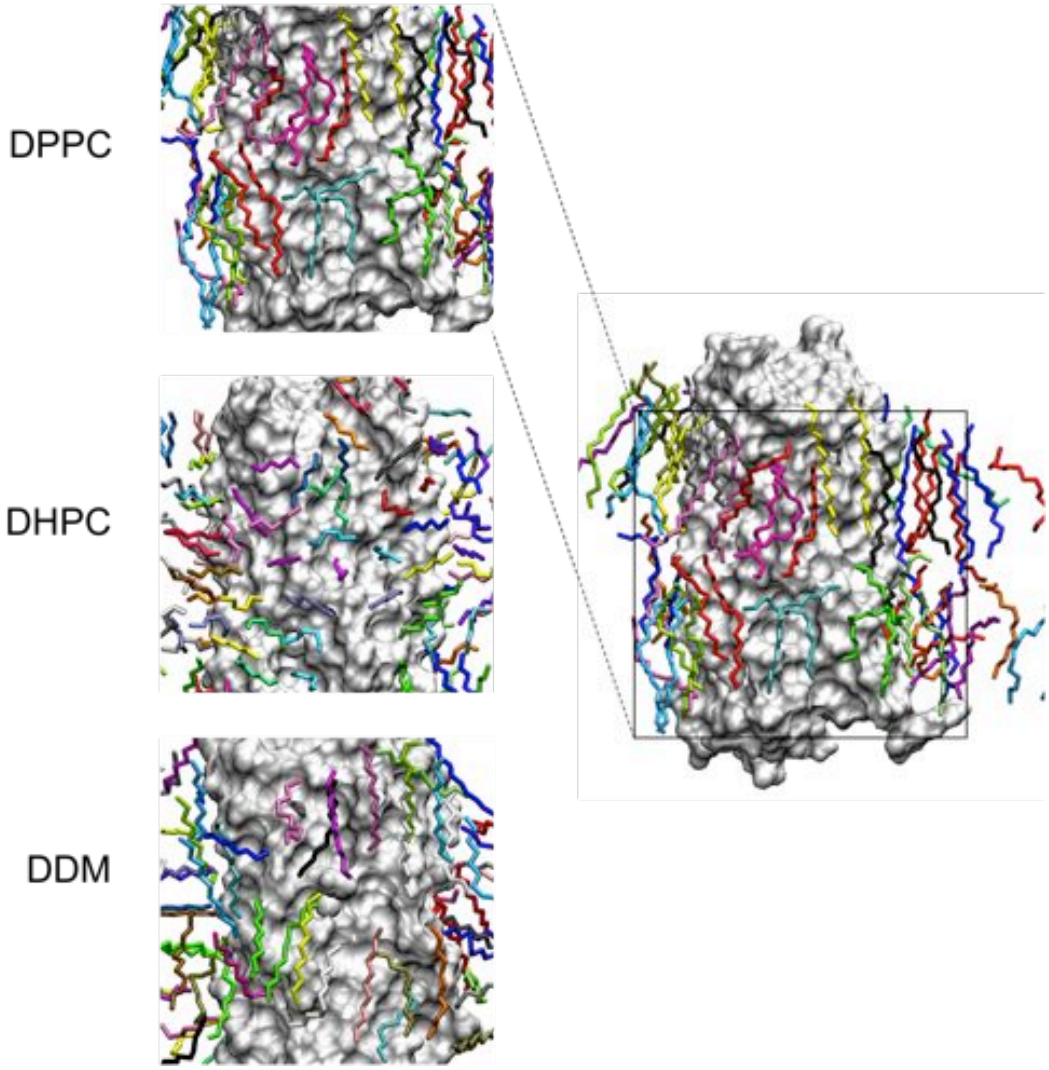


Figure 6:9. Lipid/detergent binding modes in DPPC bilayer (top) DHPC micelle (middle) and DDM micelle (bottom) simulations. The protein is displayed as a grey surface. Acyl tails of the lipids/detergents are shown in stick representation (coloured per molecule). The position of the tails is the average position adopted over the final 10 ns of each simulation performed at 323 K

Figure 6:9 shows the most occupied position of the lipid/detergent tails over the final 10 ns of each simulation. It can be seen that in the DPPC bilayer the lipid tails tend to align parallel to the bilayer surface. In the DHPC micelle, such a binding mode is not possible due to the shorter acyl chains. Instead, the majority of the lipid tails are aligned perpendicular to the surface, such that most of the interactions are between the terminal acyl carbon and the protein surface. This maximizes coverage of the protein hydrophobic surface. In the DDM micelle a slightly different picture is seen; the lipids close to the protein surface tend to align parallel to the surface as in DPPC, however, the shorter length of the DDM tail means that two extended carbon tails are not sufficient to cover the hydrophobic surface, and so slightly more distant detergent molecules align such that the end of their acyl chains are directed towards the protein surface.

To quantify the parallel *vs.* perpendicular binding mode, the order parameter S_{CD} may be used. This is typically used to measure the order parameters of lipid tails in aligned bilayers. However, it may also be used as a measure of alignment along the z-direction [4]. Here, it is used to indicate orientation parallel to the protein principle axis. A value of 1 implies parallel orientation, -0.5 corresponds to a perpendicular orientation and zero means that there is no overall alignment. The trajectories were renumbered such that the lipids were ordered in terms of the distance from the protein centre of mass for each frame. The S_{CD} order parameter may then be calculated as a function of distance from the protein centre of mass (Figure 6:10).

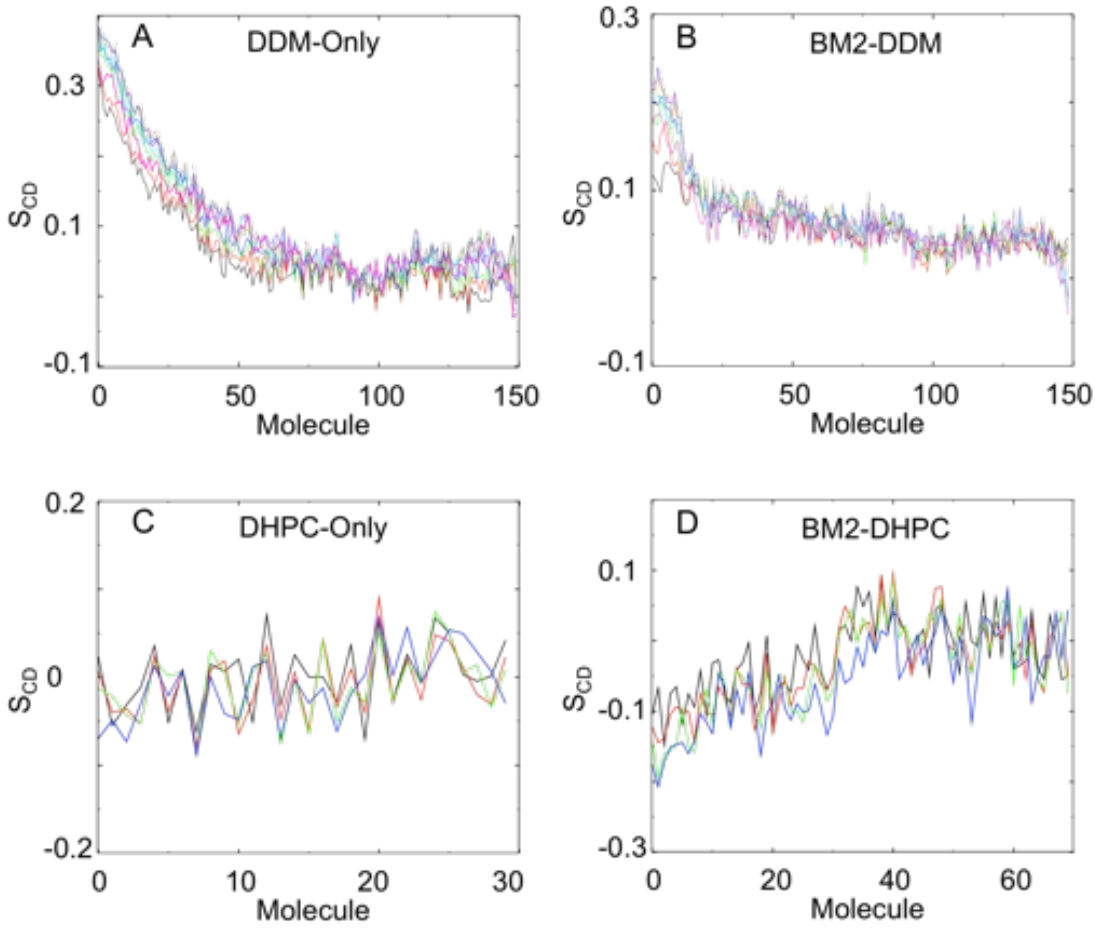


Figure 6:10. DDM and DHPC acyl chain order parameters S_{CD}. A: the DDM N=150 system and B: the BM2-DDM system. The detergent molecules are ordered in each frame in terms of distance from the protein, such that molecule 1 refers to the DDM lipid that is closest to the protein in a given frame of the trajectory and does not necessarily correspond to the same detergent molecule over time.

The plots of S_{CD} for DDM demonstrate a clear increase in alignment close to the protein, which gradually tails off to become non-aligned. Comparison of this property to the DDM-only micelle shows that in the pure detergent micelle the carbon tails are ordered and in a parallel orientation at the core of the micelle, and become less aligned further out from the centre of mass of the micelle. For the DHPC micelle, the detergent

molecules closest to the protein had a tendency to be perpendicularly aligned to the protein axis observed. It is difficult to draw conclusions from similar plots of the DHPC-only micelle (N_{agg} = 30) due to the comparatively small size of the micelle making calculations of the detergent molecule distances from the micelle centre of mass less reliable. The contrast between binding modes of the acyl tails of detergents in the DHPC and DDM micelles is clear from comparison of the S_{CD} plots.

To determine if any conserved interaction sites between the protein and the lipid/detergent acyl tails exist, the two most similar DPPC and DDM structures (based on the Cα RMSD matrix) were chosen for comparison. The average positions of the lipid/detergent tails between 0 and 10 ns for DDM and 10-20 ns for DPPC were superimposed and are shown in Figure 6:11. It appears that the lipid tails do not occupy specific regions of the protein, rather fit into the crevices on the surface.

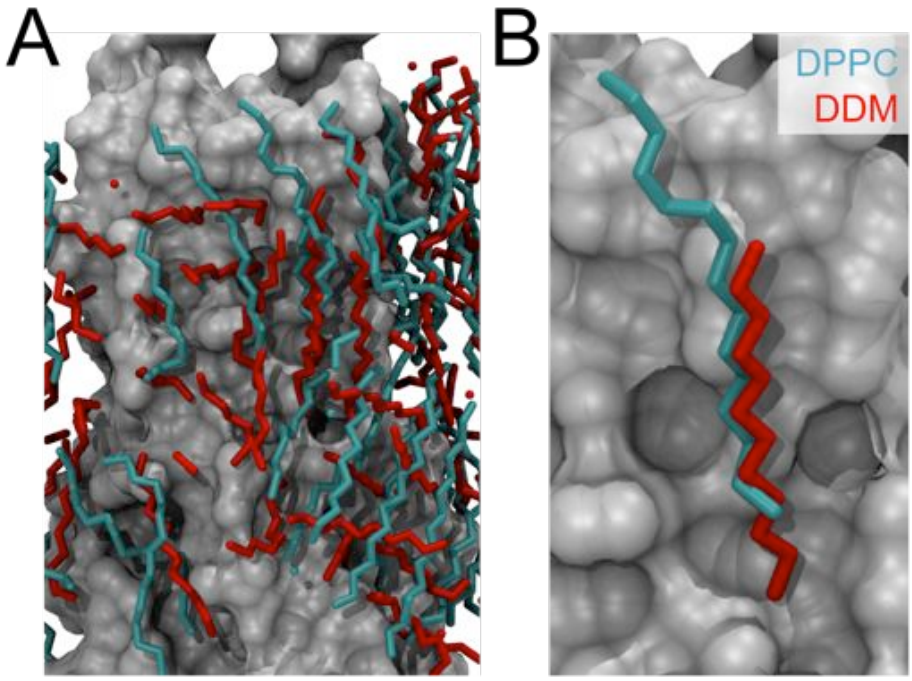


Figure 6:11. Interaction of acyl tails of DPPC and DDM with protein surface. A: Superimposition of most occupied positions of DPPC and DDM tails during 10 ns of each simulation (see text). The carbon chains of DPPC (cyan) and DDM (red) are displayed in stick representation. B: shows a rare example of where the same region on the protein surface has been occupied by an acyl tail in two different simulations. The protein is displayed as a grey surface in each case.

4. Lipid Dynamics

Although most of the specific residue-headgroup contacts between the protein and lipid are retained by the DHPC and DDM environments, visual inspection of the trajectories showed that the behaviour of the DHPC system is different in another way from the

bilayer environment. Spatial distribution plots were generated for DPPC, DHPC and DDM molecules over the 100 ns simulations performed at 323 K Figure 6:12.

A direct comparison of diffusion coefficients of lipid molecules in a bilayer compared to detergent molecules in a micelle would be limited:

i) The lipids in a bilayer are effectively restricted to lateral motion whereas in a micellar system the detergent molecules can potentially sample a spherical region

ii) The bilayer patch in question (~300 lipids) may not be large enough to avoid conserved motion of lipids between periodic images ([5])

Therefore, rather than attempt to compare diffusion coefficients of each lipid or detergent molecule, the trajectories were analysed such that the lipid or detergent molecules in each frame were ordered in terms of distance from the protein. The dynamics of lipids within 6 Å of the protein were analysed in each of the three systems.

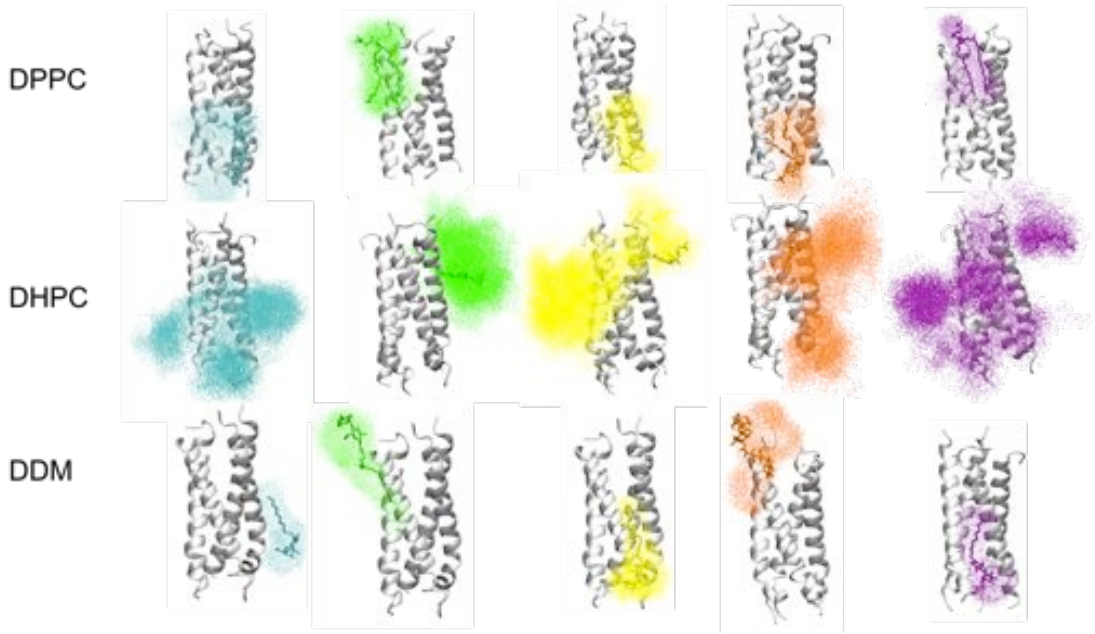


Figure 6:12. Spatial distribution plots of five representative lipid/detergent molecules within 6 A of the protein for each environment. The initial position of the lipid is shown in stick representation. The spatial distributions of each are shown as coloured dots. The lipid/detergent molecules in the DPPC and DDM simulations move a comparable amount during the simulation. Those in the DHPC simulation sample a far greater region of space.

The spatial distributions indicate that the picture of micelles being highly dynamic species is consistent with that of the small DHPC micelles. However, these motions are

considerably lower for DDM. The DPPC and DHPC headgroups are identical therefore the strength of interaction between residues and these headgroups would be expected to be similar. It appears that the very different geometry of the protein-DHPC complex leads to the higher mobility of the detergent molecules.

To establish if the less ordered lipids further from the protein in the DPPC and DDM systems display increased motion during the simulation, spatial distribution plots of representative lipids/detergents that were > 6 Å from the protein were calculated and shown in Figure 6:13. These lipid/detergents sample a far greater region of space during the simulation, comparable to that of the DHPC lipids.

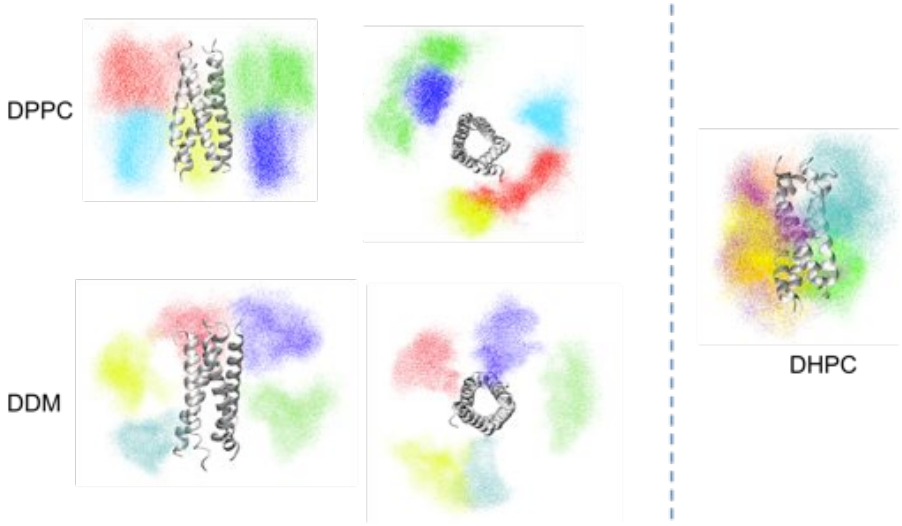


Figure 6:13. Spatial distribution plots of DPPC and DDM lipid/detergent molecules further away than 6 Å from the protein, compared to DHPC detergent molecule distributions (always less than 6 Å from the protein). Individual lipids are coloured separately.

5. Water Penetration Into the Hydrophobic Core and Solvation of Serine-rich N-terminal Region

The highly dynamic nature of the DHPC micelle might be expected to affect the degree of water penetration into the hydrophobic core. To assess if this was the case, calculations of the solvent accessible surface area (SASA) of the protein were performed for each environment and the average SASA per residue over the final 10 ns of simulation time is shown in Figure 6:14. These plots show similar features to the RMSF plots in Figure 6:3. Regions that are more exposed to solvent have correspondingly higher RMS

fluctuations. The low surface area of protein accessible to solvent is largely maintained in the hydrophobic transmembrane region with the main difference being at the C-terminal end of the channel (indicated by an asterix in Figure 6:14).

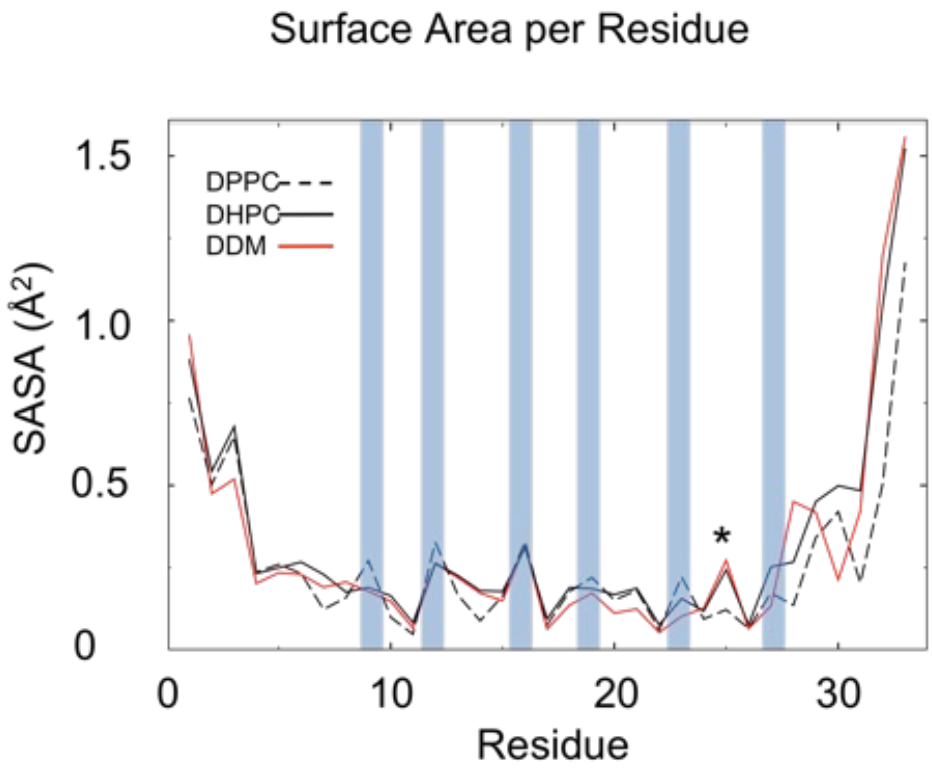


Figure 6:14. Averaged solvent accessible surface area over chain and last 10ns of each simulation. The DPPC simulation is shown as a black dashed line, the solid black line indicates the DHPC simulation and the red line corresponds to the DDM simulations. Blue bars indicate pore-lining residues. The main discrepancy between the three systems is indicated by a black asterix (*).

Radial density distribution plots of the water position relative to the protein over the simulations are shown in Figure 6:15. These highlight the hydrophobic band around the protein where no water is present in either of the micelle simulations. It is evident that there is water density within the protein pore in each of the simulations, with no clear distinction between the N-terminal region of the pore in the three simulations.

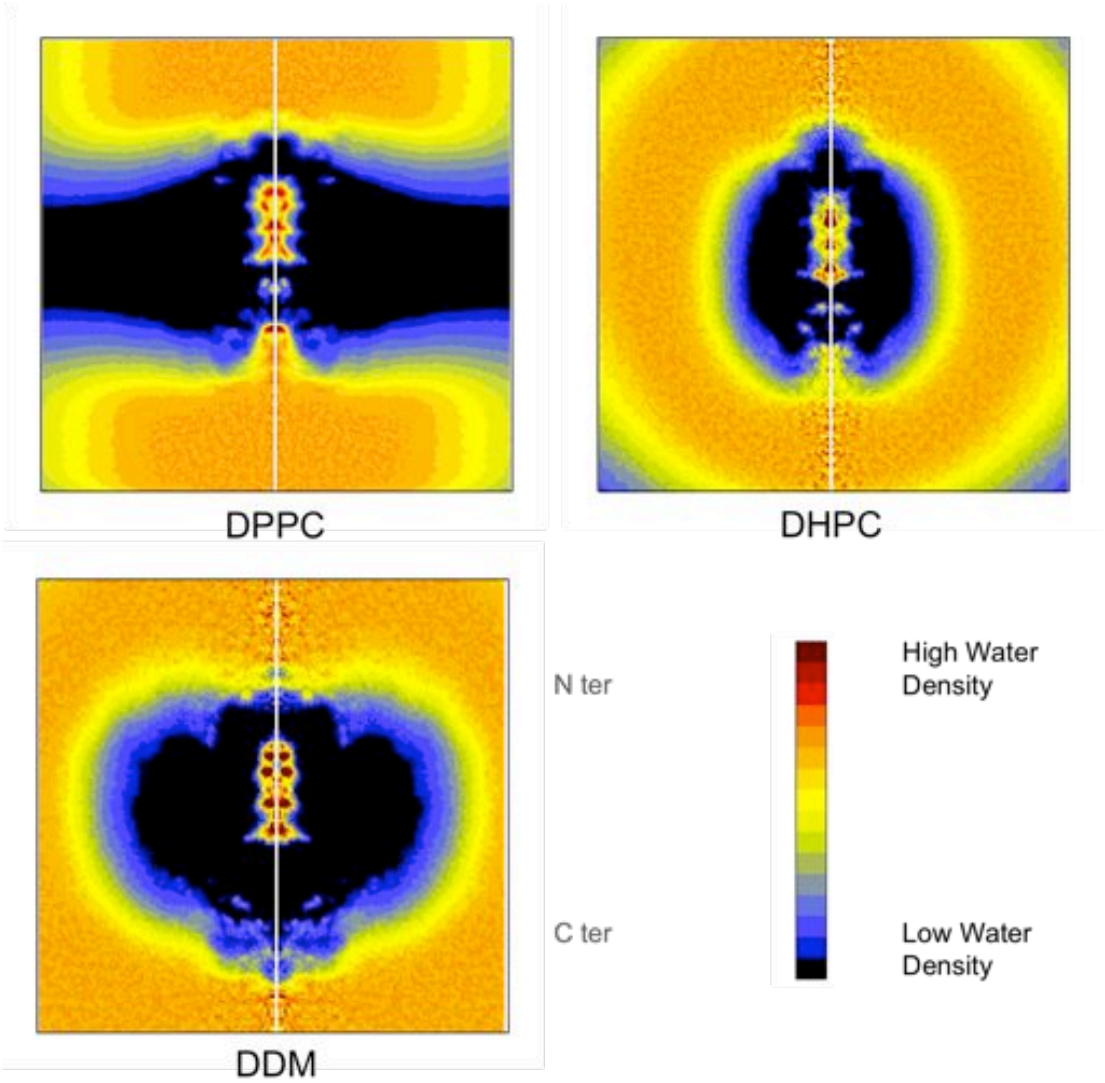


Figure 6:15. Radial density plots of water averaged over 100 ns simulation in a DPPC bilayer, DHPC micelle and DDM micelle. The axis of the BM2 tetramer is indicated by a white line. Water density is observed within the pore of the channel.

6. Discussion

The results of this Chapter allow comparisons to be made between three types of environment: a model DPPC lipid bilayer, small zwitterionic DHPC detergent micelles used in solution NMR experiment, and nonionic, larger DDM detergent micelles more commonly used in protein purification and X-ray crystallography. The observations made are discussed in the following subsections.

6.1. Architecture of DHPC and DDM Protein-Detergent Complexes

The results described in this chapter suggest that protein-detergent complexes (PDCs) appear to retain some of the features of the detergent-only complexes:

1) The ratio of the primary and secondary axes of symmetry does not change by a large amount upon incorporation of the membrane protein (see Table 8)
2) For DDM micelles, the lipids close to the centre of mass of the micelle are orientated and ordered along the principle axis of symmetry. This ordering is maintained once the protein is incorporated into the micelle and is related to the binding mode of the detergent. In DDM, the lipids close to the protein surface align along the surface, parallel with the principle axis of the complex. For DHPC, the original lipids do not show this ordering, and this is conserved in the protein-DHPC complex

That is, some of the properties of a membrane-protein detergent complex may be predicted from the properties of the pure detergent micelles. On the basis of this evidence, a detergent that forms prolate micelles is likely to form prolate PDCs, whereas a detergent that forms oblate micelles may be expected to form oblate PDCs. This has important implications on the choice of detergent for a given membrane protein as the sites of interaction of the detergent headgroups will depend on this feature. The implications of the ordering of lipids within the micelles are that *i)* for a DDM detergent micelle we may expect that the detergents close to the protein surfaces are orientated similarly to lipids in a bilayer, whereas *ii)* for a DHPC micelle, it appears that this lack of order allows the DHPC micelle to compensate for a range of hydrophobic lengths by altering its dimensions. This may contribute to the ability of DHPC to solubilise a range of membrane proteins but raises concerns over the possibility of subsequent stress inducing conformational changes in the membrane protein.

6.2. Protein Conformation, Detergent Packing, and Dynamics

The BM2 channel was shown to be sensitive to its environment, with the relatively weak helix-helix interactions readily perturbed by the presence of detergent molecules and/or solvent. Indeed, the solution NMR structure required prior simulation in a DPPC bilayer to prevent the structure unfolding within a DHPC micelle.

Micelles are often described as highly dynamic species [6]. The work in this chapter provides evidence that DHPC micelles fit this description, as the space sampled by the DHPC lipids is much greater than that by DDM or DPPC over the same timescales. However, the DDM detergents showed comparable diffusion rates to that of lipids within the DPPC bilayer. The lack of full interaction of the DHPC acyl tails with the protein surface may play a role in the faster exchange of DHPC over binding sites than for DDM [7], [8]. The ability of the acyl tails of DDM molecules to align into grooves on the protein surface in a similar fashion to the fatty acid tails of a phospholipid is perhaps a factor in the success of DDM detergent in membrane protein structural determination.

The protein did not show significantly greater RMS fluctuations in the more dynamic DHPC environment compared to DDM. It appears the presence of specific lipid/detergent-protein contacts is more important for membrane protein stability and conformation than the duration of the contacts. However, the full consequences of the increased dynamics of the detergent molecules on the protein stability and dynamics require further study over longer timescales.

The maltose-based headgroup of DDM appeared to lead to slightly different packing preferences within the protein-detergent complex compared to DPPC and DHPC phosphatidylcholine headgroups. In the BM2-DDM complex the C-terminal lysine and arginine residues were more exposed to solution than in a DHPC micelle or in the DPPC bilayer. This is presumably due to the inability of DDM to form salt bridges with charged residues, making increased solvation of these sidechains preferable.

6.3. Water Penetration and Solvation of Serine-lined Pore Region

The DHPC and DDM micelles allow a greater degree of interaction of the protein with water compared to the DPPC environment. However, the key hydrophobic regions are largely devoid of water as the hydrophobic tails arrange such that they interact with the

protein surface in each micelle despite the differing binding modes. The residues in the protein that have increased solvent accessible surface area in the micelle environments correspond to those that displayed an increase in RMSF over the last 10 ns of each simulation. This was limited to residues in the C-terminal region of the protein. The N-terminal region of the BM2 pore was found to remain hydrated for the duration of the simulations in each environment. The hydration of this region appears to be necessary for the stability of the structure, regardless of environment.

7. Further Work

7.1. Gaussian Network Modelling of the Solution NMR Structure of the BM2 TM Domain

The problem of NMR structures leading to greater structural drift upon MD simulation than X-ray structures is well-reported within the literature (see *e.g.* [9], [10]). In this work the solution NMR structure was simulated in a bilayer to give a low-energy starting structure for the micelle simulations. There are many other ways to achieve a minimised structure, including MD simulation enforcing NMR distance restraints [11], [12]. A potential method for improving the packing of the structure for further simulation would be to apply a Gaussian Network Model to each helix in the protein, allowing the major conformation of each monomer to be maintained, while helix-helix packing can alter. Preliminary simulations demonstrated that the degree of helix structure retention and overall tetrameric packing could be modulated by varying the strength of the GNM restraints. This could be a potential way to equilibrate solution NMR structures in future MD simulation studies.

7.2. Extension to Native Lipid Environment

The use of DPPC lipid in the bilayer simulations of BM2 was chosen as a model membrane as it has been extensively used in MD simulations. However, components of native membranes have been shown to have dramatic effects on membrane protein packing. The effect of unsaturated lipids and cholesterol on membrane properties has previously been studied computationally and shown to have effects on the energetics of membrane protein conformational transitions [13]. The curvature of a membrane bilayer leaflet in a vesicle rather than a flat bilayer is also known to modulate packing of

membrane proteins. The next logical step in these simulations would be to include elements of biological membranes.

8. Summary

The aim of this Chapter was to investigate the influence of environment on membrane proteins, using BM2 as a model α-helical protein. It was found that DDM micelles replicated the packing properties of lipids in the DPPC bilayer more closely than DHPC micelles, despite DHPC having the same zwitterionic headgroup as DPPC. This was due to the longer acyl tails of DDM being able to fit into grooves on the protein surface as the DPPC lipids were observed to. The longer acyl tails provided a larger interaction surface per detergent molecule for DDM than for DHPC, which appeared to lead to a slower translation of detergent molecules around the protein compared to the DHPC detergents which were observed to sample a far greater region of space over the same timescales

The binding modes of detergents within the protein-detergent complex appeared to be related to that of the pure detergent micelles, providing further evidence that nature of a membrane protein complex does not necessarily depend on the headgroup type of the lipid used, but rather the geometric properties of the detergent micelle [14].

The protein adopted a slightly compressed (along the bundle axis) structure in the bilayer environment compared to the DHPC and DDM micelle environments. It appears that BM2 is able to modulate its packing to compensate for hydrophobic mismatch in a bilayer environment (as discussed in Chapter 4) whereas in the less rigid environment of a detergent micelle the hydrophobic dimensions of the micelles preferentially alter. The results of these simulations provide further evidence that BM2, similarly to A/M2, is likely to exist in a wide range of conformations due to the relatively weak helix-helix packing interactions.

9. References

[1] M. G. Wolf, M. Hoefling, C. Aponte-Santamaría, H. Grubmüller, and G. Groenhof, "g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation.," *Journal of computational chemistry*, vol. 31, no. 11, pp. 2169-74, Aug. 2010.

[2] C. Dahl, "Bendix," *Manuscript in Press*, 2011.

[3] W. Humphrey, A. Dalke, and K. Schulten, "VMD: visual molecular dynamics.," *Journal of molecular graphics*, vol. 14, no. 1, pp. 33-8, 27-8, Feb. 1996.

[4] A. Leach, "Molecular Modelling: principles and applications," 2001.

[5] J. B. Klauda, B. R. Brooks, and R. W. Pastor, "Dynamical motions of lipids and a finite size effect in simulations of bilayers.," *The Journal of chemical physics*, vol. 125, no. 14, p. 144710, Oct. 2006.

[6] "ON THE KINETICS OF MICELLIZATION IN AQUEOUS-SOLUTIONS," *ADVANCES IN COLLOID AND INTERFACE SCIENCE*, vol. 13, no. 1-2, pp. 1-64, 1980.

[7] H. Palsdottir and C. Hunte, "Lipids in membrane protein structures.," *Biochimica et biophysica acta*, vol. 1666, no. 1-2, pp. 2-18, Nov. 2004.

[8] C. Hunte, "Specific protein-lipid interactions in membrane proteins.," *Biochemical Society transactions*, vol. 33, no. 5, pp. 938-42, Nov. 2005.

[9] H. Fan and A. E. Mark, "Relative stability of protein structures determined by X-ray crystallography or NMR spectroscopy: a molecular dynamics simulation study.," *Proteins*, vol. 53, no. 1, pp. 111-20, Oct. 2003.

[10] K. Cox, P. J. Bond, A. Grottesi, M. Baaden, and M. S. P. Sansom, "Outer membrane proteins: comparing X-ray and NMR structures by MD simulations in lipid bilayers.," *European biophysics journal: EBJ*, vol. 37, no. 2, pp. 131-41, Feb. 2008.

[11] P. Sompornpisut, B. Roux, and E. Perozo, "Structural refinement of membrane proteins by restrained molecular dynamics and solvent accessibility data.," *Biophysical journal*, vol. 95, no. 11, pp. 5349-61, Dec. 2008.

[12] A. E. Torda and W. F. V. Gunsteren, "The refinement of NMR structures by molecular dynamics simulation," *Computer Physics Communications*, vol. 62, pp. 289-296, 1991.

[13] P. S. Niemelä, S. Ollila, M. T. Hyvönen, M. Karttunen, and I. Vattulainen, "Assessing the nature of lipid raft membranes.," *PLoS computational biology*, vol. 3, no. 2, p. e34, Feb. 2007.

[14] R. M. Bill et al., "Overcoming barriers to membrane protein structure determination," *Nature Biotechnology*, vol. 29, no. 4, pp. 335-340, Apr. 2011.

Chapter 7

MD for MS: "Micelles Protect Membrane Protein Complexes from Solution to Vacuum"

1. Introduction

Mass spectrometry (MS) is a technique that has become widely used in protein characterisation, and its use is still expanding. MS has several attractive features compared to other experimental techniques. These have been referred to as the "S" advantages: speed, specificity and sensitivity [1], and more recently "stoichiometry", referring to its use in determining stoichiometry of complexes [2]. MS is highly sensitive; experiments can be performed rapidly and are compatible with an automated, high-throughput approach; and a range of quantitative and qualitative observables can be captured [3]. Initial MS experiments were limited to small, stable compounds, however the development of the 'gentle' techniques *i)* electrospray ionization (ESI) [4] and *ii)* matrix assisted laser desorption/ionization (MALDI) [5] paved the way for MS to be used in the study of peptides and larger macromolecular complexes. [6] Combining ESI/MALDI with new analyzers and instruments allows MS to be used to e.g. identify the stoichiometry of complexes [7], study conformational changes in real time [8] and quantify protein-ligand interactions. [9].

The process of electrospray ionization (ESI) may be broken down into the following stages [10], [11] :

1) The analyte is pushed through a capillary tube under applied voltage

2) This voltage leads to distortion of the meniscus at the end of the capillary tube, leading to formation of a 'Taylor cone'
3) As charge increases, the Coulombic repulsion begins to outweigh surface tension
4) At this limit, charged droplets are released as a spray into the collision cell and travel towards the detector due to the pressure and potential gradients
5) Transfer to low pressure environment, as well as collisions with neutral species, lead to evaporation of waters from the droplet
6) Evaporation of water leads to increase in surface charge density. Increases until critical point (Raleigh limit) is reached, at which ions are expelled from the droplet
7) Naked ion is detected and mass spectrum obtained

These features are shown schematically in Figure 7:1.

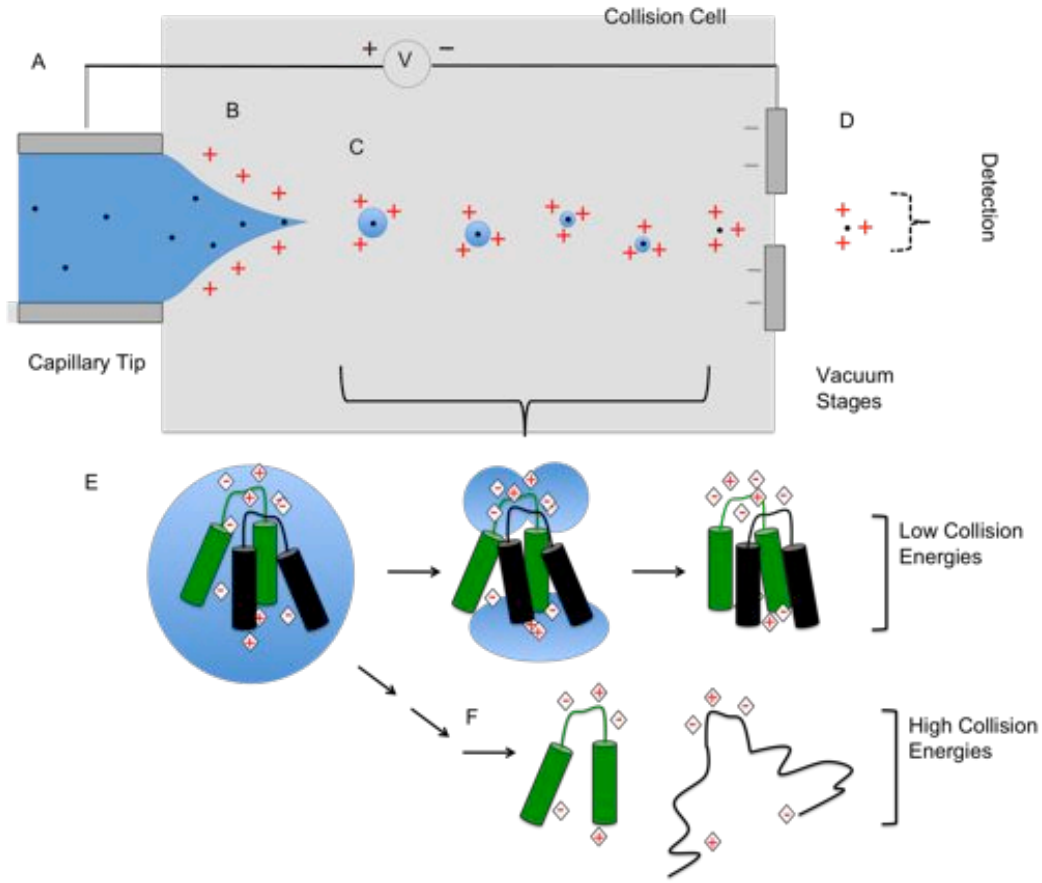


Figure 7:1. The process of electrospray ionization. A: The analyte solution is pushed through a capillary tip and charged by an applied voltage B: Formation of Taylor cone (the deformation of the meniscus due to the electric field). Above a threshold voltage a continual spray of droplets is released into the collision cell. C: These charged droplets travel thought the collision cell to the detector due to pressure and electric field gradients, gradually becoming dehydrated D: The dehydrated ion passes into the vacuum regions of the instrument for detection E: the desolvation of a dimeric protein complex under the "low collision energy" regime. Water evaporates last from the charged sidechains. The surface charges may rearrange upon desolvation to give a new, vacuum conformation F: Under the "high collision energy" regime the increased energy of collisions with neutral particles leads to dissociation and possible unfolding of the complex. In practice, careful control of the ESI conditions can be used to follow pathway E or F.

Of course, the process of ESI leads to a protein complex in an environment very different to that of solution phase experiments, raising the immediate question "is the solution state structure retained in vacuum?". Substantial evidence exists suggesting that large proteins, nucleic acids and even whole virion particles retain many aspects of their native structure upon transfer to vacuum [11-14]. Experimental and computational studies of the soluble proteins cytochrome c and ubiquitin suggest that the native structure may be stabilized transiently (millisecond timescale) by electrostatic interactions on the surface but that the conformation evolves over long timescales (seconds to minutes) to form new folds [12], [15]. The overall picture appears to be that the solution state structure may be stabilised in the gas phase, although not in all cases, and depends upon the timescales involved.

The focus of this chapter is based on the findings of Barrera *et al.*; that "micelles protect membrane complexes from solution to vacuum" [16]. This was the first reported ESI study of an intact heteromeric membrane protein complex (the bacterial ABC-transporter BtuCD) containing both transmembrane and cytoplasmic units. Most ESI protocols prior to this work involved removal of detergent molecules to study membrane proteins [17]. However, in this work (and elsewhere *e.g.* [18]) it was found that the presence of DDM detergent molecules in a protein-detergent complex appeared to stabilize the membrane protein during transition to vacuum (this was based on the fact that soluble protein complexes were stable under equivalent conditions).

Several MD studies have been used to elucidate structural changes upon dehydration of globular proteins under vacuum conditions [19-23]. These have proved informative in describing the changes that proteins undergo during the dehydration process. To date, only a single simulation study has been published of a membrane protein-detergent complex in vacuum [24]. In this study the authors simulated the outer membrane protein A (OmpA), a β-barrel protein, in a DPC micelle with water shells of varying thicknesses (0, 3 and 6 Å). The main findings of this study were that changes upon transfer to vacuum were mainly limited to detergent packing, whilst the protein remained largely unaffected. However, the nature of these simulations (manual removal of waters to give isolated protein-detergent complex) as well as the choice of a monomeric β-barrel protein (*i.e.* stable) limit the conclusions that may be drawn from this work.

In this Chapter, more of the processes involved in ESI are studied. This is achieved by three stages of simulation:

1) Steered molecular dynamics (SMD) simulations to transfer the protein-detergent complex from bulk solution to vacuum (Stages A and B in Figure 7:1)
2) Simulation under vacuum conditions to allow further dehydration of the resulting sparingly-solvated complex (Stage E in Figure 7:1)
3) Simulation at increasing (collision) energy to assess the effects on stability of the complex (Stage F in Figure 7:1)

Three systems are studied in this chapter, the choice of each may be rationalised as follows:

1) OmpA in DPC. To compare the methodology used here to that of the single published MD study of a membrane protein-detergent complex in vacuum.
2) BM2 in DHPC. This small, α-helical tetrameric protein would be expected to be more sensitive than OmpA to changes in environment. Potential monomer dissociation may be observed.
3) BM2 in DDM. This allows comparison between the non-ionic DDM used successfully in membrane protein ESI-MS experiments and zwitterionic DHPC.

2. Methods

The three classes of simulation studies performed in this chapter are described in the corresponding subsections.

2.1. Transfer of protein-detergent complex from solution to vacuum using steered MD

To replicate the initial desolvation stages that occur in the electrospray process, the protein-detergent complex (PDC) was 'pulled' from bulk solution into vacuum using steered molecular dynamics simulations (Methods Chapter). Briefly, this was achieved by *i*) extending the simulation box in the z-direction from the endpoint of a simulation of the protein-detergent complex in solution (Figure 7:2B), *ii*) coupling a virtual spring to the centre of mass of the PDC *iii*) applying a harmonic potential force to the spring such that the force increases with the distance between the spring and the centre of mass of the bulk solvent (Figure 7:2C). This results in a gradual increase of force at the start of the simulation rather than a rapid constant force. The result of these simulations is a sparingly solvated membrane protein complex in vacuum (Figure 7:2D).

The pressure coupling was switched off and the integrator chosen was stochastic dynamics. Following ionization the sprayed droplets travel through the collision cell, which is typically at pressures of 10^{-6}-10^{-7} mbar, therefore collisions and friction due to presence of gas particles is expected [25]. The use of stochastic dynamics applies a friction term to each atom, mimicking this effect. It also acts as a thermostat, keeping the overall temperature of the system constant [26]. The force constant, pull rate and 'pull groups' chosen for these simulations were calibrated using the BM2-DHPC system as a test case (see Appendix A).

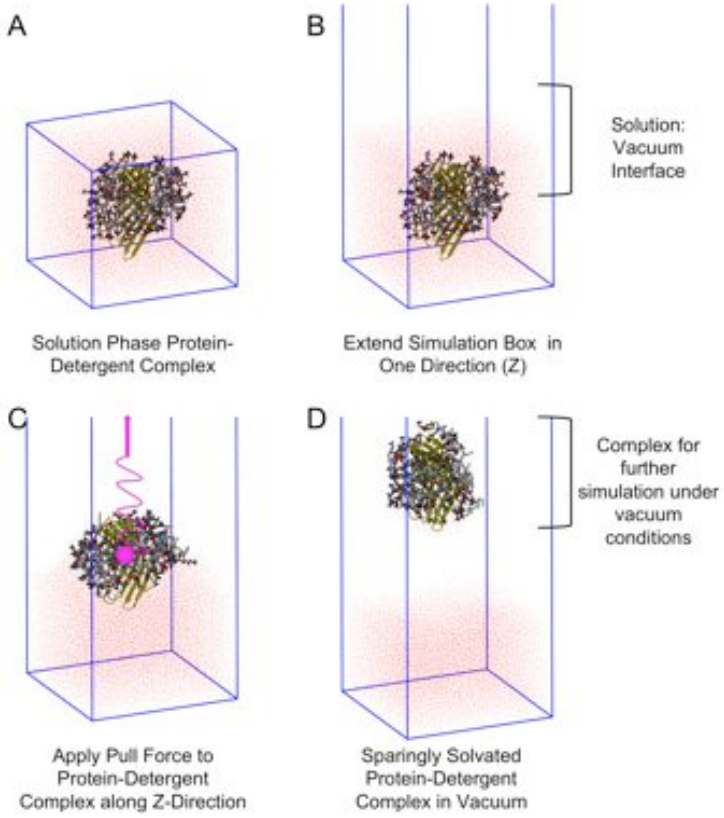


Figure 7:2. Schematic representation of the steered MD method. A: Initial system is taken from a well-equilibrated solution phase simulation of the protein-detergent complex B: The simulation box is extended in one direction (Z). C: The harmonic force is applied to a spring that is attached to the centre of mass of the protein-detergent complex. D: The spring is pulled along the Z-direction from bulk solution into vacuum, leading to a sparingly solvated complex.

2.2. Simulation in "true" vacuum

Following the transfer from bulk solution to vacuum, the main protein-detergent complex (including any water molecules directly or indirectly in contact) is extracted and simulated in a vacuum environment:

1) Dielectric constant is set to 1
2) No cut offs used (interactions between *all* particles calculated)
3) Pressure coupling switched off

In a true vacuum the temperature coupling would also be switched off. However, the removal of temperature coupling leads to fast evaporation of surface waters, taking a large amount of the internal energy with them. The remaining complex is then at low temperature (in these systems, ~270 K after a few ns from a starting temperature of 323 K), and little of the system dynamics or rearrangement may be observed. Therefore temperature coupling was used to maintain the energy of the system. This allows control of the energy of the complex in a similar fashion to the control over collision energy in the corresponding electrospray ionization experiments. Due to water being able to evaporate from the complex each component of the system (protein, detergent and water) was temperature coupled separately.

2.3. Simulation at High Temperatures

In electrospray ionization the collision energy of the cell is controlled to monitor dissociation of subunits and/or detergents from the complex. Here, the temperature of the protein-detergent complex is gradually increased (following 150 ns simulation under vacuum conditions) to investigate the resultant effects on stability and structure. Simulation at elevated temperatures is commonly used in MD studies of unfolding processes that otherwise occur too slowly to be observed with brute force MD; an increase in temperature from 298 K to 498 K can increase unfolding rates by 10^3-10^9 orders of magnitude [27].

A control simulation of the protein only was also performed. A temperature range of 323 – 750 K was used in these simulations, increasing linearly at a rate of 5 K per nanosecond. This temperature range is within those used in previous studies of proteins folding kinetics *in silico* [28].

3. Results

3.1. Transfer of Protein-Detergent Complex from Solution to Vacuum

3.1.1. OmpA in DPC Micelle

To compare the steered MD approach to published work, the same initial OmpA-DPC complex was taken as the starting point (coordinates provided by Peter J. Bond). The PDC at the beginning and end of the SMD trajectory is shown in Figure 7:3. There is an observable change to the arrangement of the detergents about the protein; from visual

inspection alone it is apparent that the detergent molecules are covering some of the loop regions that are normally accessible to solvent in the bulk solution simulations. To assess quantitatively whether the transfer from solution to vacuum using the SMD approach resulted in the same outcome as the previous simulation study (i.e. structural changes largely confined to the detergent micelle), several properties of the SMD trajectory were analyzed. These may be classed into those concerning i) changes in the protein and ii) changes to the packing of the detergent micelle.

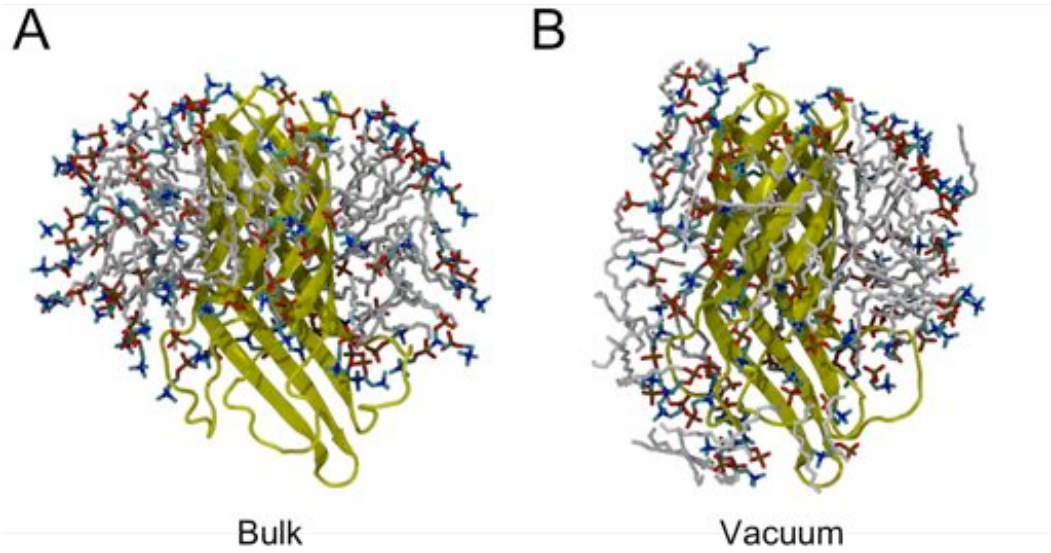


Figure 7:3. Arrangement of DPC detergent molecules about OmpA in bulk solution (A) and following transfer from solution to vacuum (B). The protein is shown in cartoon representation and coloured yellow. DPC is shown in licorice representation (headgroups coloured as follows: cyan = carbon, oxygen = red, nitrogen = blue, phosphorus = tan. Acyl tails are shown in white). Water is omitted for clarity.

The stability of OmpA may be quantified in terms of RMSD and secondary structure count. Figure 7:4 shows the variation of these properties during the transfer from bulk solution to vacuum. A temporary decrease in the number of β-sheet residues is observed part way through the transition, when approximately half of the complex is removed from the bulk solution. However this appears to be more of an instantaneous fluctuation as the difference from the average is within the standard deviation of the data, and these residues quickly (within 1 ns) regain their ideal conformation. The Cα RMSD of the β-sheet residues is relatively unchanged during the transition, as may be expected from the near-complete retention of the secondary structure of this region. The loop regions (defined here as residues not in β-sheet conformation at the beginning of the simulation) deviate to a greater degree from the initial structure, which may be expected from the greater flexibility of these regions. The RMS fluctuations of the Cα particles are shown in Figure 7:4C.

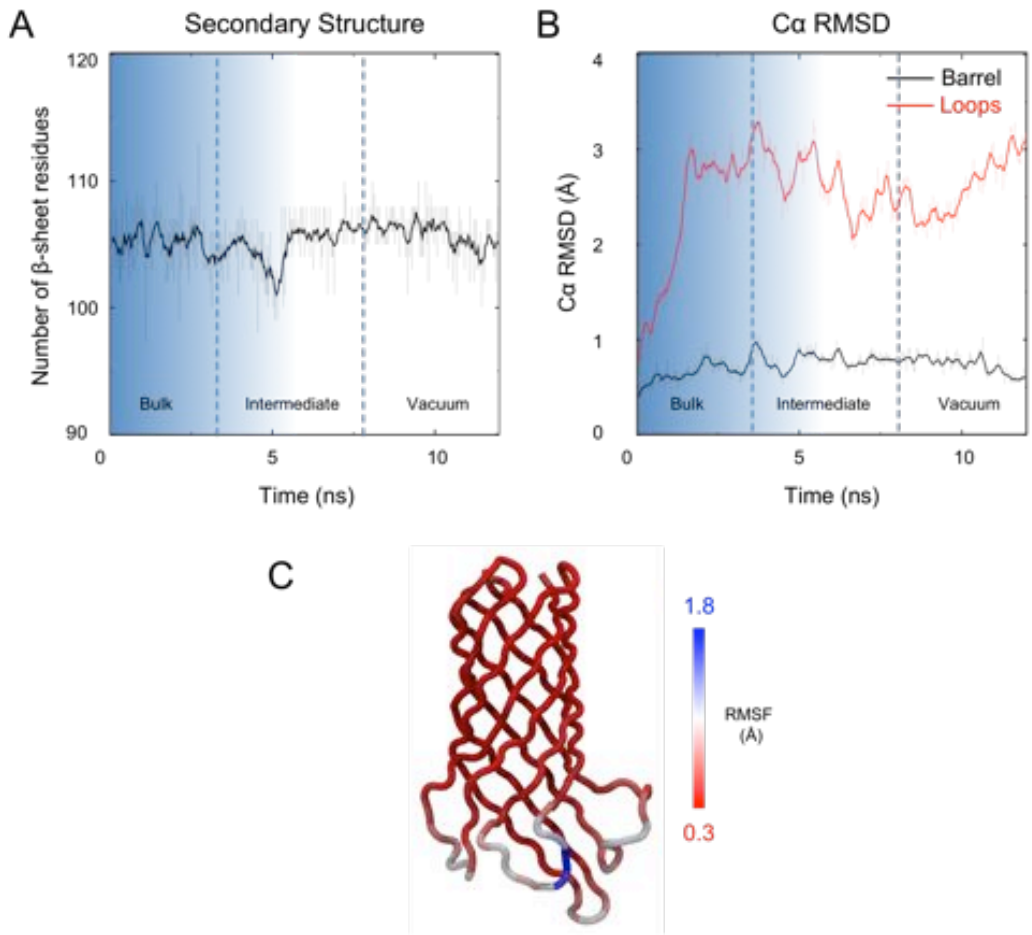


Figure 7:4. OmpA stability during transition from bulk solution to vacuum. A: Number of residues in β-sheet conformation. B: Cα RMSD of the residues initially in β-sheet conformation (black trace) and those in loop, turn or unstructured conformations (red trace). In each case the location of the dashed lines indicates the approximate boundaries at which the protein-detergent complex is first exposed to vacuum and when it is free from bulk solution. Traces correspond to running averages over 10 data points; raw data are shown as a dotted line. C: Cα RMS Fluctuation of OmpA averaged over final 5 ns of steered MD simulation. The backbone of the protein is shown in tube representation, coloured according to RMSF value

Other properties of the protein are shown in Figure 7:5. The number of intra-protein hydrogen bonds does not deviate by a great amount in vacuum or in solution. The overall solvent accessible surface areas (SASAs) of hydrophobic and hydrophilic residues (excluding detergent micelle) for the protein decrease during the transfer. This decrease in surface area appears to be due to the protein conformation altering to become more compact during the transition as a corresponding decrease in the radius of gyration is measured.

The effects of transfer on the packing of the DPC micelle are shown in Figure 7:6 The rearrangement of the DPC molecules to interact with the loop regions is measured by corresponding decreases in both hydrophobic and hydrophilic solvent accessible surface areas (SASA) of the protein (including the detergent micelle).

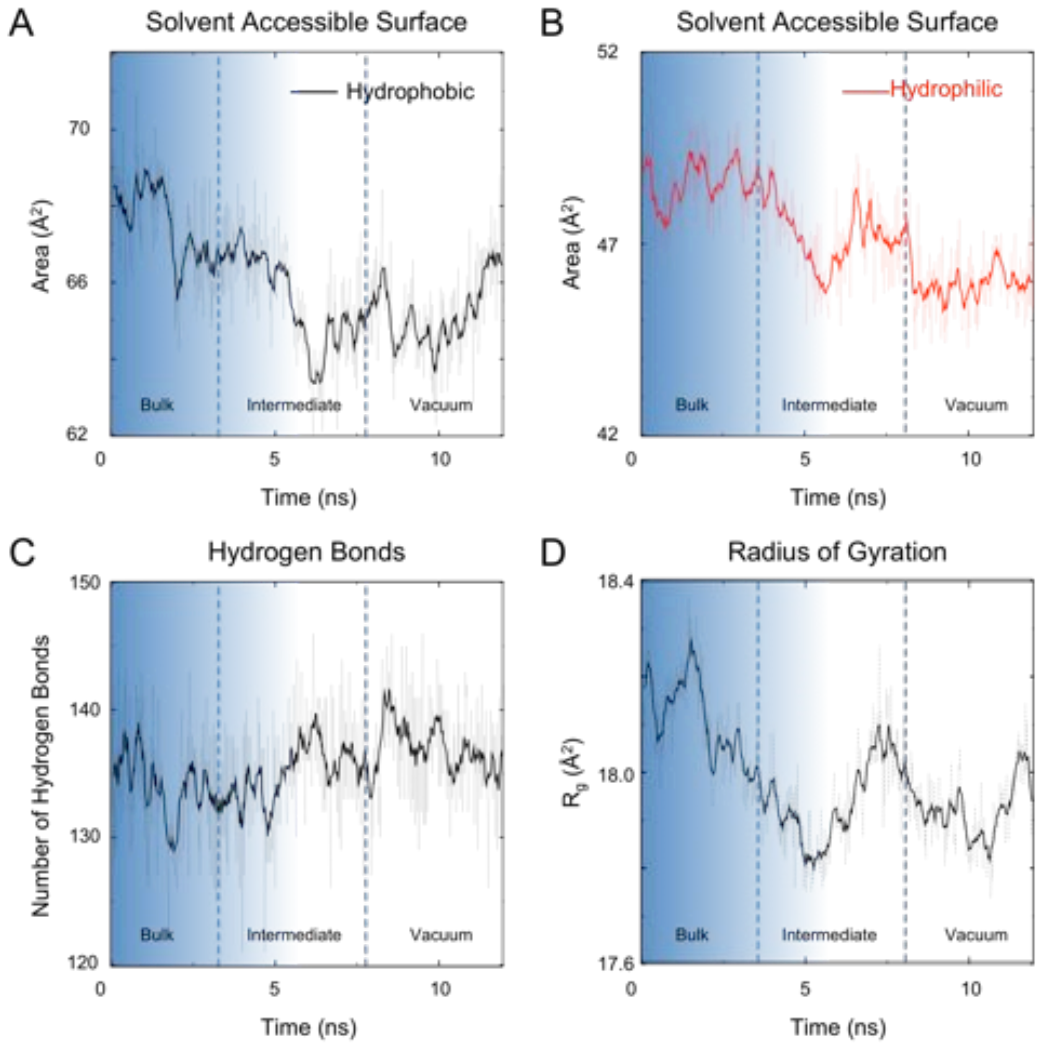


Figure 7:5. Properties of OmpA during steered MD (SMD) simulation. Hydrophobic (A) and hydrophilic (B) solvent accessible surface area of OmpA excluding detergent. C: Hydrogen bonds between protein during SMD simulation. D: Radius of gyration of protein during SMD simulation. In each case the location of the dashed lines indicates the approximate boundaries at which the protein-detergent complex is first exposed to vacuum and when it is free from bulk solution. Traces correspond to running averages over 10 data points; raw data are shown as a dotted line

The reductions in SASA are greater than that observed for the protein alone. Relatively few hydrogen bonds exist between OmpA and DPC headgroups in bulk solution. The rearrangement involves clustering of the zwitterionic headgroups; the maximum size of headgroup clusters increases dramatically following removal from bulk solution. Following transfer the number of protein-detergent hydrogen increases. The clustering and hydrogen bond increase are a result of the DPC headgroups becoming unable to hydrogen bond with bulk solution. The change in DPC micelle conformation from a normal micelle structure (headgroups on the outside, hydrophobic tails sequestered in the core) may be monitored by calculating the instantaneous radius of gyration of the headgroups and acyl tails. Upon transfer to vacuum the 'normal' micelle arrangement is clearly disrupted. The headgroups pack in closer to the centre of the micelle and the tails

are orientated further out. The rearrangement of the acyl tails appears to stabilise by the end of the simulation, whereas the headgroups appear to still be rearranging somewhat.

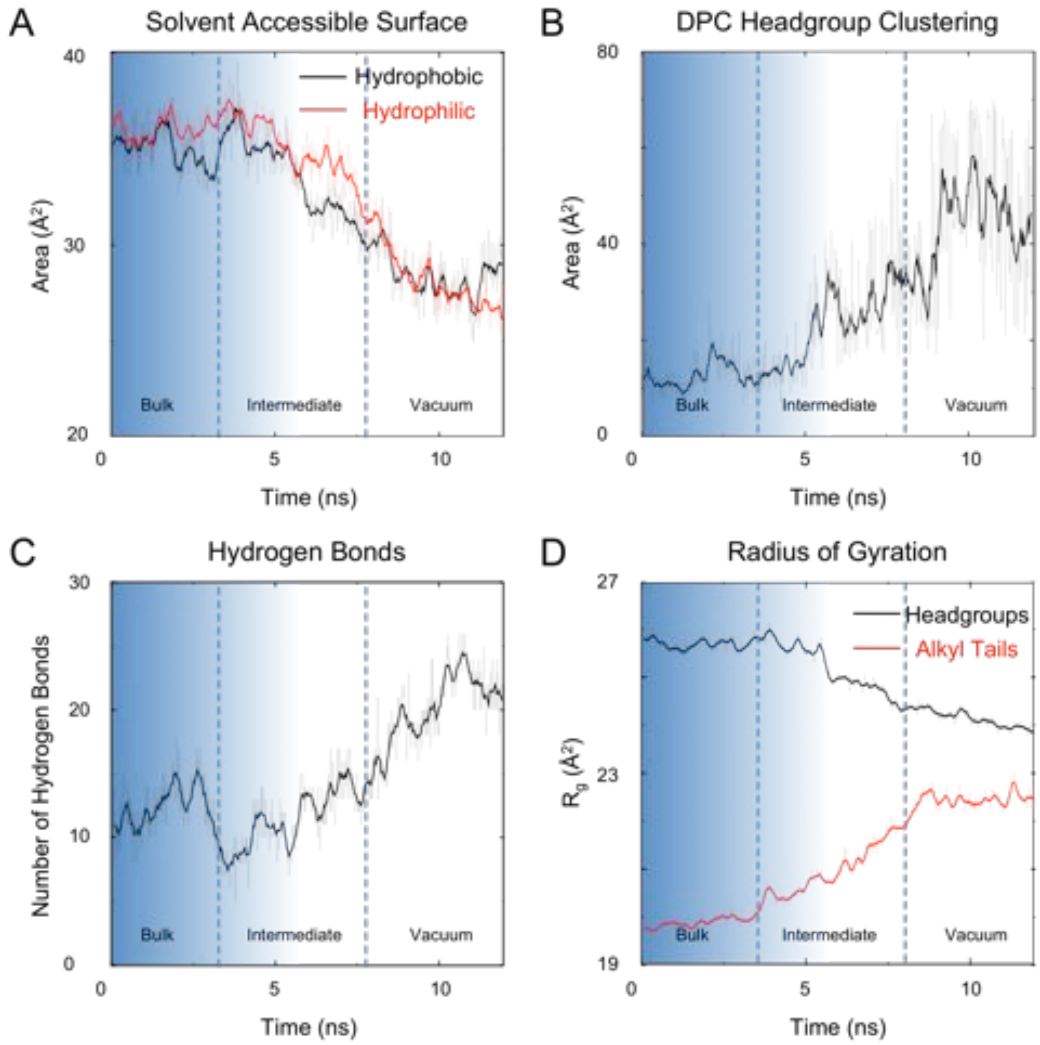


Figure 7:6. Properties of Protein-DPC micelle during transfer from solution to vacuum. A: Protein solvent accessible surface area including DPC micelle. Black trace = hydrophobic residues, red trace = hydrophilic residues. B: Maximum cluster size of DPC headgroups. Headgroups within a cut-off of 3.5 Å of any atom in a given cluster are added to the cluster. C: Protein-DPC hydrogen bonds. D: Instantaneous radii of gyration of DPC phosphocholine headgroups (black trace) and acyl tails (red trace). In each case the location of the dashed lines indicates the approximate boundaries at which the protein-detergent complex is first exposed to vacuum and when it is free from bulk solution.

Figure 7:7 shows the arrangement of DPC molecules about the OmpA protein following transfer to vacuum, highlighting the location of the remaining water molecules. The rearrangement of the DPC micelle to increase headgroup-headgroup interactions is observed. The remaining water molecules are located in the headgroup regions, creating an extended hydrogen bond network over the surface of the detergent complex. The weaker van der Waals interactions between lipid tails and protein are sufficient to align the majority of the DPC tails along the protein surface, such that little of the protein is exposed to vacuum.

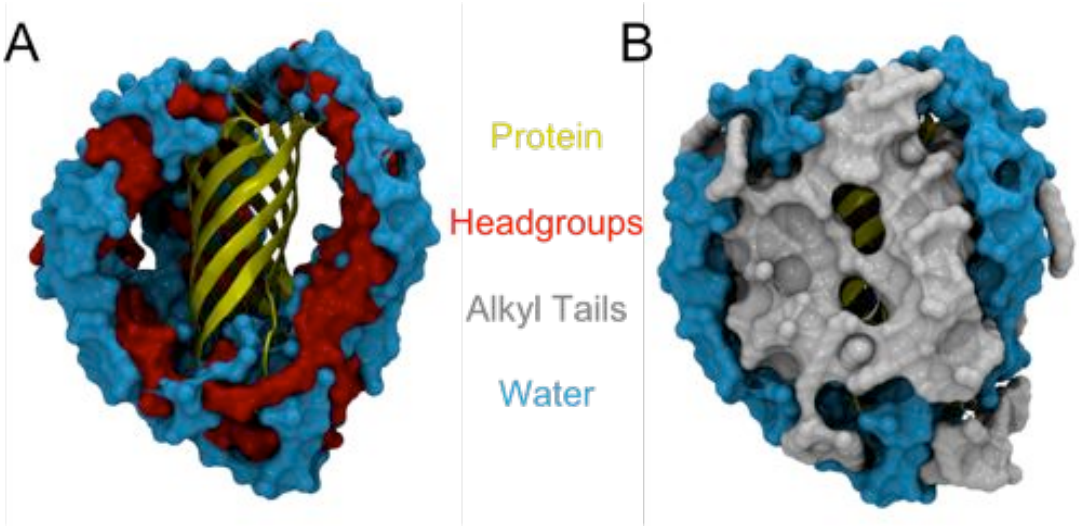


Figure 7:7. Structure of the OmpA-DPC complex following transfer to vacuum environment. In A the DPC headgroups (red) and water molecules within 5 Å of the complex (cyan) are shown as coloured surfaces. In B the DPC acyl tails are shown as a light grey surface. It is clear that the waters are largely distributed amongst the DPC headgroups.

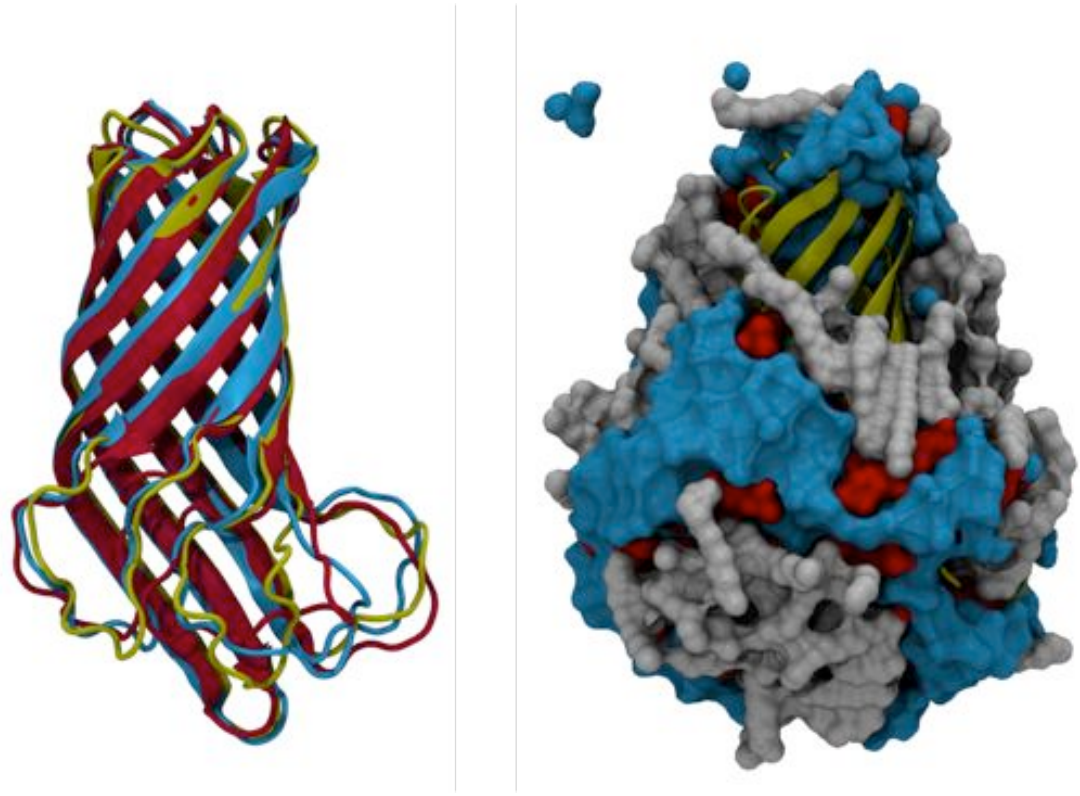


Figure 7:8. The Ompa-DPC complex following extended simulation. Left hand side: The solution phase structure (yellow), structure following steered MD simulation (cyan) and structure following 300 ns simulation (red). Right hand side: the OmpA-DPC complex after 300 ns. The protein is shown in yellow. DPC phosphocholine headgroups are shown as a red surface, and acyl tails as a grey surface. Water molecules are shown as a cyan surface. The protein is shown in cartoon representation in each case.

As 12 ns is a relatively short simulation the main PDC was simulated separately (to allow longer timescales to be reached) for a further 300 ns to evaluate any further changes following the transfer. The protein RMSD over the full 330 ns trajectory was

< 3 Å and reached a plateau within a few ns. The barrel backbone RMSD remained low (<1 Å) and the main changes were to the loop regions of the protein. Snapshots of the protein at the start and end of the 330 ns simulation superimposed on the solution structure are shown in Figure 7:8. The distribution of the detergent molecules is also shown. Over the course of the simulation the headgroups were observed to cluster together more, tending towards the formation of a water-containing inverse micelle. This was at the expense of interactions of the detergent molecules with the protein surface, resulting in more of the protein becoming exposed to vacuum. This simulation was performed using PBC conditions. This resulted in the rate of evaporation being artificially slow, as waters released from the complex are able to contact periodic images. For further study it is therefore necessary to remove the periodic boundary conditions to allow full water evaporation.

3.1.2. BM2 in DHPC and DDM

Three SMD simulations of BM2 in a DHPC micelle ('BM2-DHPC') were performed using different initial velocities. The results of these simulations are compared to a SMD simulation of BM2 in a DDM micelle ('BM2-DDM'). Analysis of the three "BM2-DHPC" simulations led to similar behaviour therefore for clarity only one of these is presented in the comparison plots.

Various properties of the protein during transfer from solution to vacuum are shown in Figure 7:9. The Cα RMSD is higher for the BM2-DHPC complex than for BM2-DDM, where quicker equilibration following the transfer is observed. The number of α-helical residues, the hydrophobic solvent accessible surface area (SASA) and the intra-protein hydrogen bonding appear to change little during the transfer in both the DHPC and DDM micellar environments. The final number of α-helical residues is similar in each, although in one repeat simulation of BM2-DHPC a greater loss in stability ($\sim$ 10 % loss in α-helical content and higher Cα RMSD) is observed. The maximum Cα RMSF during the final 5 ns of simulation is slightly lower in DDM than DHPC (1.6 Å compared to 1.8 Å), consistent with the trends observed in the solution phase fluctuations.

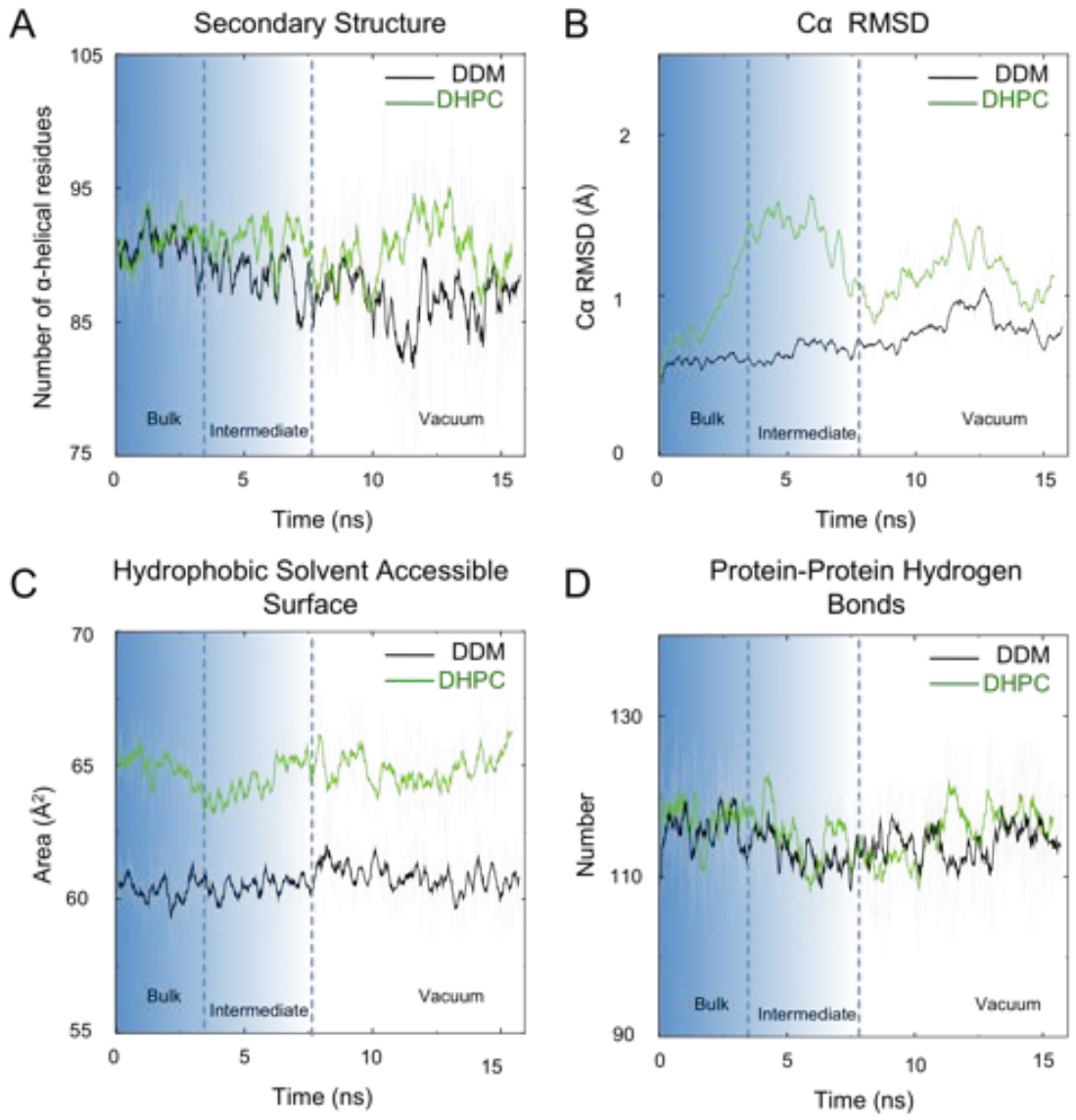


Figure 7:9. BM2 stability in DHPC and DDM micelles upon transfer from solution to vacuum. A: Number of α-helical residues. B: Cα RMSD of all residues except two terminal residues at each end. C: Hydrophobic surface accessible solvent area of protein (excluding contribution from detergent). D: Protein-protein hydrogen bonds. Black traces correspond to data from BM2-DDM simulation. Green traces correspond to DHPC simulations. Dashed lines correspond to the approximate time at which the complex is first exposed to vacuum and when it is fully removed from bulk solution.

Properties of the BM2-DHPC and BM2-DDM PDCs largely dependent on changes to detergent packing are shown in Figure 7:10. The differences between the two micelle types are clearer here. The changes to DHPC packing following transfer to vacuum are more dramatic than those of the protein alone, consistent with the OmpA-DPC results. The changes to SASA including the detergent micelle show clearer trends; the hydrophobic residues tend to become slightly more exposed to vacuum whereas hydrophilic residues (located near the termini of BM2) become less exposed. There is a general increase in the number of protein-DHPC hydrogen bonds (~40 % increase). The detergent micelle appears to lose its 'normal' solution arrangement and the headgroups

pack closer in to the core of the complex, while the DHPC tails become more external. The changes to DDM packing are much less pronounced, with little variation in hydrophobic or hydrophilic SASA and only a slight increase in the number of protein-DDM hydrogen bonds (< 20 %).

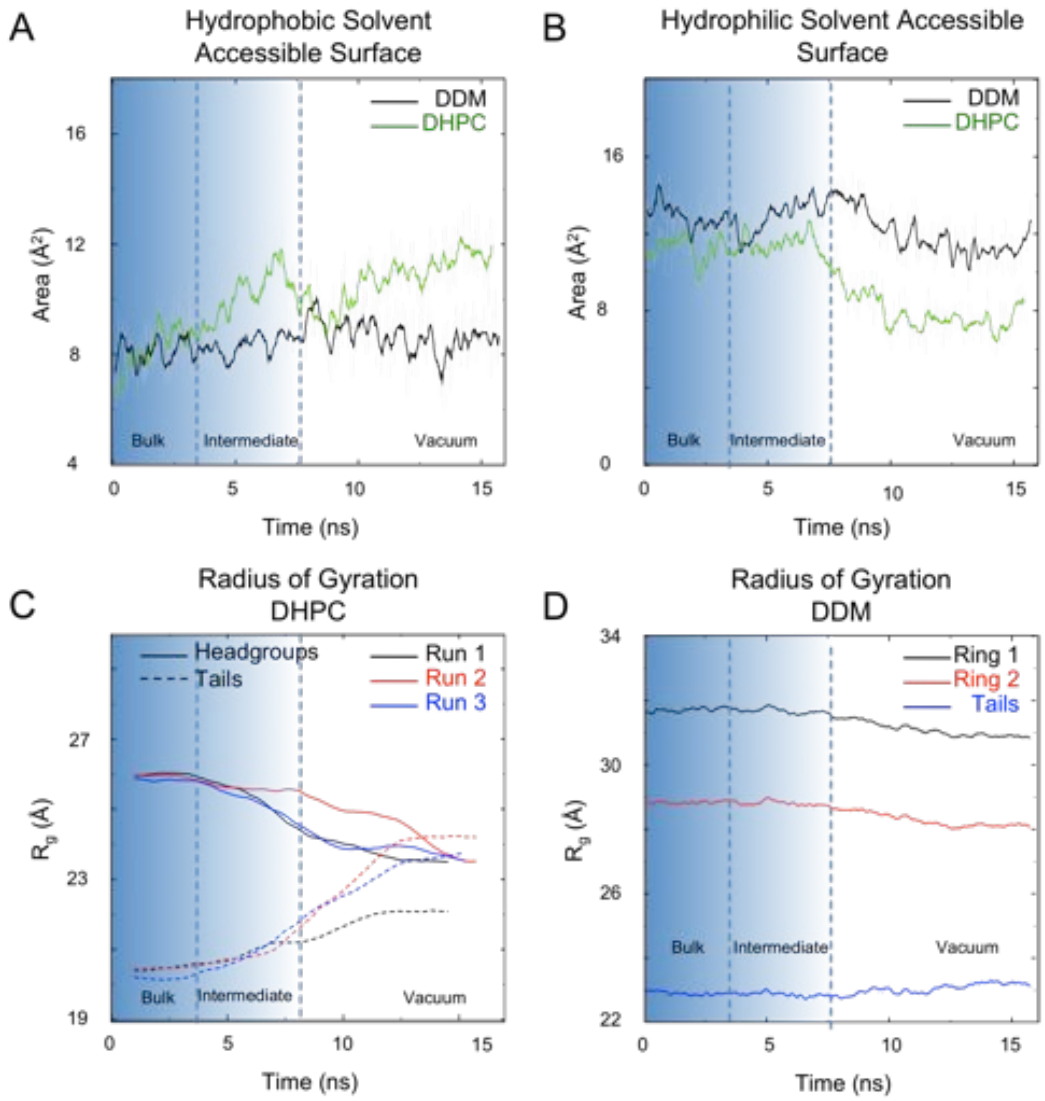


Figure 7:10. Properties of DDM micelle compared to DHPC micelle during transfer from solution to vacuum. Protein solvent accessible surface area including DDM micelle of A: hydrophobic residues and B: hydrophilic residues. C: Instantaneous radii of gyration of DHPC headgroups (solid lines) and tails (dashed lines). D: Instantaneous radii of gyration of DDM headgroups (first glucose ring = black trace, second ring = red trace) and acyl tails (blue trace). In each case the location of the dashed lines indicates the approximate boundaries at which the protein-detergent complex is first exposed to vacuum and when it is free from bulk solution. In A and B, black traces correspond to DDM data, and green traces correspond to a BM2-DHPC simulation

The degree of headgroup clustering in each simulation was measured in terms of maximum cluster size as well as pair-wise minimum distances between headgroups of detergent molecules (data not shown) In the BM2-DHPC simulation there is a general tendency towards clustering of headgroups, with over 30 of the 71 DHPC molecules belonging to a single cluster in each simulation during the transfer (compared to ~5 in

solution). The maximum cluster size corresponds to ~ 50 % of the DHPC molecules belonging to a single cluster (compared to > 75 % in OmpA-DPC). The minimum distances between DHPC headgroups show a clear decrease in vacuum compared to solution. In the BM2-DDM simulation there is very little change to the degree of headgroup clustering upon the transition to vacuum (Figure 7:11). Contrary to DHPC, the DDM headgroups tend to cluster slightly in solution simulations and so the effect of transfer to vacuum appears to be less pronounced, with little observable difference in the solution and vacuum minimum distance matrices (data not shown).

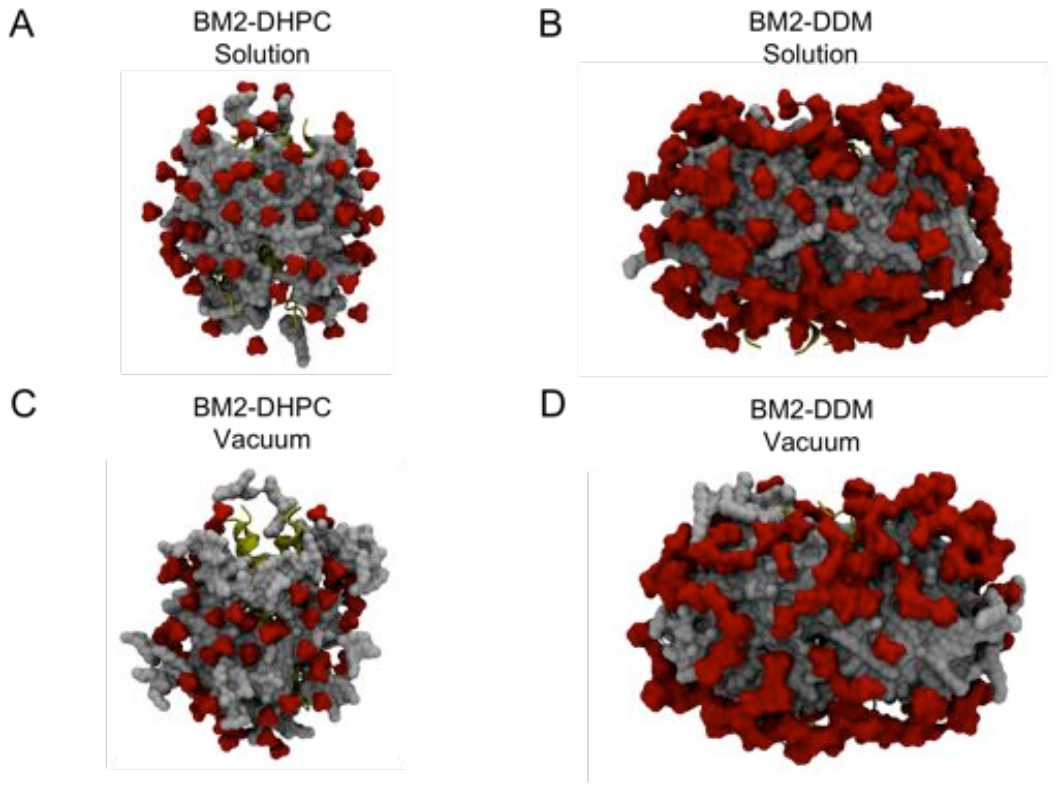


Figure 7:11. Clustering of Detergent Headgroups in Solution and Vacuum. The positions of the headgroups on the micelle surface in solution (A and B) and vacuum (C and D). In each case the headgroups (the choline group for DHPC and the first ring of the maltose headgroup for DDM) are shown as a red surface and the acyl tails as a grey surface.

The final arrangements of detergent molecules and remaining water molecules in the BM2-DHPC and BM2-DDM complexes are shown in (Figure 7:12). In each case the remaining water molecules are largely located within the headgroup regions, maximising the number of detergent-water hydrogen bonds on the PDC surface. In contrast to the OmpA-DPC complex in vacuum, the DHPC tails do not appear to align upon the protein surface but splay outwards. This may be a consequence of geometrical factors, as two chains must be incorporated per DHPC molecule compared to one per

DPC molecule. Another possibility is that this orientation is adopted to allow the glycerol oxygen atoms to partake in hydrogen bonding.

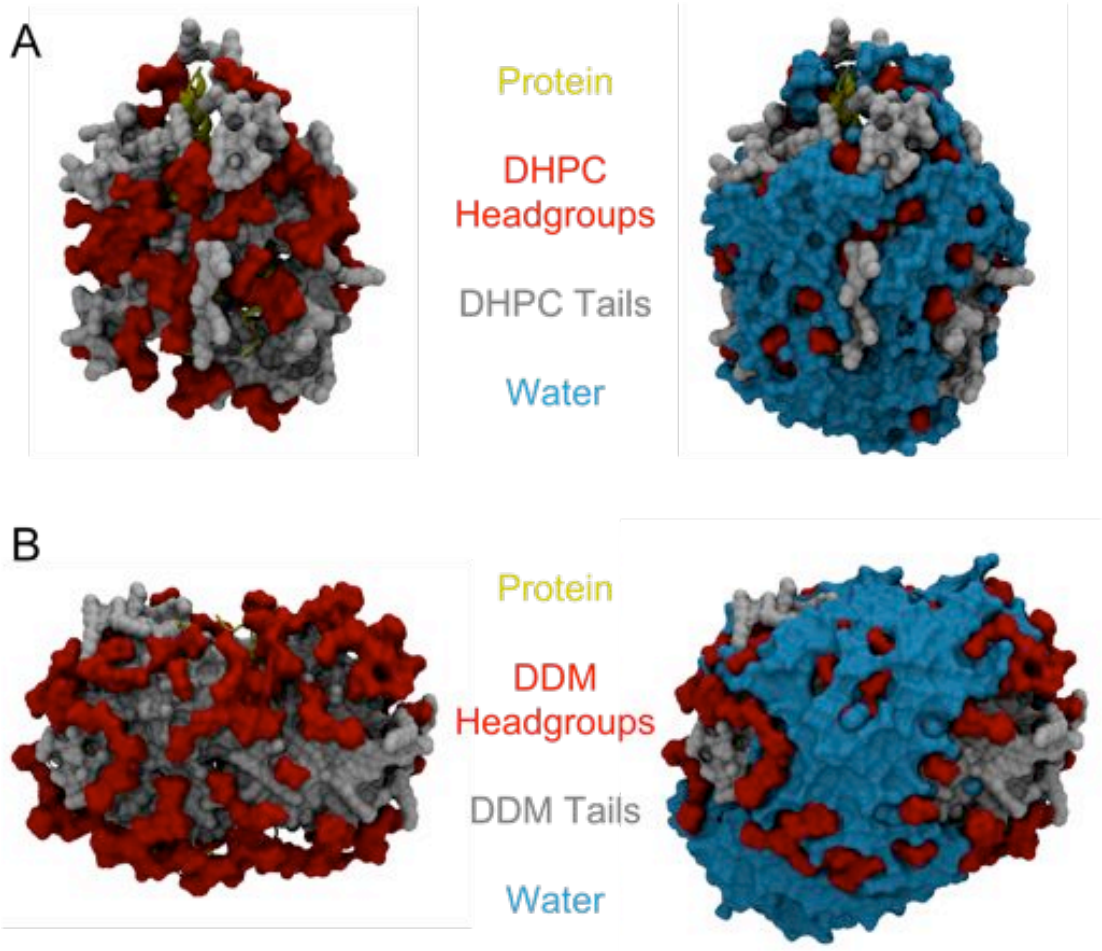


Figure 7:12 Distribution of detergent molecules around BM2 following transfer from solution to vacuum. A: The BM2-DHPC complex. B: The BM2-DDM complex. In both cases the protein is shown in cartoon representation and coloured yellow. The detergent headgroups (for DHPC the phosphatidylcholine group, for DDM the first ring of the maltose headgroup only) are shown as a red surface, and the acyl tails are shown as grey surfaces. In the left hand side figures the water is omitted, while on the right side the water distribution is shown as a cyan surface.

4. Extended Simulation Under Dehydrating Conditions

The results of the extended OmpA-DPC simulation demonstrated that the relatively short (~20 ns) steered MD simulations are sufficient to capture most of the structural changes associated with the removal from bulk solution to give partially solvated complexes. However to evaluate the effects of further dehydration, longer simulations in which water is free to evaporate are necessary. Therefore the BM2-DHPC and BM2-DDM complexes at the end of the steered MD simulations were simulated under dehydrating conditions (see Methods section) for long (> 500 ns) timescales.

4.1. Dehydration of the BM2-DHPC Complex

Figure 7:13 displays the degree of water evaporation from the protein-detergent complex over each simulation. In both simulations the majority of water molecules evaporate within the first 50 ns, with the final number of waters remaining bound within the complex being ~130 and ~35 for BM2-DHPC and BM2-DDM, respectively. The distribution of water molecules within the complex may be determined by measuring the degree of hydrogen bonding of remaining water molecules to protein and detergent molecules (Figure 7:13B). The final number of water-detergent hydrogen bonds is ~75 % less in the BM2-DDM simulation than for BM2-DHPC, whilst the final number of BM2-water hydrogen bonds is similar in each of the simulations.

The effects of dehydration on various properties of BM2 in DHPC and DDM are shown in Figure 7:14. In each simulation the initial rapid water loss results in an initial slight (~5 %) loss in α-helical content with a corresponding increase in Cα RMSD of ~1.5 Å. The protein also becomes more compact with the effect more pronounced for BM2-DHPC (R_g decreases by >2 Å) than BM2-DDM (R_g decreases by ~0.2 Å, data not shown). These properties then appear to stabilize until ~150 ns at which point the protein begins to show differing behaviour in each environment. In DDM, BM2 appears to reach a stable conformation by 200 ns, from which point onwards little change in properties of the protein are observed. In DHPC, more pronounced changes occur. The protein loses secondary structure, coinciding with a conformational change (increase in Cα RMSD, R_g and SASA) to a less compact form. The final snapshot of the protein (Figure 7:14C) shows that the main structural change is a bending of one helix away from ideal geometry.

The effects of dehydration on detergent packing were also monitored. In the BM2-DHPC simulation the degree of headgroup clustering increases rapidly within the first 100 ns of the simulation (such that almost all headgroups belong to a single cluster), coincident with water loss, and reaches a plateau by 300 ns. The increase in DHPC-DHPC interactions is at the expense of protein-DHPC interactions, with an ~35 % increase in exposed protein surface area during the simulation (data not shown). This exposure of the protein is apparent in Figure 7:15. In the BM2-DDM simulation a similar pattern in behaviour is observed where the DDM detergents rearrange following the initial rapid water loss and stabilize within 200 ns. An increase in the degree of headgroup clustering is observed (such that a third of headgroups belong to a single cluster) but the degree of micelle inversion is much less pronounced than for the BM2-DHPC system. The degree of detergent micelle rearrangement may be monitored by the radius of gyration of the

headgroup and acyl tails (see Figure 7:13). These demonstrate that while the DDM micelle relaxes slightly initially there is little rearrangement following further dehydration, while the DHPC micelle tends towards complete inversion.

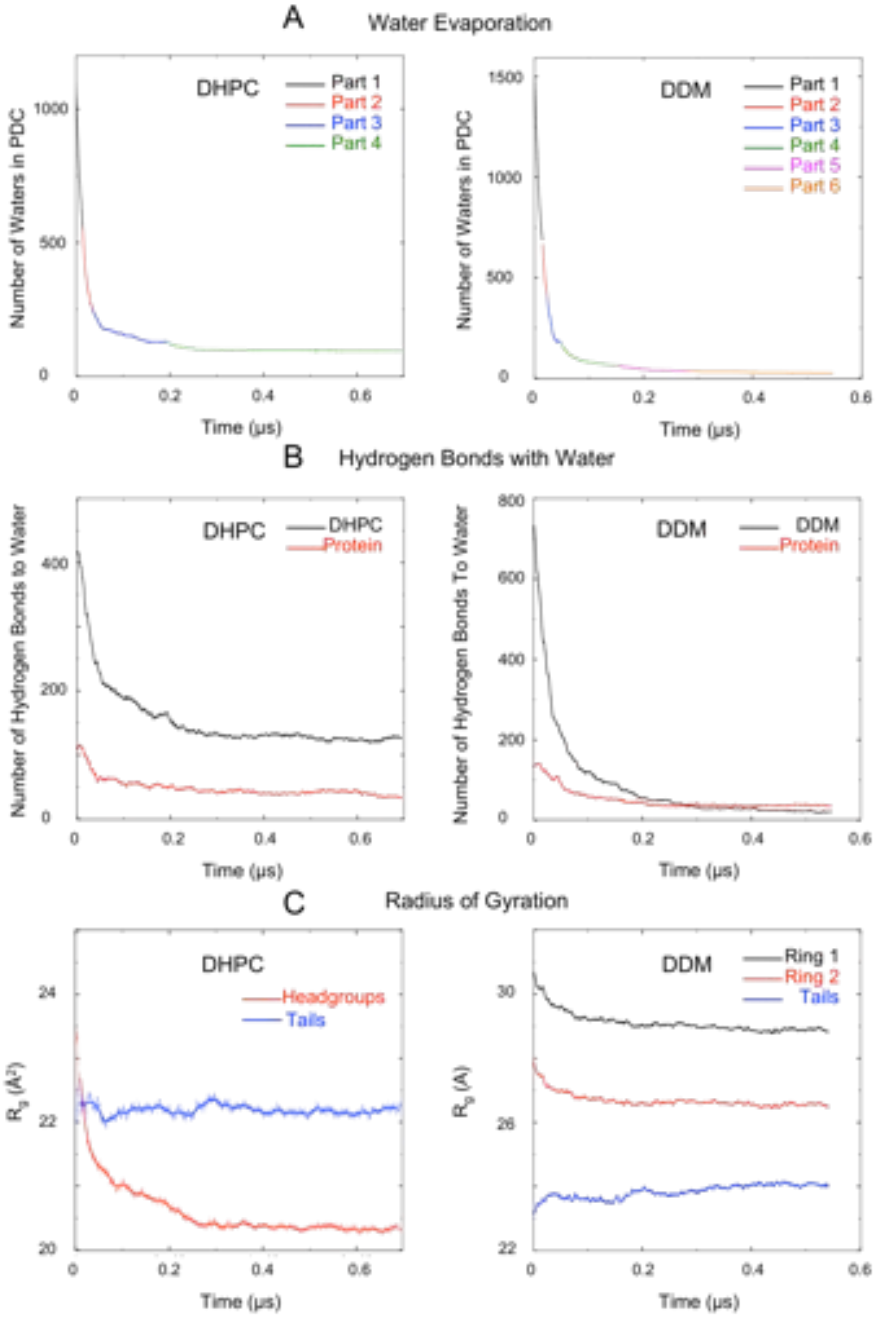


Figure 7:13. Evaporation of water molecules from BM2-DHPC (left) and BM2-DDM (right) complexes during extended simulation under dehydrating conditions. A: The number of waters remaining in the complex over simulation*. B: The hydrogen bonding interactions of water molecules between detergent (black trace) and protein (red trace). C: Radius of gyration for detergent headgroups and tails, calculated as described previously.

*Note that the simulations were restarted at various time points with removal of evaporated waters as indicated in A. This was to ensure that the evaporated water molecules (with increased translational and rotational freedom) were not leading to artificially low kinetic energy of the bound waters (related to the 'flying ice cube' effect [29]). The relatively smooth curve obtained suggests that this was not the case, and the water molecules remaining within the complex from 200 ns onwards remain bound despite having temperatures of 323 K.

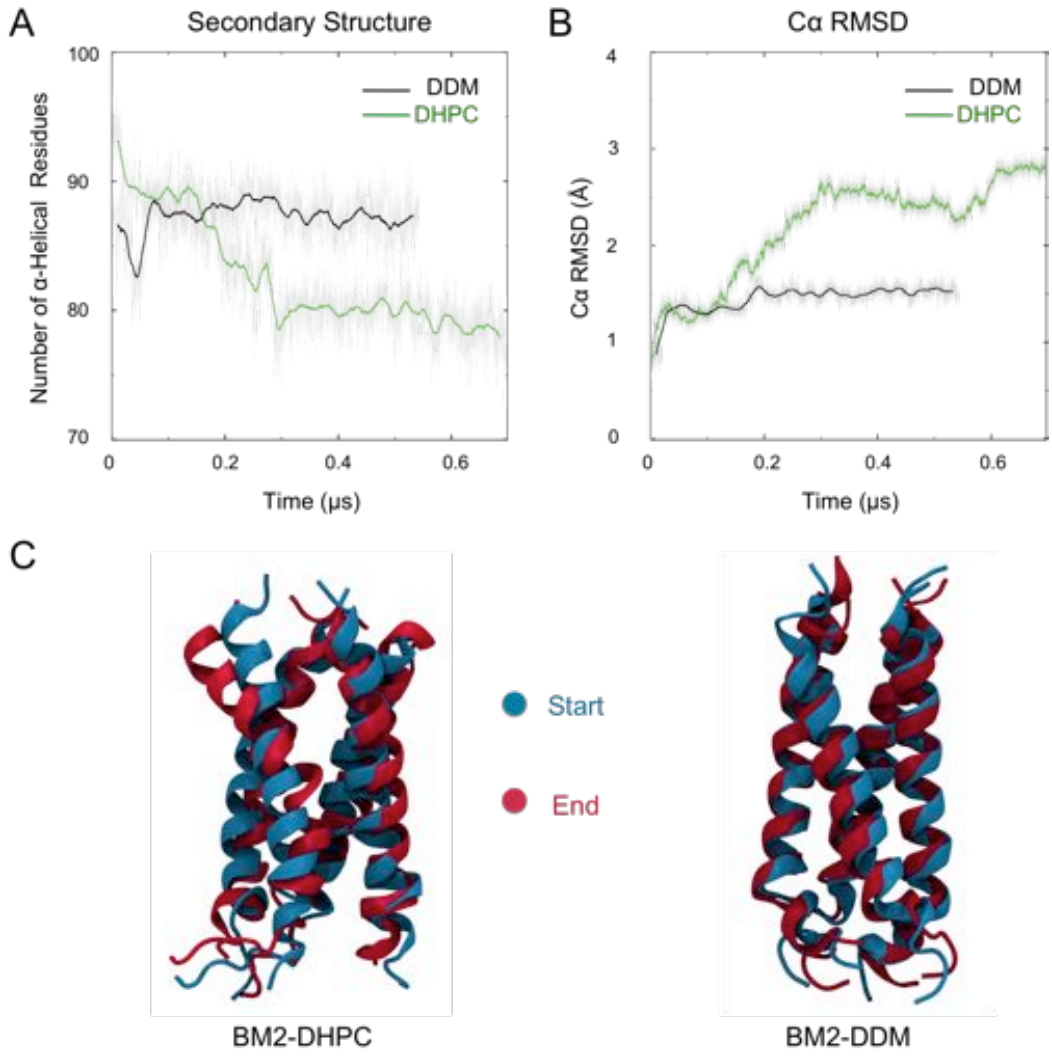


Figure 7:14. Changes to protein properties during extended simulation under vacuum conditions. A: number of α-helical residues. B: Cα RMSD of protein excluding two residues at each terminus. C: Snapshots of the protein at the beginning (cyan) and end (red) of the dehydration simulations. The protein is shown in cartoon representation.

The BM2-DHPC trajectory was inspected to determine the cause of structural changes in the protein following water evaporation. The decrease in solvation of the DHPC headgroups leads to increased electrostatic interactions between the zwitterionic headgroups and the protein residues. Figure 7:16 shows an instance in which this interaction becomes destabilising. The phosphocholine headgroup is able to penetrate between two of the helices and interact with a water molecule in the pore of the protein. This interaction eventually results in extrusion of the water molecule laterally through the protein *via* interaction of the water molecule with the positive choline group followed by transfer to the negative phospho group.

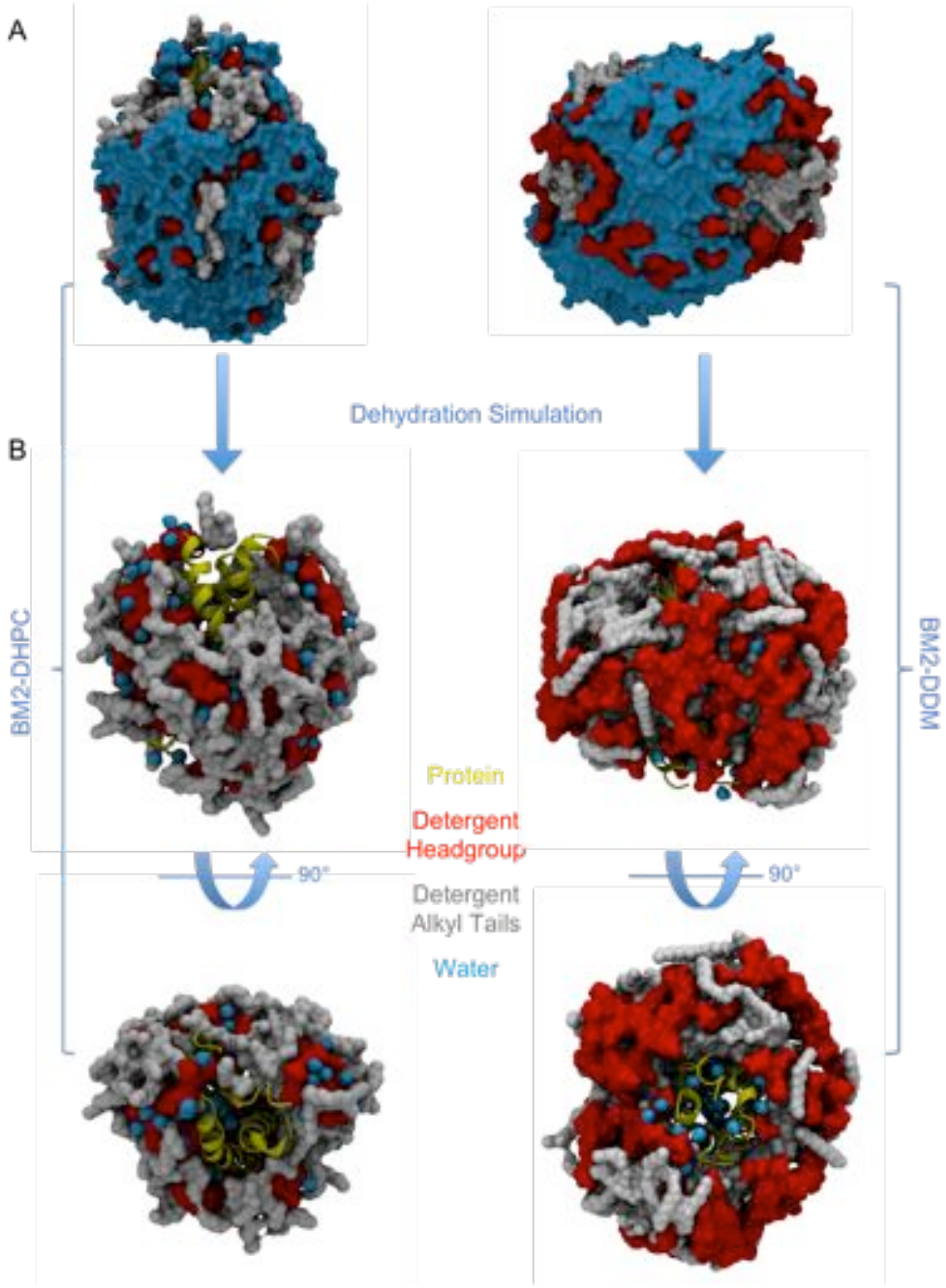


Figure 7:15. Changes to detergent arrangement during extended simulation under dehydrating conditions. A: Starting point of dehydration simulation. B: Protein-Detergent Complexes following simulation under dehydrating conditions. The protein is shown in cartoon representation (yellow). The DHPC molecules are displayed as a surface (red = phosphatidylcholine headgroups, grey = acyl tails). Water molecules are shown as blue space filling spheres.

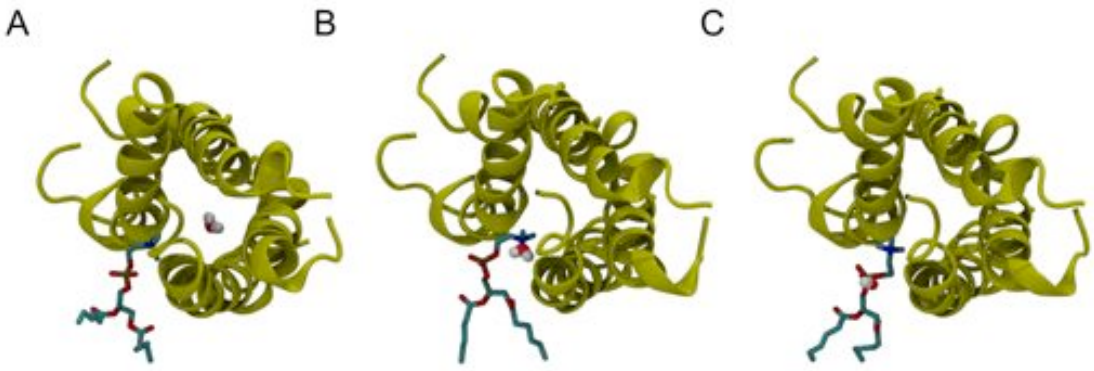


Figure 7:16. Disruption of protein packing by DHPC detergent headgroup. A: For the first 500 ns of simulation the DHPC detergents do not appear to interfere largely with the protein packing. B: Towards the end of the simulation the positively charged choline nitrogen penetrates between two of the helices and interacts with the oxygen of a water molecule within the pore. C: This interaction draws the water sideways through the protein, mediated *via* hydrogen bond transfer to the negatively charged phosphate group. Removal of this water coincides with structural changes to the protein backbone.

4.2. Increasing (Collision) Energy

The BM2-DHPC and BM2-DDM protein-detergent complexes at 150ns of the dehydration simulations were subjected to gradual temperature increase in the range of 323 to 700 K (time 0 = 323, 10ns = 350 K, 20 ns = 400 K, 30 ns = 450 K, 40 ns = 500 K, 50 ns = 550 K, 60 ns = 600 K, 70 ns = 650 K, 80 ns = 700 K). The protein alone (generated from BM2-DHPC system with manual removal of all non-protein entities) was also simulated as a control system.

The protein stability in teach simulation in terms of α-helix content and Cα RMSD is shown in Figure 7:17. It is apparent that the protein begins to lose secondary structure at temperatures approaching 450 K, and above 600 K the protein has lost the majority of its original α-helical secondary structure (although some increase in other structure *e.g.* β-turn character is observed). The decrease in α-helical content is most pronounced for the BM2-DHPC simulation, with almost complete loss above 600 K, suggesting that the presence of DHPC molecules may accelerate the unfolding process in some manner. The Cα RMSD is much higher in the BM2-DHPC and BM2-DDM simulations.

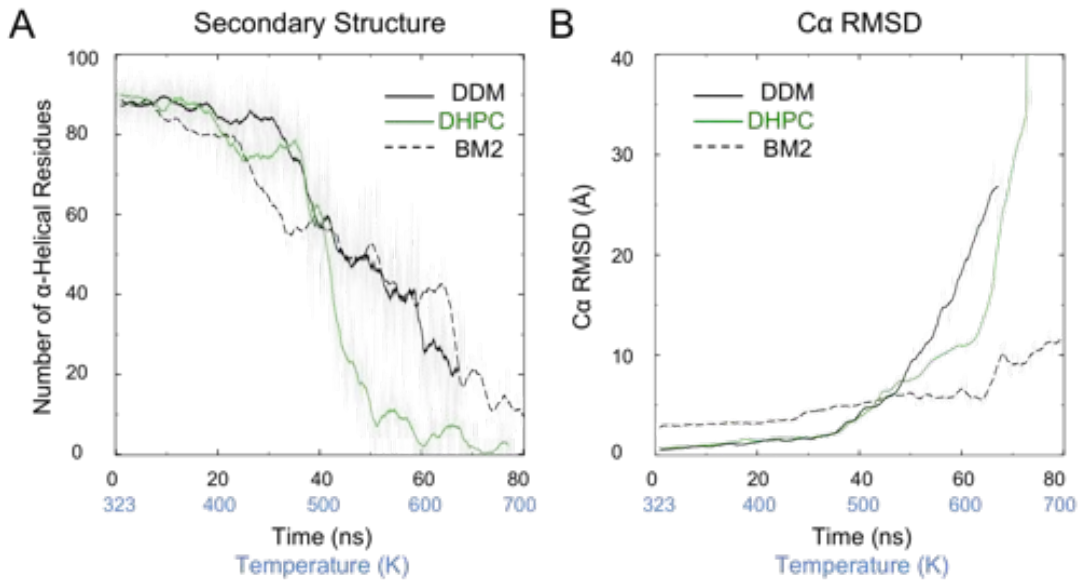


Figure 7:17. Stability of BM2 during increase of system energy. A: Number of residues in α-helix conformation B: The Cα RMSD of all particles. The dashed black trace corresponds to the protein-only simulation. The black trace corresponds to the BM2-DDM simulation. The green line corresponds to the BM2-DHPC simulation.

In Figure 7:18, snapshots of the unfolding process in each simulation are shown. In the protein-only simulation the four helices remain in contact, whereas the simulations of BM2 in the presence of detergent show different behaviour, as may be expected from the higher in Cα RMSD for these simulations. In the BM2-DHPC and BM2-DDM simulations one of the helices becomes separated from the rest of the bundle. In the BM2-DHPC simulation this occurs by a 'translation' of one of the helices; this helix elongates and moves such that the N-terminal residue interactions with the C-terminal residues of the trimer *via* electrostatic interaction. Above a threshold temperature the monomer completely detaches from the trimer, taking some DHPC molecules with it. This monomer then collapses in and refolds to a new structure. In the BM2-DDM simulation the initial unfolding occurs via 'splaying' of the helices at the N-termini, such that they bend to align along the surface of the detergent, reducing electrostatic repulsion between the charged terminal residues and allowing formation of BM2-DDM hydrogen bonds. As the temperature is increased further one of the helices begins to dissociate *via* a 'bending' mechanism. The helix continues to splay outwards until it is rotated 180° with the C-terminus remaining in contact with the C-terminal lysine and arginine residues of the three other helices. No complete separation of protein or detergent groups is observed in this simulation.

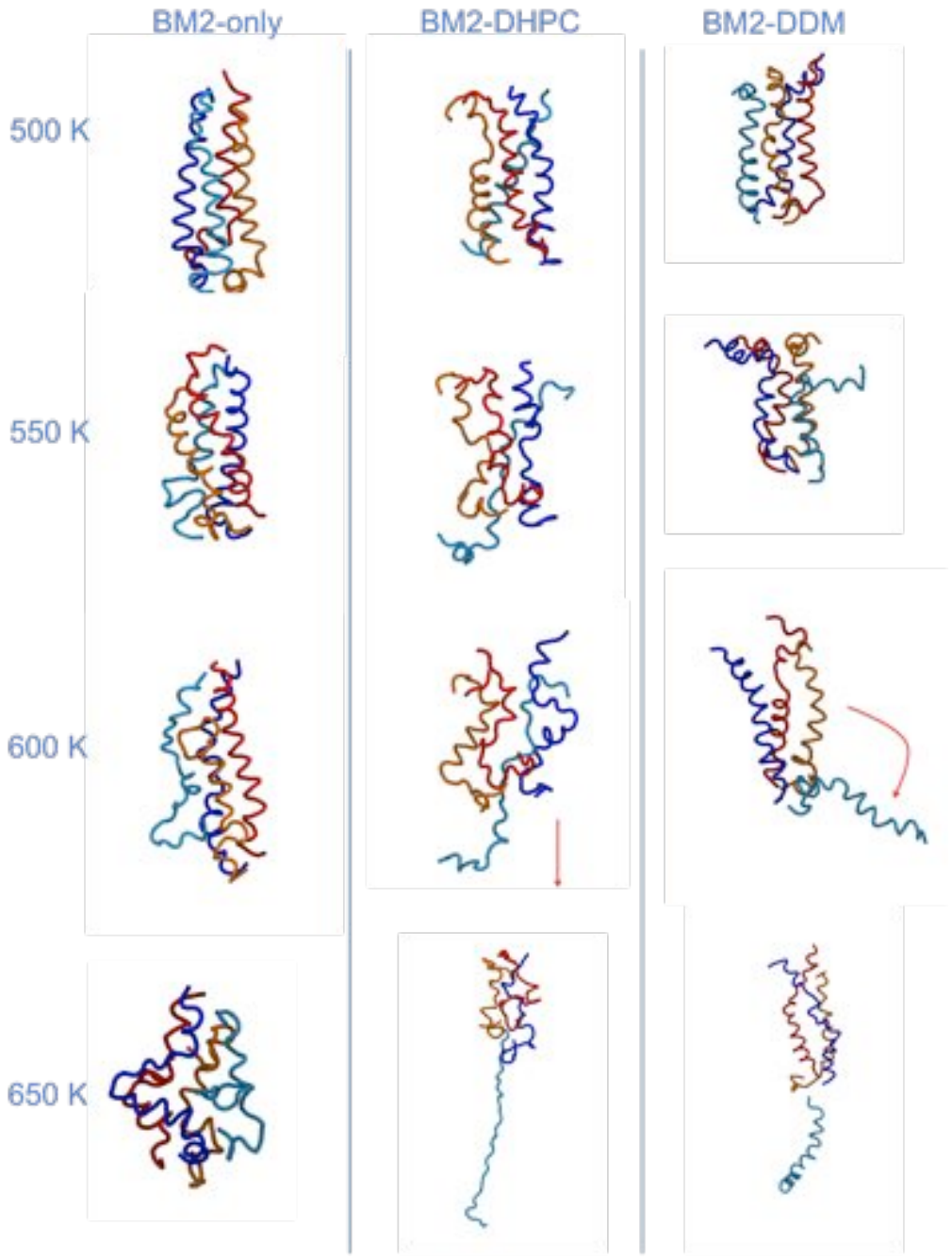


Figure 7:18. Unfolding of BM2 at increasing temperatures in the protein-only, BM2-DHPC and BM2-DDM simulations as indicated. The protein backbone is shown in tube representation and is coloured according to chain. The temperature of the system in each snapshot is indicated.

The changes to DDM arrangement during the heating process are shown in Figure 7:19. At temperatures approaching 500 K, coincident with when the protein begins to unfold, the headgroups begin to cluster and move in closer to the core of the PDC (indicated by the decrease in R_g of the headgroups) whilst the detergent tails become exposed to vacuum (increase in R_g of tails). The rearrangement process of the DDM micelle is therefore a much slower process than that of the DHPC micelle and is only observed at elevated temperatures in these simulations. The packing of the detergent micelles

following the transition towards inverse micelle formation is shown in Figure 7:19. It is apparent that rearrangement occurs at the expense of increased exposure of the protein surface to vacuum.

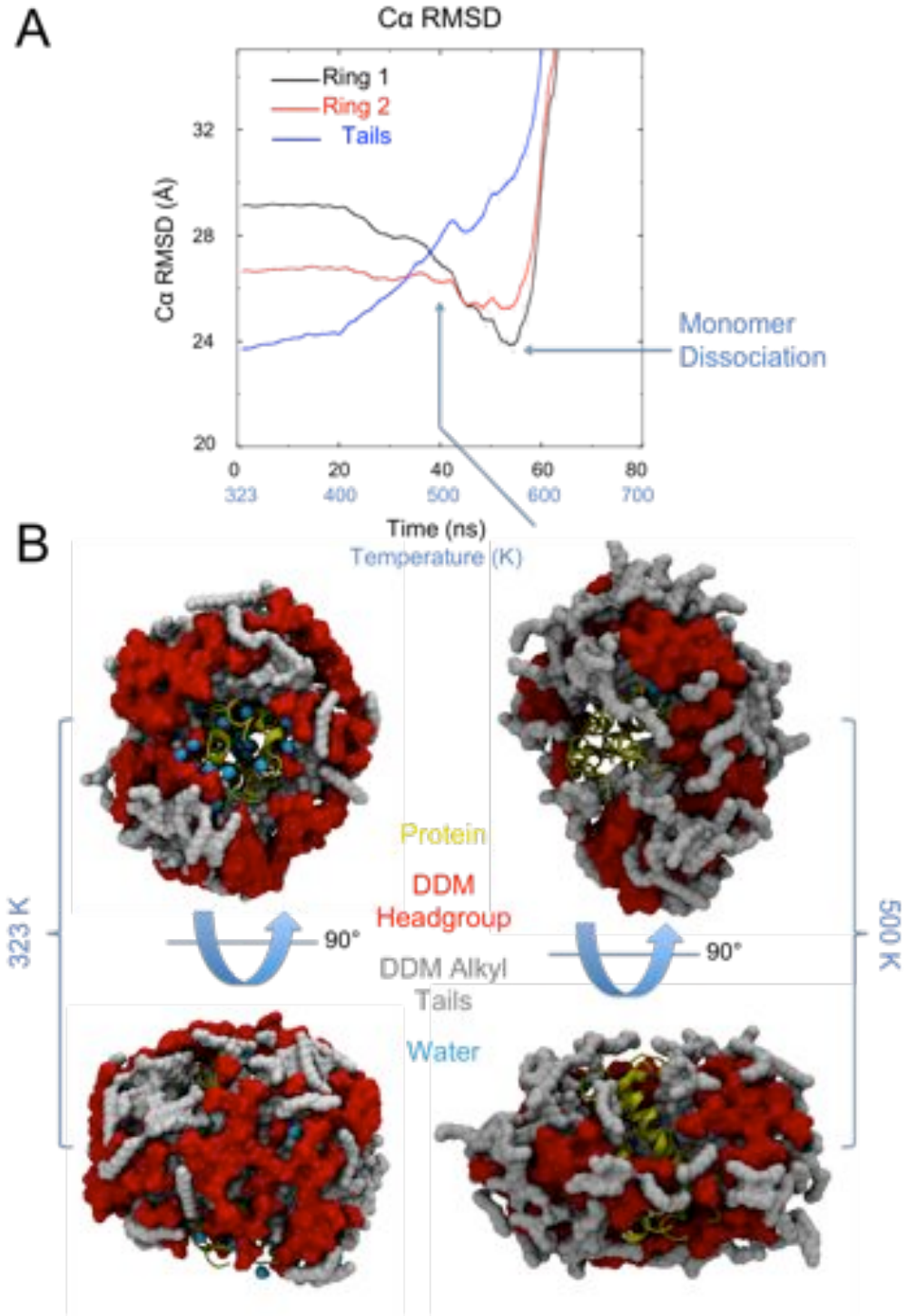


Figure 7:19. Rearrangement of DDM molecules upon temperature increase. A Instantaneous radii of gyration of the detergent micelle over time. The corresponding temperatures at each time point are indicated on the axis labels. B: Snapshots of the protein-detergent complex at 0 ns (323 K, left) and 40 ns (500 K, right) are shown. The view from the side of the complex at 500 K highlights the exposure of the protein surface to vacuum The protein is shown in cartoon representation and coloured yellow. The first glucose ring of each DDM headgroup is shown as a red surface. Acyl tails are shown as a grey surface. Water molecules are shown as space filling spheres and coloured cyan.

5. Discussion

5.1. Methods used

To study the process of ESI-MS, it was necessary for some approximations to be made, and certain limitations exist in the methodology used in this Chapter. For example, the protein-detergent complex was removed from bulk solution by applying a pulling force, whereas in ESI experiments the droplets are formed due to electric field distortion of the solution meniscus overcoming surface tension. However, the outcome of the steered MD simulations is a protein-detergent complex with only those waters interacting strongly enough to be removed from bulk solution without any force applied *i.e*. only the most strongly-interacting water molecules are retained. In the absence of the computational resources that would be required to simulate the full process (millisecond timescales) of ion expulsion due to increasing surface charge, this approach seems like a reasonable method to generate a realistic partially solvated complex.

Another aspect of full protein dehydration that cannot as yet be efficiently captured using classical mechanics is the change in charge states of surface residues. It is known that the pKas of amino acids in the gas phase are different to that in solution phase, and that protonation of basic sites occurs in the final stages of dehydration. In this work it was decided to retain the solution state charges throughout. An efficient way to incorporate proton transfer reactions into MD simulations is yet to be published. Fortunately, a recent method has been developed [30] which incorporates efficient proton transfer reactions into classical MD simulations. This will allow a more complete description of protein dehydration during ESI.

Despite the above-mentioned limitations, the main aims of the work were to study the effects of transfer of a protein-detergent complex from solution to vacuum. The results presented above allow a comparison of the two structural architectures of membrane proteins (α-helical and β-barrel) in two types of detergent micelles.

5.2. Transition from Solution to Vacuum

The results of the steered MD simulations show clear differences between complexes of proteins with zwitterionic detergents (DHPC and DPC) compared to non-ionic detergents (DDM). The zwitterionic detergents are shown to rearrange more during the transition, tending towards inverse micelle formation, at the expense of protein-

detergent interactions. In contrast, the BM2-DDM complex largely maintains its solution phase architecture following the transition.

There appears to be a correlation between the drift of the protein solution-structure and the degree of micelle disruption upon transition from solution to vacuum. Indeed, Figure 7:20 shows a clear correlation between detergent disruption (as measured by the difference between SASA of the protein including detergent) and the RMSD of the solution structure during transfer to vacuum. Note that the SASA change in the OmpA-DPC system is of the opposite sign (decrease in SASA of protein rather than increase) to the BM2 simulations yet despite this the agreement to a linear fit of the data is good.

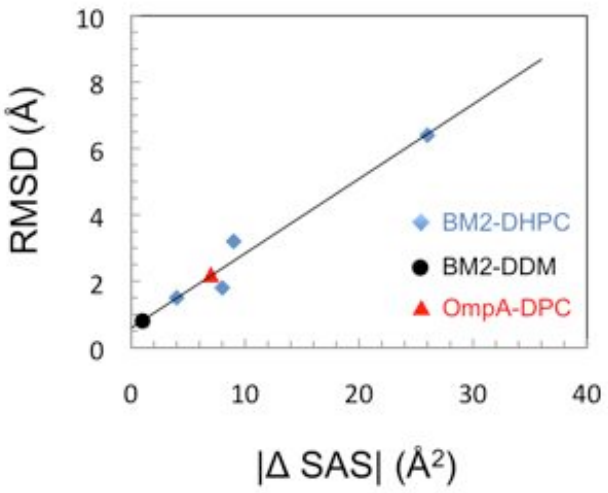


Figure 7:20. Correlation between disruption in micelle upon transfer from solution to vacuum and instability of protein solution structure. ΔSAS refers to the magnitude of the change in protein hydrophobic solvent accessible surface area (including the presence of detergent) of the complex in vacuum compared to solution. The BM2-DHPC points are from three simulations repeated with different initial velocities. The BM2-DHPC highest value corresponds to the simulation performed where only the protein was pulled from solution (see Appendix A).

The results of these simulations suggest that when planning an ESI-MS experiment the choice of detergent is important. Choice of a detergent whose solution- and gas-phase micelles are expected to be most similar should maximise chances of stabilising the protein solution structure.

5.3. Water evaporation

As was found previously, a thin water layer was sufficient to largely maintain the structure of the protein, with most structural changes confined to the detergent packing. Comparison of dehydration patterns of the BM2-DHPC and BM2-DDM upon extended simulations demonstrated contrasting behaviour in each. In the case of DHPC the number of water molecules remaining at the end of the simulation was approximately four times greater than that in the DDM micelle (despite there being more than twice as

many DDM molecules per DHPC and a correspondingly higher volume occupied by detergent headgroups). Figure 7:21 shows the dehydration rates of each system, scaled according to the number of detergent molecules. Despite the greater solvation of the DHPC complex, the reduction in water content to ~2 per detergent molecule was found to be sufficient to cause destabilising interactions of DHPC headgroups with the protein.

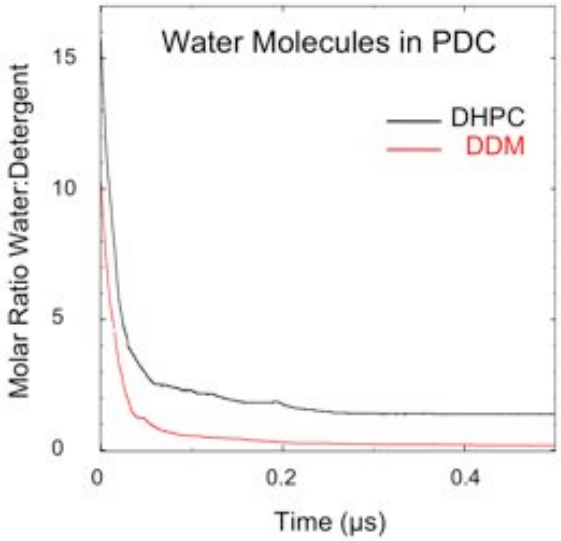


Figure 7:21. Comparison of dehydration rates of the BM2-DHPC complex (black trace) compared to the BM2-DDM complex. The number of water molecules remaining has been scaled in terms of the number of detergent molecules in each complex.

The lack of screening of electrostatic charges by water molecules in the gas phase therefore appears to lead to destabilising interactions of zwitterionic detergent headgroups with polar residues of the protein. This was not reported in the simulation of OmpA in a DPC micelle, suggesting that the monomeric β-barrel is less prone to destabilisation by detergent interactions. However it is also possible that *i)* the degree of water remaining was sufficient to screen the charges or *ii)* the shorter timescales and lower temperatures were not sufficient to observe this effect.

5.4. Increasing the "Collision" Energy

Upon increasing the temperature BM2 was observed to unfold *via* a monomer dissociation pathway in both the DHPC and DDM environments, although the mechanism of dissociation varied in each case. The presence of detergent appeared to facilitate this pathway. Once full dissociation occurred in the DHPC system the individual protein trimer and monomer were observed to refold, forming new hydrogen bonds (refolded monomer structure is shown in Figure 7:22). This is in line with other MD and experimental studies and provides further indication that following complete dehydration the protein may refold and form new hydrogen bond interactions.

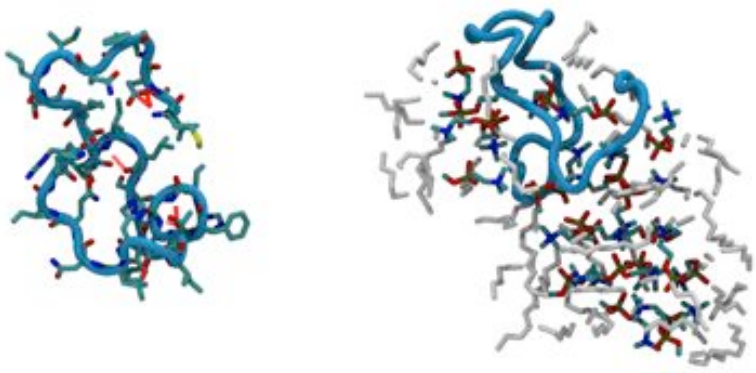


Figure 7:22. Monomer refolding in the dissociated BM2-DHPC system. Left: The protein backbone is shown in tube representation and coloured cyan. Sidechain atoms are shown explicitly. Hydrogen bonds between sidechains are shown as red dashed lines. Right: A different orientation of the snapshot shown in the left hand side panel, including the location of detergent molecules. DHPC acyl tails are shown as grey sticks. Atoms shown explicitly are coloured according to atom type (cyan = oxygen, red = oxygen, blue = nitrogen, tan = phosphorus, yellow = sulfur).

The protein appeared to retain more of its secondary structure characteristics in the DDM system, however it is difficult to draw strong conclusions from a single simulation, especially when high temperatures are involved. Although heating the system is a common approach to studying protein unfolding pathways, to replicate the true behaviour of complex dissociation during ESI a collision-based approach may be more appropriate. This is discussed in the following section.

6. Further Work

The term "collision energy" has been alluded to in this thesis already. In this chapter, the effects of increasing the cell collision energy were mimicked by increasing the overall temperature of the system. This was done by temperature-coupling each of the components to a heat bath. In effect, this guarantees that the protein will unfold whatever the surrounding environment, since the protein has an internal energy of 700 K, well above the melting point. Indeed it is somewhat surprising that any difference in stability of BM2 is observed at all in DHPC micelles compared to DDM micelles. It should also be borne in mind that the force fields used in this thesis are parameterised for ambient temperatures.

To reproduce the true environment of the protein-detergent complex as it travels through a collision cell, an approach involving physical collisions between the complex (as it moves at speed through the simulation box) with neutral gas particles would be more appropriate. This can be achieved by using steered MD to pull the protein-detergent complex through a box containing gas particles. In practice, analytes may

travel in the collision cell at a rate of several kilometres a second (over a distance of tens of centimetres).

The collision-induced dissociation of detergent molecules from a protein *via* inelastic collisions is an intrinsically stochastic process therefore a statistical approach would be preferred. To achieve this, a coarse-grained (CG) system allowing greater distances and more collisions, was set up. This contained the dehydrated BM2-DHPC complex and non-polar particles corresponding to a collision gas (e.g. argon or SF6). Steered MD was used to pull this complex along a box of length 1000 nm containing 200 gas particles. Collision with these particles was shown to dislodge individual lipid molecules from the complex (Figure 7:23). Reassuringly the detergent molecules were shown to rearrange within a few ns to give a partially inverted micelle, comparable to the ATMD simulations performed in this Chapter.

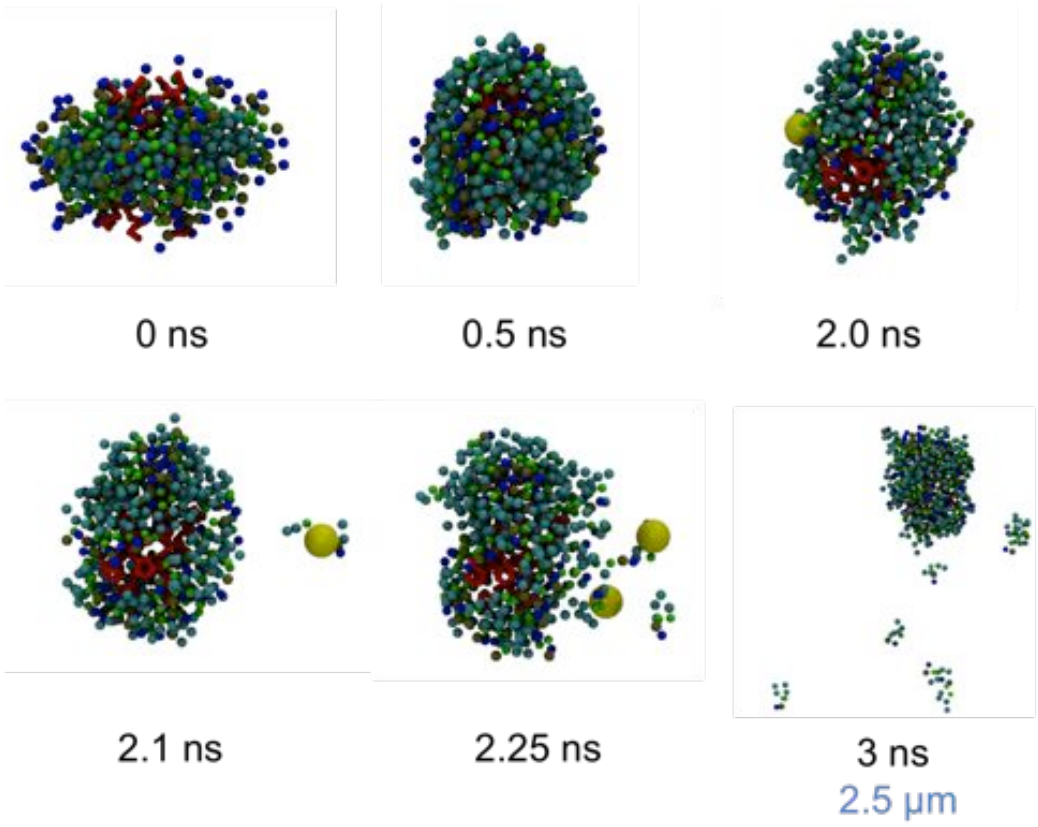


Figure 7:23. Bombardment of BM2-DHPC complex by nonpolar gas particles. The protein backbone particles are shown as a red trace. DHPC lipids are shown in space-filling representation. Non-polar gas particles are shown as yellow space-filling spheres (Dimensions increased for clarity).

These proof-of-principle simulations suggest that the process of collision-based dissociation may be studied using CGMD. The collision rates can be calibrated according to collision cross sections acquired by ESI-MS. This would be an important step towards describing the complete ESI process.

7. Summary

The main question addressed at the outset of this Chapter was "how do micelles protect membrane protein complexes from solution to vacuum?". This work provides further evidence that the major structural changes may be expected to occur in the detergent packing rather than to the protein itself: a minimal water layer was shown to protect the solution structure with any type of detergent.

Simulation under dehydrating conditions led to a clear distinction between the two classes of detergent (zwitterionic and non-ionic) investigated. Upon dehydration the zwitterionic DHPC detergent molecules appeared to cause changes to the protein packing *via* electrostatic interactions. Different behaviour was observed upon almost complete dehydration of the protein in the non-ionic DDM micelle where the solution phase packing was maintained. Comparison of multiple protein-detergent combinations suggests that the suitability of a detergent for ESI studies may be predicted from the degree of rearrangement the micelle undergoes during transfer from solution to vacuum

The effect of dehydration on solution structure stability would be expected to be most important for oligomeric protein complexes. Previous simulations of the monomeric β-barrel OmpA with zwitterionic detergent did not lead to changes in packing, providing a further example of the different properties of the two structural types of membrane proteins.

The results of this Chapter demonstrate that care should be taken in extending conclusions from one system to another. For example, the three protein-detergent complexes investigated in this Chapter when taken on their own would lead to three different conclusions of the behaviour of protein-detergent complexes in vacuum:

1) OmpA-DPC: Stable protein; detergent micelle rearranges
2) BM2-DHPC: Instable protein, detergent micelle rearranges
3) BM2-DDM: Stable protein, detergent micelle does not rearrange.

It is only by combining all three that the overall picture is understood.

8. References

[1] F. McLafferty, “Tandem mass spectrometry,” *Science*, vol. 214, no. 4518, pp. 280-287, Oct. 1981.

[2] J. A. Loo, “Studying noncovalent protein complexes by electrospray ionization mass spectrometry.,” *Mass spectrometry reviews*, vol. 16, no. 1, pp. 1-23.

[3] A.-C. Gingras, M. Gstaiger, B. Raught, and R. Aebersold, “Analysis of protein complexes using mass spectrometry.,” *Nature reviews. Molecular cell biology*, vol. 8, no. 8, pp. 645-54, Aug. 2007.

[4] J. Fenn, M. Mann, C. Meng, S. Wong, and C. Whitehouse, “Electrospray ionization for mass spectrometry of large biomolecules,” *Science*, vol. 246, no. 4926, pp. 64-71, Oct. 1989.

[5] F. Hillenkamp, “Matrix-assisted laser desorption/ionisation, an experience,” *International Journal of Mass Spectrometry*, vol. 200, no. 1-3, pp. 71-77, Dec. 2000.

[6] B. Domon and R. Aebersold, “Mass spectrometry and protein analysis.,” *Science (New York, N.Y.)*, vol. 312, no. 5771, pp. 212-7, Apr. 2006.

[7] H. Hernández and C. V. Robinson, “Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry.,” *Nature protocols*, vol. 2, no. 3, pp. 715-26, Jan. 2007.

[8] G. Ben-Nissan and M. Sharon, “Capturing protein structural kinetics by mass spectrometry.,” *Chemical Society reviews*, vol. 40, no. 7, pp. 3627-37, Jul. 2011.

[9] V. Katta and B. T. Chait, “Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry,” *Journal of the American Chemical Society*, vol. 113, no. 22, pp. 8534-8535, Oct. 1991.

[10] C. S. Ho et al., “Electrospray ionisation mass spectrometry: principles and clinical applications.,” *The Clinical biochemist. Reviews / Australian Association of Clinical Biochemists*, vol. 24, no. 1, pp. 3-12, Jan. 2003.

[11] P. Kebarle and U. H. Verkerk, “Electrospray: from ions in solution to ions in the gas phase, what we know now.,” *Mass spectrometry reviews*, vol. 28, no. 6, pp. 898-917.

[12] K. Breuker and F. W. McLafferty, “Stepwise evolution of protein native structure with electrospray into the gas phase, 10(-12) to 10(2) s.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 47, pp. 18145-52, Nov. 2008.

[13] B. Bothner and G. Siuzdak, “Electrospray ionization of a whole virus: analyzing mass, structure, and viability.,” *Chembiochem: a European journal of chemical biology*, vol. 5, no. 3, pp. 258-60, Mar. 2004.

[14] B. T. Ruotolo and C. V. Robinson, “Aspects of native proteins are retained in vacuum.,” *Current opinion in chemical biology*, vol. 10, no. 5, pp. 402-8, Oct. 2006.

[15] K. Breuker and F. W. McLafferty, “The thermal unfolding of native cytochrome c in the transition from solution to gas phase probed by native electron capture dissociation.,” *Angewandte Chemie (International ed. in English)*, vol. 44, no. 31, pp. 4911-4, Aug. 2005.

[16] N. P. Barrera, N. Di Bartolo, P. J. Booth, and C. V. Robinson, “Micelles protect membrane complexes from solution to vacuum.,” *Science (New York, N.Y.)*, vol. 321, no. 5886, pp. 243-6, Jul. 2008.

[17] J. P. Whitelegge et al., “Toward the bilayer proteome, electrospray ionization-mass spectrometry of large, intact transmembrane proteins.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 19, pp. 10695-8, Sep. 1999.

[18] L. L. Ilag, I. Ubarretxena-Belandia, C. G. Tate, and C. V. Robinson, “Drug binding revealed by tandem mass spectrometry of a protein-micelle complex.,” *Journal of the American Chemical Society*, vol. 126, no. 44, pp. 14362-3, Nov. 2004.

[19] D. van der Spoel, E. G. Marklund, D. S. D. Larsson, and C. Caleman, “Proteins, lipids, and water in the gas phase.,” *Macromolecular bioscience*, vol. 11, no. 1, pp. 50-9, Jan. 2011.

[20] E. G. Marklund, D. S. D. Larsson, D. van der Spoel, A. Patriksson, and C. Caleman, “Structural stability of electrosprayed proteins: temperature and hydration effects.,” *Physical chemistry chemical physics: PCCP*, vol. 11, no. 36, pp. 8069-78, Sep. 2009.

[21] W. D. Luedtke et al., “Nanojets, electrospray, and ion field evaporation: molecular dynamics simulations and laboratory experiments.,” *The journal of physical chemistry. A*, vol. 112, no. 40, pp. 9628-49, Oct. 2008.

[22] E. Segev, T. Wyttenbach, M. T. Bowers, and R. B. Gerber, “Conformational evolution of ubiquitin ions in electrospray mass spectrometry: molecular dynamics simulations at gradually increasing temperatures.,” *Physical chemistry chemical physics: PCCP*, vol. 10, no. 21, pp. 3077-82, Jun. 2008.

[23] S. P. Thirumuruganandham and H. M. Urbassek, “Protein isolation through impact desolvation of electrosprayed microdroplets (IDEM): Molecular dynamics simulation,” *International Journal of Mass Spectrometry*, vol. 289, no. 2-3, pp. 119-127, Jan. 2010.

[24] R. Friemann, D. S. D. Larsson, Y. Wang, and D. van der Spoel, “Molecular dynamics simulations of a membrane protein-micelle complex in vacuo.,” *Journal of the American Chemical Society*, vol. 131, no. 46, pp. 16606-7, Nov. 2009.

[25] M. G. McCammon and C. V. Robinson, “Me, my cell, and I: the role of the collision cell in the tandem mass spectrometry of macromolecules.,” *BioTechniques*, vol. 39, no. 4, pp. 447, 449, 451 passim, Oct. 2005.

[26] H. J. C. van der Spoel, D., Lindahl, E. , Hess, B. , van Buuren, A. R., Apol, E., Meulenhoff, P. J., Tieleman, D. P., Sijbers, A. L. T. M., Feenstra, K. A., van Drunen, R., Berendsen, “Gromacs User Manual version 4.5.4.” 2010.

[27] V. Daggett and M. Levitt, “Protein unfolding pathways explored through molecular dynamics simulations.,” *Journal of molecular biology*, vol. 232, no. 2, pp. 600-19, Jul. 1993.

[28] W. C. Swope et al., "Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 2. Example Applications to Alanine Dipeptide and a β-Hairpin Peptide †," *The Journal of Physical Chemistry B*, vol. 108, no. 21, pp. 6582-6594, May 2004.

[29] S. C. Harvey, R. K.-Z. Tan, and T. E. Cheatham, "The flying ice cube: Velocity rescaling in molecular dynamics leads to violation of energy equipartition," *Journal of Computational Chemistry*, vol. 19, no. 7, pp. 726-740, May 1998.

[30] E. Marklund, "Gas Phase Protein Structure Under the Computational Microscope: Hydration, Titration, and Temperature," Uppsala Unversitet, 2011.

Chapter 8

Conclusions and Future Directions

1. Introduction

This final Chapter begins by discussing the main conclusions that may be drawn from the work presented in this thesis. Subsequently, the expected future directions of this research are discussed, along with some preliminary simulations.

2. Summary

The results of this thesis demonstrate that the question of the importance of environment on membrane proteins does not have a simple black-and-white answer, but that the influence of environment is likely to depend on the properties of a given membrane protein. It appears that there are two extremes: those proteins that are particularly susceptible to environmental influences such as the influenza M2 channels, and those such as OmpA that are resistant to most changes in structure upon environmental variation, while most other proteins are likely to fall somewhere inbetween (*e.g.*. PagP, a β-barrel protein that contains an α-helix).

The multiscale simulations performed in Chapter 3 allowed the process of BM2 TM domain folding to be mimicked based on sequence and knowledge of functional oligomer state alone. The use of a serial multiscale approach allows both long-timescale processes such as protein-protein rearrangements and concerted motions of the helices to be observed, along with aspects that require atomistic resolution, such as the proposed gating mechanism *via* His19 protonation. Although there were significant differences between the MD-generated and the solution NMR structure of the TM domain, based on comparison with A/M2 and the sensitivity of BM2 conformation to environment (as demonstrated in Chapters 6 and 7), it may be expected that a structure

of the BM2 TM domain solved by *e.g.* solid state NMR would match the CGMD-generated model more closely.

The work presented in Chapter 3 shows that CGMD simulations of the two-stage folding process may be used to generate a sequence-specific tetramer of the BM2 TM domain, which was stable upon conversion to atomistic representation. Control simulations provided evidence that CGMD may potentially be used to predict the oligomer state of membrane proteins based on prediction of TM regions from sequence alone. This requires further testing, but would be an extremely powerful tool in membrane protein structural biology if successfully developed. Implementing a high-throughput approach of these types of simulations using *e.g.* Sidekick [1] could lead to a method of structural prediction by a combination of CGMD and subsequent refinement by ATMD. Part of the development of the CGMD methodology would be to include the possibility of structural transitions rather than fixed secondary structure throughout a simulation.

In Chapter 5 a serial multiscale approach was used to generate protein-detergent complexes for further study following conversion to atomistic representation. The subsequent studies of the influence of environment on BM2 suggest that BM2, similarly to A/M2, is likely to exist in a range of conformations, depending on the surrounding environment. There have been relatively few MD studies of detergent micelles, and these simulations of protein-detergent complexes led to some key observations. Detergent molecules within the smaller DHPC micelle were found to be highly dynamic, while those within the larger DDM micelle were found to display similar dynamic behaviour to DPPC lipids within a bilayer. Furthermore, the binding modes of detergent molecules on the protein surface were more 'bilayer-like' in the BM2 DDM complex compared to BM2-DHPC. Overall DDM detergent micelles appear to more closely mimic the properties of a DPPC bilayer than DHPC micelles - a somewhat surprising result considering the matching headgroups of DHPC and DPPC. The detergent-protein complexes were found to retain properties of the detergent-only micelles, with the implication that properties of a given protein-detergent complex may be predicted from the intrinsic properties of the detergent.

The distinction between BM2, a small α-helical bundle, and OmpA, a monomeric β-barrel was particularly defined in comparison of the two classes of protein upon simulation under vacuum conditions (mimicking the environment experienced by an analyte in the collision chambers of an electrospray ionisation mass spectrometry experiment). OmpA in a DPC micelle was found to be influenced relatively little by loss of water, with the main changes occurring to the detergent packing, while simulations of BM2 in a similar

detergent DHPC led to changes in both detergent packing and helix packing. The use of the maltose-based detergent DDM led to preservation of the helix-helix packing and little change in detergent arrangement. The results suggest that the choice of detergent for an electrospray ionisation mass spectrometry experiment should be based upon those that are expected to retain normal micelle packing during transfer from solution to vacuum.

Molecular Dynamics is becoming an increasingly powerful tool in membrane protein structural biology and as computational power increases it can only be expected that the use of MD simulation to both predict and corroborate experimental findings will become commonplace. The work presented in this thesis has made use of a multiscale simulation approach, combining CGMD and ATMD simulations to maximise the range of observation times of a system to allow processes occurring on sub-nanosecond to microsecond timescales to be observed. Extending this approach to include a quantum level of detail all the way up to ultra-CG (where whole proteins may be modelled as a single particle) would allow a greater spatial and temporal range to be captured by MD simulation.

3. Future Directions

3.1. Cross-Validation of Experiment and MD Simulation

Much of the work presented in this thesis was compared with experimental data in order to validate the MD simulations performed. For example, a comparison of the DDM micelles generated by multiscale simulations was found to give a good fit with recent SAXS data (Rouse and Durand, unpublished). On the basis of these comparisons it becomes possible to reverse the direction of comparison, and use MD simulation to understand and interpret experimental results. For example, protein-detergent complexes may be generated *via* this simulation method and compared to scattering data to predict the architectures of protein-detergent complexes.

An area in which the use of MD in interpreting experimental results is likely to continue is that of mass spectrometry. As discussed in Chapter 7, the inspiration for the work presented was a paper based on the bacterial ABC-transporter BtuCD, a heteromeric membrane protein consisting of two transmembrane domains and two cytoplasmic

domains. The entire protein complex was found to be stabilised by the presence of DDM micelles in the gas phase [2].

The simulation processes of PDC formation (Chapter 5), and transfer to vacuum *via* steered MD simulation (Chapter 7) have been applied successfully to the BtuCD protein. Upon CGMD simulation a preformed DDM micelle (N=150) was found to form an extended torus around the TM domain of BtuCD, excluding water from these regions. Figure 8:1 shows the protein-detergent complex following CG2AT conversion and ATMD simulation for 50 ns. Steered MD simulations were then used to pull this PDC from solution into vacuum. During the transfer slight rearrangement of the DDM packing is observed, but the hydrophobic core remains excluded from water.

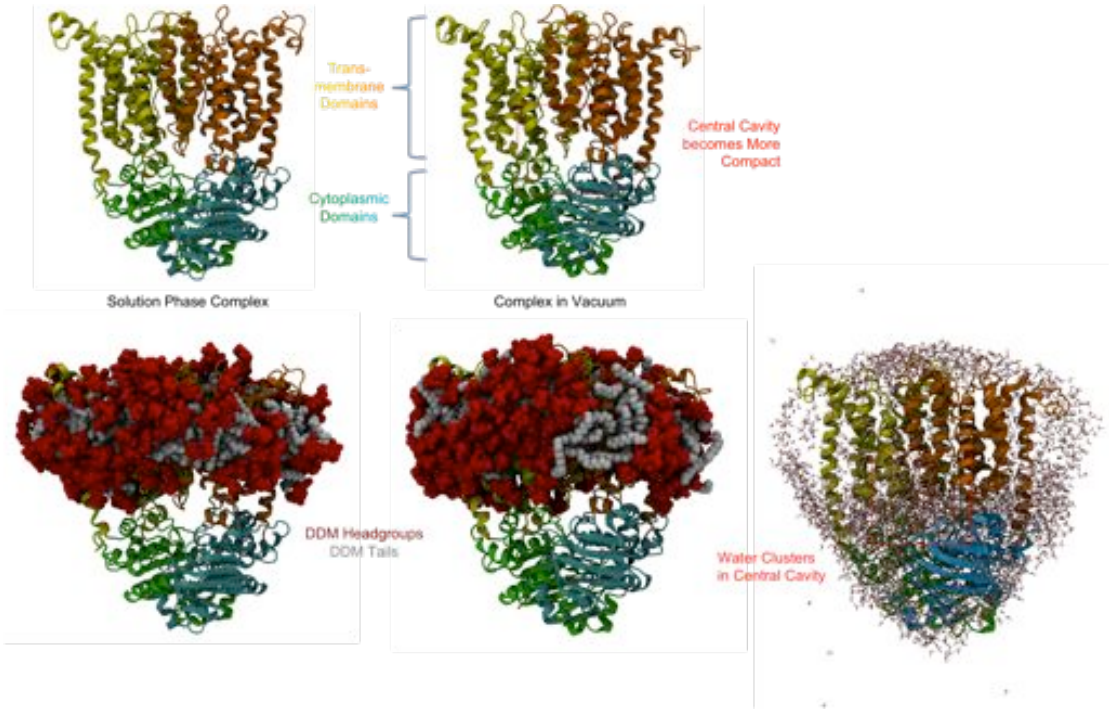


Figure 8:1. The BtuCD protein in solution (left) and vacuum (middle and right). The protein is shown in cartoon format, coloured by subunit. The DDM detergent maltose headgroups are shown as red spheres and the acyl chains as grey spheres. The right hand figure shows the partially solvated complex following a 75 ns ATMD simulation. The water molecules are shown in ball and stick representation (oxygen = red, hydrogen = white). The water molecules are observed to relocate and cluster in the central cavity between the subunits, maximising the hydrogen bonding network. A rearrangement of the BtuCD subunits accompanies this change.

Due to the increase in system size it was not feasible to perform a system under dehydrating conditions (a 500 ns ATMD simulation to match the BM2 simulations in Chapter 7 would currently take ~500 days on 64 processors). Upon a 75ns ATMD simulation of the partially solvated complex (under fixed water concentration *i.e.* with periodic boundary conditions) the main changes were a rearrangement of the subunit packing. The water molecules tended to cluster in the central cavity between the TM and cytoplasmic domains, concurrent with the protein subunits rearranging to become more

compact in this region. This thin layer of water was found to be sufficient to maintain the solution phase structure (overall Cα RMSD ≈2 Å). Manual removal of the furthest 75 % of water molecules from the protein surfaces led to the largest structural drift in the cytoplasmic domains. This is consistent with the conclusions of Chapter 7: the environment of the TM domain is affected least upon dehydration due to minimal rearrangement of the DDM detergent micelle, whereas the cytoplasmic domains are exposed to the largest change in environment, leading to increased structural drift.

Development of the CGMD collision methodology discussed in Chapter 7 would provide the next step in elucidating the structural changes that occur to protein-detergent complexes in electrospray ionisation.

3.2. Mechanism of Detergent Solubilisation

Understanding more of the process of membrane protein incorporation into detergent micelles is useful in understanding differing success rates of detergents in structural studies. The whole process of membrane protein solubilisation usually involves lysis of the cell membrane by detergent [3], [4]. Two mechanisms have been proposed for this detergent-mediated lysis. The first involves cooperative binding of detergent micelles, and the second involves association of a micelle with the bilayer and exchange of detergent and lipid molecules. The two pathways are proposed to exist for two classes of detergents – those that are able to transverse the bilayer *via* a flip-flop mechanism (first), and those that are not (second) [5].

To date no MD studies have been published that corroborate or dispute the two proposed mechanisms. To test the potential of using CGMD to study this, a system was generated with a larger (1000 lipid) DPPC bilayer patch containing BM2. Lauryldimethylamine-*N*-oxide (LDAO) detergent was added to the solution above the CMC. Some LDAO monomers and small clusters were able to penetrate the bilayer before they contacted other LDAO monomers. Rapid flip flop between the two membrane leaflets was observed. The remaining LDAO molecules formed a larger micelle (>70 LDAO molecules) which interacted with the bilayer before dissociating. This micelle was not incorporated into the bilayer until a lengthened interaction with the lipid headgroups occurred, *via* interaction with a protein. This process is shown in Figure 8:2.

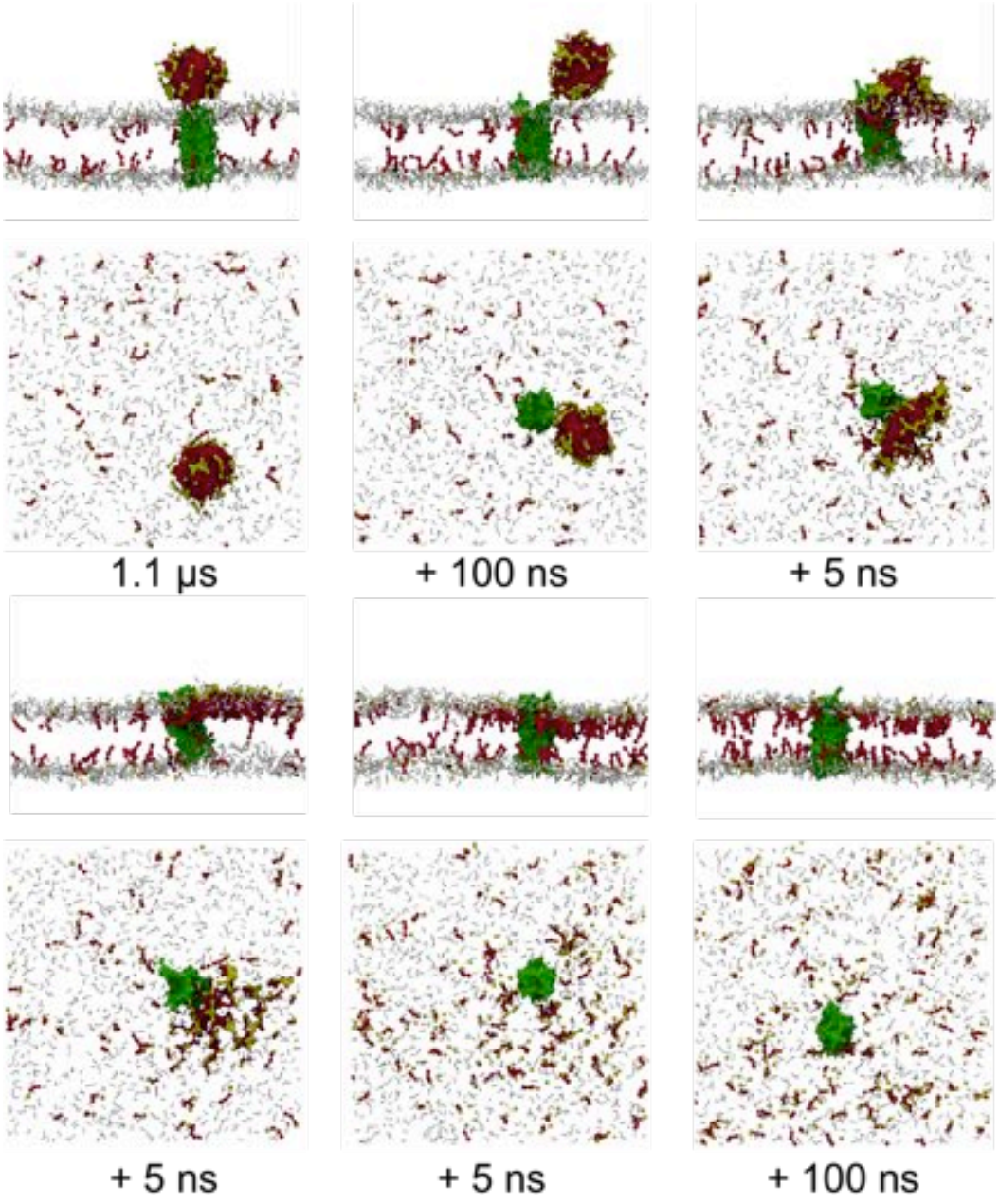


Figure 8:2. Mechanism of LDAO solubilisation of a membrane-protein containing bilayer. The snapshot at 1.1 µs shows the first time the LDAO micelle interacts with the BM2 protein. After ~100 ns of interacting the micelle moves slightly away and distorts, allowing it to collapse into the bilayer leaflet. This sudden increase in the number of lipid/detergent molecules in one leaflet leads to distortion in the bilayer, which is alleviated as the LDAO molecules equilibrate amongst the two leaflets *via* the flip-flop mechanism. BM2 is shown in surface representation (coloured green). The DPPC lipid headgroups are shown as grey spheres. The LDAO acyl tails are shown by red spheres. The LDAO headgroups are represented by a yellow surface.

Once the first LDAO micelle had become incorporated into the bilayer it appeared that a second one was much more able to bind, without requiring extended interactions with the bilayer. It is unclear if this is due to favourable interactions of the headgroups of the LDAO detergents already incorporated into the bilayer, or due to the bilayer already being disrupted somewhat by the presence of detergent. There was little evidence to suggest the presence of a detergent-rich domain forming around the protein, although it

could be that greater concentrations of detergent within the membrane are required to cause this. It appears that the bilayer solubilisation process occurs *via* a cooperative mechanism and is likely to be facilitated by the presence of projecting residues of *e.g*. membrane proteins.

These preliminary studies suggest that CGMD simulations may be used to provide a means to understanding the reasons for differing success rates of various detergents in membrane protein purification.

3.3. Modelling Full-Length BM2: Interactions of the Cytoplasmic Domain with Matrix Protein 1

The simulations of BM2 described thus far have only included the transmembrane domain of the protein. However, as discussed in the Introduction to this thesis, the cytoplasmic domain of BM2 is necessary for normal viral replication [6]. It has also been proposed that the cytoplasmic domain may modulate the packing of the TM domain. The structure of the solution NMR structure of the BM2 cytoplasmic domain is made all the more intriguing due to the absence of a structure for the cytoplasmic domain of A/M2. This structure opens up two avenues to be explored: firstly, the effect of the cytoplasmic domain on the properties (flexibility and gating) of the TM domain, and secondly, the interaction of the cytoplasmic domain with Matrix Protein 1 of *influenza* B (BM1). To explore these aspects of BM2 function, a structure of the full-length BM2 channel was generated by modelling loop regions between the two domains (described in the Methods Chapter).

ATMD simulations were performed to determine if the presence of the cytoplasmic domain had any effect on the conformation or stability of the transmembrane domain of BM2 however initial simulations of this structure in a DPPC bilayer demonstrated no increase in stability.

It was therefore decided to focus on studying the interaction of BM2 with BM1. During virus budding, the cytoplasmic domain is thought to be involved in a specific interaction with BM1, partly due to the opposing electrostatic surfaces of each protein. This interaction was studied via solution NMR where the main perturbation to the BM2 chemical shifts upon presence of BM1 was to the C-terminal residues [7]. To explore this interaction using CGMD simulations a homology model of BM1 was created by sequence alignment to AM1 (PDB id 1AA7, [8]) as described in the Methods Chapter. The full-

length BM2 model and the BM1 homology model were converted to CG resolution as described previously.

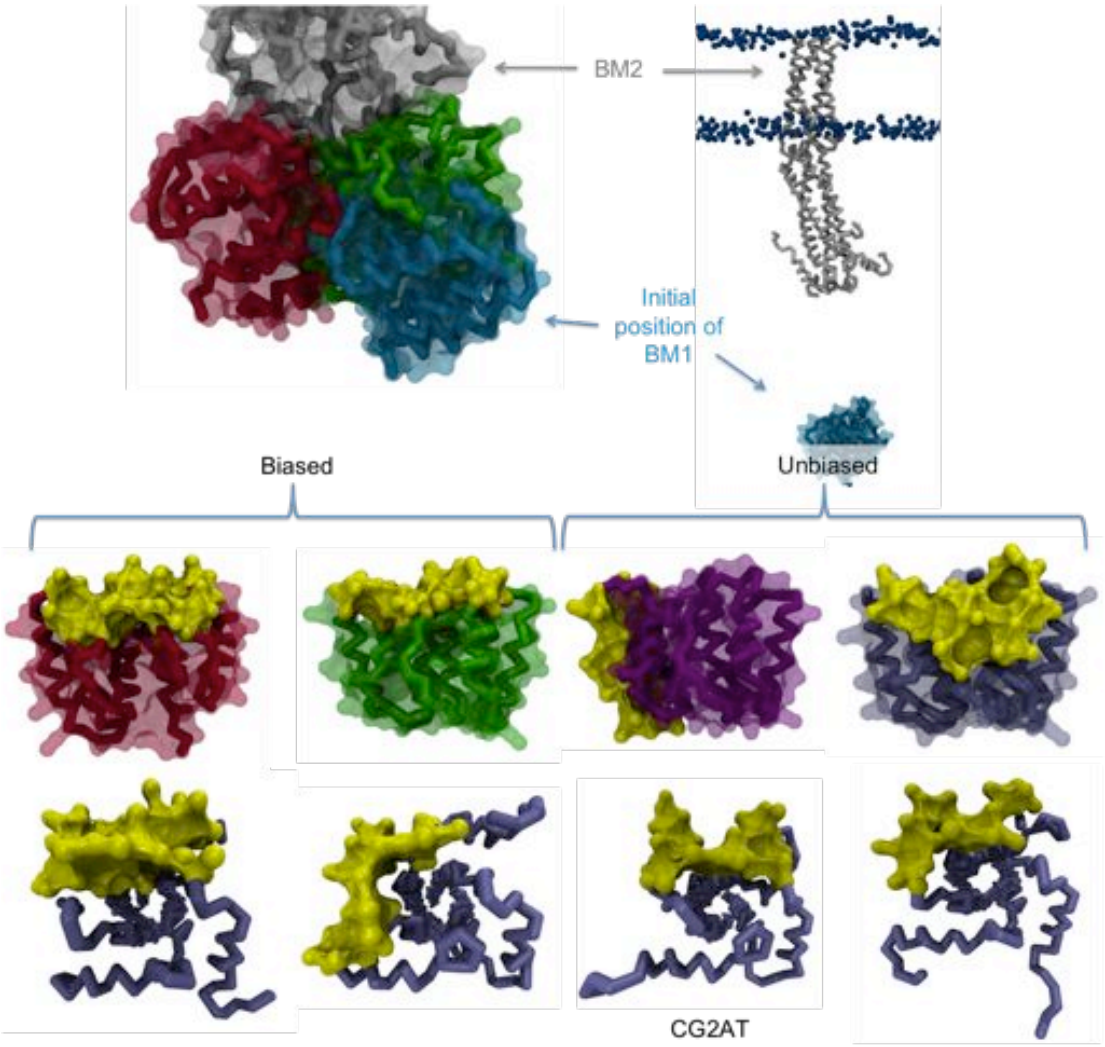


Figure 8:3. CGMD Simulations of the BM2 BM1 Interaction. Top: the initial set up of the two types of simulation performed (biased and unbiased). Bottom: the interaction surfaces of BM1 (above) and BM2 (below) are displayed as yellow surfaces..

Two types of CGMD simulation were performed, referred to as 'biased' and 'unbiased'. In the unbiased simulations the BM1 molecule was placed 40 Å from the BM2 cytoplasmic domain. Upon CGMD simulation BM1 was observed to rotate and eventually contact the BM2 cytoplasmic domain (Figure 8:3). In the biased simulations the BM1 protein was placed near the BM2 cytoplasmic domain to test whether a specific binding site was observed. In three of four simulations performed (two of each type) a conserved binding surface was obtained (corresponding to conserved charged surface residues from the AM1 structure) whereas another positively charged helix (also conserved between AM1 and BM1) was found to interact in the other simulation.

The final configuration from an unbiased simulation (labelled "CG2AT" in Figure 8:3) was converted to an atomistic representation (using CG2AT process described earlier) and simulated using ATMD. The residue-residue interaction that was generated from CGMD simulation was maintained for the duration of the ATMD simulation (although the rest of the BM2 model lost secondary structure).

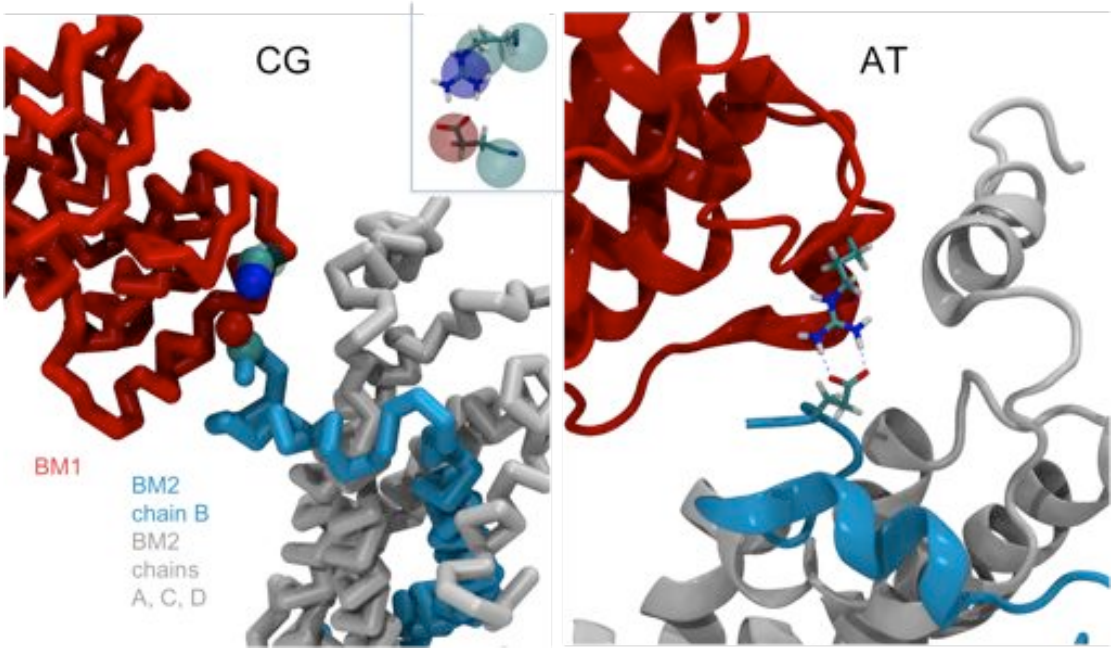


Figure 8:4. Conserved BM2-BM1 Interaction Site between CG and AT representations. Left: the configuration from a CGMD simulation chosen for conversion to an atomistic representation (using OPLS-AA force field). The inset shows the imposition of the AT residues (ball-and-stick) on the CG residues (transparent spheres). The right panel shows the AT system following ~5 ns of ATMD simulation. The maintained hydrogen-bonding interaction between BM2 and BM1 residues is shown by dotted lines.

The use of CGMD to investigate protein-protein interactions is in its infancy and more rigorous studies will be necessary to determine the interaction sites of the two proteins. However these preliminary simulations suggest that it may be possible to use a multiscale approach, combining CGMD and ATMD to incorporate protein-protein interactions into a system.

4. References

[1] M. S. Hall, Benjamin A, Sansom, "Sidekick: A pipeline and framework for high throughput molecular dynamics simulations," *In Preparation*.

[2] N. P. Barrera, N. Di Bartolo, P. J. Booth, and C. V. Robinson, "Micelles protect membrane complexes from solution to vacuum.," *Science (New York, N.Y.)*, vol. 321, no. 5886, pp. 243-6, Jul. 2008.

[3] T. Arnold and D. Linke, "The use of detergents to purify membrane proteins.," *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]*, vol. 4, pp. Unit 4.8.1-4.8.30, Aug. 2008.

[4] M. le Maire, P. Champeil, and J. V. Moller, "Interaction of membrane proteins and lipids with solubilizing detergents.," *Biochimica et biophysica acta*, vol. 1508, no. 1-2, pp. 86-111, Nov. 2000.

[5] U. Kragh-Hansen, M. le Maire, and J. V. Møller, "The mechanism of detergent solubilization of liposomes and protein-containing membranes.," *Biophysical journal*, vol. 75, no. 6, pp. 2932-46, Dec. 1998.

[6] M. Imai, K. Kawasaki, and T. Odagiri, "Cytoplasmic Domain of Influenza B Virus BM2 Protein Plays Critical Roles in Production of Infectious Virus," *Journal of Virology*, vol. 82, no. 2, pp. 728-739, 2008.

[7] J. Wang, R. M. Pielak, M. A. McClintock, and J. J. Chou, "Solution structure and functional analysis of the influenza B proton channel.," *Nature structural & molecular biology*, vol. 16, no. 12, pp. 1267-71, Dec. 2009.

[8] B. Sha and M. Luo, "Structure of a bifunctional membrane-RNA binding protein, influenza virus matrix protein M1," *Nature Structural Biology*, vol. 4, no. 3, pp. 239-244, Mar. 1997.

Appendix A

Transfer of Protein-Detergent Complex from Solution to Vacuum

Choice of Parameters for Steered MD

Several parameters were tested for the steered MD simulations transferring the protein-detergent complex from solution to vacuum. The BM2-DHPC complex was used as the control system, using the OPLS-AA force field. Table 9 summarizes the simulations performed. The aim was to generate a sparingly-solvated complex with minimal disruption to the lipids due to the directional pull force, to mimic the dehydration process. Applying the pulling force only to the protein centre of mass resulted in the majority of the DHPC lipids remaining in solution, with only a few DHPC remaining bound to the protein *via* interactions of the lipid headgroups with charged residue sidechains. Using the higher pull rate of 10 nm/ns led to some disruption of the complex, with one or more lipids becoming detached during the transition. The slowest pull rate that was feasible to use was 1 nm/ns, corresponding to simulation lengths of ~20 ns for the transition. The chosen parameters for the simulations of the solution to vacuum simulation were 'pull rate' = 1 nm/ns, 'pull force' = 1000. This lead to no lipids being removed from the complex during the transition. Any water molecules confined within the pore of a protein were included within the group to be pulled as initial simulations without doing so led to the waters moving through the pore.

Pull Rate (nm/ns)	Force Constant (kJ/mol/nm)	Pull Group	# Waters in Complex	# Lipids Lost (Total)
10	1000	Protein_DHPC	891	2 (78)
10	500	Protein_DHPC	732	1 (78)
1	1000	Protein_DHPC	1371	0 (78)
10	500	Protein	251	72 (78)

Table 9. Steered MD Parameters Tested.

For the chosen pull parameters, three parameter sets were used – GROMOS96, OPLS-UA (both united-atom) and OPLS-AA (all atom). In united atom force fields all electrostatic effects between hydrocarbon chains must be incorporated implicitly into the Lennard-Jones potentials [1] . Although these interactions are expected to be small (given the low polarisability of the acyl chains), this effect might be expected to become more significant in vacuum conditions. The resulting PDCs following transfer to vacuum are shown in Figure 0:1, along with the changes in the protein solvent accessible surface area. The OPLS-AA simulation appears to be least disrupted upon the transfer. Therefore it was decided in this work to use the OPLS-AA force field.

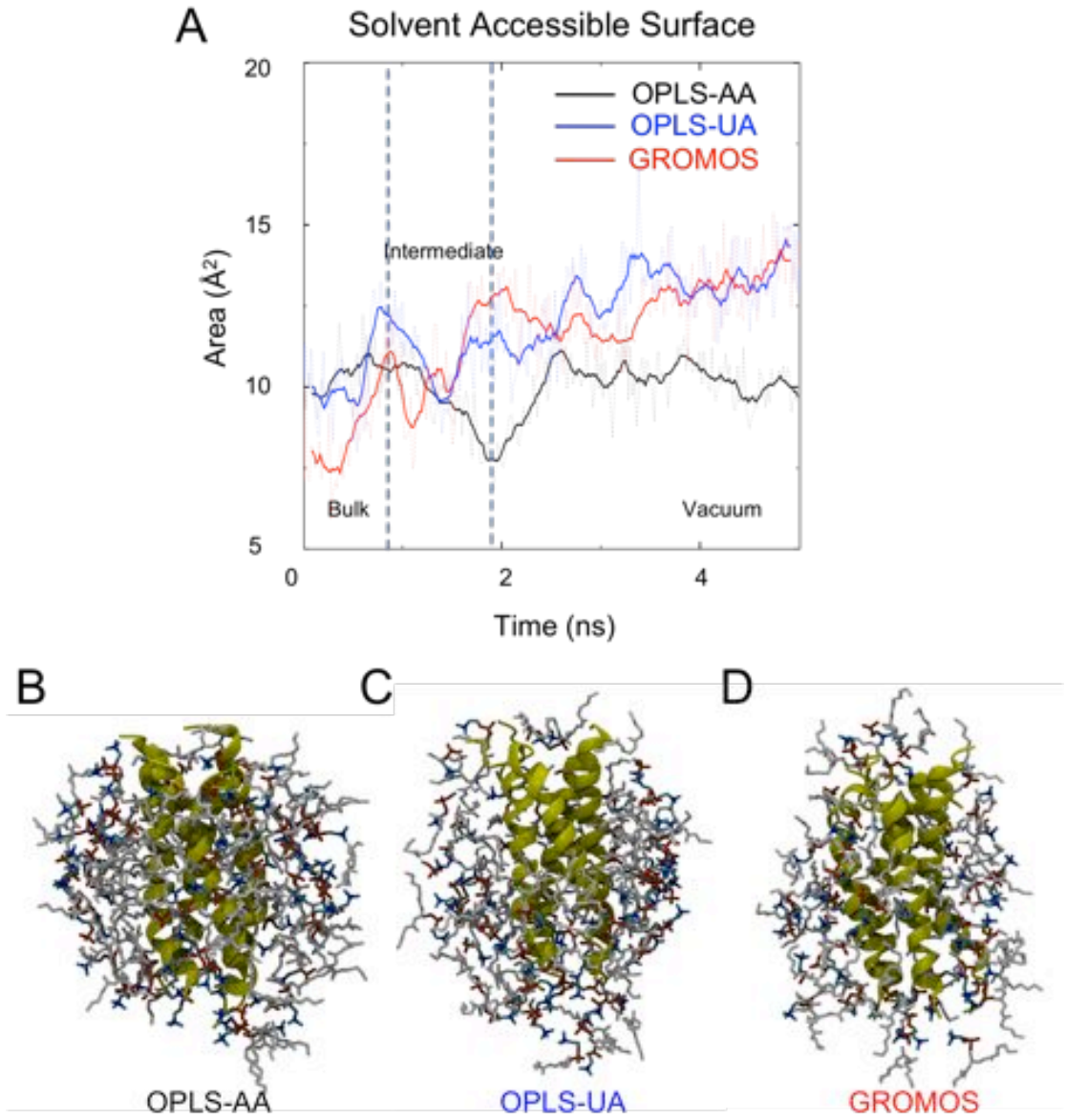


Figure 0:1. Comparison of united atom and all atom force field parameters on the transfer of the BM2-DHPC complex from solution to vacuum. A: Protein solvent accessible surface areas including presence of detergent for each system. B-D Snapshots of each protein-detergent complex at 5 ns following steered MD simulation from solution to vacuum.

Reference:

[1] J. P. Ulmschneider and M. B. Ulmschneider, "United Atom Lipid Parameters for Combination with the Optimized Potentials for Liquid Simulations All-Atom Force Field," *Journal of Chemical Theory and Computation*, vol. 5, no. 7, pp. 1803-1813, Jul. 2009.