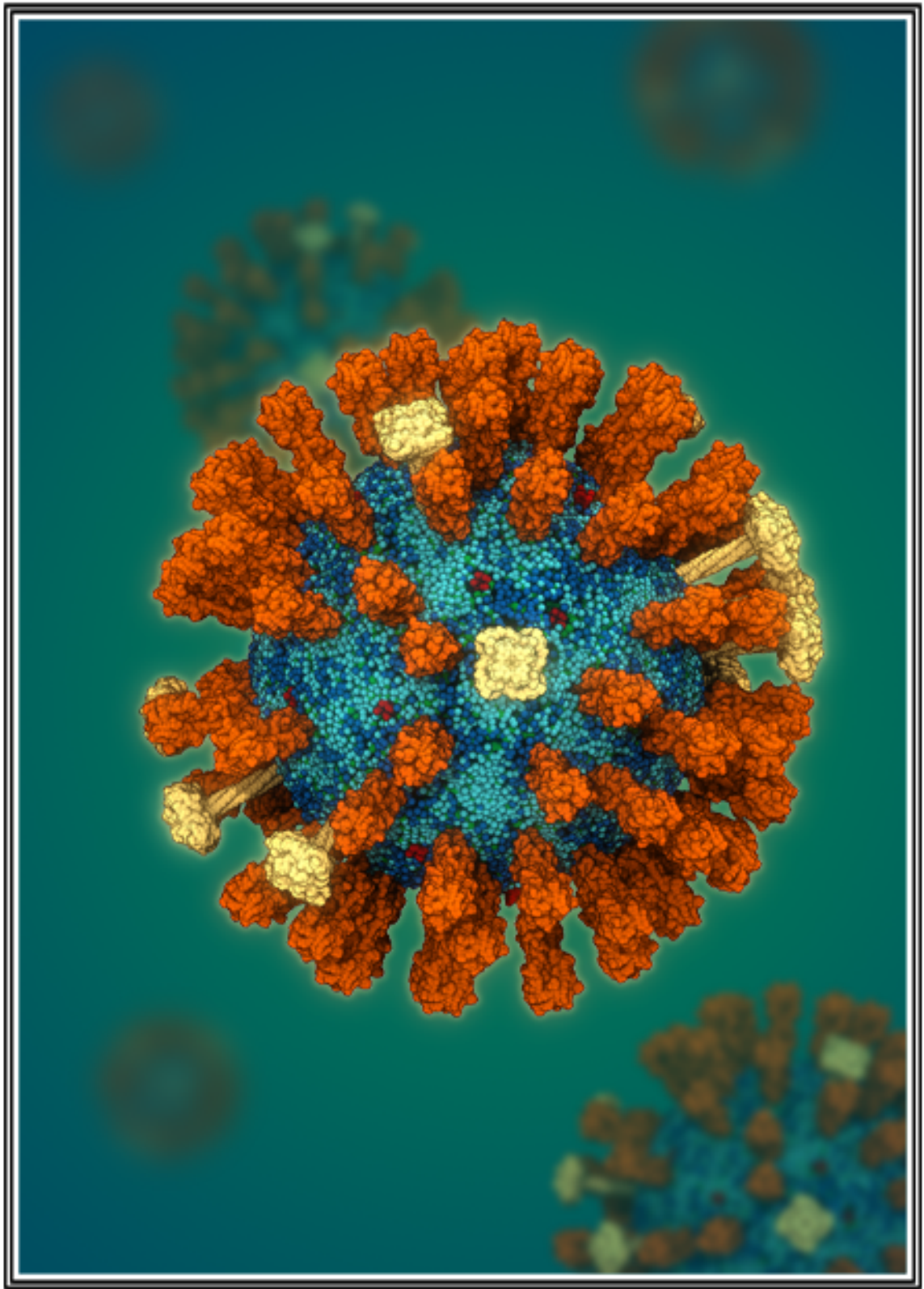


PUSHING THE BOUNDARIES:
MOLECULAR DYNAMICS SIMULATIONS OF
COMPLEX BIOLOGICAL MEMBRANES



Coarse-Grained Model of the Influenza Virus Envelope

Pushing the Boundaries: Molecular Dynamics Simulations of Complex Biological Membranes

Daniel L. Parton



Structural Bioinformatics and Computational Biochemistry Unit,
Department of Biochemistry and Wolfson 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

A copy of this thesis has been deposited in the Bodleian Library, Oxford.

Viva voce examination held on 7th November 2011 by Dr Nicole Zitzmann (University of Oxford) and Prof Richard Lavery (Université Lyon / CNRS).

© 2012 Daniel Parton

Abstract

A range of simulations have been conducted to investigate the behaviour of a diverse set of complex biological membrane systems. The processes of interest have required simulations over extended time and length scales, but without sacrifice of molecular detail. For this reason, the primary technique used has been coarse-grained molecular dynamics (CG MD) simulations, in which small groups of atoms are combined into lower-resolution CG particles. The increased computational efficiency of this technique has allowed simulations with time scales of microseconds, and length scales of hundreds of nm.

The membrane-permeabilizing action of the antimicrobial peptide maculatin 1.1 was investigated. This short α -helical peptide is thought to kill bacteria by permeabilizing the plasma membrane, but the exact mechanism has not been confirmed. Multiscale (CG and atomistic) simulations show that maculatin can insert into membranes to form disordered, water-permeable aggregates, while CG simulations of large numbers of peptides resulted in substantial deformation of lipid vesicles. The simulations imply that both pore-forming and lytic mechanisms are available to maculatin 1.1, and that the predominance of either depends on conditions such as peptide concentration and membrane composition.

A generalized study of membrane protein aggregation was conducted via CG simulations of lipid bilayers containing multiple copies of model transmembrane proteins: either α -helical bundles or β -barrels. By varying the lipid tail length and the membrane type (planar bilayer or spherical vesicle), the simulations display protein aggregation ranging from negligible to extensive; they show how this biologically important process is modulated by hydrophobic mismatch, membrane curvature, and the structural class or orientation of the protein.

The association of influenza hemagglutinin (HA) with putative lipid rafts was investigated by simulating aggregates of HA in a domain-forming membrane. The CG MD study addressed an important limitation of model membrane experiments by investigating the influence of high local protein concentration on membrane phase behaviour. The simulations showed attenuated diffusion of unsaturated lipids within HA aggregates, leading to spontaneous accumulation of raft-type lipids (saturated lipids and cholesterol).

A CG model of the entire influenza viral envelope was constructed in realistic dimensions, comprising the three types of viral envelope protein (HA, neuraminidase and M2) inserted into a large lipid vesicle. The study represents one of the largest near-atomistic simulations of a biological membrane to date. It shows how the high concentration of proteins found in the viral envelope can attenuate formation of lipid domains, which may help to explain why lipid rafts do not form on large scales in vivo.

Acknowledgements

First and foremost I must thank Prof Mark Sansom for his excellent supervision and guidance over the past four years. Despite leading a huge lab (and now a whole department), he always has time for his students, and it is thanks to him that SBCB is such a uniquely friendly environment to work in. Mark has always been on hand to give me advice when I have needed it, but has also allowed me the freedom to develop my own research skills and interests, and for these things I am immensely grateful.

I thank the BBSRC and Fujitsu Laboratories of Europe for funding my research, and am hugely grateful to Dr. Elena Akhmatskaya and Dr. Ross Nobes of Fujitsu for their helpful supervision and advice.

There are probably too many people to name in SBCB who have helped me out in work or life over the course of my DPhil, but here goes anyway (no particular order): Chezney, Goose, Rob, Rouse, Leon, Sarah Allison (née Rogers), Mitch, Janye, the Phils (and a Phill), Emi, Oli, Jochs, J-Rome, Jussi, Ranjit, Jemma, Shaz, Antreas, Alan and Ben.

I thank Syma Khalid (Southampton, ex-SBCB) for being such an excellent source of academic advice and amusing stories. Thank you to Marc Baaden for hosting my summer placement at IBCP (Paris) and for being so generous in his help with the influenza project. Also Alex and Matthieu for making me feel so welcome.

I should also acknowledge my luck in having found some great friends in college, especially Lewis, Miranda, Heather and Chris for their companionship in the midst of all the madness of academic life, and my housemates Christian and John for our various and occasionally nefarious adventures. Christian also for never missing an opportunity, as a true crystallographer, to remind me of the challenges remaining for the simulation field.

Finally, a huge thank you to my family for always being there for me, even at the times when I have forgotten to show my gratitude. Thank you to my dad for inspring my love of science, my mum for inspiring my love of music, and my little sister for... uh, being good at singing (seriously, your continuing success is a great source of encouragement for me). I love you all.

List of Publications Arising from this Thesis

Parton, D. L., A. Tek, M. Baaden, and M. S. P. Sansom, 2011. Coarse-grained molecular dynamics simulations of the envelope of influenza virus A. (*In preparation*)

Parton, D. L., A. Tek, M. Baaden, and M. S. P. Sansom, 2011. Enrichment of raft-type lipids within clusters of hemagglutinin: a coarse-grained molecular dynamics study. (*In preparation*)

Parton, D. L., E. V. Akhmatskaya, and M. S. P. Sansom, 2012. Multiscale simulations of the antimicrobial peptide maculatin 1.1: water permeation through disordered aggregates. *J. Phys. Chem. B.* (Epub ahead of print) DOI: 10.1021/jp232458y

Parton, D. L.[†], J. W. Klingelhoefer[†], and M. S. P. Sansom, 2011. Aggregation of model membrane proteins, modulated by hydrophobic mismatch, membrane curvature, and protein class. *Biophys. J.* 101:691–699

Bond, P. J., D. L. Parton, J. F. Clark, and M. S. P. Sansom, 2008. Coarse-grained simulations of the membrane-active antimicrobial peptide maculatin 1.1. *Biophys. J.* 95:3802–3815

[†] denotes joint first authorship.

Contents

Abstract	v
Acknowledgements	vi
List of Publications Arising from this Thesis	vii
List of Abbreviations	xii
1 Introduction	1
1.1 The plasma membrane	1
1.1.1 Membrane lipids	1
1.1.2 Membrane proteins	3
1.1.3 Lipid rafts	6
1.2 Coarse-grained simulations of complex membrane protein systems	9
1.3 Influenza virus A	13
1.3.1 Structure and morphology	14
1.3.2 Host cell entry	14
1.3.3 Genomic replication	16
1.3.4 Host cell exit	17
2 Theory and Methods	18

2.1	Introduction	18
2.2	Molecular mechanics force fields	18
2.2.1	Bonded interactions	19
2.2.2	Non-bonded interactions	21
2.3	Molecular dynamics simulation methods	23
2.3.1	Introduction	23
2.3.2	Numerical integration of Newton's second law	24
2.3.3	Periodic boundary conditions	25
2.3.4	Treatment of long-range intermolecular forces	27
2.3.5	Temperature and pressure control	29
2.4	Coarse-grained force fields	32
2.4.1	The Bond-Sansom model	32
2.4.2	The MARTINI model	37
3	Mechanism of Membrane Permeabilization by the Antimicrobial Peptide	
	Maculatin 1.1	41
3.1	Introduction	41
3.2	Methods	45
3.2.1	CG Simulations	45
3.2.2	Atomistic Simulations	47
3.2.3	Software and Analysis	47
3.3	Results	48
3.3.1	Multiscale Simulations of Peptide/Bilayer Interactions	48
3.3.2	Water Permeation Through Peptide Clusters	52

3.3.3	Minimum Number of Peptides Required for Formation of a Membrane-Spanning Cluster	54
3.3.4	Effect of Cluster Stoichiometry on Water Diffusion	55
3.3.5	Deformation of Vesicles Induced by Maculatin	56
3.4	Discussion	57
4	Modulation of Model Membrane Protein Aggregation by Hydrophobic Mismatch, Membrane Curvature, and Protein Class	62
4.1	Introduction	62
4.2	Methods	63
4.2.1	Protein pore models	63
4.2.2	Preparation of the protein pore-lipid bilayer systems	63
4.2.3	Preparation of the vesicle systems	64
4.2.4	Simulation details	65
4.3	Results	65
4.3.1	Simulations of Multiple TM Proteins in Planar Bilayers	65
4.3.2	Simulations of Multiple TM Proteins in Vesicles	67
4.3.3	Lateral Diffusion of Lipids and Proteins	71
4.3.4	Lipid Distribution Around the Proteins	74
4.4	Discussion	76
5	Raft-like Aggregates of Influenza Hemagglutinin Simulated by Coarse-Grained Molecular Dynamics	80
5.1	Introduction	80
5.2	Methods	82
5.2.1	Hemagglutinin model	82

5.2.2	Production simulation set-up	85
5.2.3	Simulation details	86
5.3	Results and Discussion	87
5.3.1	Single HA protein in a DUPC bilayer	87
5.3.2	Single HA protein in a domain-forming bilayer	88
5.3.3	Aggregate of ten HA proteins in a domain-forming bilayer	92
5.4	Conclusions	98
6	Coarse-Grained Molecular Dynamics Simulations of the Entire Influenza Viral Envelope	101
6.1	Introduction	101
6.2	Methods	103
6.2.1	Protein Modelling	103
6.2.2	Virion Assembly	106
6.2.3	Simulation Details	108
6.3	Results	109
6.4	Discussion	115
7	Conclusions	120
A	Influence of the integrator time step on simulation properties	124
B	Derivation of Expression for Diffusional Permeability Coefficient	127
C	Lipid vesicle leaflet equilibration	129
	Bibliography	131

List of Abbreviations

AMP Antimicrobial peptide	PDB Protein Data Bank
CFP Cyan fluorescent protein	PME Particle-mesh-Ewald
CG Coarse grained	PMF Potential of mean force
COM Centre of mass	POPC 1-palmitoyl-2-oleoyl-phosphatidylcholine
cryo-ET Cryo-electron tomography	POPE 1-palmitoyl-2-oleoyl-phosphatidylethanolamine
DIG Detergent-insoluble glycolipid complex	POPG 1-palmitoyl-2-oleoyl-phosphatidylglycerol
DMPC Dimyristoylphosphatidylcholine	RMSD Root mean square deviation
DMPG Dimyristoylphosphatidylglycerol	SFK Src-family kinase
DPD Dissipative particle dynamics	STALL Stimulation-induced temporary arrest
DPPC Dipalmitoylphosphatidylglycerol	TFE Trifluoroethanol
DRM Detergent-resistant membranes	TGN <i>Trans</i> Golgi network
DUPC Dilinoleylphosphatidylcholine	TM Transmembrane
EC Extracellular	VLP Virus-like particle
EM Electron microscopy	vRNA Viral RNA
ENM Elastic network model	YFP Yellow fluorescent protein
ESR Electron spin resonance	
FRET Förster resonance energy transfer	
GPCR G-protein coupled receptor	
GPI Glycophosphatidylinositol	
GUV Giant unilamellar vesicle	
HA Hemagglutinin	
IC Intracellular	
M1.1 Maculatin 1.1	
MD Molecular dynamics	
MOI Moment of inertia	
NA Neuraminidase	
NMR Nuclear magnetic resonance	
PBC Periodic boundary conditions	
PC Phosphatidylcholine	

Chapter 1

Introduction

1.1 The plasma membrane

The plasma membrane forms the main boundary between the interior of a cell and its surrounding environment. Some bacteria and archaea possess structural features outside this boundary (such as an outer membrane or protein S-layer), but the plasma membrane acts as the major barrier to permeation of small chemical species, providing selective control over the flow of water, ions and organic molecules into or out of the cell. It consists of two major components: lipids and membrane proteins. The huge variety in these components (Fig. 1.1) reflects the wide-ranging and complex activities which take place at the membrane, such as cell signalling, solute transport, cell adhesion and cellular motility.

1.1.1 Membrane lipids

The main structural basis of the plasma membrane is a lipid bilayer. This is composed of two “leaflets” of amphiphilic lipids, which are arranged in two-dimensional arrays and stacked so that the hydrophobic components are buried in the bilayer core. In eukaryotes, phospholipids are the major amphiphilic species. These generally contain a glycerol backbone with two hydrophobic fatty acyl chains attached via ester linkages, and a phosphate group attached at the third alcohol site. A number of polar headgroup moieties can be attached to the phosphate group, such as choline, ethanolamine, serine and *myo*-inositol. For example, 1-palmitoyl-2-oleoylphosphatidylcholine (POPC) contains a palmitoyl chain (16 carbons, fully

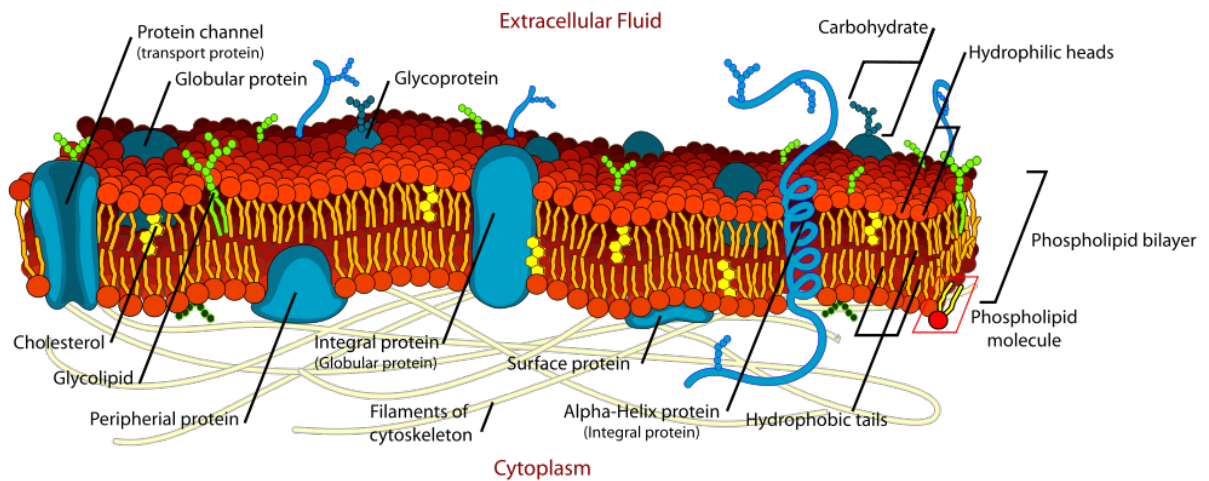


Figure 1.1: Schematic showing the major components of the plasma membrane.

Image courtesy of Marian Ruiz.

(http://commons.wikimedia.org/wiki/File:Cell_membrane_detailed_diagram_en.svg)

saturated), an oleyl chain (18 carbons, one *cis* double bond), a glycerol backbone, a phosphate group and a choline headgroup. An exception to the predominance of phospholipids with glycerol backbones are sphingolipids, which instead have a sphingosine backbone. The major sphingolipids in mammalian cells are sphingomyelin and glycosphingolipids, the latter of which contain oligosaccharide chains attached to the headgroup. The hydrocarbon tails of sphingolipids tend to be more saturated than those of glycerophospholipids [1].

Sterols are another major component of the eukaryotic plasma membrane and are largely hydrophobic, comprising a tetracyclic hydrocarbon ring structure attached to a hydrocarbon tail and alcohol headgroup. Cholesterol is the most common sterol in mammals, and ergosterol the most common sterol in yeast. Sterols modulate the properties of the bilayer in a number of ways. For example, pure phospholipid bilayers undergo a sharp phase transition between gel and liquid crystalline forms. In the gel phase, the lipid tails are extended and rigidly packed, and lateral and rotational diffusion are limited. The liquid crystalline phase is characterized by higher tail disorder and diffusion rates, as well as reduced bilayer thickness. The phase transition is typically near physiological temperature. Sterols modulate this phase behaviour, helping to maintain membrane fluidity and allowing plasma membranes to function over a wide range of temperatures [1]. Another specific example, which is important for the work reported in Chapters 5 and 6, is the cholesterol-dependent formation of lipid rafts—small, fluctuating regions of the membrane enriched in saturated phospholipids, cholesterol

and certain proteins [2]. The underlying basis for this phenomenon and its implications for membrane function are discussed in more detail later in this Chapter.

Lipids are not distributed equally between the two bilayer leaflets. In eukaryotes, sphingolipids and phosphatidylcholine (PC) lipids are enriched in the outer leaflet, while phosphatidylserine (PS) and phosphatidylethanolamine (PE) lipids are predominantly found in the inner leaflet. This asymmetry is generated in the Golgi and maintained in the plasma membrane via specific transport of PS and PE from the outer to the inner leaflet by ATP-dependent transporter proteins, and a lack of transport of sphingolipids and PC (spontaneous lipid “flip-flop” across leaflets is relatively slow) [3]. In the endoplasmic reticulum, by contrast, nonspecific transport of lipids results in transleaflet equilibrium. The distribution of lipids in the plasma membrane also results in an asymmetry in charge, with mainly neutral lipids present in the outer leaflet, and negatively charged lipids in the inner leaflet. This marks an important difference with the bacterial plasma membrane, which has an overall negative charge in both leaflets.

As well as providing a barrier to water and solutes, lipids play a specific functional role as signalling molecules. Metabolism by upstream signalling components can produce lysophospholipids, which carry only one fatty acyl chain and quickly leave the membrane [4], or diacylglycerols, which are glycerophospholipids with phosphates and headgroups removed, and which remain in the membrane [5]. These metabolites can act as secondary messenger molecules in a number of signalling pathways, interacting with other membrane receptors or recruiting cytosolic proteins. Phosphorylated derivatives of phosphatidylinositol lipids are also important components in signalling pathways, and in targetting of certain cytosolic and membrane proteins to the plasma membrane [3]. Lipids can also have important influences on membrane protein structure and function, which are discussed further later in the Chapter.

1.1.2 Membrane proteins

Protein constitutes 20–60% of the dry mass of biological membranes, or 1–2% by mole fraction [6], and membrane proteins are major drug targets [7, 8]. They can be broadly categorized as (i) integral membrane proteins, which are permanently attached to the membrane, or (ii) peripheral membrane proteins, which are only temporarily attached to a membrane. A number of methods are available for attachment of a protein to a membrane [9]. Transmembrane

(TM) proteins pass entirely through the membrane one or more times, and usually possess both extracellular and cytoplasmic domains—all TM proteins are integral membrane proteins. Other proteins may insert into only part of the membrane, for example with an amphipathic helix or hydrophobic loop, or may sit outside the membrane and interact via intermolecular forces. Certain proteins may also be anchored to the membrane via post-translational modifications. For example: glycosylphosphatidylinositol (GPI)-anchored proteins have a GPI moiety attached at the C-terminus and are typically anchored to the outer plasma membrane leaflet; acylated proteins are usually anchored to the inner leaflet by a fatty acid (e.g. palmitic acid) attached via a thioester linkage with a cysteine sidechain (*S*-acylation) or via an amide linkage with the N-terminal α -amino group (*N*-acylation).

Integral membrane proteins are estimated to constitute $\sim 30\%$ of the proteome of most organisms [10], and are highly diverse in terms of structure and function. TM domains can be categorized as two main structural types: α -helical (either single-pass helices or α -helical bundles which comprise multiple helices bundled together) and β -barrels (comprising anti-parallel β -strands forming a cylindrical sheet) (Fig. 1.2). β -barrel TM proteins are mainly found in the outer membranes of Gram-negative bacteria or mitochondria, while α -helical TM proteins are found in most other types of membrane.

Despite the large number of TM proteins identified, and their importance in biology and disease, relatively few three-dimensional structures have been solved. Of the $\sim 76,000$ structures currently deposited in the Protein Data Bank [11], only ~ 900 are for membrane proteins, of which only ~ 300 are for unique proteins (<http://blanco.biomol.uci.edu/mpstruc/listAll/list>). The difficulty of structure collection for membrane proteins is mainly due to their native membrane environment. Electron diffraction is the only method to date which can provide structures of proteins inserted in a lipid bilayer [10]. However, the approach is particularly challenging experimentally, due to the requirement for two-dimensional crystals of a protein in a membrane—only ~ 8 structures have been determined at atomic resolution, with another 30–40 determined at a resolution high enough to resolve TM α -helices [10]. It has been suggested that the specific lipids present in a membrane do not strongly effect the protein structure solved [12], e.g. for structures of aquaporin-0 in dimyristoylphosphatidylcholine (DMPC) or *Escheria coli* polar lipids, which possess different headgroups and tails [13, 14]. However, other studies have indicated that lipids can specifically influence protein function through structural changes [15–17].

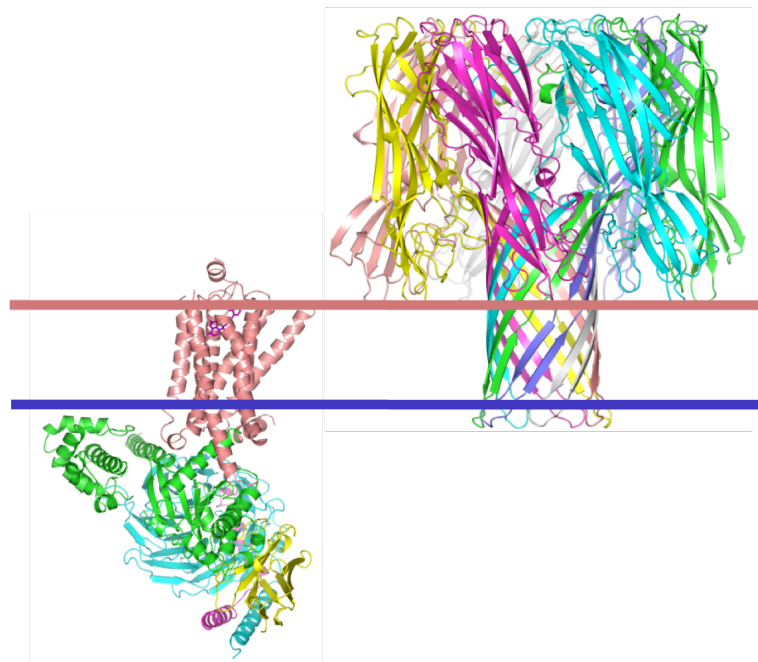


Figure 1.2: Examples of α -helical bundle and β -barrel TM domains. The left-hand image shows the human β -2 adrenergic receptor (pink) in complex with the Gs subunit of a guanine nucleotide-binding protein (G protein) (PDB code: 3SN6). The right-hand image shows the bacterial toxin α -hemolysin (PDB code: 7AHL). The protein images were downloaded from the Orientations of Proteins in Membranes (OPM) database.

(<http://opm.phar.umich.edu/>)

Membrane protein structures have also been solved using X-ray diffraction of three-dimensional crystals or NMR spectroscopy of proteins in solid or solution state [18]. However, a number of experimental issues also make these approaches challenging. Notably, proteins are typically stabilized in detergent or lipid micelles, which in some cases can affect the solved structure [19].

A major advantage of molecular dynamics simulations is the ability to observe protein behaviour in a native lipid environment. Lipids may strongly influence membrane protein dynamics and function in a number of ways [20]. For example, hydrophobic mismatch (when the hydrophobic length of a lipid bilayer and protein TM domain are not matched) and curvature stress have been shown to affect the channel activity of the aquaporin GlpF [15], as well as protein aggregation [17]. Charged lipids can also alter ion concentrations local to the membrane surface [17]. There is also increasing evidence that protein TM domains can include specific ‘binding sites’ for lipids [16], e.g. for cholesterol [21] and very recently for sphingolipids [22]. As a final example, lipid rafts are known to selectively incorporate certain types of protein, providing dynamic platforms for membrane protein function (discussed further in the following text on lipid rafts) [2].

1.1.3 Lipid rafts

Lipid rafts can be generally described as small (< 100 nm diameter), fluctuating patches of membrane enriched in saturated phospholipids, cholesterol and certain proteins, including most GPI-anchored and acylated proteins, and some TM proteins (see Fig. 1.3). They have been implicated in essentially all types of membrane bioactivity in eukaryotes, including trafficking, signalling, membrane protein sorting, endo-/exocytosis, and modulation of membrane structural properties. However, their exact nature has remained difficult to assess; much evidence for their existence has been derived from indirect methods, and rafts have so far proved impossible to visualize directly with microscopic methods.

1.1.3.1 Evidence for lipid rafts in biological membranes

The fluid mosaic model proposed by Singer and Nicholson in 1972 [23] contended that the plasma membrane is essentially a two-dimensional liquid, in which lipids and proteins can diffuse freely and randomly. However, the existence of membrane heterogeneity is clear from

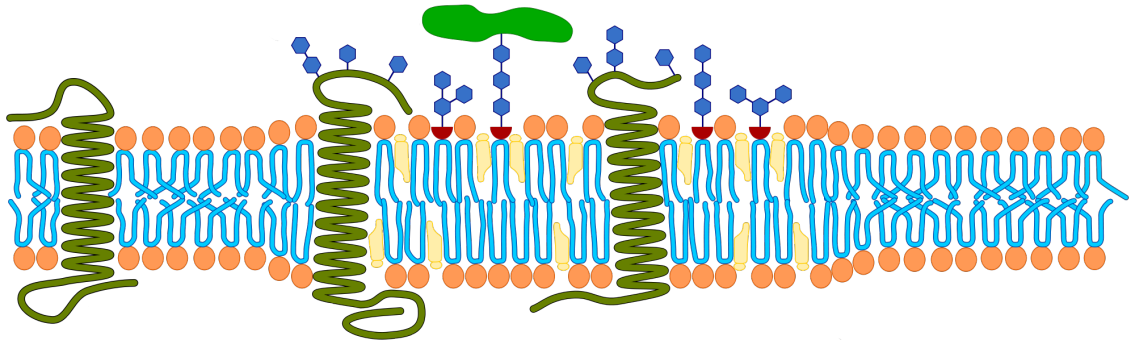


Figure 1.3: Schematic of a lipid raft. The raft is enriched in saturated lipids, which are depicted with increased ordering compared to unsaturated lipids outside the raft. Lipids within the raft are typically PC or sphingomyelin, as well as glycosphingolipids (sugars are shown in dark blue). Cholesterol is also enriched within the raft and is shown in the figure in cream. Certain proteins are associated with or incorporated into lipid rafts. GPI-anchored proteins are anchored to the outer leaflet of the plasma membrane (the green globular protein shown in the centre of the raft), while acylated proteins are typically anchored to the inner leaflet (not shown). TM proteins can also be incorporated into rafts, especially glycoproteins as shown. Many of these TM proteins also contain acylated residues within their TM domains. Image adapted from original by Artur Jan Fijalkowski.

(http://commons.wikimedia.org/wiki/File:Lipid_raft_organisation_scheme.svg)

enrichment of sphingolipids in the apical membranes of epithelial cells, and in budded viral envelopes such as influenza. Within a few years of publication of the Singer and Nicholson article, a number of theories had been presented which proposed the existence of small clusters of ordered lipids [24–26]. The next major source of evidence for the existence of rafts was the resistance of certain membrane components to cold extraction with nonionic detergents (particularly Triton X-100) [27]. These membranes were termed “detergent-resistant membranes” (DRMs) or “detergent-insoluble glycolipid complexes” (DIGs), and were found to be enriched in sphingolipids, cholesterol and certain proteins, especially those anchored to the membrane by post-translational modifications such as GPI-anchors, palmitoylations and myristoylations. However, a number of caveats must be considered when comparing DRM composition to raft composition. For example, different methods of extraction lead to incorporation of different proteins [28]. Secondary evidence, such as cholesterol dependence (modulated via cyclodextrin depletion, for example), is thus important for classification of a protein as “raft-associating” [28].

More recently, advances in fluorescence microscopy have provided further insight. For example, Förster resonance energy transfer (FRET) studies have shown that influenza hemag-

glutinin, a TM protein which is recovered from DRMs, associates with raft markers in the plasma membrane of live mammalian cells. The raft markers used were GPI-anchored cyan fluorescent protein (CFP) (which is anchored to the outer plasma membrane leaflet) [29] and acylated yellow fluorescent protein (YFP) (anchored to the inner plasma membrane leaflet) [30]. Other studies have begun to hint at functionalities for rafts. For example, single-particle tracking was used to study the role of the CD59 receptor in signal transduction in lymphocytes [31]. Cross-linking of CD59, a GPI-anchored receptor anchored to the outer plasma membrane leaflet, was known to induce formation of molecular complexes with Src-family kinases (SFKs), which are anchored to the inner plasma membrane leaflet via acylations. However, the mechanism of signal transduction across the membrane was not obvious, as the only possible contact between the proteins would be via their lipid anchors. Single-particle tracking showed that CD59 associates with Lyn (an SFK) via lipid rafts and undergoes “stimulation-induced temporary arrest” (STALL) upon binding between Lyn and the inner leaflet-anchored G α i2 protein. This STALL mechanism appears to depend upon association of Lyn or G α i2 with actin, and takes place on a time scale of ~ 0.5 s and length scale of 50 nm. Transbilayer signal transduction may occur via an unidentified TM protein which also associates with the complex.

According to a consensus developed at the 2006 Keystone Symposium of Lipid Rafts and Cell Function, lipid rafts are defined as follows: “Lipid rafts are small (10–200 nm), heterogeneous, highly dynamic, sterol- and sphingolipid-enriched domains that compartmentalize cellular processes. Small rafts can sometimes be stabilized to form larger platforms through protein-protein and protein-lipid interactions.” [32]. Caveolin proteins can also induce formation of larger lipid rafts through insertion into the plasma membrane from the cytoplasmic side. The resulting caveolae appear as small invaginations in the plasma membrane and can bud off, providing a clathrin-independent endocytic pathway. They are generally considered as a distinct type of lipid raft.

1.1.3.2 Liquid-ordered domains in model membranes

Rafts have frequently been compared with liquid-ordered domains which separate out from model membranes with ternary lipid mixtures. These membranes typically comprise saturated lipid, unsaturated lipid and cholesterol [33]. Within certain limits of temperature and lipid composition, these membranes will separate into liquid-ordered (Lo) and liquid-

disordered (Ld) domains. Ld domains are enriched in unsaturated lipids, and essentially identical to the liquid phase of the pure lipid. Lo domains are enriched in saturated lipids and cholesterol, and characterized by increased lipid tail ordering, but (compared to the equivalent solid-ordered (So or gel) phase) a relatively small decrease in lateral and rotational diffusion.

Domain segregation is thought to be driven by the mutual attraction of saturated lipid tails and cholesterol: the rigid, planar sterol ring favours interactions with the straighter and less flexible saturated hydrocarbon tails [34]. Other effects may also contribute. For example, the sterol ring induces ordering in lipid tails, increasing local bilayer thickness and therefore introducing a hydrophobic mismatch with disordered lipids [35]. Another important consideration is that sphingolipids generally increase the tendency for domain segregation compared to phosphatidyl lipids. This is due to their possession of both hydrogen-bond acceptors and donors, which increase their propensity for self-association. Sphingolipids are the main types of lipid in vivo which possess entirely saturated lipids, whereas phosphatidyl lipid tails typically include at least one double bond (e.g. POPC).

A number of important differences exist between Lo domains in model membranes and rafts in biological membranes. Most noticeably, ternary mixtures can undergo complete lipid segregation, forming two large domains. This can be visualized clearly using fluorescent markers (Fig. 1.4) [36]. Furthermore, all known raft-associating TM proteins partition into Ld domains in model membranes, rather than the Lo domain which is enriched in raft-type lipids [37–41]. A number of features of the plasma membrane may account for this disparity, such as increased protein concentration, cytoskeletal interactions, and endo-/exocytosis, as well as differences in experimental temperatures [2]. It is clear that domains in ternary-mixture model membranes are very different from rafts in biological membranes, and the terms Lo and Ld are only applied to the model membrane domains throughout this thesis.

1.2 Coarse-grained simulations of complex membrane protein systems

The use of atomistic MD simulations for the study of biological systems is well established. However, the gap in time and length scales, between fundamental chemical processes (e.g. bond vibration) and large-scale collective behaviour (e.g. protein folding), makes fully atom-

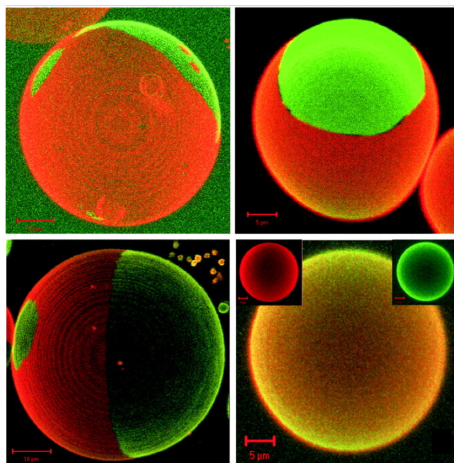


Figure 1.4: Confocal images of domains in giant unilamellar vesicles (GUVs) (figure adapted from [36]). The GUVs comprise 1 : 1 mixtures of sphingomyelin and POPC, and increasing concentrations of cholesterol: (clockwise from top-left) 10 mol%, 20 mol%, 33 mol%, and 50 mol%. Domains were visualized using a Lo domain-associating fluorescent probe (green) and a Ld domain-associating probe (red). The probes are uniformly distributed in the GUV in the bottom-right panel, as the concentration of cholesterol is too high to allow domain formation (instead, the lipids are likely to be all in the Lo phase [33]).

istic simulations computationally intractable for many biological processes of interest. CG force fields aim to bypass this problem by modelling biological species at lower levels of resolution, thus increasing the speed at which a system can be simulated (see Fig. 1.5).

The development of CG methodology has been a major focus within the biomolecular simulation field in recent years, and a number of different approaches have been taken by different groups. The MARTINI force field [43–45] models a wide range of biologically relevant molecules, such as lipids, sterols, proteins and carbohydrates. It has been made publicly available (md.chem.rug.nl/cgmartini) and is now used by a number of groups. This force field, as well as variants from the Sansom group [46] and Schulten group [47], have been parameterized to reproduce a wide range of empirical data from experiment and atomistic simulations. A different approach is to explicitly match the force profiles of CG and atomistic simulations. This has been the focus of the Voth group, who have developed the theory and methodology for such a force-matching scheme [48–51]. The method aims to produce CG force fields for which the equilibrium probability distribution of CG particles is consistent with that of the system of atomistic particles from which the force field was derived. This systematic approach shares some similarities with that of the Klein group and their early

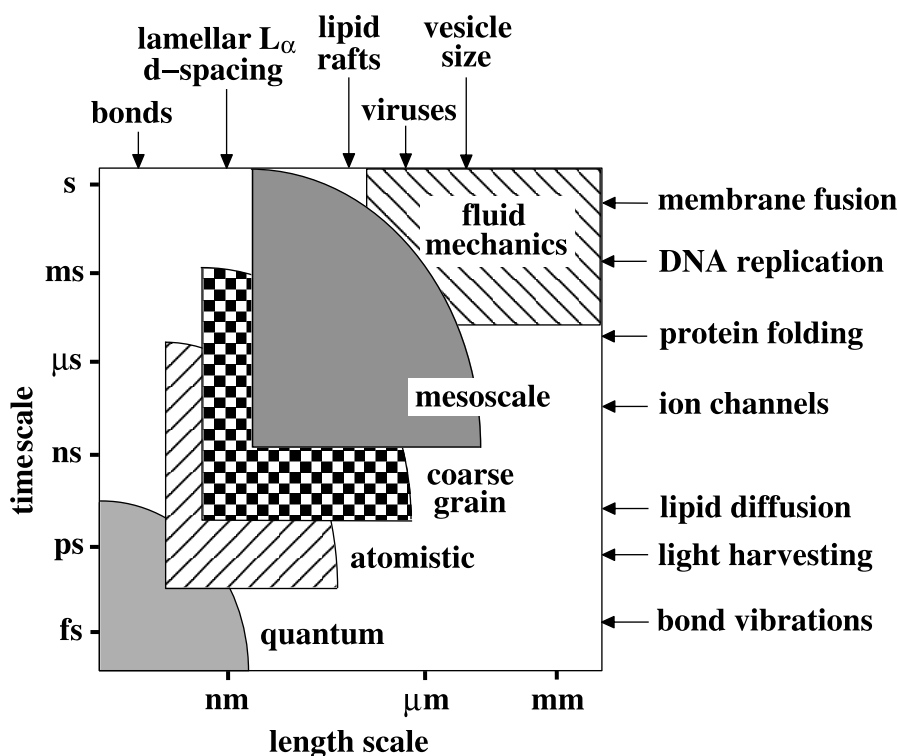


Figure 1.5: Temporal and spatial scales typically accessible with computer simulation techniques. Figure adapted from [42].

developments in CG simulation [42, 52]. Other models have been developed for use in dissipative particle dynamics (DPD) simulations (e.g. [53, 54]; see [55] for a review). The DPD technique is simpler than MD in terms of computation and chemical complexity, and is thus often used to simulate large-scale mesoscopic systems. It generally involves three types of interaction: a soft, repulsive force between particles, a drag force, and a random force [53].

Many CG models used for membrane protein simulations were originally developed for the simulation of lipid systems. Fig. 1.6 shows a number of coarse-graining approaches for lipids, using different degrees of granularity and types of potential. The basic details of these models are described in the following text. It should be noted that many other CG models have been developed for the investigation of a wide variety of biological processes, and a useful review can be found in the textbook edited by Voth [56]. The models described in the following text are those relevant to simulations of complex membrane protein systems (i.e. those with multiple TM proteins inserted in a membrane), with a few examples of other CG lipid models for comparison.

Schmidt and Weiss have used a simple, generalized lipid model for DPD simulations, com-

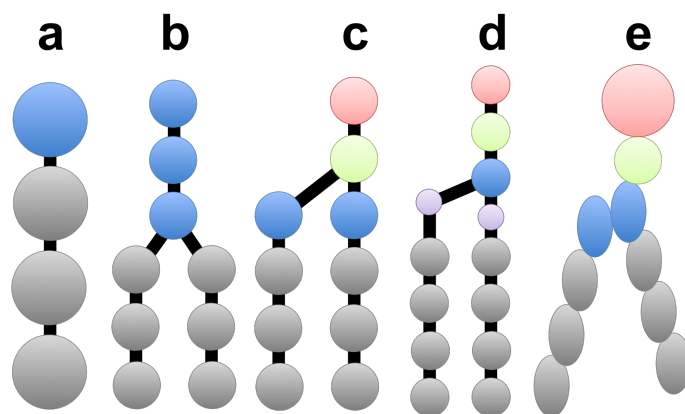


Figure 1.6: Schematics showing the CG particle mappings and connectivity of lipids from various CG force fields (models are not shown to scale). *a)* A generalized, single-tail, amphiphilic lipid used in DPD simulations by Schmidt and Weiss [54, 57]. *b)* A generalized, two-tail, amphiphilic lipid used in DPD simulations by Smit and co-workers [55]). *c–e)* models of DMPC from the force fields of (*left–right*) Marrink [44], Shelley and Klein [52], and Orsi and Essex [58]. The latter model uses ellipsoidal particles to represent the lipid tails.

prising a single hydrophilic headgroup particle (H) and a single linear tail made up of n hydrophobic particles (T_n), all connected by harmonic springs and an overall bending rigidity [54, 57]. Water is modelled explicitly as a simple sphere (particle type W). A generalized TM protein model has also been included in the force field, represented with a bundle of seven chains (one central chain surrounded by six other chains arranged hexagonally). The terminal particle of each chain is of the hydrophilic H type, and the central particles of the hydrophobic T type. The same harmonic springs and bending stiffness were applied to each chain, and additional harmonic springs were applied between particles of different chains. A number of DPD studies using this simple model have shown how hydrophobic mismatch in a membrane can modulate TM protein aggregation [54, 57, 59].

Smit and co-workers developed a two-tail lipid model for use in DPD simulations [55, 60]. The headgroup is modelled with three H type particles, and the tails with a variable number of T type particles. The particles are held together with harmonic springs, and an overall bending rigidity is applied to each tail. Water is again modelled with a separate particle type (W). They also used a similar protein model to Schmidt and Weiss, except with three H type particles at each chain terminus. This model has been used to investigate protein-membrane interactions as a function of protein length (by varying the number of central T particles in the protein chains), protein type (by varying the number protein chains), and lipid type (by

varying the number of tail particles). A study with a single inserted TM protein characterized data such as hydrophobic mismatch, protein tilt and protein-induced bilayer deformation [61]. Later studies considered aggregation of multiple TM proteins [62, 63].

The more complex Shelley and Klein CG model [52] is designed to represent specific lipids in MD simulations. The DMPC model (pictured in Fig. 1.6) comprises 13 particles, with particles of different types for each of the choline, phosphate and tail groups, and a further three particles for the glycerol group (one central particle type plus two particles representing ester linkages). A single spherical particle is used to represent three water molecules. This model has been used to simulate lipid bilayers, and matches the physical properties (radial distribution functions of lipids, and density profiles) of atomistic bilayer simulations.

The Orsi and Essex CG model [58] is notable because it represents the lipid tails with ellipsoid CG particles, using the Gay-Berne potential [64] to model intermolecular forces involving these particles. The anisotropic nature of the particles makes them in some ways more suitable for representing the lipid tails, although this comes with a cost to computational efficiency.

As described above, the CG force field originally developed by Marrink and co-workers [43, 44] has been adapted for MD simulations of membrane proteins by a number of groups [45, 47, 65]. The models developed by the Marrink and Sansom groups were used for the work reported in this thesis, and are described in more detail in the Methods Chapter. It should be noted that the protein models are far more complex and realistic than the simple models used in the DPD simulations described above. The Marrink group have conducted simulations of multiple TM proteins inserted in domain-forming membranes, and these studies are discussed in more detail in Chapters 5 and 6.

1.3 Influenza virus A

Influenza virus A is an enveloped, negative-sense RNA virus which belongs to the family *Orthomyxoviridae*. The viral genome is encoded in eight segments of single-stranded viral RNA (vRNA), which code for eleven proteins: hemagglutinin (HA), neuraminidase (NA), matrix proteins 1 and 2 (M1 and M2), nucleoprotein (NP), non-structural proteins 1 and 2 (NSP1 and NSP2—the latter is also known as nuclear export protein (NEP)), polymerase

acidic protein (PA), polymerase basic protein 1 and polymerase basic protein 1 - F2 (PB1 and PB1-F2), and polymerase basic protein 2 (PB2). Antigenic properties of HA and NA are used to subtype influenza viruses: for example, an H5N1 strain would contain HA subtype 5 and NA subtype 1. There are 15 known subtypes of HA and 9 of NA [66].

1.3.1 Structure and morphology

Individual virions can be spherical or filamentous in form. The central core of the virus is the ribonucleoprotein (RNP), which comprises the vRNA strands packed around the NP, as well as small amounts of NEP, PA, PB1 and PB2 [67]. The vRNA is surrounded by a layer of matrix protein, which is made of multiple copies of M1 and takes on the overall morphology of the virion (spherical or filamentous). Directly outside the matrix layer is the lipid envelope, which is derived from the host cell plasma membrane and densely packed with the three TM proteins HA, NA and M2.

A number of recent cryo-electron tomography (cryo-ET) studies have provided fascinating new insights into the structural and morphological characteristics of the virions, and especially the nature of the envelope [68–70]. Spherical virions of two different strains have been found to be 73–177 nm in diameter (including the glycoproteins HA and NA) [68, 69]. Filamentous virions can be up to 1 μm in length (~ 130 nm is more typical), with uniform short-axis diameters of around 90 nm [70]. HA is the most abundant envelope protein, comprising approximately 87% of the glycoprotein content in spherical virions [68]. HA and NA are often loosely clustered together in spherical virions [68, 69], whereas a study of filamentous virions showed that NA is often presented at one end (possibly to facilitate scission of an exiting virion from a host cell), and the RNP at the other [70]. M2 is present in relatively low numbers, with around 15–25 copies per virion [71]. The matrix layer is often incomplete (20% of spherical virions for one strain) or even entirely missing (14%) from budded virions [69].

1.3.2 Host cell entry

The HA protein is responsible for host cell recognition and binding. It is a homotrimeric glycoprotein and forms spikes extending ~ 14 nm from the envelope surface [68]. It comprises two disulfide-linked subunits, HA1 and HA2, which are cleaved from the HA0 precursor

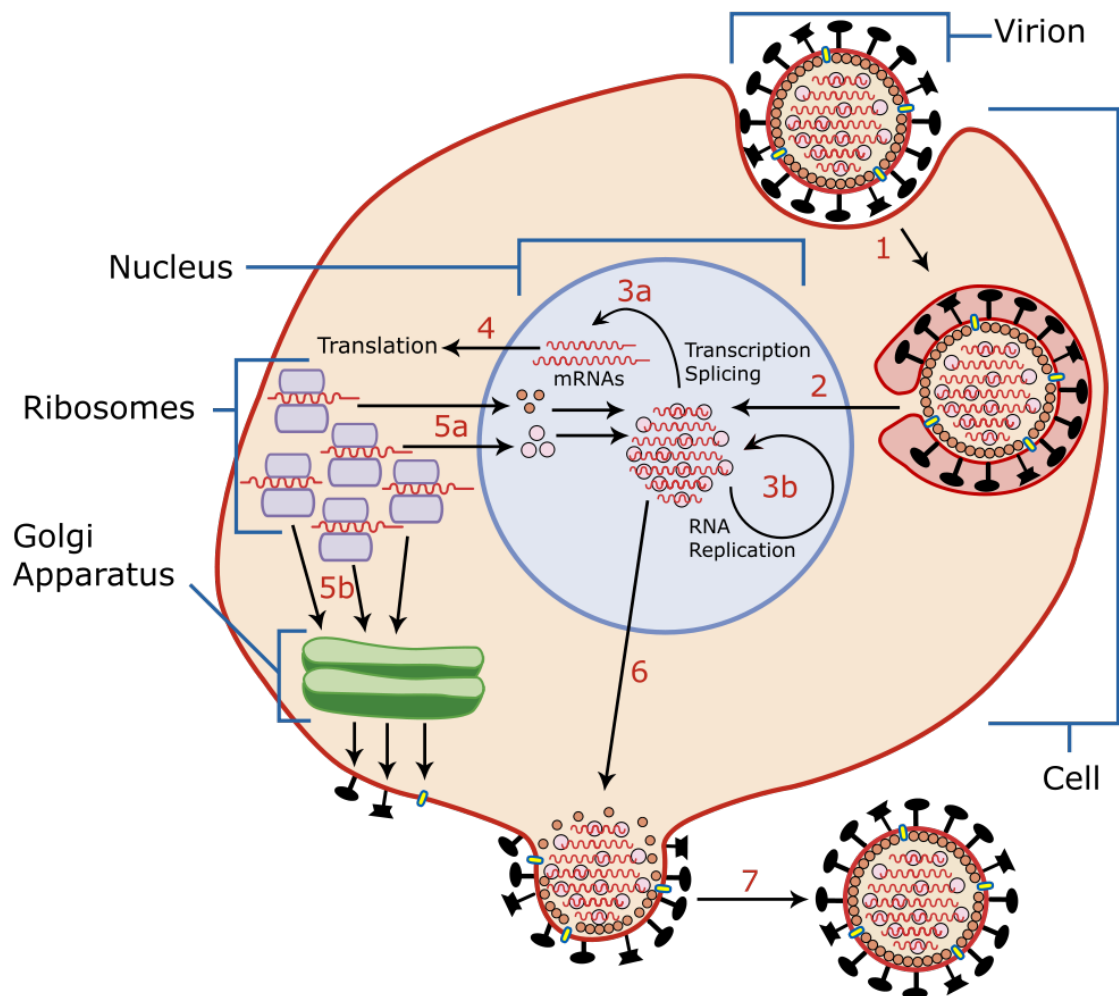


Figure 1.7: Viral replication. The steps in the process are described in the text. The image is publicly available on Wikimedia Commons.

http://commons.wikimedia.org/wiki/File:Virus_Replication_large.svg

protein. The HA1 chain contains the receptor-binding site, which binds sialic acids attached to glycoproteins on the host cell membrane surface. Sialic acids (*N*-acetylneuraminic acid, Neu5Ac, being the most common) are typically bound to the penultimate galactose residue of a glycoprotein oligosaccharide chain by one of two possible linkages: Neu5Ac $\alpha(2,3)$ -Gal or Neu5Ac $\alpha(2,6)$ -Gal. These linkages give rise to one type of specificity between different strains of virus. For example, human strain HA proteins recognize the $\alpha(2,6)$ linkage, whereas avian and equine HA proteins recognize the $\alpha(2,3)$ linkage [66]. Swine variants recognize both [66]. These differences may influence the spread of the virus between species, as well as the location of infection—human, equine and swine influenza are respiratory infections, whereas avian influenza is an enteric infection [66].

Following sialic acid binding, the virus is imported into the host cell via receptor-mediated endocytosis (Fig. 1.7, step 1). The low pH (5–6) of the endosome induces conformational changes in two of the envelope proteins. First, the HA2 chain moves to expose the fusion peptide, which inserts into the endosomal membrane. This acts to pull the endosomal and viral membranes together and facilitate fusion. The other change is in the M2 protein, which is a homotetrameric proton channel. It opens at low pH to allow entry of protons into the virus interior. The acidification of the virus interior results in disruption of the matrix layer and, in conjunction with fusion of the viral and endosomal membranes, the viral RNP is released into the host cell cytoplasm (step 2). Inhibition of the M2 channel is the mode of action used by the amantadine-type anti-influenza drugs.

1.3.3 Genomic replication

The proteins contained in the viral RNP (NP, PA, PB1 and PB2) all contain nuclear localization sequences (NLS) which target the proteins to the nuclear import machinery [72]. Following import into the cell nucleus, the RNA-dependent RNA polymerase transcribes and replicates the viral genome. Transcription produces positive-sense mRNA (step 3a), while replication produces complementary positive-sense RNA with unmodified termini which is subsequently copied back to negative-sense vRNA (step 3b) [73]. The mRNA is transported to the cytoplasm (step 4) for translation. Newly translated envelope proteins (HA, NA and M2) are transported to the plasma membrane via the Golgi apparatus (step 5b). The M2 protein performs a second important role during this process, reducing the acidity of the *trans* Golgi network (TGN) and thus preventing premature conformational rearrangement of HA.

In epithelial cells, the proteins are transported specifically to the apical surface. Translated proteins required to form the viral RNP and matrix layer are transported back to the nucleus (step 5b), where they combine with the replicated vRNA. The virion core is exported from the nucleus and then travels to the cell surface, where it combines with patches of envelope proteins which have begun the budding process (the envelope proteins are sufficient to allow budding off from the membrane [74]) (step 6).

1.3.4 Host cell exit

Following budding, viral particles must be able to travel away from the cell to proliferate the infection. The NA protein, a homotetrameric glycoprotein spike (~ 16 nm in length measured from the envelope surface [68]), cleaves sialic acid residues from host cell surface glycoproteins, preventing HA from reattaching the exiting virion to the same host cell. In budding filamentous virions, NA proteins have been observed to gather at the host cell-proximal end of the filament, providing optimal positioning for sialidase activity [70]. NA inhibitors, such as oseltamivir (Tamiflu) and zanamivir (Relenza), have been some of the most effective anti-influenza drugs.

Chapter 2

Theory and Methods

2.1 Introduction

The primary technique that has been utilised for the research reported in this thesis is coarse-grained molecular dynamics (CG MD). MD is a computational technique which allows the motions of atoms and molecules to be simulated by numerically solving Newton’s equations of motion, using a molecular mechanics force field to describe the interatomic forces. CG force fields allow for simulations of important biological processes which occur on much larger length and time scales than are accessible through traditional atomistic MD. This chapter first introduces MD from the perspective of atomistic simulations, and later describes the CG force fields used in this thesis. A number of general texts are available which give useful overviews of computer simulation techniques, including those by Leach [75], Allen and Tildesley [76], and Frenkel and Smit [77].

2.2 Molecular mechanics force fields

Molecular mechanics force fields describe the potential energy function $\mathcal{V}$ of a set of particles in terms of their coordinates $\mathbf{r}$. The interactions between particles are generally split into two main categories: bonded (intramolecular) forces and non-bonded (intermolecular) forces. In most commonly used force fields, bonded interactions are described with three terms, for bond stretching, opening and closing of bond angles, and rotation about single bonds (torsion). Non-bonded interactions can be described with two terms, for van der Waals forces and

electrostatic interactions. These interactions are represented in Fig. 2.1.

A complete atomistic force field is defined by the functional forms of the contributing energetic terms, and the values of all parameters for these terms. A number of general force fields have been developed for the simulation of biomolecules, such as GROMOS96 [78], CHARMM [79–81], AMBER [82] and OPLS [83, 84]. The energetic terms used in each case are similar, and are described in more detail in the following text.

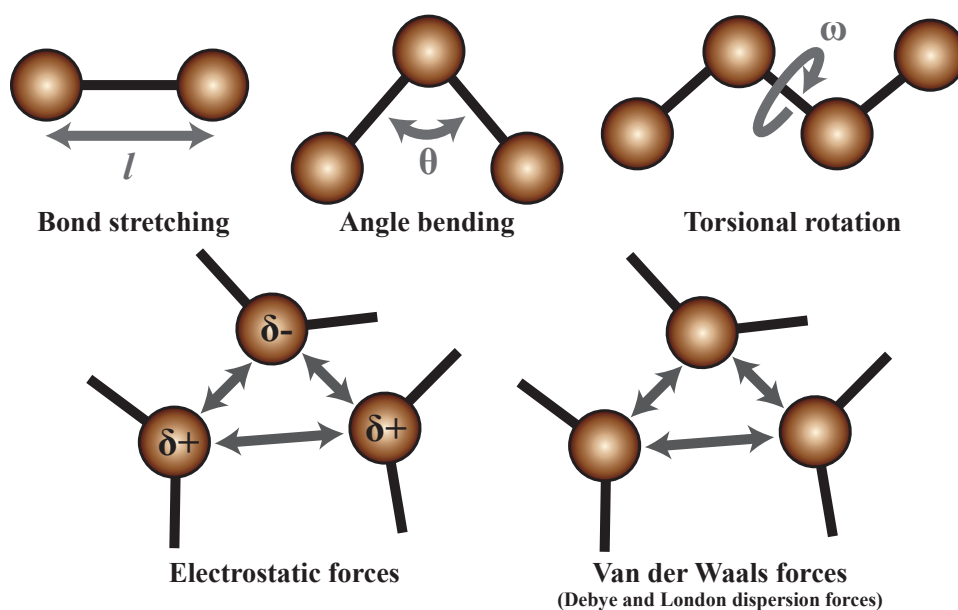


Figure 2.1: Schematic representation of the main components of a molecular mechanics force field.

2.2.1 Bonded interactions

For efficient computation, the individual components of the potential energy description are frequently chosen to be simple functions. Harmonic potentials are thus often used to describe bonded interactions. For bond stretching, a harmonic function for the potential energy $v(l)$ dependent on the bond length l would take the general form:

$$v(l) = \frac{k}{2}(l - l_0)^2 \quad (2.1)$$

The stretching constant k and the reference bond length l_0 define the nature of the potential. The reference bond length is the value associated with the minimum energy in an isolated

system, i.e. without other force field terms. A graphical representation of the potential is shown in Fig. 2.2.

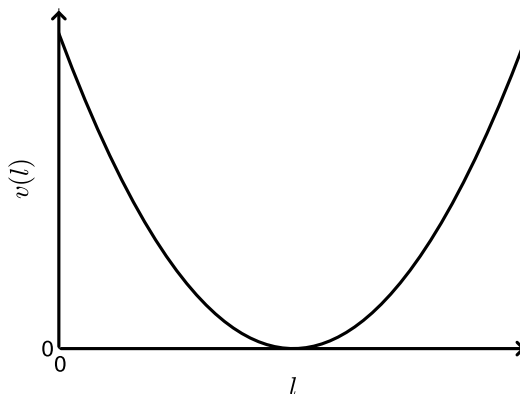


Figure 2.2: Harmonic function for the potential energy $v(l)$ of a bond in terms of its length l .

Bond angles, i.e. the angle between three bonded atoms, are also frequently described using a harmonic potential:

$$v(\theta) = \frac{k}{2}(\theta - \theta_0)^2 \quad (2.2)$$

Torsional potentials are used to describe energetic barriers to rotation about covalent bonds, and are typically applied to groups of four, linearly bonded atoms. For example, the staggered conformation of ethane, where the hydrogen atoms bonded to one carbon are rotated by 60° relative to those on the other carbon, is more stable than the eclipsed conformation in which antibonding interactions occur between the hydrogens on the opposite carbon atoms. Fig. 2.3A shows the two conformations schematically with Newman projections. Torsional potentials are usually expressed as a cosine function, for example:

$$v(\omega) = \frac{V}{2}[1 + \cos(n\omega - \gamma)] \quad (2.3)$$

ω is the torsion angle, V is the height of the barrier to rotation, γ (the phase factor) determines the location of the energy minima, and n (the multiplicity) gives the number of minima. A graphical example for ethane is shown in Fig. 2.3B. For cases where a number of different barrier heights are required, a series of cosine functions with varying parameters can be

applied and summed together. This is important for examples such as the rotation of the central carbon-carbon bond in *n*-butane, for which the anti and gauche conformations result in energy wells with different depths.

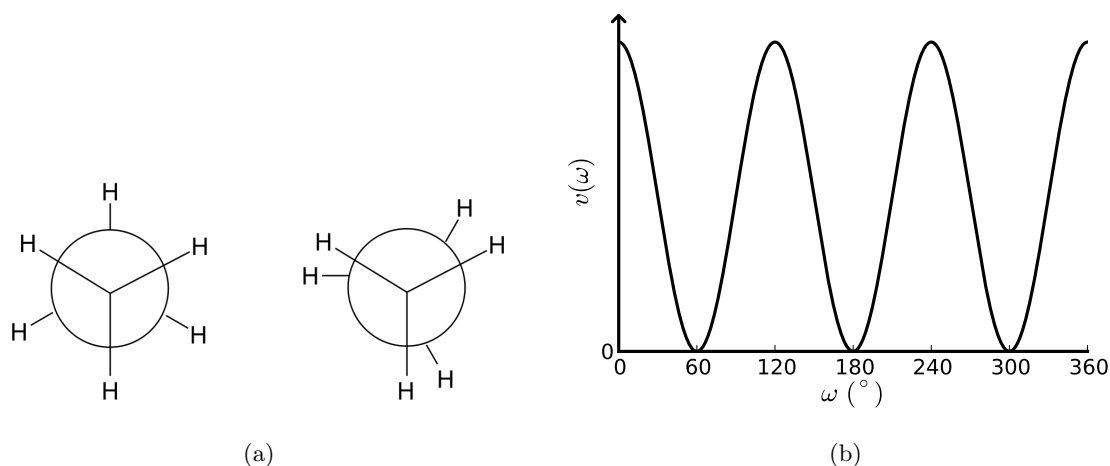


Figure 2.3: (a) Newman projections of ethane, showing the staggered conformation (left) and eclipsed conformation (right). (b) Cosine function describing the potential energy for rotation of a bond such as the carbon-carbon bond in ethane.

Improper torsional potentials are required to preserve certain important molecular features such as ring planarity and stereochemistry. In general, these functions prevent bonds from moving out of a certain plane. They can be formulated as cosine functions, similar to ‘proper’ torsional potentials, but where the atoms are not bonded in a linear sequence. Alternatively, harmonic functions can be used to prevent out-of-plane bending.

2.2.2 Non-bonded interactions

Interactions between non-covalently bound atoms are due to a number of factors. In all cases, atoms are repelled at short ranges due to the Pauli exclusion principle. At longer ranges, electrostatic effects become dominant. In common applications of MD simulations, electrostatic charge in molecules arises from either ionization or from chemical polarity. Molecular dipole moments¹ can be permanent (e.g. if there is a difference in electronegativity between two covalently bonded atoms), instantaneous (caused by fluctuations in electron density which

¹The effects of chemical polarity are described here in terms of dipoles, but in all cases the discussion can also be applied to multipoles.

can temporarily concentrate charge within areas of a molecule or atom) or induced (when any other type of dipole or ion repels the electrons in an area of a molecule or atom, resulting in a dipole moment).

Most commonly used force fields employ a two-component scheme to represent these interactions. An electrostatic term is used to describe interactions between permanently charged atoms, i.e. those with charge derived from ionization or a permanent molecular dipole moment. The van der Waals term represents the remaining dispersive forces, which involve the creation of induced dipoles by either ions, permanent dipoles or instantaneous dipoles.²

The electrostatic term is most commonly represented with Coulomb's law:

$$v_{ij} = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (2.4)$$

v_{ij} is the potential energy between two point charges i and j . It is dependent upon the values of the charges q_i and q_j , and the distance between them r_{ij} . ϵ_0 is the vacuum permittivity (8.854×10^{-12} F m⁻¹). The point charges for a polar molecule can be derived from quantum mechanical simulations, but are often adjusted using empirical data from atomistic simulations or experiment. It is worth noting that a point charge does not have to be associated with a particular atom. For example, the quadrupole of molecular nitrogen can be represented with a partial positive charge in the centre of the N-N bond and partial negative charges at the nitrogen atoms.

The van der Waals term is usually represented by the Lennard-Jones 12-6 potential:

$$v(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (2.5)$$

r is the distance between the interacting atoms, σ is the finite distance at which the potential energy is zero, and ϵ is the depth of the potential well. The function can also be expressed in terms of the distance at which the potential passes through a minimum r_m :

$$v(r) = \epsilon \left[\left(\frac{r_m}{r} \right)^{12} - 2 \left(\frac{r_m}{r} \right)^6 \right] \quad (2.6)$$

²Permanent dipole-induced dipole interactions are also known as Debye forces, and instantaneous dipole-induced dipole interactions are also known as London dispersion forces.

The function is shown graphically in Fig. 2.4. The inverse sixth power function models long-range attraction between atoms due to dispersive forces, and is related to theoretical treatments. The inverse twelfth power function models the short-range repulsion, but is not directly related to theory and an exponential form may be more accurate (as used, for example, in the Buckingham potential). However, the Lennard-Jones potential is widely used due to its ease of computation; the inverse twelfth power term can be rapidly calculated by squaring the inverse sixth power term. In atomistic molecular dynamics simulations of biomolecules, the use of the potential is ubiquitous. As for the Coulomb term, the parameters can be derived from quantum mechanical calculations or experimental data.

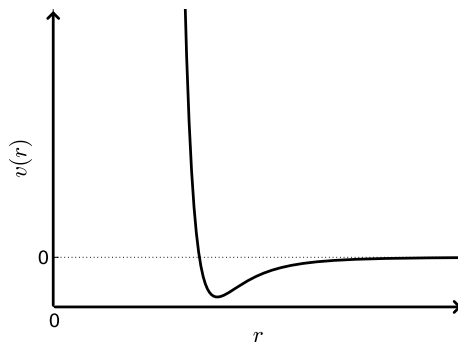


Figure 2.4: The Lennard-Jones potential.

2.3 Molecular dynamics simulation methods

2.3.1 Introduction

The potential functions discussed in the previous section form the basis for most force fields commonly used for biomolecular simulations. They can be summarized in the following equation, which represents the summation of all potentials in a system of N particles with coordinates $\mathbf{r}$:

$$\begin{aligned}
 \mathcal{V}(\mathbf{r}) = & \sum_{\text{bonds}} \frac{k_i}{2} (l_i - l_{i,0})^2 + \sum_{\text{angles}} \frac{k_i}{2} (\theta_i - \theta_{i,0})^2 + \sum_{\text{torsions}} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) + \\
 & \sum_{i=1}^N \sum_{j=i+1}^N \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right)
 \end{aligned} \tag{2.7}$$

The force $\mathbf{F}_i$ on an atom i can be calculated through differentiation of the potential:

$$\mathbf{F}_i = \frac{\delta \mathcal{V}(\mathbf{r}_1, \dots, \mathbf{r}_N)}{\delta \mathbf{r}_i} \tag{2.8}$$

In molecular dynamics simulations, a trajectory is calculated by solving the differential equation embodied in Newton's second law:

$$\frac{d^2 \mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i} \tag{2.9}$$

In almost all practical applications of MD simulations the n -body problem prohibits an analytical solution, so numerical *finite difference* methods must be used.

2.3.2 Numerical integration of Newton's second law

A number of finite difference integration methods are available for MD simulations, all of which rely the ability to approximate the positions and dynamic properties with Taylor series expansions:

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t) + \frac{1}{2} \delta t^2 \mathbf{a}(t) + \frac{1}{6} \delta t^3 \mathbf{b}(t) + \dots \tag{2.10}$$

$$\mathbf{v}(t + \delta t) = \mathbf{v}(t) + \delta t \mathbf{a}(t) + \frac{1}{2} \delta t^2 \mathbf{b}(t) + \dots \tag{2.11}$$

$$\mathbf{a}(t + \delta t) = \mathbf{a}(t) + \delta t \mathbf{b}(t) + \dots \tag{2.12}$$

$\mathbf{v}$ is the velocity, $\mathbf{a}$ the acceleration and $\mathbf{b}$ the third derivative. The *Verlet* algorithm combines the Taylor expansions for the positions at time t and at the previous timestep ($t - \delta t$):

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t) + \frac{1}{2} \delta t^2 \mathbf{a}(t) + \dots \tag{2.13}$$

$$\mathbf{r}(t - \delta t) = \mathbf{r}(t) - \delta t \mathbf{v}(t) + \frac{1}{2} \delta t^2 \mathbf{a}(t) - \dots \tag{2.14}$$

Addition of the equations results in the following expression:

$$\mathbf{r}(t + \delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \delta t) + \delta t^2 \mathbf{a}(t) \quad (2.15)$$

The Verlet algorithm thus calculates the new positions of the system using the positions at time t and at the previous timestep, and the accelerations at time t . The accelerations can be calculated by dividing the forces by the relevant atomic masses. The velocity is not explicitly included, but can be calculated as follows:

$$\mathbf{v}(t) = [\mathbf{r}(t + \delta t) - \mathbf{r}(t - \delta t)]/2\delta t \quad (2.16)$$

The *leap-frog* algorithm is used in the GROMACS software [85–88], which was used for all of the simulations reported in this thesis. It addresses some of the disadvantages of the Verlet algorithm. It involves the following expressions:

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t + \tfrac{1}{2}\delta t) \quad (2.17)$$

$$\mathbf{v}(t + \tfrac{1}{2}\delta t) = \mathbf{v}(t - \tfrac{1}{2}\delta t) + \delta t \mathbf{a}(t) \quad (2.18)$$

The name of the algorithm is derived from the way in which the velocities and positions are calculated at staggered intervals, ‘leap-frogging’ over each other.

2.3.3 Periodic boundary conditions

Periodic boundary conditions (PBC) are used to simulate bulk systems using a relatively small number of particles. In the case of a cubic simulation box, a particle that moves outside of the box at one face is instead shifted to the opposite box face with the same velocity. The system can thus be considered as a periodic array composed of infinite copies of the simulation box. Intramolecular and intermolecular forces are evaluated by applying the minimum image convention, which ensures that a particle i interacts only with the closest periodic image of any other particle j (Fig. 2.5).

It is important to note that structural fluctuations on scales larger than the box size will not be observable. For example, long wavelength undulations in bilayers are suppressed in

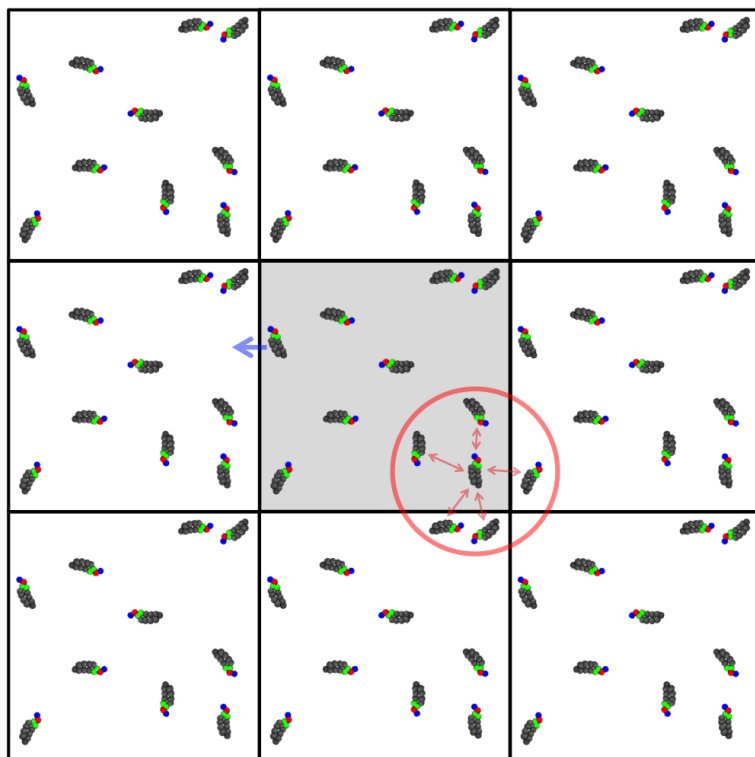


Figure 2.5: Diagram illustrating periodic boundary conditions (PBC) and how this feature operates with the use of a cutoff for long-range interactions. The lipid shown is a coarse-grained model of DPPC. The red ring represents a cutoff for long-range intermolecular interactions for a particular lipid. The red arrows show which of the surrounding lipids interact, determined using the minimum image convention. The blue arrow indicates how a lipid which leaves one side of a box is then placed on the opposite side.

simulations of small patches of membrane (although the effect on bilayer properties has been shown to be negligible [89]).

Cubic boxes are the simplest and most common type used in MD simulations, but others are available. The rhombic dodecahedron and truncated octahedron are particularly useful in simulations of systems with an overall spherical topology (Fig. 2.6). They are both a much closer approximation to a sphere than a cube; in simulations with explicitly modelled solvent they thus reduce the amount of bulk solvent required. The rhombic dodecahedron box was thus used in the simulations of vesicles described in Chapters 4, 5 and 6.

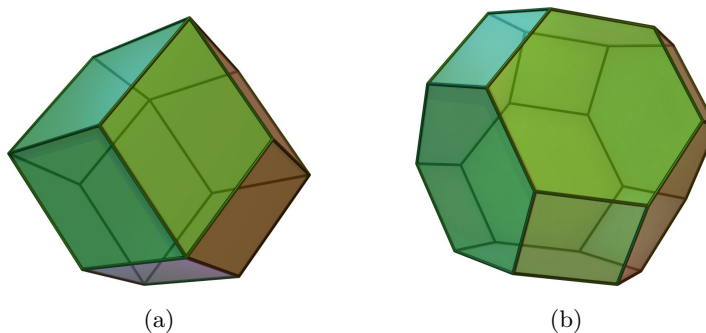


Figure 2.6: (a) Rhombic dodecahedron. (b) Truncated octahedron.

2.3.4 Treatment of long-range intermolecular forces

In theory, all non-bonded atoms within a simulation box should be able to interact with each other via intermolecular forces. This makes the electrostatics and van der Waals terms of the force field challenging to compute, because the total number of pairwise interactions scales with $\mathcal{O}(N^2)$. In practice, however, some detail of long-range intermolecular forces can be sacrificed, improving efficiency without losing important physical detail.

Firstly, it can be seen from Fig. 2.4 that the Lennard-Jones potential quickly converges towards 0. For example, at 2.5 times the equilibrium distance the potential is less than 1% of the well depth:

$$v_{LJ}(2.5r_m) = \epsilon \left[\left(\frac{r_m}{2.5r_m} \right)^{12} - 2 \left(\frac{r_m}{2.5r_m} \right)^6 \right] = -0.0082\epsilon \quad (2.19)$$

It is thus possible (and common in biomolecular simulations) to truncate the Lennard-Jones potential beyond a certain cut-off radius, greatly improving computational performance.

The Coulomb potential converges to 0 much more slowly than the Lennard-Jones potential, reducing with the inverse of the distance rather than the inverse of the sixth power of the distance. Long-range electrostatic forces can thus have significant effects on MD simulations, as demonstrated in studies of proteins, nucleotides and lipid bilayers [90–92]. Therefore, it is often inappropriate to simply truncate electrostatic forces beyond a certain cut-off. Instead a number of specialized methods have been developed, most of which perform an explicit evaluation of electrostatic forces within a cut-off and a less detailed implicit evaluation at longer ranges. The particle-mesh-Ewald (PME) method [93] is the most commonly used in biomolecular simulations and is discussed in the following text. The reaction field method

is also used in some biomolecular simulations (a popular implementation: [94]), and the Gaussian split Ewald method is another example [95].

2.3.4.1 The particle-mesh-Ewald Method

Ewald summation was first introduced as a method for the calculation of long-range electrostatic interactions in periodic crystals [96]. Firstly, a periodic array made from a box of N particles is considered. The total electrostatic energy is:

$$V = \frac{1}{2} \sum_{n_x} \sum_{n_y} \sum_{n_z^*} \sum_i^N \sum_j^N \frac{q_i q_j}{4\pi\epsilon_0 \mathbf{r}_{ij,\mathbf{n}}} \quad (2.20)$$

$\mathbf{n} = (n_x, n_y, n_z)$ is the vector of indices for the periodic boxes; the asterisk indicates that particles inside the simulation box can interact with their periodic copies, but not with themselves, i.e. terms with $i = j$ should be omitted when $(n_x, n_y, n_z) = (0, 0, 0)$. Furthermore, the distance between particles r_{ij} is the real distance, not the minimum image. The function converges slowly to a total potential. Ewald summation reformulates this as two quickly-converging terms (a direct sum V_{dir} and a reciprocal (Fourier) space sum V_{rec}) and a constant term [97]:

$$V = V_{\text{dir}} + V_{\text{rec}} + V_0 \quad (2.21)$$

$$V_{\text{dir}} = \frac{1}{2} \sum_{i,j}^N \sum_{n_x} \sum_{n_y} \sum_{n_z^*} q_i q_j \frac{\text{erfc}(\beta r_{i,j,\mathbf{n}})}{4\pi\epsilon_0 r_{i,j,\mathbf{n}}} \quad (2.22)$$

$$V_{\text{rec}} = \frac{1}{8\pi^2\epsilon_0 V} \sum_{i,j}^N q_i q_j \sum_{m_x} \sum_{m_y} \sum_{m_z^*} \frac{\exp(-(\pi\mathbf{m}/\beta)^2 + 2\pi i\mathbf{m} \cdot (\mathbf{r}_i - \mathbf{r}_j))}{\mathbf{m}^2} \quad (2.23)$$

$$V_0 = -\frac{\beta}{4\pi^{\frac{3}{2}}\epsilon_0} \sum_i^N q_i^2 \quad (2.24)$$

$\text{erfc}()$ is the complementary error function, $\mathbf{m} = (m_x, m_y, m_z)$, and β is a parameter that determines the relative weighting of the direct and reciprocal sums. Thus evaluation of the direct sum can be truncated beyond a certain cut-off (e.g. 1 nm), as well as the reciprocal sum (e.g. 10 wave vectors in each direction). However, the reciprocal term still scales as $\mathcal{O}(N^2)$. The particle-mesh-Ewald method solves this problem by re-distributing the actual particle charges onto a finite grid using cardinal B-spline interpolation. The grid is then

transformed into reciprocal space using a fast Fourier transform algorithm, allowing V_{rec} to be determined with a single sum over the grid in k -space. The PME method scales with $\mathcal{O}(N \cdot \log(N))$, and is thus substantially more efficient than ordinary Ewald summation or explicit pairwise summation with Coulomb’s law.

2.3.5 Temperature and pressure control

Under the assumption of the ergodic hypothesis, i.e. that time averages are equal to ensemble averages, a MD simulation reproduces a constant- NVE (“microcanonical”) ensemble. A MD simulation should thus in principle keep the total energy of a system constant, although numerical integration errors can cause the energy to fluctuate or drift. Interchange of potential and kinetic energy also causes the temperature to fluctuate or drift. In many cases it is preferable to keep the system temperature constant, rather than the total energy, which can be achieved with a number of thermostat algorithms. Similarly, a number of barostat algorithms are available for control of the system pressure. This allows for simulations in the constant- NVT and constant- NPT ensembles. The latter is particularly useful for biomolecular simulations because it minimizes the Gibbs free energy, which governs many important biological processes.

2.3.5.1 The Berendsen thermostat

The temperature of a simulation system is related to the time average of the kinetic energy, as follows:

$$\langle \mathcal{K} \rangle_{NVT} = \frac{1}{2} \sum_{i=1}^N m_i v_i^2 = \frac{3}{2} N k_B T \quad (2.25)$$

It is therefore possible to re-scale the velocities to control temperature. If the temperature at time t is $T(t)$ and a scaling factor λ is used to multiply the velocities, the necessary temperature change can be determined as follows:

$$\Delta T = \frac{1}{2} \sum_{i=1}^N \frac{2}{3} \frac{m_i (\lambda v_i)^2}{N k_B} - \frac{1}{2} \sum_{i=1}^N \frac{2}{3} \frac{m_i v_i^2}{N k_B} \quad (2.26)$$

$$\Delta T = (\lambda^2 - 1) T(t) \quad (2.27)$$

$$\lambda = \sqrt{T_{new}/T(t)} \quad (2.28)$$

However, one-step re-scaling of velocities can be problematic. The Berendsen thermostat maintains the temperature by coupling the system to an external heat bath fixed at the desired temperature T_{bath} [98]. This involves scaling the velocities at each time step such that the rate of change of temperature is proportional to the difference in temperature between the bath and the system:

$$\frac{dT(t)}{dt} = \frac{1}{\tau} (T_{bath} - T(t)) \quad (2.29)$$

τ is an adjustable parameter which determines the strength of the coupling between the bath and the system. The change in temperature between successive time steps is:

$$\Delta T = \frac{\delta t}{\tau} (T_{bath} - T(t)) \quad (2.30)$$

The scaling factor λ can thus be determined:

$$\lambda^2 = 1 + \frac{\delta t}{\tau} \left(\frac{T_{bath}}{T(t)} - 1 \right) \quad (2.31)$$

A large value of τ results in weak coupling, a small value represents strong coupling, and if $\tau = \delta t$ then the algorithm is equivalent to the direct velocity scaling method described earlier. Other schemes for temperature coupling include the Langevin thermostat [77], the stochastic collisions algorithm [75] and the Nosé-Hoover method [99, 100].

2.3.5.2 The Berendsen barostat

The pressure in a MD simulation can be defined by using the virial W :

$$W = \sum x_i \frac{\partial \mathbf{p}}{\partial x_i} = -3Nk_B T \quad (2.32)$$

This can be rewritten in terms of the pressure and the interaction potential $v(r)$:

$$W = -3PV + \sum_{i=1}^N \sum_{j=i+1}^N r_{ij} \frac{dv(r_{ij})}{dr_{ij}} = -3Nk_B T \quad (2.33)$$

The potential energy derivatives correspond to the interatomic forces $\mathbf{f}$:

$$P = \frac{1}{V} \left[Nk_B T - \frac{1}{3} \sum_{i=1}^N \sum_{j=i+1}^N r_{ij} f_{ij} \right] \quad (2.34)$$

The pressure can therefore be controlled by re-scaling the particle positions r and the simulation box size. This is equivalent to the control of temperature via direct re-scaling of velocities, as described earlier. The Berendsen barostat couples the system pressure to a ‘pressure bath’, analogous to the Berendsen thermostat [75, 98]:

$$\lambda = 1 - \kappa \frac{\delta t}{\tau_P} (P - P_{bath}) \quad (2.35)$$

$$\mathbf{r}'_i = \lambda^{1/3} \mathbf{r}_i \quad (2.36)$$

$\mathbf{r}'_i$ represents the re-scaled coordinates. The strength of the coupling to the ‘pressure bath’ can be adjusted using the time constant τ_P and the isothermal compressibility κ , which is related to the magnitude of volume fluctuations as follows [75]:

$$\kappa = -\frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_T = \frac{1}{k_B T} \frac{\langle V^2 \rangle - \langle V \rangle^2}{\langle V \rangle} \quad (2.37)$$

The system dimensions can be re-scaled equally and simultaneously (isotropic pressure coupling), or independently (anisotropic pressure coupling). An important technique for the simulation of bilayers is semiisotropic pressure coupling, in which the x and y dimensions (the bilayer plane) are re-scaled together, and independent from the z dimension (the bilayer normal). This allows for proper equilibration of the lateral pressure profile of the bilayer, facilitating self-assembly from randomly placed lipids.

Alternative pressure coupling algorithms include extended system methods, such as the Parrinello-Rahman scheme [77].

2.4 Coarse-grained force fields

Molecular dynamics simulations of biomolecular systems have resulted in profound new insights into many biological processes. However, many important biological processes occur on much larger time and length scales than are accessible through traditional atomistic simulations. Coarse-grained (CG) models address this problem by combining groups of atoms into CG particles. The consequent reduction in the number of degrees of freedom can greatly increase the time and length scales attainable through MD simulation.

The simulations described in Chapters 5 and 6 were conducted with the MARTINI model, and those in Chapters 3 and 4 employed the related model developed by the Sansom group [46]. These CG force fields are described in more detail in this section.

2.4.1 The Bond-Sansom model

2.4.1.1 Original CG lipid model

The CG model developed by Bond et al. [46] was originally derived from an earlier lipid model by Marrink et al. [43]. The atom mapping is based on a 4:1 ratio, i.e. approximately four heavy (non-hydrogen) atoms are mapped to a single CG site (Fig. 2.7). For example, the model for DPPC comprises 12 CG particles: a particle representing the choline moiety ('NC3'), a phosphate particle ('PO4'), two glycerol particles ('GL1' and 'GL2'), and four hydrocarbon particles per palmitoyl chain ('C1A, C2A, C3A, C4A' and 'C1B, C2B, C3B, C4B'). All particles are treated with a mass of 72 Da (equivalent to four water molecules).

Interactions between bonded CG particles are described with two harmonic potentials: a bond-stretching term and a bond angle term. The distance term has a reference bond length l_0 of 4.7 Å and a force constant k of 12.5 kJ mol⁻¹ Å⁻². An angle potential is applied between every three consecutively bonded particles, with a reference bond angle θ_0 of 180° (120° in the case of the PO4-GL1-GL2 angle) and a force constant k of 25 kJ mol⁻¹ rad⁻². A torsional angle term is not included.

Intermolecular interactions are described with a Coulomb potential term and a Lennard-Jones term, as commonly applied in atomistic simulations. However, only groups of atoms with full charges (i.e. a magnitude of > 1) are included in the electrostatics term. Taking DPPC

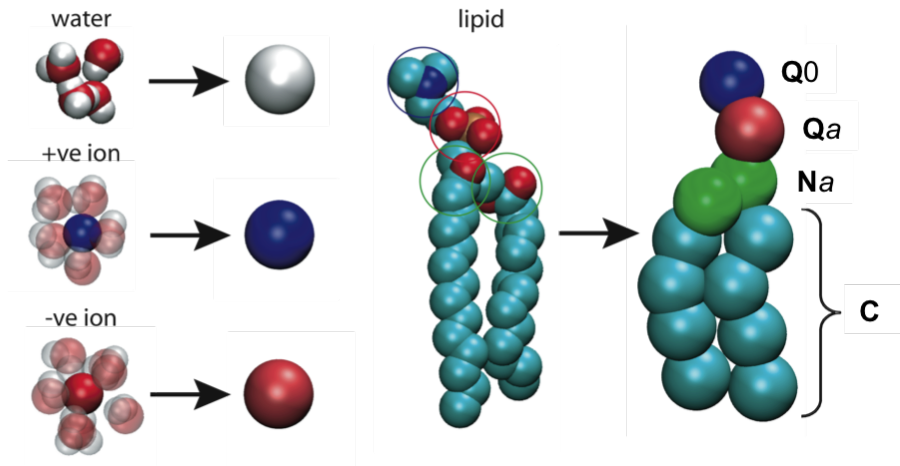


Figure 2.7: Coarse-grained mapping of water, solvated ions and DPPC. The particle types of DPPC are annotated.

as the example, this would include the positively charged choline group and the negatively charged phosphate group. The Lennard-Jones term accounts for all other interactions: those arising from partial charges associated with permanent dipoles and multipoles, as well as dispersive forces. The Coulombic potential is shifted smoothly to zero between 0 Å and 12 Å, and a long-range electrostatic approximation is usually not included. A relative dielectric constant ϵ_r of 20 is used, which was chosen during the parameterization process to reproduce empirical system properties. ϵ_r is incorporated into Coulomb's law as follows:

$$v_{ij} = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \quad (2.38)$$

There are four main types of CG particle defined for Lennard-Jones interactions: polar (**P**), nonpolar (**N**), apolar (**C**), and charged (**Q**). A typical polar particle represents a group of atoms which have partial charges and are hydrophilic, whilst apolar particles generally represent groups of neutral, hydrophobic atoms. Nonpolar particles represent a mix of both polar and apolar atoms, and are further divided into four subtypes to account for hydrogen bonding: *d* (hydrogen bond donor), *a* (hydrogen bond acceptor), *da* (hydrogen bond donor and acceptor), and 0 (no hydrogen bonding capability). Charged particles represent groups with a full charge and are also divided into four subtypes to account for hydrogen bonding.

Five interaction levels are defined (ϵ is the potential well depth): attractive (I, $\epsilon = 5$ kJ mol⁻¹), semiattractive (II, $\epsilon = 4.2$ kJ mol⁻¹), intermediate (III, $\epsilon = 3.4$ kJ mol⁻¹), semirepulsive (IV, $\epsilon = 2.6$ kJ mol⁻¹), and repulsive (V, $\epsilon = 1.8$ kJ mol⁻¹). It should be noted

that all interaction levels involve a potential well and are thus all ultimately attractive when considered in isolation. The terms “attractive” and “repulsive” are meant to describe the relative magnitudes of the well depths. All possible interactions between the different particle types are shown in Table 2.1. The parameters were calibrated using an alkane/water system, with the goal of reproducing the experimental densities of pure water and alkane systems at room temperature, the mutual solubility of oil and water, and the relative diffusion rates [43]. σ (the finite distance at which the potential is zero) is 4.7 Å in all cases. The potential is truncated at long ranges, by smoothly shifting the potential to zero between 9 Å and 12 Å.

group	subtype	P	Q				C	N				
			0	<i>d</i>	<i>a</i>	<i>da</i>		0	<i>d</i>	<i>a</i>	<i>da</i>	<i>m</i> *
P		I	I	I	I	I	V	IV	III	III	II	III
Q	0	I	III	III	III	II	V	III	III	III	III	III
	<i>d</i>	I	III	III	II	I	V	III	III	II	II	II
	<i>a</i>	I	III	II	III	I	V	III	II	III	II	II
	<i>da</i>	I	II	I	I	I	V	III	II	II	I	II
C		V	V	V	V	V	III	III	IV	IV	V	IV
N	0	IV	III	III	III	III	III	III	III	III	III	III
	<i>d</i>	III	III	III	II	II	IV	III	II	II	II	II
	<i>a</i>	III	III	II	III	II	IV	III	II	II	II	II
	<i>da</i>	II	III	II	II	I	V	III	II	II	I	II
	<i>m</i> *	III	III	II	II	II	IV	III	II	II	II	II

Table 2.1: Interaction matrix for CG force fields by Marrink et al. [43] and Bond et al. [46].

Five interaction types: I (attractive), II (semi-attractive), III (intermediate), IV (semirepulsive), and V (repulsive).

Four particle types: **P** (polar), **N** (nonpolar), **C** (apolar), and **Q** charged.

Groups **N** and **Q** have further subtypes: *d* (hydrogen bond donor), *a* (hydrogen bond acceptor), *da* (hydrogen bond donor and acceptor), and 0 (no hydrogen bonding capability)

*The **Nm** particle type was introduced for the Bond et al. [46] model.

The CG model greatly increases simulation efficiency by reducing the number of atoms, using short-range non-bonded interactions, and increasing the timestep. The increase in timestep (up to about 40 fs, depending on the simulation details) is possible because of the smoother potential landscape that results from removing high frequency motions associated

with smaller particles. This factor also increases the speed of sampling for many processes, and thus simulation times are sometimes quoted multiplied by 4. This value roughly approximates the increased sampling speed of water and of lipid diffusion in bilayers [101]. However, the simulation times quoted in this thesis have not had the correction factor applied, except where noted.

2.4.1.2 Adapted model for protein simulation

The Marrink et al. [43] lipid force field was adapted for simulation of proteins by Bond and Sansom [102] and later updated by the same group [46]. Each amino acid comprises a single CG backbone particle and up to two particles which represent the side chain. A new type of interaction site ‘ $\mathbf{N}_m$ ’ was parameterized for the backbone particles (see Table 2.1). The side chain particle types are shown in Table 2.2. The assignments were evaluated and refined through comparison with experimental data, based upon water/cyclohexane partitioning free energies for side chains (Fig. 2.8), as well as PMF calculations for insertion of helices into bilayers [46].

Bonds between particles were given a force constant of $50 \text{ kJ mol}^{-1} \text{ \AA}^{-2}$. The reference bond lengths were derived from an ensemble of protein structures from the PDB database. Thus, the length of bonds between backbone particles was derived from the mean $\text{C}\alpha$ distances, which resulted in a value of $3.6 \text{ \AA}$. Bond lengths for the side chains were chosen by evaluating the mean of the distances between $\text{C}\alpha$ atoms and the centres of mass of the relevant heavy side chain atoms for each CG particle. Arg and Lys have two side chain particles bonded consecutively, while the aromatic amino acids (His, Phe, Trp and Tyr) have all three CG particles bonded together in a triangular topology. The side chain bond lengths are shown in Table 2.2.

Bond angles between backbone particles are modelled with a harmonic potential with a force constant of $250 \text{ kJ mol}^{-1} \text{ rad}^{-2}$ for residues in random coil regions, and $350 \text{ kJ mol}^{-1} \text{ rad}^{-2}$ for structured regions. Equilibrium bond angles are 90° for α -helical segments, 130° for β -strand segments, and 120° for random coil. Side chain angles for residues with two side chain particles are harmonically restrained with a force constant of $250 \text{ kJ mol}^{-1} \text{ rad}^{-2}$ and an equilibrium angle between particles CA–CB–CG of 180° (Arg/Lys), 54° (Trp) or 60° (Phe, Tyr, His).

Amino Acid	Side chain particles	Bond lengths (Å)		
		CA-CB	CB-CG	CA-CG
Ala	C	0.9		
Arg	Nd+Qd (charge +1)	3.7	2.5	
Asn	P	2.6		
Asp	Qa (charge -1)	2.6		
Cys	Nd	1.7		
Glu	Qa (charge -1)	3.6		
Gln	P	3.6		
Gly	none			
His	Nd + Nda	3.8	3.8	3.8
Ile	C	2.6		
Leu	C	2.8		
Lys	C+Qd (charge +1)	3.1	2.1	
Met	Nd	3.1		
Phe	C+C	4.2	4.2	4.2
Pro	C	1.9		
Ser	Nda	1.6		
Thr	Nda	2.0		
Trp	C+Nd	4.7	2.8	3.8
Tyr	C+P	3.8	3.8	3.8
Val	C	2.0		

Table 2.2: Side chain particles and bond lengths for the Bond et al. [46] CG force field. ‘CA’, ‘CB’ and ‘CG’ represent the backbone particle, and first and second side chain particles respectively.

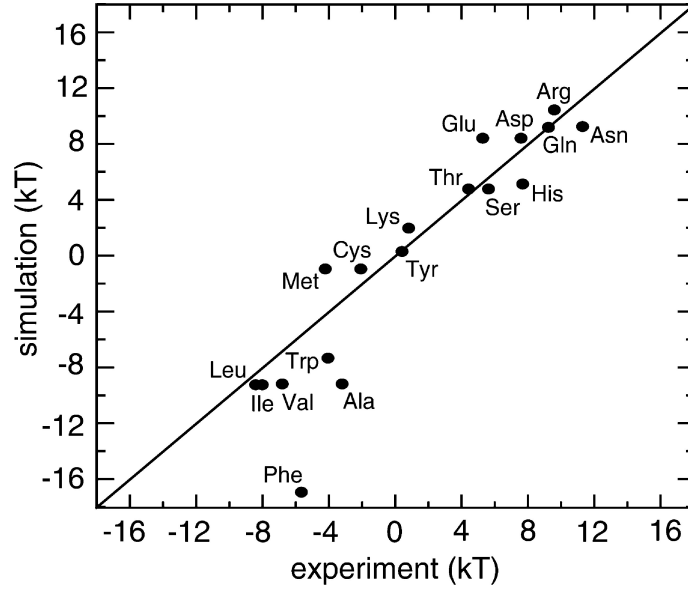


Figure 2.8: Water to cyclohexane transfer free energies of neutral CG amino acid side chain analogues derived via thermodynamic integration. The free energies are compared to experimental transfer free energies from water to cyclohexane [103]. Figure adapted from [46].

Further harmonic distance restraints can be used to conserve secondary structure, through two different schemes. The first applies only to α -helices, in which the restraints are placed between backbone particles four residues apart (with the exception of prolines), mimicking intrahelix hydrogen-bonds. The restraints are given a force constant of $10 \text{ kJ mol}^{-1} \text{ \AA}^{-2}$ and an equilibrium length of $6 \text{ \AA}$. The other distance restraint scheme can be used to conserve secondary structure in much larger and more complex proteins. It involves adding restraints between all backbone particles which are within a given cutoff and not directly bonded. The equilibrium lengths are matched to the inter-particle distances in the original structure.

2.4.2 The MARTINI model

The Marrink et al. [43] CG lipid force field was updated by the same group and named the MARTINI force field [44], and later extended to include protein parameters [45]. The MARTINI model includes a larger numbers of particle subtypes and Lennard-Jones interactions than the original force field. The **C** and **P** particle types are each divided into five subtypes, 1–5, with increasingly polar character. The original five interaction types are doubled to ten, while spanning approximately the same range of values. The interaction matrix is shown in Table 2.3. The same value for the Lennard-Jones potential parameter σ is used ($4.7 \text{ \AA}$), except for a few special cases, such as ring particles and ‘antifreeze’ particles (which are discussed in more detail below). The parameters for interactions between charged (**Q**) and the most apolar types (**C**₁ and **C**₂) are also increased to $6.2 \text{ \AA}$ to make it more favourable for charged particles to retain their hydration shells within an apolar medium. The relative dielectric constant ϵ_r for the Coulombic potential is also reduced from 20 to 15, helping to balance the increased hydration strength of the altered CG particle types.

	Q				P					N				C				
	da	d	a	0	5	4	3	2	1	da	d	a	0	5	4	3	2	1
Q ^{da}	0	0	0	II	0	0	0	I	I	I	I	I	IV	V	VI	VII	IX	IX
Q ^d	0	I	0	II	0	0	0	I	I	I	III	I	IV	V	VI	VII	IX	IX
Q ^a	0	0	I	II	0	0	0	I	I	I	I	III	IV	V	VI	VII	IX	IX
Q ⁰	II	II	II	IV	I	0	I	II	III	III	III	III	IV	V	VI	VII	IX	IX
P ₅	0	0	0	I	0	0	0	0	0	I	I	I	IV	V	VI	VI	VII	VIII
P ₄	0	0	0	0	0	I	I	II	II	III	III	III	IV	V	VI	VI	VII	VIII
P ₃	0	0	0	I	0	I	I	II	II	II	II	II	IV	IV	V	V	VI	VII
P ₂	I	I	I	II	0	II	II	II	II	II	II	II	III	IV	IV	V	VI	VII
P ₁	I	I	I	III	0	II	II	II	II	II	II	II	III	IV	IV	IV	V	VI
N ^{da}	I	I	I	III	I	III	II	II	II	II	II	II	IV	IV	V	VI	VI	VI
N ^d	I	III	I	III	I	III	II	II	II	II	III	II	IV	IV	V	VI	VI	VI
N ^a	I	I	III	III	I	III	II	II	II	II	II	III	IV	IV	V	VI	VI	VI
N ⁰	IV	IV	IV	IV	IV	IV	IV	III	III	IV	IV	IV	IV	IV	IV	IV	V	VI
C ₅	V	V	V	V	V	V	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	V	V
C ₄	VI	VI	VI	VI	VI	VI	V	IV	IV	V	V	V	IV	IV	IV	IV	V	V
C ₃	VII	VII	VII	VII	VI	VI	V	V	IV	VI	VI	VI	IV	IV	IV	IV	IV	IV
C ₂	IX	IX	IX	IX	VII	VII	VI	VI	V	VI	VI	V	IV	V	V	IV	IV	IV
C ₂	IX	IX	IX	IX	VIII	VIII	VII	VII	VI	VI	VI	VI	VI	V	V	IV	IV	IV

Table 2.3: Interaction matrix for the MARTINI force field [44, 45].

Lennard-Jones parameters: 0, $\epsilon = 5.6$ kJ mol⁻¹; I, $\epsilon = 5.0$ kJ mol⁻¹; II, $\epsilon = 4.5$ kJ mol⁻¹; III, $\epsilon = 4.0$ kJ mol⁻¹; IV, $\epsilon = 3.5$ kJ mol⁻¹; V, $\epsilon = 3.1$ kJ mol⁻¹; VI, $\epsilon = 2.7$ kJ mol⁻¹; VII, $\epsilon = 2.3$ kJ mol⁻¹; VIII, $\epsilon = 2.0$ kJ mol⁻¹; IX, $\epsilon = 2.0$ kJ mol⁻¹ ($\sigma = 6.2$ nm).

Ring particles are introduced in the MARTINI model to better represent small ring structures. Instead of the usual 4:1 mapping, a 2:1 or 3:1 mapping is used. Modifications of the normal particle types are used (e.g. **SC₁**), which have reduced σ parameters of 4.3 Å and well-depths (ϵ) scaled to 75% of their original values. Rigid constraints are typically used in place of harmonic bonds, to preserve ring planarity and avoid fast oscillations (which would be introduced by harmonic bonds with very high force constants). Improper dihedrals can also be used to preserve ring planarity. The ring particles help to represent structures such as aromatic amino acid sidechains and sterols. The latter example is particularly important for the studies described in Chapters 5 and 6, which involve the simulation of cholesterol-containing membranes. Fig. 2.9 depicts the MARTINI structure of cholesterol.

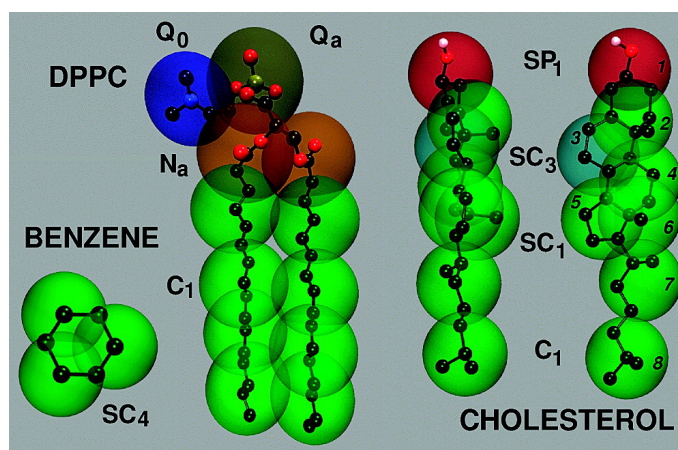


Figure 2.9: CG mappings of benzene, DPPC and cholesterol in the MARTINI force field. The ‘S...’ particles are ring-type particles. Figure adapted from [44].

‘Antifreeze’ particles help to avoid unwanted freezing of CG water, which has a melting temperature slightly too high compared to real water (between 280 and 300 K depending on the simulation conditions). They interact in the same way as normal water particles (type **P₄**), but with an increased Lennard-Jones parameter σ of 5.7 Å. The increased effective size helps to break up lattice packing in the otherwise uniformly sized water particles. A small mole fraction of 0.05 is effective in most cases. The introduction of antifreeze particles (at a mole fraction of 0.1) was found to have no effect on a number of macroscopic properties of a simulated DPPC bilayer [44].

The extension of the CG force field to proteins [45] was validated by comparing a number of simulation results with atomistic models, such as the potentials of mean force (PMFs) for

each amino acid as a function of distance from the centre of a DOPC bilayer. The partitioning of a number of model peptides into bilayers was also compared with atomistic simulation and experiment. Compared to the Bond et al. [46] CG force field for proteins, the MARTINI model is somewhat more detailed, but sacrifices some computational efficiency. The use of smaller particles for ring structures requires that the simulation timestep be no higher than 20 fs in general. In some cases (such as in the simulations in Chapters 5 and 6) a timestep of 10 fs is necessary — four times less than the 40 fs timestep commonly used with the Bond et al. [46] force field.

Chapter 3

Mechanism of Membrane Permeabilization by the Antimicrobial Peptide Maculatin 1.1

3.1 Introduction

Antimicrobial peptides (AMPs) form an important element of the innate immune system of many organisms. Most AMPs are thought to exert their antibacterial effect through permeabilization of the bacterial cell membrane, rather than by binding to a chiral receptor target. This may account for the low level of resistance to AMPs developed by bacteria [104]. For this reason, there has been interest in the development of both natural and synthetic AMPs as therapeutic agents.

A large group of AMPs form cationic, amphipathic α -helices, including the cecropins [105], magainins [106] and maculatin [107]. Various models have been proposed to explain the mechanism through which these peptides cause membrane permeabilization in bacteria. Of these, the two main models are the pore-forming and the carpet mechanisms (Fig. 3.1) [108]. The pore-forming mechanism describes the insertion and assembly of peptides to form trans-membrane channels, across which water and other cell contents can flow, disrupting the cell.

Within this general mechanism, two models further define the nature of the channel. In the barrel-stave model, a transmembrane pore is formed without significant local disruption of the surrounding bilayer structure. In the toroidal pore model, the peptides induce reorganization of nearby lipids so that both peptides and lipid headgroups line the pore lumen. The carpet model suggests that the peptides insert at the membrane surface, parallel to the bilayer, leading to bilayer thinning and, at high enough peptide concentration, membrane lysis.

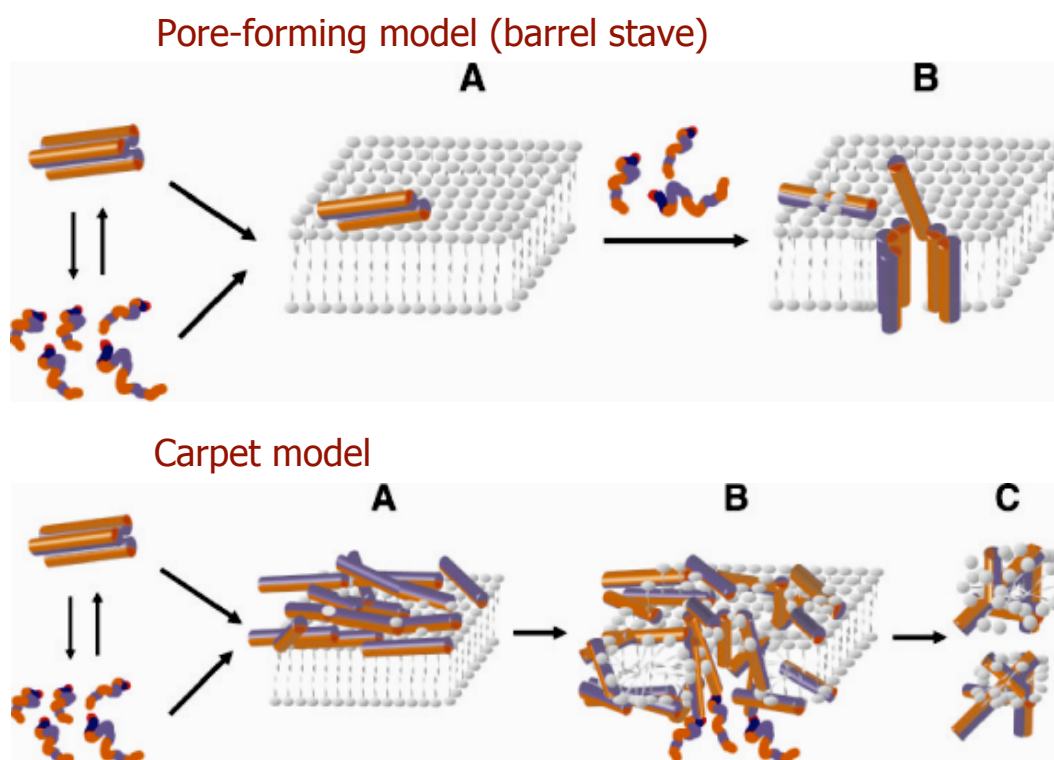


Figure 3.1: Canonical models of membrane permeabilization by antimicrobial peptides. The peptides are generally unstructured in solution or in equilibrium with a structured (often α -helical), aggregated form. They associate with the membrane surface, taking on their structured form (A). In the pore-forming model, the peptides insert into the membrane to form a transmembrane channel (B). The barrel-stave model is shown: lipids in proximity to the channel do not rearrange. In the carpet model, the peptides initially stay associated at the membrane surface, causing membrane thinning as they increase in concentration. Eventually the membrane structure is broken up and the peptides are able to cross the bilayer (B). The membrane can then be solubilized in a detergent-like manner (C). Figure adapted from [109].

The Australian green-eyed tree frog (*Litoria genimaculata*) secretes an AMP, maculatin 1.1

(M1.1), from its dorsal skin glands, which has antibacterial activity against a wide range of bacteria (Fig. 3.2) [108]. The peptide has the sequence, GLFGV⁵LAKVA¹⁰AHVVP¹⁵AIAEH²⁰F-NH₂, and its structure, as solved by NMR in both trifluoroethanol/water mixture and dodecylphosphocholine micelle environments [110], is an amphipathic α -helix of length ~ 30 Å with a kink around the central proline. This kink is suggested to be important to the mechanism of permeabilization, as the P15A mutant has much reduced activity [110]. This may indicate that the kink increases the ability of the peptide to insert into the membrane [111], or it may help to provide an optimal amphipathic surface [112].

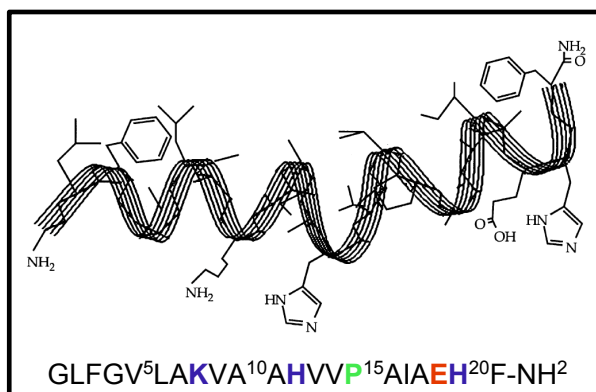


Figure 3.2: Sequence and structure of M1.1 in TFE/water, derived from NMR experiments. Figure adapted from [110].

A variety of experimental biophysical studies have been conducted to investigate the membrane-lytic properties of the peptide. A consensus on a single mechanism has not been reached, and studies have indicated that multiple modes of action may be available to the peptide, modulated the composition of the membrane. Solid-state NMR spectroscopy studies with model dimyristoylphosphatidylcholine (DMPC) membranes indicated an interfacial location for M1.1 [111, 113]. In mixed DMPC/dimyristoylphosphatidylglycerol (DMPG) bilayers there was a much stronger interaction with the membrane, which appeared to be a result of a dual mechanism: oligomerization of the peptide, possibly as a pore structure, as well as loss of lipid from the membrane [114]. This would imply that modes of action corresponding to both the pore-forming and carpet models might occur together. A carpet-type mechanism has been suggested on the basis of ³¹P NMR studies with live bacteria [115] and electron microscopy studies [116].

Other experiments have supported a pore-forming mechanism. Using confocal fluorescence

microscopy, M1.1 was shown to cause differential leakage of fluorescent marker compounds from within giant unilamellar vesicles composed of palmitoyloleylphosphatidylcholine (POPC) or POPC/palmitoyloleylphosphatidylglycerol (POPG) [112]. Out of two markers, the smaller was found to be able to escape from the vesicle, while the larger could not. Large-scale disruption of the vesicle structure was not observed, so a pore-forming mechanism was proposed as the likeliest mechanism. A study using Fourier transform infrared spectroscopy indicated stronger insertion of the peptide into DMPG than DMPC membranes [117], with an orientation of the helix axis of 35° to the membrane normal, suggestive of a transmembrane pore structure. Other studies highlight how peptide concentration and membrane composition can modify the behaviour of the peptide [114, 118].

Molecular dynamics (MD) simulations have enabled development of more detailed models of how AMPs interact with membranes. For example, simulations of magainin-H2 [119] and melittin [120, 121] have indicated that these peptides are able to form water-permeable pores. However, the time and length scales accessible with all-atom MD simulations are often below those required to observe important processes in biological systems [122]; for example, the diffusion of peptides within a membrane may be too slow to allow formation of the most stable configuration of peptides.

As discussed in the Introduction and Methods, coarse-grained (CG) models allow access to much longer time and length scales. One of the first CG models for the simulation of biological membranes was developed by Shelley and Klein [52]; it has been used in simulations of a synthetic arylamide polymer, designed to have similar properties to natural antimicrobial peptides [123]. The MARTINI force field [44, 45] has been used in simulations of magainin [45, 124, 125], melittin [120] and alamethicin [126], which are all naturally occurring α -helical AMPs. A similar force field was used in simulations of synthetic peptides by Khalfa et al. [127]. The Bond and Sansom model [46] (derived from Marrink's original lipid force field [44]) was used for this study.

To investigate the mechanism of membrane permeabilization by M1.1 in more detail, Bond et al. [65] performed CG MD simulations of the peptide at various concentrations, and in both bilayer and vesicle environments. They found that at a low peptide:lipid ratio, M1.1 adopts a kinked, membrane-interfacial location in a phospholipid bilayer, consistent with experiment [111, 113]. At higher peptide:lipid ratios, disordered membrane-spanning aggregates were formed. A simulation with a POPC vesicle revealed that the peptide is capable of inducing

membrane curvature and disruption. This simulation and a range of other vesicle simulations are presented in this chapter.

While water-flow was not observed, and well-defined pores were not evident in the clusters formed by M1.1, the coarse-grained level of representation is likely to have precluded observation of such events. More specifically, such an effect would have been due to the granularity of the CG model (the water particles, equivalent to 4 water molecules, may have been simply too large to flow through the membrane) and the lack of a water dipole. The latter factor has been shown to be important in the process of pore-formation by AMPs in MD simulations [128]. It was noted by Bond et al. [65] that conversion of a cluster to an atomistic representation could be used to investigate water flow in more detail. The results of simulations based on this multiscale methodology are reported here.

3.2 Methods

3.2.1 CG Simulations

The structure of M1.1 used in these simulations was determined by solution-state NMR in a trifluoroethanol/water mixture [110]. For the CG simulations, the structure was converted to a CG representation, as described in Bond et al. [46]. This CG structure comprised a chain of backbone particles with attached side-chain particles. Harmonic distance restraints (with equilibrium length 0.6 nm and force constant $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$) were used to restrain the i and $i+4$ backbone particles, mimicking the hydrogen-bonds formed in α -helices [102]. The i to $i+4$ restraint was omitted between Pro15 and Ala11, to model the lack of a hydrogen-bond donor in proline and the consequent greater flexibility of the helix in this region.

The CG systems were created with a pre-equilibrated, 256-lipid, dipalmitoylphosphatidylcholine (DPPC) bilayer, with peptides positioned randomly on each side of, and at least 2 nm from the bilayer. The peptides were added with equal numbers on either side of the bilayer, except for the simulation with three peptides, in which one peptide was added on one side, and two on the other. Prior to production simulation, all CG systems were energy-minimized using < 500 steps of the steepest-descent method, then equilibrated for 20 ns with positional restraints on the peptides and bilayer. A $9 \times 9 \times 26 \text{ nm}^3$ box was used for all CG simulations.

The initial POPC vesicle system was set up by Peter Bond as described in [65]. First, the coordinates of a 16 nm diameter, 877-lipid, POPE vesicle were downloaded from the Marrink CG web site (<http://md.chem.rug.nl/~marrink/MARTINI/Coordinates.html>). The lipids were converted from POPE to POPC by appropriate modification of their headgroup particles, and the solvated POPC vesicle system was energy-minimized using < 500 steps of the steepest-descent method. The box size was $21 \times 21 \times 21$ nm. A production simulation of 400 ns was performed, which served as the control simulation without peptides added. For the simulation with maculatin, 100 copies of the CG M1.1 structure were added randomly outside the vesicle at the end of the 400 ns control simulation. The same energy-minimization was performed and a 1.25 μ s production simulation was performed.

An equivalent vesicle simulation was conducted with the nonpore-forming peptide WALP23 [129]. The peptide was modelled as an idealized α -helix and then converted to a CG model. Harmonic distance restraints were used between all i and $i + 4$ backbone particles. Another simulation was carried out with M1.1 and a vesicle composed of a 3:2 mixture of POPC/POPG. The vesicle was created from the POPC vesicle at the end of the 400 ns control simulation by randomly converting the appropriate number of POPC headgroups to POPG headgroups, followed by < 500 steps of steepest-descent energy minimization and a 10 ns equilibration simulation. The M1.1 peptides were then added in the same way as for the other simulations, and a 1.25 μ s production simulation was performed.

CG parameters were taken from Marrink et al. [43] for lipids, water and ions, and from Bond et al. [65] for amino acids. CG simulations were run with Lennard-Jones interactions shifted to zero between 0.9 nm and 1.2 nm, and electrostatics shifted to zero between 0 nm and 1.2 nm, with a relative dielectric constant of 20. The nonbonded neighbor list was updated every 10 steps. Simulations were performed at constant temperature, pressure and number of particles. Temperatures of protein, lipid and solvent were coupled separately using the Berendsen algorithm at 323 K, with a coupling constant $\tau_T = 1$ ps. The system pressure was coupled using the Berendsen algorithm at 1 bar, with a coupling constant $\tau_P = 1$ ps and a compressibility of 5×10^{-6} bar $^{-1}$. The timestep for integration was 40 fs (a set of simulations which validate the use of this timestep are discussed in Appendix A). The times reported in this work are taken directly from the simulations, and not adjusted for the sampling of the CG model which can be somewhat faster than equivalent atomistic simulations.

3.2.2 Atomistic Simulations

Proteins and lipids were converted from CG to atomistic representations using a fragment-based approach, as described by Stansfeld et al. [130]. The atomistic systems were resolvated with SPC water (after reducing the box z-dimension to 11 nm to reduce the amount of bulk water) and energy-minimized, using < 500 steps of the steepest-descent method. Two equilibrations using MD were then performed, each for 2 ns. In the first, position restraints (with force constant $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ in each direction) were placed on all protein non-hydrogen atoms; in the second, only the -carbons of each residue were restrained, allowing some relaxation of the peptides without disturbing the structure of the cluster as a whole. Production simulations of 30 ns were then performed.

The force field used in the atomistic simulations was the united-atom GROMOS96 force field [131] with parameters for DPPC from Berger et al. [132]. Lennard-Jones interactions were truncated beyond 1 nm, while electrostatic interactions beyond this point were approximated using particle mesh Ewald (PME) summation. The nonbonded neighbor list was updated every 10 steps. Simulations were performed at constant temperature, pressure and number of particles. Temperatures of protein, lipid and solvent were coupled separately using the Berendsen algorithm at 323 K, with a coupling constant $\tau_T = 0.1 \text{ ps}$. The system pressure was coupled using the Parrinello-Rahman algorithm at 1 bar, with a coupling constant $\tau_P = 1 \text{ ps}$ and a compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$. The timestep for integration was 2 fs.

3.2.3 Software and Analysis

All simulations were performed using the GROMACS (www.gromacs.org) simulation package [85–88]. VMD was used for visualization [133]. Clustering of the peptide aggregates was carried out by fitting the CG trajectories (from the time of complete peptide insertion) to the final configuration of peptides, then using the RMSD-based clustering algorithm g_cluster (part of the GROMACS package) with the single linkage algorithm to populate clusters. The most populated cluster of each simulation was chosen for conversion to atomistic resolution. Secondary structure analysis was performed using the DSSP program [134]. Water permeation and flux analysis were carried out using Oliver Beckstein’s g_count and g_flux programs [135].

3.3 Results

3.3.1 Multiscale Simulations of Peptide/Bilayer Interactions

Five CG systems were set up, each with 16 M1.1 peptides positioned randomly outside a pre-formed, 256-lipid, DPPC bilayer.¹ Production simulations of 1 μ s were performed for each system. In each case, the peptides rapidly aggregated in solution and located themselves at the bilayer surface, before inserting as clusters. All 16 peptides eventually became included in a single, membrane-spanning cluster (Fig. 3.3). The clusters formed were irregular, fluctuating structures, with the peptides adopting a wide range of tilt and kink angles. Further detail on the insertion pathways can be found in the article by Bond et al. [65]. The insertion behaviour can be compared with single-peptide self-assembly simulations from the same article, in which peptides remained bound at the membrane interface with only transient insertion possible via bending around the proline kink. The insertion properties observed in the 16-peptide simulations therefore appear to be due to a cooperative mechanism.

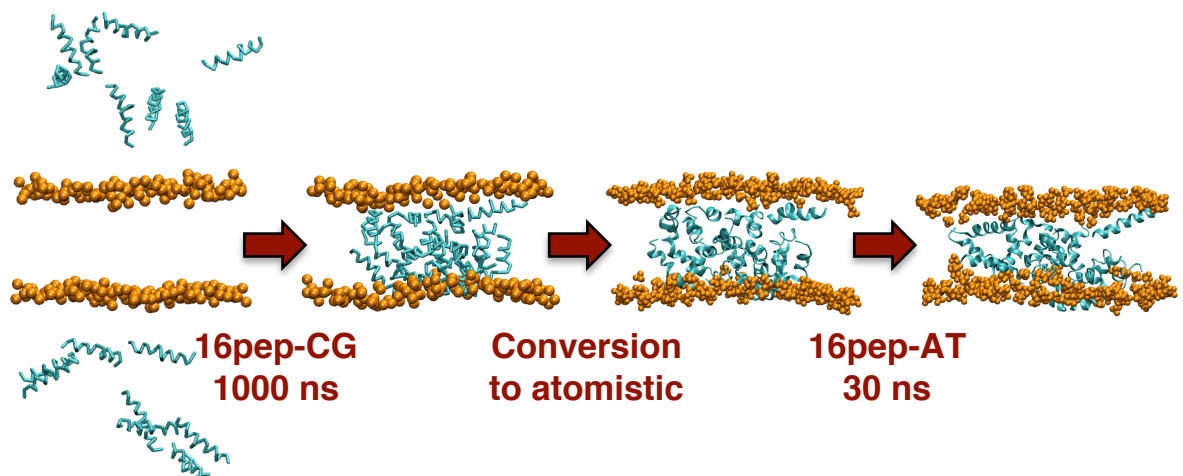


Figure 3.3: Snapshots from one of the multi-scale simulations of M1.1. The leftmost snapshot shows the CG system at $t = 0$ ns. Following a 1000 ns CG simulation, the most representative cluster (at 920 ns, shown in the second snapshot from left) is converted to an atomistic representation (third snapshot) and simulated for another 30 ns to study water flow in more detail. The final configuration is shown in the rightmost snapshot. For clarity only the peptides (cyan) and lipid phosphate particles (orange) are shown.

¹These simulations were based on the same method used in [65]. The original simulations were performed by Peter Bond, although those presented here have been conducted by the author of this thesis.

Clustering was used in each simulation to find the most representative structure of the inserted peptide aggregate. These structures were subsequently converted to atomistic representations as described in the Methods section. Production simulations (named 16pep-AT) of these five atomistic systems were run for 30 ns each. In each case, water quickly penetrated into sites within the peptide clusters during the first 2 ns equilibration stage. Over the last 15 ns of each production simulation, an average of 26 ± 6 waters were present in the bilayer region, significantly higher than in control simulations of a pure DPPC bilayer, in which an average of one water was present (Table 3.1). Fig. 3.4 shows a representative snapshot. The average density of waters across the bilayer is shown in Fig. 3.5 (green line), and is seen to be much higher in the bilayer center than in the control simulations (black line). The atomistic control system was converted from a CG simulation in the same way as for the peptide simulations, and then simulated three times (each with different starting velocities) for 30 ns.

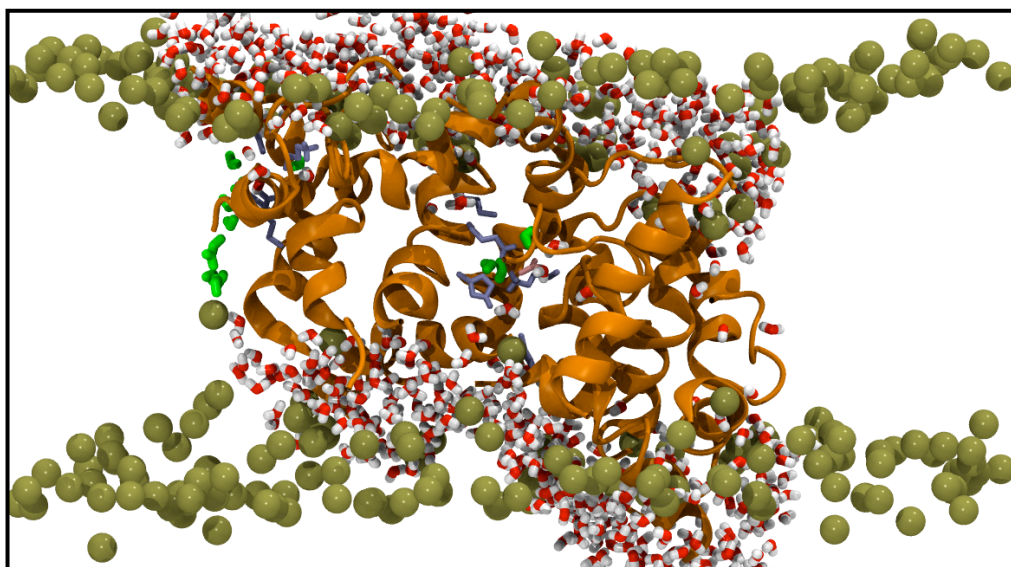


Figure 3.4: Snapshot of a 16-peptide M1.1 cluster after 24 ns of atomistic simulation. Peptides are shown in orange; DPPC phosphorus atoms in tan; waters which flow across the cluster during the simulation in green; lysine and histidine sidechains nearby these waters in pale blue; nearby glutamic acid sidechains in pink; other waters nearby the peptides in red and white. Other waters are not shown.

Peptides were able to move within the clusters individually, but remained aggregated. The overall features of the clusters were similar throughout the simulations: irregular structures with the peptides distributed approximately evenly across the bilayer normal. For each peptide, at least one terminus tended to be anchored at the membrane interface. No major

Simulation ($\times$ number of simulations)	Mean number of waters in bilayer over last 15 ns of simulation	Mean number of bidirectional water crossings over last 15 ns	Diffusional permeability for a single channel, p_P ($\times 10^{-14}$ cm ³ s ⁻¹)	Diffusional permeability coefficient P_P ($\times$ 10^{-14} cm ³ s ⁻¹)
2pep-AT ($\times 3$)	2 ± 2	7.7 ± 5.7	0.77 ± 0.57	-
4pep-AT ($\times 3$)	3 ± 3	6.3 ± 3.8	0.63 ± 0.38	-
6pep-AT ($\times 3$)	27 ± 12	48 ± 20	4.8 ± 2.0	-
8pep-AT ($\times 3$)	29 ± 6	39 ± 13	3.9 ± 1.3	-
16pep-AT ($\times 5$)	26 ± 6	49 ± 20	-	7.6 ± 3.1
DPPC ($\times 3$)	1 ± 2	$7.7 \pm 2.1^*$	-	1.2 ± 0.3

Table 3.1: The mean number of waters present in the bilayer over the last 15 ns of each simulation was calculated by counting the number of waters within 0.5 nm of the bilayer center (as defined by the average position of the lipid tail terminal carbons). Water crossings were analysed by first defining a cylinder of 2 nm length, oriented with the bilayer normal, centered laterally (in the bilayer plane) at the peptide cluster center, and longitudinally at the bilayer center, and of radius roughly equivalent to the lateral radius of the peptide cluster (ranging from 1.7 to 4.5 nm). The cylinder is thus a crude representation of a pore through the bilayer at the location of the peptide cluster. A crossing was counted if a water molecule entered one half of the cylinder, and exited from the other. The single-channel diffusional permeability, p_d , was calculated using the relationship $p_d = q_0 v_W$ [136], where q_0 is the number of unidirectional water-crossings per unit time, and v_W is the volume of a single water molecule. The diffusional permeability coefficient for a given area of membrane, P_d , can be calculated experimentally by measuring the flux of a tracer molecule across a membrane. Here, the permeability coefficient is calculated using the expression $P_d = q_0 v_W / A$, where A is the cross-sectional area of the pore, as defined by the cylinder used for the water crossing analysis. A derivation of this expression can be found in Appendix B.

* For comparison with the 16pep-AT simulations, the same cylinder radius of 4.5 nm was used for the analysis of the pure DPPC bilayer control simulations. For comparison with the 2pep-AT simulations, the analysis of the control simulations was repeated with a cylinder radius of 1.7 nm, resulting in a mean of 1 ± 1 crossings.

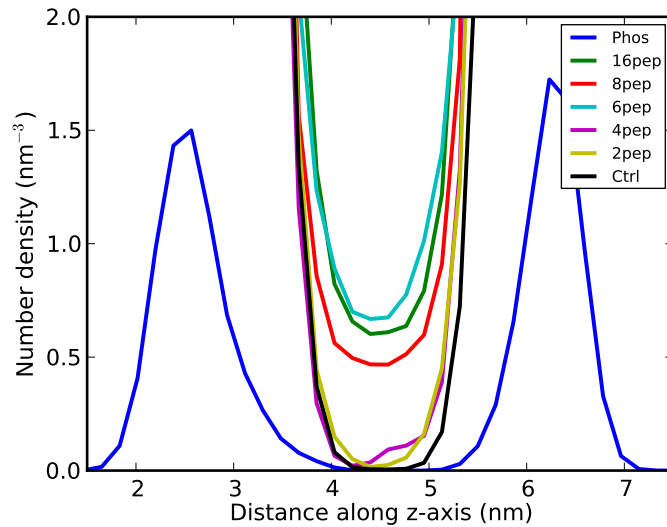


Figure 3.5: Water density along the z-axis (bilayer normal) for the last 15 ns of each of the atomistic simulations, averaged over time and each simulation repeat. The blue line represents the density of DPPC phosphorous atoms for one of the 16pep-AT simulations. Bilayers were centered in the z-coordinate in the trajectories used to calculate the density plots.

changes in the helicity of the peptides (mean 46%; Fig. 3.6) occurred over the course of the simulations.

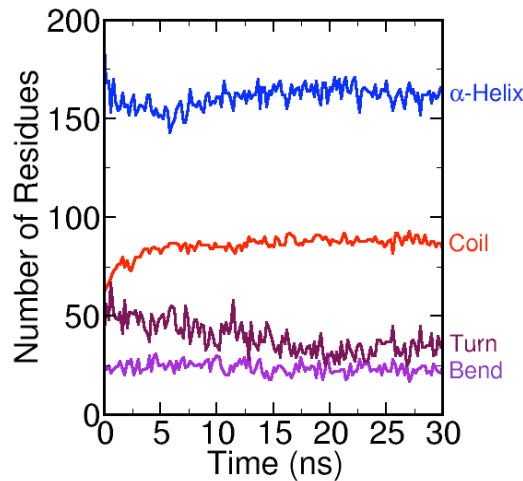


Figure 3.6: Secondary structure of all 16 peptides during one of the 16pep-AT simulations, analysed using the DSSP program [134].

3.3.2 Water Permeation Through Peptide Clusters

Water molecules were able to flow across the clusters through multiple, fluctuating channels, as indicated in Fig. 3.4. The channels were narrow, usually allowing only one water molecule to pass through at a time (although multiple channels would sometimes be available within a peptide cluster). Channels generally remained traversable for the duration of the simulations, but the exact path would often fluctuate with the movements of peptides within the cluster.

Two main factors appeared to be important in aiding the flow of water across the membrane. Polar and charged peptide sidechains (Lys8, His12, Glu19 and His20) distributed across the membrane core provided a hydrophilic pathway for waters to cross. This is reflected in the higher numbers of water contacts recorded for those residues (Fig. 3.7), and is especially evident for the central Lys8 and His12 when compared to the surrounding residues. Some hydrophobic residues also had relatively high numbers of water contacts, but these were towards the ends of the peptides, which were often anchored at the membrane and thus close to bulk water.

The other factor aiding the flow of water was the presence of inserted lipid headgroups, which were noticeable in channels formed at the edges of peptide clusters. The insertion of lipids into the membrane core is reminiscent of the disordered toroidal pores observed in simulations of magainin-H2 [119] and melittin [120, 121]. However, these inserted lipids and their associated channels were rare in the 16pep-AT simulations and the majority of water flux occurred through channels entirely within the clusters without inserted lipid headgroups present (although more lipid insertions were observed in simulations with 6 peptides, as will be described in later sections).

Water flux was quantified by counting the number of bidirectional crossing events during the simulations (Table 3.1). The diffusional permeability coefficient for the region of membrane containing the inserted cluster was calculated to be $7.6 \times 10^{-2} \text{ cm s}^{-1} \pm 3.1$ (Table 3.1). The value for the control simulations was calculated to be $1.2 \times 10^{-2} \text{ cm s}^{-1} \pm 0.3$, which is in agreement with previous atomistic simulations e.g. [137] and experimental estimates [138, 139]. Thus, the peptide cluster appears to increase the local membrane permeability by a factor of ~ 6 .

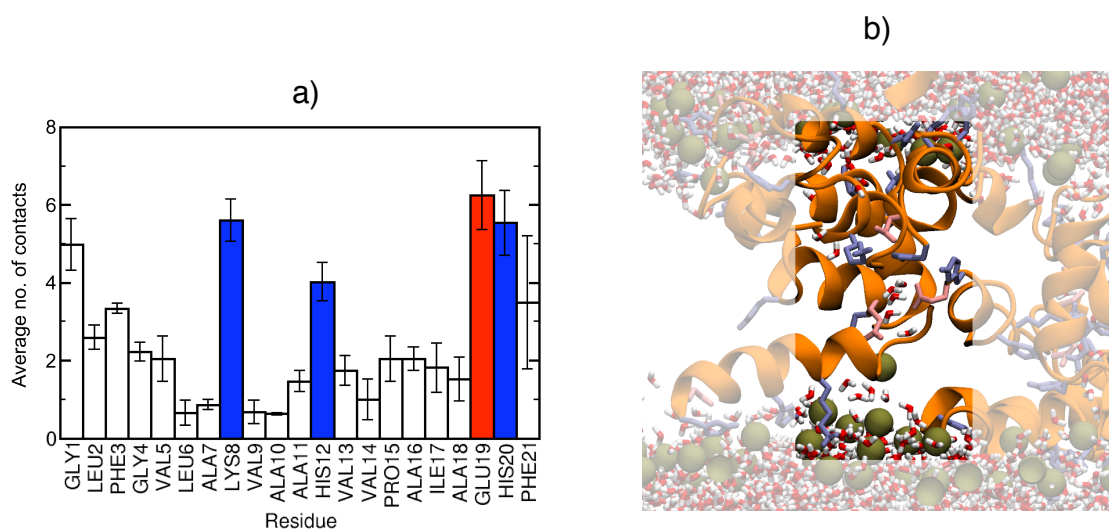


Figure 3.7: *a)* Average number of residue-water contacts per peptide at any point in time during each of the five 16pep-AT simulations. Contacts with positively charged residues are shown in blue; contacts with negatively charged residues in red; contacts with hydrophobic residues in white. A contact is defined as when any atom of a water molecule is within 0.4 nm of any atom of an amino acid residue, with a maximum possible count of one per frame for each residue-water pair.

b) Snapshot from one of the 16pep-AT simulations showing waters inserted in the membrane around charged peptide residues. Peptide backbones are shown in orange ribbon format, with the sidechains of His12, Glu19 and His20 shown in pale blue, and that of Lys8 in pink. Waters are shown in red and white licorice format, and lipid phosphate atoms as tan spheres.

3.3.3 Minimum Number of Peptides Required for Formation of a Membrane-Spanning Cluster

To investigate the effect of the number of constituent peptides on cluster behaviour, CG systems with two, four, six and eight peptides (named 2pep-CG, 4pep-CG, 6pep-CG and 8pep-CG respectively; Fig. 3.8) were each simulated three times for 1 μ s.

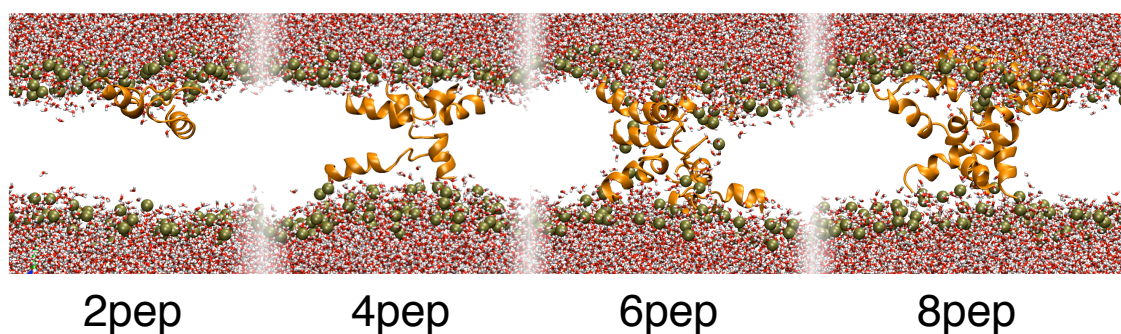


Figure 3.8: (Left to right:) Snapshots of clusters formed during the 2pep-AT, 4pep-AT, 6pep-AT and 8pep-AT simulations. Peptides are shown in orange, and water molecules in red and white. Lipid phosphorus atoms are shown as tan spheres, but the remainder of the lipid structure is not shown for clarity. In the 6pep-AT snapshot, it can be seen that a number of lipids have rearranged so that their headgroups line the peptide cluster.

In the 2pep-CG simulations, the peptides quickly inserted at the bilayer interface, with one positioned in each leaflet. The peptides behaved in a similar fashion to previous simulations of a single M1.1 peptide with a DPPC bilayer [65], adopting an orientation parallel to the bilayer plane with only transient periods of further insertion. The peptides were thus unable to contact each other through the bilayer during the simulation, and no aggregation was observed. Atomistic conversions were not performed for these systems.

The 4pep-CG simulations all resulted in four-peptide clusters, spanning the membrane with two peptides in either leaflet. Clustering was used to select snapshots from each, which were then converted to atomistic resolution (the subsequent simulations being named 4pep-AT). In one of the 4pep-CG simulations, one pair of peptides first aggregated and inserted into one leaflet of the membrane, while the other pair of peptides did not insert into the other leaflet of the membrane until after 600 ns. This gave the possibility of studying a two-peptide

cluster in atomistic detail, so a version of the first 600 ns of the trajectory was created with the non-membrane-bound peptides (which were too distant to be influencing the inserted peptides or bilayer) removed, and clustering was used to select three representative snapshots for conversion to atomistic resolution (the subsequent simulations being named 2pep-AT). The 6pep-CG and 8pep-CG simulations also resulted in membrane-spanning clusters, with roughly equal distributions of peptides across the two leaflets. The three 6pep-AT and three 8pep-AT simulations were derived from representative atomistic conversions of each of these systems.

A CG simulation with three peptides was also conducted — two peptides inserted as a cluster in one leaflet, while the other peptide inserted in the opposite leaflet. The single peptide did not reach far enough into the bilayer center to contact with the other peptides, and thus a membrane-spanning cluster did not form.

3.3.4 Effect of Cluster Stoichiometry on Water Diffusion

The atomistic systems described above, 2pep-AT, 4pep-AT, 6pep-AT and 8pep-AT, were each simulated for 30 ns. The overall structures of the clusters did not change to a large extent over this simulation time. The same analyses as used for the 16pep-AT simulations were used to determine the water density, number of water molecules present in the bilayer center, and the level of water flux (Fig. 3.5 and Table 3.1). In the 2pep-AT and 4pep-AT simulations, water penetration and flux was similar to the base rate in the control simulations. A large increase was observed for the 6pep-AT simulations, in which a number of lipids were inserted into the membrane core surrounding the peptide clusters (Fig. 3.8). The majority of water flow in these simulations occurred through channels formed by these lipid headgroups and the edges of the peptide cluster. However, lipid insertions did not occur in the 8pep-CG (or 8pep-AT) simulations, which gave rise to fairly similar levels of water flux. This finding fits with the results from the 16pep-AT simulations, in which lipid insertions could contribute to water flux, but were not essential.

Rather than calculating permeability coefficients for the smaller peptide clusters (which would be less appropriate for smaller areas of membrane), they have been treated as single water-channels, by evaluating the area-independent diffusional permeability p_d (Table 1). This is directly proportional to the flux (see caption of Table 1), but it is useful to compare it with

other simulation and experimental studies. For example, aquaporin-1 has been found to have a p_d of $\sim 6 \times 10^{-15} \text{ cm}^3 \text{ s}^{-1}$ [140, 141]. The values for the 6- and 8-peptide clusters are almost an order of magnitude higher than this.

To investigate the effect of ion concentration on peptide behaviour, atomistic simulations were carried out with a range of ion buffer concentrations (from 0.5 M to 2 M NaCl). Little difference was observed compared to the other simulations, as the ions did not penetrate the cluster within the timescales simulated. It should be noted, however, that no voltage gradient was present, and this may be an important factor in the effect of ions on the cluster, and on their rate of flux. It should also be noted that no osmotic gradient was present in any of the simulations, and thus the water flux observed is only representative of diffusional flow of water molecules across the membrane.

3.3.5 Deformation of Vesicles Induced by Maculatin

To investigate the effect of maculatin on a larger, curved membrane, a CG simulation was conducted with a POPC vesicle and 100 M1.1 peptides.² The 16 nm diameter vesicle was pre-formed (as discussed in the Methods section) and the peptides were added randomly in solution, outside the vesicle. The observations from the 1250 ns simulation can be broadly divided into three stages (Fig. 3.9, left column): 1) ($\sim 0\text{--}100 \text{ ns}$) peptides aggregate and insert at the vesicle outer leaflet interface; 2) ($\sim 100\text{--}300 \text{ ns}$) aggregates begin to insert into the membrane, contacting the inner leaflet headgroups, and the vesicle shape becomes distorted; 3) ($\sim 300 \text{ ns onwards}$) the ellipsoid shape of the vesicle has essentially stabilized, with the inserted peptides aggregated in two main clusters at the poles of the vesicle long-axis, while non-aggregated peptides are still associated at the outer leaflet interfacial region. The evolution of contacts between peptides and lipids over time can be seen in Fig. 3.10. The degree of vesicle distortion is quantified with the ratio of the principal moments of inertia of the inner leaflet lipids (Fig. 3.11), showing that the vesicle takes the shape of a prolate ellipsoid after about 300 ns. This behaviour can be compared to that of a control simulation with no peptides, in which the vesicle stays spherical over 400 ns (Figs. 3.11).

The aggregates observed were similar to those in the flat bilayer simulations in terms of

²This simulation and a control simulation with no peptides were originally performed by Peter Bond and included in [65]. The analysis reported here and other vesicle simulations have been performed by the author of this thesis.

their lack of ordering. However, within the aggregates there tended to be a larger number of peptides in the outer leaflet. This is likely to be the driving force for the vesicle deformation, especially considering that the aggregates localize at the poles where the membrane curvature is highest.

The membrane distortion induced by maculatin was compared to that of WALP23. The WALP peptides are a series of artificial transmembrane α -helices, composed of alternating alanines and leucines, with two flanking tryptophans at each end, and with the N- and C-termini blocked [129]. They do not have antimicrobial activity, nor do they permeabilize membranes [142, 143]. In a simulation with 100 WALP23 peptides and a POPC vesicle, the peptides followed similar pathways of aggregation and insertion as maculatin. However, the peptides were generally oriented in a TM manner, spanning both leaflets of the vesicle (Fig. 3.9, right panel, and Fig. 3.10). A degree of distortion of the membrane was observed, but this was considerably reduced compared to the maculatin simulation (Fig. 3.11).

To investigate the effects of maculatin on a membrane more closely related to that of a bacterium (which has negatively charged phospholipids in both leaflets), a simulation was carried out with 100 M1.1 peptides and a vesicle comprising POPC and POPG in a 3:2 ratio. The pathway of insertion and the nature of the resulting aggregates appeared similar to the simulation with pure POPC. However, the degree of distortion of the vesicle appeared to be considerably higher than in the simulation with the neutral POPC vesicle. While more simulations would be required to confirm that this phenomenon is not simply a product of the particular pathways of peptide insertion in either simulation, it raises an interesting possibility for understanding the mechanism of selectivity of maculatin for permeabilizing bacterial over eukaryotic membranes.

3.4 Discussion

The mechanism of membrane permeabilization by maculatin has been discussed widely in the literature, with various experimental studies indicating pore-forming or membrane-disrupting models, often showing dependence on conditions such as lipid composition or peptide concentration (see [144] for a full review). This work presents two possible mechanisms: water permeation through membrane-inserted disordered pores, and membrane distortion.

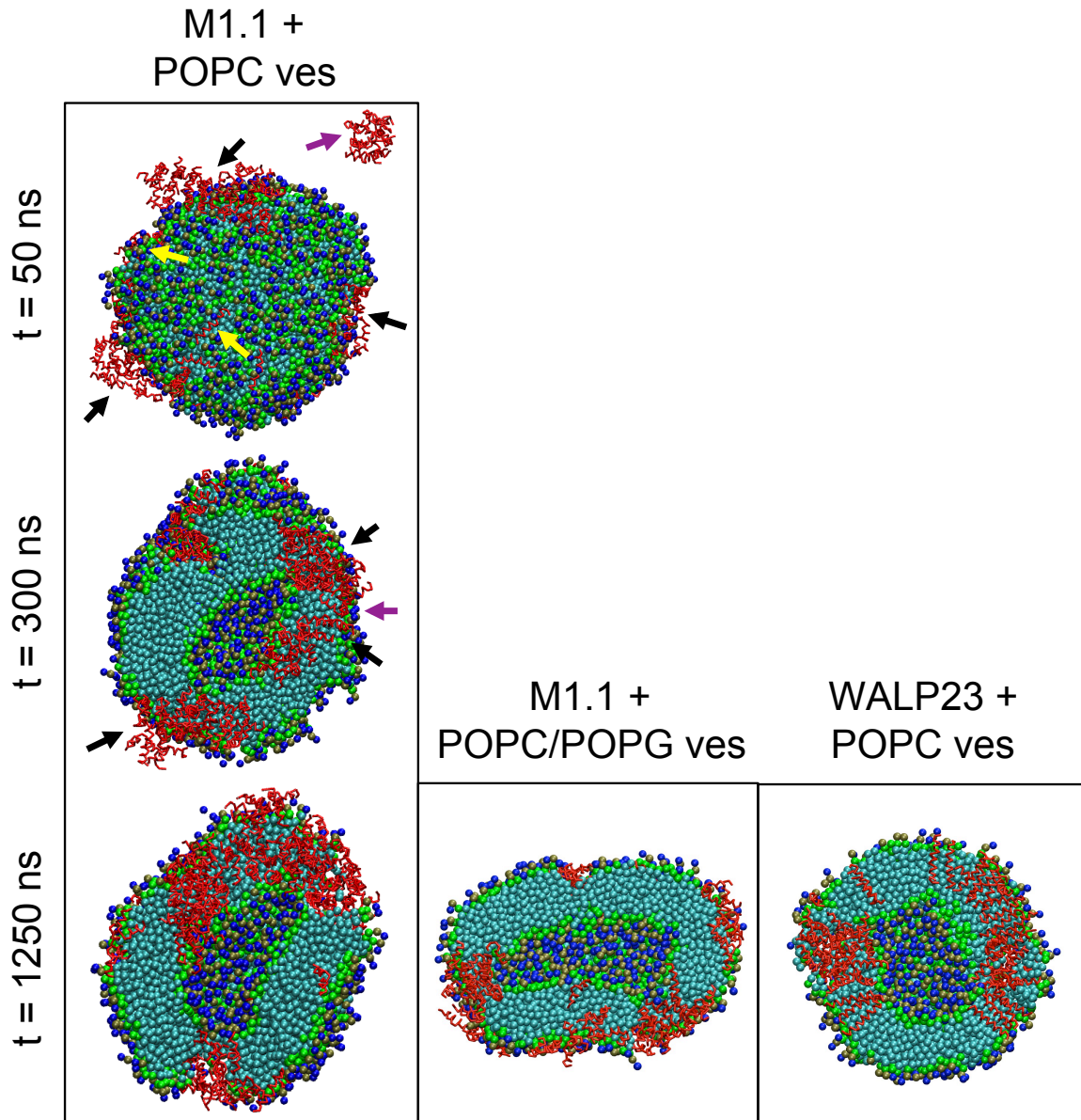


Figure 3.9: Snapshots from the M1.1 + POPC vesicle simulation at 50 ns, 300 ns, and 1250 ns, and from the M1.1 + POPC/POPG vesicle and WALP23 + POPC vesicle simulations at 1250 ns. All except the former are cross-sections through the vesicle. The 50 ns snapshot shows a peptide cluster in solution (*purple arrow*), three peptide clusters adsorbed to the vesicle surface (*black arrows*), and groups of monomeric/dimeric peptides adsorbed to the surface (*yellow arrows*). The 300 ns snapshot shows three membrane-integrating peptide clusters (*black arrows*), two of which later fuse to form a larger peptide aggregate (*purple arrow*). In the 1250 ns snapshot of the M1.1 + POPC vesicle simulation, two TM peptide clusters are evident, roughly at the poles of the elongated vesicle. In each snapshot, the M1.1 peptide is shown in red backbone format. For the POPC lipids, cyan, green, tan, and dark blue = lipid tail, glycerol backbone, phosphate, and choline particles, respectively.

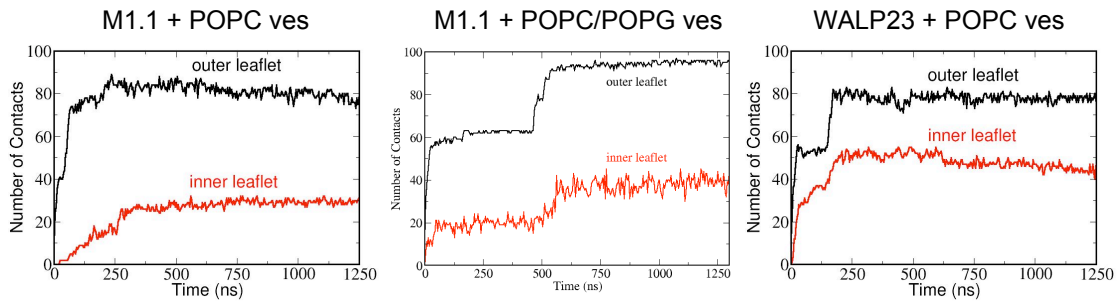


Figure 3.10: Contacts between peptides and outer and inner lipid headgroups from the vesicle simulations.

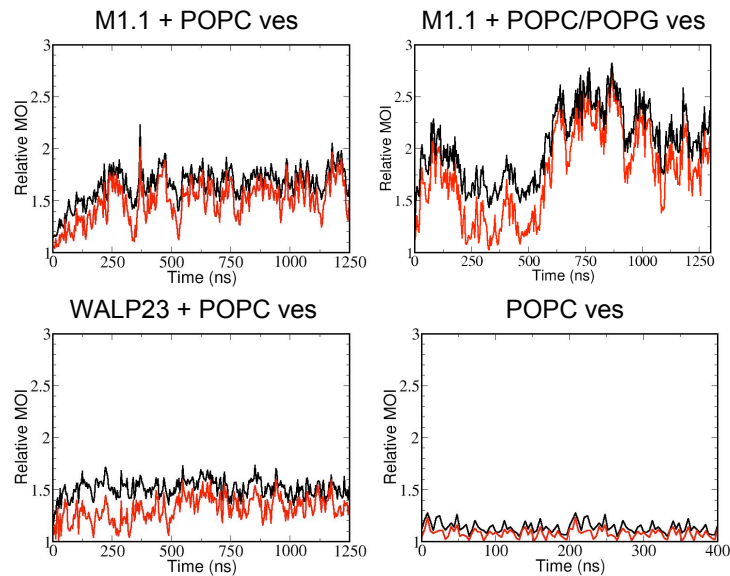


Figure 3.11: Principal moments of inertia (MOIs) for the inner leaflet lipid headgroups in the vesicle simulations. The two largest values at each timestep are plotted relative to the smallest. Where the three principal MOIs of an object are represented by a , b and c , a perfectly spherical object would have a ratio of $1 : 1 : 1$, i.e. $a = b = c$. An oblate ellipsoid (disc-shaped) would have one higher principal MOI ($a = b < c$), while a prolate ellipsoid (like a rugby ball) would have two higher principal MOIs ($a = b > c$).

Although many studies have implied the possibility of pore-formation by the peptide, little direct evidence is available regarding the structure of such aggregates. The multiscale simulations of M1.1 and POPC bilayers indicate that M1.1 will form disordered, membrane-spanning clusters, which are permeable to water when containing six or more peptides. Permeation occurs through fluctuating channels, formed either by inserted polar and charged sidechains within the peptide clusters, or between the cluster edges and inserted lipid headgroups. This model contrasts with earlier theories of antimicrobial peptide pore-formation, in which the peptides are assumed to adopt transmembrane orientations around a central pore. However, several recent simulation studies have indicated that the structures may be more disordered than originally thought, e.g. for magainin-H2 [45, 119], melittin [120, 121] and alamethicin [121, 126]. Solid-state NMR experiments in one of the same studies of alamethicin also reinforced the finding that the inserted peptides take on a large range of configurations in relation to the membrane [126].

CG simulations of maculatin with vesicles show that on a larger scale, peptide aggregates can induce membrane distortion, in contrast to the non-antimicrobial peptide WALP. This could represent the first stage of a membrane rupture mechanism. Furthermore, a simulation with a negatively charged vesicle membrane (3:2 POPC/POPG) exhibited considerably higher distortion, which suggests a possible explanation for the selectivity of maculatin for bacteria, although further work would be required to validate this observation. The effect of AMPs on membrane curvature has not been studied in the recent simulation literature to the same extent as membrane poration, due to the necessity for studying larger patches of membrane. However, a recent CG simulation study showed that magainin can induce budding of vesicles when peptides are added on one side of the membrane (or pore-formation when added on both sides) [125].

A range of permeabilization properties of M1.1 has been observed in experimental studies, depending on conditions such as peptide concentration and membrane composition. Our simulations hint at a number of possible explanations. Firstly, the permeability of the peptide clusters in the atomistic simulations was dependent on the number of peptides present. The water channels observed were also irregular and variable. It could also be hypothesized that these channels in fact represent the first step in a membrane rupture mechanism, which would only become apparent at much longer time scales and/or under different conditions. For example, turgor pressure would likely influence the rate of water flow across a peptide

cluster *in vivo*. Another mechanism which may contribute to the range of permeabilization behaviour is the membrane distortion and disruption seen in the work with vesicles. This may be supported by the possible differentiation of the curvature effect by the membrane charge. An effect on membrane curvature was also observed in a recent CG simulation study of magainin, which showed that the peptide can induce budding of vesicles when added on one side of a membrane, or pore-formation when added on both sides [125]. Overall, it seems likely that multiple modes of action are available to M1.1 (and many other AMPs), which may work together in concert or become more individually dominant depending on the conditions. This work shows that water flow across a disordered peptide cluster and induced membrane distortion are two possible mechanisms.

This work is among the first to use a multiscale simulation methodology to study antimicrobial peptides, alongside recent studies of alamethicin [126] and magainin-H2 [124]. The long time-scales attainable with CG molecular dynamics (a total of 24 μ s were simulated in the bilayer studies) made feasible the simulation of spontaneous formation of membrane-spanning peptide clusters. Subsequent transformation to an atomistic representation (with a total of 600 ns of simulation time) allowed for the detailed study of the more subtle properties of the system, especially the interactions of water with the peptides and membrane. This methodology added to the CG simulation work by showing that the inserted peptide clusters can cause permeabilization of the membrane.

Chapter 4

Modulation of Model Membrane Protein Aggregation by Hydrophobic Mismatch, Membrane Curvature, and Protein Class

4.1 Introduction

A key concept in the understanding of membrane/protein interactions is that of hydrophobic mismatch [15, 145–149], which refers to the difference between the length of the hydrophobic region of a membrane protein and the thickness of the hydrophobic core of the lipid bilayer in which it is embedded. Hydrophobic mismatch can induce a number of possible responses in a membrane, including: (*i*) local bilayer stretching if the hydrophobic length of the protein is longer than the hydrophobic thickness of the lipid bilayer (i.e. positive mismatch); (*ii*) local bilayer compression in the case of negative mismatch (i.e. when the hydrophobic length of the protein is shorter than the hydrophobic thickness of the bilayer) [150]; (*iii*) tilting of TM regions of proteins (more common for α -helical proteins than β -barrels [151]); (*iv*) TM protein deformation [152]; and (*v*) TM protein aggregation [153].

In this study the influence of hydrophobic mismatch on protein aggregation is explored using CG MD. The proteins considered are simple models of α -helix bundle and β -barrel membrane

pore proteins, and have been used in a previous study [154]. In this work, 16 copies of a protein are embedded in a ~ 7000 lipid bilayer, and a range of parameters are varied in order to investigate how protein aggregation is affected. These parameters include protein class (α -helical or β -barrel), lipid bilayer hydrophobic thickness, and membrane curvature (by comparing planar bilayer systems to vesicle systems). The focus is on α -helical bundle proteins as the majority of TM proteins are of this class [155], but the study is extended to β -barrel TM proteins [156] to explore the robustness of the findings. Protein aggregation driven by hydrophobic mismatch has been reported in a number of previous studies (e.g. [153, 157]). Other studies have been extended to explore protein-induced lipid sorting [158, 159] and lipid-induced protein sorting [54].

4.2 Methods

4.2.1 Protein pore models

The two simplified membrane protein TM domain models (Fig. 4.1) are derived from: (i) an NMR-based model of the nicotinic acetylcholine receptor M2 δ peptide helical bundle (M2; PDB ID: 1EQ8) [160], a simple α -helical protein pore; and (ii) the TM domain of Staphylococcal α -hemolysin (HL; PDB ID: 7AHL), a well studied β -barrel protein pore [161]. The structures of these proteins, which represent the two main classes of membrane protein were simplified to yield the two model protein pore structures: G α ('generalised α ') and G β ('generalised β '). These models were originally produced by Jochen Klingelhoefer (SBCB, Oxford), as detailed in a paper [154] and his thesis [162]. Each of the model proteins has a hydrophobic outer surface (leucine residues), a hydrophilic inner pore surface (serine residues), and rings of membrane-anchoring tryptophan residues [163, 164]) at each end of the protein. The G α protein resembles the sequence of Lear and DeGrados de novo designed pore-forming peptide LS3 [165]. The hydrophobic lengths of the protein pores are 2.7 nm (G α) and 2.8 nm (G β).

4.2.2 Preparation of the protein pore-lipid bilayer systems

In the CG model, each phospholipid is represented by a positively charged particle (choline), a negatively charged particle (phosphate), two polar particles (glycerol), and a variable number of hydrophobic particles to represent acyl chains of different lengths [43]. In this chapter, the

nomenclature is that $\text{CG}n$ indicates a lipid with two chains of (equal) length n hydrophobic CG particles, of which each particle corresponds approximately to four $-\text{CH}_2-$ groups. Thus, e.g. $\text{CG}4$ is approximately equivalent to dipalmitoyl (i.e. $(\text{C16:0})_2$) phosphatidylcholine. A range of bilayer hydrophobic thicknesses (from 3.1 nm to 4.4 nm) were studied by using lipid bilayer compositions ranging from 1:1 $\text{CG}2/\text{CG}3$ to $\text{CG}4/\text{CG}5$. Systems similar to $\text{CG}3$, $\text{CG}4$, and $\text{CG}5$ have been used in a previous study of membrane proteins by Marrink and colleagues [166]. The corresponding lipid bilayer systems, each comprising 512 lipids, were formed via 200 ns self-assembly simulations. Single copies of each model protein were inserted into the bilayer centers, with a small number of lipids removed from the site of protein insertion. After a subsequent energy minimization step the systems were solvated with CG water particles. These systems were energy minimized and then simulated for 200 ns. To make the larger systems with 16 proteins the final frame of each of these single protein systems was replicated onto a 4×4 grid in the bilayer plane. The final systems thus comprised 16 identical copies of one protein pore species and ~ 7000 lipids. These were used in production simulations of duration 5 μs . Parts of this work were carried out by Jochen Klingelhoefer (Table 4.1).

4.2.3 Preparation of the vesicle systems

The $\text{CG}2/\text{CG}3$ and $\text{CG}4/\text{CG}5$ mixtures were also used to form unilamellar vesicles of ~ 7000 lipids and diameter ~ 32 nm. To do this, a hollow inner shell (8.2 nm diameter) of strictly hydrophilic particles (repulsive to all CG particles other than water) was placed in the center of the simulation box. By adding CG lipids in random orientations in a surrounding shell, the inner shell thus acts as a mould around which the lipids can form a vesicle, preventing artefacts such as hemifusion. A ~ 7 nm-thick shell of water was placed between the lipids and inner shell. A rhombic dodecahedron simulation box was used. 100 ns simulations were then used to form vesicles. The method is similar in spirit to an earlier study of lipid vesicles [167] which used mean field boundary potentials instead of a shell. The mixed lipid compositions were fairly well conserved within each leaflet (data not shown). The vesicles were also well equilibrated in terms of the total number of lipids in each leaflet (see Appendix C).

Following vesicle formation, the central restrained shell was removed and replaced by waters, and 16 $\text{G}\alpha$ proteins were inserted with equal spacing around the membrane. Lipids were removed from the vesicle membrane to accommodate the proteins. Their positions within the bilayer were matched to a snapshot from a flat bilayer simulation, and their orientations

were matched to the local bilayer normal so that the C-terminal end of each helix was facing outwards from the centre of the vesicle. The resulting structures were energy minimized using 500 steps of the steepest descents method, then briefly equilibrated for 1000 MD steps with a timestep of 10 fs. Subsequent simulations were performed without the central restrained shell, and significant deviations in vesicle shape were not observed (data not shown).

4.2.4 Simulation details

The simulations were performed at 323K, with further details as described in Chapter 3, Section 3.2.2. There has recently been discussion of the timesteps used in CG simulations [168, 169]. A set of simulations which validate the timestep used in these simulations (40 fs) can be found in Appendix A. An elastic network model was used to represent protein secondary and tertiary structure [102, 170] with a force constant of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ and cut-off distances of 1 nm and 0.7 nm for α -helical and β -barrel protein pores respectively. The coupling constant for the Berendsen thermostat was 10 ps. For the Berendsen barostat, the coupling constant was 10 ps and the compressibility $5 \times 10^{-6} \text{ bar}^{-1}$.

	Set-up	Simulation	Analysis
G α bilayer	J.K.	J.K.	D.P.
G β bilayer	J.K.	J.K./D.P.*	D.P.
G α vesicle	D.P.	D.P.	D.P.

Table 4.1: Indicates the originators of the different stages of work on the simulations reported in this chapter. D.P. = Daniel Parton. J.K. = Jochen Klingelhoefer. The work by J.K. is also reported in his DPhil thesis [162].

* The G β simulations were run to 2 μs by J.K., then extended to 5 μs by D.P.

4.3 Results

4.3.1 Simulations of Multiple TM Proteins in Planar Bilayers

Two simplified models of membrane proteins were studied (Fig. 4.1). One model was a generalized α -helix bundle (G α) and the other was a generalized β -barrel (G β). Each model was homo-oligomeric: homo-pentameric in the case of G α , and homo-heptameric for G β .

(Note this corresponds to a 14-stranded β -barrel for $G\beta$, as each monomer consists of a 2-stranded β -hairpin.) For both $G\alpha$ and $G\beta$ the amino acid sequences were such as to reproduce key aspects of biological pores: (i) a central pore lined by polar (serine) sidechains; (ii) a hydrophobic exterior formed by leucine sidechains; (iii) a band of tryptophan sidechains defining the preferred region of interaction of the protein surface with lipid headgroups. A previous simulation study investigated the interaction of single copies of the proteins with bilayers, showing that the proteins can adapt to increased hydrophobic mismatch by tilting [154].

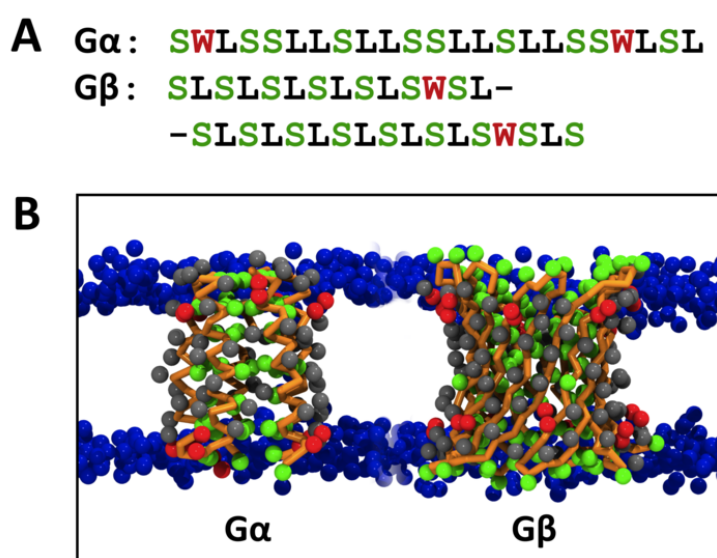


Figure 4.1: *a*) Sequences of the $G\alpha$ and $G\beta$ models, with the hydrophobic leucines in black, the polar serines in green, and the amphipathic aromatic tryptophans in red.

b) CG structures with the protein backbone in orange and the sidechains in the same colour scheme (with leucines in grey rather than black). The structures are snapshots from (separate) simulations with CG2/CG3 bilayers; the phosphate groups of the surrounding lipids are shown in blue.

In order to explore possible aggregation as a result of protein/bilayer mismatch, large ($\sim 50 \times 50 \text{ nm}^2$) bilayers were prepared containing 16 copies of the protein in a bilayer at a protein : lipid ratio $\sim 1 : 450$. The progression of the $G\alpha$ CG2/CG3 simulation (in which there was a positive hydrophobic mismatch, i.e. the hydrophobic region of the protein is longer than the bilayer thickness) is illustrated in Fig. 4.2A. Initially the 16 identical copies of the $G\alpha$ TM protein were spaced equally ($\times 12 \text{ nm}$ apart). Over the course of the $5 \mu\text{s}$ simulation the proteins sampled the whole of the lipid bilayer plane. After about $0.5 \mu\text{s}$

initial protein/protein contacts had been formed and by 5 μ s three aggregates (of 3, 5, and 8 proteins) had been formed. This behaviour may be contrasted with that of the corresponding hydrophobically matched system ($G\alpha$ CG4/CG5; Fig. 4.2D) where although random transient encounters between pairs of proteins occurred over the course of the simulation, they subsequently rapidly dissociated and did not form aggregates.

A comparable behaviour was seen for the β -barrel systems, where the more markedly mismatched system ($G\beta$ CG2/CG3) formed a small number of substantial clusters whereas the better matched system ($G\beta$ CG3/CG4) did not do so to such an extent. Interestingly, both visual inspection of the clusters formed by the end of the simulations (Figs. 4.2B and 4.2C) and measurement of the number of neighbouring proteins vs. time (Fig. 4.3) suggests a difference whereby the $G\alpha$ TM proteins form two dimensional clusters whereas the $G\beta$ proteins form elongated (linear) aggregates. This is of interest in the context of recent comparable simulation studies on clustering of rhodopsin [166], suggesting that the pattern of interacting surfaces on the protein may modulate the nature of the cluster induced by hydrophobic mismatch.

It is also of interest to examine the dynamic aspects of clustering, i.e. the formation and dissolution of contacts of each protein (Fig. 4.3 and Table 4.2). This shows another difference between the $G\alpha$ and the $G\beta$ CG2/CG3 system — dissolution events are more frequent for the $G\alpha$ CG2/CG3 system. The analysis also confirms the occurrence of transient encounters in the bilayer matched systems ($G\alpha$ CG4/CG5 and $G\beta$ CG3/CG4) which however do not lead to significant cluster formation.

4.3.2 Simulations of Multiple TM Proteins in Vesicles

The simulations of multiple TM proteins were extended to a closed membrane vesicle system, in order to explore a model of biological membranes (e.g. those of virions or organelles) which may have a pronounced curvature. To this end the $G\alpha$ CG2/CG3 and CG4/CG5 simulations were repeated in vesicles of diameter ~ 32 nm containing 16 approximately equally spaced proteins, again at a protein : lipid ratio of $\sim 1 : 450$ (Fig. 4.4). As with the planar bilayer simulations, a clear difference in aggregation was observed between the mismatched and matched systems. However, the rate and extent of protein aggregation during the vesicle $G\alpha$ CG2/CG3 simulation appeared to be reduced compared to the equivalent planar bilayer

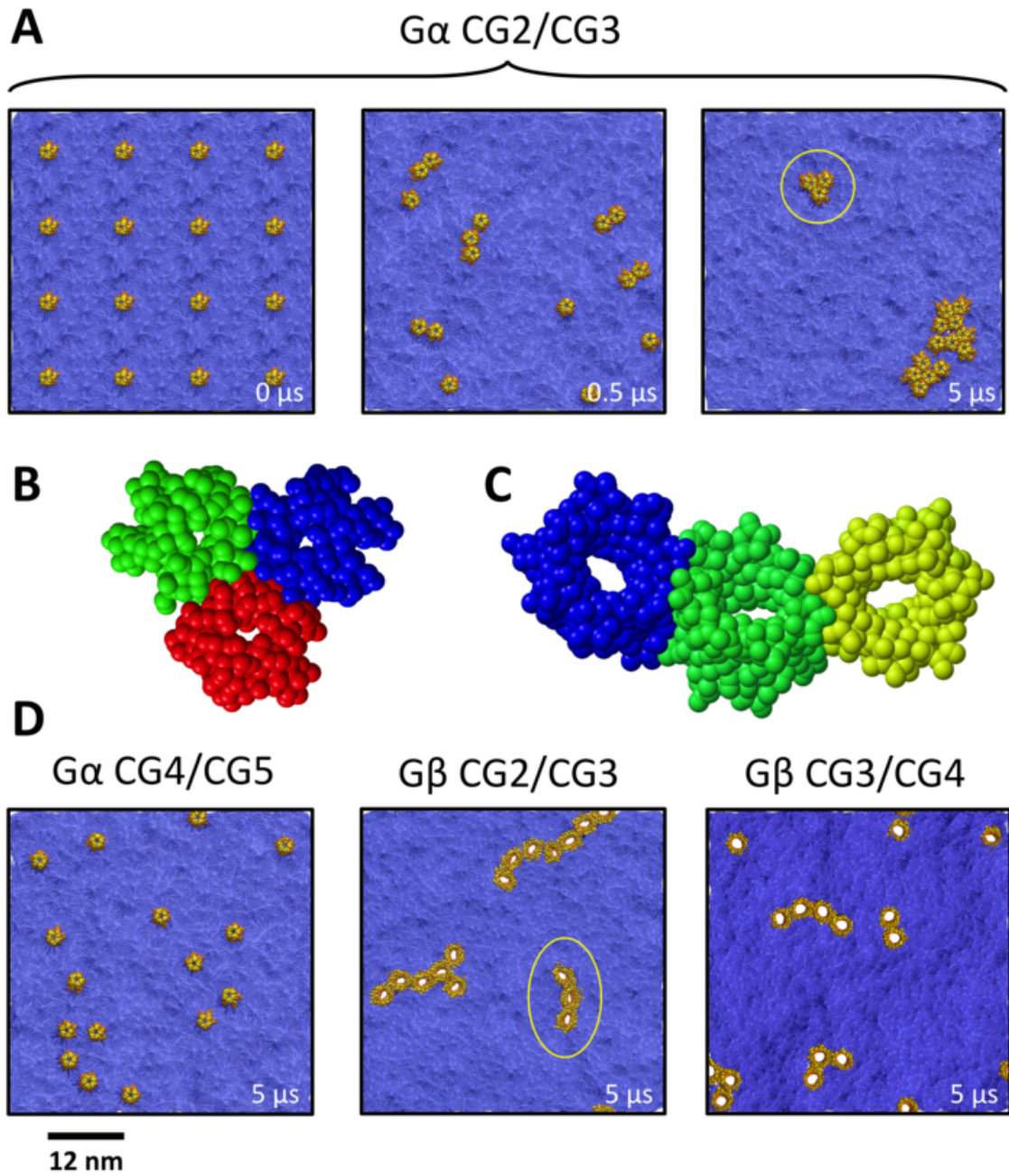


Figure 4.2: Simulations of multiple (16) TM proteins in planar lipid bilayers. In each case the protein is shown in yellow and the bilayer surface in blue. *a*) For the hydrophobically mismatched $G\alpha$ CG2/CG3 system, snapshots at the start (0 μ s), early on during (0.5 μ s), and at the end (5 μ s) of the simulation are shown, indicating the progress of the aggregation process. The yellow ring indicates the cluster shown in more detail in *b*). *d*) For the other three systems ($G\alpha$ CG4/CG5, $G\beta$ CG2/CG3 and $G\beta$ CG3/CG4) a snapshot at the end of each 5 μ s simulation is shown. The yellow ring indicates the $G\beta$ CG2/CG3 cluster shown in more detail in *c*).

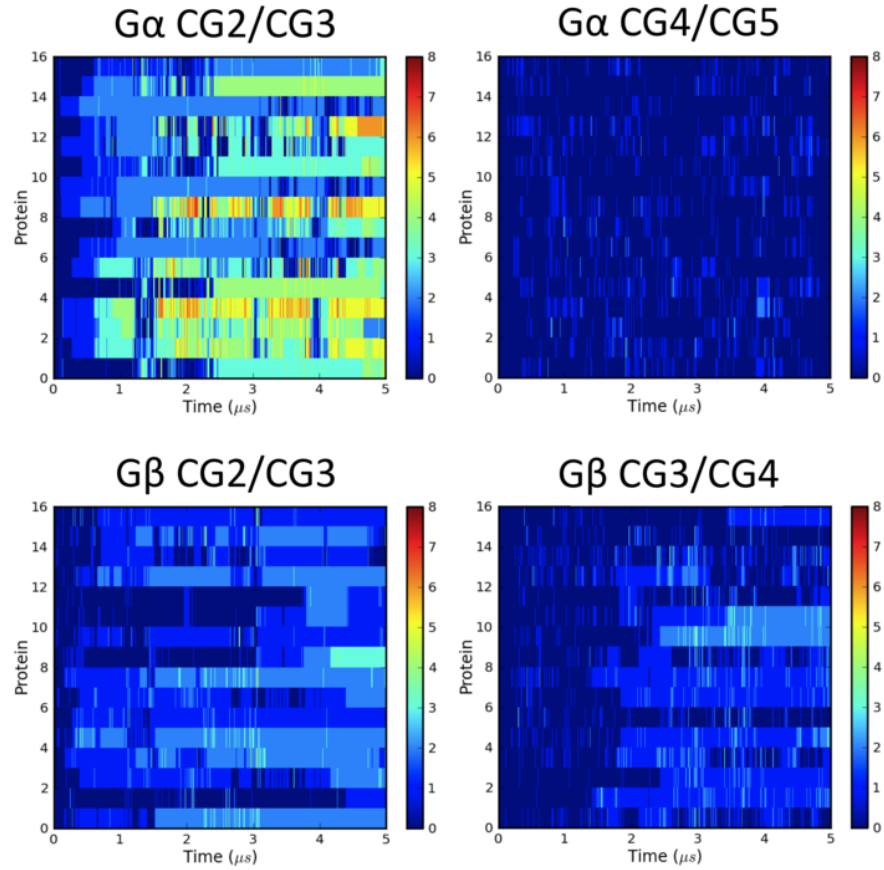


Figure 4.3: Aggregation analysis of the multiple TM protein planar bilayer simulations. For a given simulation, each row represents the number of neighbouring proteins (defined using a cutoff of 2.4 nm between the COM of two proteins) on a scale from 0 = dark blue to 8 = dark red for each of the 16 proteins in the system over the duration of the simulation.

System	Mean cluster residency time (ns)	Number of cluster residencies per protein per 1000 ns
<u>Gα bilayer</u>		
CG2/CG3	540 ± 750	1.8
CG4/CG5	54 ± 38	1.8
<u>Gα vesicle</u>		
CG2/CG3	270 ± 320	2.0
Inv CG2/CG3	2800 ± 660	0.35
CG4/CG5	62 ± 61	1.5
<u>Gβ bilayer</u>		
CG2/CG3	620 ± 970	1.2
CG3/CG4	290 ± 690	2.1

Table 4.2: Cluster residency analysis, estimated over the final 3 μ s of each simulation. A cluster is defined as having at least 2 proteins within 6 nm of each other. The mean cluster residency time is the amount of time that a protein is resident in a particular cluster, averaged for each residency over the course of the simulation, and for each protein. Residency times of 16 ns or less are discounted from the analysis. The number of cluster residencies per 1000 ns indicates the frequency with which proteins join and leave clusters.

simulation. The change in rate follows with a general reduction in diffusion rates in curved membranes, especially of lipids in the inner leaflet (see discussion below). The reduction in the extent of aggregation is confirmed by analysis of the numbers of contacting proteins (Fig. 4.4) which suggests the formation of smaller clusters with more frequent dissolution events than for the equivalent planar bilayer simulation. This can also be seen by comparing the mean cluster lifetime over the last 3 μs of simulation for the $\text{G}\alpha$ CG2/CG3 planar bilayer (mean $0.54 \pm 0.75 \mu\text{s}$) with that for the vesicle system (mean $0.27 \pm 0.32 \mu\text{s}$) (Table 4.2).

A key aspect of the behaviour of the $\text{G}\alpha$ protein models in the vesicles is that the protein is shaped approximately like a truncated cone (Fig. 4.1B). Thus packing together of $\text{G}\alpha$ proteins is likely to be influenced by the orientation of insertion of the proteins relative to the curvature of the membrane in the vesicle. In the $\text{G}\alpha$ CG2/CG3 simulations the protein was inserted with the narrower, C-terminal end of the truncated cone directed outwards, which would be anticipated to disfavour aggregation. Another 5 μs $\text{G}\alpha$ vesicle simulation ($\text{G}\alpha$ -inv CG2/CG3) was therefore performed, with the protein orientation inverted (i.e. with the narrower, C-terminal end directed inwards) which would be anticipated to favour aggregation. $\text{G}\alpha$ -inv CG2/CG3 shows greatly increased aggregation (Figs. 4.4 and 4.5), with two main clusters remaining tightly bound for the last 2 μs . It therefore seems that there is a complex interplay between protein/bilayer mismatch, protein shape, and membrane curvature which governs the stability of an aggregate. In contrast, the mean cluster duration for the hydrophobically matched $\text{G}\alpha$ CG4/CG5 system (for which the “clusters” correspond to brief encounters) is unchanged between the planar (mean $0.05 \pm 0.04 \mu\text{s}$) and vesicle (mean $0.06 \pm 0.06 \mu\text{s}$) systems.

4.3.3 Lateral Diffusion of Lipids and Proteins

Measurements of the lateral diffusion coefficients of the proteins and lipids were examined (Table 4.3), which are of interest given recent studies of protein diffusion in extended atomistic simulations of membrane proteins [171]. The trends which emerge are consistent, with slower diffusion of both protein and lipid in the longer chain lipid simulations, as has been seen in recent studies (e.g. [172]). The $\text{G}\alpha$ proteins also diffuse more quickly than the larger-diameter $\text{G}\beta$ proteins. The values obtained are comparable with those reported in CG simulations of rhodopsin [166] once one notes that the CG timescale has not been corrected by a factor of 4. Such a correction if applied would bring the diffusion rate for the lipids in the CG4/CG5

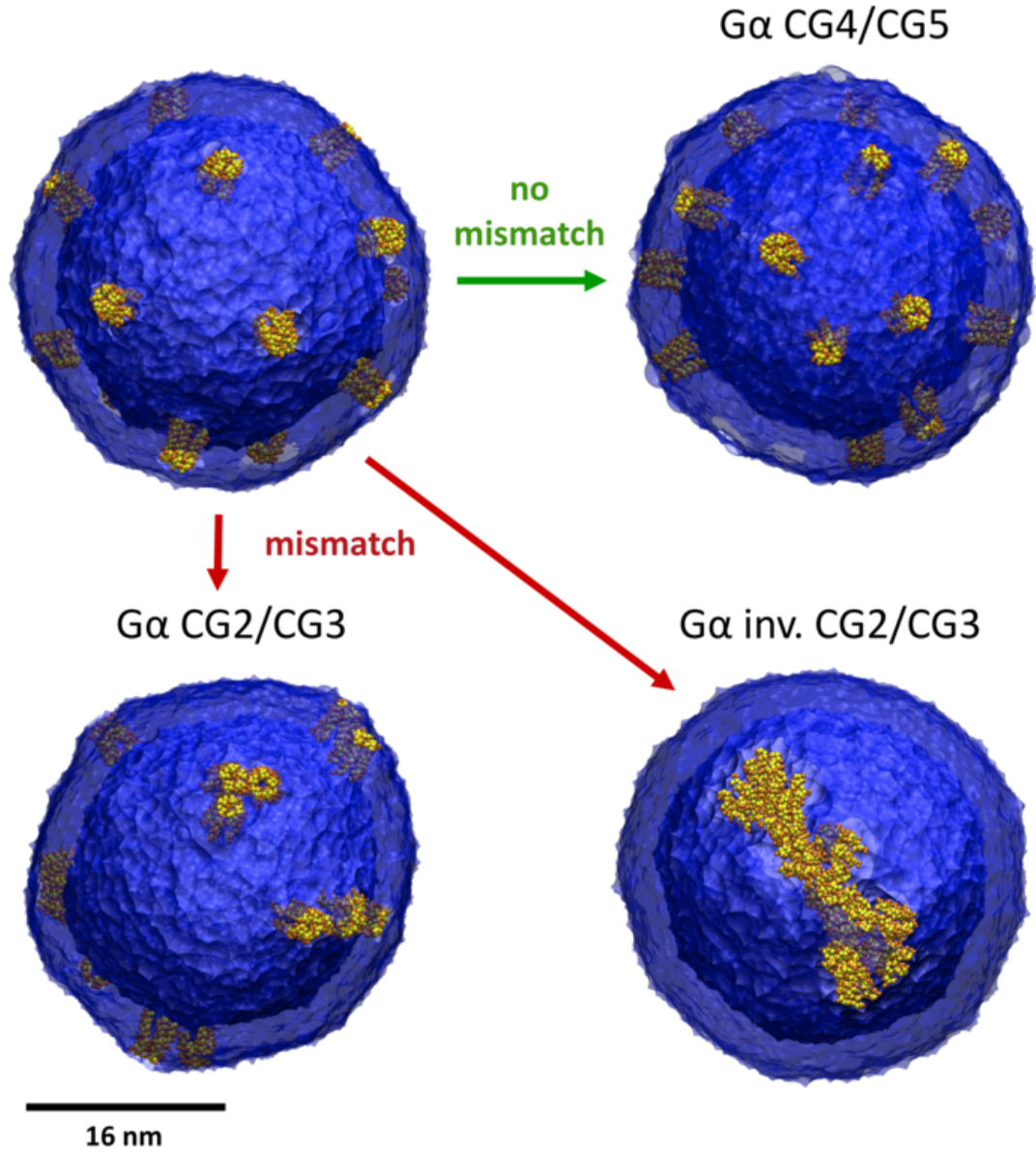


Figure 4.4: Simulations of 16 Gα TM proteins in vesicle membranes. The proteins are shown in yellow, the inner leaflet lipid phosphates in solid blue, and the outer leaflet lipid phosphates in transparent blue. The top left picture shows the starting configuration ($t = 0 \mu\text{s}$) for the Gα CG2/CG3 simulation; the proteins start as far apart from each other as possible, and this same configuration is used for the Gα CG4/CG5 and Gα-inv CG2/CG3 simulations. The other pictures show the simulations at $t = 5 \mu\text{s}$. In the Gα-inv CG2/CG3 system, the proteins are embedded in the membrane in the opposite orientation. The shape of the protein is thus a good match for the curvature of the membrane, leading to greatly increased aggregation.

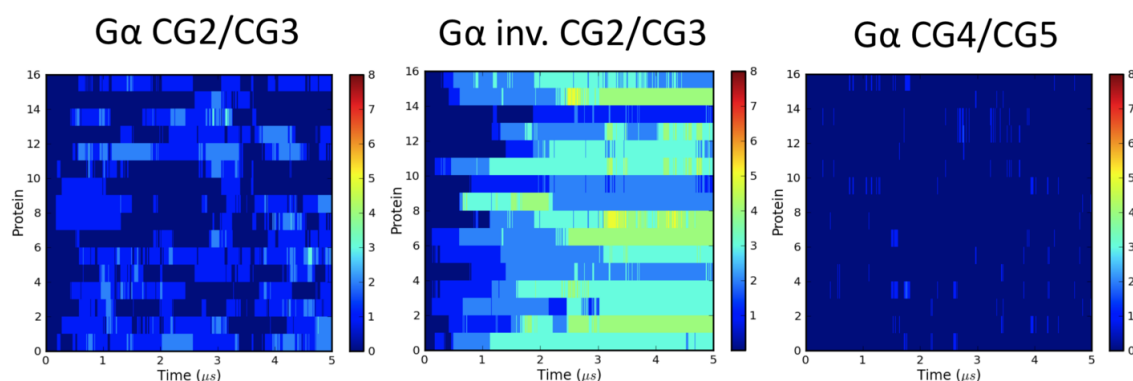


Figure 4.5: Aggregation analysis of the multiple TM protein vesicle membrane simulations. See Fig. 4.3 caption for details. The same scale of 0 to 8 is used for ease of comparison.

systems down to $\sim 3 \times 10^{-7} \text{ cm}^2\text{s}^{-1}$, which is comparable to e.g. the experimental estimate for POPC ($2 \times 10^{-7} \text{ cm}^2\text{s}^{-1}$) [173].

System	D ($10^{-7} \text{ cm}^2/\text{s}$)		
	<u>protein</u>	<u>shorter lipid</u>	<u>longer lipid</u>
<u>Gα bilayer</u>			
CG2/CG3	5.1	28	24
CG4/CG5	3.6	12	11
<u>Gα vesicle</u>			
CG2/CG3	5.6	25	21
Inv CG2/CG3	3.9	25	22
CG4/CG5	3.1	11	11
<u>Gβ bilayer</u>			
CG2/CG3	2.7	25	23
CG3/CG4	2.6	14	13

Table 4.3: Lateral diffusion coefficients, determined for the period 1 to 5 μs of each simulation. They are *not* corrected by a factor of 4 to adjust for the CG forcefield.

In the vesicle simulations, the Gα proteins diffuse at a similar rate to the equivalent bilayer simulations when oriented with their narrower C-terminal ends outwards. In the inverted orientation, however, diffusion is slowed considerably by extensive aggregation of the proteins. The correlation between aggregation and decreased diffusion rate can also be observed, for

example, in the $G\alpha$ CG2/CG3 bilayer simulation, in which the protein diffusion rate coefficient decreases from $6.5 \pm 1.2 \times 10^{-7} \text{ cm}^2\text{s}^{-1}$ over the period 1–2 μs (when the system forms mostly small clusters of 2–4 proteins) to $2.8 \pm 0.3 \times 10^{-7} \text{ cm}^2\text{s}^{-1}$ at 4–5 μs (2–3 large clusters).

Another feature observed was that the diffusion of lipids was slower when in close proximity to proteins. In the $G\alpha$ CG4/CG5 vesicle simulation, diffusion coefficients reduced from $11.3 \pm 0.4 \times 10^{-7} \text{ cm}^2\text{s}^{-1}$ for bulk lipids ($> 3 \text{ nm}$ from the centre of mass of a protein; either CG4 or CG5 type) to $7.8 \pm 0.3 \times 10^{-7} \text{ cm}^2\text{s}^{-1}$ for annular lipids ($\sim 3 \text{ nm}$ from the centre of mass of a protein). The annular lipids are not bound strongly to the proteins, and are freely exchanged for other lipids. The simulation results are in-keeping with electron spin resonance (ESR) spectroscopy measurements which found similar reductions in the rates of diffusion of annular phospholipids around a number of membrane proteins [174, 175]) and also agree with simulation studies suggesting that proteins diffuse concertedly with lipids as dynamic protein-lipid complexes [171].

4.3.4 Lipid Distribution Around the Proteins

A substantive degree of hydrophobic mismatch between a lipid bilayer and integral membrane proteins may result in a lipid sorting mechanism in mixed lipid bilayers (see e.g. [57, 101, 176]). Thus, with positive protein/bilayer mismatch (i.e. the hydrophobic length of the protein is greater than the hydrophobic thickness of the lipid bilayer) the longer tail lipids would be expected to accumulate in close proximity to the protein pore. Analysis of the nature of the lipids adjacent to the protein in these simulations (Fig. 4.6) shows that where one lipid matches the hydrophobic topology of a protein more closely than the other (i.e. $G\alpha$ CG4/CG5, $G\alpha$ -ves CG4/CG5, both $G\beta$ simulations), it will demix from others to increase its own concentration proximal to the protein. There are also cases where no de-mixing is evident (e.g. the $G\alpha$ CG2/CG3 simulations), which appears to be due to both lipids being poor hydrophobic matches for the protein. In general, the degree of lipid de-mixing appears to be independent of the extent of protein aggregation. The effect is also relatively small — lipids do not completely de-mix so that proteins are surrounded by only one type of lipid — which is in agreement with experiment [16].

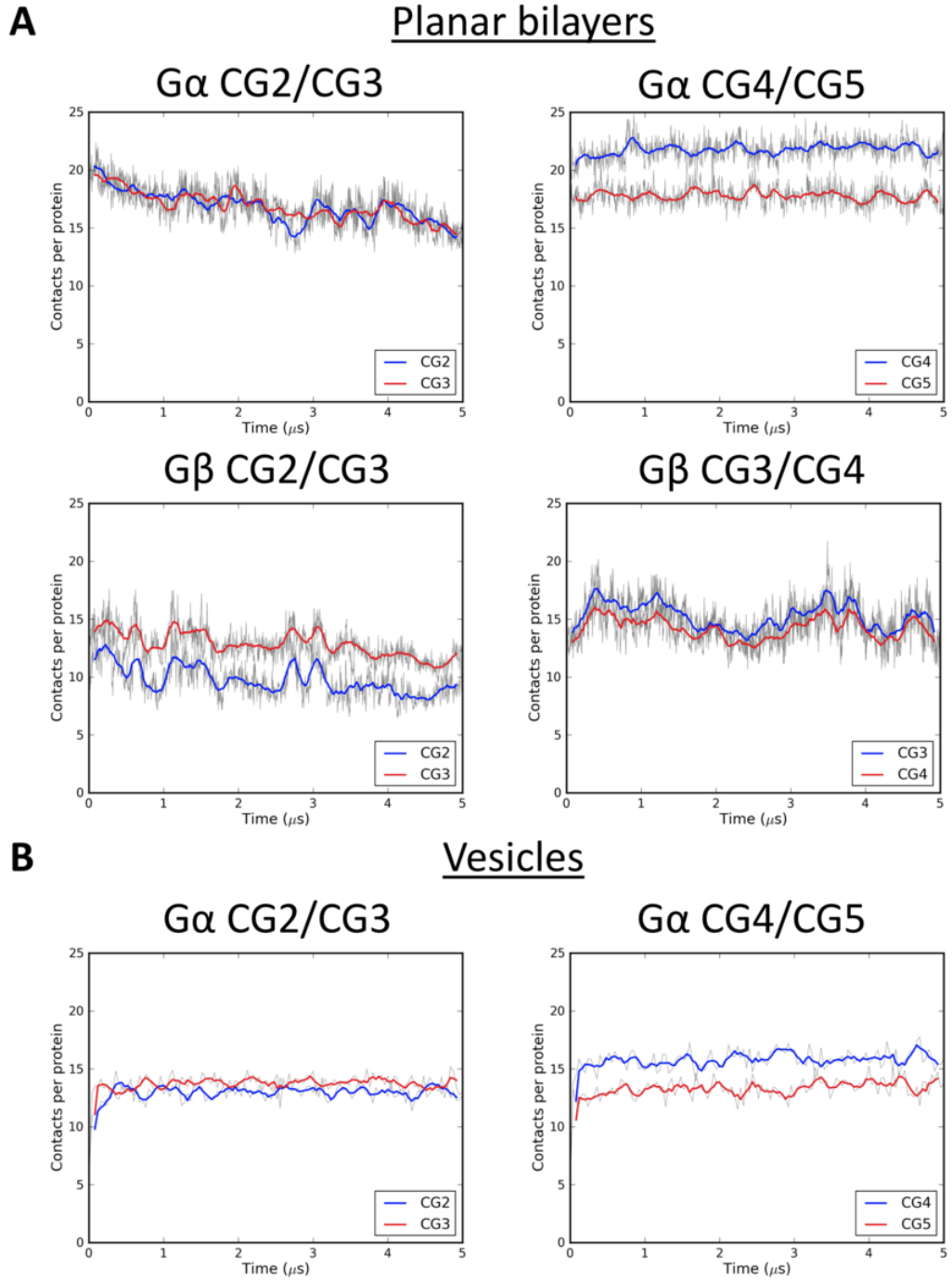


Figure 4.6: Lipids adjacent to the proteins. For each system the average (over 16 protein pores) numbers of each lipid species around the protein is shown as a function of time, using a cutoff distance determined from protein/lipid radial distribution functions. The grey lines represent the raw data, whereas the red/blue lines are calculated using a moving average over 20 ns. The results of this analysis are shown for *a*) the planar bilayer and *b*) the vesicle systems.

4.4 Discussion

This simulation study demonstrates how the extent and nature of aggregation of proteins within a lipid bilayer are sensitive to the degree of hydrophobic mismatch, the class of membrane protein (α -helix vs. β -barrel) and the curvature of the lipid bilayer. Relatively large systems have been examined, with a protein : lipid ratio of $\sim 1 : 450$. It is notable that many biological membranes have a protein : lipid mass ratio of $\sim 1 : 1$, which corresponds to a protein:lipid molar ratio of ~ 100 , and for an ‘average’ membrane protein the number of annular lipids is ca. 30 [177]. The protein models used attempted to reproduce the ‘canonical’ outer surface of a membrane protein, with hydrophobic (leucine) sidechains in the TM region and amphipathic (tryptophan) sidechains in the lipid/water interfacial region. Both the $G\alpha$ and the $G\beta$ proteins were approximately circularly symmetrical in cross section, and so this study is in many ways complementary to e.g. that of [166] who explored mismatch-driven aggregation of rhodopsin and demonstrated that there were preferred interaction regions on the surface of this more asymmetrical protein. Furthermore, mismatch-driven aggregation is considered in the absence of lateral phase separation of the lipids, in order to aid the investigation of individual features which make up part of a more complex system in biological membranes.

The level of protein aggregation was modulated primarily by the degree of hydrophobic mismatch and the membrane curvature, ranging from very strong ($G\alpha$ -inv CG2/CG3 vesicle simulation) to essentially non-existent ($G\alpha$ CG4/CG5 simulations). Intermediate levels of aggregation were also observed; in the $G\alpha$ CG2/CG3 vesicle simulation for example, clusters would form, break up and re-form multiple times (see cluster residency times in Table 4.2).

A number of other subtle effects were observed, depending on the exact nature of the protein and of the bilayer. Thus, the β -barrel protein seemed to prefer to form linear aggregates (‘strings’) within the bilayer plane, in contrast to the α -helical protein which formed extended two-dimensional clusters. This difference will merit further investigation with a wider range of membrane proteins — both simplified models (as here) and real membrane proteins. A further subtlety revealed by the vesicle simulations was that the shape and orientation of the protein relative to the vesicle bilayer could have profound influence on its aggregation properties. The relative importance of hydrophobic mismatch and curvature deformation have been discussed in the context of experimental studies of aggregation of e.g. rhodopsin

in membranes [157].

A range of other features observed can be compared with recent simulation studies. For example, aggregates were seen to diffuse more slowly than individual peptides, and annular lipids were slowed in comparison to bulk lipids. These findings are in keeping with the picture of concerted movement of protein-lipid complexes, as recently highlighted by an atomistic simulation study [171]. Local bilayer deformation was also observed in all cases where hydrophobic mismatch was present, as has been commented on in a number of previous studies (e.g. [61, 166, 178, 179]). In mixed bilayers where one lipid species matches a protein for hydrophobic length much better than another, a degree of lipid de-mixing about the protein has been observed. This is in contrast with recent simulations of a single α -helical peptide [180]. The difference between the two studies may be the larger size of the inserted protein in the current study, such that it will have a larger annulus of lipids.

There have also been a number of recent computational studies which have followed a similar overall theme of investigating mismatch-induced aggregation and related behaviours of membrane proteins. Several studies have used DPD simulations combined with more radically simplified models of membrane proteins. Thus, Smit and colleagues have conclusively demonstrated that protein aggregation can be simply a consequence of mismatch in simulations with no attractive potential between proteins [62] and that such clustering may be modulated by the presence of cholesterol in the membrane [63]. Again using DPD and a simple ‘cylindrical’ model of membrane proteins, Weiss and colleagues have explored mismatch-induced aggregation in the context of membrane protein sorting [54, 57] and have suggested that protein acylation may modulate such aggregation [59].

As mentioned above, Marrink and colleagues have used CG-MD simulations in a detailed study of mismatch-induced rhodopsin clustering and its possible relationship to the importance of GPCR dimers in signalling [166]. CG-MD has also recently been used to study hydrophobic (WALP) α -helices in a complex mixed lipid systems (diC16:0-PC/diC18:2-PC/cholesterol), revealing that the peptides prefer to cluster in the liquid-disordered (Ld) phase rather than enter the liquid-ordered (Lo) phase [101]. Single WALP helices have also been studied in atomistic simulations in terms of the effect of hydrophobic mismatch and helix tilt [181].

It is important to consider the possible limitations of the current studies. One is the use

of a CG forcefield which inevitably approximates the nature of the protein/lipid and protein/protein interactions. However, comparisons with available experimental data suggest that this forcefield accurately reproduces protein/lipid interactions for a number of membrane proteins [182]. Recent studies on the free energies of interactions of transmembrane α -helices using the Martini CG forcefield [183, 184] have resulted in potentials of mean force for helix/helix interactions comparable to those obtained by earlier all atom simulations [185]. In the context of the CG forcefield, there has also been discussion [168, 169]) of time steps in CG-MD simulations. Test simulations were therefore conducted on a G α CG4/CG5 system (see Appendix A for details) with time steps of 5, 10, 20 and (as in the main study) 40 fs. A number of structural and dynamic properties of the system were evaluated, and no statistically significant differences in behaviour were observed for the different time steps. The results therefore appear to be robust to the exact choice of time step.

The second major limitation is that the studies have been restricted to relatively simple systems, in terms of the protein and lipid species, and to relatively small systems, in terms of vesicle dimensions, compared to genuine biological complexity. The proteins simulated here are canonical models of the most common types of TM protein. While experimental work has reported mismatch-induced aggregation of simple TM peptides [153] and more complex membrane proteins [157], other studies have hinted that membrane proteins are not always affected by bilayer thickness [175, 186]). In terms of lipid complexity, it is reassuring that comparable simulations have now been performed for ternary mixtures of lipids [101], but there remains much to be done in bridging the gap between *in silico* and *in vivo* studies of protein aggregation within biological membranes. Regarding the size of vesicles in these simulations, they are comparable to e.g. synaptic vesicles (diameter 30–80 nm) [187] whilst a little smaller than e.g. the influenza virion (diameter $\sim$ 120 nm) [68]. Based on comparison with the planar membrane simulations, it seems likely that the modulation of protein aggregation by membrane curvature would reduce with increasing vesicle size. A more detailed investigation via simulation of a wider range of vesicle sizes would be of some interest.

A number of test simulations of small vesicles (see Appendix C for details) were conducted to evaluate whether the self-assembly procedure enabled correct equilibration of the number of lipids between the inner and outer leaflets. For 15 nm diameter vesicles the mean (over 10 simulations) fraction of inner leaflet lipids was 0.275, with a standard deviation of 0.004. The latter represents approximately 3–4 lipids per vesicle. This suggests that while the self-

assembly of lipid vesicles may result in small deviations from the ideal lipid composition, such deviations are not sufficient to alter the results presented.

Reflecting on the likely biological relevance of these studies and related efforts from a number of other groups and approaches (see summary above), they confirm the complex, coupled interplay of protein/protein, protein/lipid, and lipid/lipid interactions in membranes which may lead to membrane protein aggregation. Such effects are likely to be of importance both for experimental studies of simpler systems consisting of single species of membrane proteins reconstituted *in vitro* (e.g. [157]) and for more complex, biological membranes containing multiple lipid and protein species. In particular, simulations of protein aggregation in ‘simple’ systems might be usefully extended to explore aspects of the likely mechanisms of 2D crystallization of membrane proteins [188]. Indeed, these simulations with simple models have suggested different cluster geometries for different proteins, which is of possible interest in this context. However, a detailed understanding of the nature of clusters formed will require a more systematic examination of a wider range of proteins than in this study.

In terms of *in vivo* relevance, the results have bearing on the understanding of lateral complexities in membrane structures such as rafts etc. [189, 190] and are also of possible relevance to membrane protein sorting [57, 176]. In future studies it will be important to use simulations to explore the behavior of such complex systems which exhibit considerable lipid and protein heterogeneity.

Chapter 5

Raft-like Aggregates of Influenza Hemagglutinin Simulated by Coarse-Grained Molecular Dynamics

5.1 Introduction

Influenza virus A is bounded by a lipid envelope which contains three transmembrane (TM) proteins: hemagglutinin (HA), neuraminidase (NA) and the M2 channel. The envelope is derived from the plasma membrane of the host cell during virion exocytosis. It is characterized by a high protein concentration (~ 13 glycoproteins per $40 \times 40 \text{ nm}^2$ area, as shown by cryo-electron tomography images [68]) and enrichment in saturated phospholipids and cholesterol [191–194]. These features are thought to originate from the association of HA with putative lipid rafts in the plasma membrane, prior to viral budding [192].

Lipid rafts can be generally described as small ($< 100 \text{ nm}$ diameter), fluctuating patches of membrane enriched in saturated phospholipids, cholesterol and certain proteins, including most GPI-anchored and acylated proteins, and some TM proteins. However, their exact nature is subject to intense discussion in the literature, particularly due to the lack of their direct visualization *in vivo* [2].

Early evidence for the association of HA with lipid rafts arose from isolation of detergent-insoluble glycolipid complexes (DIGs) from the plasma membranes of influenza-infected cells, which were found to incorporate HA [195, 196]. More recently, FRET studies have shown HA to associate with labelled raft-markers, such as GPI-anchored proteins in the extracellular leaflet of the plasma membrane [29], and acylated proteins in the intracellular leaflet [30]. These studies also highlight how raft association can be influenced by palmitoylation of HA at residues in the cytoplasmic domain, and by mutation of hydrophobic amino acids towards the extracellular side of the TM domain. Other studies have provided direct visualization of aggregates of HA in the plasma membrane via immunogold-labelling electron microscopy (EM), but without definitive evidence of raft-association [197, 198].

Rafts have been compared to liquid-ordered (Lo) domains in model membranes, which form on much larger scales and have thus been employed in many experiments as model raft systems [2]. The archetypal domain-forming model membrane comprises a ternary mixture of saturated phospholipid (often PC or sphingomyelin), unsaturated phospholipid and cholesterol, which will undergo spontaneous temperature-dependent separation into Lo (enriched in saturated lipids and cholesterol) and Ld (liquid-disordered; enriched in unsaturated lipids) domains [33]. Lateral phase segregation is thought to be driven by the preference of cholesterol for association with saturated lipid tails, which can adopt a favourable ordered conformation when adjacent to the rigid, planar sterol ring [199]. The Lo phase is therefore distinguishable from the Ld phase by increased phospholipid tail ordering, but (unlike the solid-ordered So (gel) phase) without a drastic decrease in lateral mobility; the lateral diffusion coefficient is reduced by a factor of $\sim 2-3$ [34].

A number of important differences separate the behaviour of TM proteins in Lo domains in model membranes from their behaviour in lipid rafts in plasma membranes, and their direct comparison has been controversial. Notably, all known raft-associating TM proteins partition into the Ld domain in model membranes, rather than the Lo domain as might be expected for a true raft-mimic [38–41]. This has recently been confirmed for HA [37]. Two possible reasons for this disparity are the much higher protein concentration present in the plasma membrane, and interactions with cytoskeletal components.

Experimental approaches to date have lacked the ability for direct observation of the structure and dynamics of lipid rafts and their association with TM proteins on the molecular scale. Molecular dynamics (MD) simulations can in principle provide this information, and have

been used in a number of studies to investigate domain-forming membranes in atomistic detail [200–202]. However, many important biological processes require longer length and time scales than are accessible through traditional atomistic simulations. Coarse-grained (CG) force fields, such as the popular MARTINI model [44, 45], improve computational efficiency by reducing the number of degrees of freedom in a simulation system. A 2008 study showed that the MARTINI model can reproduce the domain-forming properties of model membranes composed of ternary lipid mixtures [203]. A number of other studies followed, investigating concerted lipid diffusion within domains [201], the influence of lipid domain properties and leaflet asymmetry on inter-leaflet coupling [204], and showing how simple TM peptides partition into Ld domains in model membranes [101, 205].

The aim of this study is to investigate the molecular details of the association between HA aggregates and raft-type lipids in domain-forming membranes. The CG MD simulations described here provide a molecular-level insight into this problem. They show that raft-type lipids are accumulated within HA aggregates at the expense of unsaturated lipids. The driving force appears to be an entropic penalty for partitioning of unsaturated lipids within the aggregate interior, introduced by an attenuation in lateral diffusion in that region. The results suggest that high local protein concentration can be sufficient for association of HA with rafts in the plasma membrane.

5.2 Methods

5.2.1 Hemagglutinin model

The model of HA used in this work was derived from a structure of the protein from the A/duck/Ukraine/1/63 (H3N8) influenza strain solved by X-ray crystallography (PDB code: 1MQM) [206]. The crystal structure includes only the ectodomain, which terminates with a short α -helix at the C-terminal end, roughly perpendicular to the protein long axis (shown in orange in Fig. 5.1). The subsequent sequence includes a twelve-residue linker, followed by an α -helical TM region and an unstructured cytoplasmic domain; these regions were modelled as described below. The TM domain sequence is WILWISFAISCLLLCVLLGFIMWACQ.

Firstly, the ectodomain structure was converted to a CG representation (as shown in orange in Fig. 5.1) using the MARTINI tools [44, 45]. The TM domain, linker, and last five residues

of the ectodomain C-terminal helix were modelled as an ideal α -helix (residues 486–527; annotated in Fig. 5.1). This was added to the ectodomain via alignment with the resolved C-terminal helix. A bend in the helix was required to orient the TM domain directly downwards toward the membrane. This was achieved by modelling the helix with a simple elastic network model, then adding harmonic restraints to bend the helix into the correct orientation, with rigid constraints to retain ideal α -helical structure in the TM domain.¹

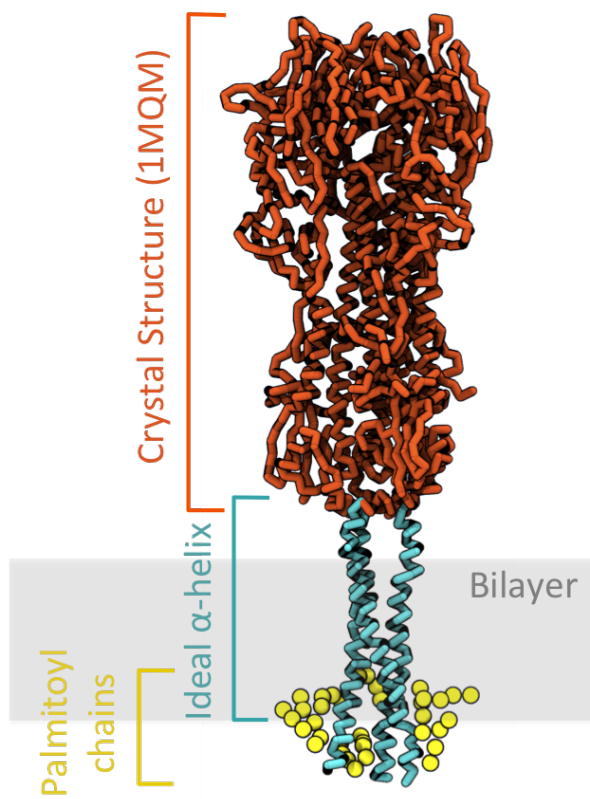


Figure 5.1: Annotated diagram of the CG HA model. The ectodomain structure (orange) was converted to the CG representation from the structure solved by X-ray crystallography (PDB code: 1MQM) [206]. The cytoplasmic domain, TM domain and TM domain-ectodomain linker are shown in light blue. These were modelled as an ideal α -helix, although the cytoplasmic domain was treated as an unstructured region in the simulations, by removing secondary structure restraints. The palmitoyl chains are shown in yellow.

The process was repeated for all three of the protein subunits, and the resulting structure was inserted into a $15 \times 15 \text{ nm}^2$ pre-formed CG bilayer (comprising 0.35 : 0.35 : 0.3

¹The ENM software was developed by Alex Tek (IBCP, Paris), who also aided in its application to the protein modelling.

DPPC/DUPC/cholesterol, and formed using the same process described below for other bilayers) using *g_membed* [207]. The initial position of the the protein in the z-axis (bilayer normal) was determined by matching the central TM domain residue (Leu516, according to [197]) to the bilayer centre. System parameters were created using the MARTINI CG force field [44, 45], with H-bond restraints applied to the TM domain and linker, and elastic network model (ENM) restraints applied to the ectodomain (the two types of restraints are described below). The system was energy minimized using < 500 steps of the steepest descents algorithm, then simulated with MD for 1000 ns. The use of H-bond restraints allowed the TM domain and linker to flex and rotate without deviating from their α -helical secondary structure. The last 500 ns of simulation were clustered using the single-linkage algorithm to find the most representative structure. As can be seen in Fig. 5.1, the three helices coiled around each other somewhat. This allowed the four polar residues in the TM domain (Ser508, Ser512, Cys513 and Cys517), which are positioned roughly on the same face of the helix, to be buried away from the hydrophobic lipid tails.

The cytoplasmic tail was initially modelled as an ideal α -helix (residues 523–539) and added to the equilibrated TM domain structure by aligning the helix backbone particles. However, restraints were not added and coil-type parameters were applied, reflecting the largely unstructured nature of the cytoplasmic region (according to a sequence analysis using the HHpred server [208]). Palmitoyl chains were then added at Cys528, Cys535 and Cys538, initially pointing directly outwards from the helix backbone. The chains each comprised four CG particles, with the same parameters as the palmitoyl chains of DPPC. The complete protein structure was then inserted into a $15 \times 15 \text{ nm}^2$ domain-forming bilayer as before. The system was energy minimized, and three 1000 ns MD simulations were carried out with different randomly generated starting velocities, with ENM restraints on the ectodomain, H-bond restraints on the linker and TM domain, and no restraints on the cytoplasmic tail. The unstructured cytoplasmic tail stayed mostly associated at the membrane interface, and was generally oriented outwards from the TM domain centre in the plane of the membrane. This allowed the attached palmitoyl chains to insert into the bilayer, adopting similar positions to those of the lower leaflet lipid tails. The final snapshot of the HA structure from one of these simulations was used for all subsequent production simulations.

The ectodomain of HA is glycosylated in a number of positions. Although MARTINI CG parameters have been published for carbohydrates [209], they have not yet been applied to

the study of glycoproteins. Glycans were therefore not included in the HA structure used in this work. The possible effects on simulated properties are discussed later in the Chapter.

5.2.2 Production simulation set-up

Three types of lipid were used in the simulation. As described previously, the MARTINI CG force field models the DPPC palmitoyl tails with four CG particles each, with equilibrium bond angles set to 180° . DUPC represents dilinoleylphosphatidylcholine (the “U” denoting poly-*unsaturated*, for the 18:2 linoleyl chains), for which the tails are modelled with four CG particles: two central double bond-type particles representing sp^2 carbons, surrounded by single bond-type particles representing sp^3 carbons. The three particles closest to the glycerol particles are given a 100° equilibrium bond angle, and the three terminal particles are given a 120° equilibrium bond angle. The MARTINI structure of cholesterol comprises eight CG particles, with a polar-type particle to represent the alcohol headgroup, five hydrophobic ring type particles to represent the planar ring motif, and two hydrophobic particles to represent the tail.

The $20 \times 20 \text{ nm}^2$ bilayers used in simulations with a single HA protein were initially formed using the Packmol software [210], which allowed copies of individual lipid molecules to be packed within the necessary spatial constraints. This method also allowed control over the lipid composition in each bilayer leaflet. The DUPC bilayer comprised 1024 lipids. The domain-forming bilayer comprised 560 DPPC molecules, 560 DUPC molecules and 480 cholesterol molecules² (a ratio of 0.35 : 0.35 : 0.3). The bilayers were each equilibrated with a 100 ns CG simulation at 323K. In the case of the domain-forming bilayer, this temperature was useful in preventing large-scale domain formation from occurring prior to protein insertion.

The larger $50 \times 50 \text{ nm}^2$ patches of domain-forming membrane, used for simulations of HA aggregates, were assembled by tiling smaller $12 \times 12 \text{ nm}^2$ patches which were themselves assembled and equilibrated in the same way as described above for the $20 \times 20 \text{ nm}^2$ patches. The $50 \times 50 \text{ nm}^2$ patches were equilibrated for 100 ns, and comprised approximately 3706 DPPC molecules, 3706 DUPC molecules and 3130 cholesterol molecules (the same 0.35 : 0.35 : 0.3 ratio). This process was conducted for each of the four membrane patches used in simulations of HA aggregates, with different random seeds for the initial packing carried out

²More lipids were required to create the same bilayer dimensions as the pure DUPC bilayer because cholesterol and DPPC take up less membrane area than DUPC.

with Packmol.

HA proteins were inserted into the pre-equilibrated bilayers using the g.membed tool [207]. This process removes a small number of lipids from the bilayer to make space for a TM protein. The initial positioning of proteins in the z-axis (the bilayer normal) was matched to that of the earlier simulations used in development of the HA model. The resulting structures were energy minimized for ~ 500 steps of the steepest descents algorithm, and then equilibrated with 100 steps of CG simulation. All simulated systems were solvated with MARTINI CG water, including 5% “antifreeze” particles to prevent unwanted nucleation of freezing (see Methods chapter for further details). Na^+ counterions were added randomly in solution to neutralize the 9 negative charges on each HA protein.

5.2.3 Simulation details

The GROMACS simulation software was used for all simulations [85–88]. The MARTINI CG parameters were used for all proteins, lipids, ions and solvent. A time step of 10 fs was used. Coulomb interactions were shifted to zero between 0 and 1.2 nm, and Lennard-Jones interactions were shifted to zero between 0.9 and 1.2 nm. The temperatures of the proteins, lipids and solvent were each coupled separately using the Berendsen algorithm [98] at 295K (or 323K for equilibration simulations) with a coupling constant τ_T of 1 ps. The system pressure was coupled semi-isotropically using the Berendsen algorithm at 1 bar with a coupling constant τ_P of 1.1 ps and a compressibility κ of $1 \times 10^{-6} \text{ bar}^{-1}$.

In all production simulations, structured regions of HA were restrained using an elastic network model (ENM). The ElnDyn ENM tools for the MARTINI model [170] were used to generate restraints between all backbone particles more than 3 residues apart, and less than 1.4 nm apart. The initial distance between the backbone particles was used as the reference distance for each restraint, and a force constant of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ was used. In certain simulations used for development of the HA model, α -helical structures were maintained using “H-bond” type restraints. These were applied between backbone particles 4 residues apart, mimicking the hydrogen-bonds which naturally occur in α -helices. Reference distances were matched to the initial distances between backbone particles, and force constants were set to $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$.

All analysis was performed using Python with the MDAnalysis library [211]. Visualization

was performed and figures produced with VMD [133].

Simulation times and time-dependent properties (such as diffusion coefficients) are quoted from the simulation time reported by the GROMACS MD software, and have *not* been adjusted to account for the faster sampling of the CG force field.

5.3 Results and Discussion

5.3.1 Single HA protein in a DUPC bilayer

To study a simple protein-membrane system in the absence of domain-forming properties, a single HA protein was simulated in a DUPC bilayer for 2000 ns (Fig. 5.2). The protein TM domain stayed situated within the bilayer as expected, and the short unstructured cytoplasmic domains (residues 530–539) were positioned at the membrane interface, roughly perpendicular to the TM domain. This allowed the nine palmitoyl chains (three per subunit, bonded to Cys528, Cys535 and Cys538) to be incorporated into the intracellular (IC) leaflet monolayer.³ The protein was able to tilt dynamically within the membrane, adopting an average tilt angle of $16^\circ \pm 6$ relative to the bilayer normal.

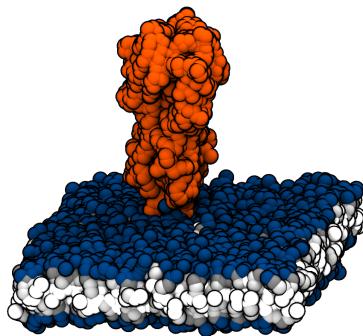


Figure 5.2: Snapshot of a single HA protein in a DUPC bilayer after 2000 ns CG simulation. The protein is shown in orange, the lipid headgroups in dark blue, and the lipid tails in grey and white.

The effect of the protein on the dynamic behaviour of the membrane was analyzed (Fig. 5.3).

³The terms “intracellular” (IC) and “extracellular” (EC) are used in this Chapter to distinguish between the two membrane monolayers, based on their position relative to HA. However, it should be noted that there was no compositional difference between the monolayers or the solvent on either side of the membrane.

Lipid diffusion coefficients, calculated at intervallic distances from the protein, show a roughly exponential decrease in diffusion rates with increasing proximity to the protein. At distances beyond 7 nm, diffusion coefficients converge to $\sim 11 \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$, approximately double that for the lipids closest to the protein ($D \simeq 5.5 \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$). This is likely to be due in part to hydrodynamic effects, i.e. the slow-diffusing protein ($D = 1.5 \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$) attenuates the diffusion of nearby lipids. Intermolecular forces between lipids and the protein may also play a role. Interestingly, there was little difference in diffusion rates between the two leaflets, despite the palmitoyl chains situated in the IC leaflet. A relatively small effect on lipid tail ordering was observed; for the closest lipids (1–2 nm from the protein), a slight increase in the range of angles adopted can be observed for the EC leaflet, and a slight decrease in average angle (relative to the bilayer normal) for the IC leaflet. The likely explanation here is that in the EC leaflet the irregular surface of the protein TM domain causes nearby lipid tails to adopt a wider range of conformations, while in the IC leaflet the protein surface is shielded by the protein-attached palmitoyl chains, which induce a slight ordering effect in nearby lipids. Lipid tails beyond ~ 3 nm from the protein appeared unaffected.

5.3.2 Single HA protein in a domain-forming bilayer

To investigate the behaviour of a domain-forming membrane, a simulation was performed with HA in a bilayer comprising 0.35 : 0.35 : 0.3 DPPC/DUPC/Chol. Ternary mixtures of saturated phospholipid, unsaturated phospholipid and cholesterol are frequently employed in model membrane studies of lipid domains, and recent CG MARTINI simulations with similar compositions have been shown to phase separate into Lo and Ld domains [203]. Initially, the lipids were randomly arranged in the bilayer. By the end of the 3000 ns simulation, the lipids were almost entirely segregated in two domains, one composed of DUPC and the other of DPPC and cholesterol, with each domain taking up approximately half of the membrane area. In contrast with previous CG MARTINI simulations of simple α -helical TM peptides which exhibited partitioning into Ld domains [101, 205], the larger and more asymmetric HA protein remained situated at the interface between domains (Fig. 5.4). In the EC leaflet, the protein was surrounded by a short, dynamic annulus of DUPC (~ 2 lipids), while in the IC leaflet it was surrounded by a short, dynamic annulus of DPPC and cholesterol. The protein diffusion coefficient was $4.7 \times 10^{-8} \text{ cm}^2 \text{ s}^{-1}$.

Analyses of lipid diffusion coefficients and tail ordering are shown in Fig. 5.5. As in the

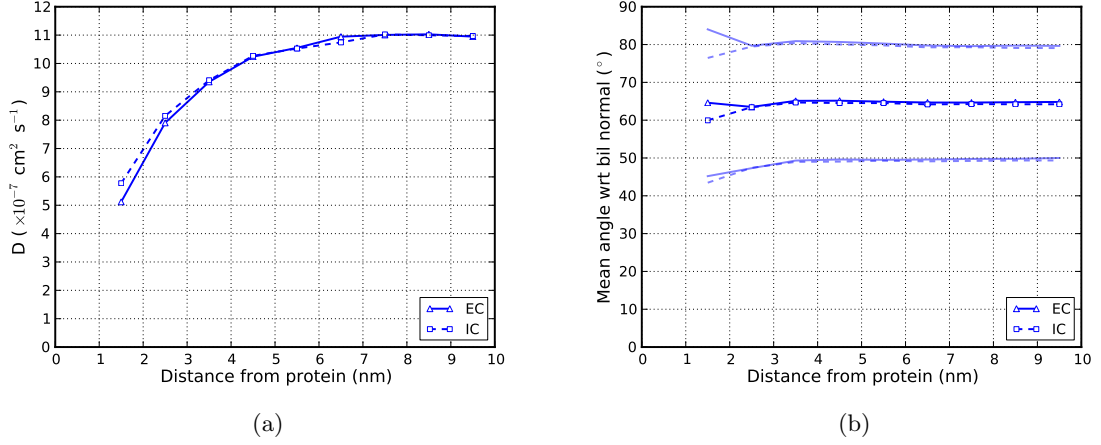


Figure 5.3: Effect of an HA protein on lipid diffusion coefficients and tail ordering in a DUPC bilayer. (a) Lipid diffusion coefficients, with respect to distance from the HA protein. Lipids were placed within two-dimensional concentric rings of width 1 nm, radiating out from the geometric centre of the protein TM domain. The approximate radius of the TM domain was 1 nm, thus data is only available for the second ring onwards. Displacements $\mathbf{r}$ were measured using the lipid geometric centres, and converted to diffusion coefficients with the following equation: $D = \frac{1}{4\Delta t} \langle |\mathbf{r}(\Delta t) - \mathbf{r}(0)|^2 \rangle$ where the averaging was conducted over all lipids and time over 1000–2000 ns. Displacements were measured over a range of timesteps Δt , and in all cases the value which gave the average displacement closest to 2 nm was used to calculate D , to avoid inclusion of fast non-diffusional movements of lipids which take place over shorter length scales. (b) Lipid tail ordering, represented by the average of the angles between the two terminal lipid tail bonds (those between the second and third tail CG particles, and the third and fourth) and the bilayer normal. The opaque solid and dashed lines show the angles averaged over 1000–2000 ns, and the transparent lines show ± 1 mean deviation. EC = extracellular leaflet, IC = intracellular leaflet.

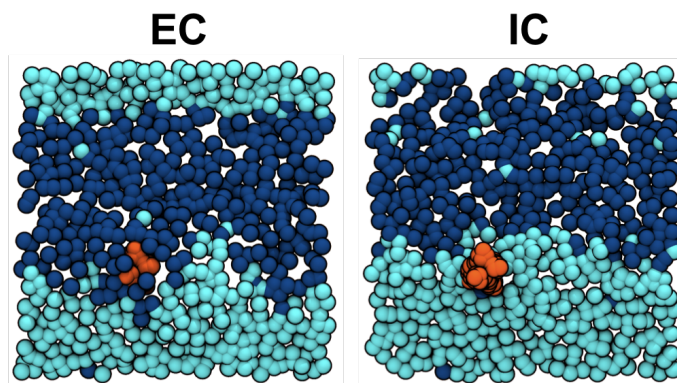


Figure 5.4: Snapshots of a single HA protein in a domain-forming bilayer after 3000 ns, shown with a top-down view from the EC side. Only the TM region of the protein is shown (orange). The lipid phosphate particles are shown in dark blue (DUPC) and light blue (DPPC), with only the EC leaflet visible in the left image, and only the IC leaflet visible in the right image. Cholesterol is not shown for clarity, but is generally associated with the same regions as DPPC (their preferential interaction being the major driver for domain segregation).

previous simulation with HA in a pure DUPC bilayer, the diffusion coefficients for DUPC plateau beyond 7 nm from the protein, and beyond 3 nm for the lipid tail ordering. Compared to the pure DUPC simulation, the diffusion coefficients in this bulk zone are decreased, while the lipid tail ordering is relatively unaffected. This behaviour is in keeping with that expected for Ld domains [2]. In proximity to the protein, a dramatic decrease in diffusion coefficient is again observed. A slightly stronger effect can be seen for the IC leaflet, which is likely due to the increased presence of DPPC and cholesterol in this area. In contrast, the diffusion coefficients for DPPC are relatively unaffected by proximity to the protein in the EC leaflet, with a small decrease in the IC leaflet. In terms of lipid tail ordering, DUPC shows similar behaviour to the pure DUPC bilayer simulation, with slightly increased ordering in the IC leaflet. DPPC shows a significant decrease in ordering in the EC leaflet (likely due to the increased concentration of DUPC around the protein and/or unfavourable interactions with the irregular protein surface), and relatively little change in the IC leaflet.

Fluorescence microscopy experiments have indicated that HA partitions into Ld domains in model membranes [37]. Partitioning of HA at the domain boundary in the simulation reported here may be due to the small size of the domains and the presence of only a single protein. Large domains which form in model membrane experiments have much shorter perimeter lengths relative to their area, so it is possible that the boundaries are easily saturated with

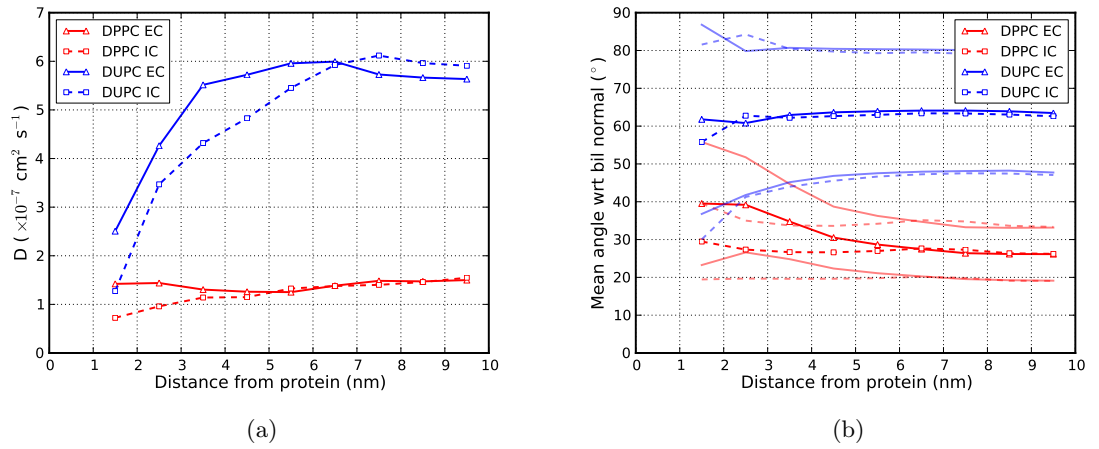


Figure 5.5: Effect of an HA protein on (a) lipid diffusion coefficients and (b) tail ordering in a domain-forming bilayer. The analyses were conducted as described in the caption of Fig. 5.3, with the exception that time averaging was applied over 2000–3000 ns, and a modification was introduced to exclude lipids which were not in the “expected” domain. Taking DPPC as the example, a lipid was only analysed in a particular frame if 60% of lipids within 1 nm (measured via phosphate-phosphate distances) were also DPPC. This removed the influence on the analyses of DPPC lipids situated in Ld domains, which were unstable and diffused much faster than those in Lo domains. Relatively few DPPC lipids were situated in Ld domains, but introduced a slight skew in the diffusion analysis data, which can be considered as anomalous for the relevant discussion below. The same explanation applies vice versa to DUPC. The modification had little effect on the lipid tail ordering analysis.

HA, and excess HA prefers to partition into the Ld domain. Alternatively, HA may be dynamically distributed between Ld domains and domain boundaries, with a slight preference for the latter. This type of behaviour (a loose preference for domain boundaries) has been directly observed in an AFM of study of the raft-associating acylated protein N-Ras [212].

5.3.3 Aggregate of ten HA proteins in a domain-forming bilayer

A set of simulations were performed to investigate the behaviour of HA proteins in a high concentration area within a domain-forming membrane. A patch of HA proteins was initially arranged in an array with an average spacing of ~ 10 nm, and inserted in a pre-formed 50×50 nm² patch of membrane comprising the same ternary mixture of 0.35 : 0.35 : 0.3 DPPC/DUPC/cholesterol (Fig. 5.6). The area of the membrane patch was made considerably larger than the area of the protein aggregate, to avoid influence of the periodic box on aggregate properties and to allow lipids within the aggregate to exchange with those outside. This allowed the investigation of HA in the context of a single aggregate within a larger domain-forming membrane, as opposed to investigating a membrane entirely crowded with protein (as in a recent study [205]).

Four 3000 ns simulations were conducted using this set-up, each starting with a different random lipid arrangement. A number of snapshots relevant to the following discussion are shown in Fig. 5.6. Most proteins quickly interacted with each other by tilting and making contacts between ectodomains. Importantly, the bulky ectodomains prevented direct contacts between TM domains, and lipids were always present between any pair of proteins. Ectodomain contacts were dynamic; they were able to separate but contacts were usually quickly made with another protein. The lateral diffusion coefficients for these proteins, $D = 2.9 \times 10^{-8} \text{ cm}^2 \text{ s}^{-1} \pm 0.6 \times 10^{-8}$ (calculated individually for each simulation, then averaged), were thus reduced compared to the previous single-protein simulations ($D = 4.7 \times 10^{-8} \text{ cm}^2 \text{ s}^{-1}$ in the domain-forming bilayer). Two cases were observed (in simulations *a* and *c*) where one protein did not make contact with any other protein during the simulation (ringed in red in Fig. 5.6). These proteins diffused slightly faster, with a coefficient of $3.9 \times 10^{-8} \text{ cm}^2 \text{ s}^{-1} \pm 1.2 \times 10^{-8}$.

As domains formed in the membrane, the proteins associated at the interfaces. As with the single-protein simulation, proteins retained an annulus of DUPC in the EC leaflet, and an

annulus of DPPC and cholesterol in the IC leaflet.

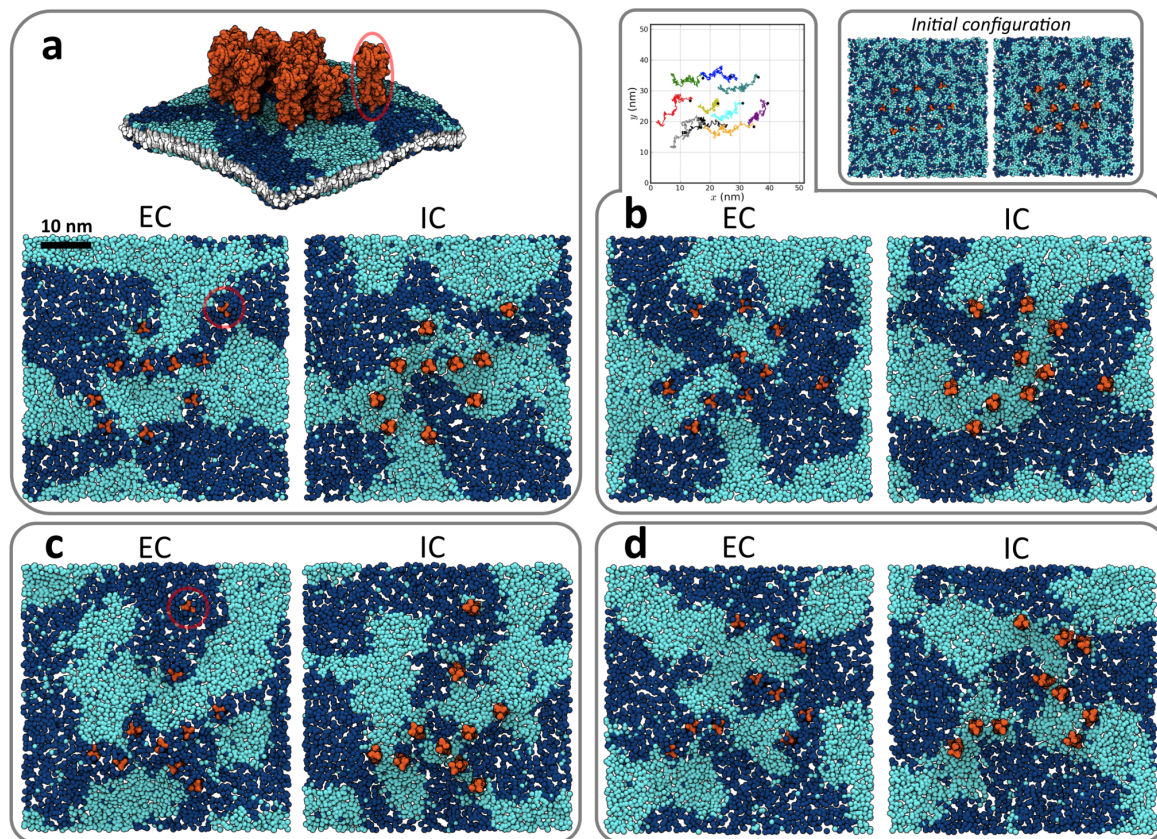


Figure 5.6: Top-down snapshots of four CG simulations (labelled *a-d*) of aggregates of ten HA proteins in domain-forming membranes after 3 μ s (scale bar = 10 nm). DUPC phosphate headgroups are shown in dark blue, DPPC phosphate headgroups in light blue, and protein TM domains in orange (ectodomains and cytoplasmic domains not shown). The protein aggregates have been centered in the images shown. For simulation *a*, an oblique view including the full protein structures is also shown. The image has been rendered without perspective and indicates a slight outward curvature of the membrane in the EC direction. For simulation *b*, the inset graph shows the trajectories of the ten proteins (shown in different colours) over time, with the black dots indicating $t = 0$. The top-right panel shows the starting configuration for one of the simulations.

The evolution of lipid composition within the protein aggregates is shown in Fig. 5.7. The areas of membrane included in this analysis were defined by drawing triangles between each protein and taking their combined area, while ignoring triangles with any side longer than 14 nm. The latter part of the algorithm was used to exclude the small number of lipids which were technically within the cluster but more than ~ 7 nm from a protein, and thus unlikely to be affected in terms of diffusion rate or tail ordering according to the data from the previous

single-protein simulations. The analysis showed a general increase in the ratio of DPPC in either leaflet over the course of the simulations.

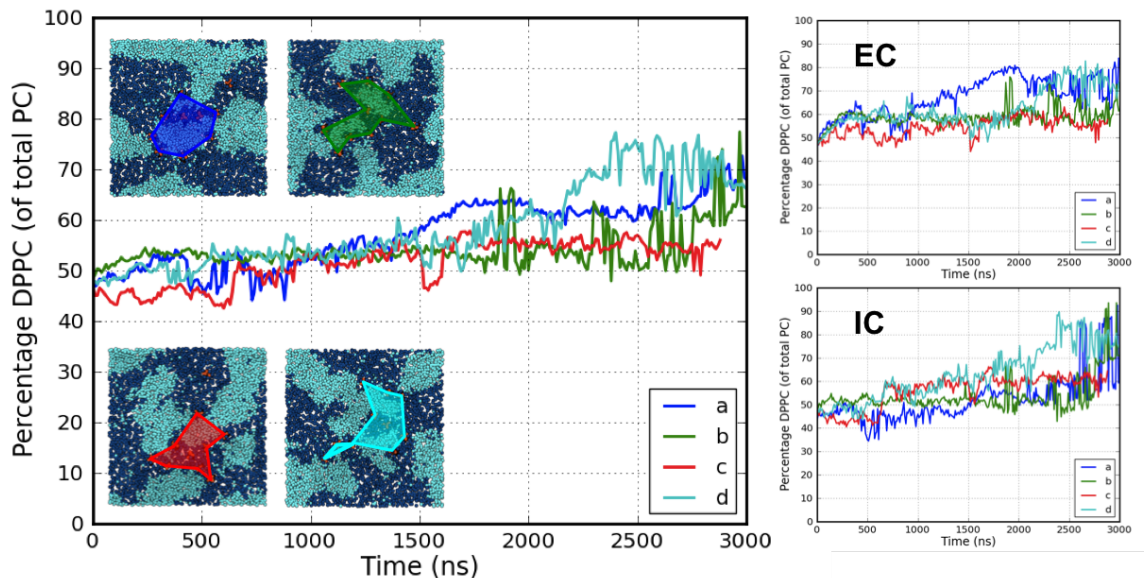


Figure 5.7: Lipid composition within the aggregate interior over time, for each of the four simulations of an HA aggregate in a domain-forming membrane. The area of the aggregate interior was defined at each timestep by drawing two-dimensional triangles in the plane of the membrane, with vertices defined by the geometric centres of the protein TM domains. Every possible combination of three proteins was used to define a triangle, then any triangle with side longer than 14 nm was removed. This distance was chosen as twice the distance at which HA proteins had no effect on lipid diffusion coefficients or tail ordering, according to the analysis with single proteins (Fig. 5.5). The aggregate interior was defined as the combined area of the remaining triangles. The simulation images shown on the graph indicate the areas included in the analysis at the final simulation frames (showing the EC leaflet).

The two graphs on the right show the same analysis for the individual leaflets, with a modification to the algorithm to ignore lipids within 2 nm of a protein. The reason for this modification was that the lipid annuli around the proteins were fairly constant throughout the single-protein and aggregate simulations, i.e. a DUPC annulus in the EC leaflet and a DPPC/cholesterol annulus in the IC leaflet. Removing these lipids from the analysis thus allows observation of only the regions of the aggregate interior which change significantly over time.

Lipid diffusion coefficients are shown in Fig. 5.8, and have been analysed in three different regions: firstly, lipids within 7 nm of any protein and inside the aggregate (defined according to the triangularisation scheme described above); secondly, lipids within 7 nm of any protein,

but outside the aggregate; thirdly, lipids more than 7 nm from any protein and outside the aggregate (i.e. bulk lipids). This analysis shows that diffusion of DUPC within the aggregate is considerably limited in comparison with DUPC outside the aggregate and within 7 nm of a protein, and even more so in comparison with bulk DUPC. In contrast, only a small effect is seen for DPPC. Comparing this behaviour with the equivalent single-protein simulation, it appears that the decrease in DUPC diffusion coefficient in proximity to the protein is amplified by increased local protein concentration, while the smaller effect on DPPC remains relatively insignificant. This behaviour provides a likely explanation for the increase in DPPC concentration within aggregates: a DUPC lipid within a protein aggregate is disfavoured due to limited diffusional capability induced by the high local concentration of slow-diffusing proteins, while DPPC (which ordinarily diffuses more slowly than DUPC) is relatively unaffected. There is thus an entropic penalty associated with the presence of DUPC within the aggregate interior, while only a relatively small penalty for DPPC. Lipid tail ordering was essentially unaffected by position inside or outside the aggregate (analysis not shown), suggesting that the attenuation of lateral (and possibly rotational) diffusion of DUPC is the major driver for accumulation of raft-type lipids within the aggregate interior.

It is clear from the snapshots in Fig. 5.6 that lipids within the aggregate interior are not completely replaced by DPPC. Firstly, the annuli of DUPC in the EC leaflet and DPPC/cholesterol in the IC leaflet are retained. Beyond this region the lipid composition varies widely within different areas of the protein aggregates and between different simulations. Indeed, it is for this reason that it is possible to measure the diffusion coefficients of both DPPC and DUPC within the aggregate interior. However, the analysis in Fig. 5.7 does indicate a gradual increase in the ratio of DPPC within the region defined by a distance of 2–7 nm from the closest protein. The following theoretical considerations aid in the description and explanation of this process.

Two important concepts are that the most stable position of HA within a domain-forming membrane is at the interface between two domains, and that HA TM domains are always separated by a region of lipids, due to the bulky ectodomains. This raises the question: how can two aggregated HA proteins be arranged in a domain-forming membrane in such a way that both conditions are satisfied? Two main types of configuration seem obvious: *i*) the proteins line up along the same domain boundary, or *ii*) a domain of one type (Lo or Ld) occupies the area between the proteins, while domains of the other type occupy the

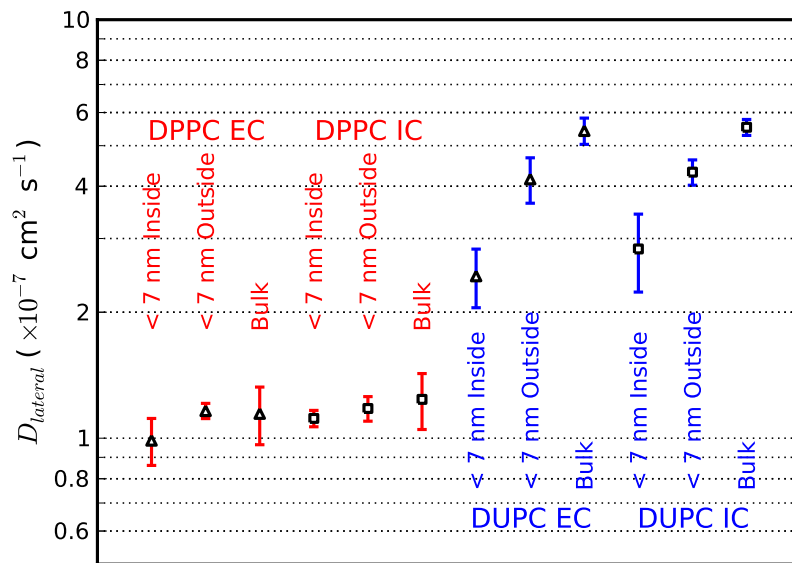


Figure 5.8: Lipid diffusion coefficients, averaged (with error bars) over the four simulations of a HA aggregate in a domain-forming bilayer. Lipids were analysed in three regions: within the aggregate (according to the triangularisation scheme described in the caption of Fig. 5.7) and < 7 nm from the nearest protein; outside the aggregate and < 7 nm from the nearest protein; and outside the aggregate and > 7 nm from the nearest protein (i.e. bulk lipids). Lipids were only included in the analysis if they were within the “expected” domain, as explained in the caption of Fig. 5.5. The details of the diffusion coefficient measurements themselves were the same as described in the caption of Fig. 5.3.

surrounding area (notwithstanding the presence of conserved lipid annuli). Looking at the snapshots in Fig. 5.6, examples of both configuration types can be seen, such as for the central proteins in simulation *b*. A second question related to the type *ii* configuration is whether Lo or Ld domains are preferred between a pair of proteins. This is essentially the same question that is addressed in the analysis in Fig. 5.7, except that it considers proteins in triplets (and the area between them as triangles) rather than pairs. Looking again at the snapshots for the type *ii* configurations, a slight preference for DPPC between proteins does seem apparent. Thus, there is an overall increase in the concentration of DPPC within the aggregates, but without complete replacement of DUPC. This is due to the preference for Lo domains separating proteins in type *ii* configurations being relatively slight, as well as the continued existence of type *i* configurations.

This behaviour may also influence the aggregative properties of HA. For example, a pair or triplet of proteins arranged according to configuration type *ii* and separated by DUPC would likely experience a potential driving the proteins apart, arising from the unstable slow-diffusing state of the DUPC molecules. Notably, the longest nearest-neighbour distances between protein TM domains in the simulations correlate with separation by DUPC (such as the proteins ringed in red in Fig. 5.6 and the rightmost protein in simulation *b*). Several examples were also observed in the trajectories where DUPC lipids would enter a protein aggregate and push apart two proteins. An example observable from the snapshots in Fig. 5.6 is the behaviour of the two proteins at the top-left of the aggregate in simulation *b*, separated by a large region of DUPC, and shown in green and red in the inset plot of protein trajectories. This pair of proteins were originally in contact with each other and separated by a DPPC/cholesterol domain. At ~ 1500 ns in the simulation, the proteins broke contact, but did not move far away from each other (and both remained in contact with other proteins). At ~ 2500 ns, a region of DUPC lipids diffused into the area between the proteins, at which point they quickly moved apart and further DUPC flowed into the area.

Another interesting feature observed in all four simulations was the association of the protein aggregates with an outward curvature of the membrane in the EC direction (example shown in Fig. 5.6a, oblique view). The reasons for this effect are not yet clear, but are of interest considering that transfected HA and NA have been shown to be sufficient for budding of virus-like particles (VLPs) from the plasma membrane [74]. This feature of the simulations will be investigated further, potentially providing new insight into the molecular details of

this process.

5.4 Conclusions

The simulations reported here have provided a novel method for studying the interactions of HA with domain-forming membranes in molecular detail. They show how raft-type lipids (saturated phospholipids and cholesterol) can spontaneously accumulate within a HA aggregate. This result implies that a high local concentration of HA proteins may be sufficient for association with lipid rafts in the plasma membrane, which is an important step in the assembly process of the influenza virus. Interactions with cytoskeletal components, which have also been proposed to play a role in this process [37, 198], may thus not be essential or may play a strengthening role.

The enrichment in raft-type lipids was slight, but was generally rising throughout the duration of the simulations. Complete replacement of unsaturated lipids by raft-type lipids would not be necessary to attain the lipid composition found in released influenza virions, which comprises $\sim 60\%$ saturated lipids (as a percentage of total phospholipids) [193] and $\sim 40\%$ cholesterol [191]. The lipid composition of the simulated aggregates matched this composition fairly closely by the end of the simulations.

Comparing the simulated systems with the *in vivo* plasma membrane of an influenza-infected cell, a number of differences are clear. Firstly, the HA ectodomain is glycosylated at a number of sites, which may alter its propensity for aggregation. For example, the increased hydrophilicity would likely reduce the entropic gain provided by aggregation (via the hydrophobic effect), while the ability for glycans to form specific hydrogen-bonding interactions might provide an enthalpic driving force. It is unlikely that ectodomain contacts alone are responsible for HA aggregation *in vivo*, particularly due to the potential for non-specific aggregation with other proteins. Interestingly however, intracellular domain contacts, as well as palmitoylation, have been implicated in the targetting of the linker for activation of T cells (LAT) to rafts [41]. The simulations did indicate an ability of lipid domains to influence the aggregative behaviour of HA, in that regions of unsaturated lipid appeared to be able to force apart proteins. In combination with the other complex interactions between proteins in lipid domains, this may provide an ability for lipid domains to drive HA aggregation. Cytoskeletal interactions may also be important. It should also be noted that the elastic

network restraints added to the protein (except for the cytoplasmic tails) hold the structure together fairly rigidly. Large-scale conformational changes would not be expected, but it is possible that parts of the protein may be more flexible in reality, for example at the TM domain-ectodomain linker region.

A second important difference between the simulations and the *in vivo* case is related to the nature of the membrane. The HA aggregate simulations have been designed to allow consideration of the case of a single, isolated HA aggregate in a domain-forming bilayer. However, large domains do not form in biological membranes, possibly due to the high protein content which has been estimated at up to $\sim 50\%$ by surface area. The nature of the membrane surrounding the simulated HA aggregates would thus be rather different *in vivo*, but for the purposes of the simulations the major necessity was that lipids could exchange freely between the aggregate interior and exterior. Furthermore, the complexity of biological membranes is much greater, including a wide range of lipid types, variety in lipid tail length and saturation, leaflet asymmetry, cytoskeletal interactions, and many other important factors. Efforts to study more realistic biological membranes are ongoing within the simulation community, but studies of simpler systems are also vital for investigating specific features of complex processes in isolation.

The simulations reported here can be related more directly to model membrane experiments. Studies of the nature of lipid domains, and their association with proteins, have provided useful comparisons with the nature of lipid rafts [2]. However, a number of fundamental differences separate the behaviour of Lo domains in model membranes from that of lipid rafts *in vivo*, most notably that lipid rafts do not coalesce to form large domains. In terms of the interactions of proteins with domains, it has been found that all known raft-associating TM proteins partition into Ld domains [37–41]. Two likely reasons for this unexpected behaviour are the much higher protein concentration found in biological membranes (and which is difficult to attain with current model membrane experiments), and the influence of membrane-associated cytoskeletal components. The simulations have thus provided a way to directly address the former hypothesis, and show that high local protein concentration can be sufficient for association with raft-type lipids. This is supported by experimental studies of live cells which have shown that cross-linking (with antibodies or cholera toxin subunit B) of raft-associating proteins, including HA, can increase raft size [213].

Relatively few simulation studies have been applied to the study of protein interactions with

lipid domains, due to the requirement for simulation of long time and length scales. However, recent advances in CG force fields have allowed the ability to conduct such simulations, and a 2008 study showed that the MARTINI model can reproduce the properties of domain-forming membranes [203]. Two recent papers have built upon this study and investigated the interaction of simple α -helical peptides with a domain-forming membrane. The first showed that the peptides partition into Ld domains, and validated this with experimental work [101]. The second showed that extreme crowding of a membrane with peptides could induce domain formation in membranes which would otherwise be mixed, that the peptides would still partition into the Ld domains, and that bacteriorhodopsin (a 7TM protein found in archaeal membranes, which comprise very different lipid types from eukaryotes) would also partition into Ld domains [205]. These studies of simple α -helical peptides make an interesting contrast with the current study of HA for which the TM domain is larger in cross-section than the peptides and is palmitoylated at multiple sites. Thus the CG model has been shown to be capable of distinguishing between raft-associating and non-raft-associating TM proteins. Simulations of non-raft mutants of HA—those which prevent attachment of palmitoyl chains, or alter the TM domain sequence in the EC leaflet region—will further help to expose the underlying mechanisms which regulate raft affinity of TM proteins.

Chapter 6

Coarse-Grained Molecular Dynamics Simulations of the Entire Influenza Viral Envelope

6.1 Introduction

Lipid rafts have been implicated in a wide range of biological processes, such as signalling, membrane protein sorting and trafficking, endo-/exocytosis and modulation of membrane structural properties [2]. The three proteins of the envelope of influenza virus A (neuraminidase (NA), hemagglutinin (HA) and M2) are thought to associate with lipid rafts in the plasma membrane, facilitating the establishment of the high local protein concentrations necessary for incorporation into the viral envelope. Although the existence and biological importance of lipid rafts is not in question [214], their exact nature has been difficult to define. A key comparative system is the model membrane with a ternary mixture of saturated lipid, unsaturated lipid and cholesterol. These membranes spontaneously separate into Lo domains and Ld domains within a certain temperature range [33]. However, a number of differences are evident between domain-forming model membranes and lipid rafts *in vivo*. Firstly, model membranes can be separated entirely into two large domains, while rafts in biological membranes are thought to be less than 100 nm in size. Secondly, a prime motivation for the investigation of lipid rafts is the importance of their interactions with membrane proteins; surprisingly however, all known raft-associating TM proteins partition into Ld domains in

model membranes [37–41]. A number of unique features of the plasma membrane may account for these disparities, such as increased protein concentration, cytoskeletal interactions, and endo-/exocytosis [2]. The difference in temperature between model membrane experiments (many of which are conducted at room temperature) and physiological conditions may also be a factor.

The influenza virus envelope can be considered as a model of more complex biological membranes. It comprises a lipid vesicle enriched in saturated lipids (particularly sphingomyelin and PC) and cholesterol [193, 194], with three types of inserted TM protein, named above. Accurate atomic-level structures are available for HA and NA via X-ray crystallography, and for M2 via X-ray crystallography and NMR spectroscopy. Electron microscopy [68, 69] and AFM [215] techniques have provided images of influenza virions at near-molecular resolution, capable of distinguishing between HA and NA proteins. Using this structural information as a basis, it has been possible to build and simulate a realistic model of the viral envelope. The study represents one of the first simulations of a complete biological membrane in near-atomistic detail, and is presented in two main contexts. Firstly, the approach allows a direct investigation from the effect of high protein concentration on lipid domain behaviour. The resolution of this problem is important for understanding the basis of raft formation *in vivo*, and has been difficult to study with model membrane experiments, in which high protein concentrations are difficult to attain.

The second context in which the simulations can be considered is that of the viral envelope structure. Although the model envelope differs in a number of ways from real virions, the simulations can provide a useful comparison with experimental studies of influenza virus. For example, the matrix protein is not included in the envelope model—although atomic structures of individual subunits are available, information on the ultrastructure is still sparse. The matrix protein is thought to interact with the cytoplasmic tails of the envelope proteins [216–218]. However, a proportion of virions are known to bud with incomplete or entirely missing matrix layers, and appear to hold the same concentration of envelope proteins [68, 69]. Furthermore, transfection of viral proteins has shown that HA and NA are required for budding of virus-like particles, but matrix protein is not [74]. Again, envelope proteins were included at concentrations similar to those of real virions. It is therefore of interest to use the simulations as a way to study the viral envelope structure in isolation of the effects of matrix protein.

The goal of simulating an entire viral envelope can be compared to an earlier simulation study of the complete tobacco satellite mosaic virus [219]. This model was created in fully atomistic detail, but even this small virus (~ 17 nm diameter) could only be simulated for ~ 50 ns. Although the hardware and software used for simulations have advanced greatly in the five years since that study, the much larger influenza virus (84–170 nm for spherical virions imaged by cryo-ET [68]) represents a significant computational challenge. The ability to study domain formation also requires much longer timescales than would be readily attainable with atomistic MD. The system was therefore modelled with the MARTINI CG force field, which greatly improves computational efficiency, allowing the viral envelope to be simulated for almost two microseconds.

6.2 Methods

6.2.1 Protein Modelling

6.2.1.1 Hemagglutinin

The modelling of the HA structure is described in the Methods section of Chapter 5.

6.2.1.2 Neuraminidase

NA is a homotetrameric TM protein, comprising a headpiece which binds and cleaves sialic acid residues, a long stalk, a TM domain, and a short cytoplasmic tail. A wide range of crystal structures have been solved, but only the headpiece has been resolvable to date. The headpiece structure used in this work was from the A/Tokyo/3/67 (H2N2) strain (PDB code: 1NN2 [220]).

The stalk structure is thought to be mostly unstructured, with inter-subunit disulfide bonds stabilizing the tetrameric arrangement [217]. The exact sequence, which is almost entirely unconserved, has been found to be unimportant for viral function, and mutations of long stretches (with the exception of the disulfide bond-forming cysteine residues) can be accepted without compromising the viability of the protein or virus [221]. However, the length of the stalk is likely to be important for positioning of the headpiece at the correct location for cleavage of HA-bound sialic acids; long deletions attenuate the growth rate and alter

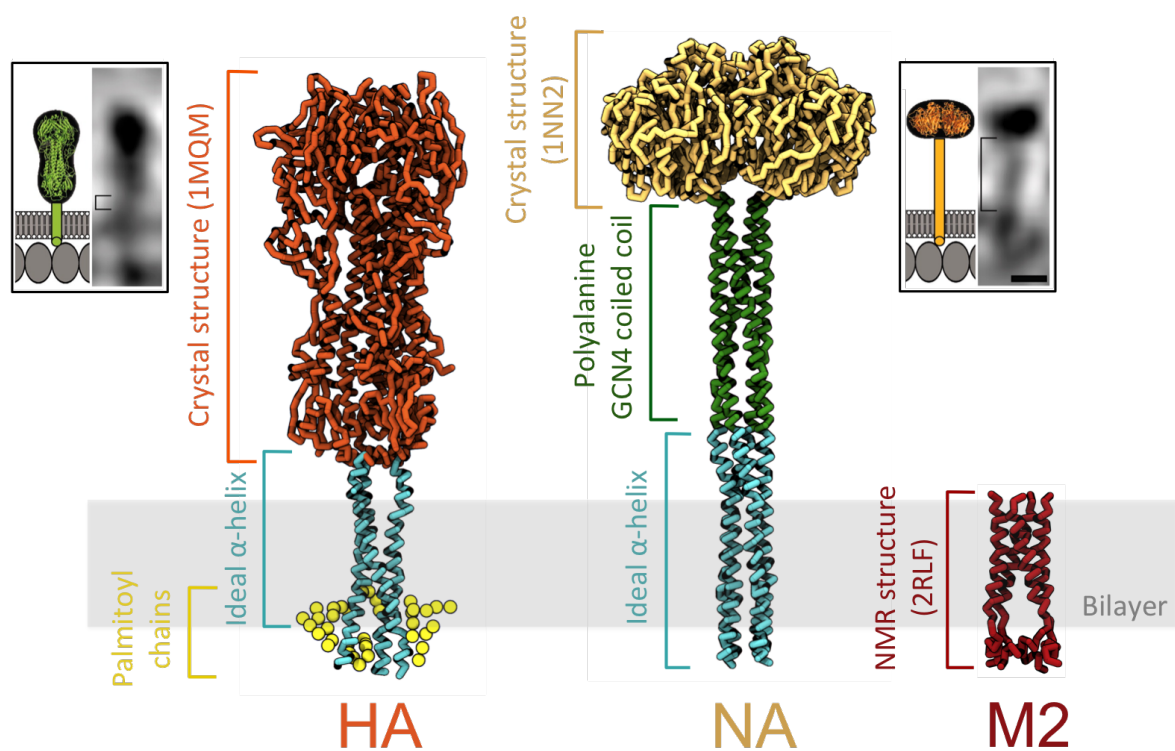


Figure 6.1: Annotated diagrams of the HA, NA and M2 models. The inset panels are adapted from [68], and show EM images of HA and NA, with schematics of the structures. The crystal structures are shown in these schematics, and the grey ovals represent the matrix protein. The length of the NA polyaniline coiled coil was adjusted to match the EM image.

the host specificity of the virus [221–223]. With these features in mind, the stalk was thus modelled as a simple polyalanine α -helix, with the length between TM domain and headpiece (10 nm) matched to cryo-EM images of the protein (which used the same N2 subtype, and are shown in Fig. 6.1). The stalk structure was based on a tetrameric coiled coil motif from the GCN4 leucine zipper protein (PDB code: 1GCL [224]). This structure, which was about 4.5 nm in length, was converted to a CG representation (using the MARTINI tools), and elongated to 8.5 nm by repeating the structure once and aligning four overlapping terminal helix residues (the completed structure is shown in green in Fig. 6.1). Residues were then manually converted to alanines. The remaining 1.5 nm of stalk proximal to the membrane were included in the TM domain helix, as described below.

The headpiece domain structure was converted to a CG representation (shown in cream in Fig. 6.1), and the stalk was attached by bending five overlapping residues into position using the same ENM technique as described for HA in the Methods section of Chapter 5.¹ To model the TM and cytoplasmic domains, a similar approach was taken to the modelling of HA. Thus, the two domains plus 16 residues of the stalk domain were modelled as continuous α -helices (shown in cyan in Fig. 6.1), attached to the membrane-proximal termini of the stalk regions, and then simulated in a bilayer using H-bond type restraints. The details of this process, including the formation of the domain-forming bilayer patch, were the same as described in the Methods section of Chapter 5. The TM domain sequence is IITIGSVSLTIATVCFLMQIAILVTTVTL. Again, polar residues within the TM domain were buried effectively, although supercoiling of the helices did not occur. The cytoplasmic domain, comprising only 6 residues, was left in its α -helical form. A secondary structure sequence analysis using the HHPred server [208] predicts that the three most N-terminal residues are coil, and the remaining three residues are α -helical. The main role of the cytoplasmic tail *in vivo* is thought to involve interactions with the sub-membranous matrix protein [225], which was not included in this study.

As a final point regarding the NA stalk structure, it is worth considering its possible influence on the findings of the study. As explained in the Results and Discussion, NA molecules frequently aggregated with other molecules of HA or NA during the simulation, but this would occur via headpiece contacts; the stalk domains did not contact other proteins at any

¹The ENM software was developed by Alex Tek (IBCP, Paris), who also aided in its application to the protein modelling.

point during the simulation, and thus would be unlikely to have affected the results. The headpiece and stalk domains are glycosylated at a number of sites *in vivo*; as for the HA model, glycans were not included in the CG model. The possible significance on protein aggregation behaviour is discussed later in the Chapter.

6.2.1.3 M2

The M2 proton channel is a 97-residue tetrameric TM protein. The CG structure (shown in red in Fig. 6.1) was converted directly from a NMR structure of the protein in a micellar environment (PDB code: 2RLF) [226]. This structure was solved for a construct comprising residues 18–60, which includes 15 residues of the C-terminal cytoplasmic tail as well as the α -helical TM domain. The protein is normally palmitoylated at Cys50 [227], but this residue was mutated to a serine in the construct used to derive the NMR structure. Palmitoyl chains were thus not included in the CG model. Palmitoylation has been found to have no effect on viral replication *in vitro*, although it may provide some contribution to virulence *in vivo* [228]. The amino acid sequence is SDPLVVAASIIGILHLILWILDRLEFFKSIYRFFEHGLK.

6.2.2 Virion Assembly

The lipid component of the virion was modelled as a 60 nm diameter vesicle with a composition of 0.36 : 0.24 : 0.4 DPPC/DUPC/cholesterol, which matches the ratio of saturated to unsaturated lipids, and the amount of cholesterol found in the virus *in vivo* [191, 193]. As described in Chapter 5, DUPC represents dilinoleylphosphatidylcholine, or 18:2-PC. The lipids were initially arranged in a vesicle of the correct diameter using Packmol, which places copies of lipids randomly and according to given spatial constraints [210]. The numbers of lipids required for each leaflet were estimated by taking the surface area of the two leaflets (assuming a bilayer thickness of 4 nm) and assuming a constant area per lipid (matched to the value for DPPC from a flat bilayer simulation, 0.64 nm²). It should be noted that a large difference in area per lipid between leaflets is not expected for vesicles of this size [167]. A 500 ns equilibration simulation was conducted, using a central shell of restrained particles (32 nm in diameter) to prevent hemifusion, as described in the Methods section of Chapter 4, and a rhombic dodecahedron box to reduce the amount of CG water solvent required. However, this simulation resulted in a shrinking of the vesicle to a diameter of 55 nm. This was

likely due to the presence of cholesterol in the lipid mixture, which takes up a smaller area in the membrane than the value for DPPC used to estimate the number of lipids required. However, the vesicle size appeared to be well equilibrated by the end of the 500 ns simulation. Therefore, the number of lipids in each leaflet were scaled by the necessary increase in surface area (from 55 to 60 nm for the outer leaflet), and a second model was built with Packmol. To allow packing of the increased number of molecules, the initial vesicle diameter was 66 nm. A subsequent 500 ns equilibration simulation resulted in a vesicle of the correct size (60 nm diameter) and with the desired lipid composition in each leaflet.

As described in the Results, the following numbers of envelope protein complexes were inserted into the pre-formed vesicle: 80 HA, 12 NA, 15 M2. The first step was to place the proteins in a spherical arrangement, maximizing the inter-protein distances to avoid contacts. The general problem of arranging a non-trivial number of points evenly on a sphere does not have a known exact geometric solution. A common approach is to model the points as repulsive charges constrained to the sphere surface, and use a simple energy minimization method to drive them as far apart as possible, resulting in an approximation of the most even spacing of points. A version of this algorithm by Simon Tatham (freely available to download from www.chiark.greenend.org.uk/~sgtatham/) was implemented into a Python script which subsequently *i*) scaled the coordinates of the energy minimized points to positions at the outer surface of the pre-formed vesicle, *ii*) placed copies of the CG protein models at these positions by matching their equivalent positions in the flat bilayer equilibration simulations, and *iii*) oriented the proteins so that their long-axes pointed outwards from the vesicle centre. Overlapping lipids were then removed from the vesicle to accommodate the protein TM domains. The final lipid composition was 11487 DPPC lipids, 7463 DUPC lipids and 13231 cholesterol molecules. A cross-section of the resulting model of the viral envelope is shown in Fig. 6.2.

The central shell of restrained particles used to help form the lipid vesicle was retained for the production simulation. A gap of ~ 9 nm was present between the outer surface of the shell and the inner surface of the vesicle. The system was solvated with MARTINI CG water (excepting the shell interior), including 5% “antifreeze” particles to prevent unwanted nucleation of freezing; the latter addition was particularly important due to the rigid central shell (see Methods chapter for details and [167] for further discussion). Na^+ counterions were added randomly in solution to neutralize the system charge.

The viral envelope system was simulated for 1800 ns. The shape of the vesicle did not deviate

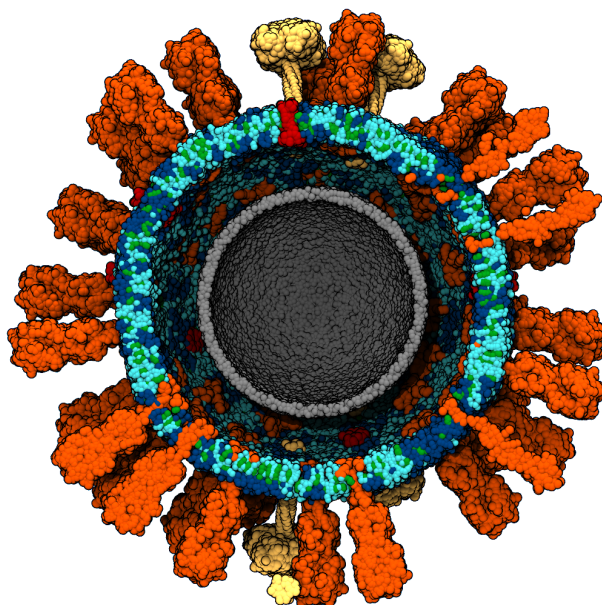


Figure 6.2: Cross-section through the completed model of the viral envelope. HA is shown in orange, NA in cream, M2 in red, DPPC in light blue and DUPC in dark blue. The central shell of restrained particles is shown in grey.

significantly from its initial spherical form. The envelope as a whole moved a small distance in the simulation box, but the closest approach between an envelope particle and a particle from the central restrained shell was ~ 6 nm.

6.2.3 Simulation Details

The GROMACS simulation software was used for all simulations [85–88]. The MARTINI CG parameters were used for all proteins, lipids, ions and solvent. A time step of 10 fs was used. Coulomb interactions were shifted to zero between 0 and 1.2 nm, and Lennard-Jones interactions were shifted to zero between 0.9 and 1.2 nm. The temperatures of the proteins, lipids and solvent were each coupled separately using the Berendsen algorithm [98] with a coupling constant τ_T of 1 ps. A temperature of 310K was used for production simulations, or 323K for equilibration simulations (preventing large-scale domain formation). The system pressure was coupled using the Berendsen algorithm at 1 bar with a coupling constant τ_P of 1.1 ps and a compressibility κ of 1×10^{-6} bar $^{-1}$.

The complete solvated envelope system comprised $\sim 5,000,000$ CG particles. It was simulated on the Jade supercomputer (CINES, Montpellier, France) using 512 cores, achieving a

performance of 57 ns/day. All analysis was performed using Python with the MDAnalysis library [211]. Visualization was performed and figures produced with VMD [133].

Simulation times and time-dependent properties (such as diffusion coefficients) are quoted from the simulation time reported by the GROMACS MD software, and have *not* been adjusted to account for the faster sampling of the CG force field.

6.3 Results

A model of the envelope of influenza virus A was built, comprising a 60 nm diameter lipid vesicle with 107 inserted TM proteins: 84 HA, 12 NA and 15 M2. A number of aspects of the envelope model were based upon the findings of a cryo-electron tomography (cryo-ET) study [68]. The overall diameter, 84 nm, was chosen as the smallest size observed for spherical virions (although spherical virions as small as 73 nm have been observed for a different viral strain [69]). The cryo-ET study allowed visualization of individual glycoprotein spikes and was able to distinguish between HA and NA molecules. The packing density of glycoprotein spikes, and the numbers of HA and NA counted from two complete three-dimensional tomograms of virions, were used to determine the required numbers of HA and NA for the envelope model. The number of M2 molecules was taken as the lower bound of an estimate of 15–25 channels per virion [71]. The vesicle lipid composition, 0.36 : 0.24 : 0.4 DPPC/DUPC/cholesterol, reflected the ratio of saturated to unsaturated lipids (~ 1.5) and the cholesterol content ($\sim 40\%$) of real virions [191, 193].

The viral envelope was simulated for 1800 ns; snapshots are shown in Fig. 6.3 and a video can be downloaded from sbc.bioch.ox.ac.uk/users/parton/videos/0-1800ns.mpg. The TM proteins were able to tilt in the membrane, as shown for HA in the previous Chapter. HA and NA were able to form contacts between ectodomains, forming HA-HA pairs, HA-NA pairs and NA-NA pairs. Many aggregates were formed comprising more than two proteins, either in a linear arrangement or in globular aggregates without directional alignment. M2, which lacks a large ectodomain, was unable to make significant contacts with other proteins. These aggregative properties were analysed quantitatively by calculating the number of contacting proteins per protein (Fig. 6.4). Thus, it can be seen that by the end of the simulation, most of the HA and NA molecules were in contact with at least one other protein, whereas M2 remained mostly disaggregated. Aggregation was thus driven primarily by contacts between

ectodomains, outside of the membrane environment. Contacts were dynamic; protein pairs were able to break apart, but in general quickly reformed contacts with the same or with other proteins.

The protein diffusion coefficients are shown in Table 6.1. Compared to the aggregates simulated in Chapter 5 ($D = 2.9 \times 10^{-8} \text{ cm}^2 \text{ s}^{-1}$), HA proteins diffused slightly faster ($D = 4.5 \times 10^{-8} \text{ cm}^2 \text{ s}^{-1}$) which was likely due to the slightly higher temperature (310 K in this study, versus 295 K in the simulations in Chapter 5). The TM domain of NA has a higher cross-sectional area than HA in the sense of having four inserted TM helices, but does not have attached palmitoyl chains. The diffusion coefficient of NA was slightly faster than HA, indicating that the palmitoyl chains attached to HA had more influence over diffusional properties than the difference in protein cross-sectional area. The diffusion coefficient for M2 was slightly faster again, which can be ascribed to the fact that the M2 proteins were generally not aggregated.

It can be seen from the snapshots in Fig. 6.3 that lipid segregation occurs over the course of the viral envelope simulations. The extent of domain formation is somewhat increased in the IC leaflet, likely due to the increased packing of headgroups [203]. In terms of lipid composition, the domains resemble Lo and Ld domains, being enriched in either DPPC and cholesterol, or DUPC. To investigate the influence of the high concentration of TM proteins on this process, a simulation was conducted with a lipid vesicle without inserted proteins. The vesicle was the same as that used to build the initial viral envelope model, and was simulated for 650 ns. The final snapshot (Fig. 6.5) indicates a similar degree of domain segregation to the final snapshot of the viral envelope simulation. The shorter duration of the simulation thus implies a significant difference in the dynamics of domain formation. Fig. 6.6 shows quantitatively how domain growth is occurring faster for the viral envelope simulation. The consistent lipid annuli which were observed for HA in the aggregate simulations in Chapter 5 (DUPC in the EC leaflet, DPPC/cholesterol in the IC leaflet), were less distinct in the viral envelope simulation, likely due to the increased temperature.

The lipid diffusion coefficients for the viral envelope and vesicle simulations are shown in Table 6.1. They indicate that the high protein concentration in the viral envelope causes a general reduction in lipid diffusion rates, particularly for lipids in proximity to the proteins. DUPC diffusion appeared to be more attenuated by proximity to protein than DPPC, behaviour which was also observed in the HA aggregate simulations in the previous Chapter.

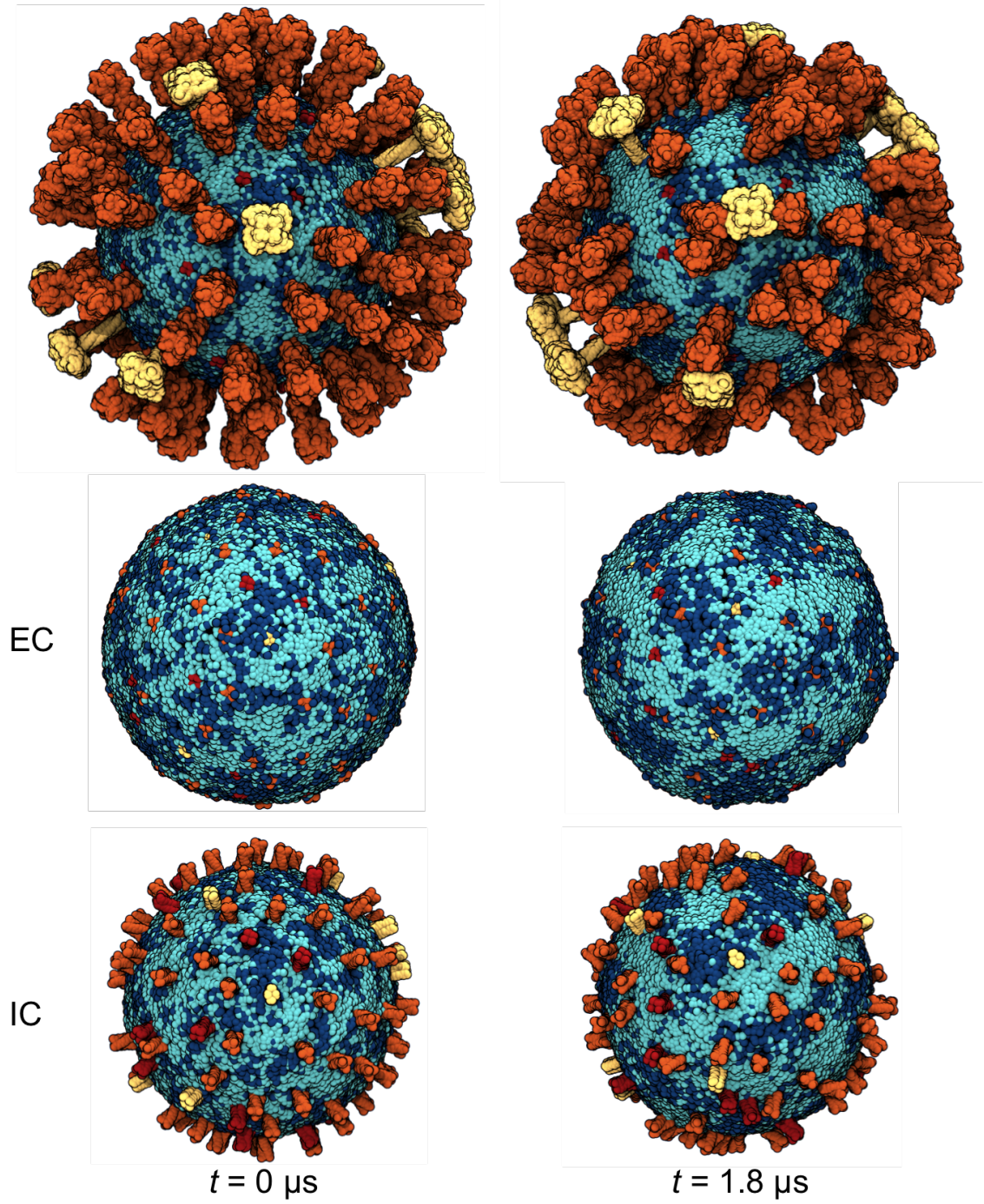


Figure 6.3: Snapshots of the viral envelope simulation at $t = 0$ (left) and $t = 1800$ ns (right). HA is shown in orange, NA in cream, M2 in red, DPPC in light blue and DUPC in dark blue. The top pair of images shown the entire envelope. In the middle pair of images, the protein TM domains have been removed for clarity. The bottom pair of images shows the IC leaflet and protein TM domains.

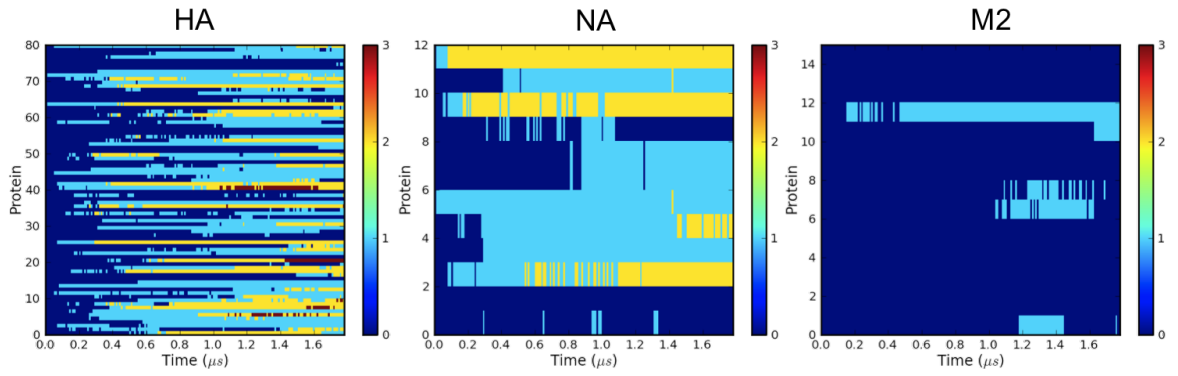


Figure 6.4: Aggregation analysis for proteins in the viral envelope simulation. The plots show the number of neighbours for each protein over time, calculated using a single-linkage clustering algorithm, whereby a pair of proteins are counted as clustered if any of their constituent particles are within 1 nm of each other. Plots are shown for each of the three TM protein types, and in each case neighbours of any protein type were counted.

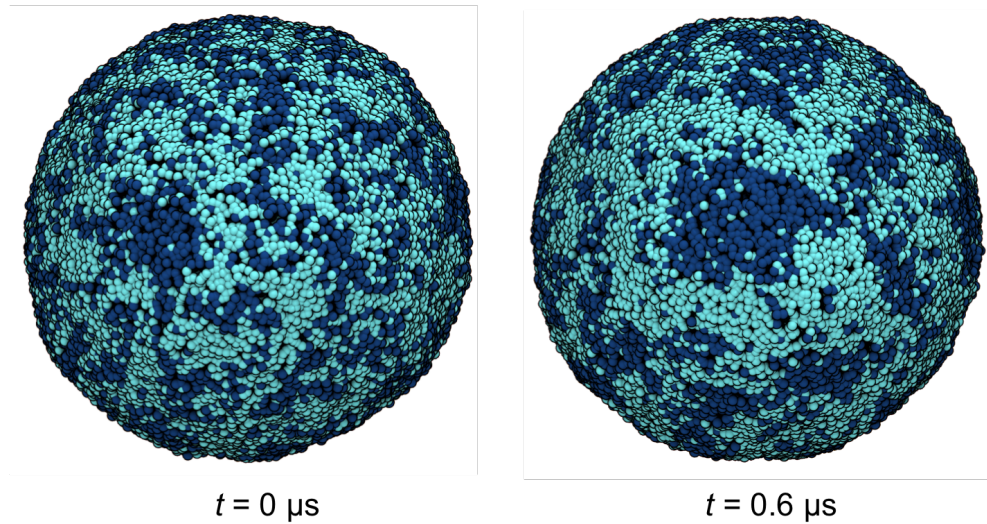


Figure 6.5: Snapshots of the vesicle simulation at $t = 0$ and $t = 614$ ns. Only the EC leaflet is shown. DPPC is shown in light blue, and DUPC in dark blue.

$D (\times 10^{-8} \text{ cm}^2 \text{ s}^{-1})$	Viral envelope			Vesicle
	All lipids	< 5 nm from protein	> 5 nm from protein	
DPPC	20.6	18.2	22.9	25.4
DUPC	45.8	33.6	55.6	56.0
Cholesterol	26.2	-	-	32.1
HA	4.5	-	-	-
NA	5.5	-	-	-
M2	6.7	-	-	-

Table 6.1: Diffusion coefficients for lipids and proteins. Lateral displacements $\mathbf{r}$ were measured using the geometric centres of lipids and protein TM domains, and converted to diffusion coefficients with the following equation: $D = \frac{1}{4\Delta t} \langle |\mathbf{r}(\Delta t) - \mathbf{r}(0)|^2 \rangle$ where the averaging was conducted over all lipids and proteins, and over the time period 900–1800 ns for the viral envelope simulation, and over 307–614 ns for the vesicle simulation. Displacements were measured over a range of timesteps Δt , and in all cases the value which gave the average displacement closest to 2 nm was used to calculate D , to avoid inclusion of fast non-diffusional movements of lipids which take place over shorter length scales. Lateral displacements were measured in the plane of the local membrane surface, defined as the plane perpendicular to the local bilayer normal. The local bilayer normal was estimated using the following algorithm: *i*) find the 10 closest EC leaflet phosphate particles, *ii*) for each EC leaflet phosphate particle, draw vectors to the 10 closest IC leaflet phosphate particles, calculate the average, and normalize, *iii*) average the 10 EC to IC vectors. Membrane curvature was not accounted for in the displacement measurements, but would have been only $\sim 4^\circ$ over the distances (2 nm) measured. For the viral envelope simulation, three diffusion coefficients were calculated: for all lipids, for lipids < 5 nm (in the plane of the local membrane surface) from the nearest protein, and for lipids > 5 nm from the nearest protein.

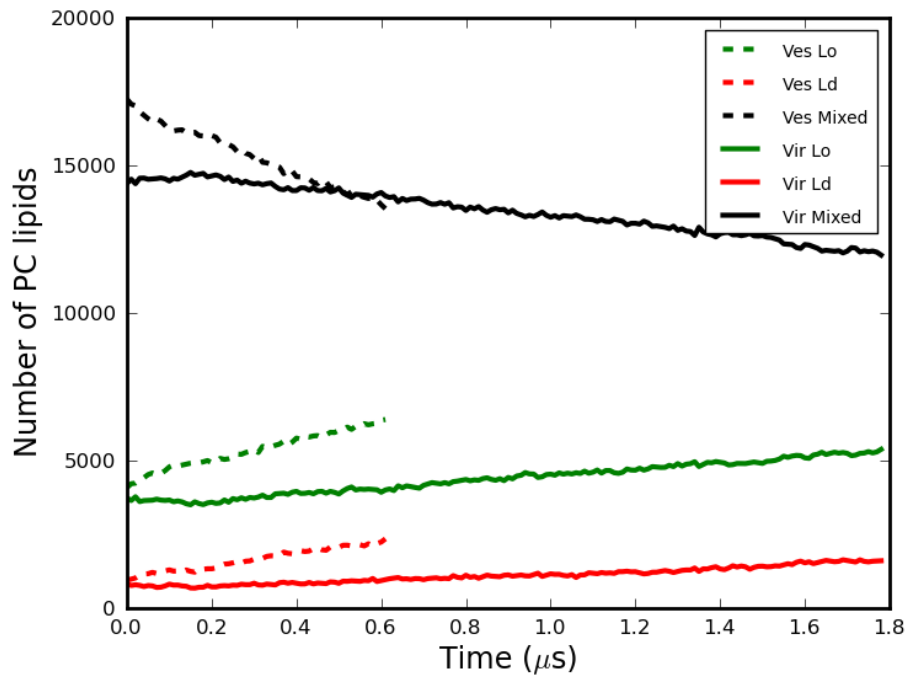


Figure 6.6: Domain growth over time in the viral envelope (Vir; solid lines) and vesicle (Ves; dashed lines) simulations. Phospholipids were categorised as being in Lo or Ld domains, or in a mixed lipid environment, by analysing the nature of the surrounding phospholipids. For each phospholipid, all phospholipids within a 2 nm ring were considered. If 85% of these were DPPC, then the phospholipid was considered to be in a Lo domain. If 85% were DUPC, then the phospholipid was considered to be in a Ld domain. Otherwise the phospholipid was considered to be in a mixed lipid environment.

6.4 Discussion

The key finding of this study is that a high concentration of TM proteins, as found in the influenza viral envelope, attenuates domain formation. This result may seem surprising given that the proteins slow lipid diffusion in the envelope. Colder temperatures increase domain formation in ternary-mixture model membranes by slowing lipid diffusion and increasing tail ordering, providing more favourable conditions for formation of Lo domains [33, 203, 229]. The simulations of HA aggregates in Chapter 5 were conducted at a lower temperature (295K vs. 310K), and exhibited slower lipid diffusion (compare Table 6.1 and Fig. 5.8) and increased domain formation (in terms of both the rate and extent of lipid segregation - compare Figs. 6.3 and 5.6). Instead, it is the specific interactions of the TM proteins with domains that holds the most likely explanation for the attenuation of domain formation in the viral envelope.

Fig. 6.7 shows an alternative view of the simulated viral envelope. It can be observed (particularly in the IC leaflet, which shows increased lipid segregation) that the proteins (including HA, NA and M2) mostly partition at the interfaces between domains. This behaviour was confirmed for HA in Chapter 5. Thus it appears that the proteins attenuate domain formation in the viral envelope by reducing the line tension of domain boundaries, increasing lipid miscibility. A preference for Lo domain boundaries has been shown for the raft-associating acylated protein N-Ras [212]. It has also been suggested from experiments on live cells that M2 associates at the periphery of rafts and alters line tension, and that this may facilitate scission of budding virions [230].

This finding indicates that high protein concentration may be the reason why domains do not form on large scales in biological membranes *in vivo*. This also helps to explain the difference in behaviour in model membranes, where lipids can completely phase separate into Lo and Ld domains, but high protein concentrations are difficult to attain. The protein:lipid ratio in the viral envelope simulation was $\sim 1 : 290$, which is within the likely bounds of protein content in the plasma membrane. For example, the protein:lipid mass ratio can be as high as 50%, which can be roughly estimated as a number ratio of $1 : 100$. Other possible contributions to the differences in domain formation behaviour are the difference in temperature between model membrane experiments (commonly conducted at room temperature) and *in vivo* membranes, cytoskeletal interactions, and endo- and exocytosis [2]. However, some recent experimental studies have suggested the possibility of an influence of protein concen-

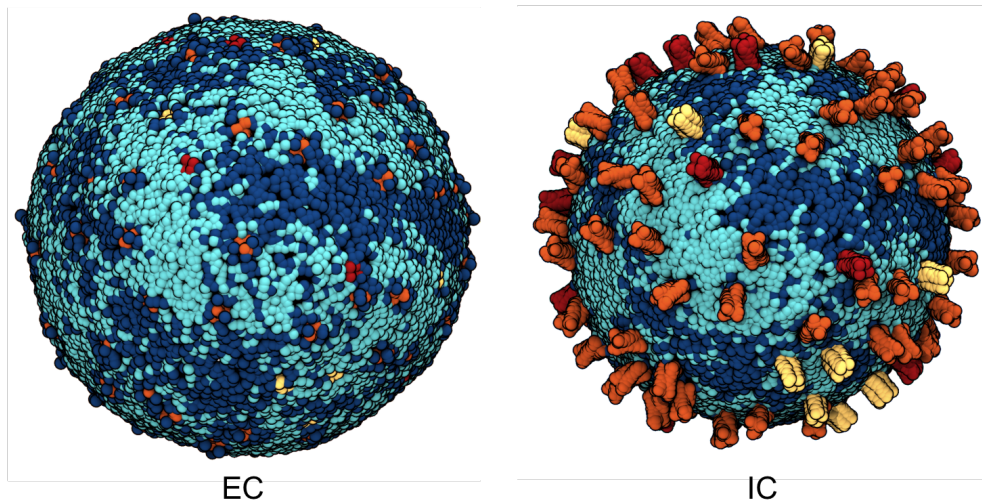


Figure 6.7: Alternative view of the viral envelope after 1.8 μs simulation, showing protein association with domain boundaries.

tration on raft formation. A study of giant plasma membrane vesicles (GPMVs)—which are isolated from the effects of cytoskeletal interactions or endo-/exocytosis, and undergo large-scale domain formation upon cooling [231]—showed that protein clustering mediated by cholera toxin subunit B could induce formation of micrometer-scale domains at 37° [232]. This supports the behaviour observed in the viral envelope simulation, in which domains formed, but not on as large a scale as the vesicle without proteins. The simulations add to the picture by indicating that this behaviour may be caused by partitioning of proteins at the boundaries of small domains, reducing line tension and the tendency for domain coalescence. The authors of the protein clustering experiment also noted that the domains observed were unlikely to be equivalent in nature to Lo domains in model membranes. This is reflected in the simulations by the clear effect of the proteins on the diffusional behaviour of proximal lipids (Table 6.1). Another study highlighted the influence of membrane protein content on lipid packing in domain-forming vesicles derived from the lipids of the plasma membrane [233].

This study can also be contrasted with a recent CG MD study which showed that crowding of a membrane with the simple α -helical peptide, WALP, could induce large-scale domain formation, with the peptides partitioning into the Ld domains. In terms of the application of CG MD simulations to the study of protein association with lipids rafts, the importance of protein type on regulation of domain formation is thus confirmed. This work shows that realistic models of proteins, with far more complex TM domains than simple peptides, can

impart a significantly different effect on domain behaviour, and better replicate the effects of TM proteins *in vivo*. It should be noted that TM peptide simulations are also useful for comparison with model membrane experiments, which are frequently conducted with TM peptides.

Another interesting observation from the simulations was the difference in aggregative properties between the three types of protein. HA and NA were able to aggregate via dynamic ectodomain contacts, whereas the much smaller M2 channel remained mostly unaggregated. This finding hints at a possible method for protein sorting in the membrane, for which a small amount of experimental evidence is available. For example, cytoplasmic domain contacts have been suggested to influence raft-targetting of the linker for activation of T cells (LAT) [41].

The relevance of the simulations to the structure of the influenza viral envelope can be considered. They imply that domains do form in the membrane, but are likely to be small and irregular. The lipid diffusional properties and tail ordering are also unlikely to match those analysed in model membrane studies of ternary lipid mixtures, as the proteins induce a strong reduction in diffusion rates on proximal lipids, especially DUPC. These findings can be usefully compared with those of an experimental study of intact virions, which used protein magic angle spinning NMR spectroscopy to show that domains did not form on large scales at physiological temperatures [234]. The study also indicated that most lipids had Ld-like characteristics, and the authors speculated that this was due to the proteins breaking up packing in nascent Lo domains. The diffusion rates of DPPC and cholesterol observed in the simulation were indeed generally higher than expected for Lo domains ($D \simeq 10 \times 10^{-8} \text{ cm}^2 \text{ s}^{-1}$ for both molecules according to CG simulations [203]), likely due to the small degree of domain formation. The simulations indicate that this effect is caused by a reduction in line tension via partitioning of TM proteins at domain boundaries. Disruption of lipid packing in Lo domains may also play a part, and analysis of the influence of HA on the properties of a small domain-forming bilayer (Fig. 5.5 in Chapter 5) showed that the EC half of the TM domain increased tail disorder in proximal DPPC lipids.

It is important to consider how the model of the viral envelope differs from that of real virions. Firstly, the envelope has been simulated in isolation from the underlying matrix protein or nucleoprotein. The cytoplasmic tails of all three envelope proteins are thought to associate with the matrix protein [216–218], which would likely affect their dynamic properties.

However, electron microscopy (cryo-ET and Zernike phase contrast cryo-electron microscopy) images show that a proportion of virions bud with substantial gaps in the matrix layer (20%) or without matrix protein at all (14%), without major changes in the appearance of the envelope [68, 69]. Furthermore, it has been shown that transfected HA and NA are sufficient for budding of virus-like particles, and are incorporated at full concentration [74]. It is difficult to definitively characterize the orientation of the glycoproteins (HA and NA) in the EM images, but they appear to be generally perpendicular to the local membrane surface, lacking the large degree of tilting seen in the simulations. They also generally appear to be unaggregated, although some aggregation does seem apparent in a three-dimensional model of a complete virus shown in the cryo-ET article [68]. It is tempting to invoke the presence of matrix protein as the likely explanation for the disparity in these features, between *in silico* and *in vivo* studies, but the similar appearance of virions lacking a matrix layer makes this hypothesis unlikely. The low temperatures required for the EM images may potentially also influence the dynamic properties of the envelope proteins.

A possible drawback to the simulations which may influence the aggregative properties of the proteins is the lack of glycosylations on HA and NA. As discussed for HA in the previous Chapter, the ectodomains of these proteins are glycosylated at a number of sites. These glycosylations were not included, as the MARTINI CG model has not yet been thoroughly tested for the simulation of glycoproteins. One possible effect on the simulation would be a reduction in the degree of aggregation, due to the high hydrophilicity of sugars. It is thus not possible to state that the level of aggregation observed in the viral envelope simulation exactly reflects the state of the proteins *in vivo*. However, the overall concentration of proteins has been matched to that of real virions, and it is not clear to what extent aggregation influences the domain-forming properties of the envelope lipids. To investigate this possible effect fully, it would be useful to conduct a simulation with the proteins loosely tethered to the central shell of restrained particles via harmonic distance restraints, preventing aggregation.

It was noted in the previous chapter that the elastic network restraints added to the HA model may restrict flexibility to some extent. This is also the case for the NA and M2 structures. The model of M2 used in this study is primarily comprised of a TM domain, which is unlikely to be highly flexible. For NA, however, the stalk domain is in reality mostly unstructured and thus likely to be somewhat flexible. With no high-resolution structural data available for this domain, or information on the degree of flexibility, it was not possible to include this

feature in the model. It is also difficult to estimate the influence of this factor on the results. However, it is worth noting that in EM images [68, 69], the stalk does generally appear to be straight.

Overall, the main finding of the study, that a high concentration of proteins can attenuate domain formation, is suggested as a general principle for biological membranes. It provides a possible explanation for why rafts do not form on large length scales *in vivo*, based on a reduction in the line tension of domain boundaries caused by specific partitioning of the envelope proteins, as well as disruption of lipid ordering in Lo domains.

Chapter 7

Conclusions

Far more than simple boundaries, membranes are a vital component of all cellular life and are responsible for a diverse range of biological processes. They are composed primarily of lipids and proteins, and much membrane activity depends upon the functional interdependence of these two components. This thesis has reported a range of simulations of complex biological membranes, i.e. those with multiple numbers and types of lipid and protein. The biological processes of interest have required simulations over extended time and scales. Coarse-grained molecular dynamics (CG MD) simulations have thus been utilized as the primary technique, due to their improved computational efficiency compared to the use of atomistic force fields. The individual studies are summarized over the following paragraphs, in order of their completion. There has been a general trend of increasing complexity in the systems studied, starting with simple peptides and homogeneous lipid bilayers, and moving towards realistic biological membranes displaying a complex diversity of lipids and proteins.

Chapter 3 reported an investigation of the mechanism of action of the antimicrobial peptide maculatin 1.1. This 21-residue α -helical peptide is secreted by a species of tree frog and kills bacteria by permeabilizing the plasma membrane. Experimental studies have pointed to a number of possible mechanisms, with the archetypal models proposing either the formation of transmembrane pores, or lysis initiated via adsorption of large quantities of peptides to the membrane surface. The simulations showed that maculatin cooperatively inserts into membranes and forms disordered aggregates which span the entire membrane thickness. By back-mapping the initial CG systems to an atomistic resolution (a simple example of a multiscale methodology), it was found that these disordered aggregates were permeable to water.

CG simulations of large concentrations of peptides with lipid vesicles resulted in substantial membrane deformation, which implies a lytic mechanism may be possible under different conditions or at longer timescales. The simulations thus imply that both membrane-forming and lytic mechanisms are available to maculatin 1.1, and that the predominance of either depends on conditions such as peptide concentration and membrane composition. This explanation fits with the seemingly conflicting evidence from experimental studies conducted under different conditions.

In the study described in Chapter 4, the cooperative interactions between transmembrane (TM) proteins and their lipid environment were investigated. Two model TM proteins were considered: a generalized α -helical bundle and a generalized β -barrel. These were simulated in a range of membrane environments, varying in lipid length and thus the degree of hydrophobic mismatch, as well as varying the degree of membrane curvature (via simulations of both planar bilayers and vesicles). The simulations indicated how protein aggregation can be modulated by a number of membrane features, such as hydrophobic mismatch, membrane curvature and protein type. Interestingly, the two types of protein were observed to aggregate in different ways, with the α -helical bundles forming isotropic clusters, and the β -barrels forming linear strings. Simulations of a wider range of protein types would help to further investigate this phenomenon. As well as its implications for membrane protein function, the nature of protein aggregation in membranes is of particular importance to the structural biology field currently; electron diffraction experiments hold the promise of obtaining structures of membrane proteins in their native environment, but are still limited by the great difficulty of producing two-dimensional crystals.

The association of influenza hemagglutinin (HA) with putative lipid rafts is thought to be vital for its inclusion in the viral envelope. However, the basis of raft composition and formation is still poorly understood, and experimental challenges have limited the additional insight that model membrane experiments can provide. A set of simulations were therefore conducted to investigate the behaviour of HA in domain-forming membranes (Chapter 5). A particular advantage compared to experimental studies was the ability to study the effects of high local protein concentration. From the simulations of HA aggregates, it was found that raft-type lipids (i.e. saturated lipids and cholesterol) spontaneously accumulate within the aggregate interior. This appears to be driven by an entropic differential, arising from the attenuation of diffusion of unsaturated lipids. The findings indicate that high local protein

concentration may be sufficient for association of HA with lipid rafts *in vivo*.

Building upon the techniques used in the previous simulations of biological membranes, a realistic CG model of the influenza viral was built (Chapter 6). The dimensions (84 nm diameter) and quantities of envelope proteins (HA, neuraminidase (NA) and M2) were based on cryo-electron tomography images [68]. The simulation is one of the largest such studies to date of a biological membrane at near-atomic resolution, and shows how the high concentration of proteins found in the viral envelope can attenuate formation of lipid domains by partitioning at the domain interfaces and thus reducing line tension. This provides a possible explanation for the limited size of lipid rafts *in vivo*, i.e. that the high concentration of proteins generally found in biological membranes prevent domain formation.

The results of the influenza studies suggest a number of possible additional simulations, which may provide further insight. Including glycosylation sites on the HA and NA proteins would be of interest, considering the influence of the aggregative properties of the proteins on membrane organization and dynamics. In the longer term, an obvious extension to the viral envelope simulation would be to include a representation of the matrix layer and/or ribonucleoprotein, although this may be dependent upon improved structural data from microscopy experiments. In the shorter term it may be possible to investigate the possible effects of anchoring of the envelope proteins to the matrix layer, by applying loose distance restraints to the central shell of restrained particles.

Reflecting on the work in the context of biological research in general, it can be observed that the simulations have been able to probe biological processes in ways that have not been achievable with current laboratory experiments. For example, the simulations of HA aggregates in domain-forming membranes (Chapter 5) were aimed at resolving a question initially raised by the limitations of model membrane experiments: how does HA associate with rafts *in vivo*, and why does it not associate with Lo domains in model membranes? The viral envelope simulation in Chapter 6 also provides a novel type of insight into an extremely important and topical system, starting with information from published cryo-electron tomography images and protein structures, and adding the ability to directly observe membrane dynamics on the molecular scale. A parallel can be drawn with simulations of membrane transporter proteins, which have been of great utility in adding dynamic detail to structures solved by diffraction or NMR [235]. It is also clear that the simulations have been highly dependent upon the availability of high-quality experimental (especially structural)

data. Furthermore, the testing of simulation findings with experiments will likely always be vital, and the experimental and simulation fields will hopefully continue to become more and more intertwined.

The use of the simulations in Chapter 5 to investigate a specific question arising from experimental work also marks the increasing maturity of the (still relatively new) simulation field; it is now able to usefully operate such “thought experiments” (arguably simulation in its purest form) in the context of biological membranes, although development and validation of techniques and force fields will continue to be vital work. The large size of the viral envelope studied in Chapter 6 also reflects the increasing ability of simulation techniques (with help from the continuing improvements in software and hardware) to study biological systems of realistic dimensions.

Finally, it was noted earlier in this Chapter that the simulations have followed a trend of increasing complexity over time. This reflects a general trend within the simulation community toward increasingly realistic simulations of biological membranes. The true complexity of biological membranes is of course still far away from what has been simulated, with relatively few considerations to date for lipid diversity on the biological level, glycoproteins, cytoskeletal structures or many other important aspects of membrane biology. However, the inclusion of this full complexity must surely be the ultimate goal for those interested in simulations of biological membranes. Very early endeavours along these lines have been undertaken in simulations of the bacterial cytosol [236]. It should also be considered that simulations of simple or generalized systems (such as those in Chapter 4) will continue to be important for the investigation of individual processes in isolation from full biological complexity, and that the development and validation of simulation techniques and force fields will continue to be a vital endeavour for the evolution of the simulation field.

Appendix A

Influence of the integrator time step on simulation properties

It has recently been suggested [168] that CG-MD simulations should not be run with time steps of more than 10 fs, due to the increase in numerical error introduced to the linear approximation of force integration, leading to some discussion by e.g. the MARTINI forcefield developers [169]. It is therefore important to assess the possible influence of the choice of time step on CG-MD simulations of complex systems.

As a test system, we took a smaller version of the G α CG4/CG5 bilayer simulation, comprising 4 proteins and ~ 1800 lipids. This was simulated using a time step of 5, 10, 20 or 40 fs (each simulation for 5 μ s). Other simulation parameters and the system set up were as reported for the main simulations in the Methods section. We analysed a number of macroscopic properties of these simulations (Table A.1 — total energy and temperature). We found small differences between each simulation, but these did not appear to be statistically significant. Notably, only a small decrease in temperature (0.5K) was observed for the 40 fs time step simulation. A much larger decrease in temperature was shown in the study which initially argued against the use of high time steps [168], with the implication that the thermostat was unable to cancel out the numerical errors. However, our results are in-keeping with simulations conducted as a response to that paper, and further discussion on possible reasons for the disparity can be found therein [169]. Overall, the analysis of macroscopic properties suggests that possible numerical errors introduced by the 40 fs time step did not have a major impact on our simulations.

	5 fs	10 fs	20 fs	40 fs
Total energy (kJ mol ⁻¹)	-830771 ± 270	-830985 ± 274	-831470 ± 277	-834694 ± 299
Temperature (K)	323.0 ± 0.8	323.0 ± 0.8	323.0 ± 0.8	322.5 ± 0.8
Area per lipid (Å ²)	66.16 ± 0.04	65.99 ± 0.04	65.90 ± 0.04	65.52 ± 0.04

Table A.1: Macroscopic quantities in the four-protein G α CG4/CG5 bilayer test simulations using time steps of 5, 10, 20 or 40 fs; data averaged from 400 to 5000 ns.

Next we checked whether these small numerical errors might have had an effect on any structural properties of the simulated systems. We had chosen the CG4/CG5 system as a test case because the non-aggregative behaviour of the protein depends on a subtle property of the membrane (i.e. it requires hydrophobic matching within a fairly tight range), whereas aggregative behaviour is essentially the norm under other conditions (i.e. hydrophobic mismatch). Within the range of statistical variation, the non-aggregative behaviour was conserved for the smaller 4-protein system and also over the different time steps used (Fig. A.1 and Table S2). The diffusion coefficients of protein and lipids were also essentially unchanged (Table S2). The area per lipid showed a very slight decrease with increasing time step, which may be explained by the slight decrease in temperature (Table S1). The density profiles showed only a small amount of deviation which was not correlated with the time step (Fig. S4). Most convincingly, the distributions of angles between the CG lipid bonds and the bilayer normal were well converged in the simulations and were the same for the different time steps (Figs. S5).

Overall, it appears that whilst numerical errors introduced by a time step of 40 fs may cause very small deviations in macroscopic system properties, they have essentially no effect on structural properties. We are therefore confident that the results we have presented in this paper are robust to the selected time step.

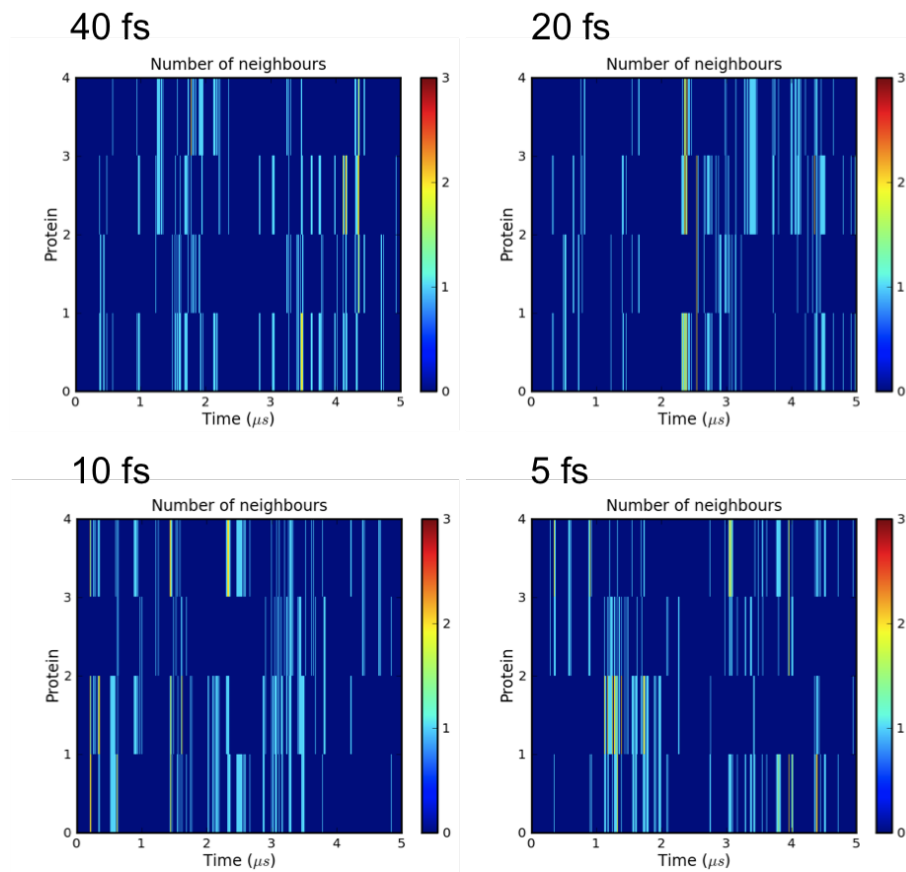


Figure A.1: Aggregation analysis for the four-protein G α CG4/CG5 bilayer test simulations, repeated with three different time steps. Each row represents the number of neighbouring proteins (defined using a cutoff of 2.4 nm between the COM of two proteins) for each of the four proteins.

Appendix B

Derivation of Expression for Diffusional Permeability Coefficient

The permeability coefficient P of a membrane relates to the situation where a solute (the permeant) is present on either side of a membrane, but in different concentrations. In this case, P is equal to the steady-state flux j of the permeant, divided by the membrane area A and the difference in permeant concentration across the membrane ΔC :

$$P = j/A\Delta C \tag{B.1}$$

Without a concentration difference, no net flux would be expected, but it is still of interest to measure diffusional permeability. For example, diffusional water permeability can be calculated experimentally by setting up two reservoirs either side of a membrane — one with a high tracer concentration C_{tr} (the tracer would be made distinct via either spin-labeling or isotopic replacement), and the other with zero concentration. The tracer flux j_{tr} can then be measured and related to the diffusional permeability coefficient P_{d} as follows:

$$j_{\text{tr}} = P_{\text{d}}AC_{\text{tr}} \tag{B.2}$$

The ratio of tracers to water molecules in the first reservoir would be $C_{\text{tr}}/C_{\text{W}}$. C_{W} is the concentration of water, and is equal to $1/V_{\text{W}}$, where V_{W} is the molar volume of water ($18 \text{ cm}^3 \text{ mol}^{-1}$). The tracer molecules should permeate the membrane in the same way as the

water molecules, and hence the tracer flux (in mol s⁻¹) can be related to the total number of unidirectional water crossings per unit time (q_0) as follows (N_A is Avogadro's number):

$$j_{\text{tr}} = \frac{C_{\text{tr}}}{C_{\text{W}}} \frac{q_0}{N_A} \quad (\text{B.3})$$

Substituting V_{W} for C_{W} , and using the expression $V_{\text{W}}/N_A = v_{\text{W}}$, where v_{W} is the volume of a single water molecule, gives:

$$j_{\text{tr}} = C_{\text{tr}} v_{\text{W}} q_0 \quad (\text{B.4})$$

The diffusional permeability coefficient can therefore be related to the rate of unidirectional water crossings by combining Eqs. B.2 and B.4:

$$P_{\text{d}} = q_0 v_{\text{W}} / A \quad (\text{B.5})$$

Appendix C

Lipid vesicle leaflet equilibration

The relatively fast kinetics of formation of lipid membranes via self-assembly, coupled with the far slower kinetics of lipid flip-flop, could potentially result in poorly equilibrated membranes. A method for equilibrating vesicle membranes was recently introduced; it involves the formation of toroidal pores in the membrane via cylindrical mean field potentials [167]. The lipids are able to flip-flop relatively quickly at the boundary of the toroidal pores, thus allowing the vesicle leaflets to be equilibrated (although it could be postulated that the introduction of cylindrical potentials into the vesicle membrane might themselves affect the equilibrated composition). Unfortunately this method is less suitable for the equilibration of the vesicles used in Chapter 4, as it would likely have disrupted the 1:1 mixture of short and long chain lipids desired in each leaflet. The vesicles were therefore formed via self-assembly around a central shell of restrained particles, and without further equilibration. The 1:1 mixture of short and long chain lipids in each leaflet was fairly well conserved in the vesicles simulated (data not shown).

To investigate whether the ratio of lipids between outer and inner leaflets was correctly equilibrated, an ensemble of ten self-assembly simulations of small (15 nm diameter), POPC vesicles was conducted. The same methods were used in these simulations as for the self-assembly of the larger vesicles used in Chapter 4, including the use of a central shell of restrained particles. It was observed that the statistical variation in the leaflet composition ratio produced by this method is relatively small (Table C.1). The mean of the fraction of inner leaflet lipids was 0.275 with a standard deviation of 0.004, which represents approximately 3–4 lipids. It would be reasonable to expect that this small number of lipids would only have a small effect

on the membrane properties. Further evidence can be derived from a second recent paper using the cylindrical potential method [237]. The study described equilibration simulations of self-assembled 16 nm vesicles, which resulted in the exchange of approximately 10–20 lipids between leaflets, i.e. $< 1\%$ of all lipids. Furthermore, after reaching apparent equilibrium, lipids continued to exchange between leaflets in numbers of approximately 5–10 and with a period of tens of nanoseconds. This implies that there is little change in energy caused by transfer of lipids between leaflets in these quantities, making it unlikely that membrane properties would be greatly affected.

Simulation	1	2	3	4	5	6	7	8	9
N lipids	873	877	872	870	874	877	875	877	875
N inner lipids	242	245	242	236	245	237	235	239	241
Fraction of inner lipids	0.277	0.279	0.278	0.271	0.280	0.270	0.269	0.273	0.275

Table C.1: Leaflet compositions for the ensemble of small (16 nm diameter) vesicle simulations. Variation in the total number of lipids occurred because some lipids do not join the membrane during the self-assembly simulations, which were initiated with 877 lipids.

Bibliography

- [1] Vance, D., and J. Vance, 2008. Biochemistry of Lipids, Lipoproteins and Membranes, 5th edition. Elsevier.
- [2] Lingwood, D., and K. Simons, 2010. Lipid rafts as a membrane-organizing principle. *Science* 327:46–50.
- [3] van Meer, G., D. R. Voelker, and G. W. Feigenson, 2008. Membrane lipids: where they are and how they behave. *Nat. Rev. Mol. Cell Biol.* 9:112–124.
- [4] Meyer zu Heringdorf, D., and K. H. Jakobs, 2007. Lysophospholipid receptors: Signalling, pharmacology and regulation by lysophospholipid metabolism. *BBA - Biomembranes* 1768:923–940.
- [5] Fernandis, A. Z., and M. R. Wenk, 2007. Membrane lipids as signaling molecules. *Curr. Opin. Lipidol.* 18:121–128.
- [6] Guidotti, G., 1972. Membrane Proteins. *Annu. Rev. Biochem.* 41:731–752.
- [7] Russ, A. P., and S. Lampel, 2005. The druggable genome: an update. *Drug Discovery Today* 10:1607–1610.
- [8] Hopkins, A. L., and C. R. Groom, 2002. The druggable genome. *Nature Reviews. Drug Discovery* 1:727–730.
- [9] Karp, G., 2009. Cell and Molecular Biology: Concepts and Experiments. Wiley, 5th edition.
- [10] Bill, R. M., P. J. F. Henderson, S. Iwata, E. R. S. Kunji, H. Michel, R. Neutze, S. Newstead, B. Poolman, C. G. Tate, and H. Vogel, 2011. Overcoming barriers to membrane protein structure determination. *Nat Biotech* 29:335–340.

- [11] Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, 2000. The Protein Data Bank. *Nucl. Acids Res.* 28:235–242.
- [12] Sanders, C. R., and K. F. Mittendorf, 2011. Tolerance to changes in membrane lipid composition as a selected trait of membrane proteins. *Biochemistry* 50:7858–7867.
- [13] Hite, R. K., Z. Li, and T. Walz, 2010. Principles of membrane protein interactions with annular lipids deduced from aquaporin-0 2D crystals. *EMBO J.* 29:1652–1658.
- [14] Gonen, T., Y. Cheng, P. Sliz, Y. Hiroaki, Y. Fujiyoshi, S. C. Harrison, and T. Walz, 2005. Lipid-protein interactions in double-layered two-dimensional AQP0 crystals. *Nature* 438:633–638.
- [15] Jensen, M. O., and O. G. Mouritsen, 2004. Lipids do influence protein function-the hydrophobic matching hypothesis revisited. *Biochim. Biophys. Acta* 1666:205–226.
- [16] Lee, A. G., 2005. How lipids and proteins interact in a membrane: a molecular approach. *Mol. Biosyst.* 1:203.
- [17] Lee, A. G., 2004. How lipids affect the activities of integral membrane proteins. *Biochim. Biophys. Acta* 1666:62–87.
- [18] White, S. H., 2004. The progress of membrane protein structure determination. *Protein Science: A Publication of the Protein Society* 13:1948–1949.
- [19] Cross, T. A., M. Sharma, M. Yi, and H. Zhou, 2011. Influence of solubilizing environments on membrane protein structures. *Trends Biochem. Sci.* 36:117–125.
- [20] Tamm, L. K., 2005. Protein-Lipid Interactions: From Membrane Domains to Cellular Networks. Wiley-VCH, Weinheim, Germany.
- [21] Hanson, M. A., V. Cherezov, C. B. Roth, M. T. Griffith, V. Jaakola, E. Y. T. Chien, J. Velasquez, P. Kuhn, and R. C. Stevens, 2008. A specific cholesterol binding site is established by the 2.8 Å structure of the human 2-adrenergic receptor in an alternate crystal form. *Structure (London, England : 1993)* 16:897–905.
- [22] Contreras, F., A. M. Ernst, P. Haberkant, P. Bjorkholm, E. Lindahl, B. Gonen, C. Tischer, A. Elofsson, G. von Heijne, C. Thiele, R. Pepperkok, F. Wieland, and B. Brugger,

2012. Molecular recognition of a single sphingolipid species by a protein/'s transmembrane domain. *Nature* advance online publication.
- [23] Singer, S. J., and G. L. Nicolson, 1972. The fluid mosaic model of the structure of cell membranes. *Science* 175:720–731.
- [24] Wunderlich, F., W. Kreutz, P. Mahler, A. Ronai, and G. Heppeler, 1978. Thermotropic fluid goes to ordered "discontinuous" phase separation in microsomal lipids of Tetrahymena. An X-ray diffraction study. *Biochemistry* 17:2005–2010.
- [25] Wunderlich, F., A. Ronai, V. Speth, J. Seelig, and A. Blume, 1975. Thermotropic lipid clustering in tetrahymena membranes. *Biochemistry* 14:3730–3735.
- [26] Lee, A. G., N. J. Birdsall, J. C. Metcalfe, P. A. Toon, and G. B. Warren, 1974. Clusters in lipid bilayers and the interpretation of thermal effects in biological membranes. *Biochemistry* 13:3699–3705.
- [27] Brown, D. A., and J. K. Rose, 1992. Sorting of GPI-anchored proteins to glycolipid-enriched membrane subdomains during transport to the apical cell surface. *Cell* 68:533–544.
- [28] Pike, L. J., 2009. The challenge of lipid rafts. *J. Lipid Res.* 50:S323–S328.
- [29] Scolari, S., S. Engel, N. Krebs, A. P. Plazzo, R. F. M. De Almeida, M. Prieto, M. Veit, and A. Herrmann, 2009. Lateral Distribution of the Transmembrane Domain of Influenza Virus Hemagglutinin Revealed by Time-resolved Fluorescence Imaging. *J. Biol. Chem.* 284:15708–15716.
- [30] Engel, S., S. Scolari, B. Thaa, N. Krebs, T. Korte, A. Herrmann, and M. Veit, 2010. FLIM-FRET and FRAP reveal association of influenza virus haemagglutinin with membrane rafts. *Biochem. J.* 425:567–573.
- [31] Suzuki, K. G., T. K. Fujiwara, F. Sanematsu, R. Iino, M. Edidin, and A. Kusumi, 2007. GPI-anchored receptor clusters transiently recruit Lyn and G for temporary cluster immobilization and Lyn activation: single-molecule tracking study 1. *J. Cell Biol.* 177:717–730.
- [32] Pike, L. J., 2006. Rafts defined: a report on the Keystone symposium on lipid rafts and cell function. *J. Lipid Res.* 47:1597–1598.

- [33] Marsh, D., 2009. Cholesterol-induced fluid membrane domains: A compendium of lipid-raft ternary phase diagrams. *BBA - Biomembranes* 1788:2114–2123.
- [34] Simons, K., and W. L. Vaz, 2004. Model Systems, Lipid Rafts, and Cell Membranes. *Annu. Rev. Biophys. Biomol. Struct.* 33:269–295.
- [35] Garca-Sez, A. J., S. Chiantia, and P. Schwille, 2007. Effect of Line Tension on the Lateral Organization of Lipid Membranes. *J. Biol. Chem.* 282:33537–33544.
- [36] Kahya, N., D. Scherfeld, K. Bacia, B. Poolman, and P. Schwille, 2003. Probing Lipid Mobility of Raft-exhibiting Model Membranes by Fluorescence Correlation Spectroscopy. *J. Biol. Chem.* 278:28109–28115.
- [37] Nikolaus, J., S. Scolari, E. Bayraktarov, N. Jungnick, S. Engel, A. P. Plazzo, M. Stekl, R. Volkmer, M. Veit, and A. Herrmann, 2010. Hemagglutinin of Influenza Virus Partitions into the Nonraft Domain of Model Membranes. *Biophys. J.* 99:489–498.
- [38] Fastenberg, M. E., H. Shogomori, X. Xu, D. A. Brown, and E. London, 2003. Exclusion of a transmembrane-type peptide from ordered-lipid domains (rafts) detected by fluorescence quenching: extension of quenching analysis to account for the effects of domain size and domain boundaries. *Biochemistry* 42:12376–12390.
- [39] Bacia, K., C. G. Schuette, N. Kahya, R. Jahn, and P. Schwille, 2004. SNAREs prefer liquid-disordered over "raft" (liquid-ordered) domains when reconstituted into giant unilamellar vesicles. *J. Biol. Chem.* 279:37951–37955.
- [40] Hammond, A. T., F. A. Heberle, T. Baumgart, D. Holowka, B. Baird, and G. W. Feigenson, 2005. Crosslinking a lipid raft component triggers liquid ordered-liquid disordered phase separation in model plasma membranes. *Proc. Natl. Acad. Sci. USA* 102:6320–6325.
- [41] Shogomori, H., A. T. Hammond, A. G. Ostermeyer-Fay, D. J. Barr, G. W. Feigenson, E. London, and D. A. Brown, 2005. Palmitoylation and intracellular domain interactions both contribute to raft targeting of linker for activation of T cells. *J. Biol. Chem.* 280:18931–18942.
- [42] Nielsen, S. O., C. F. Lopez, G. Srinivas, and M. L. Klein, 2004. Coarse grain models and the computer simulation of soft materials. *J. Phys. Condens. Matter* 16:R481–R512.

- [43] Marrink, S. J., A. H. de Vries, and A. E. Mark, 2004. Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B* 108:750–760.
- [44] Marrink, S. J., H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. d. Vries, 2007. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* 111:7812–7824.
- [45] Monticelli, L., S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S. J. Marrink, 2008. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* 4:819–834.
- [46] Bond, P. J., C. L. Wee, and M. S. P. Sansom, 2008. Coarse-grained molecular dynamics simulations of the energetics of helix insertion into a lipid bilayer. *Biochemistry* 47:11321–11331.
- [47] Shih, A. Y., A. Arkhipov, P. L. Freddolino, and K. Schulten, 2006. Coarse grained protein-lipid model with application to lipoprotein particles. *The Journal of Physical Chemistry. B* 110:3674–3684.
- [48] Izvekov, S., and G. A. Voth, 2005. A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B* 109:2469–2473.
- [49] Izvekov, S., and G. A. Voth, 2005. Multiscale coarse graining of liquid-state systems. *J. Chem. Phys.* 123:134105.
- [50] Noid, W. G., J. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, 2008. The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.* 128:244114.
- [51] Noid, W. G., P. Liu, Y. Wang, J. Chu, G. S. Ayton, S. Izvekov, H. C. Andersen, and G. A. Voth, 2008. The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models. *J. Chem. Phys.* 128:244115.
- [52] Shelley, J. C., M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, and M. L. Klein, 2001. A Coarse Grain Model for Phospholipid Simulations. *J. Phys. Chem. B* 105:4464–4470.
- [53] Kranenburg, M., M. Venturoli, and B. Smit, 2003. Phase Behavior and Induced Interdigitation in Bilayers Studied with Dissipative Particle Dynamics. *J. Phys. Chem. B* 107:11491–11501.

- [54] Schmidt, U., G. Guigas, and M. Weiss, 2008. Cluster Formation of Transmembrane Proteins Due to Hydrophobic Mismatching. *Phys. Rev. Lett.* 101:128104.
- [55] Venturoli, M., M. Maddalena Sperotto, M. Kranenburg, and B. Smit, 2006. Mesoscopic models of biological membranes. *Physics Reports* 437:1–54.
- [56] Voth, G. A., 2009. Coarse-Graining of Condensed Phase and Biomolecular Systems. CRC Press/Taylor and Francis Group, Boca Raton, 1st edition.
- [57] Schmidt, U., and M. Weiss, 2010. Hydrophobic mismatch-induced clustering as a primer for protein sorting in the secretory pathway. *Biophys. Chem.* 151:34–38.
- [58] Orsi, M., D. Y. Haubertin, W. E. Sanderson, and J. W. Essex, 2008. A Quantitative Coarse-Grain Model for Lipid Bilayers. *J. Phys. Chem. B* 112:802–815.
- [59] Morozova, D., and M. Weiss, 2010. On the Role of Acylation of Transmembrane Proteins. *Biophys. J.* 98:800–804.
- [60] Kranenburg, M., M. Venturoli, and B. Smit, 2003. Molecular simulations of mesoscopic bilayer phases. *Physical Review E* 67:060901.
- [61] Venturoli, M., B. Smit, and M. M. Sperotto, 2005. Simulation Studies of Protein-Induced Bilayer Deformations, and Lipid-Induced Protein Tilting, on a Mesoscopic Model for Lipid Bilayers with Embedded Proteins. *Biophys. J.* 88:1778–1798.
- [62] de Meyer, F. J., M. Venturoli, and B. Smit, 2008. Molecular Simulations of Lipid-Mediated Protein-Protein Interactions. *Biophys. J.* 95:1851–1865.
- [63] de Meyer, F. J., J. M. Rodgers, T. F. Willems, and B. Smit, 2010. Molecular simulation of the effect of cholesterol on lipid-mediated protein-protein interactions. *Biophys. J.* 99:3629–3638.
- [64] Gay, J. G., and B. J. Berne, 1981. Modification of the overlap potential to mimic a linear site-site potential. *J. Chem. Phys.* 74:3316.
- [65] Bond, P. J., D. L. Parton, J. F. Clark, and M. S. P. Sansom, 2008. Coarse-Grained Simulations of the Membrane-Active Antimicrobial Peptide Maculatin 1.1. *Biophys. J.* 95:3802–3815.
- [66] Skehel, J. J., and D. C. Wiley, 2000. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu. Rev. Biochem.* 69:531–69.

- [67] Nayak, D. P., R. A. Balogun, H. Yamada, Z. H. Zhou, and S. Barman, 2009. Influenza virus morphogenesis and budding. *Virus Res.* 143:147–161.
- [68] Harris, A., G. Cardone, D. C. Winkler, J. B. Heymann, M. Brecher, J. M. White, and A. C. Steven, 2006. Influenza virus pleiomorphy characterized by cryoelectron tomography. *Proc. Natl. Acad. Sci. USA* 103:19123–7.
- [69] Yamaguchi, M., R. Danev, K. Nishiyama, K. Sugawara, and K. Nagayama, 2008. Zernike phase contrast electron microscopy of ice-embedded influenza A virus. *J. Struct. Biol.* 162:271–276.
- [70] Calder, L. J., S. Wasilewski, J. A. Berriman, and P. B. Rosenthal, 2010. Structural organization of a filamentous influenza A virus. *Proc. Natl. Acad. Sci. USA* 107:10685–10690.
- [71] Takeda, M., A. Pekosz, K. Shuck, L. H. Pinto, and R. A. Lamb, 2002. Influenza a virus M2 ion channel activity is essential for efficient replication in tissue culture. *J. Virol.* 76:1391–1399.
- [72] Boulo, S., H. Akarsu, R. W. H. Ruigrok, and F. Baudin, 2007. Nuclear traffic of influenza virus proteins and ribonucleoprotein complexes. *Virus Res.* 124:12–21.
- [73] Dalton, R. M., A. E. Mullin, M. J. Amorim, E. Medcalf, L. S. Tiley, and P. Digard, 2006. Temperature sensitive influenza A virus genome replication results from low thermal stability of polymerase-cRNA complexes. *Virol. J.* 3:58.
- [74] Chen, B. J., G. P. Leser, E. Morita, and R. A. Lamb, 2007. Influenza Virus Hemagglutinin and Neuraminidase, but Not the Matrix Protein, Are Required for Assembly and Budding of Plasmid-Derived Virus-Like Particles. *J. Virol.* 81:7111–7123.
- [75] Leach, A., 2001. *Molecular Modelling: Principles and Applications*. Prentice Hall, New York, 2nd edition.
- [76] Allen, M. P., and D. J. Tildesley, 1989. *Computer Simulation of Liquids*. Clarendon Press, Oxford.
- [77] Frenkel, D., and B. Smit, 2002. *Understanding Molecular Simulation*. Academic Press, San Diego.

- [78] van Gunsteren, W. F., S. R. Billeter, A. A. Eising, P. H. Hunenberger, P. Kruger, A. E. Mark, W. R. P. Scott, and I. G. Tironi, 1996. Biomolecular Simulation: The GROMOS96 Manual and User Guide. vdf Hochschulverlag AG an der ETH Zurich and BIOMOS b.v., Zurich.
- [79] MacKerell, D. Bashford, Bellott, Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wirkiewicz-Kuczera, D. Yin, and M. Karplus, 1998. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* 102:3586–3616.
- [80] Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217.
- [81] Brooks, B. R., C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, 2009. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* 30:1545–1614.
- [82] Case, D., T. Cheatham, T. Darden, H. Gohlke, R. Luo, K. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. Woods, 2005. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* 26:1668–1688.
- [83] Jorgensen, W. L., and J. Tirado-Rives, 1988. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* 110:1657–1666.
- [84] Jorgensen, W. L., D. S. Maxwell, and J. Tirado-Rives, 1996. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* 118:11225–11236.
- [85] Berendsen, H. J. C., D. van der Spoel, and R. van Drunen, 1995. GROMACS: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.* 91:43–56.

- [86] Lindahl, E., B. Hess, and D. van der Spoel, 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* 7:306–317.
- [87] Spoel, D. V. D., E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, 2005. GROMACS: fast, exible, and free. *J. Comput. Chem.* 26:1701–1718.
- [88] Hess, B., C. Kutzner, D. v. d. Spoel, and E. Lindahl, 2008. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* 4:435–447.
- [89] Castro-Roman, F., R. W. Benz, S. H. White, and D. J. Tobias, 2006. Investigation of finite system-size effects in molecular dynamics simulations of lipid bilayers. *J. Phys. Chem. B* 110:24157–24164.
- [90] Cheatham, T. E. I., J. L. Miller, T. Fox, T. A. Darden, and P. A. Kollman, 1995. Molecular Dynamics Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins. *J. Am. Chem. Soc.* 117:4193–4194.
- [91] Norberg, J., 2000. On the Truncation of Long-Range Electrostatic Interactions in DNA. *Biophys. J.* 79:1537–1553.
- [92] Patra, M., M. Karttunen, M. T. Hyvnen, E. Falck, P. Lindqvist, and I. Vattulainen, 2003. Molecular dynamics simulations of lipid bilayers: major artifacts due to truncating electrostatic interactions. *Biophys. J.* 84:3636–3645.
- [93] Darden, T., D. York, and L. Pedersen, 1993. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089.
- [94] Tironi, I. G., R. Sperb, P. E. Smith, and W. F. van Gunsteren, 1995. A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.* 102:5451.
- [95] Shan, Y., J. L. Klepeis, M. P. Eastwood, R. O. Dror, and D. E. Shaw, 2005. Gaussian split Ewald: A fast Ewald mesh method for molecular simulation. *J. Chem. Phys.* 122:054101.
- [96] Ewald, P. P., 1921. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* 393:253–287.

- [97] Spoel, D. v. d., E. Lindahl, B. Hess, C. Kutzner, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. v. Drunen, and H. J. C. Berendsen, 2011. Gromacs User Manual version 4.5.4. www.gromacs.org.
- [98] Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.
- [99] Nose, S., 1984. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* 81:511.
- [100] Hoover, W. G., 1985. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* 31:1695.
- [101] Schfer, L. V., D. H. de Jong, A. Holt, A. J. Rzepiela, A. H. de Vries, B. Poolman, J. A. Killian, and S. J. Marrink, 2011. Lipid packing drives the segregation of transmembrane helices into disordered lipid domains in model membranes. *Proc. Natl. Acad. Sci. USA* 108:1343–1348.
- [102] Bond, P. J., and M. S. P. Sansom, 2006. Insertion and assembly of membrane proteins via simulation. *J. Am. Chem. Soc.* 128:2697–2704.
- [103] Radzicka, A., and R. Wolfenden, 1988. Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry* 27:1664–1670.
- [104] Maloy, W. L., and U. P. Kari, 1995. Structure-activity studies on magainins and other host defense peptides. *Biopolymers* 37:105–122.
- [105] Steiner, H., D. Hultmark, A. Engstrom, H. Bennich, and H. G. Boman, 1981. Sequence and specificity of two antibacterial proteins involved in insect immunity. *Nature* 292:246–248.
- [106] Zasloff, M., 1987. Magainins, a class of antimicrobial peptides from *Xenopus* skin: isolation, characterization of two active forms, and partial cDNA sequence of a precursor. *Proc. Natl. Acad. Sci. USA* 84:5449–5453.
- [107] Rozek, T., R. J. Waugh, S. T. Steinborner, J. H. Bowie, M. J. Tyler, and J. C. Wallace, 1998. The Maculatin peptides from the skin glands of the tree frog *Litoria genimaculata*:

- a comparison of the structures and antibacterial activities of Maculatin 1.1 and Caerin 1.1. *J. Pept. Sci.* 4:111–115.
- [108] Oren, Z., and Y. Shai, 1998. Mode of action of linear amphipathic alpha-helical antimicrobial peptides. *Biopolymers* 47:451–463.
- [109] Shai, Y., 2002. Mode of action of membrane active antimicrobial peptides. *Biopolymers* 66:236–248.
- [110] Chia, B. C., J. A. Carver, T. D. Mulhern, and J. H. Bowie, 2000. Maculatin 1.1, an anti-microbial peptide from the Australian tree frog, *Litoria genimaculata* solution structure and biological activity. *Eur. J. Biochem.* 267:1894–1908.
- [111] Marcotte, I., K. L. Wegener, Y. Lam, B. C. S. Chia, M. R. R. de Planque, J. H. Bowie, M. Auger, and F. Separovic, 2003. Interaction of antimicrobial peptides from Australian amphibians with lipid membranes. *Chem. Phys. Lipids* 122:107–120.
- [112] Ambroggio, E. E., F. Separovic, J. H. Bowie, G. D. Fidelio, and L. A. Bagatolli, 2005. Direct visualization of membrane leakage induced by the antibiotic peptides: maculatin, citropin, and aurein. *Biophys. J.* 89:1874–1881.
- [113] Balla, M., J. Bowie, and F. Separovic, 2004. Solid-state NMR study of antimicrobial peptides from Australian frogs in phospholipid membranes. *Eur. Biophys. J.* 33:109–116.
- [114] Gehman, J. D., F. Luc, K. Hall, T. Lee, M. P. Boland, T. L. Pukala, J. H. Bowie, M. Aguilar, and F. Separovic, 2008. Effect of Antimicrobial Peptides from Australian Tree Frogs on Anionic Phospholipid Membranes. *Biochemistry* 47:8557–8565.
- [115] Chia, B., Y. Lam, M. Dyll-Smith, F. Separovic, and J. Bowie, 2000. A ^{31}P NMR study of the interaction of amphibian antimicrobial peptides with the membranes of live bacteria. *Lett. Pept. Sci.* 7:151–156.
- [116] Pukala, T. L., J. H. Bowie, V. M. Maselli, I. F. Musgrave, and M. J. Tyler, 2006. Host-defence peptides from the glandular secretions of amphibians: structure and activity. *Nat. Prod. Rep.* 23:368–393.
- [117] Chia, C. S. B., J. Torres, M. A. Cooper, I. T. Arkin, and J. H. Bowie, 2002. The orientation of the antibiotic peptide maculatin 1.1 in DMPG and DMPC lipid bilayers. Support for a pore-forming mechanism. *FEBS Lett.* 512:47–51.

- [118] Mechler, A., S. Praporski, K. Atmuri, M. Boland, F. Separovic, and L. L. Martin, 2007. Specific and Selective Peptide-Membrane Interactions Revealed Using Quartz Crystal Microbalance. *Biophys. J.* 93:3907–3916.
- [119] Leontiadou, H., A. Mark, and S. Marrink, 2006. Antimicrobial Peptides in Action. *J. Am. Chem. Soc.* 128:12156–12161.
- [120] Sengupta, D., H. Leontiadou, A. E. Mark, and S. J. Marrink, 2008. Toroidal pores formed by antimicrobial peptides show significant disorder. *Biochim. Biophys. Acta* 1778:2308–2317.
- [121] Mihajlovic, M., and T. Lazaridis, 2010. Antimicrobial peptides in toroidal and cylindrical pores. *Biochim. Biophys. Acta* 1798:1485–1493.
- [122] Lindahl, E., and M. S. P. Sansom, 2008. Membrane proteins: molecular dynamics simulations. *Curr. Opin. Struct. Biol.* 18:425–431.
- [123] Lopez, C. F., S. O. Nielsen, G. Srinivas, W. F. DeGrado, and M. L. Klein, 2006. Probing Membrane Insertion Activity of Antimicrobial Polymers via Coarse-Grain Molecular Dynamics. *J. Chem. Theory Comput.* 2:649–655.
- [124] Rzepiela, A., D. Sengupta, N. Goga, and S. J. Marrink, 2010. Membrane poration by antimicrobial peptides combining atomistic and coarse grained descriptions. *Farad. Discuss.* 144:431–443.
- [125] Woo, H., and A. Wallqvist, 2011. Spontaneous buckling of lipid bilayer and vesicle budding induced by antimicrobial peptide magainin 2: a coarse-grained simulation study. *J. Phys. Chem. B* 115:8122–8129.
- [126] Thogersen, L., B. Schiott, T. Vosegaard, N. Nielsen, and E. Tajkhorshid, 2008. Peptide Aggregation and Pore Formation in a Lipid Bilayer: A Combined Coarse-Grained and All Atom Molecular Dynamics Study. *Biophys. J.* 95:4337–4347.
- [127] Khalfa, A., W. Treptow, B. Maigret, and M. Tarek, 2009. Self assembly of peptides near or within membranes using coarse grained MD simulations. *Chem. Phys.* 358:161–170.
- [128] Jean-Francois, F., J. Elezgaray, P. Berson, P. Vacher, and E. J. Dufourc, 2008. Pore Formation Induced by an Antimicrobial Peptide: Electrostatic Effects. *Biophys. J.* 95:5748–5756.

- [129] Killian, J., I. Salemink, M. de Planque, G. Lindblom, R. Koeppe, and D. Greathouse, 1996. Induction of Nonbilayer Structures in Diacylphosphatidylcholine Model Membranes by Transmembrane alpha-Helical Peptides: Importance of Hydrophobic Mismatch and Proposed Role of Tryptophans. *Biochemistry* 35:1037–1045.
- [130] Stansfeld, P. J., R. Hopkinson, F. M. Ashcroft, and M. S. P. Sansom, 2009. PIP2-Binding Site in Kir Channels: Definition by Multiscale Biomolecular Simulations. *Biochemistry* 48:10926–10933.
- [131] Scott, W. R. P., P. H. Hunenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Kruger, and W. F. van Gunsteren, 1999. The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A* 103:3596–3607.
- [132] Berger, O., O. Edholm, and F. Jhnig, 1997. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys. J.* 72:2002–2013.
- [133] Humphrey, W., A. Dalke, and K. Schulten, 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–8, 27–8.
- [134] Kabsch, W., and C. Sander, 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.
- [135] Beckstein, O., and M. S. P. Sansom, 2004. The influence of geometry, surface character and flexibility on the permeation of ions and water through biological pores. *Phys. Biol.* 1:42–52.
- [136] Zhu, F., E. Tajkhorshid, and K. Schulten, 2004. Theory and simulation of water permeation in aquaporin-1. *Biophys. J.* 86:50–57.
- [137] Bemporad, D., J. W. Essex, and C. Luttmann, 2004. Permeation of Small Molecules through a Lipid Bilayer: A Computer Simulation Study. *J. Phys. Chem. B* 108:4875–4884.
- [138] Finkelstein, A., 1976. Water and nonelectrolyte permeability of lipid bilayer membranes. *J. Gen. Physiol.* 68:127–135.
- [139] Paula, S., A. G. Volkov, A. N. Van Hoek, T. H. Haines, and D. W. Deamer, 1996. Permeation of protons, potassium ions, and small polar molecules through phospholipid bilayers as a function of membrane thickness. *Biophys. J.* 70:339–348.

- [140] de Groot, B. L., and H. Grubmuller, 2001. Water Permeation Across Biological Membranes: Mechanism and Dynamics of Aquaporin-1 and GlpF. *Science* 294:2353–2357.
- [141] Smolin, N., B. Li, D. A. C. Beck, and V. Daggett, 2008. Side-Chain Dynamics Are Critical for Water Permeation through Aquaporin-1. *Biophys. J.* 95:1089–1098.
- [142] Killian, J. A., and G. von Heijne, 2000. How proteins adapt to a membrane-water interface. *Trends Biochem. Sci.* 25:429–434.
- [143] Killian, J. A., and T. K. M. Nyholm, 2006. Peptides in lipid bilayers: the power of simple models. *Curr. Opin. Struct. Biol.* 16:473–479.
- [144] Fernandez, D. I., J. D. Gehman, and F. Separovic, 2009. Membrane interactions of antimicrobial peptides from Australian frogs. *Biochim. Biophys. Acta* 1788:1630–1638.
- [145] Mouritsen, O. G., and M. Bloom, 1993. Models of Lipid-Protein Interactions in Membranes. *Annu. Rev. Biophys. Biomol. Struct.* 22:145–171.
- [146] Killian, J. A., 1998. Hydrophobic mismatch between proteins and lipids in membranes. *Biochim. Biophys. Acta* 1376:401–415.
- [147] Marsh, D., 2008. Protein modulation of lipids, and vice-versa, in membranes. *Biochim. Biophys. Acta* 1778:1545–1575.
- [148] Marsh, D., 2008. Energetics of hydrophobic matching in lipid-protein interactions. *Biophys. J.* 94:3996–4013.
- [149] Petrache, H. I., D. M. Zuckerman, J. N. Sachs, J. A. Killian, R. E. Koeppe, and T. B. Woolf, 2002. Hydrophobic Matching Mechanism Investigated by Molecular Dynamics Simulations. *Langmuir* 18:1340–1351.
- [150] de Planque, M. R., E. Goormaghtigh, D. V. Greathouse, n. Koeppe, R E, J. A. Kruijtzer, R. M. Liskamp, B. de Kruijff, and J. A. Killian, 2001. Sensitivity of single membrane-spanning alpha-helical peptides to hydrophobic mismatch with a lipid bilayer: effects on backbone structure, orientation, and extent of membrane incorporation. *Biochemistry* 40:5000–5010.
- [151] O’Keeffe, A. H., J. M. East, and A. G. Lee, 2000. Selectivity in lipid binding to the bacterial outer membrane protein OmpF. *Biophys. J.* 79:2066–2074.

- [152] Soubias, O., W. E. Teague Jr., K. G. Hines, D. C. Mitchell, and K. Gawrisch, 2010. Contribution of Membrane Elastic Energy to Rhodopsin Function. *Biophys. J.* 99:817–824.
- [153] Sparr, E., W. L. Ash, P. V. Nazarov, D. T. S. Rijkers, M. A. Hemminga, D. P. Tieleman, and J. A. Killian, 2005. Self-association of transmembrane alpha-helices in model membranes: importance of helix orientation and role of hydrophobic mismatch. *J. Biol. Chem.* 280:39324–39331.
- [154] Klingelhoefer, J. W., T. Carpenter, and M. S. P. Sansom, 2009. Peptide nanopores and lipid bilayers: interactions by coarse-grained molecular-dynamics simulations. *Biophys. J.* 96:3519–3528.
- [155] Popot, J. L., and D. M. Engelman, 2000. Helical membrane protein folding, stability, and evolution. *Annu. Rev. Biochem.* 69:881–922.
- [156] Koebnik, R., K. P. Locher, and P. Van Gelder, 2000. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol. Microbiol.* 37:239–253.
- [157] Botelho, A. V., T. Huber, T. P. Sakmar, and M. F. Brown, 2006. Curvature and hydrophobic forces drive oligomerization and modulate activity of rhodopsin in membranes. *Biophys. J.* 91:4464–4477.
- [158] Dumas, F., M. M. Sperotto, M. C. Lebrun, J. F. Tocanne, and O. G. Mouritsen, 1997. Molecular sorting of lipids by bacteriorhodopsin in dilauroylphosphatidylcholine/distearoylphosphatidylcholine lipid bilayers. *Biophys. J.* 73:1940–1953.
- [159] Nielsen, S. O., C. F. Lopez, I. Ivanov, P. B. Moore, J. C. Shelley, and M. L. Klein, 2004. Transmembrane peptide-induced lipid sorting and mechanism of Lalpha-to-inverted phase transition using coarse-grain molecular dynamics. *Biophys. J.* 87:2107–2115.
- [160] Opella, S. J., F. M. Marassi, J. J. Gesell, A. P. Valente, Y. Kim, M. Oblatt-Montal, and M. Montal, 1999. Structures of the M2 channel-lining segments from nicotinic acetylcholine and NMDA receptors by NMR spectroscopy. *Nat. Struct. Biol.* 6:374–379.
- [161] Song, L., M. R. Hobaugh, C. Shustak, S. Cheley, H. Bayley, and J. E. Gouaux, 1996. Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* 274:1859–1866.

- [162] Klingelhofer, J. W., 2010. Biophysics of Nanopores - Multiscale Molecular Dynamics Simulation Studies. DPhil thesis, University of Oxford.
- [163] Yau, W. M., W. C. Wimley, K. Gawrisch, and S. H. White, 1998. The preference of tryptophan for membrane interfaces. *Biochemistry* 37:14713–14718.
- [164] de Planque, M. R., J. A. Kruijtz, R. M. Liskamp, D. Marsh, D. V. Greathouse, n. Koeppe, R E, B. de Kruijff, and J. A. Killian, 1999. Different membrane anchoring positions of tryptophan and lysine in synthetic transmembrane alpha-helical peptides. *J. Biol. Chem.* 274:20839–20846.
- [165] Lear, J. D., Z. R. Wasserman, and W. F. DeGrado, 1988. Synthetic amphiphilic peptide models for protein ion channels. *Science* 240:1177–1181.
- [166] Periole, X., T. Huber, S. Marrink, and T. P. Sakmar, 2007. G protein-coupled receptors self-assemble in dynamics simulations of model bilayers. *J. Am. Chem. Soc.* 129:10126–10132.
- [167] Risselada, H. J., A. E. Mark, and S. J. Marrink, 2008. Application of Mean Field Boundary Potentials in Simulations of Lipid Vesicles. *J. Phys. Chem. B* 112:7438–7447.
- [168] Winger, M., D. Trzesniak, R. Baron, and W. F. van Gunsteren, 2009. On using a too large integration time step in molecular dynamics simulations of coarse-grained molecular models. *Phys. Chem. Chem. Phys.* 11:1934–1941.
- [169] Marrink, S. J., X. Periole, D. P. Tieleman, and A. H. de Vries, 2010. Comment on On using a too large integration time step in molecular dynamics simulations of coarse-grained molecular models by M. Winger, D. Trzesniak, R. Baron and W. F. van Gunsteren, *Phys. Chem. Chem. Phys.*, 2009, 11, 1934. *Phys. Chem. Chem. Phys.* 12:2254–2256.
- [170] Periole, Cavalli, Marrink, and Ceruso, 2009. Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *J. Chem. Theory Comput.* 5:2531–2543.
- [171] Niemela, P. S., M. S. Miettinen, L. Monticelli, H. Hammaren, P. Bjelkmar, T. Murtola, E. Lindahl, and I. Vattulainen, 2010. Membrane Proteins Diffuse as Dynamic Complexes with Lipids. *J. Am. Chem. Soc.* 132:7574–7575.

- [172] Ramadurai, S., A. Holt, L. V. Schfer, V. V. Krasnikov, D. T. Rijkers, S. J. Marrink, J. A. Killian, and B. Poolman, 2010. Influence of Hydrophobic Mismatch and Amino Acid Composition on the Lateral Diffusion of Transmembrane Peptides. *Biophys. J.* 99:1447–1454.
- [173] Gaede, H., 2003. Lateral Diffusion Rates of Lipid, Water, and a Hydrophobic Drug in a Multilamellar Liposome. *Biophys. J.* 85:1734–1740.
- [174] East, J. M., D. Melville, and A. G. Lee, 1985. Exchange rates and numbers of annular lipids for the calcium and magnesium ion dependent adenosinetriphosphatase. *Biochemistry* 24:2615–2623.
- [175] Marsh, D., 2008. Electron spin resonance in membrane research: protein-lipid interactions. *Methods* 46:83–96.
- [176] Sharpe, H. J., T. J. Stevens, and S. Munro, 2010. A comprehensive comparison of transmembrane domains reveals organelle-specific properties. *Cell* 142:158–169.
- [177] Pali, T., D. Bashtovyy, and D. Marsh, 2006. Stoichiometry of lipid interactions with transmembrane proteins—Deduced from the 3D structures. *Prot. Sci.* 15:1153–1161.
- [178] Nielsen, S. O., B. Ensing, V. Ortiz, P. B. Moore, and M. L. Klein, 2005. Lipid Bilayer Perturbations around a Transmembrane Nanotube: A Coarse Grain Molecular Dynamics Study. *Biophys. J.* 88:3822–3828.
- [179] Guigas, G., and M. Weiss, 2008. Influence of Hydrophobic Mismatching on Membrane Protein Diffusion. *Biophys. J.* 95:L25–L27.
- [180] Yin, F., and J. T. Kindt, 2010. Atomistic Simulation of Hydrophobic Matching Effects on Lipid Composition near a Helical Peptide Embedded in Mixed-Lipid Bilayers. *J. Phys. Chem. B* 114:8076–8080.
- [181] Kim, T., and W. Im, 2010. Revisiting Hydrophobic Mismatch with Free Energy Simulation Studies of Transmembrane Helix Tilt and Rotation. *Biophys. J.* 99:175–183.
- [182] Scott, K. A., P. J. Bond, A. Ivetac, A. P. Chetwynd, S. Khalid, and M. S. P. Sansom, 2008. Coarse-grained MD simulations of membrane protein-bilayer self-assembly. *Structure* 16:621–630.

- [183] Sengupta, D., and S. J. Marrink, 2010. Lipid-mediated interactions tune the association of glycoporphin A helix and its disruptive mutants in membranes. *Phys. Chem. Chem. Phys.* 12:12987.
- [184] Janosi, L., A. Prakash, and M. Doxastakis, 2010. Lipid-modulated sequence-specific association of glycoporphin A in membranes. *Biophys. J.* 99:284–292.
- [185] Hnin, J., A. Pohorille, and C. Chipot, 2005. Insights into the Recognition and Association of Transmembrane -Helices. The Free Energy of -Helix Dimerization in Glycophorin A. *J. Am. Chem. Soc.* 127:8478–8484.
- [186] Lewis, B. A., and D. M. Engelman, 1983. Bacteriorhodopsin remains dispersed in fluid phospholipid bilayers over a wide range of bilayer thicknesses. *J. Mol. Biol.* 166:203–210.
- [187] Hu, Y., L. Qu, and T. Schikorski, 2008. Mean synaptic vesicle size varies among individual excitatory hippocampal synapses. *Synapse* 62:953–957.
- [188] Hasler, L., J. B. Heymann, A. Engel, J. Kistler, and T. Walz, 1998. 2D crystallization of membrane proteins: rationales and examples. *J. Struct. Biol.* 121:162–171.
- [189] Edidin, M., 2003. The state of lipid rafts: from model membranes to cells. *Annu. Rev. Biophys. Biomol. Struct.* 32:257–283.
- [190] Jacobson, K., O. G. Mouritsen, and R. G. W. Anderson, 2007. Lipid rafts: at a crossroad between cell biology and physics. *Nat. Cell. Biol.* 9:7–14.
- [191] Subczynski, W. K., and A. Kusumi, 2003. Dynamics of raft molecules in the cell and artificial membranes: approaches by pulse EPR spin labeling and single molecule optical microscopy. *Biochim. Biophys. Acta* 1610:231–243.
- [192] Scheiffele, P., A. Rietveld, T. Wilk, and K. Simons, 1999. Influenza Viruses Select Ordered Lipid Domains during Budding from the Plasma Membrane. *J. Biol. Chem.* 274:2038–2044.
- [193] Blough, H. A., 1971. Fatty acid composition of individual phospholipids of influenza virus. *J. Gen. Virol.* 12:317–320.
- [194] Blough, H. A., and J. P. Merlie, 1970. The lipids of incomplete influenza virus. *Virology* 40:685–692.

- [195] Skibbens, J. E., M. G. Roth, and K. S. Matlin, 1989. Differential extractability of influenza virus hemagglutinin during intracellular transport in polarized epithelial cells and nonpolar fibroblasts. *J. Cell Biol.* 108:821–832.
- [196] Fiedler, K., T. Kobayashi, T. V. Kurzchalia, and K. Simons, 1993. Glycosphingolipid-enriched, detergent-insoluble complexes in protein sorting in epithelial cells. *Biochemistry* 32:6365–6373.
- [197] Takeda, M., G. P. Leser, C. J. Russell, and R. A. Lamb, 2003. Influenza virus hemagglutinin concentrates in lipid raft microdomains for efficient viral fusion. *Proc. Natl. Acad. Sci. USA* 100:14610–14617.
- [198] Hess, S. T., M. Kumar, A. Verma, J. Farrington, A. Kenworthy, and J. Zimmerberg, 2005. Quantitative electron microscopy and fluorescence spectroscopy of the membrane distribution of influenza hemagglutinin. *J. Cell Biol.* 169:965–976.
- [199] Silvius, J. R., 2003. Role of cholesterol in lipid raft formation: lessons from lipid model systems. *Biochim. Biophys. Acta* 1610:174–183.
- [200] Niemel, P. S., S. Ollila, M. T. Hyvnen, M. Karttunen, and I. Vattulainen, 2007. Assessing the nature of lipid raft membranes. *PLoS Comput. Biol.* 3:e34.
- [201] Apajalahti, T., P. Niemela, P. N. Govindan, M. S. Miettinen, E. Salonen, S. Marrink, and I. Vattulainen, 2010. Concerted diffusion of lipids in raft-like membranes. *Farad. Discuss.* 144:411–430.
- [202] Hall, A., T. Rog, M. Karttunen, and I. Vattulainen, 2010. Role of Glycolipids in Lipid Rafts: A View through Atomistic Molecular Dynamics Simulations with Galactosylceramide. *J. Phys. Chem. B* 114:7797–7807.
- [203] Risselada, H. J., and S. J. Marrink, 2008. The molecular face of lipid rafts in model membranes. *Proc. Natl. Acad. Sci. USA* 105:17367–72.
- [204] Perlmutter, J. D., and J. N. Sachs, 2011. Interleaflet interaction and asymmetry in phase separated lipid bilayers: molecular dynamics simulations. *J. Am. Chem. Soc.* 133:6563–6577.
- [205] Domanski, J., S. J. Marrink, and L. V. Schfer, 2011. Transmembrane helices can induce domain formation in crowded model membranes. *Biochim. Biophys. Acta* .

- [206] Ha, Y., D. J. Stevens, J. J. Skehel, and D. C. Wiley, 2003. X-ray structure of the hemagglutinin of a potential H3 avian progenitor of the 1968 Hong Kong pandemic influenza virus. *Virology* 309:209–218.
- [207] Wolf, M. G., M. Hoefling, C. Aponte-Santamara, H. Grubmüller, and G. Groenhof, 2010. g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation. *J. Comput. Chem.* 31:2169–2174.
- [208] Sding, J., A. Biegert, and A. N. Lupas, 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucl. Acids Res.* 33:W244–248.
- [209] Lopez, C. A., A. J. Rzepiela, A. H. de Vries, L. Dijkhuizen, P. H. Hunenberger, and S. J. Marrink, 2009. Martini Coarse-Grained Force Field: Extension to Carbohydrates. *J. Chem. Theory Comput.* 5:3195–3210.
- [210] Martnez, L., R. Andrade, E. G. Birgin, and J. M. Martnez, 2009. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* 30:2157–2164.
- [211] Michaud-Agrawal, N., E. J. Denning, T. B. Woolf, and O. Beckstein, 2011. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* 32:2319–2327.
- [212] Nicolini, C., J. Baranski, S. Schlummer, J. Palomo, M. Lumbierres-Burgues, M. Kahms, J. Kuhlmann, S. Sanchez, E. Gratton, H. Waldmann, and R. Winter, 2006. Visualizing Association of N-Ras in Lipid Microdomains: Influence of Domain Structure and Interfacial Adsorption. *J. Am. Chem. Soc.* 128:192–201.
- [213] Harder, T., P. Scheiffele, P. Verkade, and K. Simons, 1998. Lipid Domain Structure of the Plasma Membrane Revealed by Patching of Membrane Components. *J. Cell Biol.* 141:929–942.
- [214] Hancock, J. F., 2006. Lipid rafts: contentious only from simplistic standpoints. *Nat. Rev. Mol. Cell. Biol.* 7:456–462.
- [215] Giocondi, M., F. Ronzon, M. C. Nicolai, P. Dosset, P. Milhiet, M. Chevalier, and C. Le Grimellec, 2010. Organization of influenza A virus envelope at neutral and low pH. *J. Gen. Virol.* 91:329–338.

- [216] Nobusawa, E., T. Aoyama, H. Kato, Y. Suzuki, Y. Tateno, and K. Nakajima, 1991. Comparison of complete amino acid sequences and receptor-binding properties among 13 serotypes of hemagglutinins of influenza A viruses. *Virology* 182:475–485.
- [217] Blok, J., and G. M. Air, 1982. Variation in the membrane-insertion and "stalk" sequences in eight subtypes of influenza type A virus neuraminidase. *Biochemistry* 21:4001–4007.
- [218] Chen, B. J., G. P. Leser, D. Jackson, and R. A. Lamb, 2008. The influenza virus M2 protein cytoplasmic tail interacts with the M1 protein and influences virus assembly at the site of virus budding. *J. Virol.* 82:10059–10070.
- [219] Freddolino, P. L., A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten, 2006. Molecular Dynamics Simulations of the Complete Satellite Tobacco Mosaic Virus. *Structure* 14:437–449.
- [220] Varghese, J. N., and P. M. Colman, 1991. Three-dimensional structure of the neuraminidase of influenza virus A/Tokyo/3/67 at 2.2 Å resolution. *J. Mol. Biol.* 221:473–486.
- [221] Luo, G., J. Chung, and P. Palese, 1993. Alterations of the stalk of the influenza virus neuraminidase: deletions and insertions. *Virus Res.* 29:141–153.
- [222] Matsuoka, Y., D. E. Swayne, C. Thomas, M. Rameix-Welti, N. Naffakh, C. Warnes, M. Altholtz, R. Donis, and K. Subbarao, 2009. Neuraminidase Stalk Length and Additional Glycosylation of the Hemagglutinin Influence the Virulence of Influenza H5N1 Viruses for Mice. *J. Virol.* 83:4704–4708.
- [223] Castrucci, M. R., and Y. Kawaoka, 1993. Biologic importance of neuraminidase stalk length in influenza A virus. *J. Virol.* 67:759–764.
- [224] Harbury, P. B., T. Zhang, P. S. Kim, and T. Alber, 1993. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* 262:1401–1407.
- [225] Zhang, J., G. P. Leser, A. Pekosz, and R. A. Lamb, 2000. The Cytoplasmic Tails of the Influenza Virus Spike Glycoproteins Are Required for Normal Genome Packaging. *Virology* 269:325–334.

- [226] Schnell, J. R., and J. J. Chou, 2008. Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* 451:591–595.
- [227] Sugrue, R. J., R. B. Belshe, and A. J. Hay, 1990. Palmitoylation of the influenza A virus M2 protein. *Virology* 179:51–56.
- [228] Grantham, M. L., W. Wu, E. N. Lalime, M. E. Lorenzo, S. L. Klein, and A. Pekosz, 2009. Palmitoylation of the influenza A virus M2 protein is not required for virus replication in vitro but contributes to virus virulence. *J. Virol.* 83:8655–8661.
- [229] Brown, D. A., and E. London, 1998. Structure and origin of ordered lipid domains in biological membranes. *J. Membr. Biol.* 164:103–114.
- [230] Rossman, J. S., X. Jing, G. P. Leser, and R. A. Lamb, 2010. Influenza virus M2 protein mediates ESCRT-independent membrane scission. *Cell* 142:902–913.
- [231] Baumgart, T., A. T. Hammond, P. Sengupta, S. T. Hess, D. A. Holowka, B. A. Baird, and W. W. Webb, 2007. Large-scale fluid/fluid phase separation of proteins and lipids in giant plasma membrane vesicles. *Proc. Natl. Acad. Sci. USA* 104:3165–3170.
- [232] Lingwood, D., J. Ries, P. Schwille, and K. Simons, 2008. Plasma membranes are poised for activation of raft phase coalescence at physiological temperature. *Proc. Natl. Acad. Sci. USA* 105:10005–10010.
- [233] Ge, M., A. Gidwani, H. A. Brown, D. Holowka, B. Baird, and J. H. Freed, 2003. Ordered and disordered phases coexist in plasma membrane vesicles of RBL-2H3 mast cells. An ESR study. *Biophys. J.* 85:1278–1288.
- [234] Polozov, I. V., L. Bezrukov, K. Gawrisch, and J. Zimmerberg, 2008. Progressive ordering with decreasing temperature of the phospholipids of influenza virus. *Nat. Chem. Biol.* 4:248–255.
- [235] Roux, B., S. Bernche, B. Egwolf, B. Lev, S. Y. Noskov, C. N. Rowley, and H. Yu, 2011. Ion selectivity in channels and transporters. *J. Gen. Phys.* 137:415–426.
- [236] McGuffee, S. R., and A. H. Elcock, 2010. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput. Biol.* 6:e1000694.

- [237] Risselada, H. J., and S. J. Marrink, 2009. Curvature effects on lipid packing and dynamics in liposomes revealed by coarse grained molecular dynamics simulations. *Phys. Chem. Chem. Phys.* 11:2056–2067.