Characterisation of Lipid-Protein Interactions through Computational Modelling

George Hedger

Lincoln College

University of Oxford

A thesis submitted in partial fulfilment of

the requirements for the degree of

Doctor of Philosophy

Michaelmas 2017

Abstract

Parallel advances in lipidomics and the structural biology of membrane proteins over the past decade have revealed the highly diverse and complex nature of cell membrane composition. These compositional complexities influence the behaviour of embedded proteins, and a number of biomedically important membrane proteins are now known to be modulated *via* interactions with their lipid environment. This project aims to apply an array of computational tools to probe and characterize the molecular level detail of these interactions, and obtain information to complement experimental studies.

In particular the interactions of modulatory lipid species with the transmembrane domain of the epidermal growth factor (EGF) receptor were addressed using coarse-grained potential of mean force (PMF) calculations. The results suggest this approach, widely applied in other areas of computational biology, may be successfully adapted to probe the strength and selectivity of lipid-protein interactions, and the effects of protein mutation. Subsequently these calculations were applied in tandem with equilibrium simulations to explore cardiolipin interactions with a key mitochondrial transporter, the ADP/ATP Translocase. The data show the coarse-grained model employed is capable of accurately identifying three specific cardiolipin binding sites on the protein, in agreement with prior crystallographic and NMR data. These sites were shown to be specific to cardiolipin, rather than general phospholipid interaction sites.

In the third project, prospective predictions were made as to the location of PIP_2 lipid binding sites on the membrane-exposed surface of a class A GPCR, the Neurotensin Receptor 1 (NTS1). A protein known to specifically bind these lipids from native mass spectrometry (MS) measurements at hitherto unknown regions. The results suggest PIP_2 binds to defined loop regions on the intracellular portion of the protein.

Finally, atomistic simulations are applied to examine the effect of a crystallographically resolved cholesterol molecule on the dynamics of a class F GPCR, Smoothened. The data suggest a marked influence of bound cholesterol on intra-domain dynamics of the extracellular region of Smoothened.

Acknowledgements

Firstly, my thanks go to my academic supervisor, Prof. Mark Sansom, for his advice and support over the past four years. I was afforded an unusual degree of autonomy and a diverse array of academic opportunities during my PhD for which I am extremely grateful. Secondly my thanks must go to Dr. Heidi Koldsø, whom I first met in 2012 as a young undergraduate receiving tutorials on seven-transmembrane proteins. I went on to complete my undergraduate masters project under Heidi's supervision in the SBCB unit, before beginning my PhD under her and Mark's supervision. Heidi was the chief orchestrator of my entry into the world of simulations, and a catalyst for my decision to undertake a PhD. Now at the end of my PhD I look back and can confidently assert that the achievements and skills I have are a direct function of those early teachings. I will avoid waxing lyrical, but away from the academic side, Heidi became one of my very best friends with whom I have shared so many laughs, beers, and beautiful times. I look forward to drinking more beer and watching her get married in Denmark this summer.

Academically I must also extend my thanks to Prof. Christian Siebold. I spent several months working directly with Christian on the Smoothened project as we designed and executed simulations to address the effects of cholesterol on Smoothened. The energy and enthusiasm Christian possesses is quite infectious, and being involved so closely on such a high profile project with exemplary investigators remains one of my fondest memories. My thanks also to Prof. Dame Carol Robinson, Dr Hsin-Yung Yen, and Dr Idlir Liko, with whom I collaborated on GPCR-PIP_2 interactions. Day-to-day within the lab, there are dozens of small interactions taking place which shape one as a researcher. In particular I am grateful to Dr Anna Duncan and Dr Phill Stansfeld for their time and mentorship in providing so many such interactions which doubtless had a big impact on both my practical skills and perspectives as a researcher.

One of the arguable strengths, and tragedies, of academic life is the continual rapid turnover of people, and the lab I leave in 2017 is very different to the one in which I arrived in 2013. Both environments have been incredible places to work and provided wonderful experiences, though there was a noticeable decrease in alcohol consumption with the departure of Heidi and David. From the 'old' era my thanks in particular to Dr David Shorthouse, Dr Jemma Trick, Dr Nathalie Willems, Dr Matteo Aldeghi, Dr

Amanda Buyan, Dr Lukas Stelzl, Dr Sarah Rouse, Dr Maria Musgaard, and Dr Henriette Autzen. From more recent times my thanks to Dr Anna Duncan, Sarah-Beth Amos, Dr Tere Paramo, Dr Georgios Gerogiokas, Dr Ole Andersen, Dr Wanling Song, Marc Daemgen, Nick Michelarakis, Tom Newport, Fiona Naughton, and Jan 'Janski' Domanski.

Away from the lab, I have been fortunate to have an incredible group of friends in Jack Eaton, Anna Ringblom, Christo Kaye, Anna Behrens, and Matt Pierri. These exemplary human beings have taught me a lot about the world/life, and kept me grounded during the inevitable highs and lows of research. I am grateful for the time spent with all of you - Skål!

Finally, I thank my family for their love, support, and encouragement in my various endeavours - despite still not being sure exactly what it is I do.

List of Publications

Published

1. Domański, J.; **Hedger, G**.; Best, R. B.; Stansfeld, P. J.; Sansom, M. S. P. (2016) Convergence and sampling in determining free energy landscapes for membrane protein association. *Journal of Physical Chemistry B*, 121, 3364-3375.

2. **Hedger, G**.; Rouse, S. L.; Domański, J.; Chavent, M.; Koldsø, H.; Sansom, M. S. P. (2016) Lipid-Loving ANTs: Molecular simulations of cardiolipin interactions and the organisation of adenine nucleotide translocase in model mitochondrial membranes. *Biochemistry*, 55, 6238-6249.

3. **Hedger, G**.; Sansom, M. S. P. (2016) Lipid interaction sites on channels, transporters and receptors: Recent insights from molecular dynamics simulations. *Biochimica et Biophysica Acta - Biomembranes*, 1858, 2390-2400.

4. Byrne, E. F. X.; Sircar, R.; Miller, P. S.; **Hedger, G**.; Luchetti, G.; Nachtergaele, S.; Tully, M. D.; Mydock-McGrane, L.; Covey, D. F.; Rambo, R. P.; Sansom, M. S. P.; Newstead, S.; Rohatgi, R.; Siebold, C. (2016) Structural basis of smoothened regulation by its extracellular domains. *Nature*, 535, 517-522.

5. **Hedger, G**.; Shorthouse, D.; Koldsø, H.; Sansom, M. S. P. (2016) Free energy landscape of lipid interactions with regulatory binding sites on the transmembrane domain of the EGF receptor. *Journal of Physical Chemistry B*, 120, 8154-8163.

6. Shorthouse, D.; **Hedger, G**.; Koldsø, H.; Sansom, M. S. P. (2016) Molecular simulations of glycolipids: Towards mammalian cell membrane models. *Biochimie*, 120, 105-109.

7. **Hedger, G**.; Sansom, M. S. P.; Koldsø, H. (2015) The juxtamembrane regions of human receptor tyrosine kinases exhibit conserved interaction sites with anionic lipids. *Scientific Reports*, 5, 9198.

In preparation

1. Velgy, N.; **Hedger, G**.; Biggin, P. C. (2017) Molecular dynamics simulations of G protein-coupled receptors. (*submitted*).

2. Yen, H.; Liko, I.; **Hedger, G**.; Pluckthun, A.; Sansom, M. S. P.; Tate, C.; Robinson, C. V. (2017) PIP_2 is an allosteric modular of receptors stabilising active states and enhancing G protein coupling. (*in preparation*).

3. Chavent, M.; Karia, D.; Kalli, A. C.; Domanski, J.; Duncan, A. L.; **Hedger, G**.; Stansfeld, P. J.; Seiradake, E.; Jones, E. Y.; Sansom, M. S. P. (2017) Interactions of the kinase domain of EphA2 with PIP_2-containing membranes: Implications for receptor function. (*in preparation*).

4. Rao, S.; Newport, T.; **Hedger, G.**; Stansfeld, P. J. (2017) Molecular simulations of the interactions of bacterial lipoproteins with membranes. (*in preparation*).

List of Presentations

1. **Poster Presentation:**
   Poster A: "Molecular dynamics simulations of G-protein coupled receptor lipid interactions", and Poster B: "Cardiolipin interactions and oligomerisation of ANT1"
   19th IUPAB & 11th EBSA congress, Edinburgh, UK, July 2017.

2. **Poster Presentation:** "Molecular dynamics simulations of G-protein coupled receptor lipid interactions"
   Biophysical Society 61st Annual Meeting, New Orleans, USA, February 2017.

3. **Talk:** "Lipid-protein interactions: insights from molecular simulation"
   Young Modellers Forum, London, UK, November 2016.

4. **Poster Presentation:** "Lipid loving ANTS: Cardiolipin interactions and organization of the adenine nucleotide translocase in model membranes"
   Biomembrane Days, Berlin, Germany, September 2016.

5. **Poster Presentation:** "Free energy landscape for lateral lipid interactions with regulatory binding sites on the transmembrane domain of the EGF receptor"
   American Chemical Society 251st Annual Meeting, San Diego, USA, March 2016.

6. **Poster Presentation:** "Free energy landscape for lateral lipid interactions with regulatory binding sites on the transmembrane domain of the EGF receptor"
   Biophysical Society 60th Annual Meeting, Los Angeles, USA, February 2016.

7. **Talk:** "Receptor tyrosine kinase lipid interactions: Insights from molecular simulation and free energy calculations"
   Molecular Mechanisms in Cell Signalling Symposium, Vienna, Austria, June 2015. (Awarded "Best talk" Prize)

8. **Poster Presentation:** "Local bilayer reorganization by the JM regions of all 58 human RTKs: A multiscale molecular dynamics study"
   Biophysical Society 59th Annual Meeting, Baltimore, USA, February 2015. (Student Research Achievement Award (SRAA))

9. **Poster Presentation**: "Local bilayer reorganization by the JM regions of all 58 human RTKs: A multiscale molecular dynamics study"
   CCPBioSim Annual Conference, Leeds, UK, January 2015.

Contents

List of Figures

List of Tables

List of Abbreviations

ANT	ADP/ATP translocase, also referred to as the ADP/ATP carrier
β_1AR	Beta-1-adrenergic receptor
CG	Coarse grain
CL	Cardiolipin
CRD	Cysteine rich domain
EGFR	Epidermal growth factor receptor
GM3	Monosialodihexosylganglioside
GPCR	G protein-coupled receptor
GROMACS	Groningen machine for chemical simulations
MD	Molecular dynamics
NMR	Nuclear magnetic resonance
NTS1	Neurotensin receptor 1
PBC	Periodic boundary condition
PC	Phosphatidylcholine
PDB	Protein data bank
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
PIP_2	Phosphatidylinositol-4,5-bisphosphate
PIP_3	Phosphatidylinositol-3,4,5-trisphosphate
PME	Particle mesh ewald
PMF	Potential of mean force
PO-	1-palmitoyl-2-oleoyl-
PS	Phosphatidylserine
RDF	Radial distribution function
RMSD	Root mean square deviation
RTK	Receptor tyrosine kinase
SA	Self-assembly

SMD	Steered molecular dynamics
SMO	Smoothened
TM	Transmembrane
UCP2	Uncoupling protein 2
WHAM	Weighted histogram analysis method
WT	Wild type

"Everything that is living can be understood in terms of the jiggling and wiggling of atoms."

- Richard Feynman

1

Introduction

Contents

This chapter is based upon the manuscript:

"Lipid interaction sites on channels, transporters, and receptors: Recent insights from molecular dynamics simulations" George Hedger and Mark S. P. Sansom, Biochimica Biophysica Acta - Biomembranes, 2016, 1858, 2390-2400.

1.1 Overview

Cells are separated from their environment and compartmentalised by membranes. These barriers are composed of a lipid bilayer (with the various lipid species distributed asymmetrically between the two leaflets of the bilayer), into which proteins are embedded and associated. Parallel advances in lipidomics [1] and in the structural biology of membrane proteins [2, 3] over the past decade have revealed some of the complexities of the composition of cell membranes. Thus, the structures of ca. 1500 membrane proteins have been determined (http://blanco.biomol.uci.edu/mpstruc/) and the cellular lipidome is estimated to contain 40,000 lipid species (http://www.lipidmaps.org/data/structure/) [4]. The lipidome of membranes varies according to cell age, metabolic state, stage in cell cycle, organelle, and spatial location; resulting in a complex protein-lipid interactome. In addition to providing a bilayer environment, it is increasingly appreciated that the function of integral proteins can be modulated by interactions with this complex lipid mixture [5-12]. Of particular interest, an emerging feature present within the protein-lipid interactome is that certain lipid molecules can selectively bind to specific sites on integral membrane proteins, and modulate both their structure and their function [13].

As recently reviewed [14], we now possess over 100 structures of membrane proteins containing electron density interpreted as bound lipid molecules. Structural identification of specific lipid binding sites may aid mechanistic understanding of lipid modulation of protein function, such as in the case of Kir2.2 and PIP_2 [15]. Identification of sites of allosteric modulation on proteins is also of interest for the assessment of protein druggability [16]. The majority of membrane protein structures containing bound lipid molecules were solved using x-ray crystallography. In many cases such structures have been obtained from crystals grown in the presence of detergent. It is likely that the lipids observed represent a biased sample of tight binding lipids, and in some cases the molecular identity of the observed electron density corresponding to detergent and/or lipid may be uncertain. This may change as more membrane protein structures are determined using crystals obtained from lipidic cubic phases [17], which better approximate a native membrane environment.

Molecular dynamics simulations allow membrane protein structures to be computationally re-embedded into lipid bilayers, and their dynamic interactions with surrounding lipid molecules to be characterised [18]. A number of recent simulation studies probing lipid interactions have identified specific lipid binding sites. These sites

show good agreement with those identified from a range of structural studies. A number of other, presumably weaker, binding sites can also be resolved. Whilst these weaker sites may not always be observed by x-ray crystallography, there are a number of other biophysical techniques which allow one to probe lipid interactions with membrane proteins, including e.g. fluorescence spectroscopy [19], EPR [20], NMR [21], and mass spectrometry [13]. These techniques provide further points of reference for simulation studies membrane protein/lipid interactions.

Within this chapter, I begin by discussing different modes of lipid modulation and binding to *integral* membranes proteins, before providing a brief overview of biomolecular simulations approaches, and surveying recent simulation studies which identify and characterise lipid interaction sites on integral membrane proteins. This chapter is based on my first author publication [22].

1.1.1 Lipid modulation of membrane protein function

The functions of a range of membrane proteins are known to be modulated by their lipid environment, including potassium channels [6, 7], receptor tyrosine kinases (RTKs) [8], G-protein coupled receptors (GPCRs) [9, 10], solute transporters such as BetP [5] and the ADP/ATP carrier [23], redox proteins such as cytochrome c oxidase [24], and certain P-type ATPases [25]. Such lipid modulation can influence several different aspects of protein function, including effects on the activity of a membrane protein, modulating protein-protein interactions, and altering cellular localisation by sequestering a protein to spatially defined regions of a membrane. In certain cases, a lipid may represent a native ligand for the protein rather than an allosteric modulator, as is the case for the sphingosine-1-phosphate (S1P) and lysophosphatidic acid (LPA) group of lipid-activated GPCRs [26].

In a number of cases structural, biophysical and functional assays have been combined to provide a detailed picture of lipid modulation. For example, this is the case for eukaryotic inward rectifying potassium ion (Kir) channels. Functional assays revealed Kir channels to be dependent on the presence of the anionic lipid phosphatidyl inositol-4,5-bisphosphate (PIP_2) for activation. Subsequently, simulation studies [27, 28] and crystal structures [15] revealed four specific PIP_2 binding sites and enabled the mechanism of PIP_2 channel modulation to be structurally rationalised. In other cases careful biochemical analysis has revealed a functional dependence on certain lipid

species, but the mechanism of modulation remains unclear. This is the case for the epidermal growth factor receptor (EGFR/ErbB1) which is a single-pass transmembrane receptor known to be modulated by an array of lipids, including PIP species and the glycolipid GM3 [8]. However exactly how these lipid species control receptor activity remains unclear, with proposals including an influence of receptor dimerization propensity, direct conformational stabilization and orientation effects, and larger scale lipid-induced clustering of the receptor.

1.1.2 Lipid interaction sites on integral membrane proteins

Lipids interact with membrane proteins *via* multiple modes. The presence of integral membrane proteins may induce formation of a lipid 'annulus' around the protein. Due to interactions with the protein, lipids within this annulus exhibit decreased motional freedom compared to their non-interacting bulk counterparts, and are detectable by EPR [29, 30]. This immobilizing effect of the protein may extend beyond the first shell of directly interacting annular lipids, leading to further outer shells with a lesser extent of lipid immobilisation, as suggested by MD simulations [31, 32]. In addition certain lipid species may bind to specific sites on the membrane protein surface – often described as 'non-annular' lipids. Binding may be driven by formation of physicochemical interactions between the lipid and protein surface, as well as by complementary geometry, for instance 'slotting' of lipid molecules into 'grooves' on the protein surface [33] or binding at the interface between subunits [34]. More recently, a number of studies have suggested lipid molecules may exit the bulk lipid phase of the membrane, and laterally enter into the core of integral membranes proteins. This includes the entry of entire lipid molecules into the orthosteric binding pocket of GPCRs [35, 36], as well as the protrusion of tails into the selectivity filter of ion channels [37]. Binding sites may tightly coordinate the lipid [15], or act to cause weaker and more dynamic localisation [38]. Efforts have been made to describe general features of lipid binding sites and sequence interaction motifs, such as for cholesterol [39] and cardiolipin [40].

1.1.3 Biomolecular simulation approaches for lipid binding site identification

Molecular dynamics (MD) simulations provide a powerful tool to characterise the dynamics and interactions of membrane proteins with surrounding lipid molecules [18].

However, the computational cost of the simulations is such that length scales beyond microseconds are not currently readily accessible [41], especially for extended systems containing multiple membrane proteins. This has prompted the development of more approximate coarse-grained (CG) representations of membrane lipids and proteins in MD simulations [42, 43] in which groups of atoms are represented as single particles (Figure 1.1). Reducing the number of particles in the system reduces the computational demand involved in running the simulation and thus allows access to longer time and length scales, with the caveat that the level of approximation made in a given CG model has to be matched to the underlying biological interactions being probed. CG simulations can thus allow significantly enhanced lipid exploration of the protein surface and candidate binding sites, whilst sacrificing the finer detail of lipid-protein interactions. These approximations may be reconciled to some degree by conversion of the endpoint of a CG system back to atomistic detail [44, 45] and subsequently running an atomistic simulation to assess the validity of the CG system arrangement; so called (serial) multiscale modelling [46]. The MARTINI CG force field has been most widely applied in the field of protein-lipid interactions. It should be noted that CG simulations may now extend to hundreds of microseconds [47], and contain many hundreds of protein molecules [48], whilst atomistic simulations of individual proteins may reach tens of microseconds duration using high performance computing resources [49].

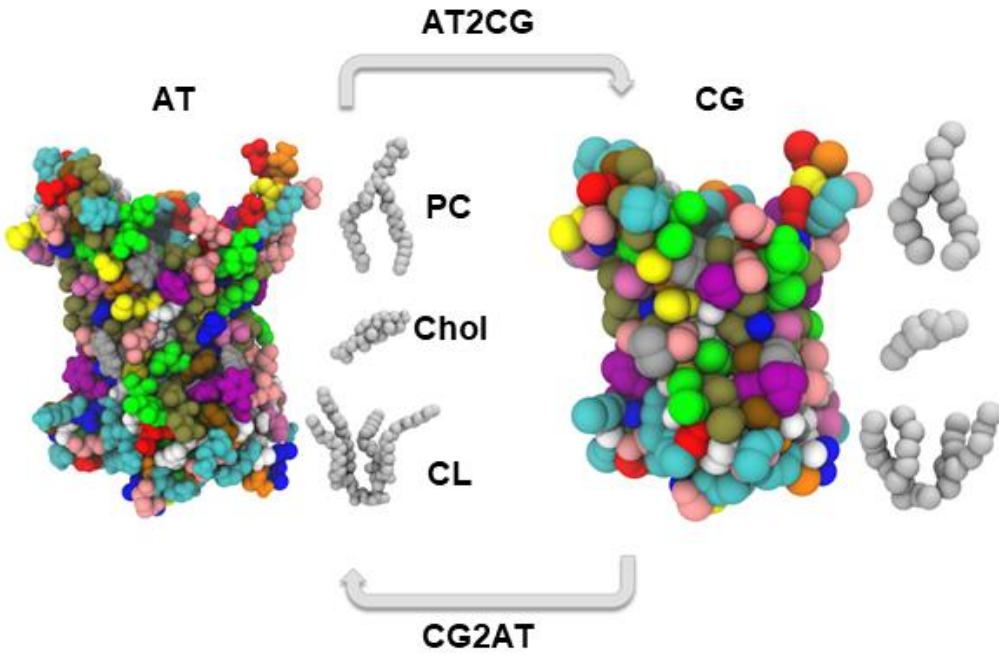

Figure 1. 1: Schematic of a multiscale approach to modelling and simulation of lipid interactions with an integral membrane protein. The ADP/ATP translocase (ANT1/AAC1; PDB: 1OKC [50]) is depicted as spheres colored by residue type, at both atomistic (left) and CG (right) resolution. Phosphatidylcholine (PC), cholesterol (Chol), and Cardiolipin (CL) molecules are shown as grey spheres.

The structure of a membrane protein used as initial input for MD simulations may be from X-ray diffraction, cryoelectron microscopy, or NMR. If the 3D structure of the protein is not known experimentally, in some cases a model may be built by modelling [32]. The membrane protein is then embedded into a lipid bilayer. This may be achieved either by self-assembly simulations [51] in which short simulations are run to allow the spontaneous formation of a bilayer around an integral membrane protein, or by a number of methods which insert a membrane protein into a pre-assembled bilayer [52-54]. Advances in lipid parameterisation [55], along with a growing appreciation of the *in vivo* compositional complexity of lipid membranes [56] is leading to simulations of proteins in complex bilayers containing multiple lipid species. At the simplest level such mixed bilayers may contain two lipid species, while at their most complex they may provide approximations of *in vivo* plasma membrane composition [57, 58]. Such mixed lipid systems allow one to address competition between different lipid species for interaction with a given protein, in addition to providing a better approximation of lipid-lipid interactions which are linked to, and may influence, protein-lipid interactions. Such

mixed lipid systems can now be routinely assembled in CG [57, 58]. Simulations of mixed lipid systems are also becoming more common at the atomistic level, utilising a number of recent tools [59-61], such as the *Membrane Builder* extension of the CHARMM-GUI [59] to facilitate automated construction of complex bilayers in all atom detail. Some of the equilibration and sampling problems arising from timescale limitations are being combated by increases in processing power and algorithm efficiency, as well as the ability to reversibly convert between CG and atomistic levels (Figure 1.1) [44, 45]. Thus in the latter approach the membrane protein of interest may first be simulated in a mixed lipid bilayer using CG methods to equilibrate the system and make an assessment of how lipids interact on extended timescales, before converting the system to atomistic detail to further refine and characterise the observed lipid-protein interactions. This serial multiscale approach has been successfully used to identify lipid binding sites on a number of membrane proteins [27, 28, 47, 62].

Potential lipid binding sites may also be identified by docking calculations, using e.g. AutoDock [63]. However these methods do not generally take into account the membrane environment in which the interactions occur. Thus protein/lipid configurations identified by docking may require refinement by subsequent molecular dynamics simulation. Combined use of molecular docking and simulation has enabled identification of lipid binding sites on several types of membrane protein, including e.g. cholesterol interactions with Cys-loop receptors [64, 65] and with Kir channels [66].

1.2 Characterization of lipid interaction sites: recent simulation studies

1.2.1 Channels

Lipids modulate the function of a number of channels. For example, the canonical bacterial potassium channel KcsA requires bound anionic lipid molecules for full activity, and both crystallographic [67] and simulation studies [68] have previously indicated the presence of a binding site for an anionic lipid molecule located between two arginine sidechains at the subunit/subunit interface of the trimeric channel protein. More recent studies have focussed on two species of lipid thought to regulate a number of channels in mammalian cell membranes, namely PIP_2 and cholesterol.

Inward rectifying potassium (Kir) channels

Kir channels are tetrameric integral membrane proteins controlling the selective permeation of K^+ ions across cell membranes. They have critical involvement in molecular processes ranging from control of the resting membrane potential to regulation of insulin secretion in pancreatic β cells [69]. Of particular interest Kir channels have a requirement for PIP_2 for maximal activation [6, 70] which may be considered as lipid agonism of these channels [71].

The interaction of PIP_2 with Kir channels has been explored using combined CG and atomistic simulations [27]. Three structures of Kir channels were simulated: KirBac1.1, a Kir3.1-KirBac1.3 chimera, and a homology model of Kir6.2. These structures were initially converted to CG representation and embedded into POPC bilayers, each of which contained 4 PIP_2 lipid molecules within the inner leaflet of the bilayer. Analysis of time-averaged density maps for PIP_2 revealed preferential localisation to a single site, present on all four subunits. The same site was observed for all three Kir channels. The binding site was formed by a cluster of basic and amphipathic aromatic residues at the interface between subunits. Once bound, PIP_2 lipids did not dissociate from the identified sites even during extended (5 μs) CG simulations, suggesting tight binding. Conversion of CG snapshots of the channel with bound PIP_2 lipids to atomistic models followed by short simulations enabled assessment of binding site interactions. Key residues seen to form contacts with the PIP_2 lipid head group in the simulations had previously been implicated in PIP_2 interactions with Kir2.1 [6, 72, 73]. This was not the case for residues seen in the simulations to form contacts with the lipid tail regions, which formed more transient interactions with the protein and were therefore not suggested to be major binding determinants.

Subsequent to this study, a high resolution X-ray structure of Kir2.2 in complex with four short chain (dioctanoyl) PIP_2 lipids was determined [15], revealing binding sites and interactions in good agreement with the simulation-based predictions [27]. This is despite the caveat that the structures used for simulations employed homology models and eukaryotic-prokaryotic chimeras. Further studies [28] extended the comparison by multiscale simulations of the PIP_2 bound structure (PDB: 3SPI[15]), as well as other Kir2.2 X-ray structures including the apo state (PDB: 3JYC[74]), and a phosphatidic acid bound state (PDB: 3SPC[15]). This study provided a 'like for like' comparison of PIP_2 binding to Kir2.2, supporting the sites initially identified *via* simulations of other Kir

subtypes. These multiscale simulations showed unambiguously that PIP_2 lipids bound at the same four sites identified in the crystal structure (Figure 1.2). The good agreement between multiple simulation and experimental approaches is suggestive of the accuracy of the binding site predictions, as well as the utility of multiscale simulations in their identification.

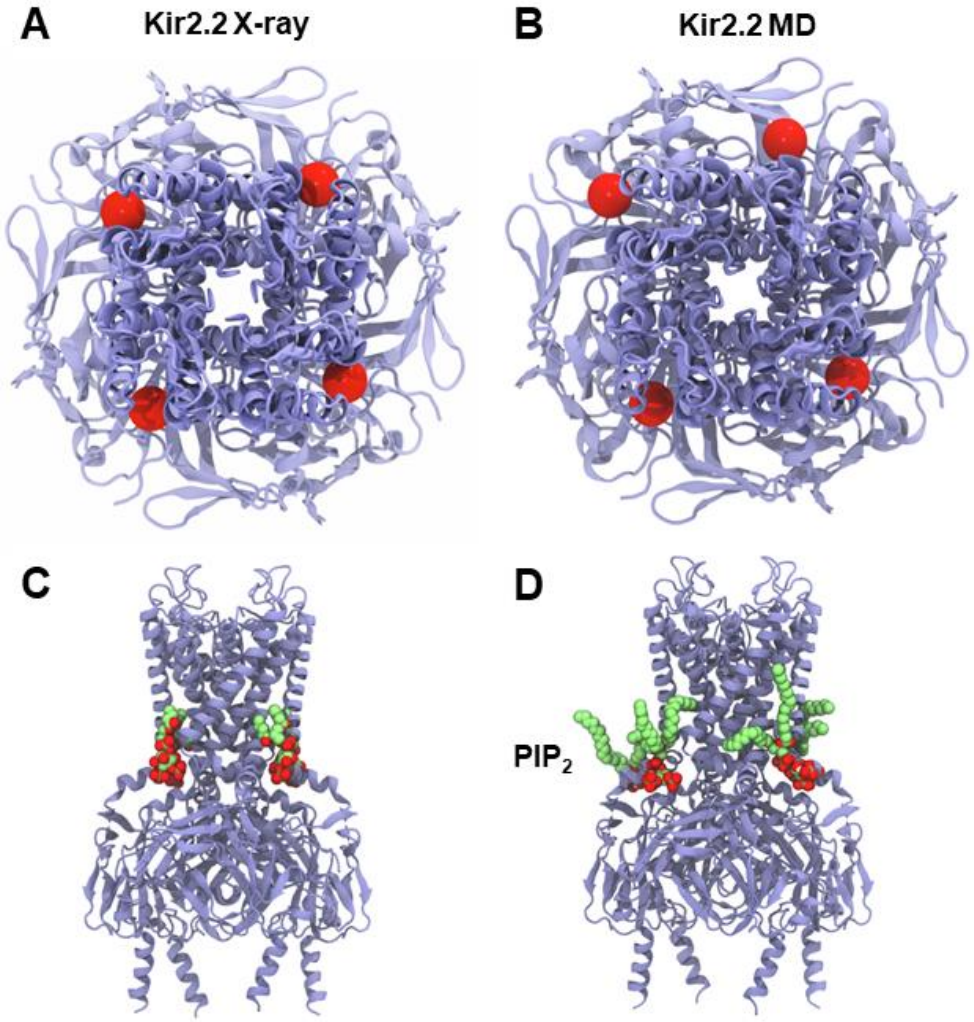

Figure 1. 2: Similarity in PIP_2 binding sites identified by crystallography [75] (A, C) and by simulation [76, 77] (B, D). View (A, B) from the extracellular side down the pore axis of the tetrameric Kir2.2 channel with the four PIP_2 α-phosphate groups (bridging the glycerol and inositol moieties) indicated as red spheres. (C, D) Side view of the channel showing bound PIP_2 molecules (lime green carbons and red oxygens).

Kir channels have also been shown to be regulated by cholesterol [78, 79], in a stereospecific fashion suggestive of the existence of a specific cholesterol binding site on

the channel proteins [7] [80]. An integrated molecular docking and simulation approach has been used to investigate cholesterol binding to Kir2.1 [81], leading to the identification of two cholesterol binding sites. The validity of the sites was tested by mutagenesis and electrophysiology, and showed good agreement with the MD predictions. Docking and simulation studies were also used to compare predicted cholesterol interactions of bacterial (KirBac) and mammalian Kir channels [79]. Overall, these two studies suggest a consensus cholesterol binding region in the cytoplasmic half of the TM domains of Kir channels with the ligand also interacting with residues on the slide helix located at the membrane surface. For a recent comprehensive review of cholesterol interactions with membrane proteins as studied by molecular dynamics see [82].

Pentameric ligand gated ion-channels

The pentameric ligand gated ion-channels (pLGICs) are a major class of neurotransmitter receptor [83]. Both the nicotinic acetylcholine receptor (nAChR) and the y-aminobutyric acid receptor ($GABA_AR$) are sensitive to membrane cholesterol [11, 12], and the nAChR also shows functional sensitivity to anionic lipids [84] and membrane hydrophobic thickness [85]. The interactions of cholesterol with the nAChR [64, 86] and the $GABA_AR$ [65] have both been explored using docking and simulations. In a study by Brannigan et al. [87], docking calculations were performed to screen the nAChR *Torpedo* cryo-EM structure (PDB: 2BG9) for possible cholesterol binding sites [64]. A total of three sites per subunit were identified (15 for the whole protein). These sites were buried both within grooves on the protein surface, and more deeply within non-membrane exposed pockets. A later study [88] identified the presence of cholesterol interaction motifs within the transmembrane domain of nAChR, which co-localised with a number of the sites suggested by Brannigan et al., as well as a number of additional solvent exposed sites. Interestingly the sites identified in these computational studies corresponded to gaps in the original cryo-EM structure (PDB: 2BG9). The possible role of cholesterol at the identified sites was addressed by 25 ns atomistic simulations both in the presence and the absence of bound cholesterol [87]. In the absence of cholesterol the binding sites were initially filled with water. During the simulation water was expelled and the structure was seen to collapse. In contrast in the presence of bound cholesterol the original

conformation was more faithfully maintained, and contacts between the pore and agonist binding domain proposed to be critical for gating were seen to be maintained. The results overall suggested 15 cholesterol molecules bind to the pentameric nAChR structure and act to provide structural integrity to the receptor.

Interactions of nAChR with phosphatidic acid (PA) have also been explored by simulations [86], which suggested potential differences in the interaction modes of anionic PA and zwitterionic lipids. However the rather short duration (10 ns) of the simulations is likely to preclude further insight.

In a second study, cholesterol interactions with the related $GABA_AR$ were explored [65]. In lieu of a high resolution structure of $GABA_AR$, the authors built a homology model based on GluCl (PDB: 3RHW [89]) and used this structure for initial protein coordinates within simulations. Five equivalent cholesterol binding pockets positioned between subunits of the pentameric receptor were identified by docking (Figure 1.3). Atomistic simulations were performed starting from a configuration with cholesterol docked into these sites. Cholesterol was seen to undergo some reorientation within all five sites. During a 200 ns simulation the cholesterol molecules at two sites were seen to dissociate within 20 ns, but were subsequently seen to rebind at around 80-110 ns with similar but presumably slightly more favourable binding modes. The remaining three cholesterol molecules remained bound. The general stability of cholesterol and the rebinding events observed during the simulations are supportive of the validity of the sites suggested from docking. A second model of cholesterol bound to $GABA_AR$ was built based on the ivermectin binding sites in the GluCl (ivermectin is a relatively hydrophobic ligand) crystal structure [89]. The binding pockets in this model were identical to those seen from docking and spontaneous rebinding events during atomistic simulation, though some differences in cholesterol orientation were observed. Thus the different approaches seemed to converge in terms of cholesterol interactions. Comparison of the behaviour of the protein in the cholesterol-bound and the apo state simulations suggested that in the absence of cholesterol the pore radius of the channel decreased, whereas with bound cholesterol the channel showed an increased tendency to adopt a more open conformation. This suggests that cholesterol may in part exert its channel modulating properties *via* direct binding and stabilisation of an open conformation of the $GABA_AR$.

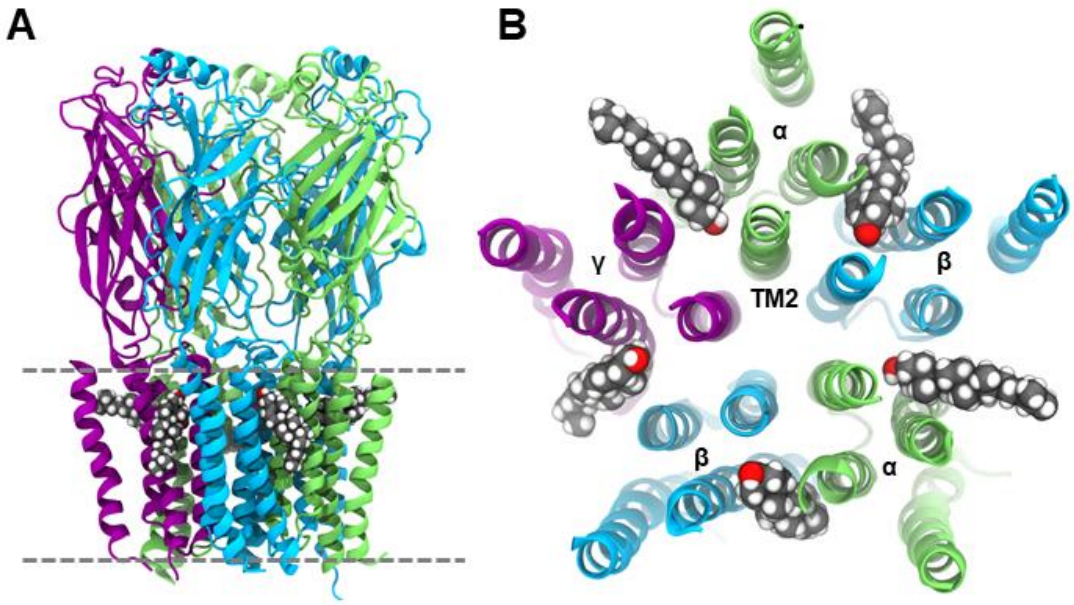

Figure 1. 3: Location of cholesterol binding sites between the five subunits of the pentameric $GABA_AR$ [90]. Initial cholesterol coordinates were predicted *via* docking, and the structure shown is post-simulation. (A) Sideview and (B) extracellular view of the five transmembrane domains, with cholesterol molecules represented as van der Waals spheres. Dr Jérôme Hénin and Dr Grace Brannigan are thanked for providing coordinates used to generate the figure.

Mechanosensitive channels

Mechanosensitive channels open in response to an increase in tension (i.e. stretching) of the membrane. A recent combined structural, biophysical and computational study has revealed the role of bound lipid molecules in mechanosensing by the *E. coli* MscS channel [91]. X-ray diffraction and simulation together identified lipid tails bound inside pockets formed by the TM helices. Significantly the number of lipid acyl chains occupying the pockets decreased upon channel opening (Figure 1.4). This suggested an activation mechanism whereby an increase in membrane tension perturbed lipid binding within the pocket, thus destabilizing the closed state and thus promoting channel opening.

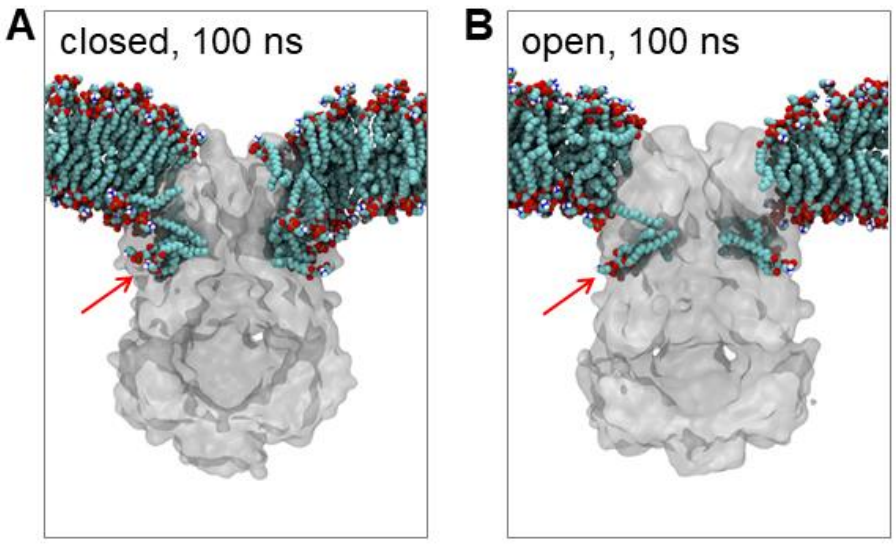

Figure 1. 4: MscS/lipid interactions comparing closed (PDB: 2OAU) and open (PDB: 5AJI) conformations. Cut-away slices are shown of snapshots at 100 ns from atomistic simulations of the closed (A) and open (B) conformations of MscS in POPE/POPG (4:1) bilayers [91]. The proteins are shown in grey and the lipids in cyan (carbons) and red (oxygens). The red arrow shows the lipid binding cavity whose occupancy changes between the closed and the open conformation. Figure modified from [91].

1.2.2 G protein-coupled receptors

G protein coupled receptors (GPCRs) are a large superfamily of integral membrane proteins composed of seven transmembrane helices [92]. The complexities of the interactions of GPCRs with cholesterol and other lipids are starting to be unravelled [93]. Cholesterol has been suggested to modulate various aspects of GPCR biology including stability [9, 10], oligomeric organisation [94], and ligand binding activity [95, 96]. Cholesterol molecules have been found co-crystallised in a number of GPCR X-ray structures [97-100], suggesting possible modulation by direct binding. Molecular simulation has been applied to identify lipid binding sites on a range of GPCRs including the β_2-adrenergic receptor (β_2AR) [101-103], the β_1-adrenergic receptor (β_1AR) [104], the Serotonin$_{1A}$ [105] and Serotonin$_{2A}$ receptors [106], the A_{2A} adenosine receptor [107], Rhodopsin [108, 109], the Cannabinoid 2 (CB2) receptor [110], and the S1P1 receptor [48].

Cholesterol binding sites have been identified on the A_{2A} adenosine receptor by long timescale all-atom MD simulations [107] and by X-ray crystallography [100]. In a study by Lee and Lyman [107], two independent 800 ns all-atom simulations of the adenosine-bound structure (PDB: 2YDO) were performed in a PC:cholesterol lipid bilayer. Over the

course of the simulations cholesterol molecules bound to three distinct sites on the receptor. Once bound, a single cholesterol molecule occupied each site for the duration of the simulation. However despite these sites remaining occupied, the binding modes of cholesterol were seen to be dynamic. The simulation data were used to estimate cholesterol interaction energies at each site *via* an inverse Boltzmann approach, yielding energies of the order of kT, suggesting cholesterol is only weakly associated with the receptor. Concomitant to this study an X-ray structure of the A_{2A} adenosine receptor was published with three co-crystallised cholesterol molecules (Figure 1.5) [100]. Both the simulation and crystallographic studies [100, 107] agreed on site II. The differences in sites I and III between the crystal structure and simulations may reflect the relatively weak and dynamic nature of the cholesterol interactions, as well as sampling constraints. Interestingly, a study later study in 2017 [35] employing all-atom MD was able to reproduce the earlier findings of Lee & Lyman on longer timescales. Additionally, the authors observed spontaneous entry of a cholesterol molecule from the membrane-phase into the orthosteric ligand binding pocket of $A_{2A}R$. Entry was seen to occurr *via* a pathway extending from the bulk membrane into the ligand bind pocket *via* an opening between helices V and VI on the extracellular side. Thus opening the possibility that cholesterol may in part exert modulatory effects on $A_{2A}R$ *via* direct competition with the native ligand for occupancy of the receptor binding site, in addition to allosteric effects.

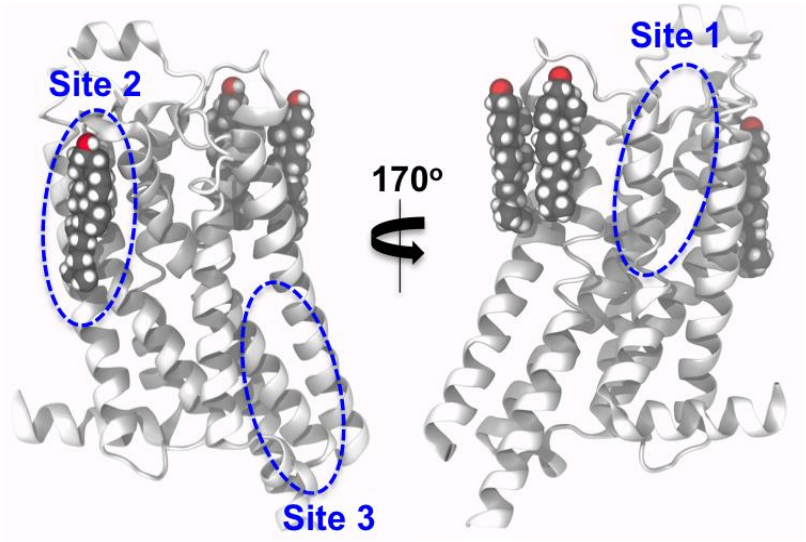

Figure 1. 5: Cholesterol interaction sites on the A_{2A} adenosine receptor. The location of cholesterol molecules (van der Waals spheres) resolved in a high resolution crystal structure of the $A_{2A}R$ (PDB: 4EIY) is shown [100]. The approximate positions of cholesterol binding sites suggested by molecular simulation [111] are encircled in blue.

CG simulations provide a means to significantly extend sampling of cholesterol/GPCR interactions. Thus, Sengupta et al. explored cholesterol binding to the Serotonin$_{1A}$ receptor [105] using CG simulations of up to 80 μs duration with the GPCR embedded in a PC:cholesterol bilayer. A number of sites were identified within both leaflets. On the basis of their lipid contact dynamics these sites were predicted to have a relatively weak propensity to bind cholesterol. Interestingly one site co-localised with a cholesterol consensus motif (CCM), and another with a cholesterol recognition amino acid consensus (CRAC) motif. However the significance of such motifs in general remains uncertain. These sites showed differences with those seen for the closely related Serotonin$_{2A}$ receptor [106], raising the question of how to best evaluate and validate simulation-based predictions.

A subsequent study utilized comparable CG simulations to investigate cholesterol interactions with the β_2AR [102]. These simulations were performed in bilayers containing two β_2AR molecules, allowing an exploration of cholesterol influence on GPCR dimerization. In the absence of cholesterol, protein dimerization was observed *via* an interface involving TM4 and TM5 from each monomer. Titration of increasing amounts of cholesterol into the lipid bilayer and subsequent simulation led to dissolution of this interface and emergence of a new interface involving helices TM1 and TM2 from each monomer. The TM4+TM5 and TM1+TM2 interfaces for β_2AR identified by CG simulation are reminiscent of those seen in atomistic simulations of β_1AR [104]. Computation of cholesterol density maps and cholesterol contact dynamics led to identification of three binding sites on the β_2AR in the outer leaflet, and four within the inner leaflet [112]. One of these sites was formed on TM4 where it formed interactions with residues of a CCM motif. The binding modes at this site were dynamic, as seen for other GPCRs [105, 107]. Notably cholesterol was seen at this position in a crystal structure of β_2AR [98]. In the cholesterol-containing bilayers this binding site became occupied by the sterol, leading to disruption of the involvement of TM4 in the TM4+TM5 interface. These results suggest that cholesterol may modulate dimerization of the β_2AR *via* direct competition between lipid-protein and protein-protein interactions. This provides an example of the use of molecular simulations to explore effects of lipid binding on protein oligomerization and organization within membranes. These aspects have also been explored in large scale simulations of rhodopsin embedded in single lipid-species bilayer containing tens of protein molecules [113-115].

Recent large simulations of the S1P1 receptor in a plasma membrane model have provided further insights into the possible influence of lipid interactions on GPCR oligomerisation [48]. CG simulations of a 140 x 140 nm^2 patch of membrane (Figure 1.6) containing 144 copies of the S1P1 receptor and with a lipid composition mimicking that of a mammalian plasma membrane (extracellular leaflet PC:PE:Sph:GM3:Cholesterol = 40:10:15:10:25; intracellular PC:PE:PS:PIP_2:Cholesterol = 10:40:15:10:25) were run for 10 μs. Analysis of the simulations revealed transient formation of S1P1 dimers, trimers, and higher order oligomers. Interactions of cholesterol and of PIP_2 with the GPCR were observed (Figure 1.6). Detailed examination of the S1P1 dimers observed in the simulation suggested that cholesterol may help to mediate the protein-protein interactions.

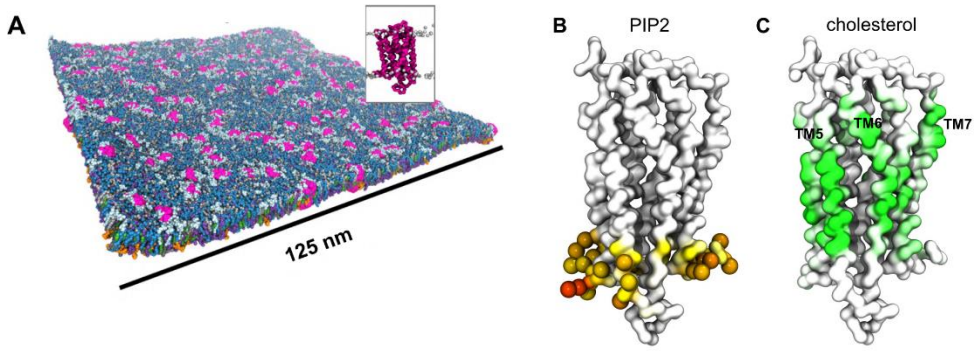

Figure 1. 6: Lipid interactions of the S1P1 receptor. (A) 144 S1P1 GPCR molecules (pink; see inset figure) embedded in a bilayer, the composition of which corresponds to that of a mammalian cell membrane. The image shown is from the end of a 10 μs CG-MD simulation [48]. (B) Model of the S1P1 receptor colored according to the level of PIP2 phosphoryl head group interaction, from white (no interaction) to orange (high interaction). Residues with high levels of PIP_2 interaction (>75% of simulation time) are shown as spheres. (C) The S1P1 receptor colored according to the degree of cholesterol interaction from white (no interaction) to green (high interaction). This figure was adapted from [48].

Simulations have also been used to explore how anionic phospholipids may modulate β_2AR function [101]. From an extensive (0.25 ms) set of all-atom simulations it was possible to observe an anionic lipid, phosphatidylglycerol (PG), entering the core of the activated β_2AR laterally *via* an opening between the cytoplasmic portions of helices TM6 and TM7 (Figure 1.7). Once bound the PG molecule formed electrostatic interactions with the protein which inhibited the formation of the ionic lock, a key interaction thought

to stabilise the inactive state of the receptor. Entry of the PG lipid thus led to an increase in stability of the active state of the transmembrane domain, providing a testable mechanism which may explain the experimental observation that anionic lipids can enhance the activity of certain GPCRs [116] including the β_2AR [117].

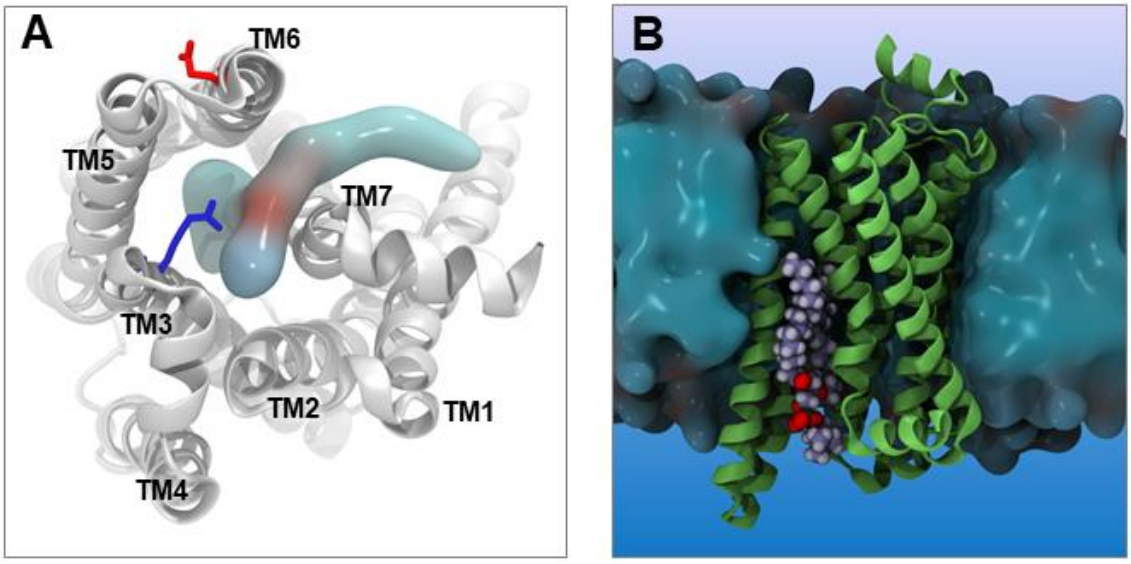

Figure 1. 7: PG/PC binding site within an activated state of the β_2AR [36]. (A) Penetration of a PC lipid into the core of the β_2AR. The residues R3.50 (blue) and E6.30 (red) of the ionic lock are indicated as sticks. (B) Side view of the receptor embedded in a lipid bilayer (blue surface) with a bound PC lipid (spheres, oxygens in red). Dr Chris Neale and Professor Angel Garcia are thanked for providing the coordinates used to produce the figure.

A number of other examples of identification of lipid binding sites on GPCRs are available [103, 104, 108, 109, 118]. These include long timescale atomistic simulations of the β_1AR[104] and CB2 receptors [110], and multiscale simulations of rhodopsin [108, 109]; as well as a recent study on the A2A adenosine and dopamine D2 receptors focussing on the structural effects of altered lipid composition on receptor oligomerisation within in disease state membranes [119]. For a recent detailed review of the application of molecular simulations to explore GPCR-cholesterol interactions see [38].

1.2.3 Receptor tyrosine kinases

A second major superfamily of membrane receptors are those which have a single membrane spanning helix, and which form functional dimers in cell membranes. These include the receptor tyrosine kinases (RTKs) [120] and also a number of other families including e.g. integrins [121]. These receptors are characterised by extensive extracellular domains, single pass transmembrane helices, and intracellular domains often incorporating kinase activity. Many crystal structures are known for extracellular and intracellular domains, but a high resolution structure of an intact RTK remains elusive. A number of NMR structures of TM helix dimers from RTKs are known [122-126].

Experimental studies have shown that lipid interactions can modulate RTK activation, especially of the EGFR [8, 127]. There have been a number of simulation studies of: (i) models of the intact "near full-length" EGFR receptor in mixed lipid bilayers [128, 129]; and of (ii) the transmembrane helix of the EGFR and its dimers [32, 130, 131]. There has also been a recent study combining experiments and simulations to explore lipid interactions of the EphA2 receptor transmembrane region and ectodomain, and their influence on the conformation and orientation of the ectodomain relative to the membranes [132].

As part of my prior Masters project I previously undertook a systematic comparative study of lipid interactions with models of the transmembrane helix plus juxtamembrane region (TM+JM) of all 58 human RTKs [32]. This revealed conservation of the interactions of the JM region with PIP_2 (Figure 1.8).

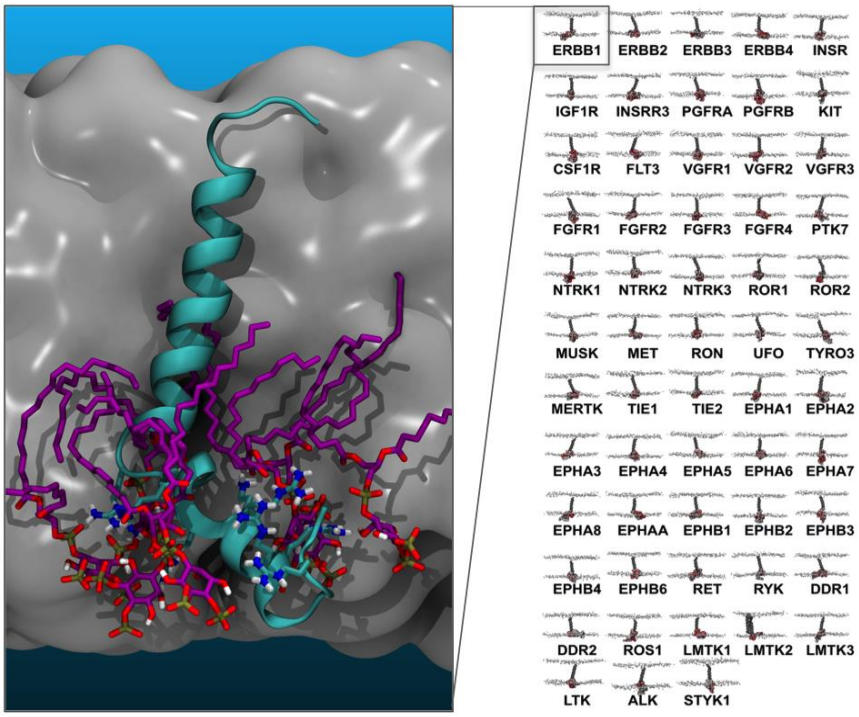

Figure 1. 8: Conserved interaction of PIP_2 lipids with TM-JM models of the human RTK superfamily [32]. Left: TM-JM model of the EGFR monomer (cyan) surrounded by a cluster of PIP_2 lipids (purple). Right: The 58 known members of the human RTK superfamily.

Furthermore, these conserved interactions were seen to induce local bilayer reorganisation and anionic lipid clustering (Figure 1.9) seen previously for the TM+JM regions of the gp130 cytokine receptor [31]. This behaviour is likely to extend to other single transmembrane domain proteins besides RTKs and related receptors. However, it is of especial interest for RTKs as PIP_2 lipids have been suggested to modulate the activity and cellular distribution of certain members of the family [8, 133, 134] by interactions with the TM-JM domain.

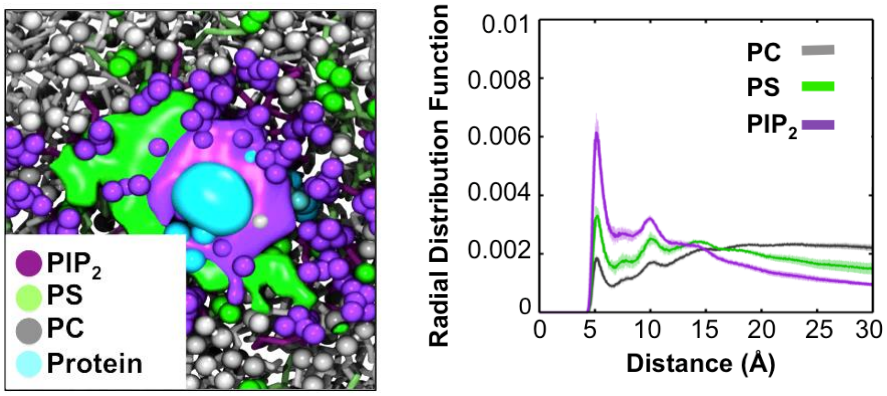

Figure 1. 9: Clustering of anionic lipids around the monomeric TM-JM domain of the insulin receptor. The image shows the average spatial occupancy of PIP_2 (purple), phosphatidylserine (green), and phosphatidylcholine (grey) lipids over the inner leaflet surface, with the JM region density is indicated in cyan (left); and the radial distribution function for each lipid species relative to the protein (right). Figure modified from [32].

This comparative simulation study demonstrated that simulations can now be used to compare protein/lipid interactions across whole families of membrane proteins (as has also proved possible for e.g. aquaporins [135]). Conservation of such interactions across a family of proteins is an indicator of their likely biological importance.

1.2.4 Transporter proteins

Lipids have been shown to play a critical role in the function of a number of solute transporters, and structural studies have revealed lipid binding sites on several transporter proteins [5]. Simulation based approaches have made a recent valuable contribution to identification and characterisation of sites on transporters including the dimeric form of LeuT [136], UraA H^+-Uracil symporter [137], LacY [138], SERCA: the canonical P-type ATPase [139], and the dopamine [140] and serotonin [141] transporters.

UraA H^+-Uracil symporter

The *Escherichia coli* UraA H^+-Uracil symporter is a member of the nucleobase-ascorbate transporter (NAT) family of transporters [142] which use proton and Na^+ gradients to enable uptake of xanthine, uric acid, and uracil, as well as vitamin C in mammals [142, 143]. A recent MD study explored lipid interactions of the bacterial UraA transporter

(PDB: 3QE7) in a bilayer, [62] the lipid composition of which mimicked that of the bacterial inner membrane, containing PE (75 mol %), PG (20 mol %) and CL (5 mol %). CG simulations of up to 10 μs duration revealed an enrichment of the anionic lipids PG and CL within the lipid annulus. In particular, CL exhibited preferential interactions with the transporter at three sites. A periplasmic site was seen to be positioned in the vicinity of the transport pathway. Given CL may act as a proton donor and acceptor at physiological pH, this suggests a putative proton-donating role for this cardiolipin site. Analysis of protein-lipid contact patterns during the simulation showed cardiolipin molecules at each site interacted predominantly *via* electrostatic interactions between negatively charged cardiolipin headgroups and basic sidechains, whilst in common with observations from other simulation studies [27, 47] the tail groups exhibited more dynamic interactions. The possible importance of interactions seen in identified lipid binding sites may be tested *via in silico* mutagenesis of binding site residues. Using this approach sidechains critical for cardiolipin binding were identified [62]. No lipid molecules have been seen in X-ray structures of UraA, which was crystallised in the presence of a detergent (n-nonyl-β-D-glucopyranoside), one molecule of which was observed bound in the crystal structure in the centre of the protein on the presumed transport pathway of the solute. This suggests possible weak association of cardiolipin with UraA, which may dissociate during detergent treatment.

Ca^{2+} ATPase: SERCA

SERCA is a canonical P-type ATPase which enables ATP-driven reuptake of Ca^{2+} ions into the endo or sarcoplasmic reticulum during calcium signalling. Enrichment of cholesterol in ER membranes of macrophages has been shown to inhibit SERCA2b [25], and lipid-like drugs such as thapsigargin modulate SERCA. Further, phospholipids and other hydrophobic compounds have been observed in x-ray crystal structures [144, 145]. In a study by Autzen et al., thirty independent CG simulations of SERCA embedded in a PC:cholesterol bilayer (molar ratio 10:1) were conducted to address possible cholesterol binding and modulation of SERCA [146]. Mapping the frequency of cholesterol contacts for each residue of the protein onto the 3D structure revealed two major cholesterol binding pockets (Figure 1.9), one of which co-localized to a position known to bind sarcolipin, a small single TM helix protein which modulates SERCA activity. However,

from the location of cholesterol within these sites and comparison of its exchange dynamics to bulk phospholipids it has proved difficult to rationalise how cholesterol binding may alter SERCA activity. Consideration of shared chemical features of thapsigargin and cholesterol lead to an expectation that cholesterol might bind in the thapsigargin binding pocket, but this was not observed in the simulations, in which the pocket instead was occupied by a phospholipid. A series of 100 ns all-atom simulations initiated from cholesterol manually positioned in the thapsigargin binding pocket (based on thapsigargin bound X-ray structure, PDB: 2C8K) did *not* result in a single stable configuration over the simulated time course, and instead exhibited a very dynamic cholesterol pose. This suggests that simulations may be used to rule out possible binding modes for lipids, although further experimental and computational studies will be needed to explore this in more detail.

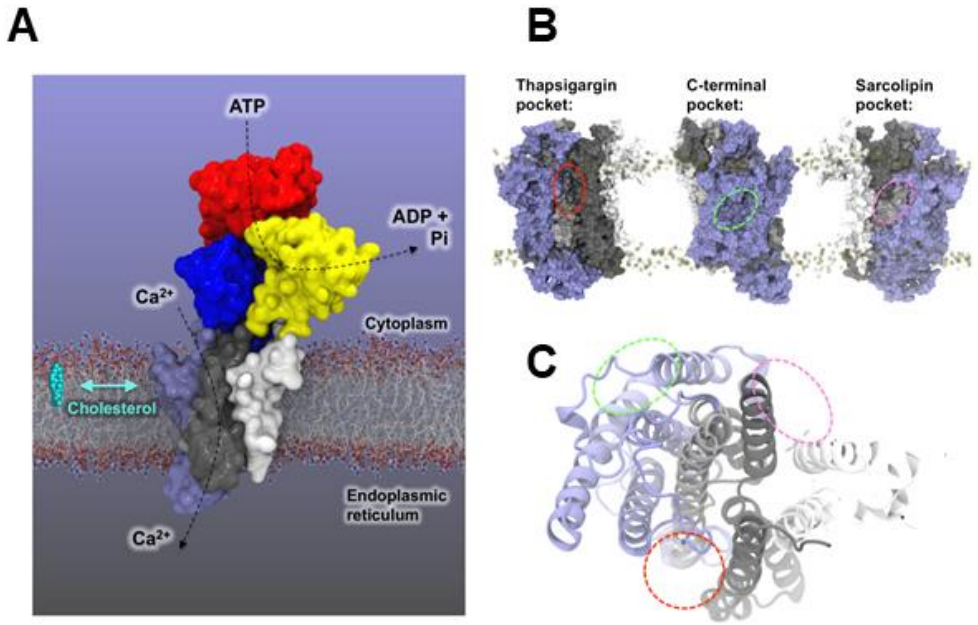

Figure 1. 10: SERCA architecture and putative cholesterol binding pockets. (A) Architecture and domain organization of SERCA in the E2 state (a simulated structure [146] derived from PDB 2C8K is shown). (B) Approximate locations of the thapsigargin (red line), C-terminal (green line), and sarcolipin (pink line) cholesterol binding pockets on the SERCA transmembrane domain (TM1-2: white, TM3-4: black, TM5-6: grey, TM7-10: blue). Dr Henriette Autzen is thanked for providing the coordinates used to produce this figure.

1.2.5 Other membrane proteins

A number of simulation studies not described in detail in this introduction have also addressed lipid interaction sites of membrane proteins. These include phospholipid binding to LacY [138], Kv channels [147], aquaporins [135, 148], the fungal lipid scrambalase TMEM [149], and the dopamine transporter [140]; cardiolipin binding to respiratory chain complexes [47, 150, 151]; and cholesterol binding to rhodopsin [108, 109], VDAC [152], β_1AR [104] and β_2AR [103].

1.3 Overall aims and motivation

This thesis describes the application of classical MD simulations to probe lipid interactions with a range of biomedically important integral membrane proteins. Much of this work has recently become possible due to complementary advances in a) experimental approaches including the lipidomics of biological membrane composition [1], the structural biology of membrane protein – lipid complexes [14], and the development of new biophysical approaches to study these interactions [13]; and b) in computational methods, particularly the continued increase in the number and quality of parameters available for lipid molecules in a range of force fields [22, 153].

In Chapter 3, free energy landscapes are computed for lateral lipid interaction with an EGFR transmembrane helix dimer. This model system is used to compare the selectivity of the EGFR toward a range of different biological lipid species, and the effects of protein and lipid mutation. The data is compared to prior mutagenesis and functional assays. The results suggest that a commonly employed computational approach, potential of mean force (PMF) calculations *via* umbrella sampling, may be adapted and applied to probe lipid-protein interactions.

In Chapter 4, a mixture of CG and atomistic equilibrium simulations are applied to probe cardiolipin (CDL) interactions with a key mitochondrial protein, the ADP/ATP Translocase. Three lipid binding sites are identified, in agreement with prior crystallographic structures and inferences made from NMR measurements [154]. Building on the work in Chapter 3, free energy calculations are employed to calculate lipid-protein PMF profiles, demonstrating that these sites are selective for CDL over other mitochondrial phospholipids. Simulation of a large patch of membrane containing 25 copies of the translocase suggests the capacity for some self-oligomerisation, with CDL molecules localising at the protein-protein interaction interfaces.

Chapter 5 employs an extensive set of CG simulations to prospectively predict the location of PIP_2 interaction sites on the membrane-exposed surface of a class A GPCR, the Neurotensin Receptor 1 (NTS1), a protein known to bind these lipids in native mass-spectrometry measurements. This work employs simulations of single proteins in single two-component lipid mixtures ranging in duration from 5 to 100 μs, as well as large-scale simulations of membranes containing 144 NTS1 molecules under a range of multi-component membrane compositions. This work was carried out in the context of a collaboration with the experimental group of Professor Dame Carol Robinson in the Department of Chemistry, University of Oxford.

Chapter 6 describes a shorter project in which atomistic simulations were applied to probe the influence of a crystallographically resolved cholesterol molecule on the dynamics of a class F GPCR, Smoothened. This work was conducted in conjunction with Professor Christian Siebold at the Wellcome Trust Centre for Human Genetics, University of Oxford, whose research group solved the Smoothened crystal structure simulated.

"*Essentially, all models are wrong, but some are useful.*"

- George E. P. Box

2

Theory and Methods

Contents

This chapter describes the theory behind key computational techniques employed within the thesis, including discussion of atomistic and coarse-grained (CG) molecular dynamics (MD), and enhanced sampling algorithms. An exhaustive description of these techniques

is beyond the scope of this thesis, with a number of excellent textbooks having been devoted to the subject [155, 156]. Rather, the purpose of this chapter is to provide a brief, yet high-level, overview of the key concepts behind these approaches.

2.1 Molecular dynamics simulations

MD simulations provide a powerful addition to the arsenal of tools employed to characterise biological systems. First developed in the 1950's [157, 158], the approach has found wide ranging application from materials science and engineering, to chemistry and biology. The first reported simulation of a protein molecule came in the 1970's in a landmark paper from Andrew J McCammon and Martin Karplus [159]. Years of theoretical development followed, accompanied by ever increasing hardware capabilities, leading to the 2013 Nobel Prize in Chemistry being awarded to Levitt, Karplus, and Warshel "for the development of multiscale models for complex chemical systems" [160].

Molecular simulations as they exist now are capable of simulating biological systems at a range of resolutions, scales, and time durations. An exact and fundamental description of molecular systems may be obtained through quantum mechanical treatment using the Schrodinger equation. Such quantum simulations have found application in understanding detailed chemical processes such as the behaviour of substrates within enzyme active sites [161]. However these approaches, and even hybrid quantum simulations [162], are very computationally expensive. Many interesting biological phenomena occur on longer time and length scales, currently intractable to quantum simulation. As such, biomolecular simulations are more commonly conducted using classical Newtonian physics, in which atoms are treated as individual particles without explicit electrons (i.e. the Born-Oppenheimer approximation). Within these systems the instantaneous interactions between particles are described by a potential energy function, to which simple molecular mechanics (MM) may be applied to calculate forces, then acceleration, and finally new particle coordinates after a small timestep, δt (typically equating to a few quadrillionths of a second). In this way particle coordinates may propagate over time, and when iterated several million times lead to dynamical motions from which observations may be made and biological predictions drawn. I begin by detailing the functional forms of the potentials describing particle interactions. For intuitive introductory purposes I discuss the principles of molecular dynamics in terms of

atomistic resolution, however it is important to realise that the same principles are extensible to CG models (Section 2.4).

2.2 Force fields

The term 'force field' refers to the functional form of the potential energy function governing particle interactions, and the set of parameters associated with it. This function can generally be broken down into bonded and non-bonded components:

$$U(r^n) = \sum U_{bonded} + \sum U_{non-bonded}$$

Where U is the potential energy of *n* particles, with positions given by r^n.

These components may be further subdivided into the energy terms shown in Figure 2.1. In order to reproduce realistic dynamics the force constants, bond lengths, and partial charges within these energy terms must be parameterised to fit experimental observables and theoretical calculations. Commonly this involves comparing to NMR and EPR spectroscopic measurements, crystallographic data, and quantum mechanical calculations. Many force fields have been developed [163-166], each with slightly different parameterisation protocols but ultimately similar underlying functional forms.

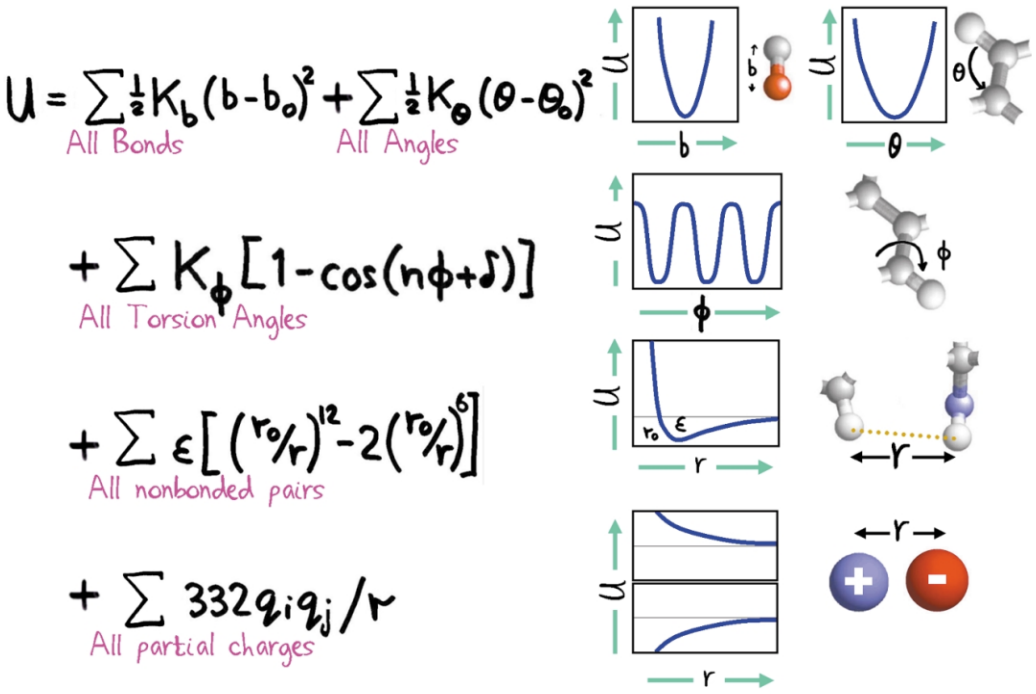

Figure 2. 1: Typical functional formalisms of potential energy terms found in a generic biomolecular force field, alongside the corresponding graphical and particle based depictions. Reproduced from [167].

Bonded interactions

Bonded interaction terms describe forces within between covalently bonded particles, which arise as a function of those covalent linkages. They include terms describing bond stretching, bending of angles, and rotations about bonds.

$$U_{bonded} = \sum U_{bonds} + \sum U_{angles} + \sum U_{dihedrals}$$

Stretching of a bond between two particles, *i* and *j*, of length b_{ij}, is generally described using a potential derived from Hooke's Law. Likewise the potential describing bending of an angle defined by three particles, *i j k*, is usually implemented using a similar term:

$$U_{bonds}(b_{ij}) = \frac{k_{ij}^{b}}{2}(b_{ij} - b_{ij}^{0})^2$$

$$U_{angles}(\theta_{ijk}) = \frac{k_{ijk}^{\theta}}{2}(\theta_{ijk} - \theta_{ijk}^{0})^2$$

Where b_{ij} is the instantaneous bond length, b^0_{ij} is the reference bond length, and k^b_{ij} is the spring force constant.

Dihedral angles, or torsions, are calculated as the angle between two planes defined by the coordinates of particles *i j k l*, where one plane corresponds to that defined by *ijk* and the second by *jkl*. Proper dihedral potentials are periodic in nature, with energy maxima representing the canonical "eclipsed" conformation of for instance, butane, and minima corresponding to sterically favourable "staggered" conformations [168].

$$U_{dihedrals}(\phi_{ijkl}) = k_{\phi}(1 + cos(n\phi_{ijkl} - \phi^{ref}_{ijkl}))$$

Where ϕ_{ijkl} is the instantaneous dihedral angle, ϕ^{ref}_{ijkl} is the reference value, k_{ϕ} is the force constant, and *n* is the multiplicity indicating the number of minima as the bond rotates 360°.

Improper dihedral potentials may also be employed to maintain planar structures, such as benzene, and to maintain chirality of tetrahedral groups. These potentials are generally harmonic, similar to those describing bond and angle terms.

Non-bonded interactions

Non-bonded interactions within classical forcefields are described by a van der Waals term, modelled using a Lennard Jones potential, and an electrostatic term, modelled using Coulomb's Law:

$$U_{non-bonded} = \sum U_{VdW} + \sum U_{Electro}$$

The Lennard Jones potential applies to all particles in the system, and is typically modelled in a 6-12 form, where the $\left(\frac{1}{r_{ij}}\right)^{12}$ term describes repulsive forces at short range, and the $\left(\frac{1}{r_{ij}}\right)^{6}$ term describes attractive forces at longer ranges (Figure 2.1):

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

Where r_{ij} is the instantaneous interatomic distance between particles *i* and *j*, ε_{ij} is the well-depth, and σ_{ij} is the distance at which the energy is zero. The values of these latter two terms are computed using the Lorentz-Berthelot rules:

$$\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj})$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}$$

Where $\sigma_{ii/jj}$ and $\varepsilon_{ii/jj}$ are the collision diameters and well-depths specific to the atom types *i* and *j*.

In addition to Lennard-Jones interactions, charged particles may interact *via* a Coulombic potential. This includes both ions, as well as atoms to which partial charges have been assigned in mimic of the differential charge distributions which may be found within certain molecules, arising from differences in the electronegativity of constituent atom types. The Coulombic potential between two particles *i* and *j*, bearing partial charges q_i and q_j, and separated by a distance r_{ij}, is given by:

$$U_{Electro}(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

Where ε_0 is the permittivity of free space.

Coulombic interactions are rather longer in range than those described by a Lennard Jones potential, with the decay of $U_{Electro}$ as a function of increasing r_{ij} being rather less steep than for U_{VdW}, which instead decays with an r_{ij}^6 dependence (Figure 2.1). Thus, whilst Lennard Jones interactions are typically neglected beyond a certain cut-off, the treatment of Coulombic interactions is generally rather different, as will be discussed in subsequent sections.

2.2.1 Energy minimization

Prior to initiation of a molecular dynamics simulation, systems are generally subjected to an energy minimization step. Such a procedure involves iteratively adjusting atomic coordinates to reach a local minimum in the potential energy landscape, described by the afore mentioned force field. This is key in alleviating incorrect geometries and steric clashes which may be present in the initial system configuration. Leaving such features uncorrected may result in the accumulation of unacceptably high forces and unstable simulations. In addition to energy minimization of the initial configuration, further rounds

of minimization may be conducted, for instance after converting a system from coarse-grained to atomistic detail (Chapters 4 & 6), or creation of a large membrane by periodic box replication (Chapters 4 & 5). Within this thesis the steepest descent algorithm [155] is used, as implemented in GROMACS [169].

2.2.2 Integration methods

The force field described previously governs the instantaneous interactions of particles within the system. Finite difference methods are then employed to integrate Newton's equations of motion and generate updated system coordinates as a function of those prior interactions. This involves breaking the integration down into multiple segments separated by a small timestep, δt. The force on each particle at time t is calculated as the vector sum of its interactions with other particles, from which acceleration may be calculated, and combined with the positions and velocities of particles at time t, to generate new positions and velocities at $t+\delta t$. This involves the assumption of constant force during the period δt, which necessitates choosing the value of δt with care as discussed in section 2.3.3.

Simulations described within this thesis utilize the leap-frog integrator [170] as implemented in GROMACS. This is an extension of the Verlet algorithm [171], so named because the particle positions are updated at integer timesteps, whilst velocities are updated at half-integer timesteps, thus essentially 'leap-frogging' over one another. The relationships for updating positions, r, and velocities, v, within the leap-frog integrator are given by:

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

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

Where $a(t)$ is the acceleration of particle i at time t calculated from $\frac{F_i(t)}{m_i}$.

This process is contingent on knowing particle velocities from the prior step. Thus, the question arises as to the origin of initial velocities. In general, such initial random velocities at the start of a simulation are canonically taken from a Maxwell-Boltzmann distribution at a user-specified temperature.

2.2.3 Choice of timestep

The frequency with which the equations of motion are integrated is determined by the time step, δt, specified by the user. Too small a value is computationally inefficient and limits sampling of phase space, whilst too high a value may fail to adequately sample bond vibrations, and may lead to deviation from their equilibrium values, resulting in the accumulation of artefactually high forces and simulation errors. The choice of an appropriate time step is determined by force field, system composition, integrators, and bond constraints. Within this thesis, a timestep of 2 fs is employed for simulations using the GROMOS forcefield, and 20 fs for simulations employing the MARTINI forcefield.

2.2.4 Bond constraints

Covalent bonds with high vibrational frequency are typically constrained to their equilibrium value within MD simulations. Constraint algorithms are implemented after particle coordinates have been updated by the integrator; and act to correct deviations from equilibrium bond lengths. This is particularly useful for bonds undergoing high frequency vibrations, such as those between heavy atoms and hydrogens, and allows for a larger timestep to be applied. The LINCS algorithm [172] was applied to constrain covalent bonds within this thesis.

2.2.5 Periodic boundary conditions

Simulation of a small, finite number of particles leads to the question of how to treat the system at its boundaries. For example, treating the boundaries as a surface from which approaching particles are forbidden from crossing would perturb the system and lead to unphysical behaviour. Periodic Boundary Conditions (PBC) represent an approach to better approximate the behaviour of particles as though they were in the bulk phase. The approach can be visualised by imagining a central simulation box, around which periodic images are replicated in each dimension (Figure 2.2).

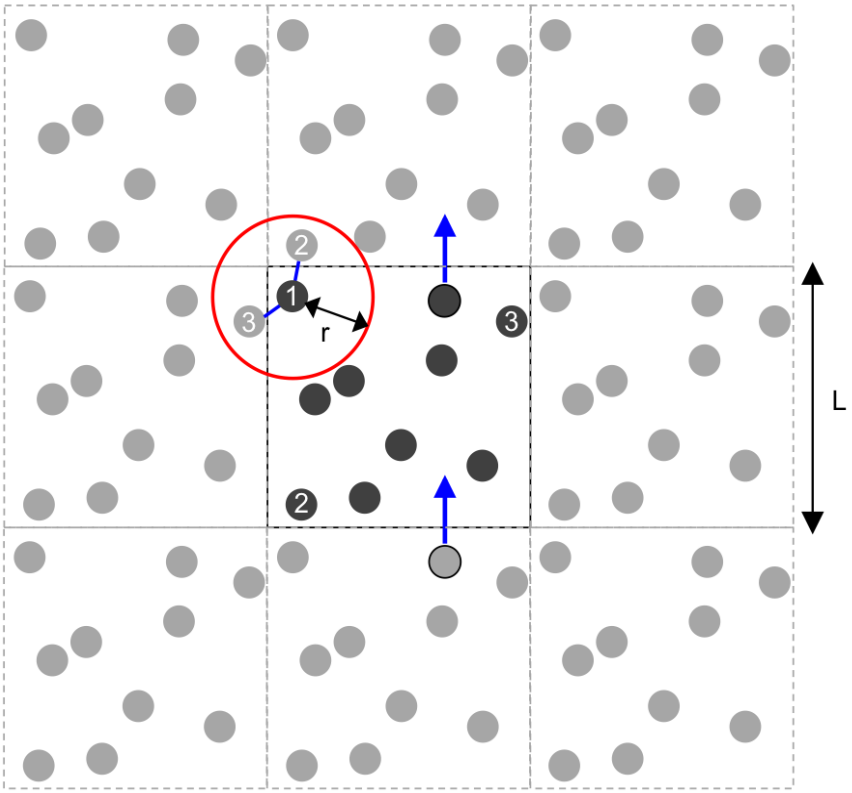

Figure 2. 2: Two dimensional representation of the periodic boundary concept. The unit cell (centre) is replicated in all directions such that a particle diffusing across a periodic boundary on one side will reappear on the other, illustrated by blue arrows. Non-bonded interactions are calculated across the periodic boundary, using a minimum image convention in conjunction with a cut-off scheme, such that the particle labelled "1" may interact with particles 2 and 3, but not its own periodic image. Consequently r must always be $< L/2$.

Particles may diffuse across one face of the box, re-appearing on the opposite face, and likewise particle interactions are calculated across boundaries. Given this, it is important that the system dimensions are chosen such that constituent particles, and indeed protein molecules, do not feel themselves across the boundary and self-interact, leading to artefacts. The minimum image convention is applied along with cut-off schemes such that each particle may interact with only those particles and image particles within less than half the unit cell dimensions. In the case of asymmetric cells the smallest distance is taken. Other periodic boundary considerations arise in simulations of large deformable structures such as bilayers, where curvature fluctuations and adaptations in box dimensions may lead to self-interactions. When simulating systems for the first time, such adaptations may be difficult to predict and so it can be prudent to build in more

generous dimensions as a buffer against such effects, with the caveat that computational demand is increased. The PBC approach is not without limitations, including undesirable system 'ordering', influences on the wavelength of membrane fluctuations, and challenges in dealing with finite size issues in membrane simulations [173-175].

2.2.6 Temperature and pressure coupling

Application of the molecular dynamics protocol so far described would lead to simulation of an NVE ensemble, in which the number of particles (N), volume (V), and energy (E) remain constant. Within biomolecular simulations it is generally desirable to control the temperature and pressure of the system, leading to simulation of an NPT or NVT ensemble, thus allowing approximation of the conditions of an *in vitro* experiment or *in vivo* biological system. Multiple algorithms have been developed to achieve these conditions. Within this thesis the velocity-rescale thermostat [176] and Parrinello-Rahman barostat [177] were applied within atomistic simulations. The Berendsen thermostat and Berendsen barostat were applied within coarse-grained simulations [178].

Thermostats

The kinetic energy of a system is related to its temperature as follows:

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

Where m_i is the molecular mass, and v_i is the velocity of a given particle i, N is the number of particles, T is the instantaneous system temperature, and k_B is the Boltzmann constant. Temperature is maintained around a desired value through application of thermostat algorithms which couple the system to an external heat bath at a chosen reference temperature, T_{ref}. Particle velocities are scaled as a function of the deviation of the systems temperature from that of the reference:

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

The coupling constant, τ_t, determines the strength of coupling and hence the possible magnitude of temperature fluctuations. This approach is implemented in the widely used Berendsen thermostat [178]. The v-rescale thermostat is an extension of this approach

and includes an additional stochastic term which generates a more accurate canonical ensemble [176].

Barostats

Pressure can be controlled in a similar manner through coupling to an external pressure bath:

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

Where *P(t)* is the pressure component at time *t*, P_{ref} is the pressure of the external bath, and τ_p is the coupling constant. This approach is implemented in the Berendsen barostat [178] and ensures the system may dynamically change volume whilst maintaining pressure around the desired value. Importantly, the algorithm involves rescaling the simulation box shape and volume. Under anisotropic schemes, and depending on the particle interactions taking place, this can lead to substantial deviations away from the initial box dimensions and can be an important consideration when setting up systems prone to large structural alterations, such as the evolution of curvature in very large bilayers.

It is sometimes desirable to implement the barostat under isotropic conditions (*x y z* box vectors rescaled uniformly), or semi-isotropic conditions (*x y* vectors rescaled uniformly and independently of *z*). The latter is particularly useful to simulations of bilayer systems as it ensures stable bilayer structures. Simulations performed within this thesis were conducted with a reference pressure of 1 bar.

2.2.7 Treatment of long-range interactions

The non-bonded energies are potentially the most computationally demanding to compute. Within a system of *N* particles, each particle intuitively interacts with all other particles, leading to the required number of calculations scaling with N^2. A number of approaches have been developed to address this challenge.

In the case of Lennard-Jones (LJ) energy terms, the potential energy decays with an r^{-6} dependence, rapidly approaching zero as particle separation increases (Figure 2.1). Thus it is common practice to only calculate LJ energies for particles within a certain distance cut-off, reasoning that the LJ energy outside of this cut-off is negligible and can hence be set to zero. Multiple cut-off schemes may be employed, including simple 'hard' cut-offs

which abruptly truncate the potential at a specified distance, as well as shifting functions which modify the form of the potential around the cut-off distance to smoothly shift it to zero and avoid discontinuities in energy. A so-called 'neighbour list' is employed to track which particles are within the cut-off distance, which is periodically updated as particles coordinates evolve [179].

Coulombic interactions are comparatively longer-range, with the energy decaying with an r^{-1} dependence (Figure 2.1). This circumscribes the accurate calculation of particle interaction energies with a simple cut-off scheme. The Particle Mesh Ewald (PME) method [180, 181] is commonly used to treat coulombic interactions. Particle interactions are grouped into short and long-range potentials, with the former calculated in real space, and the latter in Fourier space, allowing rapid convergence.

2.3 Coarse-grained simulations

The principles and protocols described so far may be more intuitively understood at an atomistic level. However a range of force fields employ more abstract representations in which multiple atoms are 'coarse-grained' into groups represented by single particles. The central benefit derived from such approaches is a reduction in the number of particles within a given system, thus reducing the computational demand and allowing access to longer time and length scales, whilst compromising on atomistic detail. In general the accessible time and length scales scale in proportion to the degree of abstraction (Figure 2.3).

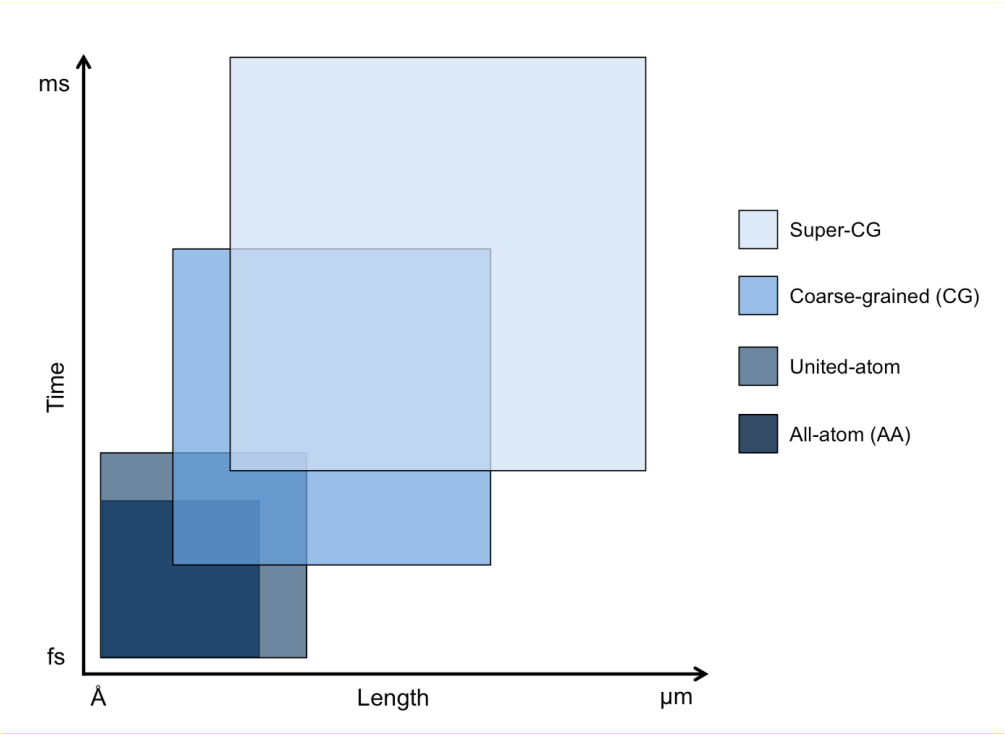

Figure 2. 3: Generalized relationship between force field type, and the accessibility of simulation time and length scales.

The most appropriate degree of abstraction should be informed by the biological question. Clearly examination of the molecular mechanism of ion passage through the selectivity filter of a pore [182] or channel [183] is a highly detailed process requiring application of an atomistic forcefields, such as AMBER [163], CHARMM [164], or OPLS-AA [165]. Likewise investigation of slow, large scale phenomena such as the development of membrane curvature and correlation with lipid clustering is more suited to coarse-grained models [184]. The GROMOS forcefield, through traditionally referred to as "atomistic" due to its explicit representation of heavy atoms, may be considered mildly coarse-grained to due it's implicit representation of non-polar hydrogen atoms within the heavy atom to which they are bonded [166]. These hydrogens are generally uninteresting and their abstraction results in a significant speed-up with a comparatively small compromise on atomic detail. More canonically the term "coarse-graining" is used to refer to the abstraction of groups of *heavy atoms* into single particles. A range of such models have been developed with various mapping schemes [185-188]. These include relatively "mild" forms of abstraction such as the MARTINI model in which groups of two to five heavy atoms may be represented by single particles [189], through to super-CG models

where entire lipids are represented by three particles [190], and residues may be grouped into single particles [191].

Mild forms of coarse-graining such as the MARTINI model result in preservation of much of the structural and molecular detail of biological systems, however this is clearly lost for super-CG models which are more suited to examination of macroscopic systems. Notably, hybrid models in which multiple levels of resolution are simultaneously present within the same simulation system have been developed [192]. The rationale behind these approaches is generally to have a more detailed representation of the principle component of interest, for instance the active site of an enzyme or ion channel, and a less detailed representation of the surrounding water/lipids, which are assumed to be less consequential. Such models may even be adaptive, with molecules dynamically switching resolution across a defined boundary [193]. However treatment of such hybrid models at the interface remains challenging, and they have yet to find wide application in the community.

2.3.1 The MARTINI force field

First developed for simulation of lipids in 2004 [188] by Marrink and co-workers, the MARTINI force field underwent an overhaul and extension in 2007 [189], where the term "MARTINI" was initially coined. The model was later expanded to include parameters for proteins [194], carbohydrates [195], nucleic acids [196, 197], and non-biological structures [198]. The force field has proved extremely popular, and found applications ranging from simulation of entire virion envelopes [199, 200], vesicles [201], planar lipid/protein membranes [48], and nanotechnology [202]. The central philosophy is that a common underlying collection of well-parameterised CG building blocks can be differentially mixed and matched to form a range of higher-order macromolecular structures without the need for continual 'bottom-up' re-parameterisation.

The forcefield employs on average a four-to-one mapping scheme with four heavy atoms represented by single CG particles. The CG particles are grouped into four major types: apolar (C), non-polar (N), polar (P), and charged (Q). Apolar particles represent groups of heavy atoms with net hydrophobicity, non-polar particles represent mixed hydrophobic/hydrophilic groups, polar particles represent hydrophilic groups, whilst charged particles represent groups containing net ionic charge. C and P particle types are further broken down into five subtypes with differing polarity, denoted by numbers 1-5,

where P1 is low polarity, and P5 is high polarity. In addition, N and Q types are broken down into four subtypes denoted by letters representing the capacity of the atoms they mimic to form hydrogen bonds. *d*=donor, *a*=acceptor, *da*=donor/acceptor, *0*=neither. In addition "S-type" particles have been developed to aid in the modelling the geometry of ring-like structures including aromatic residues, sterols, and nucleic acids, involving higher resolution 2:1 and 3:1 abstraction. The mapping of CG particles onto the atomic structure of phosphatidylcholine and phosphatidylserine lipids is illustrated in Figure X:

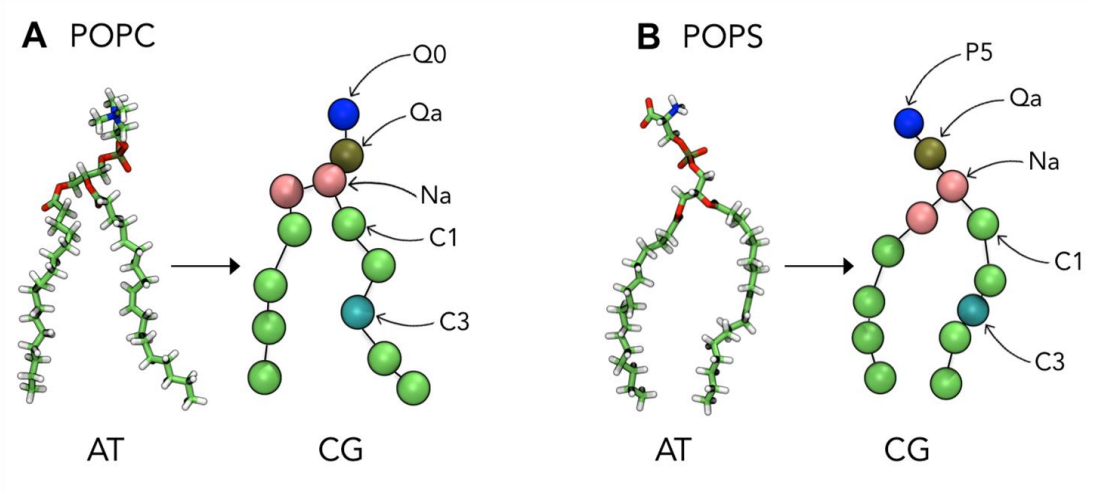

Figure 2. 4: Mapping of CG particles onto the atomic structures of A) palmitoyloleoylphosphatidylcholine (POPC), and B) palmitoyloleoylphosphatidylserine (POPS).

Thus the quaternary amine group of PC which possess a net charge of +1 but no hydrogen bonds capacity is represented by a Q0 type particle, whilst the zwitterionic serine moiety of PS containing both $-COO^-$ and $-NH3^+$ moieties but no net charge is represented by a P5 particle. In the case of proteins, each residue is assigned a single backbone particle representing the backbone heavy atoms, with the type assigned dependent on the local secondary structure. In general the polarity is decreased when residues are involved in secondary structures, in mimic of their decreased hydrogen bonding capacity compared to a disordered coil. The side chain mapping scheme for amino acids is shown in Table 2.1.

Table 2. 1: Mapping of amino acid side chains within the MARTINI forcefield v2.1.

Amino acid	CG mapping
Leu	C1
Ile	C1
Val	C2
Pro	C2
Met	C5
Cys	C5
Ser	P1
Thr	P1
Asn	P5
Gln	P4
Asp^-/Asp	Qa / P3
Glu^-/Glu	Qa / P1
Arg^+/Arg	N0-Qd / N0-P4
Lys^+/Lys	C3-Qd / C3-P1
His^+/His	SC4-SP1-Qd / SC4-SP1-SP1
Phe	SC4-SC4-SP4
Tyr	SC4-SC4-SP1
Trp	SC4-SP1-SC4-SC4

Adapted from [194]. Gly and Ala are represented by single backbone particles, with the type dependent on the secondary structure assignment. Notably the mapping of amino acid side chains was recently updated in v2.2 [203].

Bonded interaction between particles are described using the same formalisms presented in Section 2.2, with harmonic potentials applied to mimic bonds, and angles and dihedrals maintained through respective application of harmonic and periodic potentials. The primary mode of non-bonded interaction between particles is *via* 12-6 LJ potential, with the interaction strength moderated by differential well-depth (ε) values for each pairwise interaction. See [189] for the particle interaction matrix. The value of the σ parameter, denoting the distance at which the energy is zero, is set to 0.47 nm for all particles except "S-type" (σ=0.43 nm), and the interaction of Qx particles with C1 and C2 where σ=0.62 nm. Due to the higher resolution 2/3:1 mapping of S-type particles, the ε value is accordingly scaled to 75% of that of the parent particle type. Partial charge interactions may be considered implicit in the LJ interactions, whilst interactions between particles with formal net charge are modelled *via* a Coulombic potential. LJ potentials are smoothly shifted to zero between 0.9 - 1.2 nm, whilst Coulombic interactions are shifted to zero between 0.0 – 1.2 nm to avoid noise arising from discontinuities in energy.

Parameterisation of the model then relies on iterative comparison to atomistic simulations and experimental data. For instance bond lengths and angles within proteins were chosen based on structural data derived from a survey of over 2000 PDB entries [194]. Additionally the experimental partitioning free energies of small molecule analogues of amino acid side chains between aqueous solution and organic phases may be compared to those derived from CG simulations [189]. The values were optimized to within 1-2 kT for most compounds. Similarly, in the original model [188] the free energy for partitioning of water in hexane was calculated *via* simulation as 24.2 ± 2 kJ/mol, compared to an experimental estimate of 25 kJ/mol. Various properties of molecular assemblies such as lipid bilayers may also be compared, for example the area per lipid is a readily calculable observable which generally agrees with experimental estimates. As discussed in [22] the force field has had notable success in identifying lipid interaction sites on a range of integral membrane proteins, in broad agreement with crystallographic structures, NMR spectroscopic measurements, and cell-based mutagenesis assays. A full breakdown of parameterisation protocols is beyond the scope of this thesis, and the interested reader is referred to the original MARTINI publications [188, 189] as well as an interesting perspective from the developers [204].

Treatment of secondary and tertiary structure

The absence of explicitly represented hydrogen bonds within the MARTINI model necessitates additional protocols to maintain protein secondary and tertiary structure.

In the case of secondary structure, the structure "type" is assigned to particles using DSSP [205], with secondary structural dynamics then restrained around a prescribed equilibrium value through application of harmonic and periodic potentials to backbone angles and proper dihedrals, respectively. The force constants and equilibrium angles employed are detailed here [194]. Importantly a degree of simplification is used compared to strict DSSP classification, with all helical types merged into a single generic helical representation mimicking an α-helix. Perhaps most notably this parameterisation means the model does not fully model conformational transitions in secondary structure.

The two most common approaches to maintain tertiary structure within MARTINI simulations are traditional Elastic Network Models (ENMs) and the more recently developed Elnedyn [206]. The salient principle behind both approaches is to apply harmonic bond potentials between all backbone particles within a certain distance of each

particle. This creates a mesh-work of bonds which prevent significant deviations away from the initial tertiary structure, whilst preserving some local backbone dynamics. Side chain particles are generally let free. In the case of a "default" traditional ENM, harmonic bonds are applied between backbone particles within 0.5-0.9 nm of each other, using a force constant of 500 kJ/mol/nm^2. Similar harmonic bonds are applied in Elnedyn model [206], with bonds applied to all backbone particles within a 0.9 nm cut-off. The principle difference is that backbone particles in the Elnedyn model are centred on the Cα atom of each residue, whilst in canonical MARTINI backbone particles are positioned on the centre-of-mass of the backbone heavy atoms of each residue. Thus the Elnedyn model involves an alteration in the underlying structural mapping and bonded parameters, in addition to application of harmonic bonds between backbone particles. It is important to realise that the cut-offs and force constants associated with these models are tuneable and ideally parameterised *de novo* for each new protein structure by comparison to atomistic force fields. However in practice the default parameters are widely used, and in the case of Elnedyn have been shown to perform well across a range of protein structures and scales [206]. Within this thesis, all proteins with tertiary structure are modelled using the Elnedyn.

Water model

The standard MARTINI water model maps four atomic water molecules to a single P4 particle. As such the standard waters do not feel the electric fields of charged particles, and in common with a number of atomistic force fields do not reproduce polarization effects. These approximations are a trade-off one makes to access the benefits of CG models and classical MD. The development of a polarizable water model for use with MARTINI [207] went some way to addressing these challenges. In this model a further two CG particles denoted W_P and W_M are attached to the canonical P4 particle *via* harmonic bonds, and their relative orientations fixed by a harmonic angle potential. These particles bear charges of $+q$ and $-q$ respectively, and only interact with other particles *via* coulombic potentials, and therefore only visible to Q type particles. This polarizable model has been shown to more accurately model the dielectric screening properties of bulk water and its freezing point. In regard to lipid-protein interactions, the clustering behaviour of glycolipids in MARTINI simulations has recently shown to better approximate atomistic simulations when performed in the presence of polarizable

MARTINI water model [208]. Glycolipids possess large, solvent exposed headgroups and thus water may be expected to play a more prominent role in their lateral interactions compared to other lipids.

The principle limitations of the polarizable model are the uncertainty surrounding its impact on existing MARTINI parameters, which were derived in systems containing the standard water model, and its ability to interface with molecule representations which are not currently polarizable, e.g. proteins and lipids. Further, the use of additional particles results in a concomitant increase in computational load, and thus some of the benefits CG are lost. For a pure water system, polarizable water simulations are slowed by an approximate factor of three compared to standard MARTINI water. At some point, one may as well conduct an atomistic simulation, for which the properties of water remain closer to experimental estimates [207].

An implicit water model has also been developed, the so-called "Dry MARTINI" [209]. This model moves in the opposite direction, towards higher levels of abstraction and approximation. It is primarily suited for macro-scale simulations of large lipid systems, e.g. virion envelopes. In such systems bulk water can make up a significant proportion of the total particles, and much time is spent simulating these bulk waters which are generally not the component of primary interest. Thus use of an implicit water model can result in significant speed-ups for such systems, while reproducing the general properties such as area per lipid and bilayer thickness. However, proteins behave less well than in explicit water, and many of the parameterisation and compatibility concerns cited for the polarisable model remain applicable to Dry MARTINI.

The standard water model was used in all MARTINI simulations described in this thesis.

Versions

The MARTINI parameter set for lipids described in 2007 [189] is referred to as v2.0, with the subsequent expansion to proteins in 2008 [194] denoted as v2.1. In 2012 updated protein parameters were released [203], referred to as v2.2. The principle changes were decreases in the hydrophobicity of Phe, Trp, and Pro side chains, and minor alterations in bonded potentials to decrease the length of α-helices. For use with the polarizable water model [207] a variation detonated v2.2P was simultaneously introduced, which implements an off-centre charge model for residues bearing formal charge, and

polarizable side chains for Ser, Thr, Asn, and Gln [203]. These changes are expected to represent significant improvements when used in conjunction with the polarizable water model, however for simulations in this thesis using the standard water model the deviations between v2.1 and v2.2 are not expected to be significant.

Most recently, Vogel and co-workers [210] added additional periodic potentials to model dihedral angles in β-strands, leading to improved side chain dynamics as compared to atomistic simulations. Application of similar potentials to a-helices and other secondary structures did not lead to significant improvement. At the time of writing, these changes have yet to be been incorporated into the canonical MARTINI distributions. The protein structures simulated within this thesis are predominantly α-helical, with no β-strands and thus their behaviour is not expected to be significantly impacted by these results.

All MARTINI simulations described within this thesis were performed using v2.1, or v2.2, as stated within each Methods section.

2.3.2 Multiscale modelling

The approximations of coarse-graining may be reconciled to some degree through so-called multiscale modelling. Distinct from hybrid models in which multiple resolutions are simultaneously present within the same system (the "parallel" approach), multiscale modelling involves conversion of systems between CG and atomistic detail and successive application of separate MD simulation at each resolution, or the "serial" approach. Thus in the case of lipid interactions with an integral membrane protein, CG simulations may first be used to achieve adequate sampling and identify lipid binding sites, before subsequently converting [211, 212] the system to atomistic detail and running MD to refine the atomic detail of the interaction. This approach is exemplified in Chapter 4. A second common application of multiscale modelling is to embed a given membrane protein in lipid bilayer for the purpose of atomistic production run simulations, as described in Chapter 6. In this approach CG simulations may be performed to spontaneously self-assemble [213] a lipid bilayer around the protein, before subsequent conversion to atomistic coordinates. This approach has the major advantage over "cut-and-paste" methods [214, 215] in allowing the protein to dynamically adjust its position in the bilayer and adopt its optimal orientation.

CG to atomistic conversions described in this thesis were conducted using CG2AT [211], an in-house program employing a fragment-based conversion protocol. In addition to fragment based alignment, for proteins the program is also capable of more conservative alignments to user-specified whole protein structures.

2.4 Non-equilibrium methods

The timescales and sampling available within equilibrium simulations are limited, and sometimes circumscribe the investigation of certain 'rare' molecular phenomena, such as the computation of free energy surfaces for the association of transmembrane helices from an ensemble of system configurations [216], or calculation of force profiles for the passage of DNA nanopores through lipid membranes [217]. In these examples, obtaining sufficient sampling is challenging due to high energy barriers, and thus even running very long equilibrium simulations is unlikely to capture the rare events necessary for accurate computation of free energy surfaces or force profiles. In such cases, non-equilibrium methods may be employed in which a bias is introduced (e.g. by modifying the Hamiltonian) to drive the system towards certain configurations of interest, and capture events which otherwise may prove intractable.

2.4.1 Steered molecular dynamics

Within steered molecular dynamics (SMD) simulations, an external force is applied to a user-defined group of particles to drive motion along a reaction coordinate (also referred to as a 'collective variable', CV). Commonly used reaction coordinates include the center-of-mass separation between two particular groups [218], rotation about a bond [219], and root-mean-square-deviation from a reference structure [216]. A harmonic restraint may then be applied along the chosen coordinate through addition of a harmonic term to the system Hamiltonian, fixing the system at a given point along the coordinate [220]. This point is shifted with time, propagating the chosen groups along the reaction coordinate:

$$\mathrm{U}_{spring} = \frac{k_0}{2}(x(t) - x_0 - vt)^2$$

Where $x(t)$ is the position of the pulling group at time t, x_0 is its initial position, v the velocity at which it moves along the coordinate, t is the simulation time, and k_0 is the spring constant. k_0 may be modified to control the how 'closely' the pulling groups follow

the center of the potential as moves along the reaction coordinate. A range of other potentials may be used, as well as constant forces.

Within this thesis, constant velocity SMD simulations were performed using the pull-code as implemented in GROMACS [179].

2.4.2 Potential of mean force

The potential of mean force (PMF) describes the change in free energy along a particular reaction coordinate *R*, and is derived from the probably distribution along that coordinate, *p(R)*. The PMF, *W(R)*, is defined according to the formalism:

$$W(R) = -kTln(p(R))$$

A key challenge in the computation of any PMF is attaining sufficient sampling of configurational space. In theory both the distribution function and PMF may be computed directly from a straight MD simulation. However in practice sampling limitations prohibit such an approach [221]. A range of enhanced sampling methods have been developed to address this, including the afore mentioned SMD approach [222], adaptive force biasing [223], and umbrella sampling [221]. In the case of umbrella sampling, a biasing potential, *U(R')*, is introduced to enforce sampling about a particular prescribed region along the chosen reaction coordinate. Typically this biasing potential is harmonic:

$$U(R') = \frac{k}{2}(R - R_i)^2$$

A number of such simulations, so-called 'windows', are performed with the biasing potential differentially centred on successive values along *R*. The unbiased PMF, *W(R)*, is then given by:

$$W(R) = -kTln\big(p(R')\big) - U(R') + C$$

Where *p(R')* denotes the biased probably distribution, and *C* the undetermined free energy constant. The weighted histogram analysis method (WHAM) [224] is the typical approach then used to unbias and combine the potentials and generate a PMF profile. Popular implementations of WHAM include those of Grossfield [225], and the GROMACS implementation, g_wham [226]. This approach has been widely used over the years to explore a range of biomolecular processes, including the passage of ions through pores and channels [182, 227], dimerization of transmembrane helices [218, 228]

and β-barrels [229], the passage of small molecules across lipid bilayers [230, 231], soluble protein-protein interactions [232, 233], and the binding of drugs to their target macromolecule [234, 235].

Central considerations for calculation of a PMF profile *via* umbrella sampling are the choice of window separation, and strength of the biasing potential *U(R')*, controlled by the user defined spring constant *k*. For accurate computation of a PMF profile it is essential that the probability distributions of adjacent windows have sufficient overlap to allow accurate recombination, and thus the reaction coordinate is sampled in its entirety. The stronger the biasing potential, the smaller the spatial limits of sampling per window, and thus a smaller window separation with a greater total number of windows would be required to adequately sample the coordinate. However whilst weaker biasing potentials will allow an increased range of sampling along the coordinate (with a concomitant decrease in the required number of windows), that region of the coordinate will be less adequately sampled due to 'looser' biasing around the prescribed value of *R*. This is particularly true for regions of the coordinate with higher energy barriers. Smaller window separations may be used to increase sampling in such regions, with larger separations being used elsewhere [232].

PMF profiles described within this thesis were constructed using the GROMACS implementation of WHAM, g_WHAM [226].

"Simplicity is the ultimate sophistication."

- Leonardo da Vinci

3

Lipid interactions with the transmembrane domain of the EGF receptor

Contents

This chapter is based upon the manuscript:

"Free energy landscape of lipid interactions with regulatory binding sites on the transmembrane domain of the EGF receptor." George Hedger, David Shorthouse, Heidi Koldsø, and Mark S. P. Sansom, Journal of Physical Chemistry B, 2016, 120, 8154-8163.

3.1 Introduction

The epidermal growth factor receptor (EGFR/ErbB1) is perhaps the best characterized member of the human RTK family. It resides in the plasma membrane, where it serves an essential role in transmitting information on the cellular environment to intracellular signaling networks, which subsequently elicit a response [236]. Mutations of the receptor have been implicated in a variety of cancers and thus the EGFR is a target for therapeutic intervention [237]. Of particular interest, EGFR is known to be modulated by its surrounding lipid environment. The emerging picture is that this modulation may occur *via* specific lipid interactions [238, 239], in addition to the more general influences of lipids such as cholesterol on the biophysical properties of the membrane [240].

In addition to its intrinsic biological importance, the EGFR provides a well-characterized model system for the study of protein−lipid interactions. The full-length protein consists of an extracellular ligand binding ectodomain, a single transmembrane (TM) helix, a basic juxtamembrane (JM) region, and an intracellular protein kinase domain (Figure 3.1) [236]. The functional activity of the receptor is known to be modulated by a glycolipid (monosialodihexosylganglioside or GM3) and by phosphatidyl inositol-4,5-bisphosphate (PIP_2) [238, 239]. However, despite a number of biochemical and biophysical studies, the structural and energetic basis of the interactions of these lipids with the EGFR TM and JM regions remain to be fully characterized. This is in part due to the experimental challenges involved in studying the behaviour and interactions of lipids with proteins.

Molecular simulation techniques provide a powerful tool for exploring lipid interactions with membrane proteins [22] and have previously been used to successfully predict lipid interaction sites on cytochrome C oxidase [241], aquaporins [242], and Kir channels [77]. However, there is a need to quantify these interactions to enable predictions of the effects of protein mutation and to provide a fuller understanding of the lipid selectivity of proteins such as EGFR. Molecular simulations and potential of mean force (PMF) calculations represent one such route for the quantitative exploration of lipid−protein interactions. PMF methods are increasingly being applied to better understand the molecular basis of a range of biomolecular interactions (e.g., [232, 235, 241, 243, 244]) and allow one to describe the free energy profile along a particular reaction coordinate. Sufficient sampling of configurational space along this reaction coordinate is necessary for the accurate calculation of a free energy profile. Attaining sufficient sampling and convergence of the calculation is not an insignificant challenge

[245], which can inhibit its application to complex systems such as membranes. These challenges may be addressed through application of specialized sampling techniques such as umbrella sampling [246] and adaptive force biasing [223]. An additional approach is to reduce the granularity of the system (and hence the number of degrees of freedom that must be sampled) through use of coarse-grained (CG) models, which allow an increase in accessible simulation times of several orders of magnitude [204]. A number of comparative studies have indicated that estimates of the free energy for removal of lipid molecules from a bilayer obtained from CG and all-atom simulations are in agreement with one another and with experiment [189, 231]. On the basis of these considerations, umbrella sampling together with the MARTINI CG force field were applied. This force field was originally parametrized based on comparison with experimental and all-atom simulation partitioning free energies [189].

In this study, CG simulations and umbrella sampling are employed to explore the free energy landscape of interactions of a model of the wild-type (WT) and mutant EGFR TM−JM dimer with a range of different lipid species present within biological membranes.

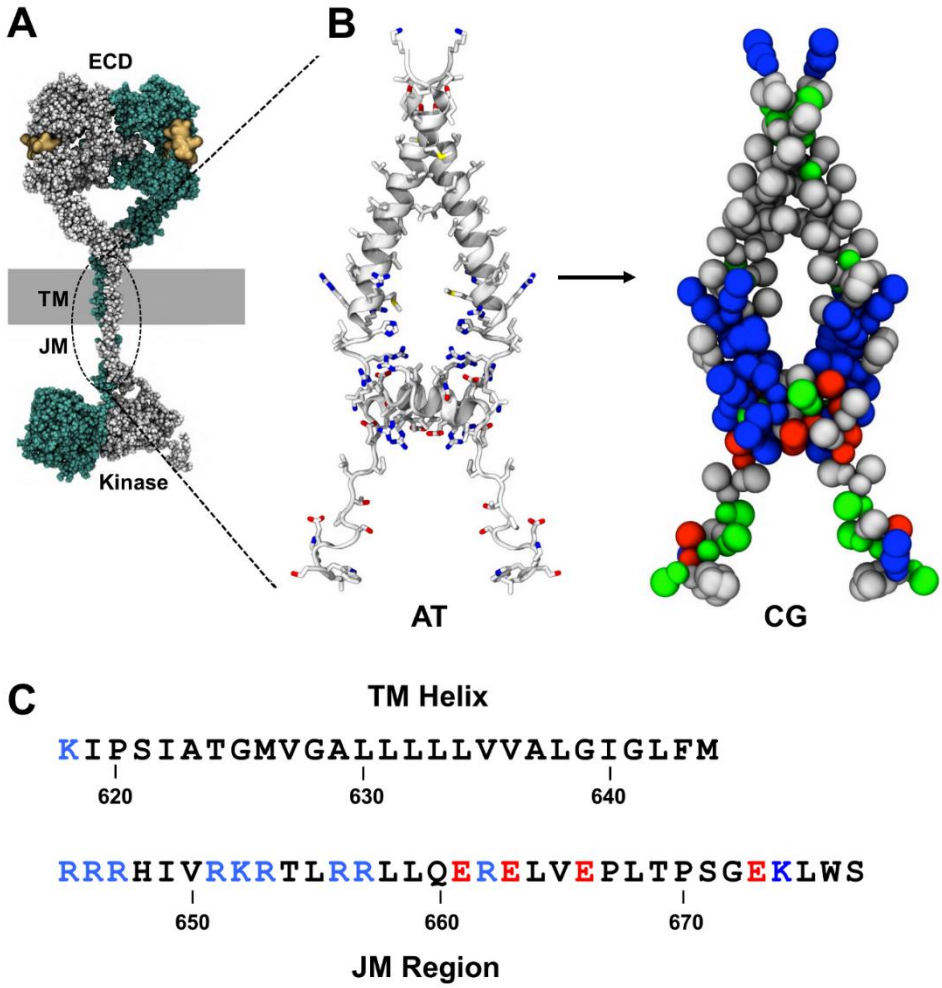

Figure 3. 1: Overview of EGF receptor structure and sequence. (A) Composite structural diagram of dimeric EGFR based on the extracellular domain (ECD) crystal structure (PDB ID: 3NJP), the transmembrane (TM) and juxtamembrane (JM) domains NMR structure (PDB: 2M20), and the asymmetric tyrosine kinase crystal structure (PDB IDs: 2GS6, 3GOP, and 2JIU). (B) NMR-derived structure of the TM−JM dimer used as the basis of the simulations. The atomistic (AT) structure was converted to a coarse-grained (CG) model (basic residues in blue, acidic residues in red, polar residues in green, and hydrophobic residues in gray). (C) Sequence of the TM helix and JM region, with the residue numbering as in the processed mature protein sequence (UniProt entry P00533). An addition of +24 should be made to convert to the preprocessed numbering.

3.2 Methods

3.2.1 NMR structure-derived molecular model

The molecular model of the EGFR helix dimer used was derived from the experimentally determined NMR structure (PDB ID: 2M20) [247]. Two native methionines (M626 and M644) had been mutated to prevent chemical cleavage of the protein during purification. These residues were restored to the WT residues using the mutagenesis tool implemented in PyMOL (https://www.pymol.org/). The sequence and structure of the model used in

simulations are indicated in Figure 1. Models of the K618G and R645−7N mutants were also obtained using the PyMOL mutagenesis tool. Energy minimized atomistic (AT) models were converted to CG representation using the MARTINI2.2 force field [203]. Assignment of dihedral restraints to model the TM and short JM-A helices was achieved in an automated fashion using the martinize.py workflow, based on DSSP assignment. The N- and C-terminus of each monomer were modelled with neutral charge.

3.2.2 CG simulation details

Simulations were performed using the GROMACS (www.gromacs.org) 5.0 and 4.6 simulation packages [179]. During the initial set of CG simulations, 50 ns self-assembly simulations [213] were performed to allow formation of a PC bilayer consisting of 700 PC lipid molecules around the TM region of the dimer (Figure 3.2).

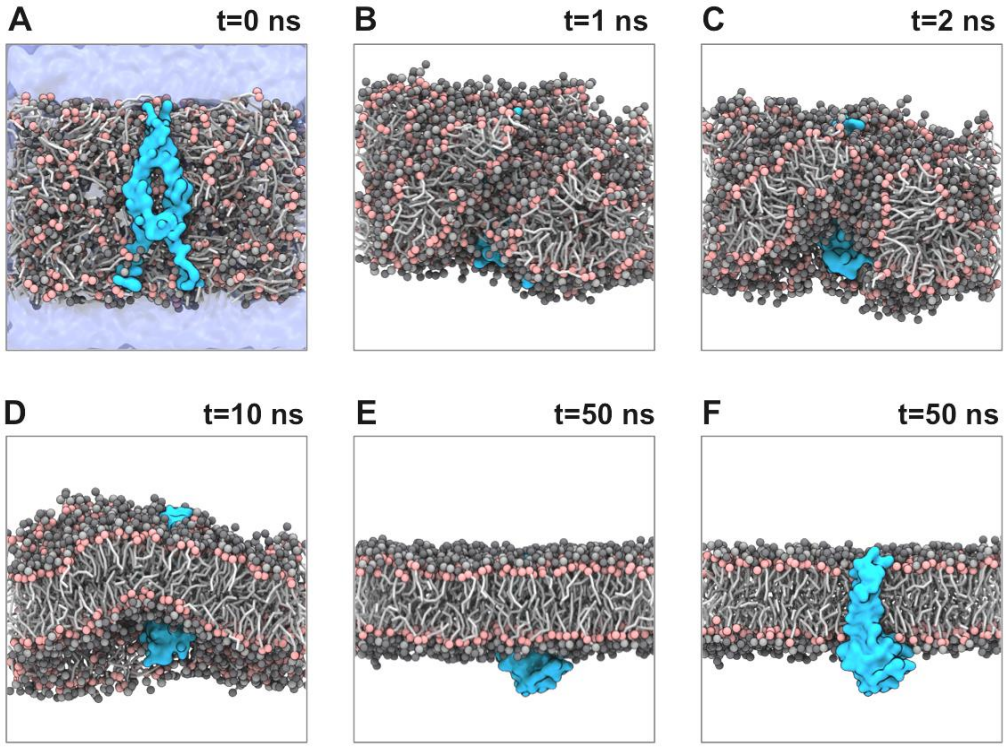

Figure 3. 2: Coarse-grain self-assembly schematic. POPC lipids are shown as white sticks (tails) with pink spheres (glycerol particles) and grey spheres (phosphate and choline particles). The TM-JM dimer is shown in cyan. Water and ions are omitted for clarity.

A locally developed script [184] was used to exchange PC molecules for other lipids to form an asymmetric bilayer containing 17 GM3 molecules (in the outer leaflet, corresponding to ca. 2.5% of all lipids within the system) and 17 PIP_2 molecules (in the inner leaflet, corresponding to ca. 2.5% of all lipids within the system). This provides a simplified model of the distribution and glycolipid/PI content of a plasma membrane [184, 248]. Five-thousand steps of steepest descent energy minimization were applied to relax the system, followed by a 10 ns equilibration simulation. Two microsecond production runs were performed with different random initial velocity seeds. Temperature was maintained at 310 K utilizing a V-rescale thermostat [176] with a coupling constant of τ_t = 1 ps. Pressure was controlled at 1 bar using a Parrinello−Rahman barostat [177] with a coupling constant of τ_p = 5 ps and a compressibility of 3 × 10−4 bar^{-1}. Protein, lipids, and solvent (water + ions) were coupled independently. Electrostatics and van der Waals interactions were shifted between 0 and 1.2 nm and 0.9 and 1.2 nm, respectively, and an integration time step of 20 was applied. Within umbrella sampling simulations, particle mesh Ewald was applied to improve modeling of long-range electrostatics [180]. Covalent bonds were constrained to their equilibrium values using the P-LINCS algorithm [172]. All simulations were run in the presence of conventional MARTINI water [189] and neutralized using 0.15 M NaCl. The MARTINI force field was used to describe all system components, other than PIP_2 and GM3, for which locally parametrized versions previously described by us [76, 184] were employed. Simulations were visualized using VMD [249]. Contact analysis was performed using VMD. Contacts were calculated between each residue of the protein and the PO3 bead of the PIP_2 lipid headgroup, and the B1A bead of the GM3 headgroup (see Figure 3.3 for lipid bead names). A 6 Å cut-off was applied to define a contact. Radial distribution functions for each lipid headgroup relative to the protein were calculated using the PO3 particle of PIP_2, PO4 particle of PC, and B1A particle of GM3.

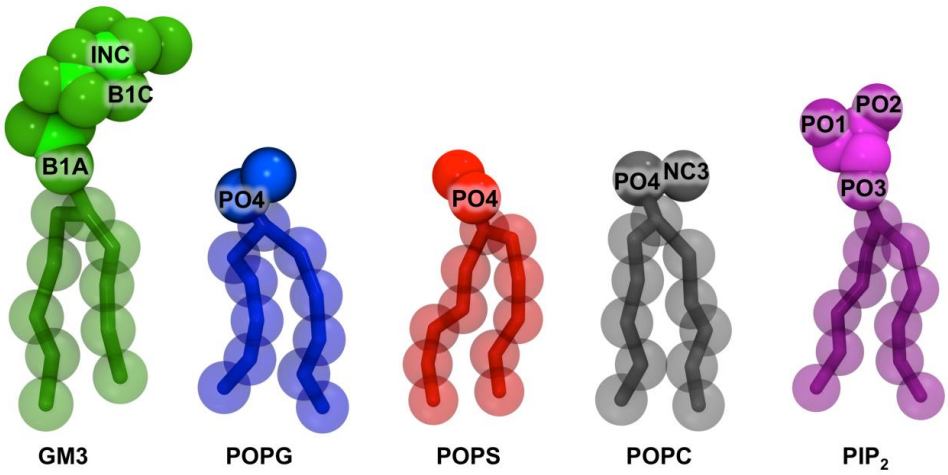

Figure 3. 3: CG lipids used in simulations and relevant bead names. CG lipids used in simulations and relevant CG bead types. B1C, PO4, and PO3 each bear a formal charge of -1, whilst PO1 and PO2 each bear a formal charge of -2, and NC3 a charge of +1. The POPG, POPC, and POPS lipids utilized in simulations were as described in the MARTINI forcefield [189], whilst the glycolipid GM3 and the phosphoinositide PIP_2 were described using locally developed versions.

3.2.3 PMF calculations

PMF is a key concept in statistical mechanics [221], which allows one to describe the free energy profile along a particular reaction coordinate (d), and is derived from the average distribution function p(d) along this reaction coordinate. A key challenge in calculating a PMF *via* MD simulation is attaining sufficient sampling of configurational space. This may be addressed *via* a number of sampling techniques including umbrella sampling [246] and adaptive force biasing [223]. In the current study, umbrella sampling is applied in conjunction with a CG model to calculate PMF profiles for lipid interactions with a TM helix dimer.

The last CG snapshot with a single bound lipid of interest was extracted from a 2 μs production run. A 50 ns self-assembly simulation was performed to permit formation of a PC bilayer around the dimer and the bound lipid. Position restraints were applied to the protein and the lipid of interest during the self-assembly, using a force constant of 1000 kJ/mol/nm^2. Protein side chains were subsequently relaxed during a 10 ns equilibration with position restraints (400 kJ/mol/nm^2 in the xy plane) only applied to the G625 and G628 backbone beads to prevent rotation and translation of the protein. All lipids of

interest remained bound over this time course, other than PC, to which a weak positional restraint of 100 kJ/mol/nm^2 was applied to the PO4 bead (Figure 3.3). Steered MD (SMD) simulations were performed to generate a series of configurations along a reaction coordinate ranging from the lipid bound to lipid unbound states (free energy profile reaches a plateau). The lipid was pulled at a rate of 0.1 nm/ns over a distance of 2−3 nm using a force constant of 1000 kJ/mol/nm^2, with the distance from the center-of-mass of G625 and G628 of the N-terminal dimerization interface of the EGFR model and a single bead of the lipid of interest defined as the 1D reaction coordinate (B1A for GM3; PO4 for PC, PS, and PG; and PO3 for PIP_2). The translational and rotational motions of the protein were reduced through application of position restraints to the backbone beads of G625 and G628 using a force constant of 400 kJ/mol/nm^2, with the z coordinate remaining unrestrained. For all inner leaflet SMD simulations, these restraints were also applied to M644 at the C-terminus of each helix to prevent the C-terminal portion of the helix from "following" the lipid. Additionally, a weak positional restraint of 100 kJ/mol/nm^2 was applied to a single bead of the lipid headgroup to limit its motional freedom along the y coordinate, in a manner similar to Arnarez et al [241, 250]. The y coordinate within the current system is defined as is the direction perpendicular to the reaction coordinate (x) and to the bilayer normal (z), and application of such restraints thus ensures sampling of the 1D reaction coordinate and prevents transition to 2D sampling and the associated convergence challenges. Application of these position restraints to the dimer, together with the standard dihedral restraints inherent to the MARTINI force field [194], prevented any major helical transitions or conformational rearrangements within the TM region of the dimer, whereas the flexible +4 N-terminus and JM were modeled in an unrestrained fashion and could freely interact with the membrane. The subject lipid was treated separately to bulk lipids for temperature and pressure coupling. Snapshots were extracted from the SMD simulation, and these configurations were used as input for umbrella sampling simulations. Windows were spaced asymmetrically with ∼0.05 nm spacing at low-center-of-mass separations and ∼0.1 nm in the bulk. Each window was run for between 500 and 1000 ns. Relative lipid−protein separations within each umbrella sampling window were maintained through application of a harmonic umbrella potential between the center-of-mass of G625 and G628 residues of the N-terminal dimerization interface and a single lipid headgroup bead, using a force constant of 1000 kJ/mol/nm^2. Additionally, the same position restraints used in the SMD simulations were applied in the umbrella sampling simulations. Within each window, temperature and pressure were

controlled using the Berendsen thermostat and Berendsen barostat [178]. The GROMACS implementation (g_wham) of the weighted histogram analysis method (WHAM) was used to combine and unbias the umbrella potentials, and the Bayesian bootstrapping method was utilized to estimate the error on each free energy profile [226]. Convergence was analyzed by comparing free energy profiles computed from nonintersecting periods of simulation time. On the basis of these observations, the first 100 ns of simulation time was excluded from analysis as equilibration. The possible introduction of bias caused by taking system configurations generated by pulling the lipid away from the dimer was tested by taking the end point of a 1 μs umbrella sampling simulation at the maximum protein−lipid separation and pulling the lipid "backward" toward the dimer from the bulk bilayer along the same 1D reaction coordinate. This backward trajectory was then used to generate a new set of umbrella sampling simulations conducted in the same manner described previously.

3.3 Results

3.3.1 Identification of lipid interaction hotspots

Lipid interactions were explored using a CG molecular model of the TM−JM dimer of EGFR based on an NMR structure (PDB ID: 2M20) [247]. The monomeric unit consisted of a 4 residue N-terminal region, a 23 residue TM helix, and a 32 residue JM region (Figure 1). This model was embedded in a phosphatidylcholine (PC) bilayer *via* self-assembly simulations [213]. Selected PC lipid molecules were subsequently exchanged for other lipid species [184] to form an asymmetric bilayer containing 2.5% GM3 (in the outer leaflet) and 2.5% PIP_2 (in the inner leaflet), providing a simplified model of the distribution of these lipids in mammalian cell membranes [248]. CG-MD simulations were performed using the MARTINI force field [189, 194, 203]. Five repeat simulations of the dimer in the asymmetric membrane were run, each 2 μs in duration (see Methods for full simulation details). The number of contacts formed over the course of the simulations between each protein residue and the headgroup of each lipid species was calculated. Within the outer leaflet, GM3 exhibited a strong preference for interaction with N-terminal membrane proximal residue K618, whereas within the inner leaflet, PIP_2 showed high frequencies of interaction with R645, R646, and R647, as well as with R656 within the cytoplasmic JM region (Figure 3.4). These lipid interaction hotspots identified in the current simulations agree well with *in vivo* mutagenesis studies and *in vitro*

functional assays that have shown that these residues are essential for the sensitivity of EGFR to GM3 and PIP_2 [194, 238]. Visual inspection of the trajectory revealed that once associated each PIP_2 lipid molecule remained bound over the 2 μs duration of the simulation. In contrast, the extracellular leaflet GM3 lipids exhibited a more dynamic pattern of transient receptor interactions such that multiple association and dissociation events were observed (Figure 3.4).

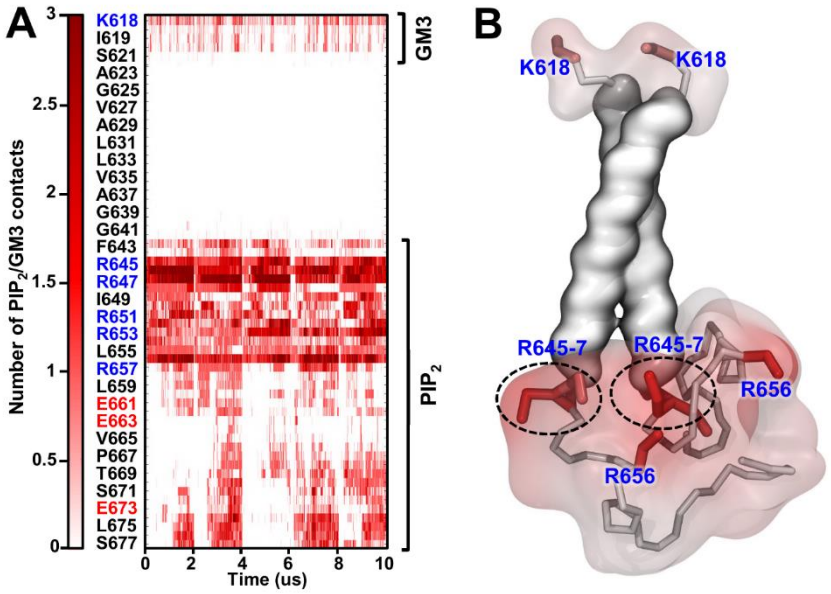

Figure 3. 4: Protein–lipid interactions within a model bilayer. Lipid interaction hotspots identified by CG simulations of the TM–JM dimer in a PC bilayer containing 5% GM3 in the extracellular leaflet and 5% PIP_2 in the intracellular leaflet. These are shown in (A) by a contact matrix of the dynamic pattern of interaction formed between the peptide and a single bead of the headgroup moieties of PIP_2 and GM3 over time (see Methods for full details of contact analysis). (B) TM helix dimer (snapshot from a CG simulation) with each residue coloured from white (no interaction with GM3 or PIP_2 headgroup) to red (high degree of interaction). Residues that formed the highest levels of contact with lipid molecules within each leaflet are labelled. Contacts were calculated over 5 × 2 μs simulation repeats, using a 6 Å cut-off.

In order to quantify the strength of the observed lipid−protein interactions and test the importance of residue identity at the predicted interaction sites, PMF profiles (i.e., free energy profiles) were calculated for lipid association with both WT and mutant receptors. A final CG snapshot with a single bound lipid of interest at either of the previously determined sites was extracted from the 2 μs simulation trajectory to provide starting structures for the PMF calculations. In each case, the snapshot corresponded to a structure

in which the lipid of interest was bound between the two TM helices, approximately equidistant from each helix axis. In the case of PIP_2, the flexible region of the JM after R647 was removed to leave just the first three JM residues (R645−7) present within a truncated JM region. This enabled expedition of data collection and to combat the technical challenge of achieving convergence in PMF calculations with a highly flexible JM region. A new PC bilayer was subsequently formed *via* self-assembly simulations around the position-restrained protein (see Methods for details of restraints used) and the bound lipid. The TM region of the dimer was thus conformationally "frozen" in terms of degrees of freedom, whereas the flexible +4 N-terminus and C-terminal JM regions remained unrestrained and free to dynamically interact with the membrane. The choice of membrane composition served to combat the additional convergence challenges that arise from calculations in more complex multicomponent bilayers. To generate a 1D reaction coordinate ranging from the lipid-bound state to the corresponding lipid (GM3 or PIP_2) being free in the PC bilayer, a steered MD simulation was performed in which a force was applied to pull the lipid of interest away from the restrained TM helix dimer over a distance of 2−3 nm (Figure 3.5). A series of 20−40 system configurations were extracted along the reaction coordinate, each spaced by ca. 0.1 nm. These configurations were used as starting points for umbrella sampling simulations. The subject lipid was maintained on the 1D reaction coordinate during umbrella sampling through application of lateral position restraints in a manner previously described in refs [241] and [250]. Simulations were performed using Gromacs 4.6 [179], and free energy profiles were constructed using the g_wham utility [226] (See the Methods for full details of how PMF calculations were conducted).

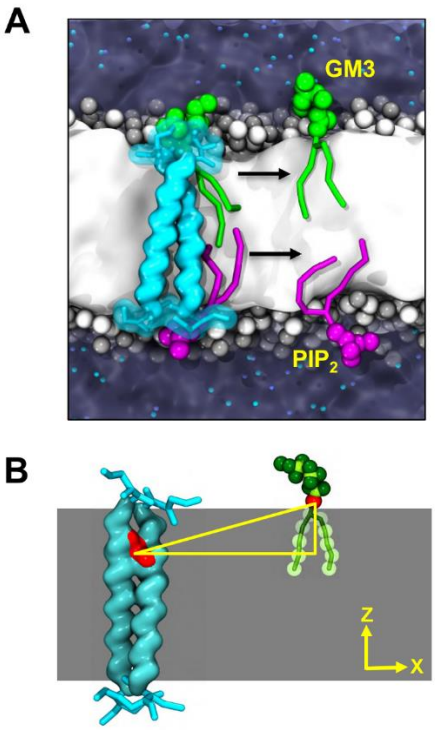

Figure 3. 5: Reaction coordinate for probing the energetics of lipid interactions with the EGFR TM–JM dimer in a lipid bilayer. (A) Schematic illustration of a dimer in a membrane, with the arrows showing the approximate pathways (i.e., reaction coordinates) for evaluating the free energy profiles (i.e., potentials of mean force, PMFs) of GM3 and PIP_2 interactions with the protein. (B) Reaction coordinate for exploring the energetics of protein–lipid interactions illustrated for GM3. The reaction coordinate is defined as the separation along the x coordinate of the center-of-mass (red) of the two glycine residues within the N-terminal GxxGA motif of the dimer and of the B1A particle (red) of GM3.

3.3.2 Binding of GM3 within the extracellular leaflet

Free energy profiles were calculated for GM3 and for phosphatidylglycerol (PG) interaction with the WT dimer within the outer leaflet and for GM3 interactions with a mutant dimer (K618G) (Figure 3.6). Additionally, the importance of the N-acetyl neuraminic acid (Neu5Ac) moiety of GM3 for protein interactions was tested by calculation of free energy profiles for GM3 lacking the Neu5Ac moiety (GM3 (−Neu5Ac)).

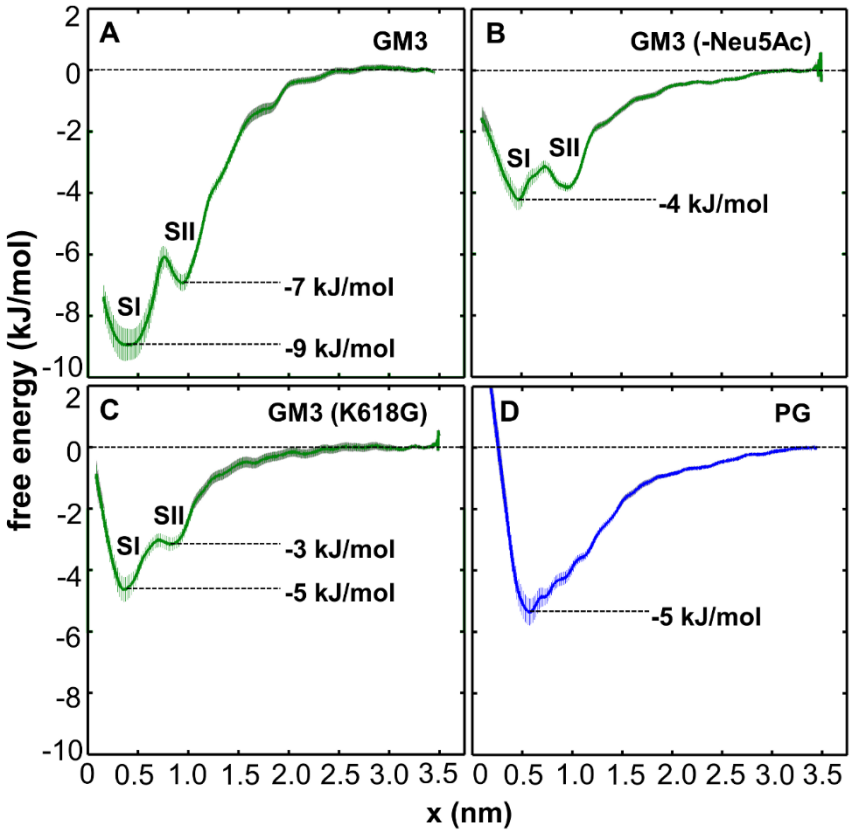

Figure 3. 6: Outer leaflet free energy profiles. (A) GM3, (B) GM3 (−Neu5Ac), i.e., GM3 from which the terminal N-acetyl neuraminic acid has been removed, (C) GM3 and the K618G mutant dimer, and (D) phosphatidylglycerol (PG). The standard deviation estimated from bootstrapping is shown as the shaded area behind the curve.

GM3 was found to interact with the EGFR dimer with an overall minimum (state I) free energy of −9 kJ/mol. A second local energy minimum (state II) with a free energy of −7 kJ/ mol was also observed (Figure 3.6A). These minima both represent bound states of GM3. In state II, the lipid headgroup is erect and parallel to the membrane normal. In contrast, in state I the GM3 lipid headgroup is tilted and penetrates further into the bilayer (Figures 3.7 and Figure 3.8; and Appendix Figure A1). This enables formation of optimal interactions between the N-terminal four residues of the helix dimer and the GM3 headgroup. In state I, there is a degree of local bilayer deformation seen as bilayer thinning (Figure 3.7 B,C). This local deformation is absent from state II. The second minimum thus most likely represents an energetically favourable adjustment of the tilt angle and depth of GM3 within the lipid bilayer upon moving away from the protein and the consequent release of local deformation in the surrounding PC bilayer. The ability of

PMF calculations to detect these subtle differences in the behaviour of a single lipid illustrates the sensitivity of the method.

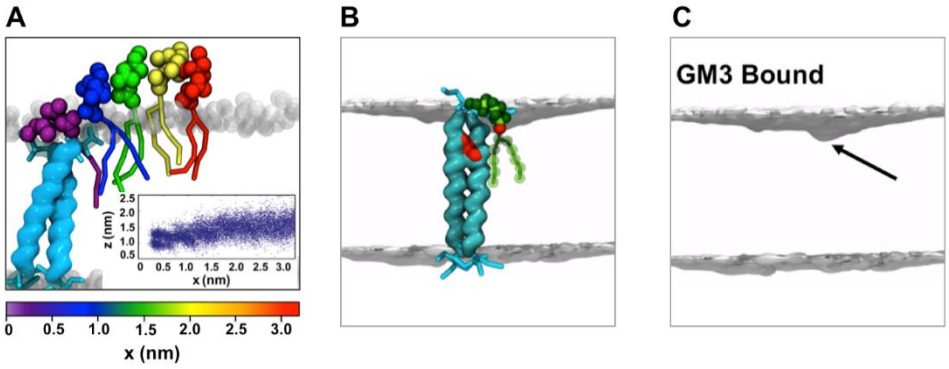

Figure 3. 7: Repositioning of GM3 at the protein–lipid interface. (A) Position of the GM3 molecule (purple to red) relative to the upper leaflet of the lipid bilayer (gray) as a function of the displacement, x, along the reaction coordinate. The inset shows the position along the bilayer normal (z) for the B1A particle (red) of GM3 as a function of the x coordinate of the lipid. (B,C) Density map of the location of the PC head groups averaged over the course of the first umbrella sampling window simulation (i.e., that with the GM3 closest to the protein). A degree of bilayer thinning is observed around the dimer and bound lipid. The approximate locations of the protein and GM3 are indicated in (B).

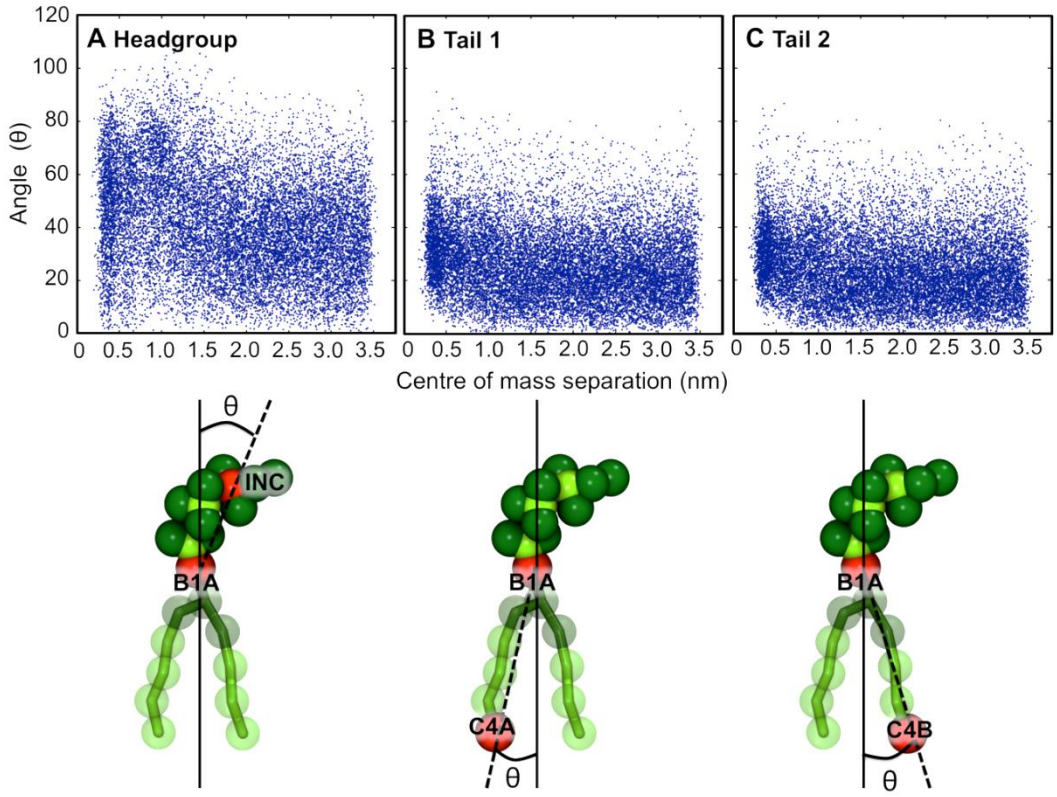

Figure 3. 8: Relationship between GM3 tilt angle and protein proximity. GM3 "tilts" relative to the membrane normal to form optimal interactions with the N-termini of the transmembrane helix dimer. (A) Distribution of headgroup angles over the reaction coordinate sampled during PMF calculations. (B+C) Angle distributions for tail 1 and tail 2. Beneath each plot is a cartoon depicting the angle reported. In each case a vector was drawn between the two beads indicated in red, and the angle between this line and the membrane normal measured for each frame of a concatenated trajectory of all window used during calculation of the GM3 PMF profile.

Mutation of residue K618 to glycine decreased the binding affinity of GM3 by 4 kJ/ mol compared to the WT (Figure 3.6). In addition, removal of the Neu5Ac sugar from the GM3 headgroup caused a reduction in binding affinity of 5 kJ/mol (Figure 3.6). Both the K618G mutant and GM3 (−Neu5Ac) truncation remove one of the formal charge components of the interaction between the lysine side chain and the carboxylate group of GM3. The reduction in binding affinity by ~50% demonstrates that both the cationic lysine side chain at position 618 and the anionic headgroup moiety of the glycolipid are significant contributors to the GM3−EGFR interaction. In agreement with this prediction, *in vitro* assays have shown that treatment of GM3 with neuraminidase (which cleaves Neu5Ac from the GM3 headgroup) abolishes its ability to modulate receptor activity and that K618 is likewise essential for GM3 interaction and the sensitivity of the receptor to

GM3 levels within membranes [238]. Significantly, the simple anionic lipid PG exhibited a weaker binding affinity than GM3 (Figure 3.6), despite both lipids bearing a formal charge of −1. This suggests that the additional polar moieties of the GM3 headgroup play a role in providing interaction sites that allow EGFR to form selective interactions with GM3. As anticipated, the zwitterionic lipid PC exhibited no significant interaction (Figure 3.9).

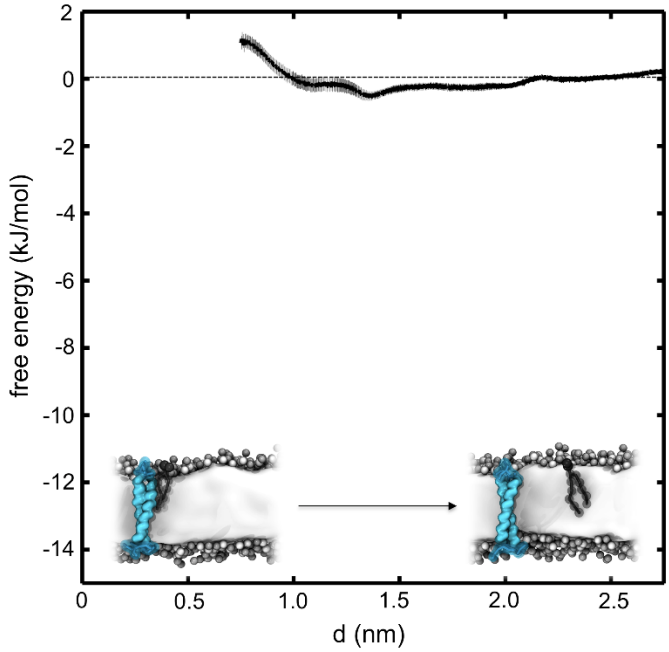

Figure 3. 9: Outer leaflet PMF profile for PC. No significant interaction of the lipid molecule (shown in back in the inset) with the protein (shown in cyan) of magnitude greater than kT was observed.

3.3.3 PIP_2 binding within the inner leaflet

Within the inner leaflet of the bilayer, free energy profiles were computed for the interactions of PIP_2^{-5}, PC, and PS. The importance of PIP_2 ionization state [251] was then evaluated by calculation of free energy profiles for PIP_2^{-3} and PIP_2^{0}, and the significance of the side chain properties at positions within the JM was investigated by calculation of a profile for the interaction of PIP_2^{-5} with a R645−7N mutant dimer (Figure 3.10). In contrast to the relatively small interaction free energies observed within the upper leaflet, PIP_2^{-5} bound to the dimer with a free energy of −42 kJ/mol, whereas PIP_2^{-3} was bound with a free energy of −35 kJ/mol. The significantly deeper wells observed in these profiles are consistent with observations from the contact analysis (Figure 3.4) and with visual

inspection of trajectories revealing that PIP_2 lipids remained bound for the duration of the initial 2 μs CG simulations. *In vivo*, PIP_2 lipids are suggested to exist in a mixture of the −3, −4, and −5 ionization states [252], although the exact charge may depend on the lipid microenvironment [253]. Thus, one would expect the interaction free energy of PIP_2 with the dimer to be about −40 kJ/mol, i.e., 4 times stronger than the interaction with GM3 in the opposite leaflet of the bilayer. Mutation of residues R645−7 within the JM cationic cluster of each monomer to three asparagines resulted in a significant reduction in the strength of interaction to about −10 kJ/mol. The free energy of the PIP_2 interaction with this asparagine (NNN) mutant is comparable of that of phosphatidylserine (PS) with the WT receptor (Figure 3.10), demonstrating that these residues are essential for the strong binding of PIP_2 lipids as well as the selectivity of EGFR for this lipid over other species present within cell membranes. Furthermore, as expected, these mutations show that the formal charge interactions between PIP_2 phosphate groups and the basic arginine residues contribute to a significant portion of the binding energy. In support of this, the interaction of neutral charge state PIP_2 lipids was ca. 30 kJ/mol weaker compared to that of the −5 ionization state (Figure 3.10). The residues identified by the simulations as important for protein−lipid interactions have been experimentally shown to be required for normal receptor function, with R-to-N and R-to-A mutants exhibiting significantly impaired activity. In addition, basic residues within the early JM region have been shown to be crucial in facilitating PIP_2 interactions and the ability of PIP_2 lipids to modulate receptor activity *in vivo* [239, 254].

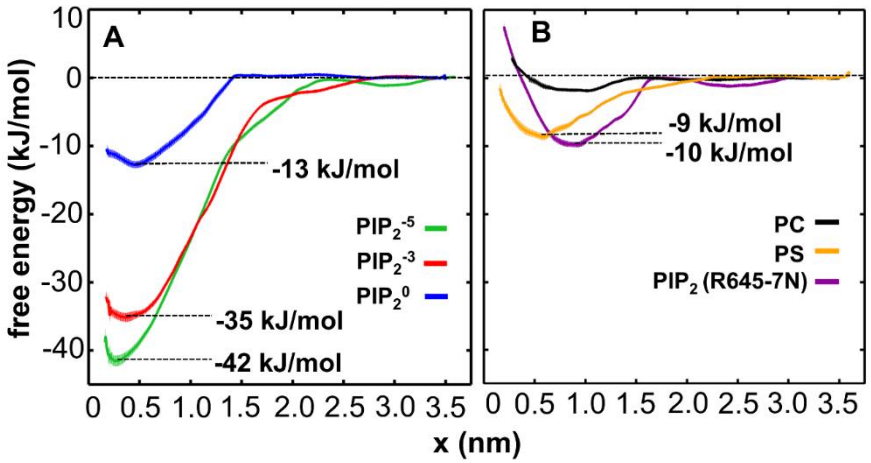

Figure 3. 10: Inner leaflet free energy profiles. (A) Free energy profiles for PIP_2^{-5}, PIP_2^{-3}, and PIP_2^{0} interactions with the WT protein. (B) Profiles for PS and PC interactions with the WT protein and PIP_2 interactions with the R645–7N mutant. The standard deviation estimated from bootstrapping is shown as the shaded area behind the curve.

When judging the utility of PMF calculations, it is important to consider their convergence [245]. The convergence of the PIP_2^{-5} PMF calculations was judged by comparison of energy profiles calculated from multiple nonoverlapping segments of simulation time (Figure 3.11A). The well depth oscillated around −42 kJ/mol, with no overall trend during the 1 μs simulation window (30 × 1 μs for the whole profile). The efficient convergence seen in these profiles is likely a function of the CG model employed, as the reduction in degrees of freedom [189] allows the system to attain convergence on faster time scales. The approach was found to be reproducible to within 5 kJ/mol (i.e., ca. 2 kT) by comparison of three independently calculated PMFs of the PIP_2^{-5} system (Figure 3.11B). The direction of the force applied in the initial steered MD simulation (i.e., whether the 1D reaction coordinate was generated by pulling the lipid away from the protein or toward it) was also tested and did not show any significant influence on the profile obtained (Figure 3.11C), indicating the absence of a hysteresis.

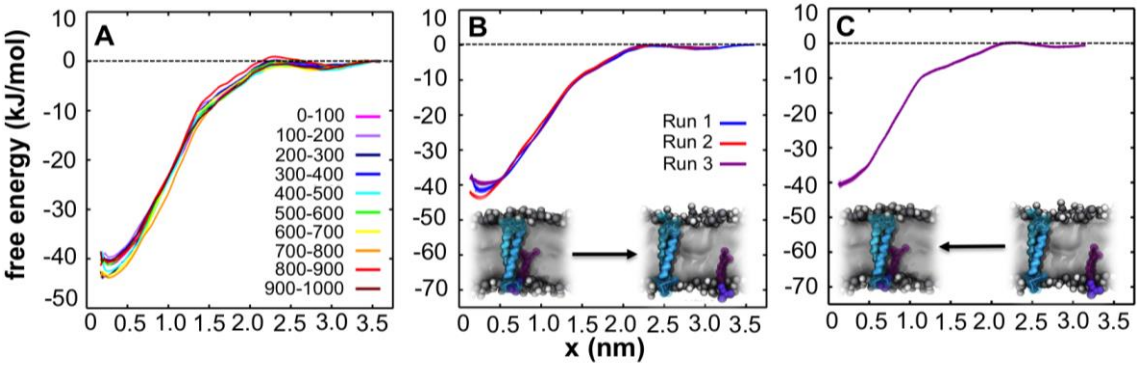

Figure 3. 11: Convergence of the free energy landscape. (A) Free energy profiles for the interaction of PIP_2^{-5} with the WT dimer, each calculated from nonoverlapping 100 ns segments of the simulation time within each 1 μs window. (B) Three PIP_2^{-5} profiles, each calculated from an independent repeat. In each case, a different structure was extracted from an equilibrium simulation, and a new set of SMD simulations and umbrella sampling calculations was conducted using that structure as the starting point. The well depths span a range of ca. 5 kJ/mol. (C) Profile obtained by pulling a PIP_2^{-5} lipid from the bulk membrane "backward" along the 1D coordinate toward the protein during an SMD simulation and using that trajectory to define the reaction pathway for a new set of umbrella sampling windows.

In addition to more specific protein−lipid interactions, we have previously revealed lipid clustering induced by disordered JM regions in simulations of the monomeric TM−JM domains of RTKs [32] and cytokine receptors [255]. Comparable clustering of anionic

lipids was seen in the current simulations of the dimeric EGFR TM−JM (Figure 3.12). Spherical radial distribution functions (RDFs) for each lipid headgroup relative to the protein exhibited peaks at 0.5 nm, a second peak extending out to 1 nm, and a third weaker peak at 1.5 nm, indicating a degree of local lipid reorganization and enrichment of certain lipid species around the protein dimer. In the outer leaflet, the RDF revealed clustering of GM3; in the inner leaflet, clustering of PIP_2 is seen. The degree of local clustering (i.e., the first peak in the RDF) is greater for PIP_2 than for GM3, reflecting the stronger electrostatic interactions for the underlying interaction, as seen in the free energy profiles.

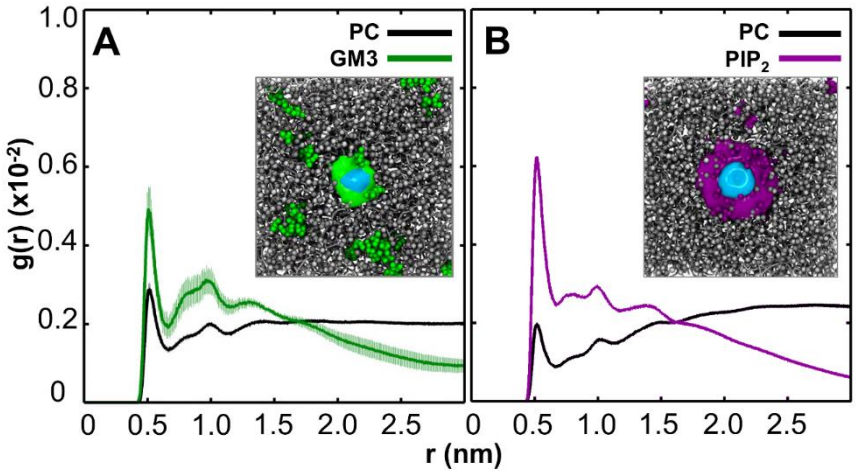

Figure 3. 12: Clustering of GM3 and PIP_2 lipids. Clustering of (A) GM3 in the outer leaflet and (B) PIP_2 in the inner leaflet around the TM + JM dimer. In each case, the radial distribution function (RDF; g(r)) for the headgroup of PC (black) and of GM3 (green) or PIP_2 (purple) is shown. Calculations were performed over 5 × 2 μs of CG simulation, with the standard deviation indicated as the shaded area behind each curve. The area under each curve is normalized to unity to aid comparison between the different lipid species. The two insets are density maps showing the average spatial occupancy of each lipid headgroup over the outer or inner leaflet surface. GM3 is indicated in green, PIP_2 is in purple, and lysine or arginine density is in cyan. Calculations were performed over 5 × 2 μs simulation repeats. In these images, the protein has been omitted to show the lysine and arginine density maps more clearly.

3.3.4 Limitations of the model

It is important to consider the possible limitations of the model employed. In particular, one might expect that inclusion of the extracellular and intracellular domains of the EGFR (along with the glycans that decorate the extracellular domain [256]) to exert some effect on the free energy landscape of lateral lipid interaction with the TM domain. However, the use of RTK TM domain constructs and mutational studies within these systems have previously been shown to be relevant to the full-length receptor [257]. Furthermore, by

its nature, the CG MARTINI model involves a degree of approximation arising from loss of atomic detail and is likely to cause deviation from binding free energies estimated from AT simulations and/or measured experimentally. In particular, some local changes in secondary structure might be induced upon binding of regulatory lipids. The standard protein backbone dihedral restraints employed within the MARTINI model [194] and the lack of hydrogen-bond directionality arising from coarse-graining the system mean that possible lipid-induced conformational changes are not be captured within the current calculations. Along the same line, it is possible that protein and lipid protonation states may change as a function of the reaction coordinate. To capture such effects, constant-pH simulations would be required [258]. In lieu of modeling dynamic protonation, some indication of the possible effects of protonation state may be provided by the calculation of PIP_2 free energy profiles for three different ionization states, along with neutralizing protein and lipid mutagenesis reported in this chapter. The standard MARTINI water model is likely to oversimplify the role of water in these interactions, particularly for lipids with large solvent-exposed headgroups such as GM3. Polarizable water models [207] have been suggested to improve the accuracy of free energy calculations for partitioning amino acid side chains into hydrophobic environments. However, there remains an incomplete understanding of its influence on lipid–water and protein–water interactions, together with a substantial increase in the convergence time associated with its use (see, e.g., ref [259]).

Despite these limitations, it should be noted that the model has been shown to correctly predict lipid interaction sites on a number of membrane proteins [22], and also that the MARTINI force field may be considered to be lipid-oriented as it was parametrized based on the reproduction of experimental partitioning free energies and atomic simulation parameters. For example, free energy profiles for phospholipid removal from a DPPC bilayer show reasonable agreement in all-atom and CG simulations, which have respectively been used to obtain estimates of ca. −80 and −90 kJ/mol [189, 260]. These values may in turn be compared with an experimental estimate of ca. −70 kJ/mol derived from the critical micelle concentration [243]. Turning to proteins, free energy profiles for the dimerization of the TM domain of glycophorin A have yielded well-depth estimates of −30 to −40 kJ/mol [218, 261] in MARTINI CG and of −45 to −60 kJ/mol [262, 263] in all-atom simulations. Considering the differences in bilayer composition, helix length, and the errors associated with both methods, this suggests that broad agreement is seen between the two levels of granularity for both lipid–

lipid and protein–protein interactions. Taking this together with a significant reduction in the sampling challenges that are routinely faced in AT simulations [245], it is expected that the current CG free energy profiles provide a reasonable first approximation of EGFR–lipid interaction free energies and of the relative effects of protein mutation.

3.4 Discussion

These simulations demonstrate that CGMD can reveal both the strength and specificity of the interactions of regulatory lipids with the TM domain of a RTK. It is shown that K618, R645, R646, and R647 play an important role in EGFR interactions with modulatory lipids GM3 and PIP_2. Comparison to experimental mutations confirms that CG simulations are indeed capable of identifying lipid interaction hotspots on integral membrane proteins. The significant (ca. 4-fold) difference in receptor interaction energies of modulatory lipids GM3 and PIP_2 and the differences in the dynamics of these interactions are suggestive of possible differential modes of receptor modulation. To date, these lipids have been proposed to modulate EGFR activity *via* influences on dimerization propensity [238], direct conformational stabilization and orientation of intracellular JM and kinase regions [264, 265], and larger scale effects on receptor clustering at the cell surface [266]. RTK helix dimerization energies have previously been estimated to be on the order of ca. −20 kJ/mol for EGFR [267] to ca. −60 kJ/mol for EphA1 [228]. The strength of interactions of PIP_2 lipids revealed in the current simulations (ca. −40 kJ/mol) is supportive of the notion that lipid modulation of EGFR activity could occur *via* direct conformational stabilization and effects on EGFR helix dimerization rather than solely *via* effects on receptor clustering [266]. In contrast, the observed weaker and more transient interactions of GM3 may be consistent with a more raft-like model of receptor modulation, which requires only weak lateral interactions between receptors and lipids [268]. This is consistent with observations that GM3 inhibits EGFR activity only within lipid-disordered/lipid-ordered (i.e., raft-like) membranes containing sphingomyelin and cholesterol [238] and not in lipid-disordered phase bilayers. Characterization of the underlying energetics of EGFR–lipid interactions is thus an essential first step toward a mechanistic understanding of these interactions. In particular, a key prediction arising from these results is that the comparable magnitudes of PIP_2 binding energies and those of helix dimerization indicate that PIP_2 lipids may be able to modulate receptor dimerization by direct competition with TM helices, as suggested by, e.g., Stangl and Schneider [269]. This could occur both by influencing the relative monomer/dimer

populations of the EGFR and by driving possible conformational transitions within existing dimers, as hinted at by recent structural studies of a presumed inactive state of the TM domain [270]. This prediction could be tested either *via* reconstitution of the EGFR in membranes of defined lipid composition [238] or *via* modulation of the lipid composition of the membranes of living cells [271]. Depending on the conformation the extracellular domain adopts and the type of glycosylation, a number of glycans can be positioned spatially close to the N-termini of the helices with which GM3 interacts [256]. It therefore seems likely that in certain states these glycans may influence lateral lipid interaction with the TM domain. Further combined simulation and experimental studies would be welcome in elucidating the influence of glycans and their relationship with regulatory lipid species and the TM domain. Ultimately, structural studies may reveal lipid-binding sites at high resolution and provide further insight into the mechanism of lipid modulation.

3.5 Conclusions

In summary, this well-studied model system demonstrates that calculation of free energy profiles for lipid interactions *via* CG simulation represents a powerful generalizable tool for the dissection and prediction of protein–lipid interactions and selectivity within cell membranes. A comparable approach has been previously applied to probe the specific binding of cardiolipin to cytochrome C oxidase [241] and, more recently, for the binding of other lipid species [250]. Taken together, these studies suggest that the CG lipid PMF methodology may be generally applicable to a wide range of membrane proteins and lipid types. This approach provides important biophysical quantification of the relative strengths of lipid interaction at specific binding sites and thus complements structural insights from crystallographic [272], computational [273], and biophysical [13] approaches.

"Real knowledge is to know the extent of one's ignorance."

- Confucius

4

Cardiolipin interactions of the ADP/ATP Translocase in model mitochondrial membranes

Contents

This chapter is based upon the manuscript:

"Lipid-Loving ANTs: Molecular simulations of cardiolipin interactions and the organisation of adenine nucleotide translocase in model mitochondrial membranes." George Hedger, Sarah L. Rouse, Jan Domanski, Matthieu Chavent, Heidi Koldsø, Mark S. P. Sansom. Biochemistry, 2016, 55, 6238-6249.

4.1 Introduction

The mitochondrial adenine nucleotide translocase (ANT; also known as the ADP/ATP carrier, AAC) facilitates export of ATP outwards across the inner membrane into the cytoplasm, in exchange for import of ADP back into the matrix [274]. This is driven by the electrochemical gradient across the inner membrane. The translocase is one of the most abundant proteins in the mitochondrial inner membrane [275]. Its abundance and the importance of this transporter has rendered it one of the best characterised members of the mitochondrial carrier (MC) family, which facilitate the movement of a range of metabolites in and out of mitochondria [276].

The translocase has an intimate relationship with a key lipid in mitochondrial membranes, cardiolipin (CL). The presence of CL is essential for maximal stability of the translocase *in vitro*, and CL has been suggested to modulate interactions of ANT with other proteins [23, 277]. Knock down of CL synthesis in humans results in Senger's syndrome, a condition in which the carrier is depleted from the inner membrane, leading to impaired oxidative phosphorylation and symptoms ranging in severity from mild difficulties with exercise to neo-natal fatality [278, 279].

To date, atomic resolution structures of bovine and yeast forms of ANT have been determined in a cytoplasmic facing state [33, 280]. An NMR structure has been determined for another MC family member, uncoupling protein 2 (UCP2) [281]. These translocase structures revealed a fold in which six transmembrane helices (H1-6) form a bundle around a central substrate binding cavity, abutted on the matrix side by three amphipathic helices (MH1-3) arranged around the outer rim of the helix bundle (Figure 1). There is an internal pseudo-three fold symmetry, seen also in the UCP2 structure, which sequence comparisons suggest to be conserved across the MC family. The presence of CL was essential for reconstitution and crystallisation of ANT, and CL molecules have been resolved in all crystal forms of the translocase, bound between the short matrix

helices. This agrees with earlier ^{31}P NMR measurements which indicated tight binding of CL molecules to bovine heart translocase [154]. Although CL molecules were not resolved in the structure of UCP2, the presence of this lipid was essential to obtain interpretable spectra [281].

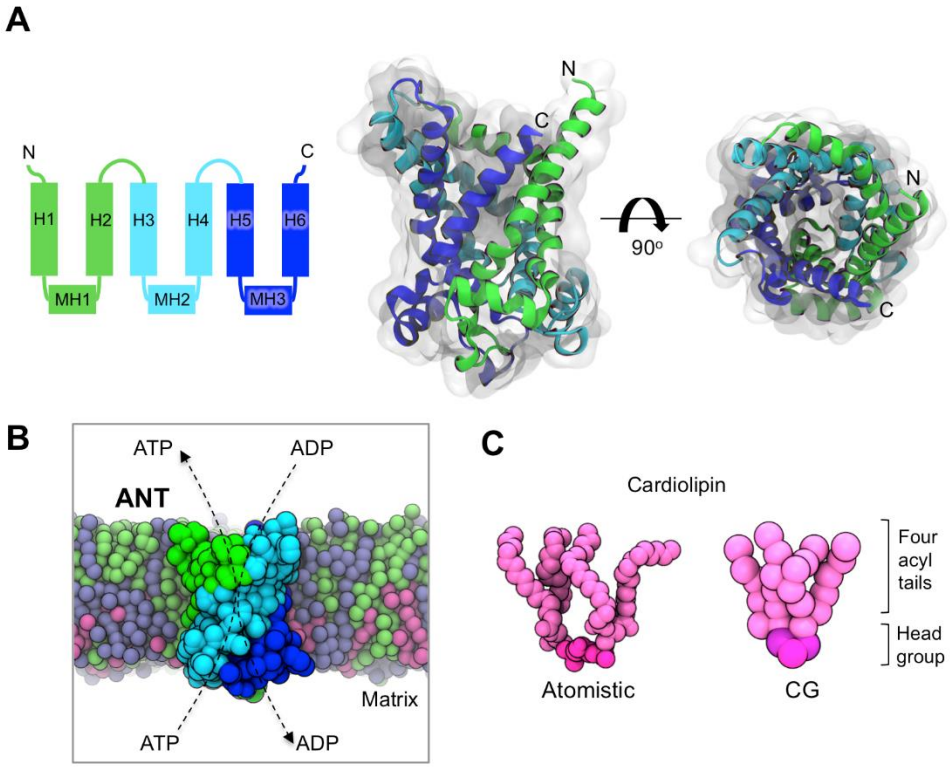

Figure 4. 1: ANT1 and CL. (A) Topology and structure of bovine ANT1 colored to indicate the internal three-fold repeat symmetry. The structure shown corresponds to the cytoplasmic facing state of ANT1 (PDB id: 1OKC) [33]. (B) A coarse-grained (CG) representation of ANT1 embedded in a lipid bilayer. (C) Atomistic and CG models of CL.

Despite the importance of CL in the structure and function of the ADP/ATP translocase and other MC family proteins, characterisation of these interactions *within a membrane environment* is relatively scarce. Here multiscale MD simulations are applied to investigate the binding of CL to ANT and to UCP2 in model mitochondrial membranes. Structural, energetic, and dynamic aspects of lipid interactions with a single copy of each protein are assessed, before the interplay of protein-lipid and protein-protein interactions in large scale simulations of membrane patches containing 25 copies of the ANT molecule is evaluated.

4.2 Methods

4.2.1 Coarse-grained simulations

All simulations were performed using the GROMACS 4.6 (www.gromacs.org) simulation package [179] and the MARTINI (http://md.chem.rug.nl/) force field [189, 194, 203]. A summary of the simulations performed is provided in Table 4.1.

Table 4. 1: Overview of the simulations performed.

Description	Membrane composition	Duration
1 x bovine ANT1[a], CG	Outer leaflet: PC(100%) Inner leaflet: PC(90%)/CL(10%)	5 x 6 μs
1 x bovine ANT1[a], CG	Outer leaflet: PC(50%)/PE(50%) Inner leaflet: PC(42.5%)/PE(42.5%)/CL(15%)	5 x 1 μs
1 x yeast Aac2p [b], CG	Outer leaflet: PC(50%)/PE(50%) Inner leaflet: PC(42.5%)/PE(42.5%)/CL(15%)	5 x 1 μs
1 x UCP2[c], CG	Outer leaflet: PC(50%)/PE(50%) Inner leaflet: PC(42.5%)/PE(42.5%)/CL(15%)	5 x 1 μs
PMF calculations, 1 x ANT1[a], CG	PC (100%)[d]	32 x 1- 2 μs per PMF
25 x bovine ANT1[a], CG	Outer leaflet: PC(60%)/PE(40%) Inner leaflet: PC(50%)/PE(40%)/CL(10%)	20 μs
25 x bovine ANT1[a], CG	Outer leaflet: PC(60%)/PE(40%) Inner leaflet: PC(60%)/PE(40%)	20 μs
1 x bovine ANT1[a], AT	Outer leaflet: PC(50%)/PE(50%) Inner leaflet: PC(42.5%)/PE(42.5%)/CL(15%)	0.5 μs

Simulations were based on the crystal structures of [a]bovine ANT1 (PDB: 1OKC [33]), [b]yeast Aac2p (PDB: 4C9G [282]) and of [c]UCP2 (PDB: 2LCK [281]). [d]See the Potential of Mean Force Calculations & Methods sections for further details regarding membrane composition.

Protein PDB coordinate sets were processed using MemProtMD [149] to remove co-crystallised lipids, waters, and carboxyatractyloside inhibitors. CG simulations used the MARTINIv2.1 force field [189]. Simulations of ANT utilized an ElNeDyn elastic

network [206]. Each protein was placed in a box of dimensions 12 x 12 x 10 nm^3, containing 460 randomly orientated PC lipids. The system was solvated using standard MARTINI water particles, and neutralised with ~0.15 M NaCl. A 50 ns self-assembly simulation [51] was performed to allow formation of a lipid bilayer around the protein. PC molecules within the bilayer were subsequently exchanged with other lipid species to form an asymmetric bilayer of specified lipid composition as described previously [58]. For repeat simulations the lipids were independently exchanged such that each repeat contained a different initial distribution of lipid molecules consistent with the same specified lipid composition. CL parameters were from [283]. CL was modelled with a net ionization state of -2*e* (i.e. -1*e* per phosphate group) based on experimental estimates [284]. The acyl tails of PC and PE lipids were modelled as 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) and 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (POPE), whilst CL was modelled as 1,3-bis(1-palmitoyl-2-oleoyl-*sn*-3-phosphatidyl)-*sn*-glycerol.

Extended systems containing 25 copies of the ANT1 molecule were generated using the GROMACS *genconf* tool to concatenate a system containing a single ANT1 molecule embedded in a PC bilayer onto a 5 x 5 grid (Figure 4.2).

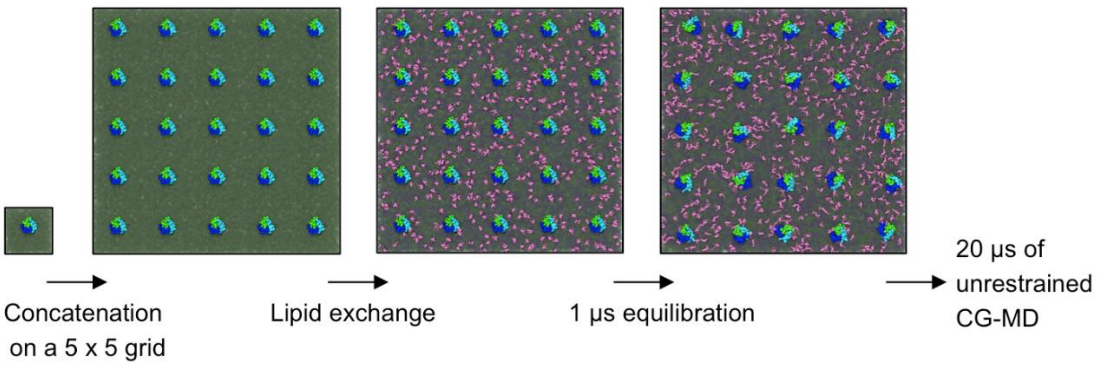

Figure 4. 2: Assembly of large membrane patches. Small systems consisting of a single copy of bovine ANT1 embedded in a POPC bilayer were concatenated onto a 5 x 5 grid using the GROMACS *genconf* tool[179], forming a 65 x 65 nm^2 membrane patch. The lipids of the system were subsequently exchanged, and the resultant system was resolvated, reionized and energy minimized[184]. A 1 µs equilibration was performed to randomize protein orientations and equilibrate lipid interactions, during which the proteins were translationally restrained but allowed to rotate freely. This was followed by 20 µs of unrestrained CG-MD.

The lipids of the energy minimized concatenated PC membrane were then exchanged for a model mitochondrial membrane as described above [58]. The resultant system dimensions were 65 x 65 x 14 nm^3, containing 25 proteins and ~11,000 lipids, solvated by standard MARTINI water and neutralised by ~0.15 M NaCl. The systems contained ca. 4.7 x 10^5 particles. An initial 1 μs equilibration period was performed to facilitate equilibration of lipid-protein interactions and facilitate randomisation of protein orientations. During this period the backbone particle of a single proline residue in each ANT1 molecule (Pro27) was restrained in the XY plane, thus allowing rotation and randomisation of protein orientations, but preventing their translational movement. Both with and without CL systems were then run for 20 μs of unrestrained production run simulation (see Table 4.1).

Simulations were performed under NPT conditions at a temperature of 310 K and pressure of 1 bar with periodic boundary conditions. Temperature and semi-isotropic pressure were controlled using the Berendsen thermostat and Berendsen barostat [178] with coupling constants of 4 ps. Equations of motion were integrated using the leapfrog algorithm with a timestep of 20 fs. Electrostatics were modelled as reaction field with a Coulomb cut-off of 1.2 nm, using a potential-shift modifier. Van der Waals interactions were smoothly shifted off between 0.9 to 1.2 nm. The LINCS algorithm [172] was employed to constrain covalent bonds to their equilibrium values.

4.2.2 Potential of mean force calculations

PMFs were calculated *via* umbrella sampling, utilizing a protocol previously described [285]. A 2 μs production run was performed with bovine ANT1 (PDB:1OKC) [33] embedded in a PC bilayer containing 10% CL within the inner leaflet. For each of the three sites, a snapshot containing a bound CL/PC molecule was extracted. A new PC bilayer was subsequently assembled around the position restrained (F_c=1000 kJ/mol/nm^2) protein and lipid of interest during a 50 ns self-assembly simulation [51]. This simple composition was chosen to accelerate convergence of calculations and has previously been shown to cause no significant deviation compared to protein-lipid PMFs conducted in two-component PC:PE bilayers [250]. A 1D reaction coordinate ranging from the lipid bound state to the lipid unbound state was then generated *via* a steered MD (SMD) simulation. Each CL/PC molecule was pulled away from the protein over a distance of 3-4 nm at a rate of 0.1 nm/ns (F_c=1000 kJ/mol/nm^2). The force was applied to the GL0

headgroup particle of CL, and the PO4 headgroup particle of PC. Weak positional restraints (F_c=400 kJ/mol/nm^2) were also applied in the XY plane to the backbone particle of each proline residue within each of the three Px[D/E]xx[K/R] motifs of the protein. Application of such restraints acted to prevent rotation of the protein and translational 'following' of the tight binding CL molecules as they were pulled away. In addition, weaker positional restraints (F_c=50 kJ/mol/nm^2) were applied to the GL0/PO4 particle of the CL/PC lipid in the Y direction. This served to retain the lipid on the 1D reaction coordinate, and thus enhance sampling of the chosen pathway. Snapshots spaced by ~0.1 nm were extracted from the SMD simulation and used as input for umbrella sampling simulations. Each window of the umbrella sampling was run for between 1-2 μs, with umbrella biasing potentials (F_c=1000 kJ/mol/nm^2) applied between the centre of mass of the three restrained proline residues and the GL0/PO4 particle of the CL/PC lipid. Within umbrella sampling simulations, long range electrostatics were modelled using Particle Mesh Ewald (PME) [180]. The subject lipid was treated separately to bulk lipids for temperature and pressure coupling. Approximately 32 umbrella sampling simulations were run per PMF. PMF profiles were constructed using the GROMACS implementation (g_wham) of the weighted histogram analysis method (WHAM) [226]. Bayesian bootstrapping with 200 bootstraps was used to estimate the errors for each profile. Convergence was assessed by comparing profiles calculated from independent 100 ns segments of simulation time (Figure 4.3), as well as by calculating profiles *via* replica exchange initiated from different starting configurations [216].

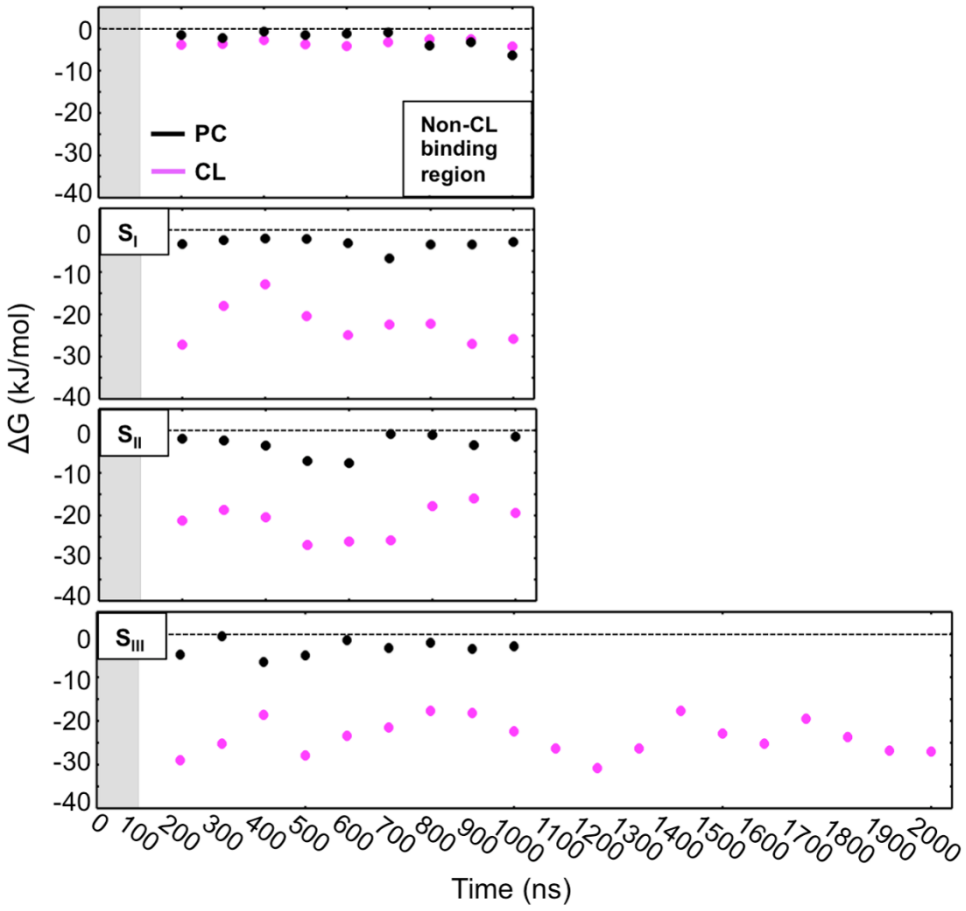

Figure 4. 3: Convergence of PMF calculations. Convergence of the calculations was assessed by calculating PMF profiles for multiple, non-overlapping 100 ns segments of simulation time. The first 100 ns of each window (i.e. 0-100 ns) was excluded as equilibration. Extending the simulated time from 1 to 2 μs for $Site_{III}$ revealed no overall trend in the value of the profile well-depths.

The reproducibility of this protocol has been tested previously for EGFR-PIP_2 interactions, and it was found the well-depths initiated from independent repeats spanned a range of 5 kJ/mol (i.e. 2RT) [285]. This was tested for the present system by conducting two independent repeats of the entire calculation for $Site_I$, along with two repeats utilizing different lateral lipid restraints (Figure 4.4). The well-depths agreed to within 5 kJ/mol, suggesting this range is applicable to the current system, and indicating the robustness of the protocol to the choice of initial snapshot and lipid restraint.

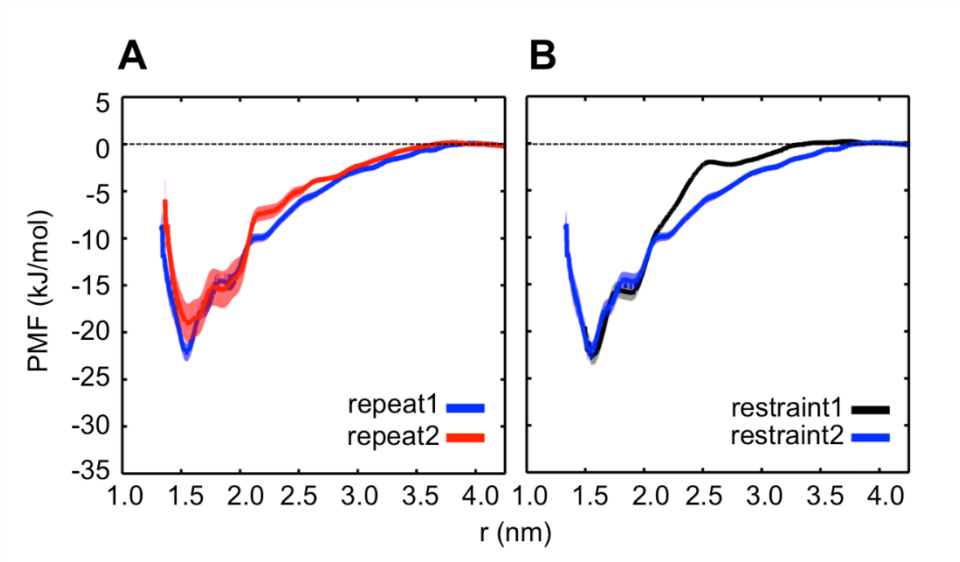

Figure 4. 4: Repeat and restraint dependency for $Site_I$ PMF calculations. (A) Two PMF profiles for CL interactions with $Site_I$, each calculated from independent repeats of the simulation protocol, starting from different initial snapshots. (B) Two profiles calculated using different lateral y-restraints on the lipid during umbrella sampling. Blue (F_c=50 kJ/mol/nm^2), and black (F_c=100 kJ/mol/nm^2).

4.2.3 Atomistic simulations

Atomistic simulations employed the GROMOS96 53A6 force field [166]. A final snapshot of a 1 μs CG simulation was converted to atomistic detail using a fragment based protocol [44]. The system was solvated using the SPC water model and neutralised with ~0.15 M NaCl. A 1 ns equilibration simulation was performed with protein Cα particles restrained (F_c=1000 kJ/mol/nm^2). This was followed by 0.5 μs of unrestrained simulation (see Table 4.1). Temperature was controlled at 310 K using a V-rescale thermostat [176] with a coupling constant of τ_t = 0.1 ps. Pressure was maintained at 1 bar using a Parrinello-Rahman barostat [177] with a compressibility of 4.5 x 10^{-5} bar^{-1}, and coupling constant of τ_p = 1 ps. Electrostatic forces were modelled using Particle Mesh Ewald (PME) [180]. Van der Waals interactions were cut-off at 1.2 nm. A timestep of 2 fs was employed and covalent bond lengths were constrained using the LINCS algorithm [172].

4.2.4 Analysis and visualization

2D density maps for lipids and proteins were calculated as described previously [286]. 3D lipid density isosurfaces were computed using the VolMap plugin of VMD [249]. Diffusion coefficients were extracted from the mean square displacement (MSD) using

g_msd and a time period of 20 ns. Contact and protein clustering analysis was performed using in-house scripts and MDAnalysis [287]. Lipid contacts were calculated between each residue of the protein and the head group particle of CL (defined by GL0) using a cut-off of 0.6 nm to define a contact. A cut-off of 4.4 nm between protein centres of mass was used for protein clustering analysis, while per residue protein-protein contacts were calculated utilizing the approach described in [286]. The VMD [249] software package was used for visualization of simulation data.

4.3 Results

4.3.1 Identification of CL Binding Sites on Bovine ANT1

A CG molecular model of bovine ANT1 was embedded into a simple model membrane containing PC(90%)/CL(10%) in the matrix facing leaflet, and PC(100%) in the cytoplasmic facing leaflet. This CL content lies within the range estimated for mitochondrial inner membranes from lipidomics [56, 288, 289]. Five repeat simulations of bovine ANT1 were conducted, each of 6 μs duration and beginning from a different distribution of lipids within the bilayer (see Table 4.1 and Methods for details).

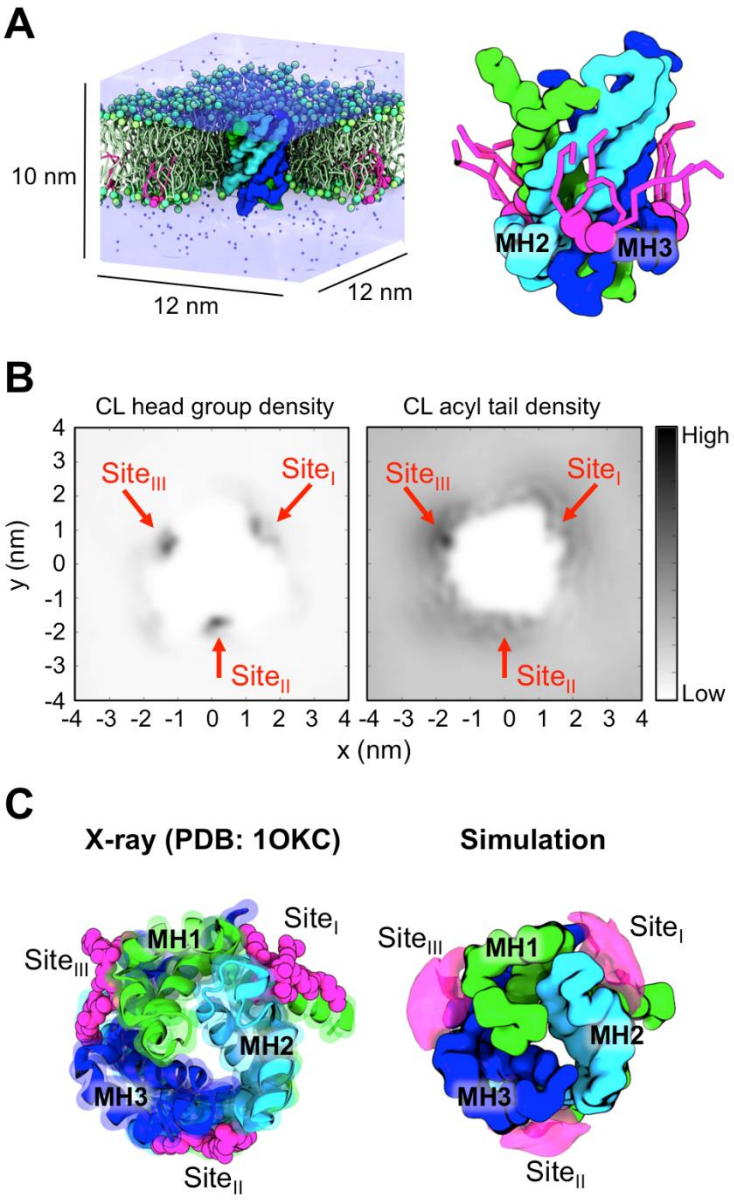

Figure 4. 5: CL binding sites on bovine ANT1. (A) Left: The initial system simulated consisted of a single molecule of bovine ANT1 embedded in a mixed PC (lime) and CL (magenta) bilayer, solvated with standard MARTINI water particles (transparent surface) and neutralized with ~0.15 M NaCl (blue spheres). Right: Final snapshot of a 6 μs simulation showing three bound CL molecules (magenta). (B) Time-averaged 2D density maps of CL in the membrane plane for the head group (left) and acyl tail moieties (right). (C) View onto the base of the translocase from the matrix showing the positing of the three CL (magenta) binding sites around the protein within the X-ray structure (left), and CG simulations (right). The CL binding sites within the CG simulations are shown by the time averaged probability density surface for the CL molecules, contoured to show the three binding sites. The density surface was calculated from 5 x 6 μs simulations repeats, each starting from different random lipid distributions. See Appendix Figure B1 for per simulation density maps.

CL molecules were seen to encounter and bind to three specific sites on the protein ($Site_{I-III}$: Figure 4.5). Each site was composed of a groove on the protein surface formed by the N-terminal end of the even numbered transmembrane helices (H2, H4, H6) of each repeat, and the short amphipathic matrix helices (MH1-3). The location of the sites thus possesses a symmetry related to the internal three-fold structural repeat. The location of these sites is in excellent agreement with those identified in X-ray crystal structures [33, 280]. This indicates that the sites occupied in the crystal structure persist within a membrane environment. Interactions were predominantly mediated by the head group moieties of CL, as seen in the time-averaged 2D density maps for CL (Figure 4.5B). In contrast the acyl tails exhibited more dynamic interactions, which resulted in less well defined density, though a small degree of ordering was seen at $Site_{III}$. This is consistent with a number of crystal structures in which only the head group moieties and adjacent atoms of the acyl chains were resolved [282]. The dominance of head group interactions seen here is also consistent with functional assays performed in yeast showing Aac2p assembles normally regardless of the acyl chain composition of CL [290].

The number of contacts formed between each residue of the protein and the CL headgroup were calculated over the course of each simulation. Within each site, the CL headgroup showed a preference for interaction with the N-termini of the even numbered TM helices and matrix helices. In $Site_{I}$, the highest degree of contact was seen with S68, G72, T154 and G155. Residues showing high levels of contact in $Site_{II}$ were Q174, G175, G252 and T253; whilst in $Site_{III}$ G52, I53, G272, and W274 (Figure 4.6A). A number of the interacting residues form part of the highly conserved [YWF][RK]G and [YF]xG motifs [275], respectively located at the N-termini of the even numbered TM helices and matrix helices (Figure 4.6B). In particular, the glycine residues within these motifs were seen to form frequent contacts with CL in the simulations. The co-localisation of CL interactions to these conserved sequence motifs is encouraging, given also the suggested role of the glycine residues [275] within the motifs in mediating CL interactions within crystal structures [280]. PC lipids showed comparatively less specificity for these regions (Appendix Figure B2). To assess the robustness of the results to the presence of PE lipids (present *in vivo* [56]) and changes in CL concentration, 5 x 1 μs simulations were conducted in mixed membranes with a composition of: PC(50%)/PE(50%) in the cytoplasmic facing leaflet, and PC(42.5%)/PE(42.5%)/CL(15%) in the matrix facing leaflet. The simulations yielded very similar patterns of CL contacts to the PC:CL only

membranes, suggesting the interaction is not sensitive to the presence of PE (Figure 4.6C), as has been seen for CL interactions with other mitochondrial proteins [291].

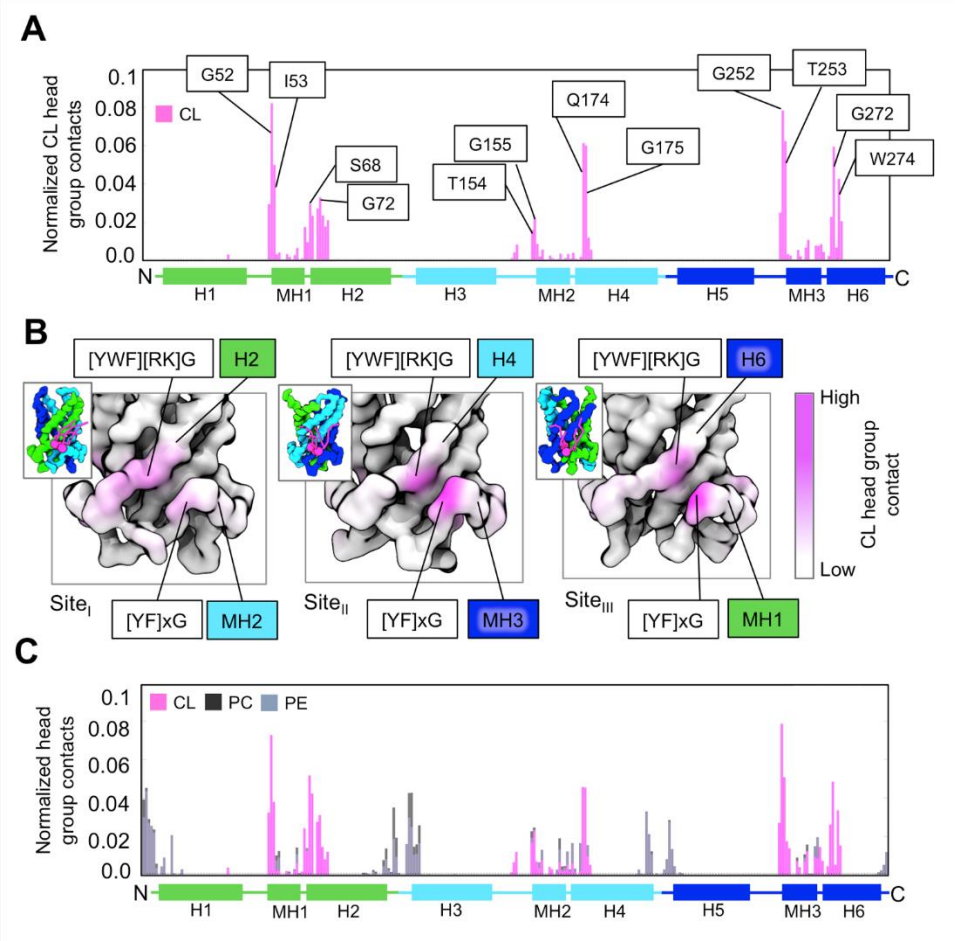

Figure 4. 6: Interactions of CL with Bovine ANT1. (A) Normalized average frequency of CL headgroup contacts per residue. Residues within each of the even numbered helices (H2, H4 and H6) and the matrix helices (MH1-3) exhibiting the highest degree of contact are labelled (See Methods for full details of contact analysis). (B) Lateral view into each CL binding site with the backbone trace of the protein coloured by the relative extent of contacts formed with CL, from white (no contact) to magenta (high contact). Insets show the corresponding view onto the whole translocase coloured according to three-fold repeat topology. The positions of the conserved motifs are labelled. (C) Normalized average frequency of phospholipid headgroup contacts per residue, for the lipid species within three-component membranes.

To further assess interactions within the identified binding sites a CG snapshot of ANT1 was converted [44] to atomistic detail and simulated for 0.5 μs. Sites I and II of the chosen snapshot were occupied by CL molecules, whilst Site$_{III}$ was unoccupied, thus allowing one to test both the stability of bound CL molecules as well as explore possible atomistic binding events within the same extended simulation. Calculation of Cα root mean square

deviation (RMSD) during the simulation revealed a stable value of ~0.3 nm for the core structural elements (i.e. helices) and ~0.4 nm for all residues, indicating that, despite the absence of the carboxyatractyloside inhibitor, the apo-cytoplasmic facing structure is conformationally stable on the timescale simulated (Figure 4.7).

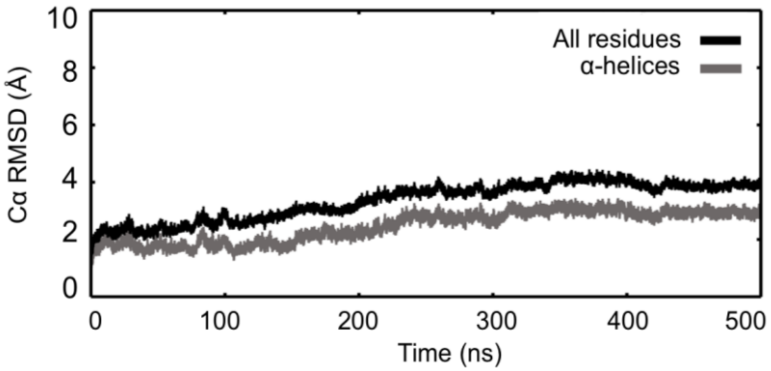

Figure 4. 7: Bovine ANT1 stability within atomistic simulation. Cα root mean square deviation (RMSD) for all residues (black) and α-helical elements (grey). The RMSD reaches a stable plateau after ca. 300 ns, which may be compared to values obtained in molecular simulation of other transporters [137].

The CL molecules within Sites I and II remained bound (Figure 4.8), with the phosphate moieties contacting those conserved motifs seen to form frequent contacts in the CG simulations. In contrast the acyl chain tails exhibited more dynamic interactions and adopted a range of conformations (Figure 4.8C). Of interest within $Site_I$, the sidechain of R71 remained pointing downwards, consistent with its position within crystal structures [280]. No lipid binding events were observed at $Site_{III}$, despite one CL being positioned in the vicinity of the site. This suggests longer timescales may be required to capture the full process of lipid binding in atomistic resolution.

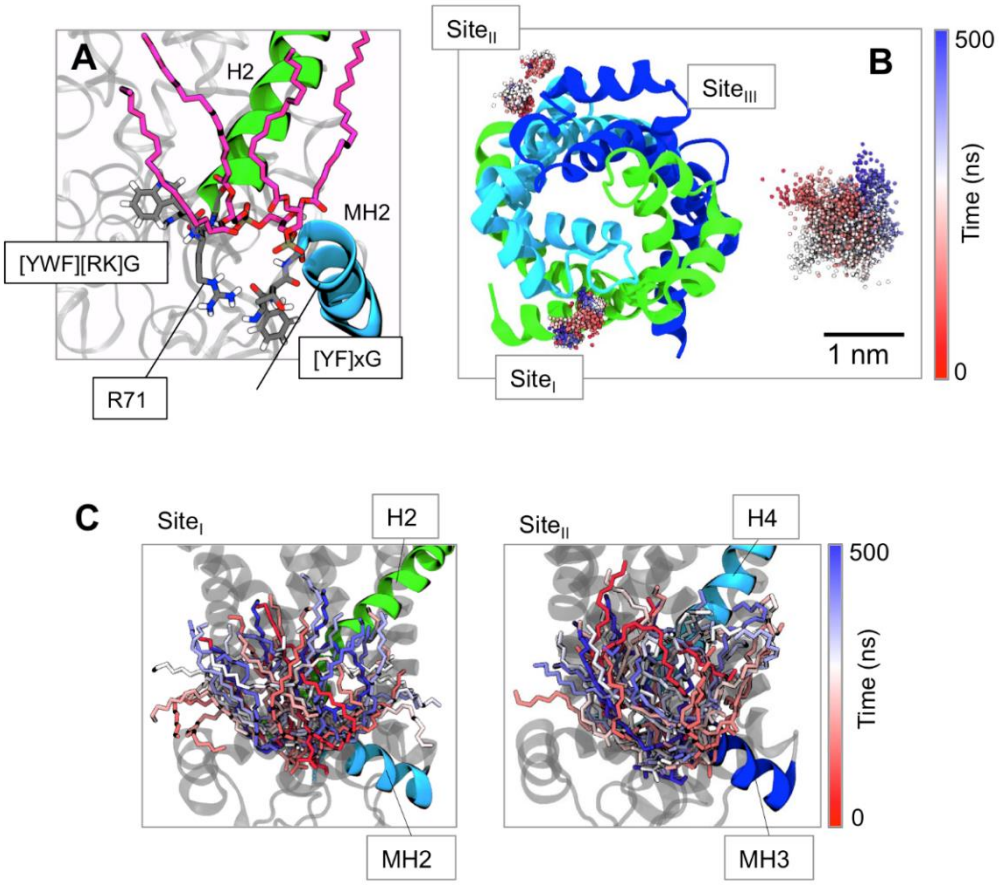

Figure 4. 8: Atomistic simulation of Bovine ANT1. (A) Final snapshot of Site$_{I}$ showing arrangement of conserved motifs (grey), N-terminal ends of the helices (green, cyan), and CL (magenta). (B) view from the matrix side, with the phosphate particles of the bound CL molecules shown as spheres. For each phosphate particle 1000 evenly distributed snapshots are shown over the 0.5 µs simulation, coloured according to simulation time. This may be compared to (right) the phosphate distribution for a non-interacting bulk-CL. The bulk CL shown is the one spatially nearest Site$_{III}$, however despite this proximity no binding events were observed over the simulated timecourse. (C) Lateral view into Sites$_{I}$ and $_{II}$ showing ten evenly distributed CL structures over the 0.5 µs timecourse. The structures are shown as sticks coloured by simulation time. The lipid tails adopt a wide range of conformations.

4.3.2 Energetics of CL Binding

To assess the strength and selectivity of CL interactions, potential of mean force (PMF) profiles for the CL/ANT1 interaction were calculated *via* CG simulation and umbrella sampling. The PMF (or free energy profile) between two species describes the change in free energy along a particular reaction coordinate, and is calculated from the probability distribution of the two species along that coordinate [292]. This approach has recently been used to define the binding of CL to mitochondrial respiratory chain complexes III and IV [47, 293], and to characterise lipid interactions of the epidermal growth factor

receptor (EGFR) transmembrane domain [285]. For each of the three sites, a steered MD simulation was conducted in which a force was applied to pull the CL molecule out of its binding site into the bulk membrane along a 1D reaction coordinate (*r*) perpendicular to the protein surface in the plane of the bilayer (Figure 4.9A). The free energy profile along this coordinate was then calculated *via* umbrella sampling, with *r* defined as the distance between the centre-of-mass of three conserved prolines within ANT1 and the β-glycerol moiety of CL.

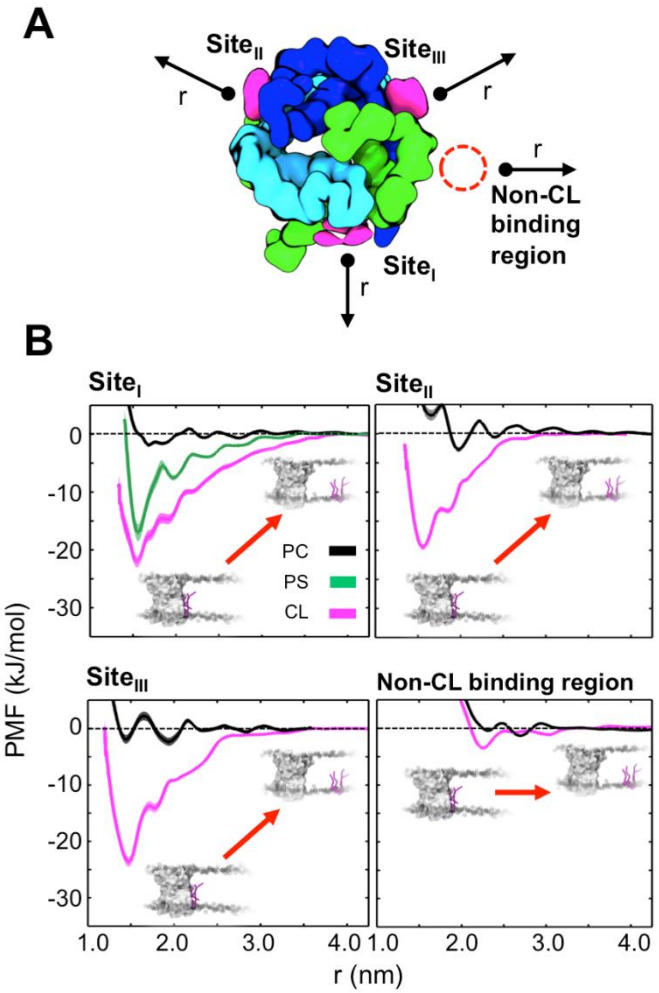

Figure 4. 9: Potentials of mean force for lipid binding to ANT1. (A) View onto the base of ANT1 illustrating the reaction coordinates used in free energy calculations. ANT1 is colored according to its three-fold repeat topology, with the position of each binding site indicated by the time averaged probability density isosurfaces for CL phosphates (magenta). The arrows indicate the approximate 1D reaction coordinates along which the PMF was calculated. (B) PMF profiles for CL (magenta) and PC (black) interactions with each of the three binding sites, and a control non-CL binding region. For Site$_I$ a PMF profile was also calculated for a further anionic lipid species, PS (green). The standard deviations estimated from bootstrapping [294] are shown as the shaded area behind each curve.

The profiles revealed comparable tight binding of CL molecules to each of the three sites, with well-depths of approximately -20 kJ/mol (Figure 4.9B). In contrast pulling a CL molecule away from a region *not* predicted to bind CL revealed a well-depth of ~ -4 kJ/mol, indicating no significant interaction at this region. Conducting the same calculations for zwitterionic PC lipids showed no significant interaction of this species beyond the value of RT (2.5 kJ/mol), demonstrating the marked selectivity of these sites for CL over zwitterionic phospholipids. These values may be compared to those obtained for cytochrome c oxidase (C_{IV}) *via* the same method [150], revealing that the binding sites on bovine ANT1 have a higher affinity for CL than most of the sites on C_{IV}, other than for two high affinity sites on C_{IV} which are estimated to bind with an interaction free energy of ~ -32 kJ/mol. Interestingly, calculation of the PMF profile for a second anionic lipid species, phosphatidyl serine (PS), at $Site_I$ revealed a well-depth of ~ -15 kJ/mol, i.e. intermediate between those of CL and of PC. Although the abundance of PS [56] within the inner mitochondrial membrane is comparatively low, these calculations suggest the capacity of the binding sites to interact with other anionic lipid species, albeit less favourably than with CL.

4.3.3 Conservation of Binding Sites

The strong binding of CL to bovine ANT1 together with the structural conservation [295] of the internal three-fold repeat of mitochondrial carriers suggests that these sites are likely to be conserved over other members of the family. This possible conservation was tested by running CG simulations of yeast Aac2p and of mouse UCP2. Multiple (5 x 1 μs) simulations were run for Aac2p and UCP2 in the same manner as for ANT1 (Table 4.1). Analysis of the average probability distribution of CL over the course of the simulations showed preferential localisation to the equivalent three sites seen for ANT1 (Figure 4.10). These positions compare well with the multiple crystal structures of bovine ANT1 [33, 280], and of yeast Aac2p and ANT3 [282]. The simulations are supportive of direct CL interactions with UCP2, and predict that these interactions occur at sites equivalent to those seen in ANT1 and Aac2p.

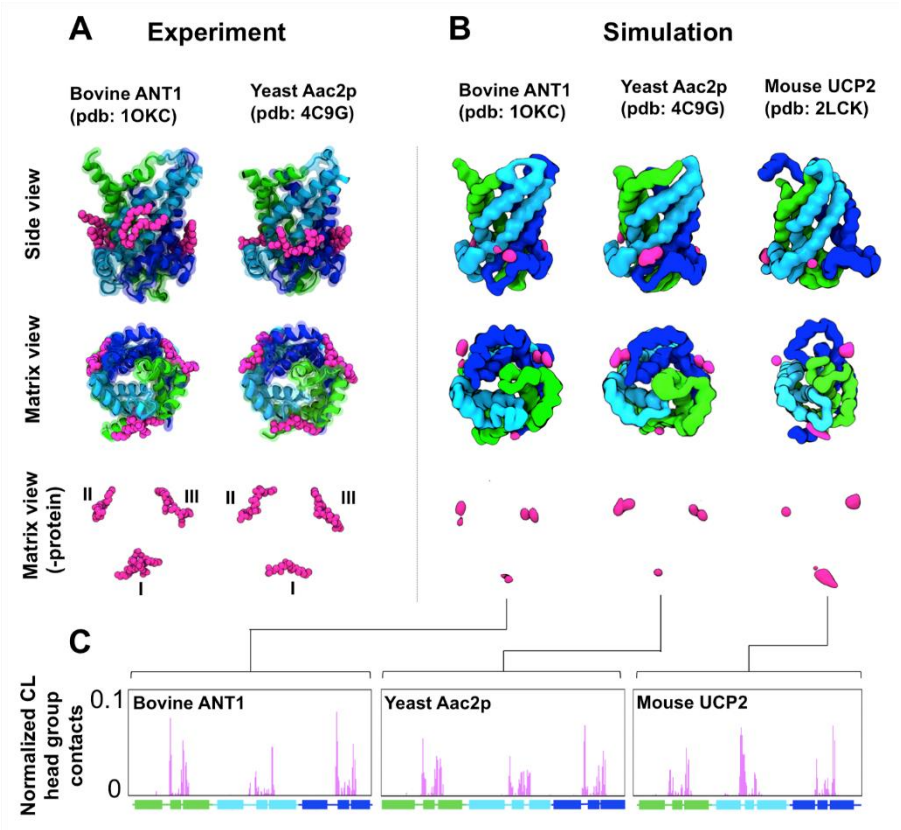

Figure 4. 10: Conservation of CL binding sites. (A) Location of co-crystallised lipids (magenta) in bovine ANT1 (PDB id: 1OKC)[33] and yeast Aac2p (PDB id: 4C9G)[282]. (B) CL binding sites identified by CG-MD simulations. The magenta surface shows the probability density for the CL headgroup phosphate particles around a snapshot of each protein, contoured to indicate the positions of the phosphate moieties of CL within each site. (C) Normalized average number of CL head group contacts per residue within simulations of Bovine ANT1, Yeast Aac2p, and Mouse UCP2.

Analysis of lipid contacts formed in the yeast Aac2p and UCP2 simulations showed spatially similar patterns to bovine ANT1 (Figure 4.10C), despite differences in sequence and the somewhat 'looser' packing of the transmembrane helices in the UCP2 structure. This suggests the groove like architecture of the sites is also an important contributor to CL binding, and that a degree of sequence substitution may be tolerated. The conservation of these sites within both crystals (obtained in the presence of detergent) and a simulated membrane environment, along with their presence in three MC proteins is supportive both of their biological significance and their likely conservation in other members of the family.

4.3.4 Lipid Interactions and Organization of Bovine ANT1 in Large Membrane Patches

Large-scale simulations of membrane patches containing multiple protein copies have recently proved useful in understanding the dynamic organisation of lipids and proteins [296]. In light of this, a single molecule of bovine ANT1 embedded in a small membrane patch was replicated (Figure 4.2) to create a large membrane patch with 25 ANT1 molecules in a ~ 65 x 65 nm^2 area of membrane containing ~11,000 lipid molecules (see Table 4.1), with a cytoplasmic facing leaflet containing PC(60%)/PE(40%), and a matrix leaflet containing PC(50%)/PE(40%)/CL(10%). This composition lies within the range estimated for the inner mitochondrial membrane from lipidomics [56, 297]. This membrane was used as the basis of 20 μs of unrestrained molecular dynamics.

Examination of the final snapshot showed a degree of protein oligomerisation, (Figure 4.11A), with a mixture of monomers, dimers, and higher order oligomers observed. Analysis of the dynamics of protein clustering showed the self-assembly of oligomers began on a sub-microsecond timescale (Figure 4.11B), with ~50% of proteins participating in an oligomeric state by 8 μs. Calculation of the averaged 2D protein density maps for ANT1 and CL showed that this interaction occurred at three specific interaction interfaces, which co-localized with the three CL binding sites identified previously (Figure 4.11C). Indeed visual inspection of the simulations showed that CL molecules were often found at these interfaces, loosely bridging between proteins. However in contrast to the crystallographic dimers [282], a stable arrangement in which two CL molecules were present between the two proteins was not observed. Such an arrangement may be unfavourable due to the proximity of the negatively charged (-2*e*) CL headgroups, as well as a degree of steric conflict. This is suggestive of possible competition between protein-protein and protein-lipid interactions. Calculation of 2D density maps and ANT-CL contacts for the first 1 μs of the simulation (ANT molecules are monomeric), and the last 1 μs (ANT molecules are predominantly oligomeric) revealed similar profiles, indicating the broad patterns of CL interaction remain similar under the two regimes (Appendix Figure B3). However further simulations with improved sampling and resolution are warranted to unpack the finer details of CL behaviour within oligomers, and assess the statistical significance of these differences.

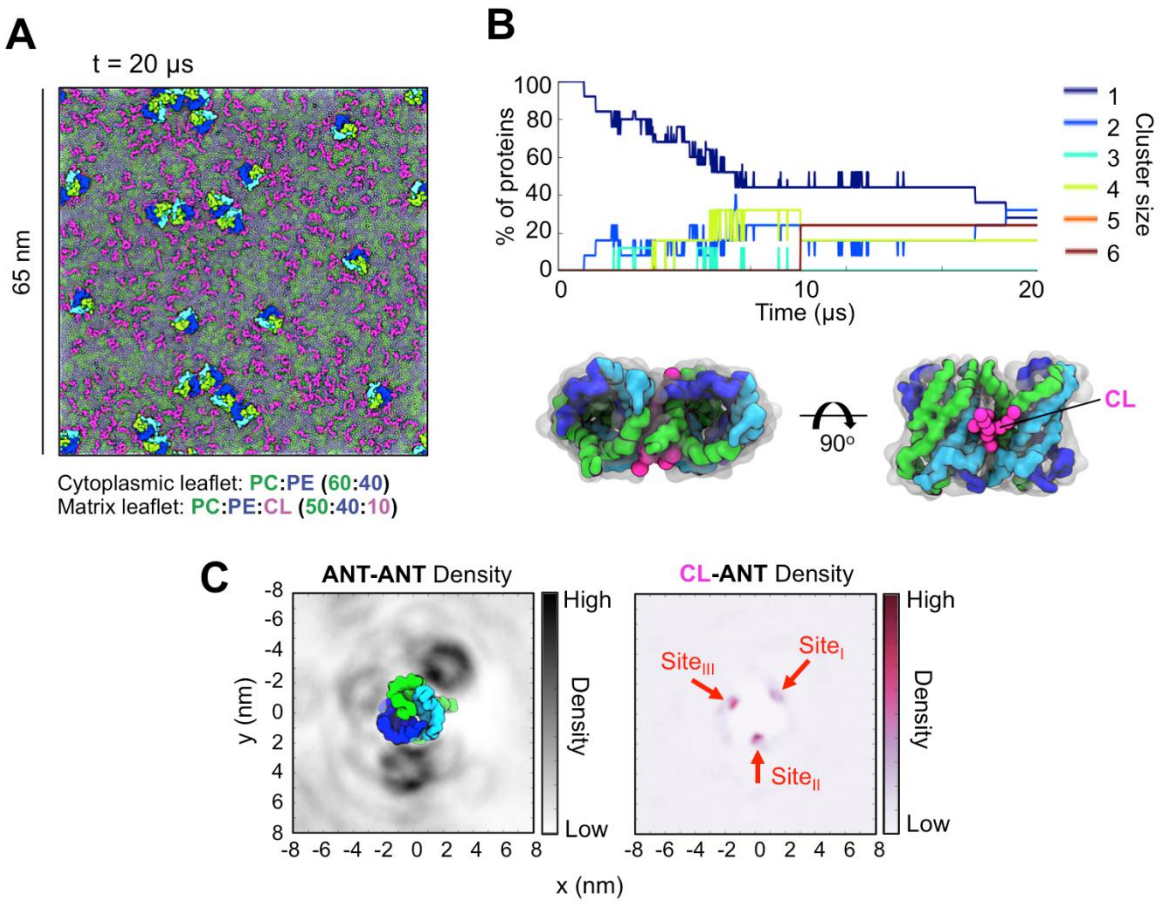

Figure 4. 11: Organization of bovine ANT1 in large membrane patches. (A) Final snapshot (t=20 μs) of the matrix leaflet showing the organization of 25 ANT1 proteins in a PC:PE:CL mixed membrane. (B) Clustering dynamics shown as the percentage of ANT1 cluster sizes over the simulated time course, along with a snapshot of an ANT1 dimer with CL mediating the interaction. Other lipid molecules have been omitted for clarity. (C) Density plots showing the frequency of occurrence of ANT around ANT, and CL around ANT. The protein has been omitted from the CL plot for clarity.

The interfaces at $Site_I$ and $Site_{II}$ showed a greater preference for protein-protein interaction, with comparatively weaker density seen at $Site_{III}$. However a clear preference for certain permutations of protein-protein interaction interfaces was not observed, but rather a mixture of interfaces were seen. For example, the interface at $Site_I$ was able to interact with all other interfaces. For all regions, the contact area between interacting proteins was relatively small, with contacts to neighbouring protein molecules seen almost exclusively at the cytoplasmic facing ends of the TM helices, and the short matrix helices (Figure 4.12). Residues exposed to the core of the membrane on the concave surface of the translocase formed comparatively few direct protein-protein contacts, in

keeping with the 'hourglass' shape of ANT1. Contacts of this hydrophobic surface were instead mediated by lipid tails.

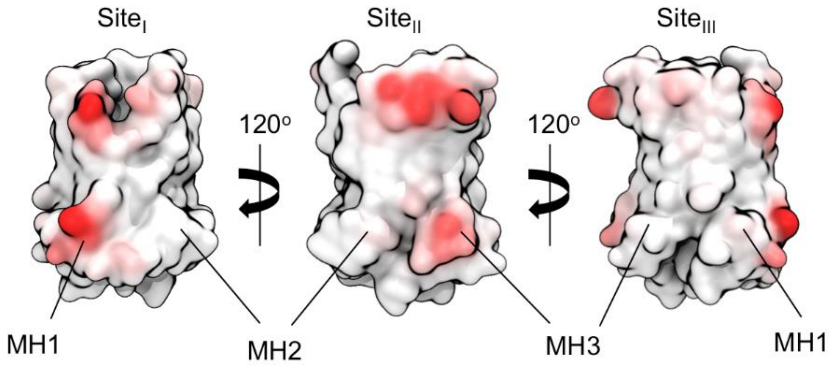

Figure 4. 12: Protein-Protein interactions for Bovine ANT1. ANT1 structures colored from white to red according to the degree of contact formed with other copies of the ANT1 protein. Contact between ANT1 proteins is predominately mediated by the short MH helices on the matrix side, and the N-terminal ends of the transmembrane helices on the cytoplasmic side.

Interestingly simulating a comparably sized membrane patch *without* CL in the bilayer (Table 4.1) resulted in a greater degree of oligomerisation of the ANT population over the 20 μs time course (Figure 4.13A). This suggests that CL may have the capacity to modulate the oligomerisation process. Calculation of diffusion coefficients for the ANT1 protein molecules showed comparable values to those seen for GPCRs [48], and did not exhibit a significant difference in the presence and absence of CL (Figure 4.13B). This suggests the difference in oligomerisation is unlikely to be due to differences in diffusive behaviour in the two lipid mixtures, and may instead be indicative of modulation *via* direct interaction.

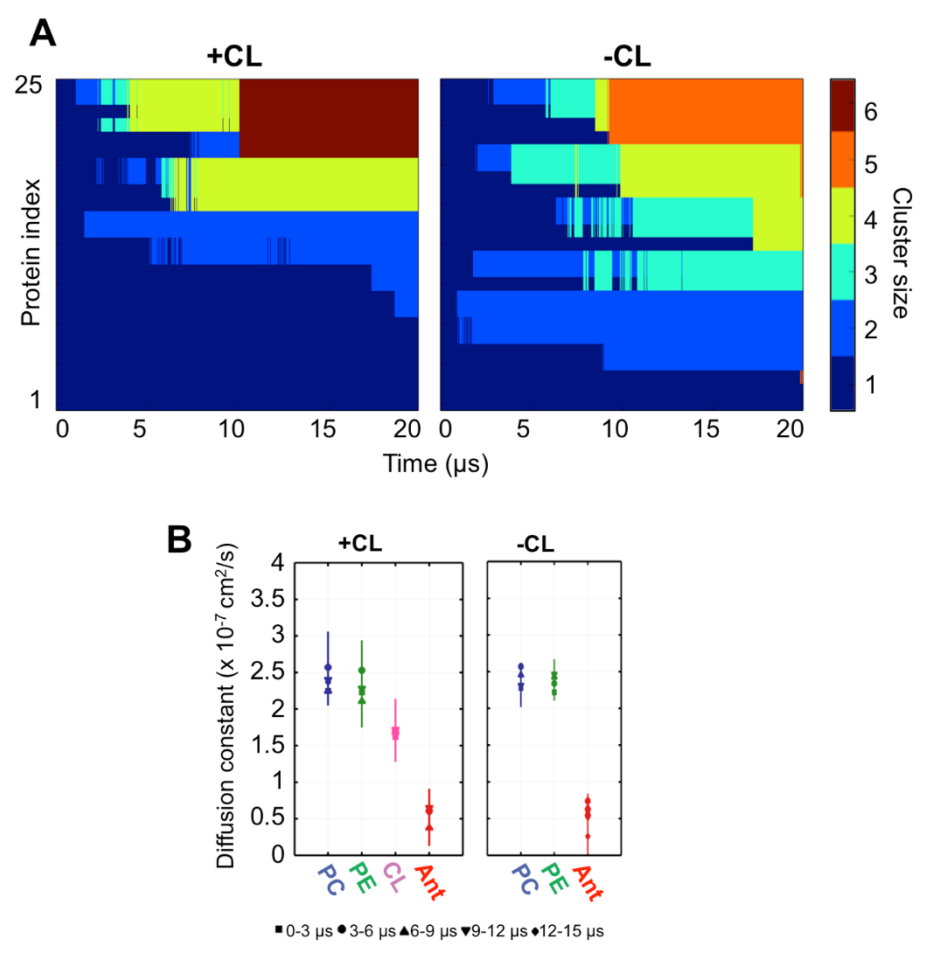

Figure 4. 13: ANT1 oligomerization dynamics within + and – CL membranes. (A) Matrices showing the oligomerization sate of each of the 25 ANT1 molecules as a function of simulation time. (B) Diffusion constants for different sections of simulation time for lipid and protein species, extracted from mean square displacements computed using *g_msd*.

4.4 Discussion

Cardiolipin molecules have been observed bound to several crystal forms of the bovine ANT1 [33, 280]. Three CL molecules are bound per ANT1 monomer, with the phosphate groups positioned within grooves formed by the short matrix helices. The sites identified within the current simulations are in complete agreement with the experimentally determined sites, demonstrating that the latter persist within a membrane environment, and thus are likely to be present *in vivo*. The significance of these sites was recently also supported by NMR measurements showing persistence under two detergent regimes [298]. Calculation of free energy profiles for lateral lipid interaction with these sites shows moderately tight binding of CL molecules compared to other regions on the

protein, along with clear selectivity for CL over zwitterionic phospholipids, in agreement with indications from ^{31}P NMR measurements [154]. The well-depths for these profiles are -20 to -25 kJ/mol, and may be compared to profiles calculated *via* the same method for other protein-lipid interactions, including glycolipid and phosphoinositide interactions with the EGFR [285], phosphoinositide binding to PH-domains [299], and in particular to CL interactions with mitochondrial C_{III} and C_{IV} [47, 293]. This overall suggests the sites on bovine ANT1 have a clear affinity for CL, which is greater than the majority of CL binding sites on C_{IV}, and is superseded only by two particularly high affinity sites on C_{IV} which exhibit interaction free energies of ca. -32 kJ/mol [150].

Extending the simulations to the two other members of the mitochondrial carrier family of known structure, yeast Aac2p and mouse UCP2, revealed comparable direct interaction of CL with both proteins, and demonstrated conservation of binding site location. For yeast Aac2p, this agrees well with the location of co-crystallised CL molecules [282]. At present there are no structural data for CL binding to UCP2. However the presence of CL was required to obtain useable NMR spectra, suggestive of direct interaction [281]. The simulations reported here support this and predict CL to bind directly to three sites on UCP2 equivalent to those seen in ANT1 and Aac2p. Interestingly, recent biochemical work [300] on the related protein, UCP1, suggested a binding stoichiometry of three CL molecules per monomer, in support of the simulation based prediction for UCP2 reported here. The conservation in binding site location over several proteins suggested by simulation, structural and biochemical studies, along with the predicted conservation of the internal three-fold repeat is strongly indicative of likely conservation of the CL binding sites in other members of the mitochondrial carrier family. It will be interesting to see how far this conservation extends as further structures emerge.

Advances in simulation methodology [42, 301] permit access to extended time and length scales. These advances allowed investigation of the larger (~0.1 μm) scale organization and protein-lipid interactions of bovine ANT1. The simulations suggest the cytoplasmic facing conformation of bovine ANT1 has the capacity to form self-interactions at specific interfaces and thus to form oligomers within model membranes. The protein-protein interfaces co-localise with the CL binding sites, and CL is often seen to bridge between adjacent proteins. However in contrast to crystallographic dimers [280], a stable arrangement in which two CL molecules may be present at an interface was not observed, with the interfaces rather being mediated by a single CL. This may be related to the unfavourable electrostatics of having two negatively charged (-2*e* each

[302]) CL head groups in close proximity, in a membrane environment which allows for more dynamic and transient protein-protein and protein-lipid assembles than may be observed than in a crystal. It would therefore be attractive to analyse the relative free energies of the different modes of protein-protein and protein-CL-protein interactions observed within these dynamic complexes. However, this would be likely to require the development of more effective simulation-based sampling regimes before such free energies may be reliably estimated [216]. Interestingly, simulations of membrane patches in the absence of CL showed a greater degree of oligomerisation over the 20 μs simulation compared to the membrane with CL present. This suggests a more complex relationship between CL and ANT-ANT interactions, with a single CL molecule bound between interfaces able to mediate protein interactions as a molecular glue [303], but the presence of two CL molecules between interfaces possibly acting as a block to disfavour oligomerisation.

The oligomeric state of ANT1 has been much debated [277, 304, 305]. While the functional unit of the translocase is widely accepted to be monomeric, this does not preclude oligomerisation under certain conditions. It is important to remember that the observations reported here pertain to the cytoplasmic facing state of the translocase within simple model membranes: both conformational changes and biologically complex lipid mixtures may also influence oligomerisation, as observed for GPCRs [114, 118]. In particular other anionic lipids, e.g. PS, PA and PI, at low levels in the inner mitochondrial membrane [56, 288, 289] may interact with ANT within living cells. Other *in vivo* factors such as the high degree of membrane curvature of the mitochondrial inner membrane, along with a more heterogeneous protein population may also influence ANT-ANT interactions. Further, a number of *in vitro* model membrane studies have suggested CL may regulate the activity of certain proteins *via* modulation of curvature and other biophysical properties of the inner membrane [306-309]. Larger scale simulations capable of assessing such macroscopic properties would be of interest in this regard. Likewise, it should be noted that key inner membrane-resident compounds such as electron transferring quinones have been suggested to influence membrane properties [310]. Structural data is now available for a number of other mitochondrial membrane proteins, along with EM data on their *in vivo* arrangement and membrane curvature [311]. As the number of lipid parameters and time and length scales accessible to simulations continue to increase [296], it may be possible to address these factors in large scale simulations. Regarding mechanistic insights into the influence of CL on the ANT transport cycle, as

information on the possible nature of ANT conformational states continues to emerge [275, 312] a key extension of the current simulations would be to examine the differential interactions of CL with multiple conformational states of ANT, in order to reveal the detailed mechanism of CL influence on ANT function.

4.4.1 Methodological considerations

It is important to be aware of possible limitations in the simulation methods employed. Use of CG models results in an inherent loss of chemical detail, which may be expected to influence the accuracy of protein-lipid interactions. Although the ionization state of the CL head group has been estimated as -2e in solution [284], it is possible this may change depending on microenvironment, as may protein ionization states. Capturing such effects would require constant-pH simulations, with a concomitant increase in computational load [313]. Despite the absence of such effects, good agreement is seen between the CG and atomistic simulations, together with previously resolved crystal structures. Further, it should be noted the method has previously been shown to correctly predict lipid binding sites on an increasing number of other membrane proteins [28, 47, 135], in agreement with experiment [301], and is capable of predicting the effects of mutation on lipid binding [285, 299].

A further key limitation of current simulations is the accessible timescale, which remains a challenge even when using CG models. This is particularly pertinent for large systems containing multiple copies of proteins. These limitations may impede the investigation of slowly converging system properties such as protein-protein interactions [48, 291]. In the current study on a 20 μs timescale only part of the self-assembly process is captured, and an equilibrated state for protein-protein interactions has likely not been reached. Thus, insights into the strength of protein-protein interactions are circumscribed due to limited statistics for association and dissociation events. Nonetheless, important information can be gleaned in identifying protein interaction interfaces and the influence of lipids on the process, as seen for e.g. respiratory complexes III and IV [291], and for the S1P1 receptor [48].

Regarding protein-protein interactions, the MARTINI model has been suggested to overestimate interaction strengths in aqueous solution [314]. However, within a membrane environment the model performs well, with the free energy profiles for glycophorin A transmembrane helix dimerization estimated as ca. -30 to -40 kJ/mol in

MARTINI CG [315, 316], and -45 to -60 kJ/mol in all-atom simulations [317, 318], which agrees reasonably with available experimental data [319]. The application of MARTINI dihedral restraints together with an ElNeDyn elastic network [320] prohibits conformational transitions and thus only the cytoplasmic facing state of the translocase is considered, for which crystal structures are known. One might anticipate that potential conformational transitions [321] within the protein could influence lateral protein-protein interactions, as well as protein-CL interactions.

4.5 Conclusions

In summary, three CL binding sites have been characterised on the cytoplasmic facing conformation of bovine ANT1 within model mitochondrial membranes. Free energy calculations reveal the ~100 μM affinity of these sites for CL, and their selectivity over zwitterionic phospholipids. Equivalent sites are present for mouse UCP2 and yeast Aac2p, indicative of both their biological significance and conservation within the transporter family, as further supported by recent biochemical studies of UCP1 [300]. Simulations in extended membrane patches demonstrate that CL interactions at these sites persist in the presence of protein-protein interactions, and indeed CL molecules may mediate interactions between adjacent ANT1 molecules, enabling formation of a range of oligomeric states. More generally, these simulations demonstrate further the ability of computer simulations in identifying lipid binding sites on membrane protein, providing near-atomic resolution characterisation of protein-lipid interactions to complement biochemical and structural studies.

"Measure what is measurable, and make measurable what is not so."

- Galileo Galilei

5

PIP_2 lipid interactions with the Neurotensin receptor 1

Contents

5.1 Introduction

Recent native mass spectrometry measurements from the Robinson laboratory (Department of Chemistry, University of Oxford) suggest PIP_2 lipids bind directly and specifically to a Class A GPCR, the Neurotensin Receptor 1 (NTS1). Purified and detergent solubilized NTS1 (derived from overexpression in *E.coli*) was incubated with

defined concentrations of PIP_2 lipids, before being subjected to native mass spectrometry measurements. Analysis of the subsequent spectra revealed peaks corresponding to the mass of NTS1+n(PIP_2), where n=1,2, or 3. This result was unaffected by the acyl tail saturation patterns, but binding was essentially abolished for PIP lipids, suggestive of a head group dominant binding mode, mediated *via* the terminal phosphoryl groups of PIP_2.

Key questions emerging from this work are a) the behaviour of PIP_2-NTS1 interactions within a membrane environment, and b) the location of the putative PIP_2 interaction sites on the surface of NTS1. The simulations conducted within this Chapter aimed to address these questions, and provide prospective predictions for the experimental mutagenesis of NTS1, toward structurally rationalising the proposed interaction sites. I begin by introducing the biology of NTS1, and PIP_2, a key member of the phosphoinositide family.

Phosphoinositides

The phosphoinositides (also referred to as phospatidylinositides) are a major family of eukaryotic cell membrane lipids [322]. Structurally, they consist of a glycerol backbone, covalently attached to two fatty acid tails, and a phospho-inositol head group (Figure 5.1). In addition to the bridging phosphate group, the hydroxyl groups of the inositol moiety may be covalently modified with further phosphoryl groups to form a range of different phosphoinositides including phosphatidylinositol-4-phosphate (PIP), phosphatidylinsitol-4,5-bisphosphate (PIP_2: the most abundant plasma membrane phosphoinositide), phosphatidylinositol-3,4,5-trisphosphate (PIP_3).

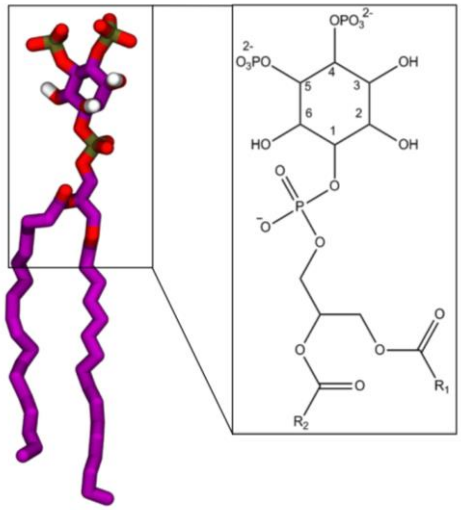

Figure 5. 1: Chemical structure of $PI(4,5)P_2$, the most abundant member of the phosphoinositide family [248]. The structure consists of a glycerol backbone, covalently attached to two fatty acid tails, and a phospho-inositol group. The inositol ring of phosphoinositides is differentially decorated with phosphate groups to form a range of lipid species [252].

Located predominately within the inner leaflet of the plasma membrane, they serve important functions in regulating the activity of ion channels [253], receptors [239], and transporters [140]. In a number of cases this modulation is thought to occur *via* direct interaction with integral membrane proteins at specific sites [22]. In particular phosphoinositide lipids have been observed co-crystallised with a number of membrane proteins [75, 323, 324], known to be functionally modulated by PIP_2. In addition to this, phosphoinositides play central roles in the recruitment of soluble proteins to the membrane surface [322], the lateral organisation of transmembrane proteins such as syntaxin-1A [325] and EGFR [266], and as secondary messengers in key signalling cascades such as the phosphatidylinositol-3-kinase (PI3K) and phospholipase C (PLC) signal transduction cascades [322]. During PI3K signalling, PIP_2 is chemically converted to PIP_3 by PI3K, which facilitates the recruitment of protein kinase B (PKB) and phosphoinositide-dependent kinase 1 (PDK1) *via* interaction of their pleckstrin homology domains with PIP_3. Whilst in the PLC cascade, PIP_2 is chemically cleaved by PLC to form the diacyl glycerol (DAG) and the soluble second messenger inositol-trisphosphate (IP3), which subsequently stimulates release of Ca^{2+} from the sarcoplasmic reticulum and activation of protein kinase C (PKC).

Neurotensin receptor 1 (NTS1)

NTS1 is a Class A GPCR expressed particularly in the central and peripheral nervous system, in addition to a range of other tissues [326]. At the molecular level it is responsible for binding the 13-residue peptide, neurotensin [327], leading to activation of intracellular G-proteins *via* exchange of GDP for GTP, and subsequent activation of intracellular signalling cascades. Binding of spin-labelled lipids to NTS1 [328], modulation of G-protein coupling by lipids [329, 330], and the influence of lipids on ligand binding [331] have been previously studied, however interactions with phosphoinositide lipids remain to be explored.

In common with other Class A GPCRs, NTS1 contains the canonical seven transmembrane helices arranged around a central ligand binding pocket (Figure 5.2).

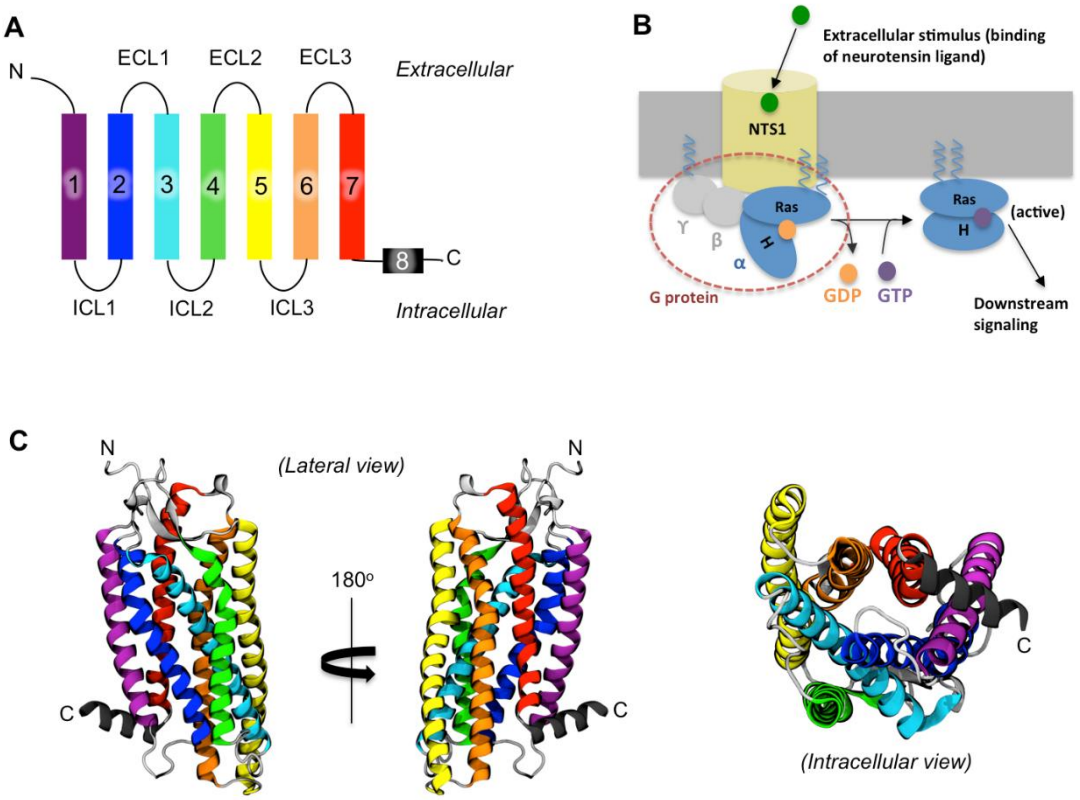

Figure 5. 2: NTS1 structure and function. (A) Topology of NTS1. (B) NTS1 is a canonical Class A GPCR, capable of coupling to a Gq proteins [330] upon binding of its native ligand, neurotensin, leading to exchange of GDP for GTP within the Gq protein α subunit, and activation of downstream signalling [332]. (C) Crystal structure of NTS1 (PDB id: 4BUO) [333] coloured according to the scheme presented in (A).

The first structure of NTS1 emerged in 2012 from the group of Reinhard Grisshammer, bound to a truncated portion of the native neurotensin ligand and resolved to a resolution of 2.8 Å (PDB id: 4GRV) [334]. The structure contained 6 stabilizing point mutations, and a T4-lysozyme (T4L) insertion within ICL3 (Figure 5.3A). The structure was suggested to represent an 'active-like' conformation on the basis of an outward tilt of the intracellular end of TM6 seen in other active GPCR structures [335]. However biochemically the construct is signalling inactive both in the presence and absence of the T4L fusion protein, raising a degree of uncertainty as to the biological impact of the stabilizing mutations. Subsequently in 2014 the group of Andreas Pluckthun solved the structures of four thermostabilized NTS1 constructs, obtained *via* a high-throughput directed evolution approach [333]. The highest resolution structure was solved to 2.75 Å (PDB id: 4BUO). Importantly the constructs used for structure determination contained a 11 stabilizing point mutations, truncation of the flexible ICL3, and of the N- and C-termini, but no fusion protein (Figure 5.3B). The presence of such fusions in ICL3 is a common strategy to provide adequate crystal contacts for GPCR crystallisation, but introduce a degree of uncertainty as to the constructs physiological relevance as a crucial functionality, G protein binding, is precluded. The omission of a fusion in the Pluckthun structures is therefore to be welcomed, and indeed the constructs used exhibited wild-type like agonist and antagonist binding, and were capable of binding G proteins and catalysing GDP/GTP exchange. In contrast to the 2012 structure [334], the Pluckthun structures exhibited comparatively less outward movement of TM6 (Figure 5.3C), and are thought to represent an inactive conformation [333].

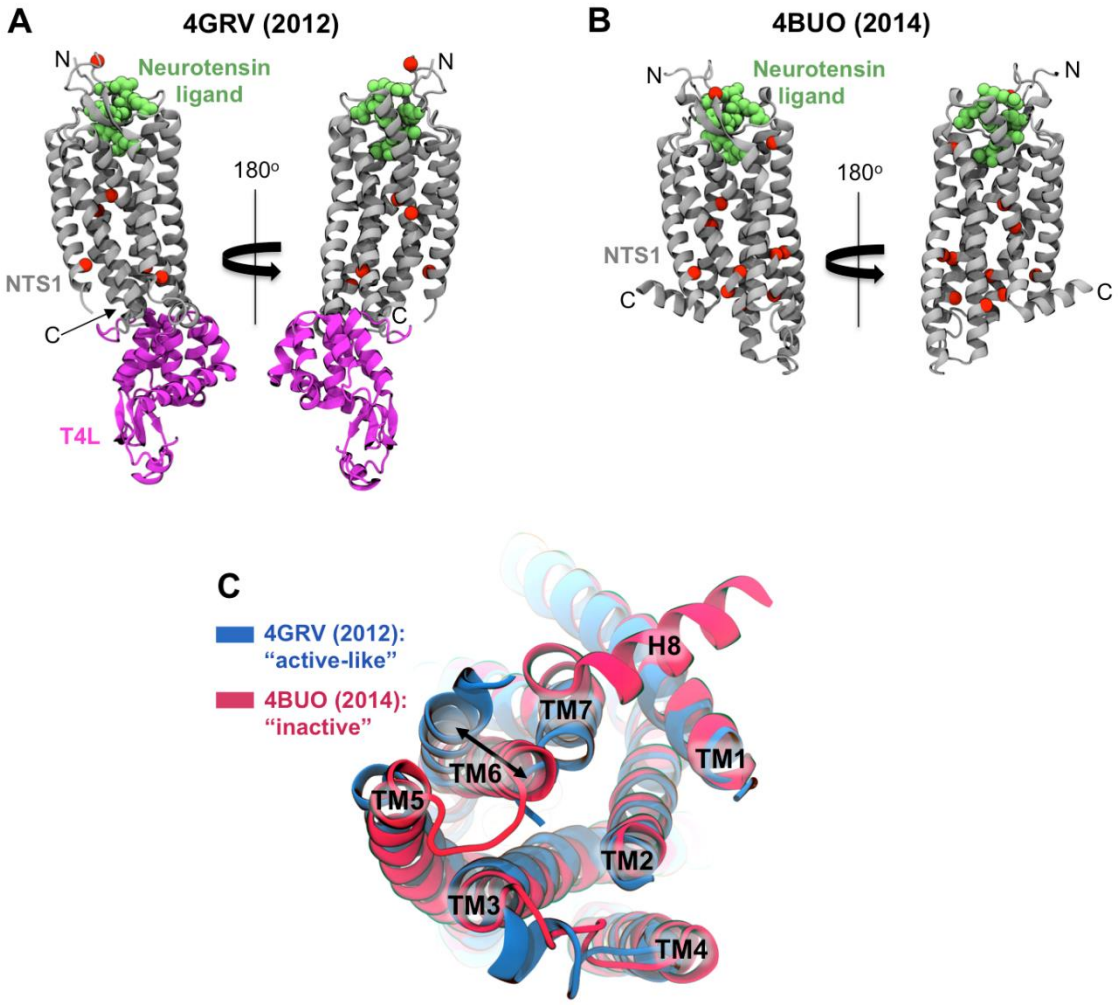

Figure 5. 3: Comparison of NTS1 structures. Lateral view of (A) 4GRV and (B) 4BUO, depicting the neurotensin ligand (green), NTS1 (grey), and inserted T4L (magenta). The location of stabilizing point mutations within each structure are indicated by red spheres. (C) Structural alignment viewed from the intracellular side. The most pronounced differences between the two structures are the outward movement of TM6, and the absence of H8 in 4GRV.

Within this Chapter a thermostabilized version of NTS1 (PDB id: 4BUO) was chosen for simulations, on the basis that it a) is similar in sequence to the construct used for mass spectrometry measurements (Appendix Figure C1), b) represents the highest resolution structure available, c) contains part of H8, and d) is devoid of fusion proteins and retains the ability to bind G proteins and catalyse GDP/GTP exchange.

5.2 Methods

5.2.1 Modelling and system setup

Simulations were performed using the GROMACS v4.6.3 simulation package [179]. Initial protein coordinates for NTS1 were obtained from PDB id: 4BUO [333] (NTS1), with missing atoms added using MODELLER [336]. Comparative simulations of β_1AR employed PDB id: 2Y03 [337]. A model of β_1AR was also constructed in which S68 in the thermostabilised structure 2Y03 was back-mutated to the wildtype R68, using the mutagenesis tool implemented in PyMOL (Schrodinger, L.L.C. The PyMOL Molecular Graphics System, Version 1.3r1 (2010)). Side chain ionization states were modelled using pdb2gmx [179]. The N and C-termini were treated with neutral charge. Each protein structure was then energy minimized using the steepest descents algorithm implemented in GROMACS, before being converted to a coarse-grained (CG) representation using the MARTINI 2.1 force field [203]. The energy minimized CG structure was centred in a periodic simulation box with dimensions 11 x 11 x 12 nm^3. 280 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipids were randomly placed around the protein and the system solvated and neutralised to a concentration of 0.15 M NaCl. An initial 50 ns of coarse-grained simulation was applied to permit the self-assembly of a POPC lipid bilayer around the GPCR. POPC lipids were randomly exchanged [184] to create a mixed-species bilayer of specified composition (Table 1). A cut-off distance of 2.5 nm was applied, with only molecules outside this distance being subject to exchange. The exchange protocol was conducted independently for each repeat simulation; such that different random initial configurations of lipids around the protein were generated for each simulation repeat. A summary of simulations performed is provided in Table 5.1.

Table 5. 1: Overview of simulations performed in Chapter 5.

Description	**Membrane composition**	**Duration**
1 x NTS1	PC(95%)/PIP_2(5%)	10 x 5 μs
1 x NTS1	PC(75%)/PIP_2(25%)	10 x 5 μs
1 x NTS1 (extended)	PC(95%)/PIP_2(5%)	100 μs
1 x NTS1	PC(90%)/PS(5%)/PIP_2(5%)	3 x 5 μs
1 x NTS1	PC(95%)/PS(5%)	3 x 5 μs
144 x NTS1[a] (asymmetric)	Outer leaflet: PC(40%)/PE(40%), Inner leaflet: PC(30%)/PE(40%)/PS(20%)/PIP_2(8%)/ PIP_3(2%)	10 μs
144 x NTS1 (symmetric)	PC(30%)/PE(40%)/PS(20%)/PIP_2(8%)/PI P_3(2%)	10 μs
144 x NTS1 (No PIPs)	PC(40%)/PE(40%)/PS(20%)	10 μs
No protein	PC(30%)/PE(40%)/PS(20%)/PIP_2(8%)/PI P_3(2%)	10 μs
1 x β_1AR	PC(95%)/PIP_2(5%)	10 x 5 μs
1 x β_1AR	PC(90%)/PIP_2(10%)	10 x 5 μs
1 x β_1AR (S68R)	PC(90%)/PIP_2(10%)	10 x 5 μs

Lipids were symmetrically distributed in all cases, other than for the simulation marked [a], for which per leaflet compositions are indicated.

Large membrane patches containing 144 NTS1 molecules were generated using the *genconf* tool of GROMACS [179] to periodically replicate a small system with a single mixed-membrane-embedded NTS1 molecule onto a 12 x 12 grid in the XY plane. The energy-minimized concatenated system contained ~40,000 lipids and 144 NTS1 molecules within a 138 x 138 nm^2 patch of membrane, solvated with standard MARTINI water and neutralised utilizing 0.15 M NaCl. The final system contained ~2.5 x 10^6 particles. An initial 5 ns of equilibration was performed to relax the system. Subsequently each system was run for 10 μs of CGMD.

5.2.2 CG simulation details

The standard MARTINI force field [189] was used to describe all system components. An ELNEDYN network [206] was applied to the protein using force constant of 500 kJ/mol/nm^2 and a cut-off of 1.5 nm. Simulations were performed as an NPT ensemble, with temperature maintained at 310 K using a Berendsen thermostat [178] using a coupling constant of τ_t = 4 ps, and semi-isotropic pressure controlled at 1 bar using a Berendsen barostat [178] with a coupling constant of τ_p = 4 ps and a compressibility of 5 x 10^{-6} bar^{-1}. Electrostatics were modelled using the reaction field coulomb type [338], and smoothly shifted between 0 and 1.2 nm. Van der Waals interactions were treated using a shifting function between 0.9 and 1.2 nm. Covalent bonds were constrained to their equilibrium values using the LINCS algorithm [172]. Equations of motion were integrated utilizing the leap-frog algorithm, with a 20 femtosecond timestep. All simulations were run in the presence of conventional MARTINI water, and neutralized to a concentration of 0.15 M NaCl.

Analysis of simulation data was conducted using VMD11, PyMOL (Schrodinger, L.L.C. The PyMOL Molecular Graphics System, Version 1.3r1 (2010)), and tools implemented in GROMACS [179], and in-house protocols. Protein-lipid contact analysis employed a cut-off distance of 0.6 nm, based on radial distribution functions for CG lipid molecules [32]. Density maps were calculated using the VolMap plugin of VMD [249].

5.3 Results

5.3.1 PIP_2 interaction sites on NTS1

To identify possible PIP_2 interaction sites on the membrane-exposed surface of NTS1, the CG MARTINI model was chosen based on the good agreement seen for lipid interaction sites in Chapters 3 & 4 with experimental data, in addition to a range of other studies [22]. Crystal structure coordinates (PDB id: 4BUO) [333] were converted to CG resolution using the MARTINI 2.1 force field . The structure was subsequently embedded into a PC bilayer via self-assembly simulations [213], before being randomly exchanged [184] for a mixed lipid bilayer of specified composition (Table 5.1). An initial 10 x 5 μs of CGMD was conducted for NTS1 embedded in a PC:PIP_2(95%:5%) membrane. Each repeat simulation contained a different initial random configuration of PIP_2 lipids around NTS1. Further, a distance cut-off of 2.5 nm was applied during the exchange step, such

that only PC lipids outside of this cut-off were subject to exchange, so as to avoid the introduction of bias through pre-placement of PIP$_2$ lipids adjacent to NTS1.

Although *in vivo* PIP$_2$ is located within predominately the inner leaflet [248] of mammalian cells, no such asymmetry exists for detergent solubilized NTS1 used within mass spectrometry experiments and it is thus possible PIP$_2$ may bind to outer leaflet regions of NTS1. As such PIP$_2$ was symmetrically distributed between leaflet within simulations.

Calculation of the number of contacts formed between each residue of NTS1 and PIP$_2$ revealed a high degree of interaction with defined regions on the intracellular portion of the protein, as well as a small degree of interaction at the extracellular side (Figure 5.4).

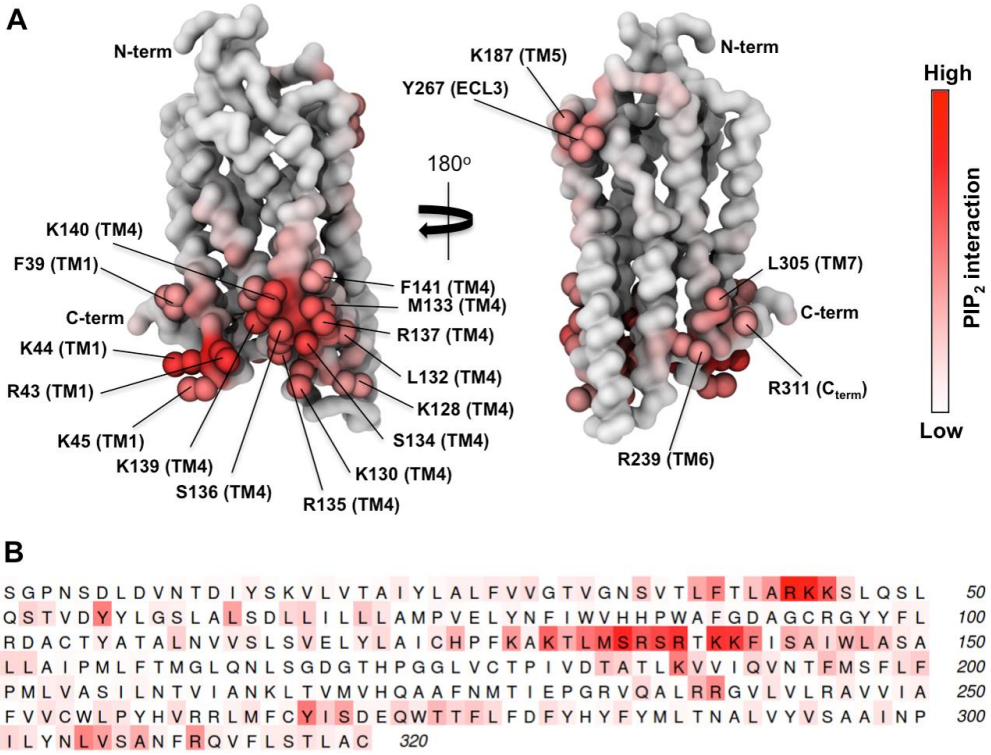

Figure 5. 4: PIP$_2$ interactions with NTS1. (A) Lateral view of the protein backbone trace shown as a surface, and (B) primary sequence of NTS1, in both cases coloured from white to red according to the degree of contact with PIP$_2$. Residues which formed especially high levels of contact (> 2.5 mean contacts per simulation frame) are labelled and shown as spheres. Contacts were calculated over 5 x 10 μs of CGMD, using a 6 Å distance cut-off to define contact.

In particular on the intracellular side, PIP$_2$ molecules were seen to bind at the interface of TM1,2, 4, of the NTS1. A degree of interaction was also observed around the intracellular

side of TM6, 7, as well as the extracellular end of TM5. In all cases, comparatively little interaction was seen for the acyl chain moieties of PIP_2, with binding mediated *via* interactions between triphosphorylated inositol headgroups of PIP_2, and basic protein side chains located at helical termini and loop regions. Increasing the PIP_2 content from 5% to 25% yielded increased total levels of interaction, but no significant deviation in the relative interaction patterns (Figure 5.5).

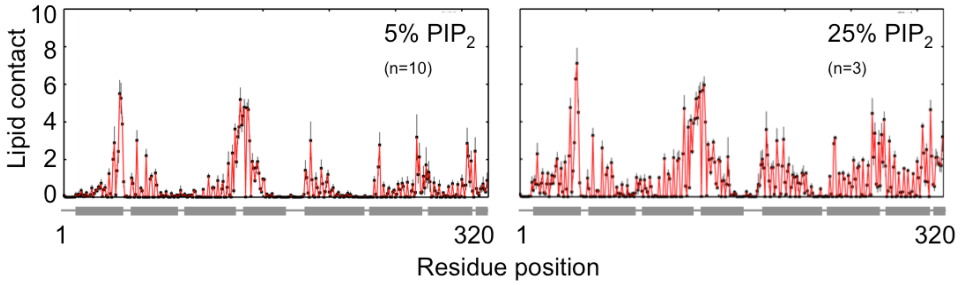

Figure 5. 5: Concentration dependence of PIP_2 contacts with NTS1. Contact maps are shown for simulations in 5% PIP_2 membranes (left), and 25% PIP_2 membranes (right).

Calculation of interaction matrices between each PIP_2 CG particle the residues seen to form particularly high levels of interaction revealed, as expected, it was the anionic phosphate and inositol moieties of the PIP_2 which tended to form the highest levels of interaction with these residues (Figure 5.6).

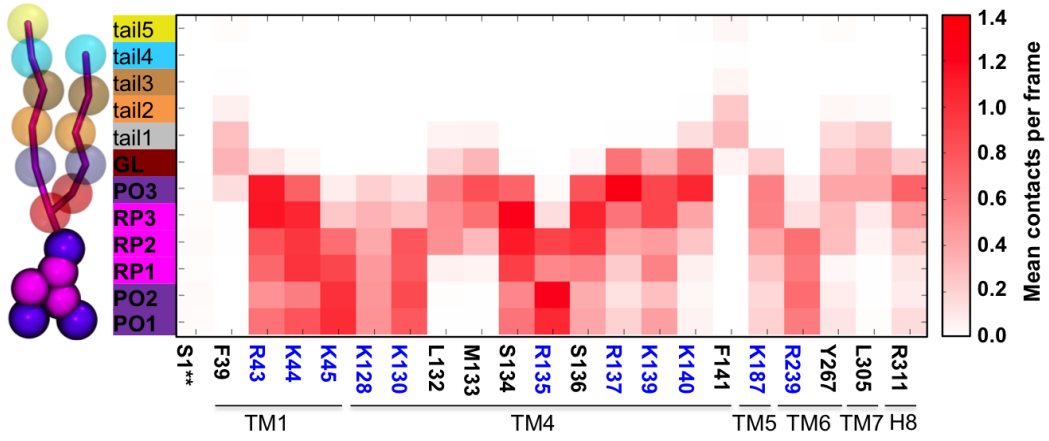

Figure 5. 6: Contact matrix showing the mean contacts per frame between each particle of PIP_2, and the major PIP_2 interacting residues on NTS1. Contacts were calculated over 10 x 5 μs of CGMD, using a 6 Å cut-off. Phosphate particles are labelled PO1, PO2, PO3; whilst inositol particles are labelled RP1, RP2, RP3. **denotes a control residue, distal from the membrane and not expected to interact with PIP_2.

The preferential interactions of PIP_2 lipids seen in the contact analysis were supported by density map calculations, which indicated the TM1,2,4 interface, alongside TM6,7 as the predominant PIP_2 interaction sites (Figure 5.7). Likewise, the density maps supported a headgroup dominant interaction mode, with less well-defined densities seen in the acyl tail region. Extending one simulation repeat to 100 μs showed no overall evolution in the patterns of PIP_2 density. Interestingly the density maps revealed four lipid interaction sites at TM1,2,4 (Figure 5.7), whilst only a single site was observed at TM6,7. Visual inspection revealed simultaneous occupancy of the four sites at TM1,2,4, with the multivalent PIP_2 headgroups coordinated by multiple residues, and capable of bridging between helices (Figure 5.7 D,E).

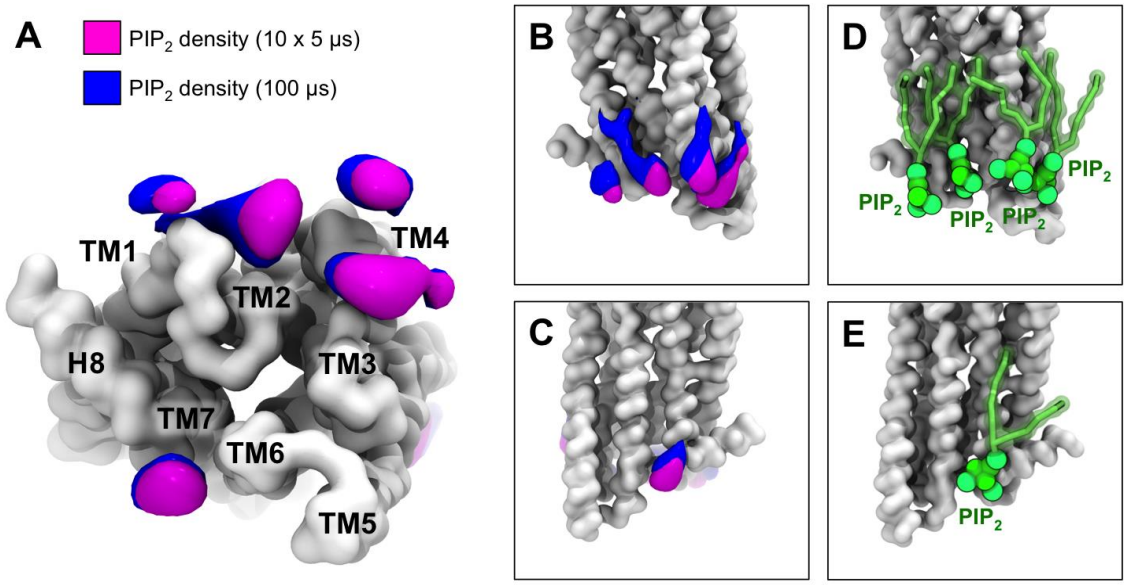

Figure 5. 7: Density maps for PIP_2. (A) Intracellular view onto NTS1 showing the time averaged density maps for the average spatial occupancy of PIP_2 around the protein. Maps were calculated from 10 x 5 μs (magenta) and 100 μs (blue) of CGMD, and have been contoured to indicate individual PIP_2 interaction sites. (B, C) Lateral views, with the 100 μs maps contoured to show tail density. Density maps were calculated using the VolMap plugin of VMD [249]. (D, E) Representative simulation snapshots showing PIP_2 lipids (green) bound to NTS1.

To assess the impact of the presence of a second anionic lipid found *in vivo* within the inner leaflet [248], PS, simulations were setup in bilayers containing PC:PS:PIP_2(90%:5%:5%), as well as PC:PS(95%:5%) (Table 5.1).

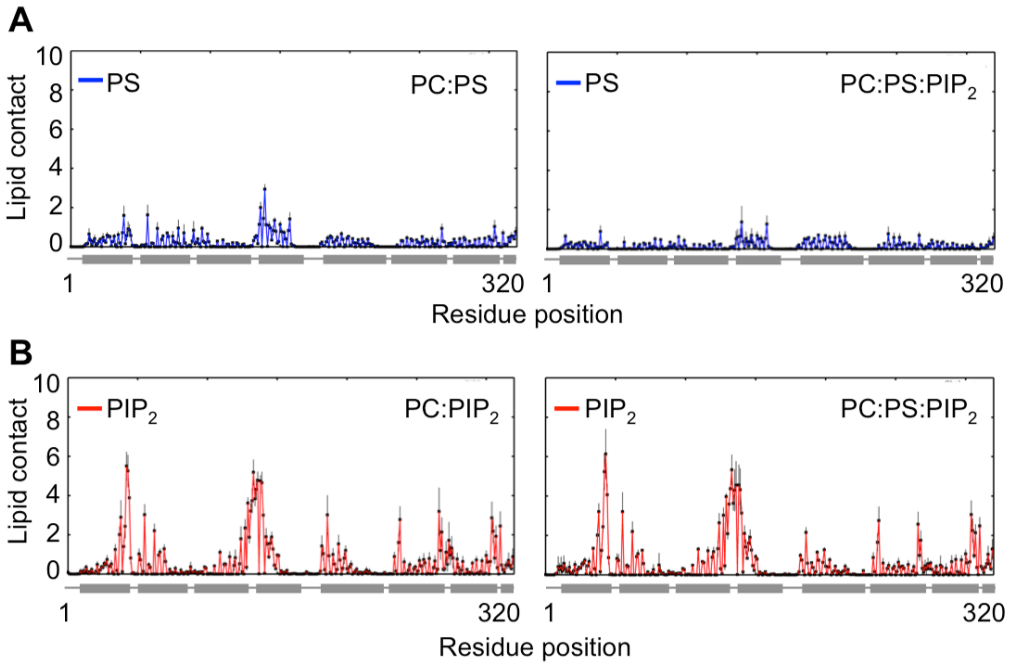

Figure 5. 8: PS and PIP_2 interactions with NTS1. (A) PS and (B) PIP_2 contact with NTS1 as a function of residue position, for PC:PS membranes (left) and PC:PS:PIP_2 membranes (right). The position of helices is denoted by horizontal grey bars. Lipid contact is calculated as the mean number of contacts between each residue and a given lipid species per frame, using a 6 Å distance cut-off. Black error bars denote the standard deviation between simulations repeats. Data from PC:PS:PIP_2 membranes was derived from 10 x 5 μs of CGMD, whilst data from PC:PS and PC:PIP_2 were each derived from 3 x 5 μs of CGMD.

Within PC:PS two-component membranes PS exhibited overall significantly less interaction with NTS1 (Figure 5.8A). In the presence of PIP_2, interaction of PS with the protein was essentially abolished, whilst PIP_2 interactions within PC:PS:PIP_2 membranes (Figure 5.8B) showed no significant deviation compared to PC:PIP_2 membranes. This suggests the lipid-protein interactions observed are specific to PIP_2 rather than broadly applicable to anionic lipid species.

5.3.2 Protein-lipid interactions within large membrane systems

In addition to these simple model membranes containing a single NTS1 molecule, simulations of large (ca. 138 x 138 nm^2) membrane patches containing 144 NTS1 molecules were performed. Membranes were constructed by periodic concatenation of single protein systems onto a 12 x 12 grid in the XY plane (See 5.2 Methods). Four

simulations were performed: PIP-symmetric, PIP-asymmetric, No-PIPs, and a PIP-symmetric(-protein) control, with system compositions detailed in Table 5.1. Each system was simulated for 10 μs of CGMD.

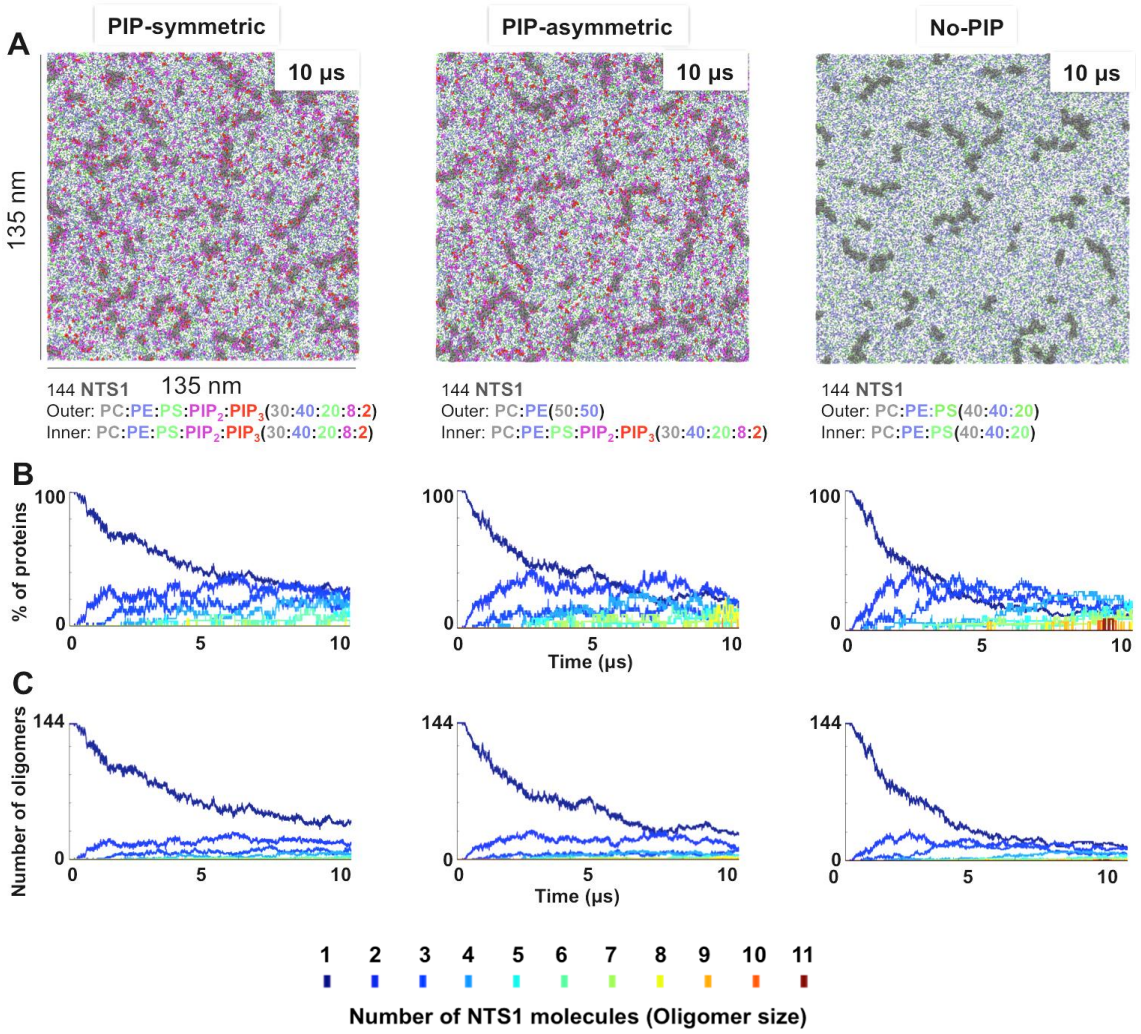

Figure 5. 9: Oligomerisation of NTS1 within large membrane patches. (A) Inner leaflet snapshots of the final frame from the PIP-symmetric, PIP-asymmetric, and no-PIP simulations, with NTS1 molecules shown in grey. (B,C) Oligomerisation as a function of simulation time expressed as the % of the protein population participating in oligomers of different sizes (B), and the absolute number of oligomers of different sizes (C). Two proteins were defined as oligomerised if the distance between their centers of mass were < 6 nm, in line with [48].

Within each system NTS1 molecules were seen to oligomerise on a sub-microsecond timescale, with a range of oligomeric states observed by 10 μs (Figure 5.9). A slightly faster rate of oligomerisation was seen in membranes the No-PIP membrane, which

contains comparatively less lipid complexity. A clear preference for certain oligomerisation interfaces was not seen (Figure 5.10).

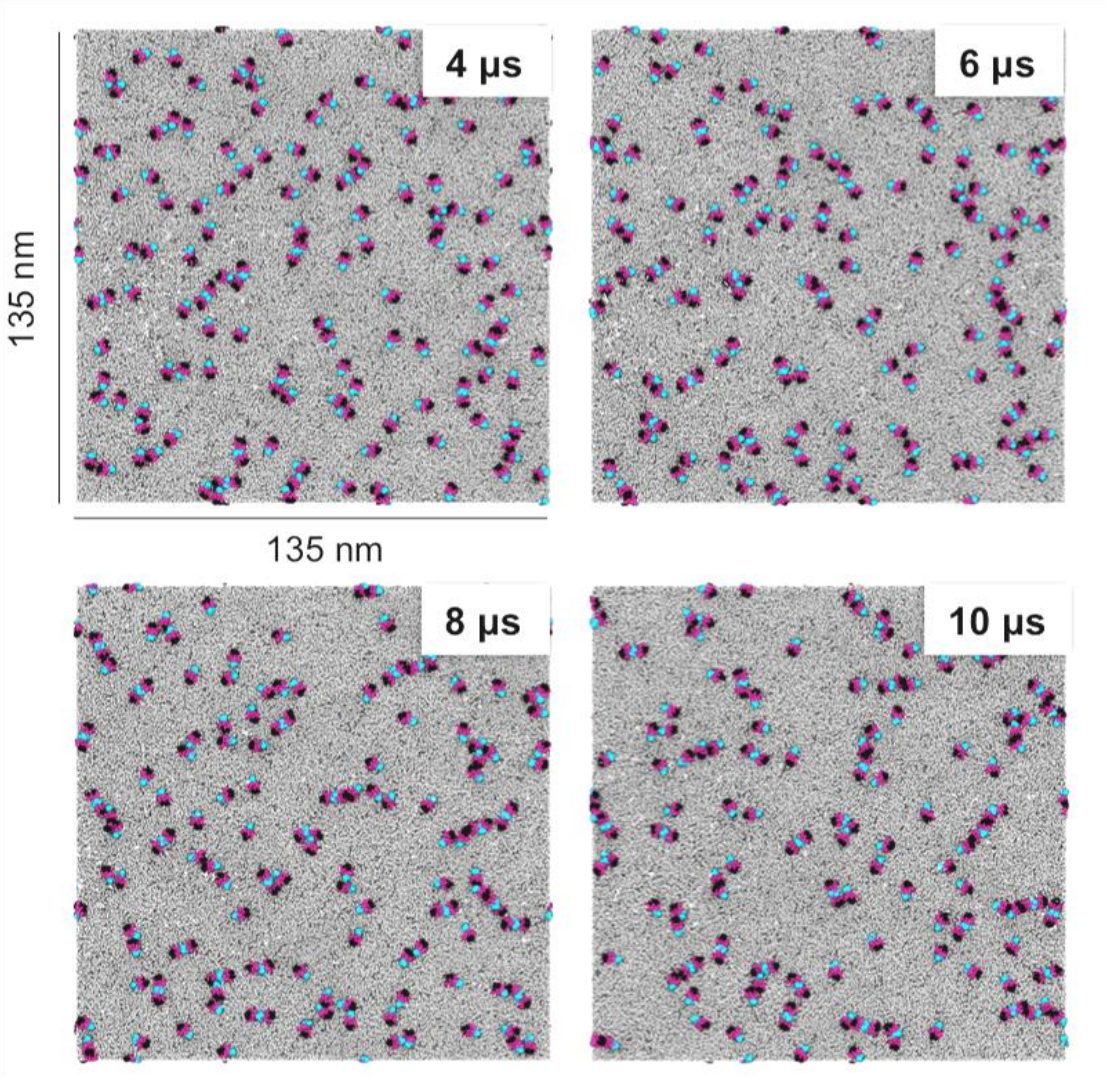

Figure 5. 10: Lateral NTS1 protein-protein interactions. NTS1 molecules are coloured in magenta, with TM1 in cyan and TM5 in black to provide perspective of relative orientations. Snapshots depict the inner leaflet of the PIP-symmetric system at different timepoints.

Calculation of diffusion constants (Figure 5.11) within each system revealed PC, PE, and PS tend to exhibit increases lateral mobility compared to PIP_2 and PIP_3. This was seen both in the presence and absence of NTS1. This slower diffusion of phosphoinositide previously been observed in CG simulations [48], and attributed to the bulkier headgroups of PIP_2 and PIP_3 and their propensity to form self-interactions [184]. Consistent with this,

the slowest rates of lipid diffusion were seen in the PIP-symmetric membrane, followed by the PIP-asymmetric membrane, and then the no-PIP membrane; with the overall lateral mobility of the system increasing as phosphoinositides were titrated out. Comparing the PIP-symmetric and PIP-symmetric(-protein) membranes, removal of slowly diffusing NTS1 molecules resulted in an overall increase in the mobility of all lipid species.

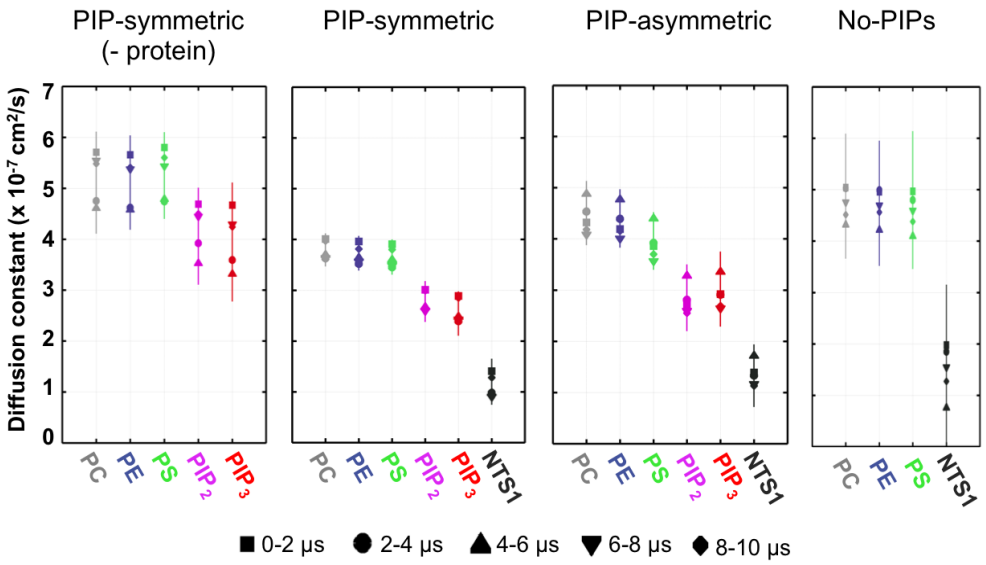

Figure 5. 11: Diffusion analysis. Lateral diffusion constants for each lipid species and NTS1, calculated from mean-square displacements using the g_msd tool of GROMACS [179].

Importantly, the pattern of PIP$_2$ interaction with NTS1 seen in the single protein systems, showed no significant evolution compared to that seen in the large NTS1 membranes (Figure 5.12). A significant increase in sampling is obtained in these large membranes, and thus this observation suggests the PIP$_2$-NTS1 interaction predictions are well converged within the limitations of the model. Further, it suggests that these interactions are not perturbed by the presence of protein-protein interactions.

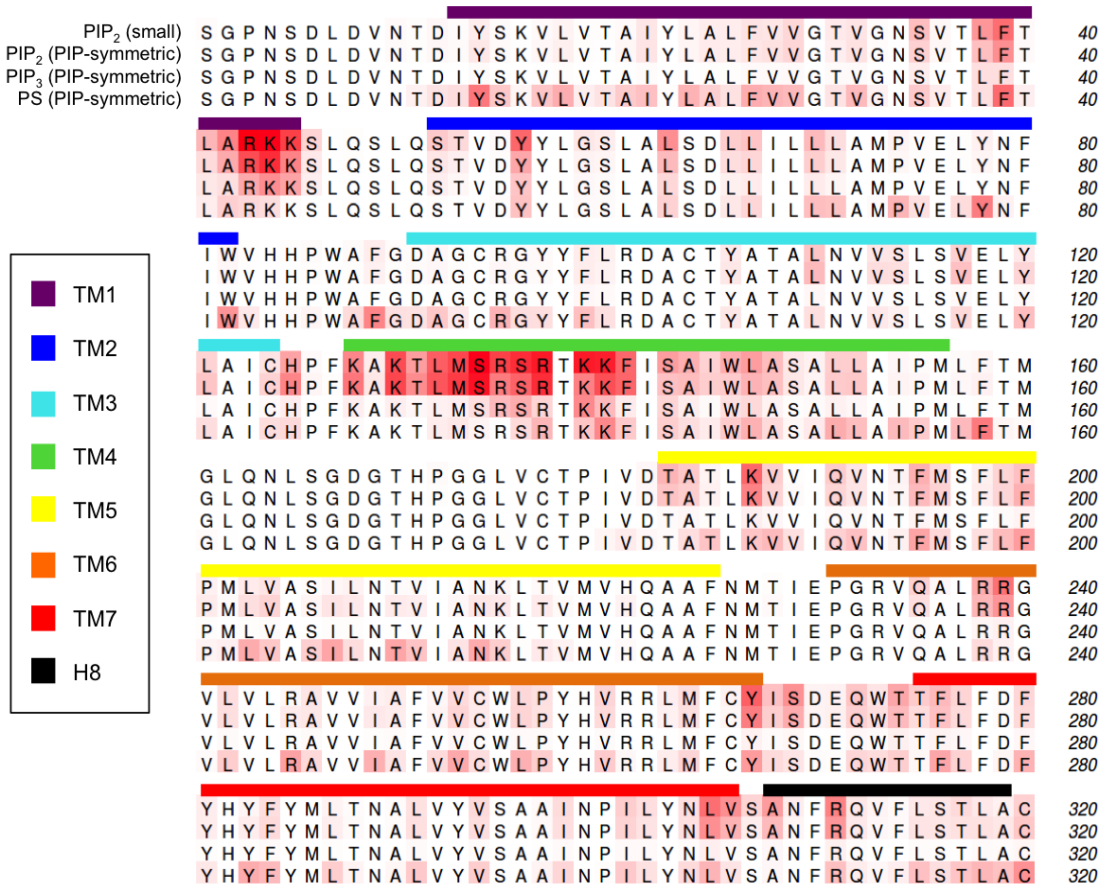

Figure 5. 12: Lipid-protein interactions. Primary sequence of NTS1 coloured from white to red according to the degree of lipid interaction. The top line shows PIP_2 interaction extracted from 10 x 5 µs CGMD of a single membrane embedded NTS1, the second line shows PIP_2 interactions extracted from the 10 µs PIP-symmetric membrane containing 144 NTS1 proteins. The third and fourth lines show interactions for PIP_3 and PS respectively, extracted from the PIP-symmetric simulation. Contacts were calculated using a 6 Å cut-off. The positions of TM1-7 and H8 within 4BUO [333] is denoted by coloured horizontal bars.

Regarding other anionic lipid species, relatively little interaction was observed for PS, despite being present at a 2.5-fold higher concentration compared to PIP_2, and 10-fold higher concentration compared PIP_3. This is consistent with the small system PS-PIP_2 competition assays. Likewise, PIP_3, present at a 4-fold lower concentration than PIP_2, exhibited relatively little interaction with NTS1. These relative compositions were chosen based on lipidomics estimations for the composition of a generic plasma membrane [48, 248] and thus suggest PIP_2 is likely to be the primary NTS1 interaction partner in more complex *in vivo*-like membranes.

5.3.3 Comparison to β_1AR

The structural conservation of Class A GPCR architecture [92, 100] raises the prospect that PIP$_2$ interaction patterns seen for NTS1 may also occur in other Class A GPCRs. To test this hypothesis, CGMD simulations were performed for a second Class A GPCR, the Beta-1-adrenergic receptor [337]. Simulations were conducted in the same manner as for NTS1 (Table 5.1), using a thermostabilised version of β_1AR (PDB id: 2Y03) [337].

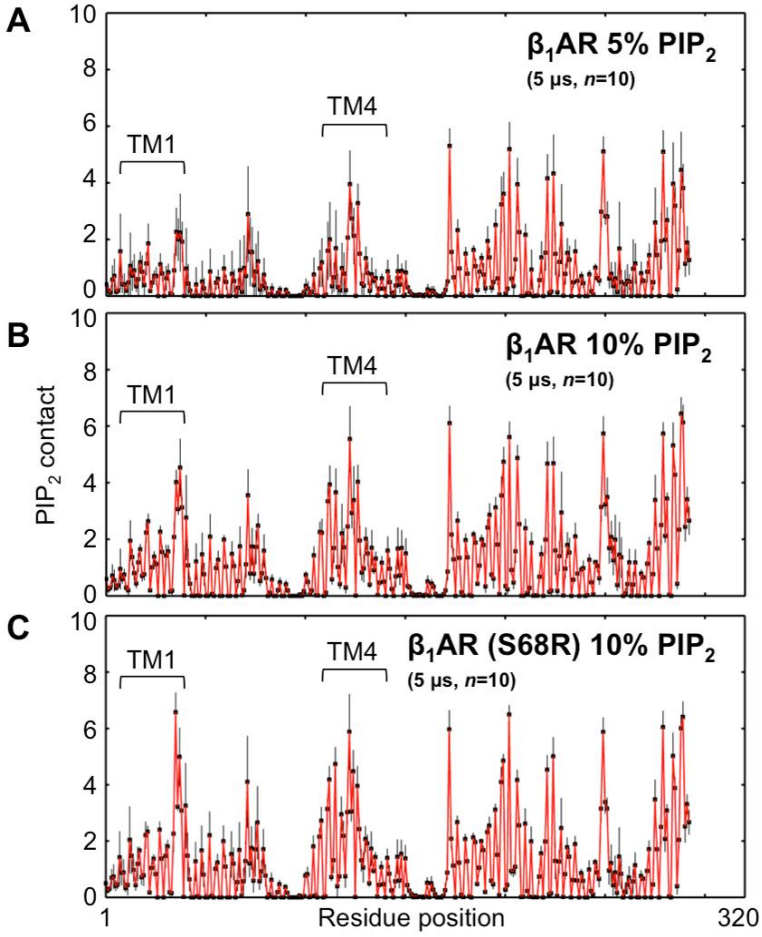

Figure 5. 13: PIP$_2$ interactions with β_1AR. PIP$_2$ interactions as a function of residue position for (A) β_1AR in a 5% PIP$_2$ bilayer, (B) β_1AR in a 10% PIP$_2$ bilayer, and (C) β_1AR containing S68R mutation in a 10% PIP$_2$ bilayer. PIP$_2$ contact was calculated as the mean number of PIP$_2$ contacts per frame for a given residue, using a 6 Å cut-off. The standard deviation is denoted as black error bars.

Within simulations containing 5% PIP$_2$, a high degree of NTS1 interaction was observed, however in contrast to NTS1 these interactions showed a higher intensity around TM5, 5, and 7, with comparatively less interaction at TM1, 2, and 4 (Figure 5.13A). Repeating

the simulations at 10% PIP_2 yielded a significant relative increase in interaction at TM1, 2, and 4 (Figure 5.13B), similar to the levels seen for NTS1. Further, the thermostabilised structure of β_1AR used in simulations contained a stabilising point mutation (R68S) at the intracellular end of TM1, which in NTS1 is a position which forms critical PIP_2 interaction. Back mutating this position to the wild-type R68 resulted in a further increase in PIP_2 interaction at the intracellular end of TM1 (Figure 5.13C). Overall this suggests β_1AR forms a significant level of PIP_2 interaction around TM5, 6, and 7, and also has the capacity to form PIP_2 interaction at TM1, 2, and 4, in a concentration dependent manner, suggestive of weaker binding at this region.

5.4 Discussion

The CGMD simulations presented in this Chapter support the notion that binding of PIP_2 lipids to NTS1 - first observed *via* native spectrometry measurements - persists within a membrane environment. The simulations suggest a number of possible interaction sites on the intracellular side of the receptor, in particular TM1,2,4 and TM6,7. Contact analysis and density calculations together suggest a headgroup dominant mode of interaction, with binding of PIP_2 to NTS1 predominantly mediated *via* interaction of the phosphate and inositol moieties of the PIP_2 headgroup with basic residues located on intracellular helical termini. The simulations revealed four interactions sites around TM1,2,4 and a single site at TM6,7. Importantly, whilst the limitations of a CG model prohibit obtaining detailed atomic models of PIP_2 interactions, the simulations do suggest that multivalent PIP_2 lipids may be coordinated by multiple sidechains simultaneously. As such it seems likely that efforts to experimentally rationalise these sites by mutagenesis may benefit from performing clusters of mutations to avoid compensatory adjustments in the interactions of neighbouring residues, and observe tangible deviations in lipid association.

Analysis of binding patterns for a second anionic lipid species, PS, revealed comparatively little interaction with NTS1. Further, competition-assays in mixed PC:PS:PIP_2 membranes showed no significant impact on PIP_2 binding. This suggests the regions on NTS1 seen to bind PIP_2 are not general anionic lipid interaction sites, and possess a degree of specificity. Simulations were also performed for large membrane patches containing 144 NTS1 molecules in a plasma membrane-like bilayers containing PE:PC:PS:PIP_2:PIP_3, with relative concentrations estimated from lipidomics

measurements [184, 248]. Importantly, PIP_2 interactions in this plasma membrane-like environment were not significantly perturbed compared to PC:PIP_2 membranes with a single NTS1 molecule. This indicates the robustness of the results to increased sampling and the presence of protein-protein interactions. It also suggests PIP_2 is the primary anionic lipid interaction partner for NTS1 within a plasma-membrane-like environment, outcompeting the other anionic lipid species, PS and PIP_3. Previous simulation studies have indicated comparatively low levels of direct cholesterol interaction with NTS1 [328], and no functional impact on ligand binding [331]. However, given the importance of cholesterol more broadly in GPCR structure and function [98, 339], it would be interesting to see how PIP_2 binding observed here may be influenced by cholesterol interactions (See Chapter 7: Future directions).

Regarding oligomerisation of NTS1, a large degree of self-interaction was observed, initiated on a sub-microseond timescale. A range of string-like oligomeric states were seen by 10 μs. NTS1 has been experimentally shown to form dimers based on FRET measurements [340], however evidence for higher order oligomers is lacking. Comparing to prior CGMD simulations of other GPCRs, string-like oligomerisation has been observed for rhodopsin [341] and more recently for the μ-opiod receptor [342]. The rate and degree of NTS1 oligomerisation was greater in membranes without PIP lipids. Based in calculation of diffusion coefficients within each system, it seems likely this results from the increased global rates of lateral diffusion in membranes lacking slowly diffusing [184] PIP molecules. In contrast to the data presented for the ADP/ATP Translocase in Chapter 3, specific protein-protein interaction interfaces were not seen, indicative of possible weak non-specific lateral interactions. Further sampling and investigation in the form of e.g. PMF calculations are required to comment on oligomerisation of NTS1 and the involvement of lipids in this process (See Chapter 7: Future directions).

The PIP_2 interaction hotspots revealed in this work show a degree of similarity to recent CGMD simulations of a second Class A GPCR, the S1P1 receptor [48]. To assess possible conservation of PIP_2 binding to Class A GPCRs, a set of simulations were conducted for the β_1AR. This GPCR was chosen due to the availability of constructs in partner laboratories, which has allowed comparison to native mass spectrometry measurements. The simulation results suggest high levels of PIP_2 binding to TM5,6,7, in addition to weaker interactions around TM1,2,4. Interestingly the interactions at TM1,2,4 were concentration-dependent, with higher PIP_2 concentrations required to observe binding at this region as compared to NTS1. This is suggestive of differences in the

precise interaction pattern between the two receptors, in the context of possible conservation of overall global interaction.

Given the increase in the number of available GPCR structures [343], and high-throughput simulation methodology [273], this analysis may now be applied to a range of GPCR structures to assess lipid interaction across the ensemble of known structures. Subsequent to the work described here, I co-supervised a Masters student who began work to expand the simulations of NTS1 and β_1AR to all GPCRs of known structure. This work is now being continued as the main project for another DPhil student.

5.5 Conclusions

The simulations performed in this Chapter suggest that binding of PIP_2 to NTS1 persists within a membrane environment. A number of interaction hotspots and key residues have been predicted, which will provide guidance for future experimental mutagenesis of NTS1 toward structurally rationalising PIP_2 binding to NTS1. Further, simulations of large membrane patches with plasma membrane-like compositions suggest that PIP_2 is the primary anionic lipid interaction partner for NTS1, outcompeting PS and PIP_3 when present at concentrations suggested from lipidomics analysis of plasma membranes. Likewise, PIP_2 binding was not significantly influenced by NTS1 oligomerisation over the timescales simulated. Interaction of PIP_2 was also observed for a second Class A GPCR, β_1AR, indicative of the possible global conservation of the interaction. A number of additional questions have arisen as a result of this work. In particular how PIP_2, and other lipids, may be involved in NTS1 oligomerisation, and b) to what degree these interactions may be conserved across a broad range of GPCRs. These questions are now being actively pursued in the lab and have formed the basis of several further projects (See Chapter 7: Future directions).

"An unexamined life is not worth living."

- Socrates

6

Cholesterol interactions with Smoothened

Contents

This chapter contains simulation data also published in the manuscript:

"Structural basis of Smoothened regulation by its extracellular domains" Byrne, E. F. X.; Sircar, R.; Miller, P. S.; **Hedger, G.**; Luchetti, G.; Nachtergaele, S.; Tully, M. D.; Mydock-McGrane, L.; Covey, D. F.; Rambo, R. P.; Sansom, M. S. P.; Newstead, S.; Rohatgi, R.; Siebold, C. Nature, 2016, 535, 517-522.

6.1 Introduction

Smoothened (SMO) is a class F GPCR with a critical role in controlling developmental processes *via* the hedgehog signalling pathway [344]. In the basal signalling state Smoothened is present in the plasma membrane in an inactive form, maintained as such through unknown mechanism(s) by of a 12 transmembrane protein, patched. During activation, the binding of soluble extracellular hedgehog proteins to patched relieves inhibition of SMO, allowing it to associate with motor proteins which result in trafficking to the primary cilium. Once compartmentalised in the plasma membrane of cilia, SMO activates a series of proteins ultimately leading to activation of GLI transcription factors and the hedgehog gene program (Figure 6.1). By modulating the level of hedgehog morphogen secreted, and the differential positioning of cells in the resultant concentration gradient, differential levels of activation may be achieved and give rise to diverse developmental outcomes and tissue patterning [345].

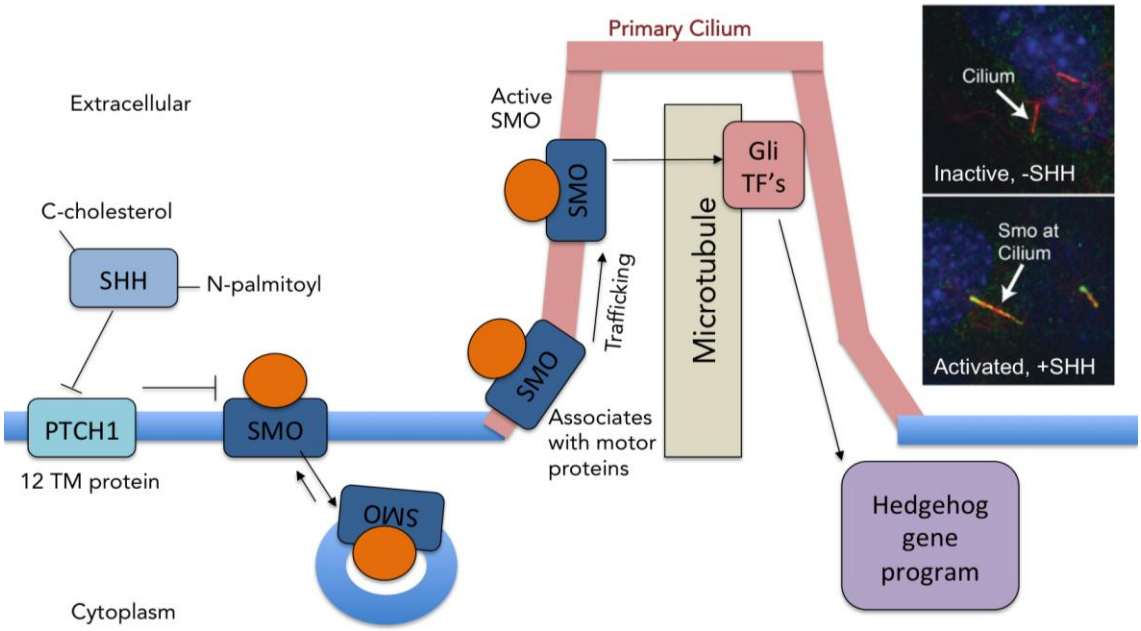

Figure 6. 1: SMO in hedgehog signalling. In the basal signalling state SMO is maintained in the plasma membrane cell body and intracellular vesicles (blue lines), through the inactivating action of patched (PTCH1). Inhibition of PTCH1 by binding of extracellular hedgehog ligands, such as SHH, results in SMO associating with motor proteins, and trafficking into the primary cilium (pink) from where it activates the canonical hedgehog signalling pathway. Inset images show cilia proteins fluorescently labelled in red and SMO in green, with untreated cells (inactive) and cells treated with SHH (active) showing differential localisation of SMO. Imaging conducted in the Rohatgi laboratory, Stanford University.

Structurally, SMO is composed of an extracellular cysteine-rich domain (CRD), a short linker domain, and the canonical seven transmembrane (7TM) domain. The first structure of the 7TM domain emerged in 2013 from the Stevens laboratory [346]. The crystal construct was truncated to include only the 7TM domain, bound to a small molecule antagonist, and solved to a resolution of 2.5 Å. Despite possessing < 10% sequence identity to Class A GPCRs, lacking key motifs including D[E]R$^{3.50}$Y (part of the so-called "ionic lock") and NP$^{7.50}$XXY, and a number of key proline residues, the structure showed remarkable spatial conservation [346]. Also in 2013, the Siebold and Rohatgi laboratories published the 2.3 Å crystal structure of the soluble CRD [347]. However an understanding of how these domains interact in the full-length structure remained elusive until 2016, when the Siebold laboratory solved the structure of near full-length construct to 3.2 Å [348], supporting a number of earlier observations from isolated structures and showing how the CRD and linker domain stack atop and interact with the canonical 7TM domain (Figure 6.2). Crystals were obtained by LCP (10% cholesterol), with one inactivating mutation in the 7TM domain (V329F), truncation of the flexible C-terminal intracellular tail, and insertion of a BRIL fusion protein in intracellular loop 3 [348]. In common with other GPCR structures lacking G-protein partners or mimetics, the 7TM domain assumed an inactive conformation. Key features of the structure include an extracellular network of nine rigidifying disulphide bonds, two N-linked glycans, attached to the sidechains of N188 and N493, and the presence of a cholesterol molecule bound to a conserved hydrophobic groove on the CRD.

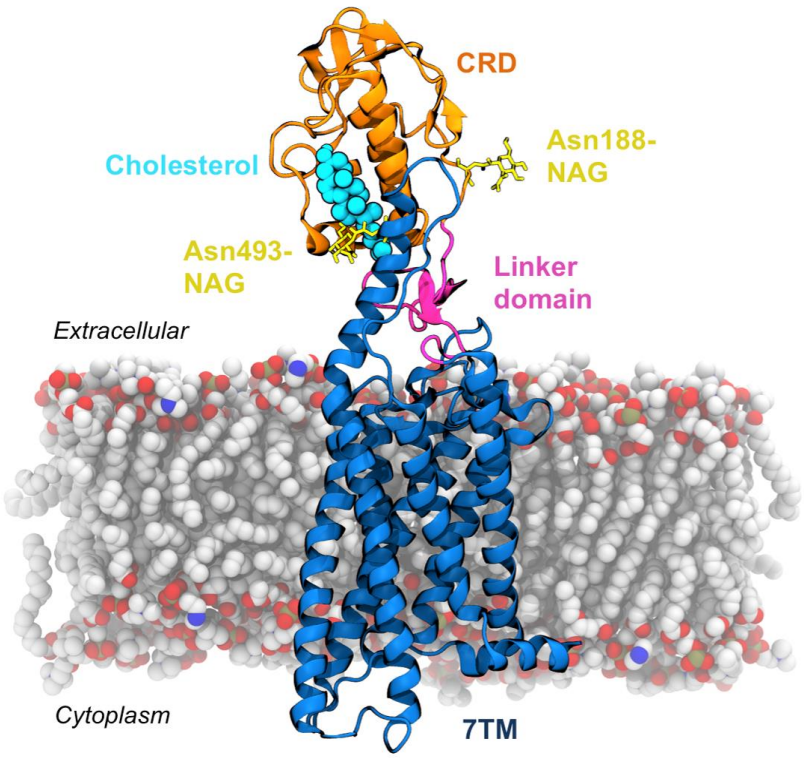

Figure 6. 2: SMO domain architecture. Image shows a crystal structure of SMO (PDB id: 5L7D) [348] superimposed on a lipid bilayer, with the N-terminal CRD in orange, the Linker domain in magenta, and the canonical 7TM domain in blue. Two N-glycosylation sites are shown as yellow sticks, and the crystallographically resolved cholesterol molecule is depicted as cyan spheres.

To assess the structural impact of the bound cholesterol molecule, a series of atomistic MD simulations were undertaken in the presence and absence of cholesterol, and the effect on SMO dynamics analysed. The simulation based predictions observed showed a good fit with subsequently available experimental data. A portion of these simulations are described in [348]. Secondly, although cholesterol molecules were not resolved on the 7TM domain, cholesterol co-purifies with SMO constructs lacking the CRD and linker domains (unpublished data), suggesting cholesterol also associates with the 7TM domain. In light of this, together with a number of GPCR crystal structures with co-crystallised cholesterol molecules [97-100], and the putative binding of cholesterol within MD simulations [48, 349], a series of CG simulations were performed which aimed to predict possible cholesterol interaction sites on the TM domain. These predictions provide input for experimental mutagenesis and activity assays to be carried out by collaborators at the Rohatgi and Siebold laboratories.

6.2 Methods

6.2.1 Modelling and system setup

Simulations were performed using the GROMACS v4.6.3 simulation package [179]. Side chain ionization states were modelled using pdb2gmx (Histidine) and PropKa (All other residues) [350, 351]. The N and C-termini were treated with neutral charge. Intracellular loop 3 occupied by the BRIL fusion in the crystal structure was modelled using coordinates from the PDB entry 4N4W [352]. The protein structure was then energy minimized using the steepest descents algorithm implemented in GROMACS, before being converted to a coarse-grained (CG) representation using the MARTINI 2.2 force field [203]. The energy minimized CG structure was centred in a simulation box with dimensions 100 x 100 x 180 $Å^3$. 270 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipids were randomly placed around the protein and the system solvated and neutralised to a concentration of 0.15 M NaCl. An initial 1 μs of CG simulation was applied to permit the self-assembly [213] of a POPC lipid bilayer around the GPCR. The endpoint of the self-assembly simulation was converted back to atomic detail using a fragment based protocol [211]. Equilibration of the atomic system was achieved *via* 1 ns of NPT simulation with the protein coordinates restrained, before the system was subjected to 100 ns of unrestrained atomistic molecular dynamics. Simulations were performed both in the presence and absence of the cholesterol ligand. Three repeat simulations were run for each case.

To assess membrane cholesterol interactions with the 7TM domain, the endpoint of the CG self-assembly simulation was randomly exchanged [184] for a two-component membrane containing POPC:cholesterol in a 3:1 molar ratio. This exchange step was conducted eight times to generate eight random initial configurations of cholesterol around SMO. Each configuration was subjected to 5 μs of CGMD to yield a total simulation set of 8 x 5 μs. This process was repeated for the plasma membrane mimetic simulation. In this case the asymmetric bilayers contained in the outer leaflet: PC:PE:cholesterol (60:15:25), and in the inner leaflet: PC:PE:PS:PIP_2:cholesterol (10:40:15:10:25).

6.2.2 Coarse-grained simulations

The standard MARTINI force field [189] and its extension to proteins [194, 203] was used to describe all system components. During the CG self-assembly simulation an ELNEDYN network [206] was applied to the protein using force constant of 500 kJ/mol/nm^2 and a cut-off of 1.5 nm. Temperature was maintained at 310 K using a Berendsen thermostat [178] with a coupling constant of τ_t = 1 ps, and pressure was controlled at 1 bar using a Berendsen barostat [178] with a coupling constant of τ_p = 1 ps and a compressibility of 5 x 10^{-6} bar^{-1}. Electrostatics and van der Waals interactions were shifted between 0 and 1.2 nm, and 0.9 and 1.2 nm respectively, and an integration time step of 20 femtoseconds was applied. Covalent bonds were constrained to their equilibrium values using the LINCS algorithm [172]. All simulations were run in the presence of conventional MARTINI water [189], and neutralized to a concentration of 0.15 M NaCl. CGMD simulation analysis was performed using the VolMap plugin of VMD [249], GROMACS [179], and in-house analysis tools.

6.2.3 Atomistic simulations

Atomistic simulations were run using the GROMOS53a6 force field [166], and its extension to glycans [353]. The system was solvated using the SPC water model, and neutralized with NaCl to a concentration of 0.15 M. Systems contained approximately 140,000 atoms including 270 POPC lipids, ~41,000 water molecules, 149 Sodium ions, and 154 Chloride ions. Periodic boundary conditions were applied, with a simulation time step of 2 femtoseconds. Temperature was maintained at 310 K using a V-rescale thermostat [176] with a coupling constant of 0.1 ps, whilst pressure was controlled at 1 bar through coupling to a Parrinello-Rahman barostat [177] with a coupling constant of 1 ps. Particle Mesh Ewald (PME) [180] was applied to model long-range electrostatics. The LINCS algorithm was used to constrain covalent bond lengths [172]. The g_dist, g_rmsf and g_rms tools implemented in the GROMACS v4.6.3 software package [179] were applied to analyse the simulations, with VMD [249] used for visualisation.

6.3 Results

6.3.1 The influence of CRD bound cholesterol on intradomain dynamics

In order to explore the influence of CRD bound cholesterol of the behaviour of SMO in a lipid bilayer, the recently resolved near full-length crystal structure (PDB id: 5L7D) [348] was embedded into a POPC lipid bilayer, *via* CG self-assembly [213] simulations. Subsequently the system was converted [211] to atomic detail and simulated for 100 ns using the GROMOS53a6 force field [166] (See 6.2 Methods for full details). Ten repeat simulations were conducted, in the presence (5 x 100 ns) and absence (5 x 100 ns) of CRD cholesterol.

To assess the impact of cholesterol on the stability of SMO, the root mean square deviation (r.m.s.d.) was calculated for the CRD and the Linker + 7TM domain (Figure 6.3). In each case the domains for which the r.m.s.d. were calculated were fitted to the initial crystal structure coordinates of that domain for each frame of the trajectory, such that the fluctuations observed represent intradomain dynamics relative to the initial crystal structure.

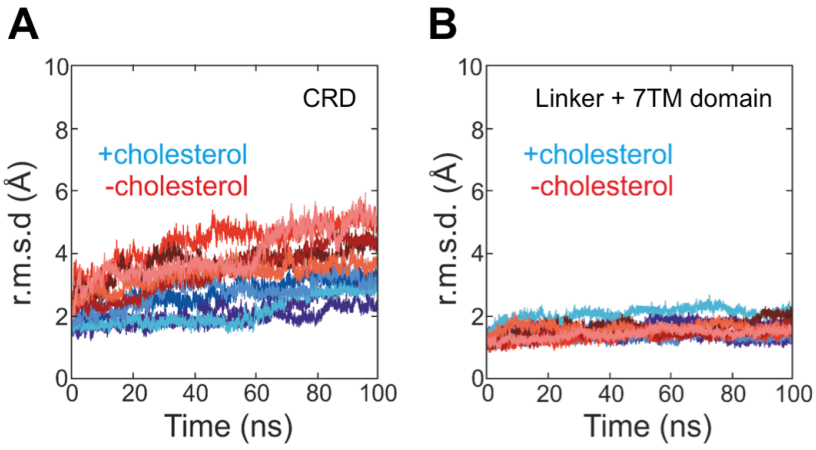

Figure 6. 3: Cholesterol stabilises the extracellular domain of SMO. Root mean square deviation (r.m.s.d) of C_α atoms for the extracellular CRD (A) and Linker + 7TM domains (B), over the course of 5 x 100 ns simulations in the presence (blue) and absence (red) of CRD cholesterol. R.m.s.d values were computed from trajectories fitted to the crystal structure coordinates of the domain under consideration, such that the values obtained pertain to intra-domain dynamics relative to the crystal structure (See 6.2 Methods for full details). Adapted from [348].

In the absence of cholesterol, the CRD exhibited a degree of conformational flexibility, with r.m.s.d values of between 3.9-5.8 Å by 100 ns (Figure 6.3 A). However in the

presence of cholesterol a marked stabilisation was seen, with r.m.s.d values of between 2.2-3.8 Å by 100 ns. In contrast, this cholesterol dependence was not seen for the Linker + 7TM domains, which exhibited significantly less conformational fluctuation under both +/- cholesterol conditions (Figure 6.3 B).

Calculation of per residue r.m.s. fluctuations and DSSP secondary structure fluctuations revealed the cholesterol induced stabilisation could be assigned to specific regions of the structure (Figure 6.4). In particular, cholesterol induced stabilisation was seen predominantly around the hydrophobic cholesterol binding pocket (Figure 6.4 A). A smaller degree of propagation to more distal regions was also observed. The most marked cholesterol induced stabilisation was seen around helix four (denoted by an asterisk in Figure 6.4), involved in interactions with the iso-octanyl tail of cholesterol and forming key interactions with the linker domain and ECL3. However on the timescale simulated the conformational variability was not transmitted to these regions. Interestingly, helix four along with the adjacent loop were also seen to show a degree of conformational flexibility in recently resolved crystal structures of the CRD under different ligand-bound states [354].

To probe whether this fluctuation at helix four resulted in loss of secondary structure, the DSSP assigned secondary structure was calculated as a function of simulation time for both the extracellular domain and membrane domain (Figure 6.4 B,C). The extracellular domain secondary structure showed a degree of fluctuation at certain regions. This was particularly evident at helix four, which exhibited a loss of helicity under both +/- cholesterol conditions. This loss was most pronounced in the absence of cholesterol. In contrast, the 7TM membrane domain secondary structure remained stable under both conditions (Figure 6.4 C).

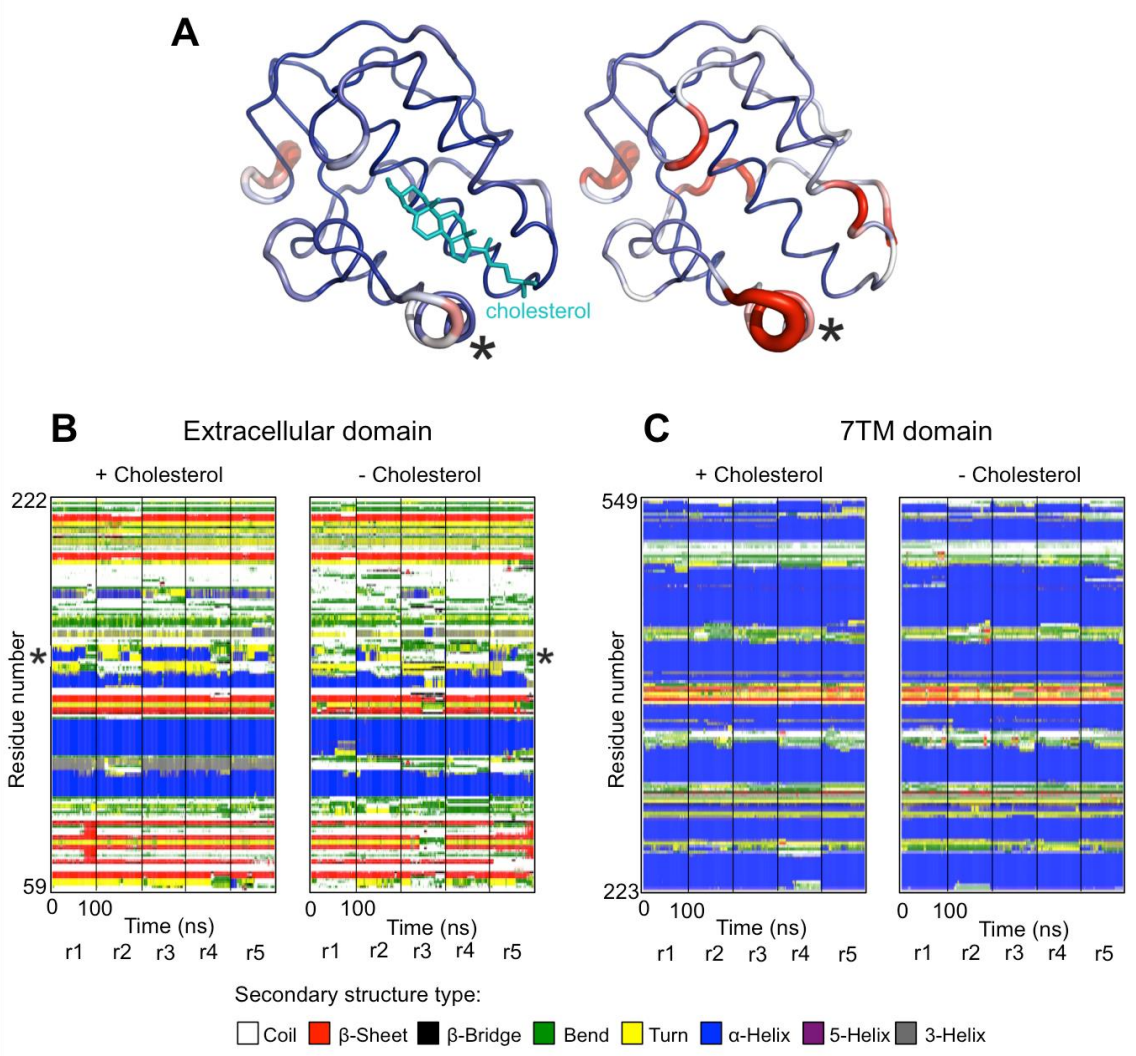

Figure 6. 4: RMSF and DSSP analysis. (A) r.m.s. fluctuations of Cα atoms over 5 x 100 ns atomistic simulations in the presence (left) and absence (right) of CRD cholesterol. The structures are shown as putty representations colored from high conformational stability (low r.m.s. fluctuations; blue/thin) to low conformational stability (high r.m.s fluctuations; red/thick). Secondary structure DSSP matrices for (A) the CRD+Linker domain, and (B) the 7TM membrane domain. Figure adapted from [348].

6.3.2 Inter-domain dynamics

Calculation of the center-of-mass separation (Z_{sep}) between the CRD and 7TM domain over the course of each simulation revealed a degree of inter-domain flexibility (Figure 6.5). In the majority of simulations (8/10) the CRD remained stably orientated atop the Linker and 7TM domains. However in 2/10 simulations the CRD showed significant movement relative to the Linker+7TM domains, moving as a rigid body down to contact the membrane. This movement was seen in one simulation in the presence of cholesterol, and one simulation without cholesterol. In the absence of cholesterol this process began

around 15 ns, and reached a plateau around 35 ns, whilst in the presence of cholesterol movement initiated at 30 ns and reached a plateau by 70 ns. In both cases the relative domain orientations showed no significant continued evolution once a plateau had been reached. Interestingly, the CRD adopted different orientations relative to the Linker+7TM domains in the two simulations (Figure 6.5 B: compare conformations β, γ and ε, ζ), with the CRD in the -cholesterol simulation rotated ~90^{o} in the membrane plane compared to the +cholesterol orientation. Whilst limited sampling circumscribes a more thorough analysis of this process and its possible cholesterol dependence, it does suggest the potential for some inter-domain flexibility between the CRD and Linker+7TM domains of SMO.

Interestingly, the only difference between each of the ten simulation repeats was the assignment of initial random velocity seeds (See 6.2 Methods), thus suggesting a carefully poised arrangement and a degree of stochasticity in the occurrence of major domain movements.

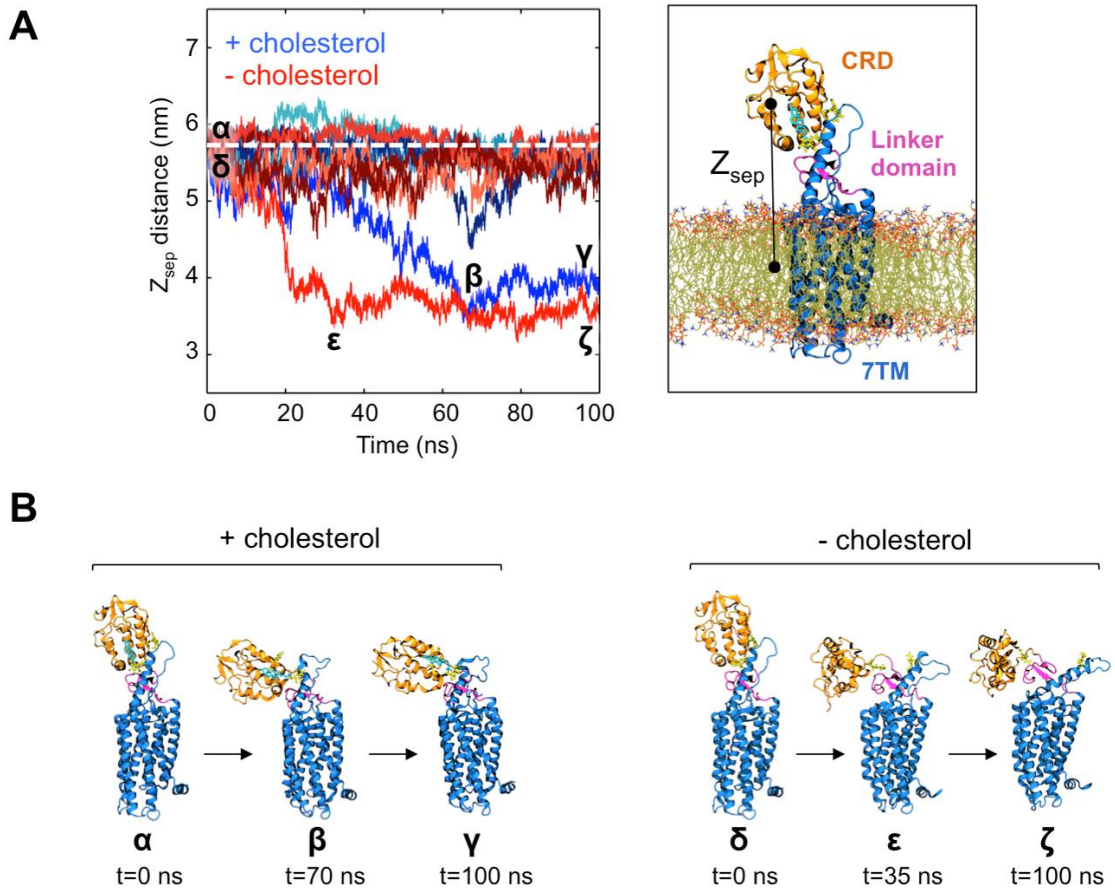

Figure 6. 5: Inter-domain flexibility. (A) Distance between the center-of-mass of the CRD and the 7TM domain (Z_{sep}) as a function of simulation time. Simulations in the presence of cholesterol (n=5) are shown in shades of blue, and those without cholesterol (n=5) in shades of red. The white dashed line indicates Z_{sep} within the crystal structure (PDB id: 5L7D) [348]. (B) Structures extracted from the two simulations in which the CRD showed significant inter-domain movement. The structures shown correspond to the time points indicated in (A), with α β γ taken from the simulation with cholesterol, and δ ε ζ extracted from the simulation without cholesterol. In the latter case, the CRD remains conformationally stable (See CRD r.m.s.d. in Figure 6.3), but collapses into the plane of the figure such that the orientation shown reveals the base of the CRD (rotated ~90º in the plane of the membrane compared to the +cholesterol orientation), in contact with the linker domain in the initial crystal structure.

6.3.3 Prediction of cholesterol interaction sites on the 7TM domain

Although density corresponding to membrane cholesterol molecules has not been observed in available crystal structures of the SMO 7TM domain [346, 348, 355], cholesterol co-purifies with SMO constructs lacking the CRD (unpublished data), suggesting it may also interact with the 7TM domain. Further, cholesterol molecules have been observed to interact with the 7TM domains of a range of GPCRs in crystal structures [97-100].

To identify possible cholesterol interaction hotspots on the 7TM domain of SMO a series of CG-MD simulations were performed. A model of the near full-length structure of SMO (PDB id: 5L7D) [348] was converted to a CG representation using the MARTINI 2.2 force field [189], before being embedded into a POPC lipid bilayer *via* self-assembly simulations [213]. POPC lipids were subsequently randomly exchanged [184] for cholesterol molecules to form a symmetric membrane with 3:1 molar ratio of POPC:cholesterol, approximating the phospholipid:sterol ratio found in a generic plasma membrane [184]. This process was repeated eight times to form eight different initial bilayer configurations. Each system was then run for 5 μs of CGMD (See Methods 6.2 for further details).

Calculation of time averaged 3D density maps for the spatial occupancy of cholesterol around SMO (Figure 6.6 A), as well as computation of the number of contacts formed between each residue of the protein and cholesterol (Figure 6.6 B) revealed a number of interaction hotspots. In particular, a preference was seen for cholesterol interaction with TM helices 2, 3, and 4 within the extracellular leaflet. A degree of interaction was also seen on the intracellular leaflet portions of TM helices 4, 5, 6, and 7.

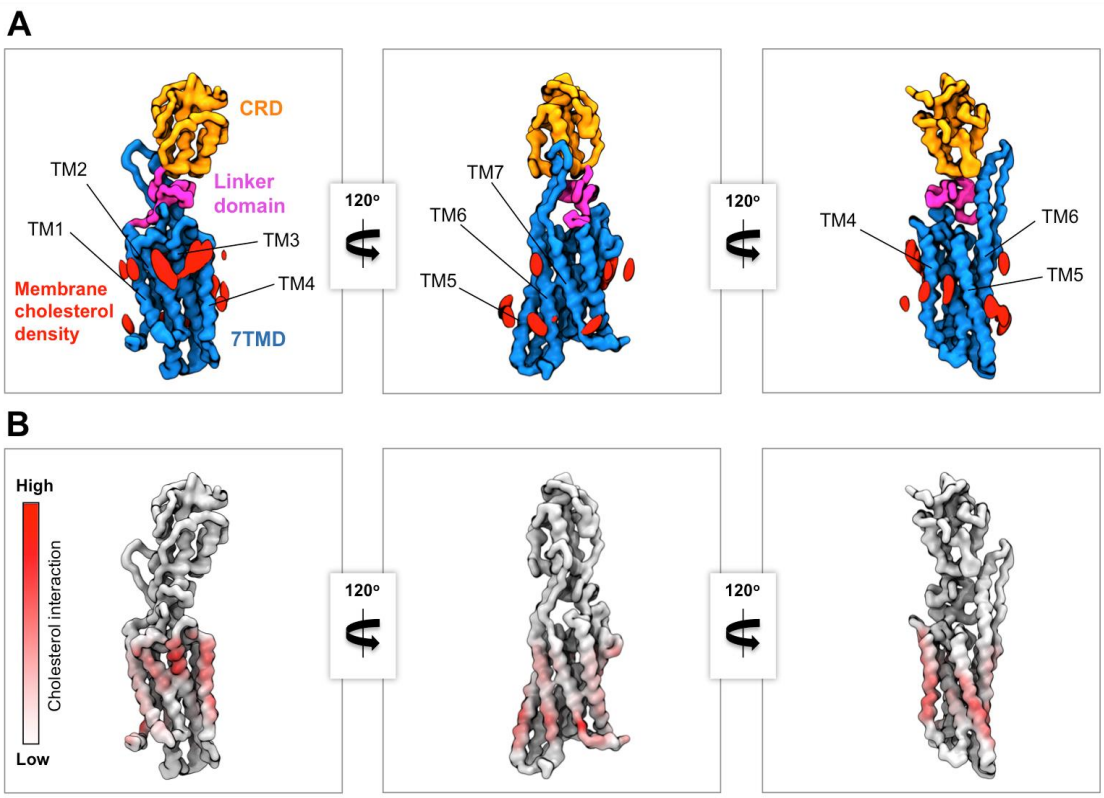

Figure 6. 6: Membrane cholesterol interactions with SMO. (A) Density maps showing the time-averaged spatial occupancy of membrane cholesterol (red surface) around SMO. Densities were calculated using the VolMap tool of VMD and are contoured to indicate densities attributable to individual cholesterol molecules. (B) Cholesterol contacts mapped onto structure. Each residue is coloured from white (low interaction) to red (high interaction) according to the degree of interaction with membrane cholesterol. Contacts were calculated over 8 x 5 μs of CGMD, using a 6 Å cut-off to define contact.

Mapping of cholesterol contact data onto sequence revealed the cholesterol localisation around TM helices 2, 3, and 4, was predominantly mediated *via* interaction with a triad of Isoleucine residues, Ile316, Ile317, and Ile320, located at the extracellular tip of TM3. A high degree of interaction was observed with a range of other residues, particularly Trp537 at the intracellular tip of TM7.

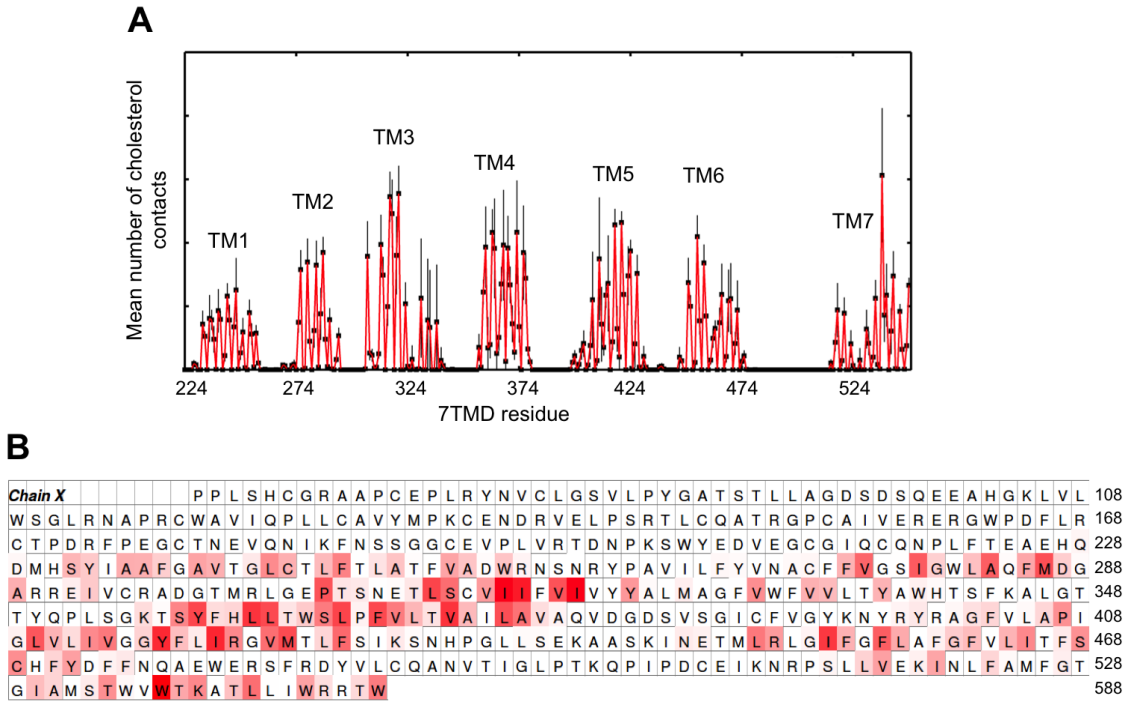

Figure 6. 7: Per residue SMO cholesterol contacts. (A) Cholesterol contacts shown as a function of residue position. Contacts were calculated over 8 x 5 μs of CGMD. Black error bars indicate the std. (B) The contact data shown in (A) mapped onto primary sequence. Each residue is coloured from white (low interaction) to red (high interaction).

Convergence of the observed cholesterol interaction patterns was assessed by comparing contact maps calculated from 3 x 5 μs, and 8 x 5 μs of CGMD (Figure 6.8). The maps showed no significant deviation, indicating that the prediction is unlikely to evolve upon additional of further repeat simulations.

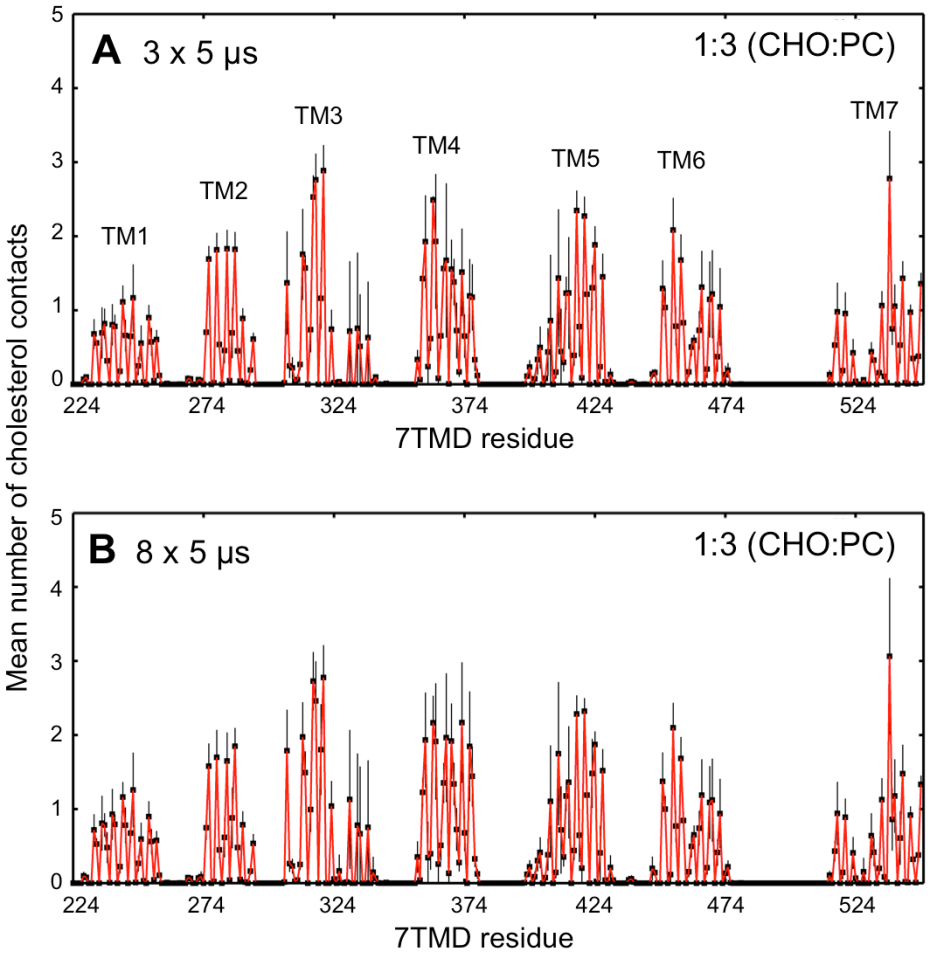

Figure 6. 8: Convergence of cholesterol interaction patterns. Cholesterol contacts with SMO are shown as a function of residue position for the 7TM domain. (A) Contacts calculated over 3 x 5 μs of CGMD, and (B) contacts calculated over 8 x 5 μs of CGMD. Black error bars indicate the std. A 6 Å cut-off distance was applied to define contacts.

These two component PC:cholesterol mixtures represent a relatively simple bilayer composition. To assess the robustness of the results to increasing lipid complexity, the set of 8 x 5 μs CGMD simulations were repeated in more complex asymmetric bilayers approximating some of the lipid complexity found *in vivo* [248].

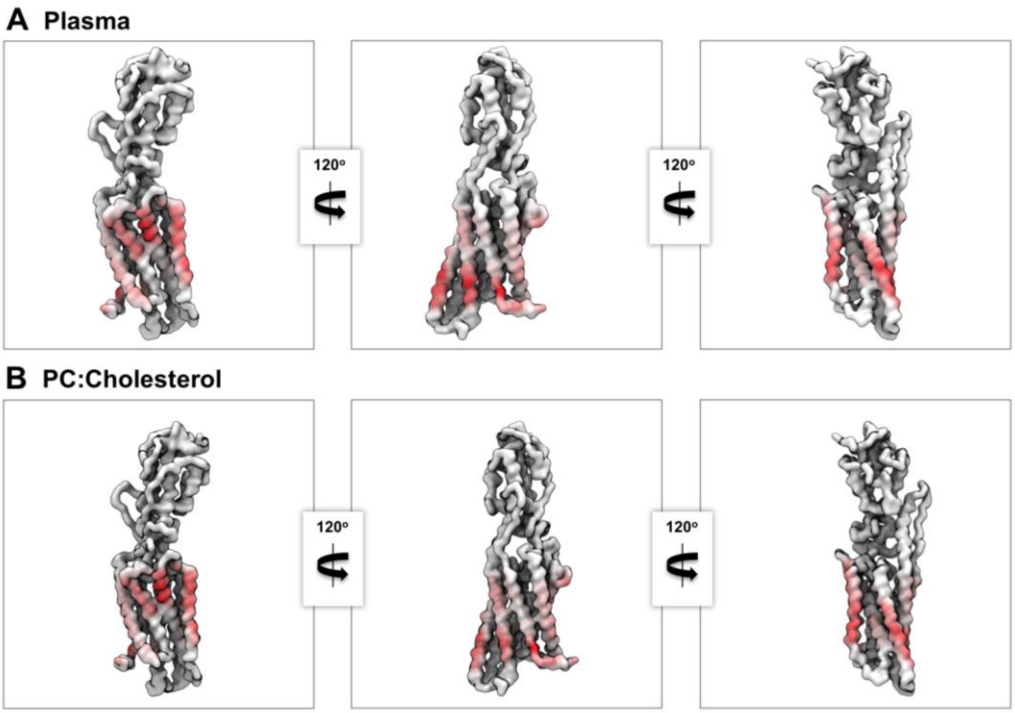

Figure 6. 9: Cholesterol interactions within more complex membrane compositions. The backbone trace of SMO is shown as a surface coloured from white (no interaction) to red (high interaction) according to the degree of contact formed between each residue and membrane cholesterol. The patterns obtained in a plasma membrane mimetic bilayer environment (A) are broadly similar to those seen in a simple two-component PC:Cholesterol bilayer (B).

The patterns of cholesterol interaction with SMO in the plasma membrane mimetic simulations (Figure 6.9 A) showed good agreement with those observed in simpler two-component mixtures (Figure 6.9 B; See also: Figure D1), indicating the robustness of the predictions to inclusion of other lipid species.

6.4 Discussion

These atomistic simulations of the near full-length structure of SMO embedded in a lipid bilayer suggest a degree of conformational flexibility within the extracellular CRD. This flexibility showed a marked reduction in the presence of cholesterol, suggesting cholesterol stabilises the CRD. Consistent with this observation, thermostability assays subsequently performed by collaborators showed a decrease in SMO stability upon MβCD depletion of cholesterol [348]. Analysis of per residue r.m.s. fluctuations revealed this cholesterol induced stabilisation occurred predominantly around the hydrophobic

pocket interacting with cholesterol, in addition to some propagation to more distal regions. In particular, helix four showed the most significant stabilisation upon addition of cholesterol. Interestingly, subsequent to this observation a number of crystal structures emerged of the isolated CRD of SMO, in an apo form and under a range of ligand bound states [354]. The major structural deviation between these structures occurred in helix four and the adjacent loop regions. Consistent with the simulation-based predictions that this may be a point of conformational flexibility within the structure. Further, helix four forms interactions with the iso-octanyl tail of cholesterol, as well as key interactions with the Linker domains and ECL3 of the 7TM domain and is thus an attractive candidate for transmitting possible conformational changes between domains. However on the timescales simulated the Linker+7TM domains remained conformationally stable. Further microsecond scale MD simulations are likely to be beneficial in understanding possible pathways for conformational communication between domains. Ultimately, a structure of SMO with an activated 7TM domain may reveal the mechanism of inter-domain communication.

In the majority of simulations the CRD remained stacked atop the Linker+7TM domains. However in 2/10 simulations the CRD exhibited significant rigid body movement relative to the Linker+7TM domains. In both cases this process involved bending at the CRD-Linker interface and ca. 2 nm downwards movement of the CRD as a rigid body to contact the membrane surface. Such movements have previously been seen for membrane proteins with soluble ectodomains, e.g. EGFR [265], the EphrinA2 receptor [132], and the glucagon receptor [356]. Differing 'bent' conformations were adopted in these two simulations, with the CRD in the -cholesterol simulation undergoing an ~90^{o} rotation in the plane of the membrane relative to the +cholesterol simulation. Whilst sampling limitations prohibit a more thorough analysis of this process and its possible cholesterol dependence, it does suggest the capacity for some stochastic inter-domain flexibility within the structure. Interestingly, a number of developments emerged subsequent to these simulations which also suggest the capacity for inter-domain flexibility. Firstly, two further structures of near full-length SMO were solved bound to the 7TM domain antagonist vismodegib (PDB id: 5L7I), and to TC114 (PDB id: 5V56), respectively. In both structures the CRD underwent a rotation [348] compared to the original structure (PDB id: 5L7D) (See Figure 6.10). Secondly, small angle X-ray scattering experiments on SMO under differing ligand regimes suggest differences in the globularity of SMO conformation with different ligands [348], indicative of inter-domain

movement influenced by ligand binding. Thirdly, recently reported 1 μs duration atomistic simulations of near-full length SMO from the iHuman institute in Shanghai also report a degree of CRD flexibility [355]. Taken together, it seems likely the CRD possesses the capacity for some inter-domain movement, the challenge now will be to make a quantitative assessment as to the degree of movement and tie this to function (See Chapter 7: Future directions).

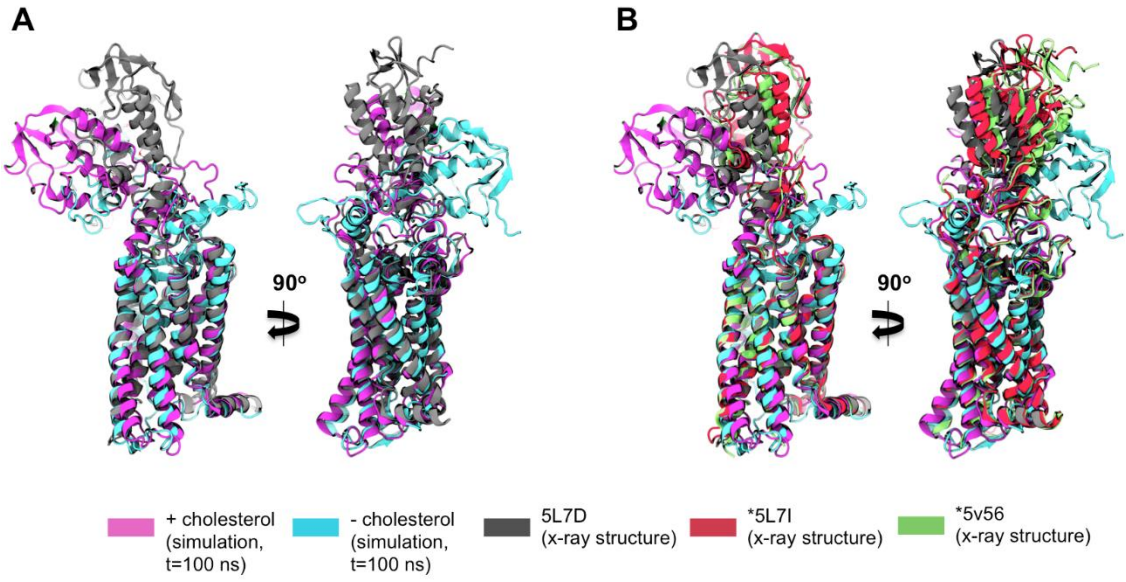

Figure 6. 10: Structural alignment comparing inter-domain flexibility. SMO structures are shown aligned to the Cα atoms of TM helices. (A) Alignment of the final snapshots (t=100 ns) from the two simulations exhibiting significant inter-domain movement (magenta, cyan), and the original crystal structure (5L7D, grey) used to initiate those simulations. (B) The corresponding alignment, including the *two SMO structures to emerge subsequent to the simulations (5L7I, red; 5V56, green).

The CGMD simulations performed in this thesis suggest the capacity for membrane cholesterol interaction with the 7TM domain, and further that these interactions are localised to specific hotspots rather than occurring in an annular manner. In particular the extracellular tips of TM2, 3, and 4 showed high levels of cholesterol interaction, along with the intracellular portions of TM4, 5, 6, and 7. Comparing to available cholesterol-containing crystal structures of other GPCRs (Table 6.1), SMO appears to show a distinct pattern of cholesterol interaction. This perhaps not surprising, given 1) the deviation between available crystal structures, 2) SMO possesses <10% sequence identity with

Class A GPCRs [346], and 3) available functional data [35] suggest that cholesterol binding and modulation appears GPCR-type specific.

Table 6. 1: Cholesterol binding sites on GPCRs.

GPCR	Intracellular	Extracellular
Smoothened[a] (Class F)	TM4, 5, 6, 7	TM2, 3, 4
β_2AR (Class A) [97, 98]	TM1, 2, 3, 4	None
Adenosine A_{2A}R (Class A) [100]	None	TM3, 6
$mGlu_1$R (Class C) [99]	None	TM1, 2

Comparison of simulation-based predictions (a) for SMO cholesterol interaction sites, with those in available crystal structures of other GPCRs.

These predictions have been passed to collaborators at the Rohatgi laboratory in Stanford, to provide guidance for cluster based mutagenesis of SMO.

6.5 Conclusions

Multiscale simulations have been performed to explore the domain dynamics and cholesterol interactions of SMO. Atomistic simulations of the near full-length structure in a lipid bilayer revealed a marked stabilising effect of extracellular cholesterol on intra-domain dynamics of the CRD. In a minority of simulations, the CRD also exhibited the capacity for inter-domain rigid-body bending relative to the Linker+7TM domains. This flexing is qualitatively consistent with a range of subsequently published experimental and simulation data [355]. Whilst limited sampling in the current set of simulations circumscribed predictions as to the cholesterol dependence of this process, it nonetheless suggests the capacity for inter-domain flexibility within smoothened worthy of further investigation. Establishing methodology for characterising such domain movements may also have utility in studying the extracellular domains of other GPCRs, such as the recently determined full-length structure of the glucagon receptor [357]. Finally, the methodology outlined in Chapter 5 was applied to predict cholesterol interaction sites on the 7TM domain of SMO. These predictions will inform efforts to experimentally identify membrane cholesterol interaction sites *via* cluster based mutagenesis of the 7TM domain.

"Life is finite."

- Mark S. P. Sansom

7

Future directions

In this chapter I discuss some of the pre-eminent questions arising from the work described within this thesis, and present my thoughts on topics worthy of further investigation.

Competitive effects of lipids on helix dimerization

In Chapter 3, calculation of 1D PMF profiles for lateral lipid interactions with the EGF receptor revealed a range of well-depths. Of particular interest, the tight binding of PIP_2 lipids (ca. -40 kJ/mol for the -5 ionization state) may be considered in the context of equivalent 1D PMF calculations previously performed for the dimerization of a range of TM helices [216, 218, 267]. These studies report values in the order of ca. -20 kJ/mol [267] to ca. -60 kJ/mol [228] depending on helix identity. It seems plausible therefore that one mode of lipid modulation of receptor function [239] may be *via* direct competition between tight binding lipid species and TM helices. Such a model has been previously proposed by Stangl and Schneider [269], and the calculations reported here would seem to support its credibility. A highly attractive future study could involve conducting PMF calculation for the dimerization of lipid modulated TM helix dimers (such as EGF) in membranes of different lipid composition, in the first instance + PIP_2 and – PIP_2 membranes. Such a study could be complemented by *in vitro* FRET measurements under defined lipid compositions [358].

Cardiolipin and ANT

In Chapter 4, a variety of simulation approaches were employed to show that CL binds to ANT at three specific sites, and provide additional molecular and energetic detail to complement existing experimental data. However the question remains as to how CL mechanistically stabilises and promotes ANT function. Simulations in large membranes suggested a possible influence on oligomerisation, with the three CL binding sites co-localising with favoured protein-protein interactions interfaces, and a greater degree of oligomerisation seen in membranes lacking CL. One can therefore speculate as to whether CL is involved in preventing aggregation *in vivo*, where ANT1 can constitute more than 10% of inner mitochondrial membrane proteins by mass. However the limited sampling in the current simulation prohibit more detailed analysis of CL involvement in ANT oligomerisation. These questions may benefit from better sampling of the large membrane systems and non-equilibrium approaches, e.g. PMF calculations for ANT lateral interactions. Experimentally, EM studies under defined CL regimes would be of value. Further, in the work described here only a single conformational state is simulated, the cytoplasmic facing conformation. Should the matrix facing state, or indeed possible occluded states, of ANT merge in the future, then the above simulations may be worth revisiting. In particular examining possible differential CL interactions between conformational states, and atomistic simulations to investigate possible mechanistic aspects of CL interactions on these conformations.

Conservation of PIP_2 interactions with GPCRs

The conformational conservation of GPCRs of known structure [343] raises the question as to whether the interactions of PIP_2 with NTS1 seen in Chapter 5 may be applicable to other GPCRs. The positive inside rule for membrane proteins intuitively suggests some degree of PIP_2 interaction is likely. The exact degree of possible interaction, the interaction pattern (i.e. where interactions occur on the protein), and how this relates to GPCR type and functional state are not immediately obvious from existing structures, and merits further investigation *via* high-throughput simulations of the type previously employed for aquaporins [242] and receptor tyrosine kinases [32]. Indeed prior to the work described here I co-supervised a Masters student working to create a framework for expanding simulations to other GPCRs of known structure. This work is now ongoing as the main project of a DPhil student and I am hopeful will yield interesting results.

As novel native mass spectrometry techniques are being rapidly developed in partner laboratories, it is likely we may be able to assess the possible functional consequences (e.g. G protein coupling) of these interactions in the not-too-distant future.

SMO inter-domain dynamics, and membrane cholesterol interactions

The rigid body inter-domain movements of the CRD seen in Chapter 6 are intriguing. Firstly, there is precedent for similar movements, e.g. in the glucagon receptor [356], and in the EGF receptor [265]. Secondly, subsequent to the simulations described here, additional crystallographic and SAXS data emerged also suggesting the capacity for CRD movements relative to the 7TM domain (See Chapter 6 for further discussion of this data). It would be extremely interesting to run long timescale simulations to further examine inter-domain dynamics and the possible influence of ligands and glycans on this process. Should one pursue this route, membrane composition and particularly the presence of glycolipids is likely to exert an influence. E.g. the bulky headgroups of glycolipids extend the effective head group region of the bilayer by as much as ca. 1 nm, depending on the exact species. This may be expected to influence domain dynamics and possibly circumscribe large movements. Similar glycan 'cushioning' effects have been seen for N-glycosylated EGF receptor [256].

Importantly, significant inter-domain movement was only seen in 2/10 simulations reported here, suggesting a degree of stochasticity. Achieving sufficient sampling of these slow, stochastic, movements is therefore likely to be a serious challenge to future investigations. Enhanced sampling approaches such as metadynamics [359] may be required.

Regarding membrane cholesterol interactions with the 7TM domain, the preliminary predictions made here suggest possible cholesterol interaction sites and represent a useful first-pass guide for current experimental mutagenesis efforts. However I believe there is scope to refine the predictions, in particular by testing their robustness to an alternative cholesterol model [360], cholesterol concentration, and bilayer thickness. Further, whilst the tight binding of PIP_2 lipids seen in Chapter 3 and 6 precludes a meaningful analysis of residence times on the simulated timescales, visual inspection of SMO-cholesterol trajectories suggestions more transient patterns of interactions which may provide sufficient sampling of association/dissociation to prove amenable to such residence time calculations. This could provide an important temporal perspective to complement existing analysis. Finally, it would be attractive to see if atomistic simulations of the 7TM

domain in the presence and absence of cholesterol can be correlated to the results of experimental thermostability assays of the isolated 7TM domain under the corresponding +/- cholesterol conditions, possibly providing mechanistic insight into the observation that cholesterol stabilises SMO [348].

Final remarks and remaining challenges

The emergence of a large body (>100) of x-ray and cryoEM structures of membrane proteins containing densities attributed to specific lipid molecules [14], together with continued demonstration of the functional significance of lipid binding [75, 238, 239], demands complementary approaches to dissect the dynamical, energetic, and mechanistic details of these interactions in defined membrane environments. I hope that the work presented in this thesis helps demonstrate the utility of MD simulations in fulfilling these roles. In particular the scope for the CG MARTINI force field to provide valuable information on lipid-protein interactions; information which is often less tractable to atomistic simulation due to sampling constraints. The field is now going in two major directions, a) toward high-throughput studies and systematic analysis [273], and b) toward increasing complexity in terms of lipids, proteins, and supporting structures (e.g. the cytoskeleton [361] and bacterial cell walls [362]). The former case is to be welcomed. I believe there is sufficient support in the literature demonstrating the utility and accuracy of simulations to warrant expansion and scaling-up, and indeed such studies may reveal shortcomings which can be improved upon. The latter case regarding complexity is to be theoretically welcomed, with a healthy dose of caution. In CG MARTINI the components (be they lipids, proteins, glycans etc.) are typically parameterised in simple environments, and it is assumed the parameters may be extrapolated and remain sufficiently accurate in complex environments. At a more fundamental level, atomistic force fields often parameterise lipid and proteins separately, and then unite the components under the arguable assumption of compatibility [363]. This often results from limitations in the availability of structural and energetic experimental data with which to compare. As the complexity of membrane protein/lipid systems continues to increase [296], the field would benefit immensely from rigorous studies to address these factors. Doing so will ultimately lead to more accurate biological predictions, and increased confidence and application of the methodology.

Regarding the assembly of complex membranes, the accurate choice of composition is often limited by the availability of high quality lipidomics data on which to base the

model. The compositions in Chapters 5 and 6 represent generic plasma membrane models [184, 364] which are an amalgamation of signalling states, metabolic states, ages, and membrane locations. However significant advances have been made in lipidomics technologies, with e.g. state-of-the-art bespoke 'on-order' lipidomics analysis services now commercially available. A particularly attractive possibility presented by such advances is not only the creation of various 'generic' membranes from different cells and organelles already taking place, but also the possibility to create membrane compositions under e.g. 'diseased' and 'healthy' conditions, as well as different signalling states (e.g. enriched/depleted in PIP_2). These developments hold the promise to open exciting new avenues for simulations of lipid-protein interactions.

References

1. Shevchenko, A., and Simons, K. (2010). Lipidomics: coming to grips with lipid diversity. Nat. Rev. Mol. Cell Biol. *11*, 593-598.
2. Pedersen, B.P., and Nissen, P. (2015). Membrane proteins - do we catch up with the breathless pace of soluble protein structural biology? Biochim. Biophys. Acta *1850*, 447-448.
3. Vinothkumar, K.R. (2015). Membrane protein structures without crystals, by single particle electron cryomicroscopy. Curr. Opin. Struct. Biol. *33*, 103-114.
4. Fahy, E., Subramaniam, S., Murphy, R.C., Nishijima, M., Raetz, C.R.H., Shimizu, T., Spener, F., van Meer, G., Wakelam, M.J.O., and Dennis, E.A. (2009). Update of the LIPID MAPS comprehensive classification system for lipids. J. Lipid Res. *50*, S9-S14.
5. Koshy, C., and Ziegler, C. (2015). Structural insights into functional lipid–protein interactions in secondary transporters. Biochim. Biophys. Acta *1850*, 476-487.
6. Lopes, C.M.B., Zhang, H.L., Rohacs, T., Jin, T.H., Yang, J., and Logothetis, D.E. (2002). Alterations in conserved Kir channel-PIP2 interactions underlie channelopathies. Neuron *34*, 933-944.
7. Romanenko, V.G., Rothblat, G.H., and Levitan, I. (2002). Modulation of endothelial inward-rectifier K^+ current by optical isomers of cholesterol. Biophys. J. *83*, 3211-3222.
8. Michailidis, I.E., Rusinova, R., Georgakopoulos, A., Chen, Y., Iyengar, R., Robakis, N.K., Logothetis, D.E., and Baki, L. (2011). Phosphatidylinositol-4,5-bisphosphate regulates epidermal growth factor receptor activation. Pflugers Arch. *461*, 387-397.
9. Zocher, M., Zhang, C., Rasmussen, S.G.F., Kobilka, B.K., and Muller, D.J. (2012). Cholesterol increases kinetic, energetic, and mechanical stability of the human beta(2)-adrenergic receptor. Proc. Natl. Acad. Sci. USA *109*, E3463-E3472.
10. Saxena, R., and Chattopadhyay, A. (2012). Membrane cholesterol stabilizes the human serotonin1A receptor. Biochim. Biophys. Acta *1818*, 2936-2942.
11. Barrantes, F.J. (2007). Cholesterol effects on nicotinic acetylcholine receptor. J. Neurochem. *103*, 72-80.
12. Sooksawate, T., and Simmonds, M.A. (2001). Effects of membrane cholesterol on the sensitivity of the $GABA_A$ receptor to GABA in acutely dissociated rat hippocampal neurones. Neuropharmacol. *40*, 178-184.
13. Laganowsky, A., Reading, E., Allison, T.M., Ulmschneider, M.B., Degiacomi, M.T., Baldwin, A.J., and Robinson, C.V. (2014). Membrane proteins bind lipids selectively to modulate their structure and function. Nature *510*, 172-175.
14. Yeagle, P.L. (2014). Non-covalent binding of membrane lipids to membrane proteins. Biochim. Biophys. Acta *1838*, 1548-1559.
15. Hansen, S.B., Tao, X., and Mackinnon, R. (2011). Structural basis of PIP_2 activation of the classical inward rectifier K^+ channel Kir2.2. Nature *477*, 495-498.
16. Durrant, J., and McCammon, J.A. (2011). Molecular dynamics simulations and drug discovery. BMC Biol. *9*, 71.

17. Caffrey, M. (2015). A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes. Acta Cryst. F *71*, 3-18.
18. Stansfeld, P.J., and Sansom, M.S.P. (2011). Molecular simulation approaches to membrane proteins. Structure *19*, 1562-1572.
19. Powl, A.M., East, J.M., and Lee, A.G. (2007). Different effects of lipid chain length on the two sides of a membrane and the lipid annulus of MscL. Biophys. J. *93*, 113–122.
20. Marsh, D. (2010). Electron spin resonance in membrane research: protein-lipid interactions from challenging beginnings to state of the art. Eur. Biophys. J. *39*, 513-525.
21. Das, N., Murray, D.T., and Cross, T.A. (2013). Lipid bilayer preparations of membrane proteins for oriented and magic-angle spinning solid-state NMR samples. Nature Protocols *8*, 2256-2270.
22. Hedger, G., and Sansom, M.S.P. (2016). Lipid interaction sites on channels, transporters and receptors: Recent insights from molecular dynamics simulations. Biochim. Biophys. Acta *1858*, 2390-2400.
23. Claypool, S.M., Oktay, Y., Boontheung, P., Loo, J.A., and Koehler, C.M. (2008). Cardiolipin defines the interactome of the major ADP/ATP carrier protein of the mitochondrial inner membrane. J. Cell Biol. *182*, 937-950.
24. Robinson, N.C., Zborowski, J., and Talbert, L.H. (1990). Cardiolipin-depleted bovine heart cytochrome c oxidase: binding stoichiometry and affinity for cardiolipin derivatives. Biochem. *29*, 8962-8969.
25. Li, Y., Ge, M., Ciani, L., Kuriakose, G., Westover, E.J., Dura, M., Covey, D.F., Freed, J.H., Maxfield, F.R., Lytton, J., et al. (2004). Enrichment of endoplasmic reticulum with cholesterol inhibits sarcoplasmic-endoplasmic reticulum calcium ATPase-2b activity in parallel with increased order of membrane lipids: implications for depletion of endoplasmic reticulum calcium stores and apoptosis in cholesterol-loaded macrophages. J. Biol. Chem. *279*, 37030-37039.
26. Hanson, M., and Peach, R. (2014). Structural Biology of the S1P1 Receptor. In Sphingosine-1-Phosphate Signaling in Immunology and Infectious Diseases, Volume 378, M.B.A. Oldstone and H. Rosen, eds. (Springer International Publishing), pp. 23-53.
27. Stansfeld, P.J., Hopkinson, R.J., Ashcroft, F.M., and Sansom, M.S.P. (2009). The PIP_2 binding site in Kir channels: definition by multi-scale biomolecular simulations. Biochem. *48*, 10926–10933.
28. Schmidt, M.R., Stansfeld, P.J., Tucker, S.J., and Sansom, M.S.P. (2013). Simulation-based prediction of phosphatidylinositol 4,5-bisphosphate binding to an ion channel. Biochem. *52*, 279-281.
29. Marsh, D., and Pali, T. (2004). The protein-lipid interface: perspectives from magnetic resonance and crystal structures. Biochim. Biophys. Acta *1666*, 118-141.
30. Marsh, D., and Horváth, L.I. (1998). Structure, dynamics and composition of the lipid-protein interface. Perspectives from spin-labelling. Biochim. Biophys. Acta *1376*, 267-296.
31. Koldsø, H., and Sansom, M.S.P. (2012). Local lipid reorganization by a transmembrane protein domain. J. Phys. Chem. Lett. *3*, 3498-3502.
32. Hedger, G., Sansom, M.S.P., and Koldsø, H. (2015). The juxtamembrane regions of human receptor tyrosine kinases exhibit conserved interaction sites with anionic lipids. Scientific Reports *5*, 9198.

33. Pebay-Peyroula, E., Dahout-Gonzalez, C., Kahn, R., Trezeguet, V., Lauquin, G.J., and Brandolin, G. (2003). Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside. Nature *426*, 39-44.
34. Shinzawa-Itoh, K., Aoyama, H., Muramoto, K., Terada, H., Kurauchi, T., Tadehara, Y., Yamasaki, A., Sugimura, T., Kurono, S., Tsujimoto, K., et al. (2007). Structures and physiological roles of 13 integral lipids of bovine heart cytochrome c oxidase. EMBO J. *26*, 1713-1725.
35. Guixà-González, R., Albasanz, J.L., Rodriguez-Espigares, I., Pastor, M., Sanz, F., Martí-Solano, M., Manna, M., Martinez-Seara, H., Hildebrand, P.W., Martín, M., et al. (2017). Membrane cholesterol access into a G-protein-coupled receptor. Nature Communications *8*, 14505.
36. Neale, C., Herce, Henry D., Pomès, R., and García, Angel E. (2015). Can Specific Protein-Lipid Interactions Stabilize an Active State of the Beta 2 Adrenergic Receptor? Biophys. J. *109*, 1652-1662.
37. Aryal, P., Abd-Wahab, F., Bucci, G., Sansom, M.S.P., and Tucker, S.J. (2015). Influence of lipids on the hydrophobic barrier within the pore of the TWIK-1 K2P channel. Channels *9*, 44-49.
38. Sengupta, D., and Chattopadhyay, A. (2015). Molecular dynamics simulations of GPCR–cholesterol interaction: An emerging paradigm. Biochim. Biophys. Acta *1848*, 1775-1782.
39. Fantini, J., Yahi, N., and Garmy, N. (2013). Cholesterol accelerates the binding of Alzheimer's beta-amyloid peptide to ganglioside GM1 through a universal hydrogen-bond-dependent sterol tuning of glycolipid conformation. Front. Physiol. *4*, DOI: 10.3389/fphys.2013.00120
40. Planas-Iglesias, J., Dwarakanath, H., Mohammadyani, D., Yanamala, N., Kagan, Valerian E., and Klein-Seetharaman, J. (2015). Cardiolipin interactions with proteins. Biophys. J. *109*, 1282-1294.
41. Dror, R.O., Dirks, R.M., Grossman, J.P., Xu, H.F., and Shaw, D.E. (2012). Biomolecular simulation: a computational microscope for molecular biology. Ann. Rev. Biophys. *41*, 429-452.
42. Marrink, S.J., and Tieleman, D.P. (2013). Perspective on the Martini model. Chem. Soc. Rev. *42*, 6801-6822.
43. Voth, G.A. (2008). Coarse-Graining of Condensed Phase and Biomolecular Systems, (CRC Press).
44. Stansfeld, P.J., and Sansom, M.S.P. (2011). From coarse-grained to atomistic: a serial multi-scale approach to membrane protein simulations. J. Chem. Theor. Comp. *7*, 1157–1166.
45. Wassenaar, T.A., Pluhackova, K., Böckmann, R.A., Marrink, S.J., and Tieleman, D.P. (2014). Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models. J. Chem. Theor. Comput. *10*, 676-690.
46. Ayton, G.A., Noid, W.G., and Voth, G.A. (2007). Multiscale modeling of biomolecular systems: in serial and in parallel. Curr. Opin. Struct. Biol. *17*, 192-198.
47. Arnarez, C., Mazat, J.-P., Elezgaray, J., Marrink, S.-J., and Periole, X. (2013). Evidence for cardiolipin binding sites on the membrane-exposed surface of the cytochrome bc1. J. Amer. Chem. Soc. *135*, 3112–3120.
48. Koldsø, H., and Sansom, M.S.P. (2015). Organization and Dynamics of Receptor Proteins in a Plasma Membrane. J. Am. Chem. Soc. *137*, 14694-14704.

49. Dror, R.O., Mildorf, T.J., Hilger, D., Manglik, A., Borhani, D.W., Arlow, D.H., Philippsen, A., Villanueva, N., Yang, Z.Y., Lerch, M.T., et al. (2015). Structural basis for nucleotide exchange in heterotrimeric G proteins. Science *348*, 1361-1365.
50. Pebay-Peyroula, E., Dahout-Gonzalez, C., Kahn, R., Trezeguet, V., Lauquin, G.J.M., and Brandolin, G. (2003). Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside. Nature *426*, 39-44.
51. Scott, K.A., Bond, P.J., Ivetac, A., Chetwynd, A.P., Khalid, S., and Sansom, M.S.P. (2008). Coarse-grained MD simulations of membrane protein-bilayer self-assembly Structure *16*, 621-630.
52. Faraldo-Gómez, J.D., Smith, G.R., and Sansom, M.S.P. (2002). Setup and optimisation of membrane protein simulations. Eur. Biophys. J. *31*, 217-227.
53. Wolf, M.G., Hoefling, M., Aponte-Santamaría, C., Grubmüller, H., and Groenhof, G. (2010). g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation. J. Comput. Chem. *31*, 2169–2174.
54. Jefferys, E., Sands, Z.A., Shi, J., Sansom, M.S., and Fowler, P.W. (2015). Alchembed: a computational method for incorporating multiple proteins into complex lipid geometries. J. Chem. Theory Comput. *11*, 2743-2754.
55. Wassenaar, T.A., Ingolfsson, H.I., Boeckmann, R.A., Tieleman, D.P., and Marrink, S.J. (2015). Computational lipidomics with Insane: a versatile tool for generating custom membranes for molecular simulations. J. Chem. Theor. Comput. *11*, 2144-2155.
56. van Meer, G., and de Kroon, A.I.P.M. (2011). Lipid map of the mammalian cell. J. Cell Sci. *124*, 5-8.
57. Ingolfsson, H.I., Melo, M.N., van Eerden, F.J., Arnarez, C., Lopez, C.A., Wassenaar, T.A., Periole, X., de Vries, A.H., Tieleman, D.P., and Marrink, S.J. (2014). Lipid organization of the plasma membrane. J. Amer. Chem. Soc. *136*, 14554-14559.
58. Koldsø, H., Shorthouse, D., Hélie, J., and Sansom, M.S.P. (2014). Lipid clustering correlates with membrane curvature as revealed by molecular simulations of complex lipid bilayers. PLoS Comp. Biol. *10*, e1003911.
59. Wu, E.L., Cheng, X., Jo, S., Rui, H., Song, K.C., Dávila-Contreras, E.M., Qi, Y., Lee, J., Monje-Galvan, V., Venable, R.M., et al. (2014). CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. J. Comput. Chem. *35*, 1997-2004.
60. Martínez, L., Andrade, R., Birgin, E.G., and Martínez, J.M. (2009). PACKMOL: A package for building initial configurations for molecular dynamics simulations. J. Comput. Chem. *30*, 2157-2164.
61. Sommer, B., Dingersen, T., Gamroth, C., Schneider, S.E., Rubert, S., Krüger, J., and Dietz, K.-J. (2011). CELLmicrocosmos 2.2 MembraneEditor: A Modular Interactive Shape-Based Software Approach To Solve Heterogeneous Membrane Packing Problems. J. Chem. Inf. Model. *51*, 1165-1182.
62. Kalli, A.C., Sansom, M.S.P., and Reithmeier, R.A.F. (2015). Molecular dynamics simulations of the bacterial UraA H^+-uracil symporter in lipid bilayers reveal a closed state and a selective interaction with cardiolipin. PLoS Comput. Biol. *11*, e1004123.
63. Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., and Olson, A.J. (2009). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J. Comp. Chem. *30*, 2785-2791.

64. Brannigan, G., Hénin, J., Law, R., Eckenhoff, R., and Klein, M.L. (2008). Embedded cholesterol in the nicotinic acetylcholine receptor. Proc. Natl. Acad. Sci. USA *105*, 14418-14423.
65. Hénin, J., Salari, R., Murlidaran, S., and Brannigan, G. (2014). A predicted binding site for cholesterol on the $GABA_A$ receptor. Biophys. J. *106*, 1938-1949.
66. Rosenhouse-Dantsker, A., Noskov, S., Durdagi, S., Logothetis, D.E., and Levitan, I. (2013). Identification of Novel Cholesterol-binding Regions in Kir2 Channels. Journal of Biological Chemistry *288*, 31154-31164.
67. Valiyaveetil, F.I., Zhou, Y., and MacKinnon, R. (2002). Lipids in the structure, folding and function of the KcsA channel. Biochem. *41*, 10771-10777.
68. Deol, S.S., Domene, C., Bond, P.J., and Sansom, M.S.P. (2006). Anionic phospholipids interactions with the potassium channel KcsA: simulation studies. Biophys. J. *90*, 822-830.
69. Doupnik, C.A., Davidson, N., and Lester, H.A. (1995). The inward rectifier potassium channel family. Curr. Opin. Neurobiol. *5*, 268-277.
70. Suh, B.C., and Hille, B. (2008). PIP_2 is a necessary cofactor for ion channel function: How and why? Ann. Rev. Biophys. *37*, 175-195.
71. Hansen, S.B. (2015). Lipid agonism: The PIP2 paradigm of ligand-gated ion channels. Biochim. Biophys. Acta *1851*, 620-628.
72. Zhang, H.L., He, C., Yan, X.X., Mirshahi, T., and Logothetis, D.E. (1999). Activation of inwardly rectifying K+ channels by distinct PtdIns(4,5)P-2 interactions. Nature Cell Biology *1*, 183-188.
73. Huang, C.L., Feng, S.Y., and Hilgemann, D.W. (1998). Direct activation of inward rectifier potassium channels by PIP2 and its stabilization by G beta gamma. Nature *391*, 803-806.
74. Tao, X., Avalos, J.L., Chen, J., and MacKinnon, R. (2009). Crystal Structure of the Eukaryotic Strong Inward-Rectifier K+ Channel Kir2.2 at 3.1 Å Resolution. Science *326*, 1668-1674.
75. Hansen, S.B., Tao, X., and MacKinnon, R. (2011). Structural basis of PIP2 activation of the classical inward rectifier K+ channel Kir2.2. Nature *477*, 495-498.
76. Stansfeld, P.J., Hopkinson, R., Ashcroft, F.M., and Sansom, M.S.P. (2009). PIP2-Binding Site in Kir Channels: Definition by Multiscale Biomolecular Simulations. Biochemistry *48*, 10926-10933.
77. Schmidt, M.R., Stansfeld, P.J., Tucker, S.J., and Sansom, M.S.P. (2013). Simulation-Based Prediction of Phosphatidylinositol 4,5-Bisphosphate Binding to an Ion Channel. Biochemistry *52*, 279-281.
78. Singh, D.K., Rosenhouse-Dantsker, A., Nichols, C.G., Enkvetchakul, D., and Levitan, I. (2009). Direct regulation of prokaryotic Kir channel by cholesterol. J. Biol. Chem. *284*, 30727-30736.
79. Fürst, O., Nichols, C.G., Lamoureux, G., and D'Avanzo, N. (2014). Identification of a cholesterol-binding pocket in inward rectifier K^+ (Kir) channels. Biophys. J. *107*, 2786-2796.
80. D'Avanzo, N., Hyrc, K., Enkvetchakul, D., Covey, D.F., and Nichols, C.G. (2011). Enantioselective protein-sterol interactions mediate regulation of both prokaryotic and eukaryotic inward rectifier K^+ channels by cholesterol. PLoS ONE *6*, e19393.
81. Rosenhouse-Dantsker, A., Noskov, S., Durdagi, S., Logothetis, D.E., and Levitan, I. (2013). Identification of novel cholesterol-binding regions in Kir2 channels. J. Biol. Chem.

82. Grouleff, J., Irudayam, S.J., Skeby, K.K., and Schiøtt, B. (2015). The influence of cholesterol on membrane protein structure, function, and dynamics studied by molecular dynamics simulations. Biochim. Biophys. Acta *1848*, 1783-1795.
83. Thompson, A.J., Lester, H.A., and Lummis, S.C.R. (2010). The structural basis of function in Cys-loop receptors. Quart. Rev. Biophys. *43*, 449-499.
84. Baenziger, J.E., Morris, M.-L., Darsaut, T.E., and Ryan, S.E. (2000). Effect of membrane lipid composition on the conformational equilibria of the nicotinic acetylcholine receptor. J. Biol. Chem. *275*, 777-784.
85. daCosta, C.J.B., Dey, L., Therien, J.P.D., and Baenziger, J.E. (2013). A distinct mechanism for activating uncoupled nicotinic acetylcholine receptors. Nat. Chem. Biol. *9*, 701-707.
86. Cheng, M.H., Xu, Y., and Tang, P. (2009). Anionic Lipid and Cholesterol Interactions with α4β2 nAChR: Insights from MD Simulations. The Journal of Physical Chemistry B *113*, 6964-6970.
87. Brannigan, G., Hénin, J., Law, R., Eckenhoff, R., and Klein, M.L. (2008). Embedded cholesterol in the nicotinic acetylcholine receptor. Proceedings of the National Academy of Sciences *105*, 14418-14423.
88. Baier, C.J., Fantini, J., and Barrantes, F.J. (2011). Disclosure of cholesterol recognition motifs in transmembrane domains of the human nicotinic acetylcholine receptor. Sci. Rep. *1*, 69.
89. Hibbs, R.E., and Gouaux, E. (2011). Principles of activation and permeation in an anion-selective Cys-loop receptor. Nature *474*, 54-60.
90. Hénin, J., Salari, R., Murlidaran, S., and Brannigan, G. (2014). A Predicted Binding Site for Cholesterol on the GABAA Receptor. Biophys. J. *106*, 1938-1949.
91. Pliotas, C., Dahl, A.C.E., Rasmussen, T., Mahendran, K.R., Smith, T.K., Marius, P., Gault, J., Banda, T., Rasmussen, A., Miller, S., et al. (2015). The role of lipids in mechanosensation. Nat. Struct. Mol. Biol. *22*, 991-998.
92. Rosenbaum, D.M., Rasmussen, S.G.F., and Kobilka, B.K. (2009). The structure and function of G-protein-coupled receptors. Nature *459*, 356-363.
93. Oates, J., and Watts, A. (2011). Uncovering the intimate relationship between lipids, cholesterol and GPCR activation. Curr. Opin. Struct. Biol. *21*, 802-807.
94. Ganguly, S., Clayton, A.H.A., and Chattopadhyay, A. Organization of higher-order oligomers of the serotonin1A receptor explored utilizing homo-FRET in live cells. Biophys. J. *100*, 361-368.
95. Harikumar, K.G., Puri, V., Singh, R.D., Hanada, K., Pagano, R.E., and Miller, L.J. (2005). Differential effects of modification of membrane cholesterol and sphingolipids on the conformation, function, and trafficking of the G protein-coupled cholecystokinin receptor. J. Biol. Chem. *280*, 2176-2185.
96. Gimpl, G., Reitz, J., Brauer, S., and Trossen, C. (2008). Oxytocin receptors: ligand binding, signalling and cholesterol dependence. In Prog. Brain Res., Volume Volume 170, D.N. Inga and L. Rainer, eds. (Elsevier), pp. 193-204.
97. Cherezov, V., Rosenbaum, D.M., Hanson, M.A., Rasmussen, S.G.F., Thian, F.S., Kobilka, T.S., Choi, H.-J., Kuhn, P., Weis, W.I., Kobilka, B.K., et al. (2007). High-Resolution Crystal Structure of an Engineered Human β2-Adrenergic G Protein–Coupled Receptor. Science *318*, 1258-1265.
98. Hanson, M.A., Cherezov, V., Griffith, M.T., Roth, C.B., Jaakola, V.-P., Chien, E.Y.T., Velasquez, J., Kuhn, P., and Stevens, R.C. (2008). A Specific Cholesterol Binding Site Is Established by the 2.8 Å Structure of the Human β2-Adrenergic Receptor. Structure *16*, 897-905.

99. Wu, H., Wang, C., Gregory, K.J., Han, G.W., Cho, H.P., Xia, Y., Niswender, C.M., Katritch, V., Meiler, J., Cherezov, V., et al. (2014). Structure of a Class C GPCR Metabotropic Glutamate Receptor 1 Bound to an Allosteric Modulator. Science *344*, 58-64.
100. Liu, W., Chun, E., Thompson, A.A., Chubukov, P., Xu, F., Katritch, V., Han, G.W., Roth, C.B., Heitman, L.H., IJzerman, A.P., et al. (2012). Structural Basis for Allosteric Regulation of GPCRs by Sodium Ions. Science *337*, 232-236.
101. Neale, C., Herce, Henry D., Pomès, R., and García, Angel E. (2015). Can specific protein-lipid interactions stabilize an active state of the beta 2 adrenergic receptor? Biophys. J. *109*, 1652-1662.
102. Prasanna, X., Chattopadhyay, A., and Sengupta, D. (2014). Cholesterol modulates the dimer interface of the β2-adrenergic receptor via cholesterol occupancy sites. Biophys. J. *106*, 1290-1300.
103. Cang, X., Du, Y., Mao, Y., Wang, Y., Yang, H., and Jiang, H. (2013). Mapping the functional binding sites of cholesterol in β2-adrenergic receptor by long-time molecular dynamics simulations. J. Phys. Chem. B *117*, 1085-1094.
104. Cang, X., Yang, L., Yang, J., Luo, C., Zheng, M., Yu, K., Yang, H., and Jiang, H. (2014). Cholesterol-β1AR interaction versus cholesterol-β2AR interaction. Proteins: Struc. Func. Bioinf. *82*, 760-770.
105. Sengupta, D., and Chattopadhyay, A. (2012). Identification of Cholesterol Binding Sites in the Serotonin1A Receptor. J. Phys. Chem. B *116*, 12991-12996.
106. Shan, J.F., Khelashvili, G., Mondal, S., Mehler, E.L., and Weinstein, H. (2012). Ligand-dependent conformations and dynamics of the serotonin 5-HT_{2A} receptor determine its activation and membrane-driven oligomerization properties. PLoS Comp. Biol. *8*, e1002473.
107. Lee, J.Y., and Lyman, E. (2012). Predictions for cholesterol interaction sites on the A2A adenosine receptor. J. Amer. Chen. Soc. *134*, 16512-16515.
108. Khelashvili, G., Grossfield, A., Feller, S.E., Pitman, M.C., and Weinstein, H. (2009). Structural and dynamic effects of cholesterol at preferred sites of interaction with rhodopsin identified from microsecond length molecular dynamics simulations. Proteins: Struc. Func. Bioinf. *76*, 403-417.
109. Horn, J., Kao, T.-C., and Grossfield, A. (2014). Coarse-Grained Molecular Dynamics Provides Insight into the Interactions of Lipids and Cholesterol with Rhodopsin. In G Protein-Coupled Receptors - Modeling and Simulation, Volume 796, M. Filizola, ed. (Springer Netherlands), pp. 75-94.
110. Hurst, D.P., Grossfield, A., Lynch, D.L., Feller, S., Romo, T.D., Gawrisch, K., Pitman, M.C., and Reggio, P.H. (2010). A Lipid Pathway for Ligand Binding Is Necessary for a Cannabinoid G Protein-coupled Receptor. J. Biol. Chem. *285*, 17954-17964.
111. Lee, J.Y., and Lyman, E. (2012). Predictions for Cholesterol Interaction Sites on the A2A Adenosine Receptor. J. Am. Chem. Soc. *134*, 16512-16515.
112. Prasanna, X., Chattopadhyay, A., and Sengupta, D. (2014). Cholesterol Modulates the Dimer Interface of the β2-Adrenergic Receptor via Cholesterol Occupancy Sites. Biophys. J. *106*, 1290-1300.
113. Periole, X., Huber, T., Marrink, S.J., and Sakmar, T.P. (2007). G protein-coupled receptors self-assemble in dynamics simulations of model bilayers. J. Amer. Chem. Soc. *129*, 10126-10132.
114. Periole, X., Knepp, A.M., Sakmar, T.P., Marrink, S.J., and Huber, T. (2012). Structural determinants of the supramolecular organization of G protein-coupled receptors in bilayers. J. Amer. Chem. Soc. *134*, 10959-10965.

115. Wassenaar, T.A., Pluhackova, K., Moussatova, A., Sengupta, D., Marrink, S.J., Tieleman, D.P., and Böckmann, R.A. (2015). High-throughput simulations of dimer and trimer assembly of membrane proteins. The DAFT approach. J. Chem. Theory Comput., 150326134217003.
116. Inagaki, S., Ghirlando, R., White, J.F., Gvozdenovic-Jeremic, J., Northup, J.K., and Grisshammer, R. (2012). Modulation of the interaction between neurotensin receptor NTS1 and Gq protein by lipid. J. Mol. Biol. *417*, 95-111.
117. Dawaliby, R., Trubbia, C., Delporte, C., Masureel, M., Van Antwerpen, P., Kobilka, B.K., and Govaerts, C. (2015). Allosteric regulation of G protein-coupled receptor activity by phospholipids. Nature Chem. Biol. *12*, 35-39.
118. Koldsø, H., and Sansom, M.S.P. (2015). Organization and dynamics of receptor proteins in a plasma membrane. J. Amer. Chem. Soc. *137*, 14694−14704.
119. Guixà-González, R., Javanainen, M., Gómez-Soler, M., Cordobilla, B., Domingo, J.C., Sanz, F., Pastor, M., Ciruela, F., Martinez-Seara, H., and Selent, J. (2016). Membrane omega-3 fatty acids modulate the oligomerisation kinetics of adenosine A2A and dopamine D2 receptors. Sci. Rep. *6*, 19839.
120. Lemmon, M.A., and Schlessinger, J. (2010). Cell signaling by receptor tyrosine kinases. Cell *141*, 1117-1134.
121. Luo, B.-H., Carman, C.V., and Springer, T.A. (2007). Structural basis of integrin regulation and signaling. Ann. Rev. Immunol. *25*, 619-647.
122. Endres, N.F., Das, R., Smith, A.W., Arkhipov, A., Kovacs, E., Huang, Y.J., Pelton, J.G., Shan, Y.B., Shaw, D.E., Wemmer, D.E., et al. (2013). Conformational coupling across the plasma membrane in activation of the EGF receptor. Cell *152*, 543-556.
123. Bocharov, E.V., Lesovoy, D.M., Goncharuk, S.A., Goncharuk, M.V., Hristova, K., and Arseniev, A.S. (2013). Structure of FGFR3 transmembrane domain dimer: implications for signaling and human pathologies. Structure *21*, 2087-2093.
124. Manni, S., Mineev, K.S., Usmanova, D., Lyukmanova, E.N., Shulepko, M.A., Kirpichnikov, M.P., Winter, J., Matkovic, M., Deupi, X., Arseniev, A.S., et al. (2014). Structural and functional characterization of alternative transmembrane domain conformations in VEGF receptor 2 activation. Structure *22*, 1077-1089.
125. Bragin, P., Mineev, K., Bocharova, O., Volynsky, P., Bocharov, E., and Arseniev, A. (2015). HER2 transmembrane domain dimerization coupled with self-association of membrane-embedded cytoplasmic juxtamembrane regions. J. Mol. Biol. *(in press)*, doi: 10.1016/j.jmb.2015.1011.1007.
126. Bocharov, E.V., Bragin, P.E., Pavlov, K.V., Bocharova, O.V., Mineev, K.S., Polyansky, A.A., Volynsky, P.E., Efremov, R.G., and Arseniev, A.S. (2017). The Conformation of the Epidermal Growth Factor Receptor Transmembrane Domain Dimer Dynamically Adapts to the Local Membrane Environment. Biochemistry *56*, 1697-1705.
127. Coskun, U., Grzybek, M., Drechsel, D., and Simons, K. (2011). Regulation of human EGF receptor by lipids. Proc. Natl. Acad. Sci. USA *108*, 9044–9048.
128. Arkhipov, A., Shan, Y., Das, R., Endres, N.F., Eastwood, M.P., Wemmer, D.E., Kuriyan, J., and Shaw, D.E. (2013). Architecture and membrane interactions of the EGF receptor. Cell *152*, 557-569.
129. Kaszuba, K., Grzybek, M., Orlowski, A., Danne, R., Rog, T., Simons, K., Coskun, U., and Vattulainen, I. (2015). N-Glycosylation as determinant of epidermal growth factor receptor conformation in membranes. Proc. Natl. Acad. Sci. USA *112*, 4334-4339.

130. Prakash, A., Janosi, L., and Doxastakis, M. (2010). Self-association of models of transmembrane domains of ErbB receptors in a lipid bilayer. Biophys. J. *99*, 3657-3665.
131. Halim, K.B.A., Koldsø, H., and Sansom, M.S.P. (2015). Interactions of the EGFR juxtamembrane domain with PIP-containing lipid bilayers: Insights from multiscale molecular dynamics simulations. Biochim. Biophys. Acta *1850*, 1017-1025.
132. Chavent, M., Seiradake, E., Jones, E.Y., and Sansom, Mark S.P. (2016). Structures of the EphA2 Receptor at the Membrane: Role of Lipid Interactions. Structure *24*, 337-347.
133. Matsushita, C., Tamagaki, H., Miyazawa, Y., Aimoto, S., Smith, S.O., and Sato, T. (2013). Transmembrane helix orientation influences membrane binding of the intracellular juxtamembrane domain in Neu receptor peptides. Proc. Natl. Acad. Sci. USA *110*, 1646-1651.
134. Wang, Y., Gao, J., Guo, X., Tong, T., Shi, X., Li, L., Qi, M., Wang, Y., Cai, M., Jiang, J., et al. (2014). Regulation of EGFR nanocluster formation by ionic protein-lipid interaction. Cell Res. *24*, 959-976.
135. Stansfeld, P.J., Jefferies, E.E., and Sansom, M.S.P. (2013). Multiscale simulations reveal conserved patterns of lipid interactions with aquaporins. Structure *21*, 810-819.
136. Gupta, K., Donlan, J.A.C., Hopper, J.T.S., Uzdavinys, P., Landreh, M., Struwe, W.B., Drew, D., Baldwin, A.J., Stansfeld, P.J., and Robinson, C.V. (2017). The role of interfacial lipids in stabilizing membrane protein oligomers. Nature *541*, 421-424.
137. Kalli, A.C., Sansom, M.S.P., and Reithmeier, R.A.F. (2015). Molecular Dynamics Simulations of the Bacterial UraA H+ Uracil Symporter in Lipid Bilayers Reveal a Closed State and a Selective Interaction with Cardiolipin. PLoS Comput. Biol. *11*, e1004123.
138. Lensink, M.F., Govaerts, C., and Ruysschaert, J.-M. (2010). Identification of specific lipid-binding sites in integral membrane proteins. J. Biol. Chem. *285*, 10519-10526.
139. Autzen, H.E., Siuda, I., Sonntag, Y., Nissen, P., Møller, J.V., and Thøgersen, L. (2015). Regulation of the Ca^{2+}-ATPase by cholesterol: a specific or non-specific effect? Molec. Memb. Biol. *32*, 75-87.
140. Khelashvili, G., Stanley, N., Sahai, M.A., Medina, J., LeVine, M.V., Shi, L., De Fabritiis, G., and Weinstein, H. (2015). Spontaneous Inward Opening of the Dopamine Transporter Is Triggered by PIP2-Regulated Dynamics of the N-Terminus. ACS Chem. Neurosci. *6*, 1825-1837.
141. Ferraro, M., Masetti, M., Recanatini, M., Cavalli, A., and Bottegoni, G. (2016). Mapping Cholesterol Interaction Sites on Serotonin Transporter through Coarse-Grained Molecular Dynamics. PLoS One *11*, e0166196.
142. Gournas, C., Papageorgiou, I., and Diallinas, G. (2008). The nucleobase-ascorbate transporter (NAT) family: genomics, evolution, structure-function relationships and physiological role. Molec. BioSys. *4*, 404-416.
143. Kosti, V., Lambrinidis, G., Myrianthopoulos, V., Diallinas, G., and Mikros, E. (2012). Identification of the Substrate Recognition and Transport Pathway in a Eukaryotic Member of the Nucleobase-Ascorbate Transporter (NAT) Family. PLoS One *7*, e41939.

144. Drachmann, N.D., Olesen, C., Møller, J.V., Guo, Z., Nissen, P., and Bublitz, M. (2014). Comparing crystal structures of Ca^{2+}-ATPase in the presence of different lipids. FEBS J. *281*, 4249-4262.
145. Jensen, A.M.L., Sørensen, T.L.M., Olesen, C., Møller, J.V., and Nissen, P. (2006). Modulatory and catalytic modes of ATP binding by the calcium pump. EMBO J. *25*, 2305-2314.
146. Autzen, H.E., Siuda, I., Sonntag, Y., Nissen, P., Møller, J.V., and Thøgersen, L. (2015). Regulation of the Ca2+-ATPase by cholesterol: A specific or non-specific effect? Mol. Membr. Biol. *32*, 75-87.
147. Chen, L., Zhang, Q., Qiu, Y., Li, Z., Chen, Z., Jiang, H., Li, Y., and Yang, H. (2015). Migration of PIP2 lipids on voltage-gated potassium channel surface influences channel deactivation. Sci. Rep. *5*, 15079.
148. Aponte-Santamaria, C., Briones, R., Schenk, A.D., Walz, T., and de Groot, B.L. (2012). Molecular driving forces defining lipid positions around aquaporin-0. Proc. Natl. Acad. Sci. USA *109*, 9887-9892.
149. Stansfeld, P.J., Goose, J.E., Caffrey, M., Carpenter, E.P., Parker, J.L., Newstead, N., and Sansom, M.S.P. (2015). MemProtMD: automated insertion of membrane protein structures into explicit lipid membranes. Structure *23*, 1350-1361.
150. Arnarez, C., Marrink, S.J., and Periole, X. (2013). Identification of cardiolipin binding sites on cytochrome c oxidase at the entrance of proton channels. Sci. Rep. *3*, 1263.
151. Pöyry, S., Cramariuc, O., Postila, P.A., Kaszuba, K., Sarewicz, M., Osyczka, A., Vattulainen, I., and Róg, T. (2013). Atomistic simulations indicate cardiolipin to have an integral role in the structure of the cytochrome bc1 complex. Biochim. Biophys. Acta *1827*, 769-778.
152. Weiser, B.P., Salari, R., Eckenhoff, R.G., and Brannigan, G. (2014). Computational investigation of cholesterol binding sites on mitochondrial VDAC. J. Phys. Chem. B *118*, 9852-9860.
153. Domański, J., Stansfeld, P.J., Sansom, M.S.P., and Beckstein, O. (2010). Lipidbook: A Public Repository for Force-Field Parameters Used in Membrane Simulations. The Journal of Membrane Biology *236*, 255-258.
154. Beyer, K., and Klingenberg, M. (1985). ADP/ATP carrier protein from beef heart mitochondria has high amounts of tightly bound cardiolipin, as revealed by phosphorus-31 nuclear magnetic resonance. Biochemistry *24*, 3821-3826.
155. Leach, A. (2001). Molecular Modelling: Principles and Applications (2nd Edition), (Prentice Hall).
156. Frenkel, D., and Smit, B. (2001). Understanding Molecular Simulation, (Academic Press, Inc.).
157. Alder, B.J., and Wainwright, T.E. (1957). Phase Transition for a Hard Sphere System. The Journal of Chemical Physics *27*, 1208-1209.
158. Alder, B.J., and Wainwright, T.E. (1959). Studies in Molecular Dynamics. I. General Method. The Journal of Chemical Physics *31*, 459-466.
159. McCammon, J.A., Gelin, B.R., and Karplus, M. (1977). Dynamics of folded proteins. Nature *267*, 585-590.
160. Zhou, H.-X. (2014). Theoretical frameworks for multiscale modeling and simulation. Curr. Opin. Struct. Biol. *25*, 67-76.
161. Lonsdale, R., Harvey, J.N., and Mulholland, A.J. (2012). A practical guide to modelling enzyme-catalysed reactions. Chem. Soc. Rev. *41*, 3025-3038.
162. Senn, H.M., and Thiel, W. (2007). QM/MM studies of enzymes. Curr. Opin. Chem. Biol. *11*, 182-187.

163. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., and Case, D.A. (2004). Development and testing of a general amber force field. J. Comput. Chem. *25*, 1157-1174.
164. MacKerell, A.D., Brooks, B., Brooks, C.L., Nilsson, L., Roux, B., Won, Y., and Karplus, M. (2002). CHARMM: The Energy Function and Its Parameterization. In Encyclopedia of Computational Chemistry. (John Wiley & Sons, Ltd).
165. Jorgensen, W.L., Maxwell, D.S., and Rives, T. (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.
166. Oostenbrink, C., Villa, A., Mark, A.E., and Van Gunsteren, W.F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. J. Comput. Chem. *25*, 1656-1676.
167. Levitt, M. (2001). The birth of computational structural biology. Nat. Struct. Mol. Biol. *8*, 392-393.
168. Mo, Y. (2010). A Critical Analysis on the Rotation Barriers in Butane. The Journal of Organic Chemistry *75*, 2733-2736.
169. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., and Berendsen, H.J.C. (2005). GROMACS: Fast, flexible, and free. J. Comput. Chem. *26*, 1701-1718.
170. Hockney, R.W., Goel, S.P., and Eastwood, J.W. (1974). Quiet high-resolution computer models of a plasma. J. Comput. Phys. *14*, 148-158.
171. Verlet, L. (1967). Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. Phys. Rev. *159*, 98-103.
172. Hess, B., Bekker, H., Berendsen, H.J.C., and Fraaije, J.G.E.M. (1997). LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem. *18*, 1463-1472.
173. Tieleman, D.P. (2010). Chapter 1 Methods and Parameters for Membrane Simulations. In Molecular Simulations and Biomembranes: From Biophysics to Function. (The Royal Society of Chemistry), pp. 1-25.
174. Vögele, M., and Hummer, G. (2016). Divergent Diffusion Coefficients in Simulations of Fluids and Lipid Membranes. The Journal of Physical Chemistry B *120*, 8722-8732.
175. Dolan, E.A., Venable, R.M., Pastor, R.W., and Brooks, B.R. (2002). Simulations of Membranes and Other Interfacial Systems Using P21 and Pc Periodic Boundary Conditions. Biophys. J. *82*, 2317-2325.
176. Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. The Journal of Chemical Physics *126*, 014101.
177. Parrinello, M., and Giaquinta, P.V. (1981). Analytical solution of a new integral equation for triplet correlations in hard sphere fluids. The Journal of Chemical Physics *74*, 1990-1997.
178. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., and Haak, J.R. (1984). Molecular dynamics with coupling to an external bath. The Journal of Chemical Physics *81*, 3684-3690.
179. Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J. Chem. Theory Comput. *4*, 435-447.

180. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G. (1995). A smooth particle mesh Ewald method. The Journal of Chemical Physics *103*, 8577-8593.
181. Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. The Journal of Chemical Physics *98*, 10089-10092.
182. Trick, J.L., Wallace, E.J., Bayley, H., and Sansom, M.S.P. (2014). Designing a Hydrophobic Barrier within Biomimetic Nanopores. ACS Nano *8*, 11268-11279.
183. Köpfer, D.A., Song, C., Gruene, T., Sheldrick, G.M., Zachariae, U., and de Groot, B.L. (2014). Ion permeation in K+ channels occurs by direct Coulomb knock-on. Science *346*, 352-355.
184. Koldsø, H., Shorthouse, D., Hélie, J., and Sansom, M.S.P. (2014). Lipid Clustering Correlates with Membrane Curvature as Revealed by Molecular Simulations of Complex Lipid Bilayers. PLoS Comput. Biol. *10*, e1003911.
185. Izvekov, S., and Voth, G.A. (2005). A Multiscale Coarse-Graining Method for Biomolecular Systems. The Journal of Physical Chemistry B *109*, 2469-2473.
186. Shih, A.Y., Arkhipov, A., Freddolino, P.L., and Schulten, K. (2006). Coarse Grained Protein−Lipid Model with Application to Lipoprotein Particles. The Journal of Physical Chemistry B *110*, 3674-3684.
187. Bond, P.J., and Sansom, M.S.P. (2006). Insertion and Assembly of Membrane Proteins via Simulation. J. Am. Chem. Soc. *128*, 2697-2704.
188. Marrink, S.J., de Vries, A.H., and Mark, A.E. (2004). Coarse Grained Model for Semiquantitative Lipid Simulations. The Journal of Physical Chemistry B *108*, 750-760.
189. Marrink, S.J., Risselada, H.J., Yefimov, S., Tieleman, D.P., and de Vries, A.H. (2007). The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. The Journal of Physical Chemistry B *111*, 7812-7824.
190. Srivastava, A., and Voth, G.A. (2013). Hybrid Approach for Highly Coarse-Grained Lipid Bilayer Models. J. Chem. Theory Comput. *9*, 750-765.
191. Saunders, M.G., and Voth, G.A. (2012). Coarse-graining of multiprotein assemblies. Curr. Opin. Struct. Biol. *22*, 144-150.
192. Wassenaar, T.A., Ingólfsson, H.I., Prieß, M., Marrink, S.J., and Schäfer, L.V. (2013). Mixing MARTINI: Electrostatic Coupling in Hybrid Atomistic–Coarse-Grained Biomolecular Simulations. The Journal of Physical Chemistry B *117*, 3516-3530.
193. Zavadlav, J., Melo, M.N., Cunha, A.V., de Vries, A.H., Marrink, S.J., and Praprotnik, M. (2014). Adaptive Resolution Simulation of MARTINI Solvents. J. Chem. Theory Comput. *10*, 2591-2598.
194. Monticelli, L., Kandasamy, S.K., Periole, X., Larson, R.G., Tieleman, D.P., and Marrink, S.-J. (2008). The MARTINI Coarse-Grained Force Field: Extension to Proteins. J. Chem. Theory Comput. *4*, 819-834.
195. López, C.A., Rzepiela, A.J., de Vries, A.H., Dijkhuizen, L., Hünenberger, P.H., and Marrink, S.J. (2009). Martini Coarse-Grained Force Field: Extension to Carbohydrates. J. Chem. Theory Comput. *5*, 3195-3210.
196. Uusitalo, J.J., Ingólfsson, H.I., Akhshi, P., Tieleman, D.P., and Marrink, S.J. (2015). Martini Coarse-Grained Force Field: Extension to DNA. J. Chem. Theory Comput. *11*, 3932-3945.
197. Uusitalo, J.J., Ingólfsson, H.I., Marrink, S.J., and Faustino, I. Martini Coarse-Grained Force Field: Extension to RNA. Biophys. J.

198. Lee, H., de Vries, A.H., Marrink, S.-J., and Pastor, R.W. (2009). A Coarse-Grained Model for Polyethylene Oxide and Polyethylene Glycol: Conformation and Hydrodynamics. The Journal of Physical Chemistry B *113*, 13186-13194.
199. Reddy, T., Shorthouse, D., Parton, Daniel L., Jefferys, E., Fowler, Philip W., Chavent, M., Baaden, M., and Sansom, Mark S.P. (2015). Nothing to Sneeze At: A Dynamic and Integrative Computational Model of an Influenza A Virion. Structure *23*, 584-597.
200. Marzinek, Jan K., Holdbrook, Daniel A., Huber, Roland G., Verma, C., and Bond, Peter J. (2016). Pushing the Envelope: Dengue Viral Membrane Coaxed into Shape by Molecular Simulations. Structure *24*, 1410-1420.
201. Risselada, H.J., and Marrink, S.J. (2009). Curvature effects on lipid packing and dynamics in liposomes revealed by coarse grained molecular dynamics simulations. PCCP *11*, 2056-2067.
202. Lelimousin, M., and Sansom, M.S.P. (2013). Membrane Perturbation by Carbon Nanotube Insertion: Pathways to Internalization. Small *9*, 3639-3646.
203. de Jong, D.H., Singh, G., Bennett, W.F.D., Arnarez, C., Wassenaar, T.A., Schäfer, L.V., Periole, X., Tieleman, D.P., and Marrink, S.J. (2013). Improved Parameters for the Martini Coarse-Grained Protein Force Field. J. Chem. Theory Comput. *9*, 687-697.
204. Marrink, S.J., and Tieleman, D.P. (2013). Perspective on the Martini model. Chem. Soc. Rev. *42*, 6801-6822.
205. Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers *22*, 2577-2637.
206. Periole, X., Cavalli, M., Marrink, S.-J., and Ceruso, M.A. (2009). Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. J. Chem. Theory Comput. *5*, 2531-2543.
207. Yesylevskyy, S.O., Schäfer, L.V., Sengupta, D., and Marrink, S.J. (2010). Polarizable Water Model for the Coarse-Grained MARTINI Force Field. PLoS Comp. Biol. *6*, e1000810.
208. Gu, R.-X., Ingólfsson, H.I., de Vries, A.H., Marrink, S.J., and Tieleman, D.P. (2017). Ganglioside-Lipid and Ganglioside-Protein Interactions Revealed by Coarse-Grained and Atomistic Molecular Dynamics Simulations. The Journal of Physical Chemistry B *121*, 3262-3275.
209. Arnarez, C., Uusitalo, J.J., Masman, M.F., Ingólfsson, H.I., de Jong, D.H., Melo, M.N., Periole, X., de Vries, A.H., and Marrink, S.J. (2015). Dry Martini, a Coarse-Grained Force Field for Lipid Membrane Simulations with Implicit Solvent. J. Chem. Theory Comput. *11*, 260-275.
210. Herzog, F.A., Braun, L., Schoen, I., and Vogel, V. (2016). Improved Side Chain Dynamics in MARTINI Simulations of Protein–Lipid Interfaces. J. Chem. Theory Comput. *12*, 2446-2458.
211. Stansfeld, P.J., and Sansom, M.S.P. (2011). From Coarse Grained to Atomistic: A Serial Multiscale Approach to Membrane Protein Simulations. J. Chem. Theory Comput. *7*, 1157-1166.
212. Wassenaar, T.A., Pluhackova, K., Böckmann, R.A., Marrink, S.J., and Tieleman, D.P. (2014). Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. J. Chem. Theory Comput. *10*, 676-690.

213. Scott, K.A., Bond, P.J., Ivetac, A., Chetwynd, A.P., Khalid, S., and Sansom, M.S.P. (2008). Coarse-Grained MD Simulations of Membrane Protein-Bilayer Self-Assembly. Structure *16*, 621-630.
214. Nugent, T., and Jones, D.T. (2013). Membrane protein orientation and refinement using a knowledge-based statistical potential. BMC Bioinformatics *14*, 276.
215. Jo, S., Lim, J.B., Klauda, J.B., and Im, W. (2009). CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. Biophys. J. *97*, 50-58.
216. Domański, J., Hedger, G., Best, R.B., Stansfeld, P.J., and Sansom, M.S.P. (2017). Convergence and Sampling in Determining Free Energy Landscapes for Membrane Protein Association. The Journal of Physical Chemistry B *121*, 3364-3375.
217. Maingi, V., Burns, J.R., Uusitalo, J.J., Howorka, S., Marrink, S.J., and Sansom, M.S.P. (2017). Stability and dynamics of membrane-spanning DNA nanopores. Nature Communications *8*, 14784.
218. Sengupta, D., and Marrink, S.J. (2010). Lipid-mediated interactions tune the association of glycophorin A helix and its disruptive mutants in membranes. PCCP *12*, 12987-12996.
219. Shamsudin Khan, Y., Kazemi, M., Gutiérrez-de-Terán, H., and Åqvist, J. (2015). Origin of the Enigmatic Stepwise Tight-Binding Inhibition of Cyclooxygenase-1. Biochemistry *54*, 7283-7291.
220. Grubmüller, H., Heymann, B., and Tavan, P. (1996). Ligand Binding: Molecular Mechanics Calculation of the Streptavidin-Biotin Rupture Force. Science *271*, 997-999.
221. Roux, B. (1995). The calculation of the potential of mean force using computer simulations. Comput. Phys. Commun. *91*, 275-282.
222. Park, S., and Schulten, K. (2004). Calculating potentials of mean force from steered molecular dynamics simulations. The Journal of Chemical Physics *120*, 5946-5961.
223. Hénin, J., and Chipot, C. (2004). Overcoming free energy barriers using unconstrained molecular dynamics simulations. The Journal of Chemical Physics *121*, 2904-2914.
224. Kumar, S., Rosenberg, J.M., Bouzida, D., Swendsen, R.H., and Kollman, P.A. (1992). THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J. Comput. Chem. *13*, 1011-1021.
225. Grossfield, A., Feller, S.E., and Pitman, M.C. (2006). A role for direct interactions in the modulation of rhodopsin by ω-3 polyunsaturated lipids. Proc. Natl. Acad. Sci. U.S.A. *103*, 4888-4893.
226. Hub, J.S., de Groot, B.L., and van der Spoel, D. (2010). g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. J. Chem. Theory Comput. *6*, 3713-3720.
227. Ostmeyer, J., Chakrapani, S., Pan, A.C., Perozo, E., and Roux, B. (2013). Recovery from slow inactivation in K+ channels is controlled by water molecules. Nature *501*, 121-124.
228. Chavent, M., Chetwynd, A.P., Stansfeld, P.J., and Sansom, M.S.P. (2014). Dimerization of the EphA1 Receptor Tyrosine Kinase Transmembrane Domain: Insights into the Mechanism of Receptor Activation. Biochemistry *53*, 6641-6652.
229. Dunton, T.A., Goose, J.E., Gavaghan, D.J., Sansom, M.S.P., and Osborne, J.M. (2014). The Free Energy Landscape of Dimerization of a Membrane Protein, NanC. PLoS Comp. Biol. *10*, e1003417.

230. Hanson, Sonya M., Newstead, S., Swartz, Kenton J., and Sansom, Mark S.P. (2015). Capsaicin Interaction with TRPV1 Channels in a Lipid Bilayer: Molecular Dynamics Simulation. Biophys. J. *108*, 1425-1434.
231. Bennett, W.F.D., MacCallum, J.L., Hinner, M.J., Marrink, S.J., and Tieleman, D.P. (2009). Molecular View of Cholesterol Flip-Flop and Chemical Potential in Different Membrane Environments. J. Am. Chem. Soc. *131*, 12714-12720.
232. Lemkul, J.A., and Bevan, D.R. (2010). Assessing the Stability of Alzheimer's Amyloid Protofibrils Using Molecular Dynamics. The Journal of Physical Chemistry B *114*, 1652-1660.
233. Rashid, M.H., Heinzelmann, G., Huq, R., Tajhya, R.B., Chang, S.C., Chhabra, S., Pennington, M.W., Beeton, C., Norton, R.S., and Kuyucak, S. (2013). A Potent and Selective Peptide Blocker of the Kv1.3 Channel: Prediction from Free-Energy Simulations and Experimental Confirmation. PLoS One *8*, e78712.
234. Alhadeff, R., Assa, D., Astrahan, P., Krugliak, M., and Arkin, I.T. (2014). Computational and experimental analysis of drug binding to the Influenza M2 channel. Biochim. Biophys. Acta *1838*, 1068-1073.
235. Huang, X., Zhao, X., Zheng, F., and Zhan, C.-G. (2012). Cocaine Esterase–Cocaine Binding Process and the Free Energy Profiles by Molecular Dynamics and Potential of Mean Force Simulations. The Journal of Physical Chemistry B *116*, 3361-3368.
236. Bessman, N.J., Freed, D.M., and Lemmon, M.A. (2014). Putting together structures of epidermal growth factor receptors. Curr. Opin. Struct. Biol. *29*, 95-101.
237. Gschwind, A., Fischer, O.M., and Ullrich, A. (2004). The discovery of receptor tyrosine kinases: targets for cancer therapy. Nat. Rev. Cancer *4*, 361-370.
238. Coskun, Ü., Grzybek, M., Drechsel, D., and Simons, K. (2011). Regulation of human EGF receptor by lipids. Proceedings of the National Academy of Sciences *108*, 9044-9048.
239. Michailidis, I., Rusinova, R., Georgakopoulos, A., Chen, Y., Iyengar, R., Robakis, N., Logothetis, D., and Baki, L. (2011). Phosphatidylinositol-4,5-bisphosphate regulates epidermal growth factor receptor activation. Pflügers Archiv - European Journal of Physiology *461*, 387-397.
240. Pike, L.J., and Casey, L. (2002). Cholesterol Levels Modulate EGF Receptor-Mediated Signaling by Altering Receptor Function and Trafficking. Biochemistry *41*, 10315-10322.
241. Arnarez, C., Marrink, S.J., and Periole, X. (2013). Identification of cardiolipin binding sites on cytochrome c oxidase at the entrance of proton channels. Sci. Rep. *3*, 1263.
242. Stansfeld, Phillip J., Jefferys, Elizabeth E., and Sansom, Mark S.P. (2013). Multiscale Simulations Reveal Conserved Patterns of Lipid Interactions with Aquaporins. Structure *21*, 810-819.
243. Tieleman, D.P., and Marrink, S.-J. (2006). Lipids Out of Equilibrium: Energetics of Desorption and Pore Mediated Flip-Flop. J. Am. Chem. Soc. *128*, 12462-12467.
244. Corry, B., and Thomas, M. (2012). Mechanism of Ion Permeation and Selectivity in a Voltage Gated Sodium Channel. J. Am. Chem. Soc. *134*, 1840-1846.
245. Gumbart, J.C., Roux, B., and Chipot, C. (2013). Standard Binding Free Energies from Computer Simulations: What Is the Best Strategy? J. Chem. Theory Comput. *9*, 794-802.

246. Torrie, G.M., and Valleau, J.P. (1974). Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. Chem. Phys. Lett. *28*, 578-581.
247. Endres, Nicholas F., Das, R., Smith, Adam W., Arkhipov, A., Kovacs, E., Huang, Y., Pelton, Jeffrey G., Shan, Y., Shaw, David E., Wemmer, David E., et al. (2013). Conformational Coupling across the Plasma Membrane in Activation of the EGF Receptor. Cell *152*, 543-556.
248. van Meer, G., and de Kroon, A.I.P.M. (2011). Lipid map of the mammalian cell. J. Cell Sci. *124*, 5-8.
249. Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J. Mol. Graph. *14*, 33-38, 27-38.
250. Arnarez, C., Marrink, S.J., and Periole, X. (2016). Molecular mechanism of cardiolipin-mediated assembly of respiratory chain supercomplexes. Chemical Science.
251. Kooijman, E.E., King, K.E., Gangoda, M., and Gericke, A. (2009). Ionization Properties of Phosphatidylinositol Polyphosphates in Mixed Model Membranes. Biochemistry *48*, 9360-9371.
252. McLaughlin, S., Wang, J., Gambhir, A., and Murray, D. (2002). PIP2 AND PROTEINS: Interactions, Organization, and Information Flow. Annu. Rev. Biophys. Biomol. Struct. *31*, 151-175.
253. Suh, B., and Hille, B. (2008). PIP2 Is a Necessary Cofactor for Ion Channel Function: How and Why? Annual Review of Biophysics *37*, 175-195.
254. Red Brewer, M., Choi, S.H., Alvarado, D., Moravcevic, K., Pozzi, A., Lemmon, M.A., and Carpenter, G. (2009). The Juxtamembrane Region of the EGF Receptor Functions as an Activation Domain. Mol. Cell *34*, 641-651.
255. Koldsø, H., and Sansom, M.S.P. (2012). Local Lipid Reorganization by a Transmembrane Protein Domain. The Journal of Physical Chemistry Letters *3*, 3498-3502.
256. Kaszuba, K., Grzybek, M., Orłowski, A., Danne, R., Róg, T., Simons, K., Coskun, Ü., and Vattulainen, I. (2015). N-Glycosylation as determinant of epidermal growth factor receptor conformation in membranes. Proceedings of the National Academy of Sciences *112*, 4334-4339.
257. Sarabipour, S., and Hristova, K. (2013). FGFR3 Transmembrane Domain Interactions Persist in the Presence of Its Extracellular Domain. Biophys. J. *105*, 165-171.
258. Lee, M.S., Salsbury, F.R., and Brooks, C.L. (2004). Constant-pH molecular dynamics using continuous titration coordinates. Proteins: Structure, Function, and Bioinformatics *56*, 738-752.
259. Naughton, F.B., Kalli, A.C., and Sansom, M.S.P. (2016). Association of Peripheral Membrane Proteins with Membranes: Free Energy of Binding of GRP1 PH Domain with Phosphatidylinositol Phosphate-Containing Model Bilayers. The Journal of Physical Chemistry Letters *7*, 1219-1224.
260. Sapay, N., Bennett, W.F.D., and Tieleman, D.P. (2009). Thermodynamics of flip-flop and desorption for a systematic series of phosphatidylcholine lipids. Soft Matter *5*, 3295-3302.
261. Janosi, L., Prakash, A., and Doxastakis, M. (2010). Lipid-Modulated Sequence-Specific Association of Glycophorin A in Membranes. Biophys. J. *99*, 284-292.
262. Kuznetsov, A.S., Polyansky, A.A., Fleck, M., Volynsky, P.E., and Efremov, R.G. (2015). Adaptable Lipid Matrix Promotes Protein–Protein Association in Membranes. J. Chem. Theory Comput. *11*, 4415-4426.

263. Hénin, J., Pohorille, A., and Chipot, C. (2005). Insights into the Recognition and Association of Transmembrane α-Helices. The Free Energy of α-Helix Dimerization in Glycophorin A. J. Am. Chem. Soc. *127*, 8478-8484.
264. Abd Halim, K.B., Koldsø, H., and Sansom, M.S.P. (2015). Interactions of the EGFR juxtamembrane domain with PIP2-containing lipid bilayers: Insights from multiscale molecular dynamics simulations. Biochimica et Biophysica Acta (BBA) - General Subjects *1850*, 1017-1025.
265. Arkhipov, A., Shan, Y., Das, R., Endres, Nicholas F., Eastwood, Michael P., Wemmer, David E., Kuriyan, J., and Shaw, David E. (2013). Architecture and Membrane Interactions of the EGF Receptor. Cell *152*, 557-569.
266. Wang, Y., Gao, J., Guo, X., Tong, T., Shi, X., Li, L., Qi, M., Wang, Y., Cai, M., Jiang, J., et al. (2014). Regulation of EGFR nanocluster formation by ionic protein-lipid interaction. Cell Res. *24*, 959-976.
267. Prakash, A., Janosi, L., and Doxastakis, M. (2010). Self-Association of Models of Transmembrane Domains of ErbB Receptors in a Lipid Bilayer. Biophys. J. *99*, 3657-3665.
268. Lingwood, D., and Simons, K. (2010). Lipid Rafts As a Membrane-Organizing Principle. Science *327*, 46-50.
269. Stangl, M., and Schneider, D. (2015). Functional competition within a membrane: Lipid recognition vs. transmembrane helix oligomerization. Biochim. Biophys. Acta *1848*, 1886-1896.
270. Bocharov, E.V., Lesovoy, D.M., Pavlov, K.V., Pustovalova, Y.E., Bocharova, O.V., and Arseniev, A.S. (2016). Alternative packing of EGFR transmembrane domain suggests that protein–lipid interactions underlie signal conduction across membrane. Biochim. Biophys. Acta *1858*, 1254-1261.
271. Coban, O., Zanetti-Dominguez, Laura C., Matthews, Daniel R., Rolfe, Daniel J., Weitsman, G., Barber, Paul R., Barbeau, J., Devauges, V., Kampmeier, F., Winn, M., et al. (2015). Effect of Phosphorylation on EGFR Dimer Stability Probed by Single-Molecule Dynamics and FRET/FLIM. Biophys. J. *108*, 1013-1026.
272. Palsdottir, H., and Hunte, C. (2004). Lipids in membrane protein structures. Biochim. Biophys. Acta *1666*, 2-18.
273. Stansfeld, Phillip J., Goose, Joseph E., Caffrey, M., Carpenter, Elisabeth P., Parker, Joanne L., Newstead, S., and Sansom, Mark S.P. (2015). MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes. Structure *23*, 1350-1361.
274. Klingenberg, M. (2008). The ADP and ATP transport in mitochondria and its carrier. Biochim. Biophys. Acta *1778*, 1978-2021.
275. Kunji, E.R.S., Aleksandrova, A., King, M.S., Majd, H., Ashton, V.L., Cerson, E., Springett, R., Kibalchenko, M., Tavoulari, S., Crichton, P.G., et al. (2016). The transport mechanism of the mitochondrial ADP/ATP carrier. Biochim. Biophys. Acta *1863*, 2379-2393.
276. Palmieri, F. (2014). Mitochondrial transporters of the SLC25 family and associated diseases: a review. J. Inher. Metab. Dis. *37*, 565-575.
277. Klingenberg, M. (2009). Cardiolipin and mitochondrial carriers. Biochim. Biophys. Acta *1788*, 2048-2058.
278. Mayr, J A., Haack, T B., Graf, E., Zimmermann, F A., Wieland, T., Haberberger, B., Superti-Furga, A., Kirschner, J., Steinmann, B., Baumgartner, M R., et al. (2012). Lack of the mitochondrial protein acylglycerol kinase causes Sengers syndrome. Amer. J. Hum. Genet. *90*, 314-320.

279. Haghighi, A., Haack, T.B., Atiq, M., Mottaghi, H., Haghighi-Kakhki, H., Bashir, R.A., Ahting, U., Feichtinger, R.G., Mayr, J.A., Rötig, A., et al. (2014). Sengers syndrome: six novel AGK mutations in seven new families and review of the phenotypic and mutational spectrum of 29 patients. Orphanet J. Rare Dis. *9*, 1-12.
280. Nury, H., Dahout-Gonzalez, C., Trezeguet, V., Lauquin, C.J.M., Brandolin, G., and Pebay-Peyroula, E. (2005). Structural basis for lipid-mediated interactions between mitochondrial ADP/ATP carrier monomers. FEBS Lett. *579*, 6031-6036.
281. Berardi, M.J., Shih, W.M., Harrison, S.C., and Chou, J.J. (2011). Mitochondrial uncoupling protein 2 structure determined by NMR molecular fragment searching. Nature *476*, 109-113.
282. Ruprecht, J.J., Hellawell, A.M., Harding, M., Crichton, P.G., McCoy, A.J., and Kunji, E.R.S. (2014). Structures of yeast mitochondrial ADP/ATP carriers support a domain-based alternating-access transport mechanism. Proc. Natl. Acad. Sci. USA *111*, E426-E434.
283. Dahlberg, M., and Maliniak, A. (2010). Mechanical properties of coarse-grained bilayers Formed by cardiolipin and zwitterionic lipids. J. Chem. Theor Comp. *6*, 1638-1649.
284. Olofsson, G., and Sparr, E. (2013). Ionization Constants pK_a of cardiolipin. PLoS ONE *8*, e73040.
285. Hedger, G., Koldsø, H., and Sansom, M.S.P. (2016). Free energy landscape of lipid interactions with regulatory binding sites on the transmembrane domain of the EGF receptor. J. Phys. Chem. B *120*, 8154–8163.
286. Rassam, P., Copeland, N.A., Birkholz, O., Toth, C., Chavent, M., Duncan, A.L., Cross, S.J., Housden, N.G., Kaminska, R., Seger, U., et al. (2015). Supramolecular assemblies underpin turnover of outer membrane proteins in bacteria. Nature *523*, 333-336.
287. Michaud-Agrawal, N., Denning, E.J., Woolf, T.B., and Beckstein, O. (2011). MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. J. Comput. Chem. *32*, 2319-2327.
288. Colbeau, A., Nachbaur, J., and Vignais, P.M. (1971). Enzymic characterization and lipid composition of rat liver subcellular membranes. Biochim. Biophys. Acta *249*, 462-492.
289. Zinser, E., and Daum, G. (1995). Isolation and biochemical-characterization of organelles from the yeast, *Saccharomyces cerevisiae*. Yeast *11*, 493-536.
290. Baile, M.G., Sathappa, M., Lu, Y.W., Pryce, E., Whited, K., McCaffery, J.M., Han, X.L., Alder, N.N., and Claypool, S.M. (2014). Unremodeled and remodeled cardiolipin are functionally indistinguishable in yeast. J. Biol. Chem. *289*, 1768-1778.
291. Arnarez, C., Marrink, S.J., and Periole, X. (2016). Molecular mechanism of cardiolipin-mediated assembly of respiratory chain supercomplexes. Chemical Science *7*, 4435-4443.
292. Roux, B. (1995). The calculation of the potential of mean force using computer simulations. Comp. Physics Comm. *91*, 275-282.
293. Arnarez, C., Marrink, S.J., and Periole, X. (2016). Molecular mechanism of cardiolipin-mediated assembly of respiratory chain supercomplexes. Chem. Sci. *DOI: 10.1039/C5SC04664E*
294. Hub, J.S., de Groot, B.L., and van der Spoel, D. (2010). g_wham A free weighted histogram analysis implementation Including robust error and autocorrelation estimates. J. Chem. Theor. Comput. *6*, 3713-3720.

295. Kunji, E.R.S., and Robinson, A.J. (2006). The conserved substrate binding site of mitochondrial carriers. Biochimica et Biophysica Acta - Bioenergetics *1757*, 1237-1248.
296. Ingólfsson, H.I., Arnarez, C., Periole, X., and Marrink, S.J. (2016). Computational 'microscopy' of cellular membranes. J. Cell Sci. *129*, 257.
297. Osman, C., Voelker, D.R., and Langer, T. (2011). Making heads or tails of phospholipids in mitochondria. The Journal of Cell Biology *192*, 7-16.
298. Zhao, L., Wang, S., Run, C., OuYang, B., and Chou, J.J. (2016). Specific Lipid Binding of Membrane Proteins in Detergent Micelles Characterized by NMR and Molecular Dynamics. Biochemistry *55*, 5317-5320.
299. Naughton, F.B., Kalli, A.C., and Sansom, M.S.P. (2016). Association of peripheral membrane proteins with membranes: Free energy of binding of GRP1 PH domain with PIP-containing model bilayers. J. Phys. Chem. Lett. *7*, 1219-1224.
300. Lee, Y., Willers, C., Kunji, E.R.S., and Crichton, P.G. (2015). Uncoupling protein 1 binds one nucleotide per monomer and is stabilized by tightly bound cardiolipin. Proc. Natl. Acad. Sci. USA *112*, 6973-6978.
301. Hedger, G., and Sansom, M.S.P. (2016). Lipid interaction sites on channels, transporters and receptors: recent insights from molecular dynamics simulations. Biochim. Biophys. Acta *1858*, 2390–2400.
302. Malyshka, D., Pandiscia, L.A., and Schweitzer-Stenner, R. (2014). Cardiolipin containing liposomes are fully ionized at physiological pH. An FT-IR study of phosphate group ionization. Vibrat. Spectro. *75*, 86-92.
303. Zhang, M., Mileykovskaya, E., and Dowhan, W. (2002). Gluing the respiratory chain together - cardiolipin is required for supercomplex formation in the inner mitochondrial membrane. J. Biol. Chem. *277*, 43553-43556.
304. Hackenberg, H., and Klingenberg, M. (1980). Molecular-weight and hydrodynamic parameters of the adenosine 5'-diphosphate-adenosine 5'-triphosphate carrier in Triton X-100. Biochem. *19*, 548-555.
305. Kunji, E.R.S., and Crichton, P.G. (2010). Mitochondrial carriers function as monomers. Biochim. Biophys. Acta *1797*, 817-831.
306. Lewis, R., Zweytick, D., Pabst, G., Lohner, K., and McElhaney, R.N. (2007). Calorimetric, X-ray diffraction, and spectroscopic studies of the thermotropic phase behavior and organization of tetramyristoyl cardiolipin membranes. Biophys. J. *92*, 3166-3177.
307. Frias, M., Benesch, M.G.K., Lewis, R., and McElhaney, R.N. (2011). On the miscibility of cardiolipin with 1,2-diacyl phosphoglycerides: Binary mixtures of dimyristoylphosphatidylethanolamine and tetramyristoylcardiolipin. Biochim. Biophysic. Acta *1808*, 774-783.
308. Zeczycki, T.N., Whelan, J., Hayden, W.T., Brown, D.A., and Shaikh, S.R. (2014). Increasing levels of cardiolipin differentially influence packing of phospholipids found in the mitochondrial inner membrane. Biochem. Biophys. Res. Comm, *450*, 366-371.
309. Phan, M.D., and Shin, K. (2015). Effects of cardiolipin on membrane morphology: a Langmuir monolayer study. Biophys. J. *108*, 1977-1986.
310. Kaurola, P., Sharma, V., Vonk, A., Vattulainen, I., and Rog, T. (2016). Distribution and dynamics of quinones in the lipid bilayer mimicking the inner membrane of mitochondria. Biochim. Biophys. Acta *1858*, 2116-2122.
311. Kühlbrandt, W. (2015). Structure and function of mitochondrial membrane protein complexes. BMC Biol. *13*, 1-11.

312. Pietropaolo, A., Pierri, C.L., Palmieri, F., and Klingenberg, M. (2016). The switching mechanism of the mitochondrial ADP/ATP carrier explored by free-energy landscapes. Biochim. Biophys. Acta *1857*, 772-781.
313. Lee, M.S., Salsbury, F.R., and Brooks, C.L. (2004). Constant-pH molecular dynamics using continuous titration coordinates. Proteins: Struct. Func. Bioinf. *56*, 738-752.
314. Stark, A.C., Andrews, C.T., and Elcock, A.H. (2013). Toward optimized potential functions for protein-protein interactions in aqueous solutions: osmotic second virial coefficient calculations using the MARTINI coarse-grained force field. J. Chem. Theor. Comp. *9*, 4176-4185.
315. Sengupta, D., and Marrink, S.J. (2010). Lipid-mediated interactions tune the association of glycophorin A helix and its disruptive mutants in membranes. Phys. Chem. Chem. Phys. *12*, 12987-12996.
316. Janosi, L., Prakash, A., and Doxastakis, M. (2010). Lipid-modulated sequence-specific association of glycophorin A in membranes. Biophys. J. *99*, 284-292.
317. Kuznetsov, A.S., Polyansky, A.A., Fleck, M., Volynsky, P.E., and Efremov, R.G. (2015). Adaptable lipid matrix promotes protein-protein association in membranes. J. Chem. Theor. Comput. *11*, 4415-4426.
318. Hénin, J., Pohorille, A., and Chipot, C. (2005). Insights into the recognition and association of transmembrane a-helices. The free energy of a-helix dimerization in glycophorin A. J. Amer. Chem. Soc. *127*, 8478-8484.
319. Hong, H., and Bowie, J.U. (2011). Dramatic destabilization of transmembrane helix interactions by features of natural membrane environments. J. Amer. Chem. Soc. *133*, 11389-11398.
320. Periole, X., Cavalli, M., Marrink, S.J., and Ceruso, M.A. (2009). Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. J. Chem. Theory Comput. *5*, 2531-2543.
321. Pietropaolo, A., Pierri, C.L., Palmieri, F., and Klingenberg, M. (2016). The switching mechanism of the mitochondrial ADP/ATP carrier explored by free-energy landscapes. Biochim. Biophys. Acta *1857*, 772-781.
322. Di Paolo, G., and De Camilli, P. (2006). Phosphoinositides in cell regulation and membrane dynamics. Nature *443*, 651-657.
323. Whorton, Matthew R., and MacKinnon, R. (2011). Crystal Structure of the Mammalian GIRK2 K+ Channel and Gating Regulation by G Proteins, PIP2, and Sodium. Cell *147*, 199-208.
324. Gao, Y., Cao, E., Julius, D., and Cheng, Y. (2016). TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action. Nature *534*, 347-351.
325. van den Bogaart, G., Meyenberg, K., Risselada, H.J., Amin, H., Willig, K.I., Hubrich, B.E., Dier, M., Hell, S.W., Grubmuller, H., Diederichsen, U., et al. (2011). Membrane protein sequestering by ionic protein-lipid interactions. Nature *479*, 552-555.
326. Vincent, J.-P., Mazella, J., and Kitabgi, P. (1999). Neurotensin and neurotensin receptors. Trends Pharmacol. Sci. *20*, 302-309.
327. Leeman, S.E., and Carraway, R.E. (1982). NEUROTENSIN: DISCOVERY, ISOLATION, CHARACTERIZATION, SYNTHESIS AND POSSIBLE PHYSIOLOGICAL ROLES*. Ann. N. Y. Acad. Sci. *400*, 1-16.
328. Bolivar, J.H., Muñoz-García, J.C., Castro-Dopico, T., Dijkman, P.M., Stansfeld, P.J., and Watts, A. (2016). Interaction of lipids with the neurotensin receptor 1. Biochim. Biophys. Acta *1858*, 1278-1287.

329. Dijkman, P.M., and Watts, A. (2015). Lipid modulation of early G protein-coupled receptor signalling events. Biochim. Biophys. Acta *1848*, 2889-2897.
330. Inagaki, S., Ghirlando, R., White, J.F., Gvozdenovic-Jeremic, J., Northup, J.K., and Grisshammer, R. (2012). Modulation of the Interaction between Neurotensin Receptor NTS1 and Gq Protein by Lipid. J. Mol. Biol. *417*, 95-111.
331. Oates, J., Faust, B., Attrill, H., Harding, P., Orwick, M., and Watts, A. (2012). The role of cholesterol on the activity and stability of neurotensin receptor 1. Biochim. Biophys. Acta *1818*, 2228-2233.
332. Liu, P., Cheng, H., Roberts, T.M., and Zhao, J.J. (2009). Targeting the phosphoinositide 3-kinase pathway in cancer. Nat. Rev. Drug Discov. *8*, 627-644.
333. Egloff, P., Hillenbrand, M., Klenk, C., Batyuk, A., Heine, P., Balada, S., Schlinkmann, K.M., Scott, D.J., Schütz, M., and Plückthun, A. (2014). Structure of signaling-competent neurotensin receptor 1 obtained by directed evolution in Escherichia coli. Proceedings of the National Academy of Sciences *111*, E655-E662.
334. White, J.F., Noinaj, N., Shibata, Y., Love, J., Kloss, B., Xu, F., Gvozdenovic-Jeremic, J., Shah, P., Shiloach, J., Tate, C.G., et al. (2012). Structure of the agonist-bound neurotensin receptor. Nature *490*, 508-513.
335. Rasmussen, S.G.F., Choi, H.-J., Fung, J.J., Pardon, E., Casarosa, P., Chae, P.S., DeVree, B.T., Rosenbaum, D.M., Thian, F.S., Kobilka, T.S., et al. (2011). Structure of a nanobody-stabilized active state of the [bgr]2 adrenoceptor. Nature *469*, 175-180.
336. Šali, A., and Blundell, T.L. (1993). Comparative Protein Modelling by Satisfaction of Spatial Restraints. J. Mol. Biol. *234*, 779-815.
337. Warne, T., Moukhametzianov, R., Baker, J.G., Nehme, R., Edwards, P.C., Leslie, A.G.W., Schertler, G.F.X., and Tate, C.G. (2011). The structural basis for agonist and partial agonist action on a [bgr]1-adrenergic receptor. Nature *469*, 241-244.
338. Tironi, I.G., Sperb, R., Smith, P.E., and Gunsteren, W.F.v. (1995). A generalized reaction field method for molecular dynamics simulations. The Journal of Chemical Physics *102*, 5451-5459.
339. Zocher, M., Zhang, C., Rasmussen, S.G.F., Kobilka, B.K., and Müller, D.J. (2012). Cholesterol increases kinetic, energetic, and mechanical stability of the human β2-adrenergic receptor. Proceedings of the National Academy of Sciences *109*, E3463–E3472.
340. Harding, P.J., Attrill, H., Boehringer, J., Ross, S., Wadhams, G.H., Smith, E., Armitage, J.P., and Watts, A. (2009). Constitutive Dimerization of the G-Protein Coupled Receptor, Neurotensin Receptor 1, Reconstituted into Phospholipid Bilayers. Biophys. J. *96*, 964-973.
341. Periole, X., Knepp, A.M., Sakmar, T.P., Marrink, S.J., and Huber, T. (2012). Structural Determinants of the Supramolecular Organization of G Protein-Coupled Receptors in Bilayers. J. Am. Chem. Soc. *134*, 10959-10965.
342. Marino, K.A., Prada-Gracia, D., Provasi, D., and Filizola, M. (2016). Impact of Lipid Composition and Receptor Conformation on the Spatio-temporal Organization of μ-Opioid Receptors in a Multi-component Plasma Membrane Model. PLoS Comp. Biol. *12*, e1005240.
343. Grisshammer, R. New approaches towards the understanding of integral membrane proteins: a structural perspective on G protein-coupled receptors. Protein Sci., n/a-n/a.
344. Robbins, D.J., Fei, D.L., and Riobo, N.A. (2012). The Hedgehog Signal Transduction Network. Science Signaling *5*, re6-re6.

345. Ingham, P.W., and McMahon, A.P. (2001). Hedgehog signaling in animal development: paradigms and principles. Genes Dev. *15*, 3059-3087.
346. Wang, C., Wu, H., Katritch, V., Han, G.W., Huang, X.-P., Liu, W., Siu, F.Y., Roth, B.L., Cherezov, V., and Stevens, R.C. (2013). Structure of the human smoothened receptor bound to an antitumour agent. Nature *497*, 338-343.
347. Nachtergaele, S., Whalen, D.M., Mydock, L.K., Zhao, Z., Malinauskas, T., Krishnan, K., Ingham, P.W., Covey, D.F., Siebold, C., and Rohatgi, R. (2013). Structure and function of the Smoothened extracellular domain in vertebrate Hedgehog signaling. eLife *2*, e01340.
348. Byrne, E.F.X., Sircar, R., Miller, P.S., Hedger, G., Luchetti, G., Nachtergaele, S., Tully, M.D., Mydock-McGrane, L., Covey, D.F., Rambo, R.P., et al. (2016). Structural basis of Smoothened regulation by its extracellular domains. Nature *535*, 517-522.
349. Sengupta, D., and Chattopadhyay, A. (2015). Molecular dynamics simulations of GPCR–cholesterol interaction: An emerging paradigm. Biochim. Biophys. Acta *1848*, 1775-1782.
350. Olsson, M.H.M., Sondergaard, C.R., Rostkowski, M., and Jensen, J.H. (2011). PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK(a) Predictions. Journal of chemical theory and computation *7*, 525-537.
351. Sondergaard, C.R., Olsson, M.H.M., Rostkowski, M., and Jensen, J.H. (2011). Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK(a) Values. Journal of chemical theory and computation *7*, 2284-2295.
352. Wang, C., Wu, H., Evron, T., Vardy, E., Han, G.W., Huang, X.P., Hufeisen, S.J., Mangano, T.J., Urban, D.J., Katritch, V., et al. (2014). Structural basis for Smoothened receptor modulation and chemoresistance to anticancer drugs. Nature communications *5*, 4355.
353. Pol-Fachin, L., Verli, H., and Lins, R.D. (2014). Extension and Validation of the GROMOS 53A6(GLYC) Parameter Set for Glycoproteins. Journal of computational chemistry *35*, 2087-2095.
354. Huang, P., Nedelcu, D., Watanabe, M., Jao, C., Kim, Y., Liu, J., and Salic, A. (2016). Cellular Cholesterol Directly Activates Smoothened in Hedgehog Signaling. Cell *166*, 1176-1187.e1114.
355. Zhang, X., Zhao, F., Wu, Y., Yang, J., Han, G.W., Zhao, S., Ishchenko, A., Ye, L., Lin, X., Ding, K., et al. (2017). Crystal structure of a multi-domain human smoothened receptor in complex with a super stabilizing ligand. Nature Communications *8*, 15383.
356. Yang, L., Yang, D., de Graaf, C., Moeller, A., West, G.M., Dharmarajan, V., Wang, C., Siu, F.Y., Song, G., Reedtz-Runge, S., et al. (2015). Conformational states of the full-length glucagon receptor. *6*, 7859.
357. Zhang, H., Qiao, A., Yang, D., Yang, L., Dai, A., de Graaf, C., Reedtz-Runge, S., Dharmarajan, V., Zhang, H., Han, G.W., et al. (2017). Structure of the full-length glucagon class B G-protein-coupled receptor. Nature *546*, 259-264.
358. Chen, L., Novicky, L., Merzlyakov, M., Hristov, T., and Hristova, K. (2010). Measuring the Energetics of Membrane Protein Dimerization in Mammalian Membranes. J. Am. Chem. Soc. *132*, 3628-3635.
359. Lelimousin, M., Limongelli, V., and Sansom, M.S.P. (2016). Conformational Changes in the Epidermal Growth Factor Receptor: Role of the Transmembrane Domain Investigated by Coarse-Grained MetaDynamics Free Energy Calculations. J. Am. Chem. Soc. *138*, 10611-10622.

360. Daily, M.D., Olsen, B.N., Schlesinger, P.H., Ory, D.S., and Baker, N.A. (2014). Improved Coarse-Grained Modeling of Cholesterol-Containing Lipid Bilayers. J. Chem. Theory Comput. *10*, 2137-2150.
361. Koldsø, H., Reddy, T., Fowler, P.W., Duncan, A.L., and Sansom, M.S.P. (2016). Membrane Compartmentalization Reducing the Mobility of Lipids and Proteins within a Model Plasma Membrane. The Journal of Physical Chemistry B *120*, 8873-8881.
362. Boags, A., Hsu, P.-C., Samsudin, F., Bond, P.J., and Khalid, S. (2017). Progress in Molecular Dynamics Simulations of Gram-Negative Bacterial Cell Envelopes. The Journal of Physical Chemistry Letters *8*, 2513-2518.
363. Sandoval-Perez, A., Pluhackova, K., and Böckmann, R.A. (2017). Critical Comparison of Biomembrane Force Fields: Protein–Lipid Interactions at the Membrane Interface. J. Chem. Theory Comput. *13*, 2310-2321.
364. Ingólfsson, H.I., Melo, M.N., van Eerden, F.J., Arnarez, C., Lopez, C.A., Wassenaar, T.A., Periole, X., de Vries, A.H., Tieleman, D.P., and Marrink, S.J. (2014). Lipid Organization of the Plasma Membrane. J. Am. Chem. Soc. *136*, 14554-14559.
365. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*.

Appendix A

This appendix contains supplemental data pertaining to Chapter 3.

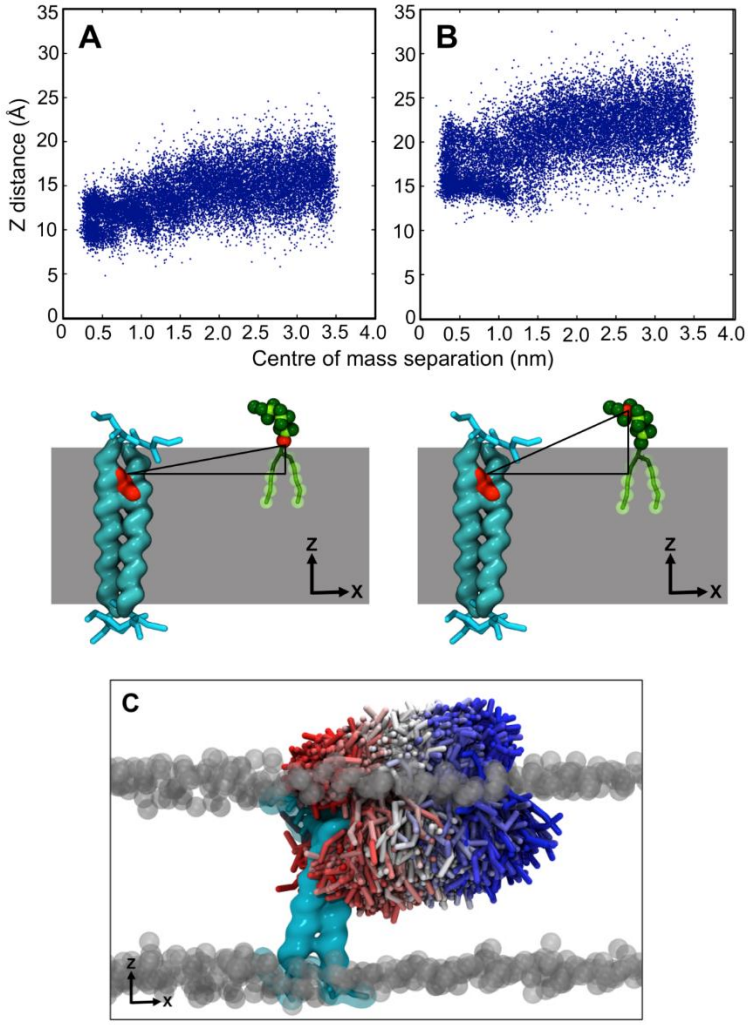

Figure A. 1: Relationship between GM3 Z positioning in the bilayer and lateral separation from the protein. GM3 "sinks" deeper into the bilayer in the bound state. (A+B) The z component of the relative distance between the centre-of-mass of the Gly residues within the N-terminal GxxGA dimerization motif, and the B1A (A) and INC (B) beads of GM3 over the reaction coordinate sampled during PMF calculations. Shown underneath each plot is a cartoon depicting the distance measured. The GxxGA motif and the beads used for distance calculations are indicated in red. Distances were computed for each frame of a concatenated trajectory of all windows used during calculation of the GM3 PMF profile. (C) 320 evenly distributed GM3 stick models colored by timestep from red to blue along the concatenated trajectory. The positioning of GM3 relative to the membrane surface (POPC headgroups shown as grey spheres) and the protein (cyan) can be seen to evolve over time as the lipid moves away from the protein.

Appendix B

This appendix contains supplemental data pertaining to Chapter 4.

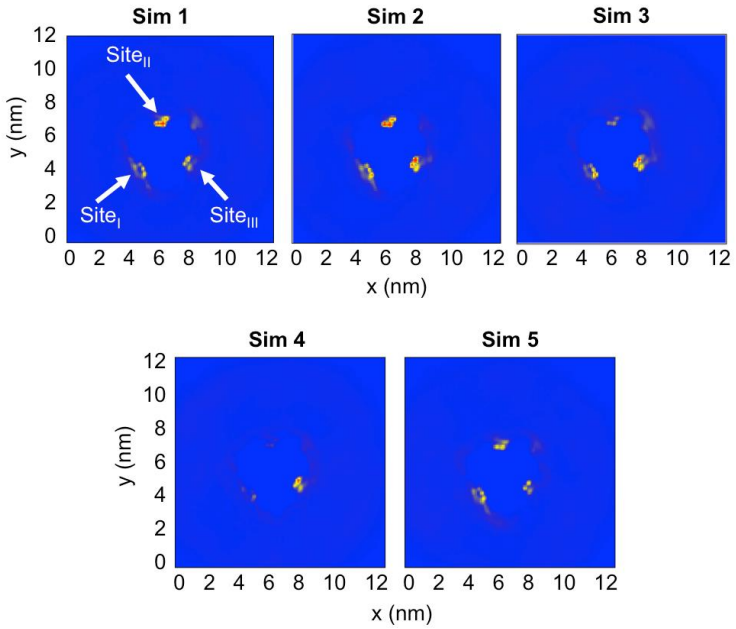

Figure B. 1: Convergence of density profiles. Heat maps showing the probability of CL head group occurrence in the membrane plane, from blue (low density) to red (high density). Data is shown for five independent repeat simulations, each starting from different initial lipid configurations and simulated for 6 μs.

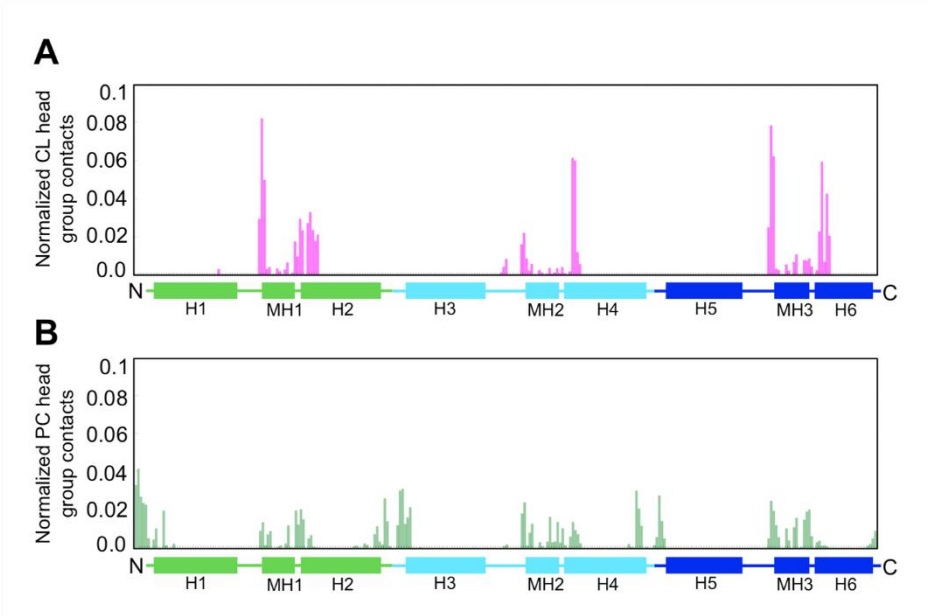

Figure B. 2: Lipid contact maps. Normalized average number of head group contacts per residue for (A) CL, and (B) PC lipids. Contacts were calculated over 5 x 6 μs repeat simulations, using a 6 Å cut-off. See Methods for full details of contact analysis.

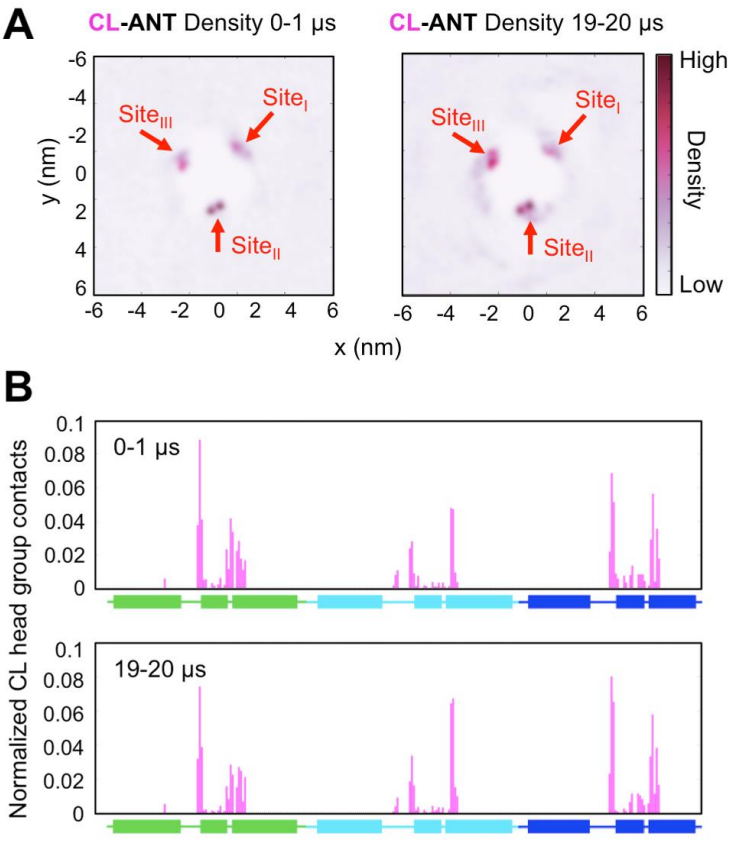

Figure B. 3: Oligomerization dependence of CL interactions. (A) Density plots showing the frequency of occurrence of CL around ANT, calculated for 0-1 µs (left: ANT molecules are monomeric) and 19-20 µs (right: ANT molecules are predominately oligomeric). (B) Normalized average number of CL head group contacts per residue, calculated for 0-1 µs and 19-20 µs. Contacts were calculated using a 6 Å cut-off, in the same manner as described for the small systems.

Appendix C

This appendix contains supplemental data pertaining to Chapter 5.

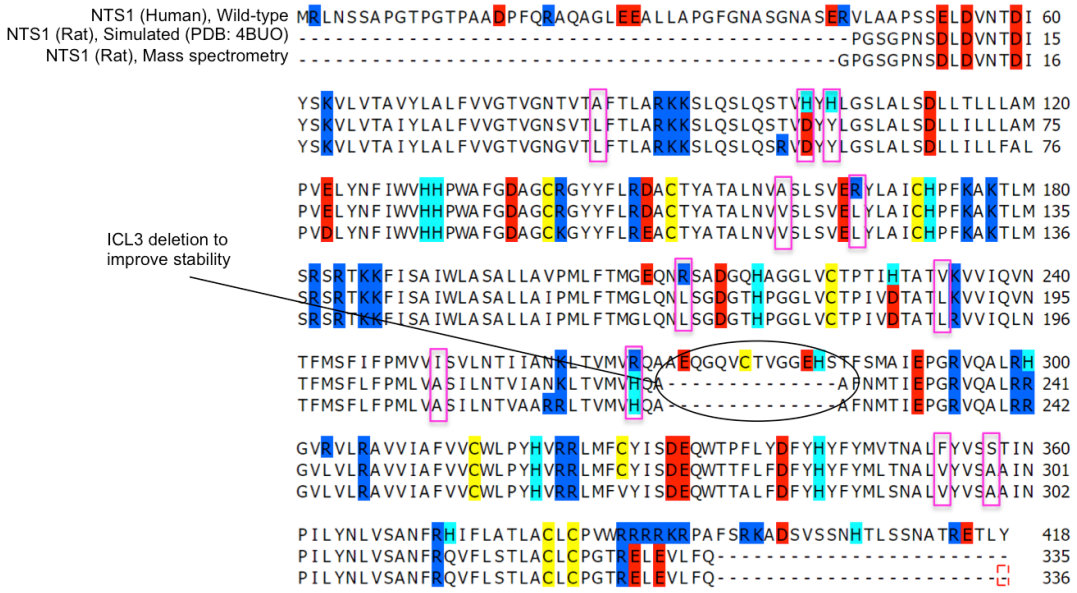

Figure C. 1: NTS1 sequence alignment. Primary sequences of wild-type human NTS1 (top), thermostabilised rat NTS1 (PDB: 4BUO) [333] used in simulations (middle), and thermostabilised rat NTS1 used for mass spectrometry analysis (bottom). Arg/Lys are indicated in blue, Glu/Asp in red, His in cyan, and Cys in yellow. The positions of the 11 thermostabilising mutations are indicated by magenta boxes. Alignments were produced using ClustelOmega [365]. The sequences PG, and LCPGTRELEVLFQ, respectively located at the N- and C-termini of the crystal structure construct, were not resolved in the crystal structure itself (4BUO) [348], and were omitted from simulations**.**

Appendix D

This appendix contains supplemental data pertaining to Chapter 6.

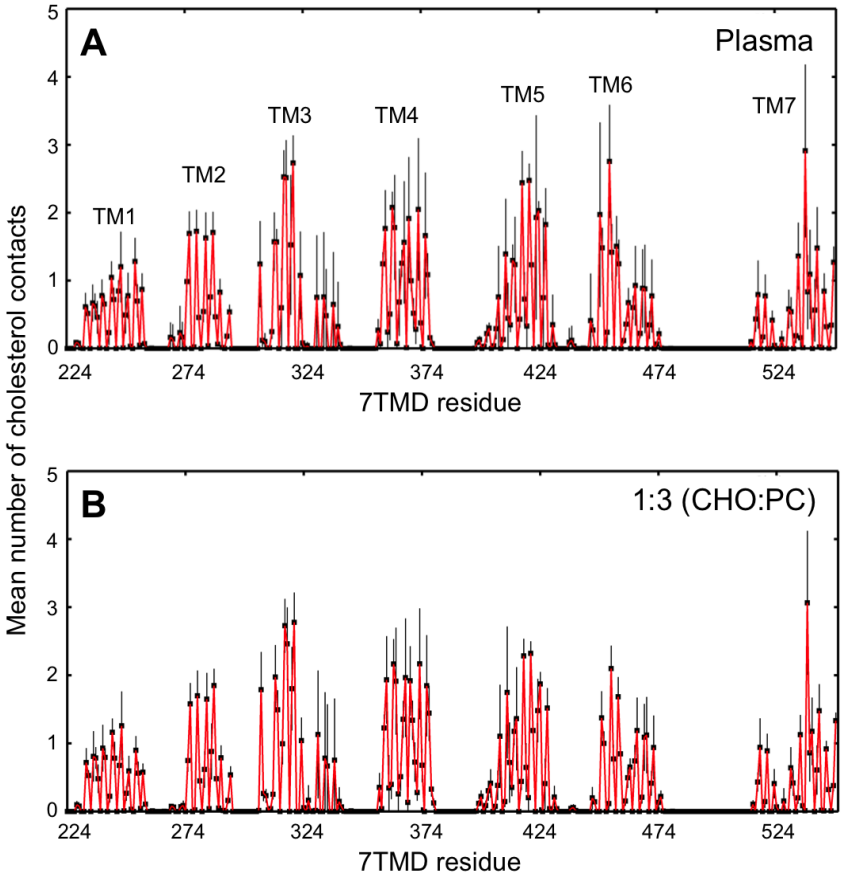

Figure D. 1: Robustness of cholesterol interaction patterns to increased lipid complexity. Contact maps are shown for plasma membrane simulations (A), and two component PC:Cholesterol simulations (B). The maps are broadly similar, with some variation seen around TM5 and TM6.