

Molecular Dynamics Simulations of Membrane Proteins in Nanodiscs with Investigation of Protein-Lipid Interactions



Michael R. Horrell

St Hugh's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2022

Abstract

Membrane proteins perform various important and wide-ranging functions. They exist within a lipid bilayer and their interactions with the components of this bilayer are of great scientific and pharmacological interest due to the ability of these components to regulate protein structure and function. Computational approaches such as multiscale molecular dynamics simulations are well suited to the study of these systems. As such, the research presented here uses an array of these approaches to examine and characterise the interactions between membrane proteins and lipids in two environments: nanodiscs and lipid bilayers.

There has been exponential growth in the use of nanodiscs for solubilisation of membrane proteins in a native-like environment, where the membrane protein is embedded in a patch of lipid bilayer constrained by helical proteins. However, the impact of constraining the nanodisc components in this manner is currently not well understood. In this research molecular dynamics simulations are used to investigate these effects; the results presented here describe the changes in lipid properties which occur as a consequence of constraining the nanodisc components and subsequent variations in protein-lipid interactions.

One family of membrane proteins, the G-protein-coupled receptors, was selected for further study. A simulation approach was used to examine the binding of a membrane lipid, cholesterol, to potential allosteric modulatory sites on these proteins, including the previously identified TM345 site. Interactions with cholesterol at this site were found to be common to several members of this family despite limited sequence conservation.

The results presented here demonstrate the importance of understanding the environment in which membrane proteins exist and the impact of changes to this environment upon protein behaviour, properties and interactions.

Acknowledgements

Firstly, my thanks go to my supervisors Mark Sansom and Phill Stansfeld, for providing so much advice, input and guidance over the course of my four years working towards this DPhil. I must also thank the BBSRC's Oxford DTP, UCB for funding my iCASE studentship, and Jiye Shi, my industrial supervisor.

The SBCB unit in the Department of Biochemistry has been a wonderful place to work (both when we could actually go in without lockdowns interfering and even when working remotely thanks to the frequent SBCB virtual coffee meetups). A very talented great group of people has been brought together, creating a supportive and constructive environment.

The post docs of SBCB require a huge thankyou from me for putting up with the many questions and distractions I have pestered them with over the years. Chiefly among this group is Robin Corey, who has always had a useful insight or helpful comment no matter the topic and who must be thanked for his very enthusiastic and detailed thesis proof reading skills. Several other post docs, past and present, have played a big part in helping me understand this field: Andreas Larsen, Anna Duncan, Keith Cassidy, Charlotte Lynch, Owen Vickery and Wanling Song; thank you for helping me develop and grow as a scientist. Irfan Alibay also deserves special thanks for our many technical discussions about the ins and outs of computing, for making sure that the lab computers run smoothly, and for making emergency repairs to my workstation after I 'fixed it'.

My fellow DPhil students must also be thanked for keeping me (relatively) sane over the course of my degree and sharing so many memories over lunches, coffee breaks, and expeditions to ATS or the Gloucester Green food market. In particular my thanks go to Laura, Patrick, Mina, Jodie, Charlie, Tanadet, Nick, Shanlin, Bertie, Dan, Rocco, Linda, Alex, and to Josh, my fellow self-appointed member of the 'lunch committee'.

Outside SBCB much of my time has been spent with the members of OURC, who have made a huge impact on me and given me many happy memories over my nearly ten years in Oxford. In particular, my good friend Jonny Page must be acknowledged for many years of friendship and encouragement, though when I informed him of my place to study for a DPhil his response was ‘You clearly didn’t learn from my mistake’.

My family must also be thanked for a lifetime of love and encouragement; I have been very fortunate to have benefited from their support. In particular, my Mum, Dad, Gran and of course my brother Will, who have supported me over the course of this (rather long) journey.

My other half, Harriet, whose love has supported me through the highs and lows of the past ten years, needs more thanks than I could ever say, especially for proof reading this thesis more times than I could count. Without her friendship, good humour and patience I would have not been able to come this far.

Finally, my thanks go to the other people who so very kindly proof read this thesis and thus have probably been exposed to more than a lifetime’s worth of nanodiscs and simulations. In particular, Harriet’s parents, Ann and Matthew, were as sharp eyed as ever and their contributions are greatly appreciated. Lastly to Carol, you have been a hugely influential support over my 10 years in Oxford and thank you for introducing me to the world of punctuation; you are quite probably responsible for every semi-colon in this thesis.

Michael Horrell

February 2022, University of Oxford

List of Publications

Published:

- 1) Yen, H. Y.; Hoi, K. K.; Liko, I.; Hedger, G.; **Horrell, M. R.**; Song, W. L.; Wu, D.; Heine, P.; Warne, T.; Lee, Y.; Carpenter, B.; Pluckthun, A.; Tate, C. G.; Sansom, M. S. P.; Robinson, C. V., PtdIns(4,5)P-2 stabilizes active states of GPCRs and enhances selectivity of G-protein coupling. *Nature* 2018, 559 (7714), 424-427.
- 2) Ansell, T. B.; Curran, L.; **Horrell, M. R.**; Pipatpolkai, T.; Letham, S. C.; Song, W.; Siebold, C.; Stansfeld, P. J.; Sansom, M. S. P.; Corey, R. A., Relative Affinities of Protein–Cholesterol Interactions from Equilibrium Molecular Dynamics Simulations. *J. Chem. Theory Comput.* 2021, 17 (10), 6548-6558.
- 3) Song, W.; Corey, R. A.; Ansell, T. B.; Cassidy, C. K.; **Horrell, M. R.**; Duncan, A. L.; Stansfeld, P. J.; Sansom, M. S. P., PyLipID: A Python package for analysis of protein-lipid interactions from MD simulations. *J. Chem. Theory Comput.* 2022. DOI: 10.1021/acs.jctc.1c00708 (*in press*)

In preparation:

- 1) **Horrell, M. R.**; Corey, R. A.; Larsen, A. H; Shi, J; Sansom, M. S. P.; Stansfeld, P. J., Multiscale Simulations of Membrane Proteins in Nanodiscs. (*Described in Thesis Chapters 3 and 4*)
- 2) **Horrell, M. R.**; Corey, R. A.; Michelarakis, N.; Shi, J; Stansfeld, P. J.; Sansom, M. S. P., Investigation of Cholesterol TM345 Binding Sites on G-Protein-Coupled Receptors using Molecular Dynamics Simulations. (*Described in Thesis Chapter 5*)

Contents

Chapter 1 - Introduction	1
1.1 Overview of Membrane Proteins	2
1.1.1 Functions of Membrane Proteins	2
1.2 Methods for Structural Determination of Membrane Proteins	7
1.2.1 Historical Background	7
1.2.2 X-Ray Crystallography	7
1.2.3 Single Particle Cryo-EM	8
1.2.4 Experimental Methods for Membrane Protein Solubilisation	9
1.2.4.1 Use of Nanodiscs for Membrane Protein Solubilisation	12
1.3 Use of Molecular Dynamics to Study Protein-Lipid Interactions	14
1.3.1 MD Simulations and the Coarse-Grained Approach	17
1.3.2 Non-Specific Lipid Interactions with Membrane Proteins	20
1.3.3 Lipid Regulation of Membrane Proteins and Specific Protein-Lipid Interactions	21
1.3.3.1 The Interaction of PIP ₂ with GPCRs	22
1.3.3.2 The Interaction of Cholesterol with GPCRs	23
1.3.3.3 The Interaction of Cardiolipin with Membrane Proteins	25
1.4 Nanodisc Simulations	27
1.5 Research Overview	29
Chapter 2 - Theory and Methods	31
2.1 Molecular Dynamics Simulations	32
2.2 Force Fields	34
2.2.1 Bonded Interactions	35
2.2.1.1 Bond Stretching	35
2.2.1.2 Angle Bending	36
2.2.1.3 Dihedral Angles	36
2.2.2 Non-Bonded Interactions	37
2.2.2.1 Van der Waals Interactions	37
2.2.2.2 Electrostatic Interactions	38
2.2.3 Integration Methods	39
2.2.4 Energy Minimisation	40
2.2.5 Timesteps	41
2.2.5.1 Virtual Sites	42

2.2.6 Bond Constraints	42
2.2.7 Periodic Boundary Conditions	43
2.2.8 Temperature and Pressure Coupling	45
2.2.8.1 Thermostats	46
2.2.8.2 Barostats	47
2.2.9 Long-Ranged Interactions	48
2.3 Coarse-Grained Simulations	50
2.3.1 The Martini Force Field	52
2.3.1.1 Martini Structural Mapping	52
2.3.1.2 Martini Particle Interactions	54
2.3.1.3 Secondary and Tertiary Structural Elements in Martini	56
2.3.1.4 The Martini Water Model	57
2.3.2 The Setup of CG Systems	59
2.3.3 Multiscale Modelling	60
Chapter 3 - An Automated Nanodisc Pipeline Methodology	62
3.1 Introduction	63
3.1.1 Solubilisation Techniques	63
3.1.2 Nanodiscs Overview	66
3.1.3 Status of Simulation Studies of Nanodiscs	68
3.1.4 Pipeline Approach	71
3.2 Methods	72
3.2.1 Coarse-Grained Simulations	72
3.2.2 Atomistic Simulations	74
3.2.3 Analysis	75
3.2.3.1 SAXS Analysis	75
3.2.3.2 Lipid Analysis	76
3.3 Results	78
3.3.1 Alternative Nanodisc Simulation Methodologies	78
3.3.2 A Multiscale Pipeline	80
3.3.3 Elastic Network Models	83
3.3.3.1 Effects of ENM choice on CG Systems: Comparison with SAXS Data	86
3.3.3.2 Effects of ENM Choice on Atomistic Systems: Comparison with SAXS Data	88
3.3.4 Analysis of Pipeline Output in Atomistic Detail	89
3.3.4.1 Root Mean Square Fluctuation	90

3.3.4.2 Axial Ratio	92
3.3.4.3 Area Per Lipid	93
3.3.4.4 Order Parameters.....	95
3.4 Conclusions	98
3.4.1 ENM Use in Nanodisc Simulations	99
3.4.2 Lipid Properties	101
Chapter 4 - Simulation Studies of Nanodiscs with Membrane Proteins	105
4.1 Introduction	106
4.1.1 Background	106
4.1.2 Existing Approaches	107
4.1.2.1 Constructing Nanodiscs Using CHARMM-GUI.....	107
4.1.2.2 Constructing Nanodiscs Using Self-Assembly.....	109
4.1.2.3 Constructing Nanodiscs Using insane.py	111
4.1.3 Simulation of Membrane Proteins in Nanodiscs	113
4.2 Methods	114
4.2.1 Coarse-Grained Simulations	114
4.2.2 Atomistic Simulations.....	116
4.2.3 Analysis.....	116
4.2.3.1 Lipid Analysis	117
4.2.3.2 Protein-Lipid Interaction Analysis.....	117
4.3 Results	120
4.3.1 Membrane Protein Pipeline Overview and Output	120
4.3.2 Membrane Protein Interactions with MSPs.....	126
4.3.3 Analysis of Membrane Protein System Properties	129
4.3.3.1 Area Per Lipid in Nanodiscs with Membrane Proteins	129
4.3.3.2 Lipid Deformation	131
4.3.3.3 Order Parameters for Bilayer and Nanodisc Systems	132
4.3.3.4 Root Mean Square Fluctuation Analysis for TRPV6	134
4.3.4 Protein-Lipid Interactions	136
4.3.4.1 Protein-Lipid Interactions in Nanodiscs and Bilayers	136
4.3.4.2 The Effect of Saturation on Protein-Lipid Interactions	142
4.4 Conclusions	145
4.4.1 Membrane Protein-MSP Interactions	145
4.4.2 Lipid Properties	147
4.4.3 Protein Properties and Protein-Lipid Interactions	149

Chapter 5 - Interactions Between Cholesterol and G-Protein-Coupled Receptors.....	152
5.1 Introduction	153
5.1.1 G-Protein-Coupled Receptors.....	154
5.1.1.1 The Activation of GPCRs.....	156
5.1.2 GPCR Interactions with Lipophilic Ligands	157
5.1.2.1 The TM345 Site on Free Fatty Acid Receptor 1	158
5.1.2.2 GPCR Interactions with Cholesterol	159
5.2 Methods	162
5.2.1 Coarse-Grained Simulations	162
5.2.2 Analysis.....	162
5.3 Results.....	164
5.3.1 Cholesterol Interaction with the TM345 Site on FFAR1	164
5.3.1.1 Residues of Interest in the TM345 Binding Site on FFAR1.....	167
5.3.2 Investigation of the TM345 Site on Additional GPCRs	170
5.3.2.1 General Properties of the TM345 Binding Site	183
5.3.2.2 The TM345 Binding Site on ADRB2.....	188
5.3.2.3 TM345 Binding Site Specificity	191
5.3.2.4 Comparison of Residence Times and GPCR State.....	192
5.3.2.5 Protein Conformation and Cholesterol Binding	194
5.4 Conclusions	196
5.4.1 Cholesterol Binding Variation within Class A GPCRs	199
5.4.2 Cholesterol Binding Specificity at the TM345 Site.....	201
5.4.3 Cholesterol Binding Variation with Conformational State	201
Chapter 6 - Summaries and Future Directions	204
6.1 Chapter Summaries	205
6.1.1 Simulation Models of Nanodiscs	205
6.1.2 Protein-Lipid Interactions in Nanodiscs	206
6.1.3 Interaction of Cholesterol with GPCRs	208
6.2 Future Directions.....	211
6.2.1 General Remarks	211
6.2.2 Nanodiscs	211
6.2.3 Membrane Protein-Lipid Interactions.....	214
6.3 Closing Remarks.....	216
Bibliography	217

List of Figures

Chapter 1 – Introduction

Figure 1.1 – The increasing number of structurally determined membrane proteins since the formation of the PDB	3
Figure 1.2 – Three techniques for membrane protein solubilisation.....	11
Figure 1.3 – A nanodisc viewed from above the plane of the disc	13
Figure 1.4 – Structures of common mammalian lipids.....	16
Figure 1.5 – The coarse-graining process of a lipid.....	19
Figure 1.6 – The structure of cardiolipin	25
Figure 1.7 – Models for the configuration of MSPs in nanodiscs.....	28

Chapter 2 – Theory and Methods

Figure 2.1 – A two-dimensional representation of periodic boundary conditions	44
Figure 2.2 – A two-dimensional representation of a cut-off scheme	45
Figure 2.3 – System size and length for four different simulation methodologies.....	51
Figure 2.4 – Martini 2 mapping of amino acids	54
Figure 2.5 – The coarse-graining process of a membrane protein	60
Figure 2.6 – The conversion of CG cholesterol to atomistic detail.....	61

Chapter 3 – An Automated Nanodisc Pipeline Methodology

Figure 3.1 – Further techniques for solubilisation of membrane proteins.....	65
Figure 3.2 – An atomistic representation of a nanodisc	67
Figure 3.3 – The nanodisc self-assembly methodology	79
Figure 3.4 – An overview of the automated pipeline methodology.....	82
Figure 3.5 – A visual representation of the restraints applied to the MSPs by the ElNeDyn force field.....	84
Figure 3.6 – Effects of ENM choice on MSP stability	85
Figure 3.7 – Predicted SAXS profiles of CG systems constructed with different ENMs	87
Figure 3.8 – Predicted SAXS profiles of systems constructed with different ENMs after conversion to atomistic detail	88
Figure 3.9 – The final structures from two atomistic simulations, with and without an ENM	90
Figure 3.10 – RMSF plots for two atomistic simulations with and without an ENM.....	91

Figure 3.11 – Variation in APL with distance from the centre of a nanodisc	94
Figure 3.12 – Lipid order variation in nanodisc and bilayer environments.....	97
Chapter 4 – Simulation Studies of Nanodiscs with Membrane Proteins	
Figure 4.1 – Trends in nanodisc utilisation	106
Figure 4.2 – The product of the CHARMM-GUI Nanodisc Builder with an embedded membrane protein.....	109
Figure 4.3 – The product of the self-assembly methodology with an embedded membrane protein	111
Figure 4.4 – The product of assembly using the insane.py script.....	112
Figure 4.5 – Application of the automated pipeline methodology to an integral membrane protein	121
Figure 4.6 – The product of the automated pipeline methodology when applied to TRPV6	122
Figure 4.7 – The products of the automated pipeline methodology for multiple systems of varying sizes	125
Figure 4.8 – Interactions between a membrane protein and the MSPs	128
Figure 4.9 – Variation in APL with distance from the centre of a nanodisc with an embedded membrane protein.....	130
Figure 4.10 – Visual representation of lipid patch deformation	132
Figure 4.11 – Variations in lipid order in nanodisc and bilayer environments with embedded membrane proteins.....	133
Figure 4.12 – RMSF plots for TRPV6 in nanodisc and bilayer environments	135
Figure 4.13 – Occupancy and residence times of three lipid species during nanodisc and bilayer simulations with embedded GltTk.....	137
Figure 4.14 – Occupancy and residence times of lipid species during nanodisc and bilayer simulations with embedded membrane proteins	139
Figure 4.15 – Effect of lipid saturation on occupancy and residence times during nanodisc and bilayer simulations with a TmrAB ABC exporter.....	144
Chapter 5 – Interactions Between Cholesterol and G-Protein-Coupled Receptors	
Figure 5.1 – Schematic representation of the structure of a GPCR	155
Figure 5.2 – The TM345 binding site	159

Figure 5.3 – Structure of the allosteric regulator AP8 alongside cholesterol.....	164
Figure 5.4 – The TM345 binding site of AP8 and cholesterol on FFAR1	166
Figure 5.5 – The TM345 binding site of cholesterol on FFAR1 with residence times and residues of interest	169
Figure 5.6 – Phylogenetic analysis of the GPCRs used in this research, with GPCR class and ligand type	171
Figure 5.7 – Sequence alignments of TM 3, 4 and 5, with residues coloured according to residence time.....	176
Figure 5.8 – The binding of cholesterol to the TM345 site on ADRB2.....	190
Figure 5.9 – Specificity of cholesterol binding at the TM345 site	192
Figure 5.10 – The effect of conformational change on cholesterol binding to the TM345 site on NTR1.....	195

List of Tables

Chapter 3 - An Automated Nanodisc Pipeline Methodology

Table 3.1 - Nanodisc axial ratios for atomistic simulations.....	92
--	----

Chapter 4 - Simulation Studies of Nanodiscs with Membrane Proteins

Table 4.1 - Application of the automated pipeline methodology to several membrane proteins with individual lipid patch compositions	124
---	-----

Chapter 5 - Interactions Between Cholesterol and G-Protein-Coupled Receptors

Table 5.1 - Structures of the GPCRs used for investigation of cholesterol binding at the TM345 site.....	172
Table 5.2 - Cholesterol binding at the TM345 site on a variety of GPCRs, with residence times and residues of particular interest.....	184
Table 5.3 - Activation state of the GPCRs simulated.....	193

List of General Abbreviations

APL	Area per lipid
BAM	Beta-barrel assembly machinery
CCM	Cholesterol consensus motif
CG	Coarse-grained
CRAC	Cholesterol recognition amino acids consensus
Cryo-EM	Cryo-electron microscopy
ECL	Extracellular loop
ENM	Elastic network model
EPR	Electron paramagnetic resonance
GDP	Guanosine-5'-diphosphate
GPCR	G-protein-coupled receptor
GROMACS	Groningen machine for chemical simulations
GTP	Guanosine-5'-triphosphate
H8	Helix eight
ICL	Intracellular loop
L _d	Lipid disordered
LINCS	Linear constraint solver
L _o	Lipid ordered
MD	Molecular dynamics
MSP	Membrane scaffold protein
NMR	Nuclear magnetic resonance
PDB	Protein data bank
PME	Particle mesh Ewald
RMSF	Root mean square fluctuation
SANS	Small angle neutron scattering
SAXS	Small angle X-ray scattering
SMALP	Styrene-maleic acid lipid particle
TM	Transmembrane

List of Protein and Lipid Abbreviations

AA2AR	Adenosine A _{2A} Receptor
ACM2	Muscarinic Acetylcholine Receptor M2
ADRB2	β ₂ Adrenergic Receptor
CARD	Cardiolipin
CCR2	C-C Chemokine Receptor type 2
CCR5	C-C Chemokine Receptor type 5
CCR9	C-C Chemokine Receptor type 9
CHOL	Cholesterol
CRFR1	Corticotropin Releasing Factor Receptor 1
DLPC	1,2-dilauroyl-sn-glycero-3-phosphocholine
DO-	1,2-dioleoyl-
DP-	1,2-dipalmitoyl-
DRD3	D (3) Dopamine Receptor
FFAR1	Free Fatty Acid Receptor 1
GLR	Glucagon Receptor
GLP1R	Glucagon-Like Peptide 1 Receptor
GlyR	Strychnine-sensitive Glycine Receptor
NTR1	Neurotensin Receptor Type 1
P2RY1	P2Y Purinoceptor 1
-PA	-phosphatidic acid
-PC	-phosphatidylcholine
-PE	-phosphatidylethanolamine
-PG	-phosphatidylglycerol
-PI	-phosphatidylinositol
PIP ₂	Phosphatidylinositol 4,5-bisphosphate
PO-	1-palmitoyl-2-oleoyl-
-PS	-phosphatidylserine
S1PR1	Sphingosine-1-Phosphate Receptor 1
TRP	Transient Receptor Potential cation channel

1

Chapter 1 - Introduction

Contents:

1.1 Overview of Membrane Proteins.....	2
1.1.1 Functions of Membrane Proteins	2
1.2 Methods for Structural Determination of Membrane Proteins	7
1.2.1 Historical Background	7
1.2.2 X-Ray Crystallography	7
1.2.3 Single Particle Cryo-EM	8
1.2.4 Experimental Methods for Membrane Protein Solubilisation	9
1.2.4.1 Use of Nanodiscs for Membrane Protein Solubilisation.....	12
1.3 Use of Molecular Dynamics to Study Protein-Lipid Interactions	14
1.3.1 MD Simulations and the Coarse-Grained Approach.....	17
1.3.2 Non-Specific Lipid Interactions with Membrane Proteins.....	20
1.3.3 Lipid Regulation of Membrane Proteins and Specific Protein- Lipid Interactions	21
1.3.3.1 The Interaction of PIP ₂ with GPCRs	22
1.3.3.2 The Interaction of Cholesterol with GPCRs.....	23
1.3.3.3 The Interaction of Cardiolipin with Membrane Proteins	25
1.4 Nanodisc Simulations.....	27
1.5 Research Overview	29

1.1 Overview of Membrane Proteins

The cell is the basic unit of life, and every cell is enclosed by a membrane which separates its interior from the exterior environment. This is true for both prokaryotes and eukaryotes, which, having diverged more than 1,800 million years ago, comprise all living organisms. In addition, the compartments of eukaryotic cells are generally separated by membranes¹. Eukaryotic membranes are typically 3 nm to 5 nm thick² and are composed of lipids in a bilayer arrangement. Growth in the field of lipidomics has revealed the complexity of the bilayer environment, which is evident from the identification of over 47,000 unique lipid structures, as determined by the LIPID MAPS database³⁻⁴. Proteins represent further complexity in the structure of membranes, being attached to or embedded in the lipid bilayer and performing a wide variety of functions.

1.1.1 Functions of Membrane Proteins

Membrane proteins are responsible for the execution of various vital processes within the human body, including transport across cell membranes, signal transduction, cell adhesion, enzymatic functions and both the innate and adaptive immune responses. It has been estimated that of the over 22,000 protein coding genes in humans⁵, approximately 23% encode membrane proteins⁶. Due to their abundance and varied function, membrane proteins are considered the largest class of potential drug target and it has been estimated that approximately 50% of all druggable small molecule drug binding sites are located on membrane proteins⁷. Despite this importance, membrane protein

structures make up less than 3% of those deposited in the Protein Data Bank (PDB) ⁸, though there has been exponential growth in the number of membrane protein structures solved in recent years. The number of unique membrane protein folds and unique proteins deposited in the PDB has also been increasing annually, as demonstrated in **Figure 1.1**.

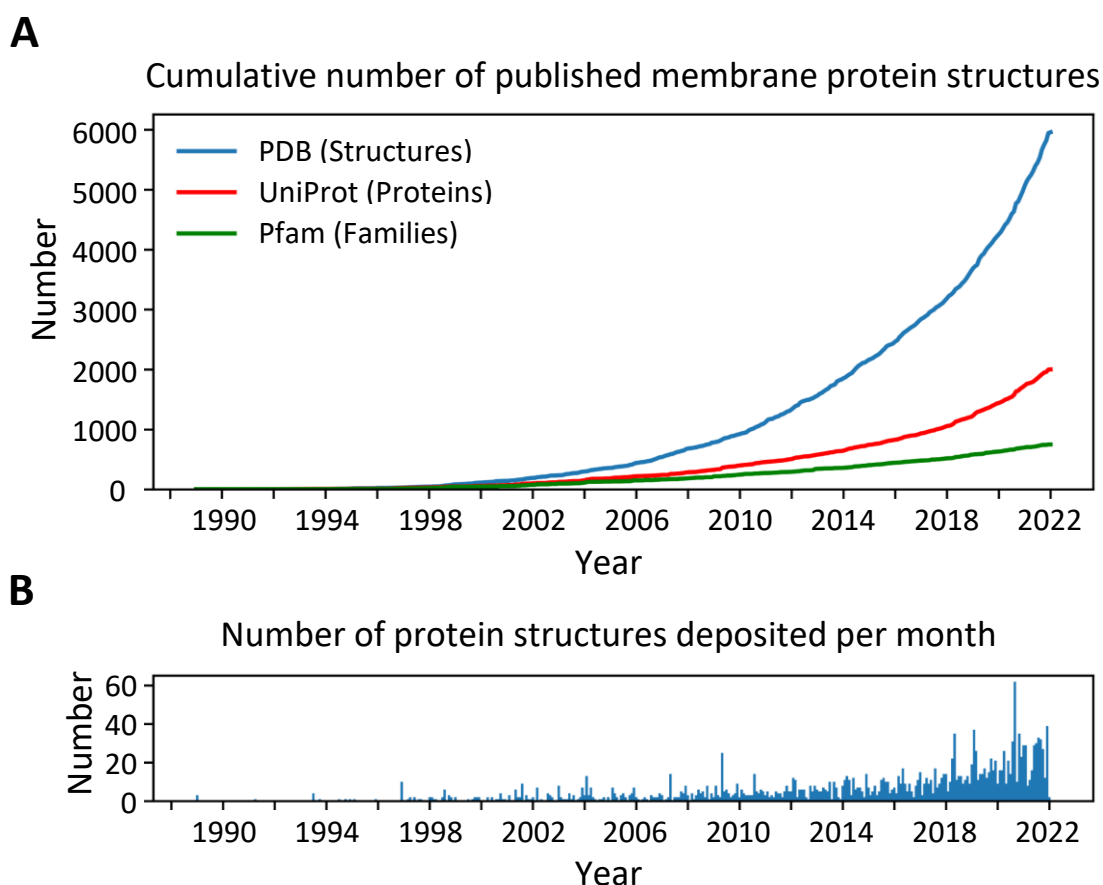


Figure 1.1 – The increasing number of structurally determined membrane proteins since the formation of the PDB

A: The number of membrane protein structures that have been solved, as measured by deposition of structures in the PDB (shown in blue), has increased exponentially over the past several decades due to technological advances and improvements to solubilisation techniques. Increasing deposition rate of unique proteins and protein families is shown by the red and green lines respectively.

B: The increasing monthly number of structures deposited in the PDB.

Data collated by memprotmd.bioch.ox.ac.uk ⁹⁻¹⁰

Structural knowledge of this important class of proteins is essential for understanding their functions and regulatory processes, and can be used to inform structure-based drug design and screening of potential drug fragments and molecules ¹¹. Membrane proteins can be categorised into two main groups: peripheral and integral membrane proteins, which differ by the nature of their association with the membrane. Peripheral membrane proteins can interact with the membrane both directly, for example through the use of hydrophobic regions or covalently attached lipids or GPI-anchors ¹², and indirectly via electrostatic or hydrophobic interactions. Integral membrane proteins use α -helices or β -sheets to span the membrane sometimes doing so multiple times within their structure. These integral membrane proteins will be the focus of the research presented here.

Membrane proteins are a common target during drug discovery and development on account of their wide-ranging functions, clinical importance and accessible location on the surface of cells. One family of membrane proteins which is of particular interest is the G-Protein-Coupled Receptors (GPCRs), due to their association with a large variety of disorders, ranging from cancer to diabetes and hypertension ¹³. Currently approved drugs which have been designed to target GPCRs include: Rosiglitazone, which acts on the Free Fatty Acid Receptor 1 to help prevent hyperglycaemia in type 2 diabetes mellitus; Fingolimod, which acts on the Sphingosine 1-Phosphate Receptor 1 to help reduce frequency of symptomatic episodes and slow the worsening of disability in patients with multiple sclerosis; and Maraviroc, which acts on the

C-C Chemokine Receptor type 5 to inhibit virus entry into cells in patients with Human Immunodeficiency Virus ¹⁴. In addition, Suramin 2, an antiparasitic which acts on the P2Y Purinoceptor and has long been used to treat sleeping sickness, is currently being investigated for potential repurposing due to its inhibition of SARS-CoV-2 ¹⁵.

The interactions between the two major components of cellular membranes, membrane proteins and lipids, are increasingly appreciated to be essential when understanding the function, dynamics and regulation of these membrane components ¹⁶. Identification of specific lipid binding sites on the surfaces of membrane proteins can provide a mechanistic understanding of lipid modulation of membrane protein function, as has been demonstrated in various membrane proteins including the Kir 2.2 channel, which is modulated by the lipid PIP₂ ¹⁷. This lipid has also been demonstrated to mediate electrostatic interactions between the N-terminal and intracellular loops of the human dopamine transporter, acting as an allosteric regulator ¹⁸. Another example of the role of protein-lipid interactions is the mediation of the interface between dimers of the GDP-mannose transporter, Vrg4, by two monoolein lipids ¹⁹. These discovery of lipids which are able to allosterically modulate the properties of membrane proteins are also of interest for protein druggability and drug discovery ²⁰.

In some cases, a bound lipid is identified during structural determination of a membrane protein ²¹, such as the cholesterol which was identified during structural determination of the 5-HT receptor ²². However, most protein

structures have been solved in the presence of detergents, which may have removed the physiological binding partners of these proteins, preventing identification of the particular ligand species. This is further compounded by difficulties assigning specific ligand species to detected densities in the structure. In other cases, the techniques used for structural determination of a membrane protein may lock the protein in a conformation which renders it unable to bind its ligand ²⁰.

Several approaches have been developed for identification of lipid binding sites on membrane proteins including native mass spectrometry ²³, Electron Paramagnetic Resonance (EPR) ²⁴⁻²⁵, and Nuclear Magnetic Resonance (NMR) ²⁶. Molecular Dynamics (MD) simulations are also well suited for the investigation of protein-lipid interactions within the membrane and can provide insight into interactions which may not be observed during structural studies due to the presence of detergents or weak ligand binding. These simulation approaches have been able to successfully identify lipid binding sites on several membrane proteins, and have shown good agreement with the results from structural studies ²⁷, for example in the case of PIP₂ binding to inward rectifier K⁺ channels ^{17, 28}.

1.2 Methods for Structural Determination of Membrane Proteins

1.2.1 Historical Background

Various methods have been utilised to determine the structures of membrane proteins. Some of the earliest structural insights into membrane proteins were provided in 1975 using electron crystallography. These studies identified the presence of the α -helices that make up the membrane-spanning components of bacteriorhodopsin, though at a relatively low resolution of 7 Å²⁹. The first integral membrane protein structure determined using X-ray crystallography was the bacterial photosynthetic reaction centre and associated co-factors, which was reported in 1985³⁰. The structure of another bacterial protein, a porin, was published in 1991³¹ and subsequently, in 1994, the structure of bacteriorhodopsin was determined in more detail using cryo-electron microscopy (cryo-EM)³². The full-length structure of bacteriorhodopsin was solved using X-ray crystallography in 1997, at a resolution of 2.5 Å³³. More recently, NMR has been utilised for the determination of the structure of smaller membrane proteins³⁴ for example a mitochondrial barrel protein, the voltage-dependent anion channel³⁵.

1.2.2 X-Ray Crystallography

Determination of membrane protein structures has occurred at a slower pace than that of soluble proteins. This is for several reasons, including difficulties in purification and crystallisation of membrane proteins from a detergent solution, with only some membrane proteins being amenable to detergent

solubilisation³⁶ and many not forming regular crystal lattices due to their flexible regions.

The development of lipidic cubic phase crystallisation in 1996³⁷ mitigated some of these issues. This membrane-mimetic matrix³⁸ is an arrangement of lipids not found in cellular membranes; instead it is a structured liquid crystal which forms spontaneously upon the mixing of certain lipid species with water, enabling membrane proteins to be solubilised in a bilayer-like environment. This technique, along with improvement of detergents³⁶ and advances in hetero-expression systems for membrane protein expression, has enabled the utilisation of X-ray crystallography for structural determination of members of over 750 protein families (memprotmd.bioch.ox.ac.uk), including many members of the GPCR family. GPCR structural determination in particular has also recently been facilitated by the creation of fusion proteins, commonly by insertion of T4 lysozyme into a flexible intracellular loop³⁹.

These advances in X-ray crystallography enabled the first structural determination of an ion channel, the potassium channel KcsA, which was reported in 1998⁴⁰. This structure provided insights into the assembly of this tetrameric membrane protein as well as information about the selectivity filter which is formed from the backbone carbonyl oxygens, allowing selection by coordination of de-solvated K⁺ ions.

1.2.3 Single Particle Cryo-EM

Cryo-EM has increasingly been utilised for structural determination of membrane proteins. Recent technological advances have facilitated huge

improvements in the resolution of cryo-EM determined membrane protein structures resulting in the so-called 'resolution revolution' ⁴¹⁻⁴². This has been primarily driven by the development of direct electron detectors ⁴³, as well as improvements in data processing techniques ⁴⁴⁻⁴⁶. As a result of these improvements, many cryo-EM structures have now reached near atomic resolutions.

One of the major advantages of cryo-EM is that it does not require positioning of the sample in a regular repeating crystal, making this technique particularly useful for membrane proteins which have been resistant to structural determination by X-ray crystallography ⁴⁷. These membrane proteins often contain disordered regions which have typically hindered crystallographic methods ⁴⁸. Another benefit of cryo-EM is the reduced likelihood of crystal artefacts in protein structures such as the kinking of helices due to strong crystal contacts ⁴⁹. Examples of protein structures solved using cryo-EM methods include the Transient Receptor Potential cation channel subfamily V1 (TRPV1) ion channel ⁵⁰, the strychnine-sensitive Glycine Receptor (GlyR) ⁵¹ and the Magnesium Channel CorA ⁵².

1.2.4 Experimental Methods for Membrane Protein Solubilisation

Several solubilisation techniques have been developed for the structural determination of membrane proteins, often using detergents to create discoidal conformations such as micelles ⁵³. Unfortunately, use of detergents has several drawbacks, primarily resulting from the removal of the native lipids that

surrounded the protein in its physiological state. This can cause changes in the lateral pressure exerted upon the protein, causing variation in protein structure⁵⁴. Insertion of membrane proteins into micelles, which are illustrated in **Figure 1.2 A**, can result in protein instability and as such, screening of detergents is often required to select for conditions promoting protein stability in order to aid crystallisation⁵⁵. Bicelles⁵⁶, shown in **Figure 1.2 B**, and liposomes⁵⁷, shown in **Figure 1.2 C**, were later developed in order to minimise some of these detrimental effects, but these assemblies are not suited for use with X-ray crystallographic methods and can be too large for use in cryo-EM⁵⁸.

Extraction of the membrane protein from its bilayer environment in preparation for solubilisation with these techniques may also result in the loss of those lipids interacting with its transmembrane (TM) regions³⁶ and acting as supports or co-factors. Two examples of such co-factors are phospholipids which are required for the conduction of ions through the KcsA channel⁵⁹⁻⁶⁰ and cholesterol which is an allosteric regulator of the GPCR oxytocin receptor⁶¹. Loss of these important co-factors and the associated disruption of protein structure and function may therefore detrimentally impact upon the results of studies using these techniques.

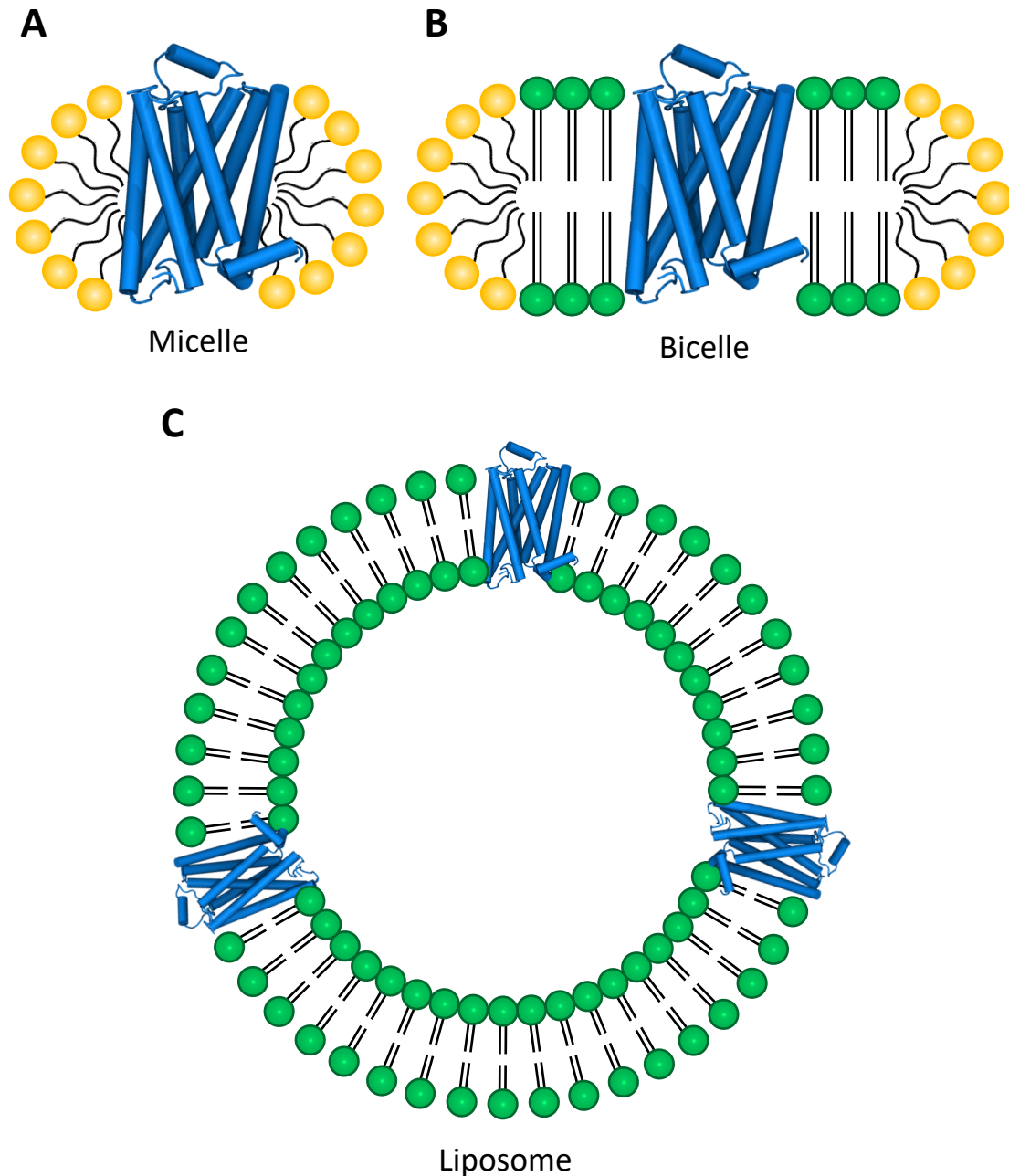


Figure 1.2 – Three techniques for membrane protein solubilisation

Several techniques have been developed for the solubilisation of membrane proteins. Illustrated here are three such techniques; detergent molecules are shown in yellow, lipids in green and membrane proteins in blue.

A: Micelles make use of detergent molecules to shield the hydrophobic parts of the membrane protein from the surrounding aqueous environment.

B: Bicelles are comprised of a central lipid bilayer region surrounded by detergent molecules.

C: Liposomes are relatively large spherical assemblies consisting of phospholipid bilayers into which membrane proteins can be inserted.

For cryo-EM, as is the case for X-ray crystallography, purified membrane protein samples are required, though sample crystallisation is not necessary due to suspension of the sample in a thin layer of vitreous ice. Detergent-free solubilisation techniques can therefore be utilised. These techniques, which are discussed in more detail in **Section 3.1.1**, include amphipols ⁶²⁻⁶³, saposin-lipoprotein nanoparticles ⁶⁴, styrene-maleic acid lipid particles (SMALPS) ⁶⁵ and nanodiscs ⁶⁶⁻⁶⁷, which are the primary environment assessed here.

1.2.4.1 Use of Nanodiscs for Membrane Protein Solubilisation

A nanodisc consists of a discoidal arrangement of lipids arranged into a bilayer configuration, encircled and constrained by two copies of an amphipathic α -helical belt protein called the Membrane Scaffold Protein (MSP), as shown in **Figure 1.3**. The MSPs shield the hydrophobic acyl chains of the lipids from the polar solution which surrounds the nanodisc.

A membrane protein can be embedded into the central patch of lipids in order to study its structure, function and interactions with surrounding system components. The membrane protein is first purified using detergents, which are then removed as lipids are reintroduced to the sample. Several constructs have been produced with MSPs of different lengths, enabling inclusion of varying numbers of lipids ranging from approximately 100 to 600 lipids per nanodisc ⁶⁸. The lipids reintroduced can be varied in order to re-create a native lipid environment or to study key lipids or lipid compositions.

These nanodisc systems are able to self-assemble when the constituent detergent-solubilised components are combined and the detergent is then removed ⁶⁹. Utilisation of nanodiscs in combination with cryo-EM enables structural determination of membrane proteins in a bilayer-like environment, allowing preservation of protein-lipid interactions ⁷⁰.

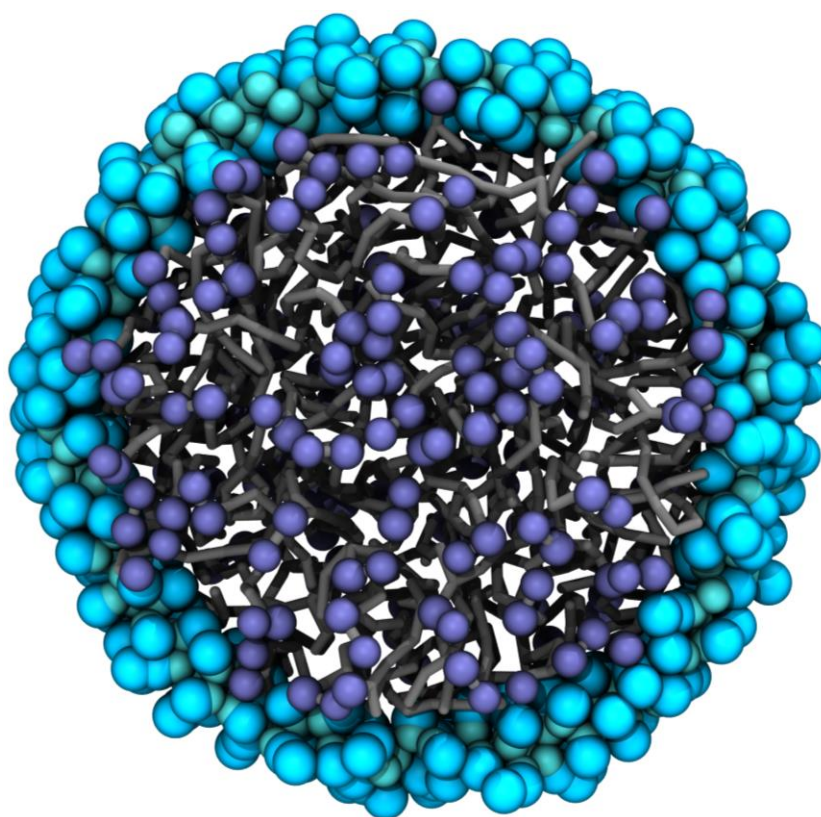


Figure 1.3 – A nanodisc viewed from above the plane of the disc

The structure of a nanodisc is shown from above with the lipid patch encircled by the MSPs. The lipid patch comprises a selection of lipids with head groups in dark blue and tails in grey, arranged in a bilayer-like configuration. The constraining MSPs, shown in cyan, serve to shield the hydrophobic acyl chains from the aqueous environment. A membrane protein can be embedded into the lipid patch of the nanodisc to enable study of its properties, structure, and function.

1.3 Use of Molecular Dynamics to Study Protein-Lipid Interactions

The lipid environment in which membrane proteins exist affects their properties in two main ways: variation in the physical properties of the bilayer, and through molecular interactions between the membrane components ⁷¹. Typically, in cellular membranes, lipids adopt different but coexisting lipid phases: the L_o (lipid ordered) and L_d (lipid disordered) phases ⁷². The different lipid properties of these phases can be observed using fluorescence microscopy and atomic force microscopy, which have been combined to study the L_o and L_d phases and measure differences in membrane thickness, fluidity and lipid order ⁷³. These physical properties of the bilayer can all affect the function of a membrane protein. For example, the mechanosensitive channel of large conductance has demonstrated alterations to ion flux when the lipid acyl chains in its environment are varied in length ⁷⁴. The second way in which the lipid environment can affect the function of a membrane protein is through molecular interactions occurring between the bilayer components; for example, the formation of hydrogen bonds between a specific lipid partner and the membrane protein.

The lipids of mammalian membranes are most commonly phospholipids, of which phosphatidylcholines (PC) are the most abundant, accounting for more than half of the lipids in the majority of eukaryotic membranes. Other lipids of this type differ by their head groups, which include phosphatidic acid (PA), phosphatidylethanolamine (PE), phosphatidylinositol (PI) and phosphatidyl-

serine (PS) ⁷⁵. The acyl chains, which vary in length and degree of unsaturation, comprise the hydrophobic region of the lipid. The rest of the molecule, especially the phosphate head group, is hydrophilic. For convenience the full phospholipid names are usually abbreviated, for example 1-palmitoyl-2-oleoyl-phosphatidylcholine is known as POPC. The structures of the most common phospholipids in mammalian membranes are shown in **Figure 1.4 A**.

Another type of lipid species found in mammalian membranes is the sphingolipid class, characterised by a sphingoid base linked to a fatty acid chain and a polar head group ⁷⁶⁻⁷⁷, for example sphingomyelin as shown in **Figure 1.4 B**. A third class of lipids present in plasma membranes is sterols; in mammalian cells this is most commonly cholesterol ⁷⁸. This is comprised of a non-polar central cyclic region made of four hydrocarbon rings with a hydroxyl group and a hydrocarbon chain as shown in **Figure 1.4 C**. The cyclic region and hydrocarbon chain are non-polar, while the terminal hydroxyl group is polar and can therefore interact with the aqueous environment or polar residues on a membrane protein.

In non-mammalian membranes, a different variety of lipids is most abundant. For example in bacteria, while some of the some lipid species are the same as those in mammalian membranes, they are usually less abundant than other lipid types, such as phosphatidylglycerol (PG) and cardiolipin (CARD) ⁷⁹. Cardiolipin, which is also present in mammals as part of the inner membranes of mitochondria due to their bacterial origin ⁸⁰, has an unusual structure and will be discussed further in **Section 1.3.3.3**.

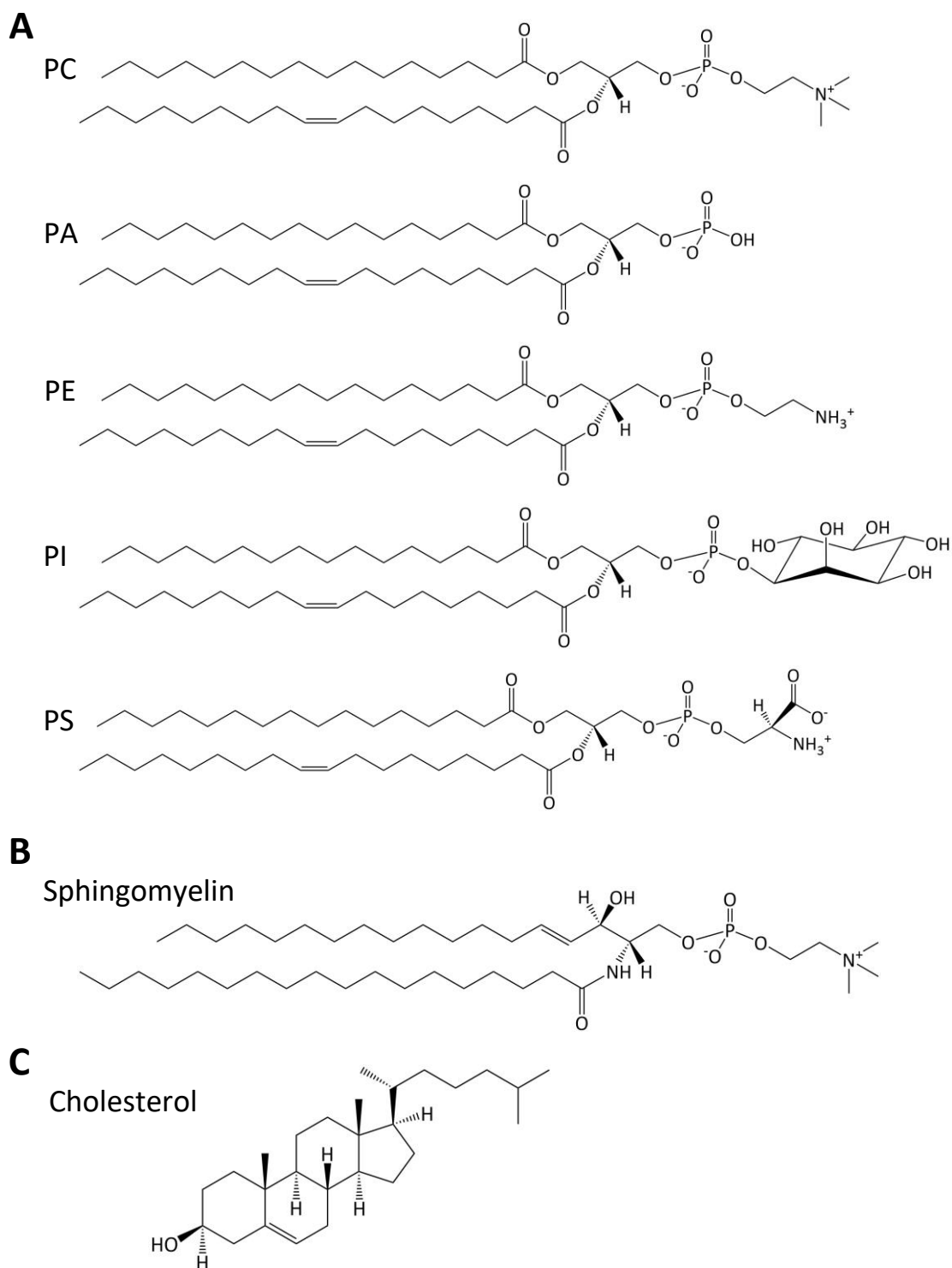


Figure 1.4 – Structures of common mammalian lipids

A: The five most abundant phospholipids in mammalian plasma membranes. The glycerol moiety can be identified at the centre of each structure, with acyl chains on the left and the polar head group on the right.

B: Sphingomyelin, a member of the sphingolipid class of lipids, has a general structure similar to that of the glycerol derived phospholipids.

C: Cholesterol has four hydrocarbon rings comprising a cyclic region in the centre, with a hydroxyl group on the first ring and a hydrocarbon chain on the fourth.

The formation of a lipid bilayer is dependent upon the ability of the constituent lipids to pack together ⁸¹. Many lipid species, such as POPC and POPS, are approximately cylindrical in shape, with head groups and tail regions which occupy an area of similar size. As these lipids aggregate together to shield their hydrophobic tail regions from the aqueous solution, they form planar sheets which make up the bulk of the membrane lipid bilayer. In some lipids however, the tail regions pack less efficiently than the lipid head groups. These conically shaped lipids, such as cardiolipin, result in a negative curvature of the bilayer ⁸². Some other lipid species have tail regions which can pack into a smaller area than that required by the head groups. This inverse conical shape can result in regions of positive curvature in the bilayer and facilitates the formation of micelles ⁸². The degree of lipid tail saturation can also affect the packing of lipids in the bilayer; the more unsaturated the lipid tails, the less efficient the packing of this region is. This results in an inverse conical shape, promoting positive curvature of the bilayer when the tails of the constituent lipids have a higher degree of unsaturation ⁸³.

1.3.1 MD Simulations and the Coarse-Grained Approach

MD simulations are a powerful tool for investigation of the interactions between membrane proteins and their lipid environments ²⁷. The availability of protein structures from sources such as the PDB ⁸⁴, combined with the increasing simulation lengths that can be accessed through recent hardware improvements, allow this approach to investigate more systems over longer periods of time. However, in spite of these advances, the time spans over which

these MD simulations can be performed are still only in the order of microseconds and simulation of larger systems with extended time periods remains a challenge, particularly for simulations in atomistic detail. Utilisation of the coarse-grained (CG) MD method, which was initially developed to study lipids⁸⁵⁻⁸⁶, overcomes some of these issues of restricted system size and simulation length. The most widely applied CG MD approach for the investigation of protein-lipid interactions is the Martini CG force field. Unlike in atomistic simulations where every atom is simulated, in a Martini CG approach groups of heavy (non-hydrogen) atoms in the system are represented by single particles, with one particle representing approximately four heavy atoms, as shown in **Figure 1.5**. This reduction in the number of particles to be simulated for each system reduces the computational demands thereby allowing longer simulations of larger systems than would be possible using atomistic detail.

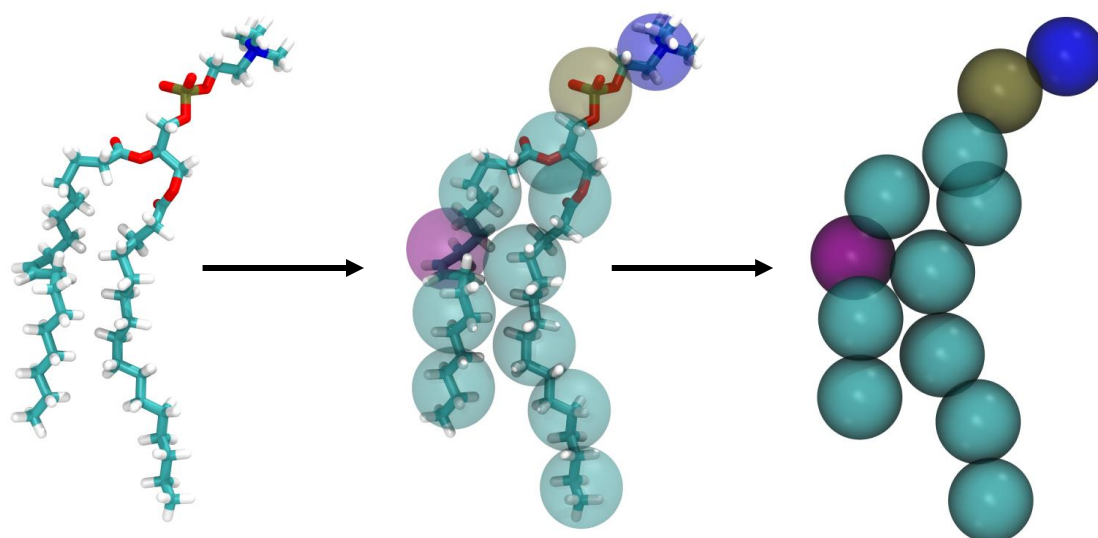


Figure 1.5 – The coarse-graining process of a lipid

A POPC lipid is used here to illustrate the coarse-graining process which is performed for each molecule in the simulation system. The atomistic structure is shown on the left in liquorice representation, with the polar head group at the top, joined to the two acyl chains by the glycerol moiety. This structure undergoes mapping according to the parameterised values of the force field, in this case Martini 2, with the resulting CG particles overlaid on the liquorice structure in the central image. The completed coarse-graining process results in the representation of, on average, four heavy atoms by one CG particle, as shown on the right.

Utilisation of the CG approach does come with the caveat that there is a level of approximation of the biological system which must also be taken into account when considering the interactions observed. However, this approximation allows the CG system to neglect some of the atomistic degrees of freedom⁸⁷, resulting in a smoother energy landscape and allowing more rapid diffusion of particles around the system. This, and the increased simulation lengths available to the CG approach, enable more protein-lipid interactions to occur and be sampled over the course of the simulation.

The application of multiscale MD simulations offers benefits over both CG and atomistic methodologies, by combining the increased simulation speed

accessible to CG MD with the increased biological detail that can be resolved by atomistic simulations. This multiscale approach involves first simulating a system in CG detail, then converting the endpoint into atomistic detail and subsequently running atomistic simulations ⁸⁸⁻⁹⁰.

1.3.2 Non-Specific Lipid Interactions with Membrane Proteins

The lipid components of the bilayer can be categorised depending upon their interactions with the membrane protein ⁷¹. The majority of lipids in a bilayer can be described as acting as a bulk solvent, interacting with the membrane protein in a non-specific manner. These lipids have also been described as boundary or annular lipids, as they form a shell around the membrane protein ⁹¹. The lipids of the annular shell exchange rapidly with the other bulk lipids present in the bilayer, as observed by EPR ⁹². Those lipids which are in closest proximity to the membrane protein experience a reduction in mobility when compared with the non-interacting bulk lipids as can also be observed from nitroxide EPR measurements ⁹³. However, modification of lipid properties is not exclusive to the lipids in the first annular shell; MD simulations have shown that the properties of lipids further from the membrane protein are also altered by its presence ⁹⁴⁻⁹⁵.

The impact of non-specific protein-lipid interactions on protein function reinforces the requirement for membrane protein reconstitution in a bilayer environment, with appropriate properties such as membrane thickness and Area Per Lipid (APL), prior to studies of protein function ⁹⁶⁻⁹⁷. In addition, other physical properties of bilayers have also been demonstrated to affect the

function of membrane proteins, including the fluidity of the bilayer ⁹⁸, bilayer curvature ⁹⁹ and lateral pressure ¹⁰⁰, all of which should be considered when investigating membrane protein function.

1.3.3 Lipid Regulation of Membrane Proteins and Specific Protein-Lipid Interactions

The regulation of membrane protein function and activity by lipids is not limited to the effects arising from the physical properties of the bulk bilayer. Specific binding of particular lipids at definite binding locations on membrane proteins has also been demonstrated to influence the structure of these proteins as well as modulate their activities ¹⁰¹. Lipids which form specific interactions with membrane proteins have been termed co-factor lipids, tending to form longer lasting interactions with membrane proteins and often binding at the intersection of TM helices or between protein subunits. The binding of co-factor lipids has been demonstrated to affect the properties of membrane proteins and in some cases these lipids must be present in order for the protein to function.

This interaction of membrane proteins and lipids, and the resulting regulation of function, has been documented in several protein classes. For example, an anionic phospholipid acts to promote the activity of the bacterial homotetrameric potassium channel KcsA ⁷¹. Indeed, anionic lipids must be present in the environment of the KcsA channel in order to enable protein function, a requirement which has been observed both experimentally ¹⁰² and computationally ¹⁰³. MD simulations are particularly well suited to the investigation of these interactions and have confirmed that the entry of lipids

to orthosteric binding pockets on GPCRs affects protein activity¹⁰⁴. These techniques have also been used to investigate several specific protein-lipid interactions, including those of eukaryotic inward rectifying potassium (Kir) channels, where PIP₂ acts as an agonist¹⁰⁵ and is required for maximum activation¹⁰⁶⁻¹⁰⁷.

1.3.3.1 The Interaction of PIP₂ with GPCRs

Due to the widespread physiological and pharmacological importance of GPCRs, this family has been the subject of significant interest and research into protein regulation, including by protein-lipid interactions. The application of native mass spectrometry to identify endogenous lipids binding to GPCRs has identified preferential binding of PIP₂ when compared to other related membrane lipids. For example, a study published in 2018 found that the complex of the Adenosine A_{2A} Receptor (AA2AR) and its G-protein subunits was stabilised by the binding of PIP₂¹⁰⁸. This study also observed that binding of PIP₂ to the Neurotensin Receptor 1 (NTR1) was responsible for facilitating cognate G-protein coupling. Simulation of AA2AR in a mixed lipid bilayer provided additional information about the role of PIP₂ in stabilisation of the active state of the receptor and in assisting G-protein association¹⁰⁹.

Members of the GPCR family possess several clusters of basic residues at the intracellular rim of the protein which could potentially bind to PIP₂. The binding of PIP₂ to this region has been shown to increase the binding affinity of a GPCR kinase, GRK5, to the β_2 Adrenergic Receptor (ADRB2)¹¹⁰. Studies of AA2AR, which combined large-scale MD simulations with Markov state

models, have also demonstrated that binding of PIP₂ to a TM helix stabilises protein-protein interactions, and the resulting PIP₂ mediated oligomerisation enables the formation of supramolecular signalling complexes ¹¹¹.

1.3.3.2 The Interaction of Cholesterol with GPCRs

Cholesterol is a lipid found in the plasma membranes of eukaryotic cells in large quantities, often comprising 20% to 40% of the bilayer ¹¹². It is responsible for modulating the physical properties of bilayers as well as regulating the activity of membrane proteins. In some cases cholesterol can even be used to anchor proteins to the membrane surface, as in the case of the Hedgehog protein ¹¹³.

The structure of cholesterol is unusual when compared to that of other membrane lipids, as was demonstrated in **Figure 1.4 C**. It is small, with a relatively hydrophobic tail region ¹¹², and its hydroxyl head group is located closer to the centre of the bilayer than that of other longer-tailed lipids. Its rigidity, produced by a planar tetracycle ring system, increases the overall lipid order within the bilayer. Typically, cholesterol forms interactions with saturated lipids in preference to unsaturated lipids, which is thought to result in the formation of microdomains of enriched saturated lipids within the bilayer ¹¹⁴, sometimes termed lipid rafts ¹¹⁵. These regions are thicker than the surrounding bilayer as measured using atomic force microscopy ¹¹⁶ and X-ray diffraction measurements ¹¹⁷. It has been proposed that these lipid rafts may be involved in various processes including cellular signalling and protein trafficking.

Cholesterol is able to affect GPCRs in both specific and non-specific ways. Non-specific effects upon GPCRs result from the changes that the presence of cholesterol has upon the physical properties of the bilayer. These changes to the membrane fluidity, curvature and thickness can result in conformational alteration of the GPCRs. Localised changes to the membrane thickness resulting from the presence of cholesterol can cause recruitment of lipids. This can result in tilting of TM helices causing changes to conformation and oligomerisation states of these proteins ¹¹⁸⁻¹¹⁹.

Cholesterol has also been observed forming specific interactions with membrane proteins. The presence of cholesterol has often been detected during structural determination of GPCRs and the binding of cholesterol has been observed at several different sites on these proteins ¹²⁰. A structure of ADRB2 (PDB ID 3D4S) was solved showing the presence of two bound cholesterol molecules on the intracellular region of the protein at two clefts between the TM helices. The prevalence of bound cholesterol led to the prediction of a cholesterol consensus binding motif ¹²¹. Simulations of ADRB2 in the presence of cholesterol have suggested a mechanism for allosteric regulation of this GPCR ¹²², by limiting the conformational variability of the protein. The cholesterol was observed to wedge between the TM helices resulting in a reduction in conformational dynamics, thereby affecting the coupling of the GPCR to its G-protein binding partner. Several other GPCR structures have been solved with a bound cholesterol at the Helix 8 region of the protein,

typically at the location of covalently attached palmitoyl chains found at the protein's C terminus.

The presence of cholesterol at the binding interface between GPCR dimers has also been observed in structures of ADRB2, where it has been proposed to modulate the dimerisation of the two monomers ¹²³. MD simulations have also implicated the presence of cholesterol in the dimerisation of other GPCRs ¹²⁴.

1.3.3.3 The Interaction of Cardiolipin with Membrane Proteins

The allosteric regulation of membrane proteins by lipids is not exclusive to eukaryotic proteins. Cardiolipin is a structurally unusual lipid component of bacterial and mitochondrial inner membranes, consisting of two phosphatidyl moieties joined together by a linking glycerol group ¹²⁵, as shown in **Figure 1.6**. It can therefore be thought of as a dimeric lipid joined by a glycerol molecule, and as a result of the presence of the four acyl chains, cardiolipin can form enriched domains ¹²⁶.

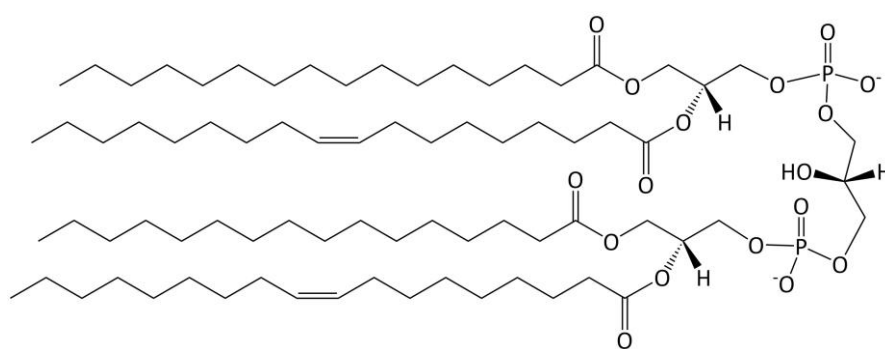


Figure 1.6 – The structure of cardiolipin

Cardiolipin is a structurally unusual anionic dimeric phospholipid. It can be considered as two phosphatidic acid monomers joined at their head groups by a glycerol molecule ¹²⁷.

Bound cardiolipin has been observed in each of the trimers of the *E. coli* ammonia channel, AmtB, and protein-lipid interactions were allosterically modulated by the presence of cardiolipin. This was demonstrated by mutation of a single residue at the cardiolipin binding site, which resulted in abolition of the allosteric effect ¹²⁸, confirming that specific interactions between cardiolipin and membrane proteins can act as allosteric modulators.

Though the components of bacterial membranes are complex, recent advances have enabled the simulation of these systems ¹²⁹. Simulation studies of membranes containing cardiolipin with the mitochondrial Adenine Nucleotide Translocase (ANT) have identified binding sites which were in agreement with experimental results from NMR and crystal structures ¹³⁰. The binding of cardiolipin at these sites was proposed to be involved with mediation of interactions between translocases.

1.4 Nanodisc Simulations

MD simulations can be utilised to provide insight into the structure and arrangement of nanodisc components, improving our understanding of the impact of experimental techniques for structural determination on protein function. Early simulation studies assumed that the arrangement of the MSPs around the nanodisc could be described by a picket fence model. This was suggested based on primary sequence analysis and circular dichroism studies ¹³¹ in addition to early molecular modelling ¹³². The picket fence model, shown in **Figure 1.7 A**, consisted of two MSPs on opposite sides of the nanodisc, with the lipid patch enclosed by short vertical helices, arranged parallel to the membrane normal.

An alternative model, termed the hairpin model, was proposed following insights from Fluorescence Resonance Energy Transfer (FRET) ¹³³. This model proposed that the two MSPs were indeed on opposite sides of the lipid patch of the nanodisc but the helices were perpendicular to the membrane normal as shown in **Figure 1.7 B**. However, the double belt model, **Figure 1.7 C**, is now accepted to be the arrangement adopted by the MSPs ¹³⁴⁻¹³⁶, as has been confirmed by observations of MSP1D1 with DMPC lipids when investigated using NMR spectroscopy, EPR spectroscopy and transmission electron microscopy ¹³⁷.

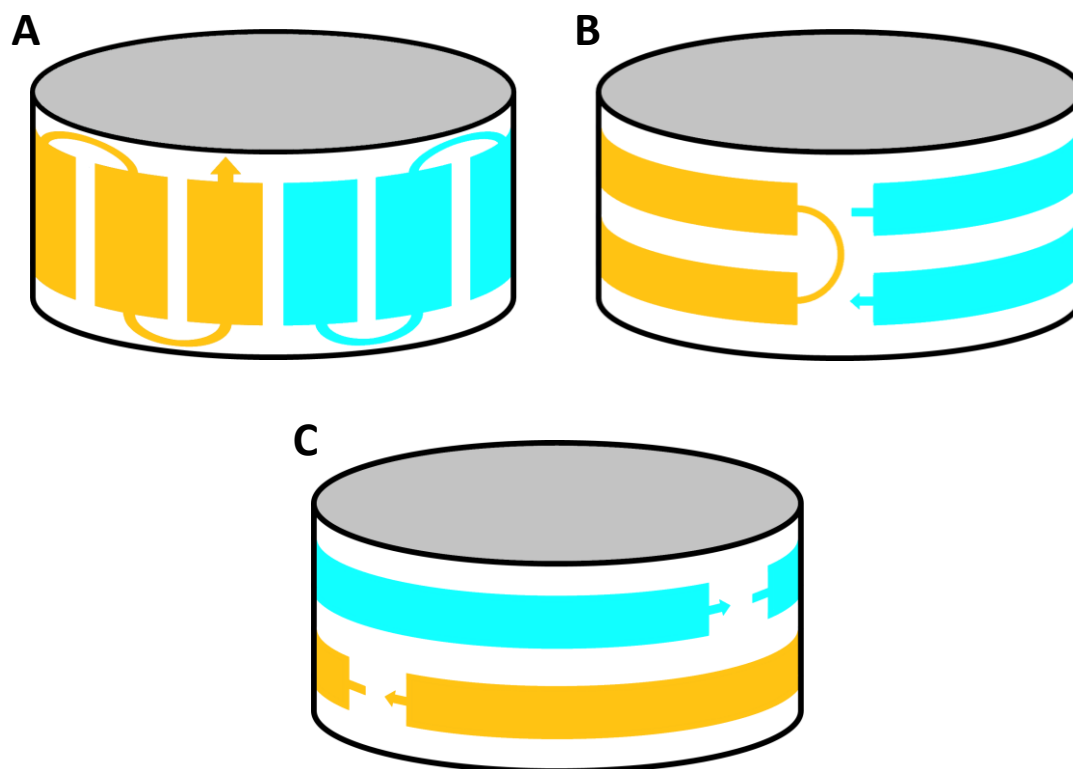


Figure 1.7 – Models for the configuration of MSPs in nanodiscs

A: The picket fence model; the MSPs are at opposite sides of the nanodisc and are arranged into short helices which lie perpendicular to the membrane normal.

B: The hairpin model; like the previous model, the MSPs are on opposite sides of the nanodisc, but have longer helices which lie parallel to the membrane normal.

C: The double belt model; the accepted arrangement of the MSPs in nanodiscs. The MSPs are arranged into long helices which wrap around the entirety of the nanodisc, with one stacked on top of the other and the sequences running parallel or antiparallel to each other.

Figure adapted from ¹³⁸

MD simulations of nanodiscs have been performed using both atomistic ¹³⁹ and CG levels of detail ¹⁴⁰, but simulations performed to date have typically focused on nanodiscs which are comprised only of lipids, without an embedded membrane protein, typically with short length MSPs. MD has also been utilised to aid interpretation of small angle X-ray scattering (SAXS) and small angle neutron scattering (SANS) data ¹⁴¹, which has provided information about the structures and composition of nanodiscs ^{66, 142}.

1.5 Research Overview

The aim of this research was to investigate the interactions which occur between membrane proteins and their surrounding lipids by utilising MD simulations. The simulation studies discussed here aimed to further the understanding of recent advances in membrane protein structural determination, with particular focus on cryo-EM techniques and developments from the application of nanodiscs for the solubilisation of membrane proteins. In addition, the membrane proteins chosen for simulation included medically important and physiologically significant examples. This research has been made possible by recent developments and advances in both software and hardware, which has enabled MD simulations to access larger systems and longer simulation lengths.

In **Chapter 2**, the methodologies for MD simulations are discussed in further detail, with particular focus on the techniques used in this research.

In **Chapter 3**, a multidimensional simulation methodology is presented for the automated simulation of nanodisc systems. This MD approach uses CG and atomistic simulations to enable the automated creation and simulation of customised nanodisc systems, allowing user choice of MSPs and lipid bilayer composition. In order to validate this pipeline approach, comparisons are made between the output of this methodology and experimental results including those from SAXS and SANS experiments. The effect of constraining MSPs on the lipid properties of a nanodisc bilayer is also investigated and comparisons made with the properties of lipid bilayers of the same composition.

In **Chapter 4**, the pipeline approach from **Chapter 3** is extended to enable simulation of nanodiscs which contain membrane proteins. Several membrane proteins with structures solved using nanodiscs are selected for MD simulation in a nanodisc environment, mimicking the experimental conditions used for structural determination. In addition, simulations of the same proteins are conducted in a bilayer environment, allowing comparison of the effects of the two environments on the membrane protein, both in terms of protein structure and with respect to the interactions with the lipid environment.

In **Chapter 5**, protein-lipid interactions are further investigated using a CG approach, with a focus on the GPCR family of membrane proteins. Interactions between these membrane proteins and cholesterol are investigated, and the validity of the simulation approach is supported by successful identification of a lipophilic binding site on the Free Fatty Acid Receptor 1, which had been previously found to bind an allosteric regulator. Finally, cholesterol binding at this site is further investigated with other members of the GPCR family, with comparison of binding specificities and the impact of protein conformational state.

2

Chapter 2 - Theory and Methods

Contents:

2.1 Molecular Dynamics Simulations	32
2.2 Force Fields	34
2.2.1 Bonded Interactions	35
2.2.1.1 Bond Stretching.....	35
2.2.1.2 Angle Bending.....	36
2.2.1.3 Dihedral Angles	36
2.2.2 Non-Bonded Interactions	37
2.2.2.1 Van der Waals Interactions	37
2.2.2.2 Electrostatic Interactions.....	38
2.2.3 Integration Methods.....	39
2.2.4 Energy Minimisation	40
2.2.5 Timesteps.....	41
2.2.5.1 Virtual Sites.....	42
2.2.6 Bond Constraints	42
2.2.7 Periodic Boundary Conditions	43
2.2.8 Temperature and Pressure Coupling	45
2.2.8.1 Thermostats	46
2.2.8.2 Barostats	47
2.2.9 Long-Ranged Interactions.....	48
2.3 Coarse-Grained Simulations.....	50
2.3.1 The Martini Force Field	52
2.3.1.1 Martini Structural Mapping.....	52
2.3.1.2 Martini Particle Interactions.....	54
2.3.1.3 Secondary and Tertiary Structural Elements in Martini.....	56
2.3.1.4 The Martini Water Model.....	57
2.3.2 The Setup of CG Systems	59
2.3.3 Multiscale Modelling	60

2.1 Molecular Dynamics Simulations

The field of MD simulations has advanced considerably since its beginnings in the late 1950s, when simulations of simple gases were first performed ¹⁴³. In the late 1970s the first simulation of a small protein, the bovine pancreatic trypsin inhibitor, was carried out ¹⁴⁴. Establishment of the techniques which formed the foundations for MD simulation of biomolecules was recognised by the Nobel Prize in Chemistry in 2013 ¹⁴⁵⁻¹⁴⁶, and since then it has grown in both popularity and scope. This has been further aided by increased availability of protein structures deposited in databases such as the PDB ¹⁴⁷, which has collated the structures of proteins solved by various techniques including X-ray crystallography and cryo-EM.

Advances in MD techniques have facilitated simulation of biological systems at various resolutions, system sizes and lengths of time, and different simulation methodologies have been developed to investigate particular aspects of these systems ¹⁴⁸. Quantum mechanical models allow the description of chemical bond formation and breaking by explicit consideration of the electrons in the model. These models typically involve finding approximations to the Schrödinger equation, which are necessary because the Schrödinger equation can only be solved for systems containing a single electron, and these models are many-body systems with multiple electrons. However, these quantum mechanical approaches are computationally expensive. This means that many biological phenomena are inaccessible to the timescales and system sizes investigated by quantum mechanical models, though recent

developments have allowed these models to simulate some systems containing thousands of atoms including the binding of ligands to small proteins¹⁴⁹⁻¹⁵⁰.

Biomolecular simulations therefore often use a molecular mechanics model, allowing access to larger system sizes and longer time scales than are currently available with quantum mechanical models. This approach employs classical Newtonian physics and ignores electronic motions; it calculates the energy of the system as a function of the nuclear locations only, the Born-Oppenheimer approximation. This enables simulation of systems containing significantly more particles for greater periods of time. However, as a result of these approximations, details which depend upon the properties of electronic distribution within a molecule, including ionisation and the making and breaking of bonds, are largely inaccessible to these methods.

The underlying theory of MD discussed in this chapter follows that detailed by Leach¹⁵¹. Molecular mechanics utilises simplified interactions between the components of a system, including contributions from the stretching of bonds, opening and closing of bond angles, and rotation around single bonds. In these systems the instantaneous inter-particle interactions are described by a potential energy function, to which molecular mechanics is applied in order to calculate the forces acting on the particles. The acceleration and updated coordinates of the particles are then calculated after a short timestep, δt , which is typically of the order of femtoseconds. This allows the particle coordinates to propagate and when this procedure is iterated millions of times, molecular motions can be observed which can be used for biological inferences.

2.2 Force Fields

A force field is a mathematical expression which describes dependence of the energy of the system on the coordinates of its constituent particles ¹⁵². The description of interactions between system components by the force field contains energetic penalties derived from the deviation of bond lengths and angles from their equilibrium or reference values. It also includes description of energy changes which occur as bonds are rotated, as well as the interactions which occur between the non-bonded components of the system. The force field can be broadly divided into the bonded and non-bonded components:

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

Where U is the potential energy of n number of particles at position r . These components can then be further subdivided into terms which correspond specifically to the energy associated with bond lengths, angles, and dihedral angles, along with non-bonded van der Waals and electrostatic interactions.

To enable reproduction of expected dynamics, the bond lengths, force constant and partial charges within these energy terms undergo parameterisation in order to fit experimental observations and theoretical calculations. This involves comparison with results from NMR, EPR, and crystallographic data, as well as quantum mechanical calculations. Many force fields have now been developed and combined with reliable and user-friendly software to enable MD simulations to be performed, with examples including: CHARMM ¹⁵³, NAMD ¹⁵⁴, Amber ¹⁵⁵ and GROMACS ¹⁵⁶⁻¹⁵⁸.

2.2.1 Bonded Interactions

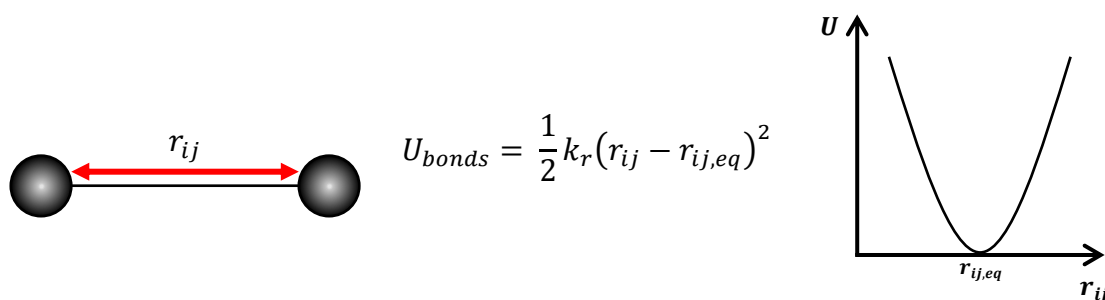
The bonded terms in the force field model the interactions which occur between particles that are covalently bonded. These include terms which describe the stretching of bonds, changes in bond angles, and rotation about bonds, otherwise known as dihedrals. These bonded interactions are summarised in the equation below ^{151, 159}:

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

Where U_{bonded} is the potential energy arising from each of the interaction types: bond stretching (*bonds*), angle bending (*angles*) and rotation about bonds (*dihedrals*). In order to facilitate the understanding of these interactions, the equation above has been broken down into sections as follows.

2.2.1.1 Bond Stretching

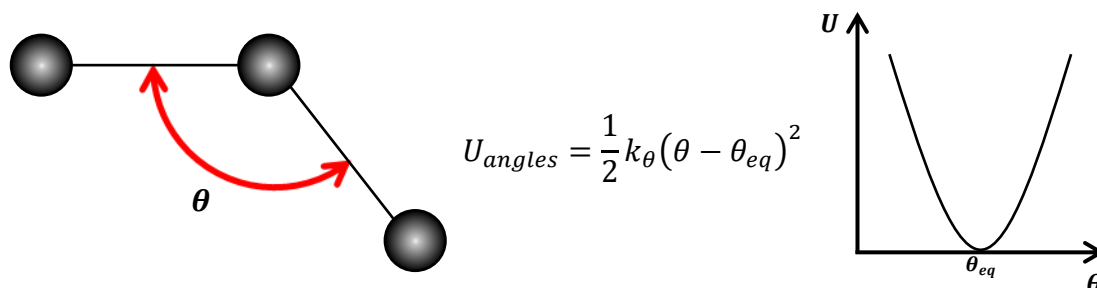
An energy potential derived from Hooke's law is used to describe stretching of the bond which exists between two particles. A harmonic potential gives the energy increase as the length of the bond deviates from the equilibrium length.



Where k_r is a force constant, r_{ij} is the length of the bond between two particles, i and j , and $r_{ij,eq}$ is the equilibrium position. As shown in the graph, this term has a single minimum and deviation from this equilibrium value causes a rapid rise in potential energy.

2.2.1.2 Angle Bending

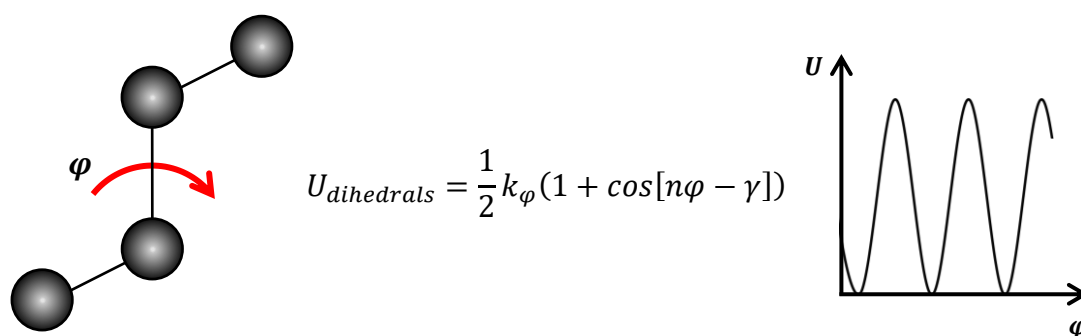
Harmonic potentials are also used to model the energy potential resulting from bending or oscillations of an angle involving three particles.



Where k_{θ} is a force constant, θ is the atomic angle, and θ_{eq} is the equilibrium position. This term also has a single minimum as shown on the graph, with deviation from the equilibrium causing a rapid rise in potential energy.

2.2.1.3 Dihedral Angles

A periodic term allows calculation of the energy potential resulting from dihedral angles, also known as torsions. It describes how energy changes occur as two horizontal planes, containing two particles each, rotate relative to one another around a single bond.



Where k_{φ} is a force constant, n is the number of minima, φ is the dihedral angle and γ is the phase shift. As shown in the graph, this term can have multiple minima separated by positions of higher potential energy.

2.2.2 Non-Bonded Interactions

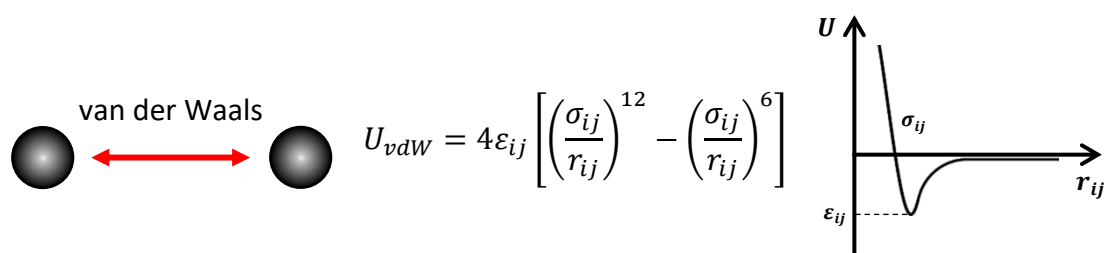
The non-bonded interactions are calculated between all pairs of atoms which are in different molecules, or within the same molecule but separated by at least three bonds. These non-bonded interactions are not dependent upon a specific bonding relationship between the atoms, but can be categorised into van der Waals and electrostatic interactions as summarised in the equation below:

$$U_{non-bonded} = \sum U_{vdW} + \sum U_{electro}$$

Where $U_{non-bonded}$ is the potential energy arising from the two different interaction types: van der Waals forces (*vdW*) and electrostatic forces (*electro*). As before, in order to facilitate the understanding of these interactions, the equation has been broken down into two sections as described below.

2.2.2.1 Van der Waals Interactions

A Lennard-Jones potential is typically used to calculate van der Waals interactions. It is modelled as a 6-12 potential; the terms $\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}$ and $\left(\frac{\sigma_{ij}}{r_{ij}}\right)^6$ model short ranged repulsive and longer ranged attractive forces respectively.



Where i and j are two atoms, r_{ij} is the instantaneous distance between them, σ_{ij} is the collision diameter (the distance at which the potential reaches 0), and $4\epsilon_{ij}$ is related to the Lennard-Jones well depth. As shown in the graph, the minimum potential energy occurs at the Lennard-Jones well, with a rapid rise

in potential energy as r_{ij} decreases beyond this point. If r_{ij} increases there is an initial rise in potential energy, though this tends toward 0 as the particles move further apart and cease to interact.

The values of σ_{ij} and ε_{ij} are calculated from the Lorentz-Berthelot rules as shown below:

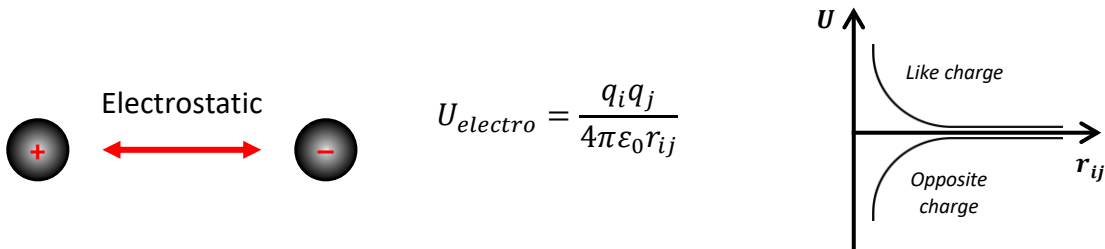
The Lorentz rule: $\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2}$

The Berthelot rule: $\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}$

Where σ_{ii} and σ_{jj} are the collision diameters for atoms i and j respectively, and ε_{ii} and ε_{jj} the well depths of atoms i and j respectively.

2.2.2.2 Electrostatic Interactions

A Coulomb potential is used to calculate the interaction between charged particles, including both ions and particles with partial charges, to represent differential charge distributions due to differences in electronegativity.



Where q_i and q_j are the charges on the respective atoms, r_{ij} is the instantaneous distance between them and ϵ_0 is the dielectric constant representing the permittivity of free space. The graph illustrates that with increasing proximity of two particles of the same charge there is rapidly increasing potential energy, while if the two particles have opposite charges the potential energy rapidly decreases with increasing proximity of the particles.

Coulombic interactions are longer ranged than van der Waals forces, due to a reduced rate of decay with distance (r_{ij}) in $U_{electro}$ when compared with that in U_{vdW} , which is dependent upon $(r_{ij})^6$. As a result of this decay, past a certain cut-off distance the electrostatic interactions are small and can be approximated to 0. However, the treatment of Coulombic interactions is different and will be discussed later, in **Section 2.2.9**.

2.2.3 Integration Methods

The instantaneous interactions of the particles in a system are governed by the force field. Finite difference methods are then used to integrate Newton's equations of motion and produce updated system coordinates dependent upon the prior interactions of the system. The integration is separated into multiple segments which are separated by a small timestep, δt . The force acting on each particle at time t is determined from the vector sum of its interactions with other particles. This is used to calculate the acceleration of the particle at time t , and then to produce updated velocities and accelerations at $t + \delta t$. This assumes that the force acting on the particle is constant over the period of δt , and the value of δt must therefore be appropriate for the system, as will be discussed in **Section 2.2.5**.

The simulations here use the GROMACS implementation of the leap-frog integrator¹⁶⁰, which extends the Verlet method¹⁶¹ for integration of Newton's equations of motion. The integrator alternates between updating the position and velocity of each particle, with positions updated at integer timesteps and velocities at half-integer timesteps, resulting in these sets of results

‘leap-frogging’ over each other. The relationship for updating the positions and velocities of a particle is calculated by the following equations:

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

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

Where r is the position, t is the instantaneous time, δt is the timestep, v is the velocity of the particle, and $a(t)$ is the acceleration of particle i at time t , calculated using:

$$a(t) = \frac{F_i(t)}{m_i}$$

Where F_i is the force acting on the particle and m_i is the mass of the particle.

The utilisation of the leap-frog integrator is therefore dependent upon knowledge of the particle velocities from the preceding step in the simulation, and therefore is not suitable for use during the initial configuration of the system. As such for this initialisation step, velocities are typically taken from a Maxwell-Boltzmann distribution at a specified temperature.

2.2.4 Energy Minimisation

Before an MD simulation is performed, the initial configuration of the system components typically undergoes energy minimisation. This involves iterative steps where the positions of the particles in the system are adjusted in order to reach a local minimum in the potential energy landscape as described by the force field. A robust and easily implemented integrator used for energy

minimisation is steepest descent, which specifies that a system has reached an energy minimum after either a specified number of force evaluations have been calculated, or when the change in system energy is smaller than a specified value. A second integrator, conjugate gradient, is less efficient at the start of energy minimisation but improves as the system energy nears a minimum. However, this technique is incompatible with constraints¹⁵⁸. Reductions in the potential energy landscape result from resolution of steric clashes and incorrect geometries and if this does not occur, it can cause accumulation of excessively high forces, resulting in system instability. A series of energy minimisations are therefore also performed upon conversion of a CG system to atomistic detail.

2.2.5 Timesteps

The selection of an appropriate timestep, δt , for integration of the laws of motion is important in order to ensure the stability of the simulation, and has an impact upon the overall time required for a simulation to be carried out. If the timestep is too large instabilities might arise, but if the timestep is too small the simulation will be inefficient, progress too slowly and not sample the phase space required. Therefore, a timestep which represents a balance between these two outcomes is required. In this research, for atomistic simulations with the CHARMM36m force field¹⁶² a timestep of 2 fs is used, except in cases where virtual sites have been employed. In these cases, an increased timestep of 4 fs is used, for reasons that will be discussed in **Section 2.2.5.1**. For CG simulations using the Martini 2 force field a timestep of between 20 fs and 40 fs is commonly used.

2.2.5.1 Virtual Sites

The size of the timestep is limited by the fastest degrees of freedom in a molecule¹⁶³, which in atomistic simulations are usually hydrogen bond vibrations. With the application of virtual sites to a system, the mass of the hydrogen atom and the forces acting on it are transferred to the heavy atom to which the hydrogen is bonded, and the laws of motion are then calculated on the updated heavy atoms. After the integration step, the positions of the massless hydrogen atoms are calculated using linear or nonlinear combinations of the positions of heavy atoms that are located one or two bonds away from the hydrogen atom.

For some hydrogen atoms, such as those found in methyl and amine groups which are able to rotate rapidly around the heavy atom to which they are bonded, the position of the hydrogen cannot be calculated from the neighbouring heavy atoms. In these cases, two dummy particles are added to which the masses of the group are transferred. This enables calculation of the position of the hydrogen atoms from the positions of the dummy particles. Application of virtual sites and insertion of dummy particles therefore allows the timestep to be increased, as described in the previous section.

2.2.6 Bond Constraints

In MD simulations constraints are often applied to covalent bonds which have high vibrational frequencies, such as those between heavy atoms and hydrogen atoms. The application of a constraint algorithm occurs after the coordinates of the particles in the system have been updated by the integrator.

This allows deviations in covalent bond lengths from their equilibrium values to be corrected. The application of a constraint algorithm, such as the LINear Constraint Solver (LINCS) algorithm used in this research, enables the utilisation of a larger timestep by the integrator ¹⁶⁴⁻¹⁶⁵.

2.2.7 Periodic Boundary Conditions

Simulation systems are composed of a limited number of particles, far fewer than would present in an *in vitro* experiment or on the bilayer of a cell. Due to the comparatively reduced number of particles in a simulation, a much greater proportion of them may be in contact with the edge of the simulation at any one time. The treatment of the particles at the system boundary is therefore crucial for the system to be able to reproduce the properties of a bulk solution, since if the particles were able to interact with the system boundaries it would cause system perturbation and result in non-physiological behaviour.

Periodic boundary conditions enable macroscopic properties of a system to be calculated with a relatively small number of particles, allowing them to experience forces as if they were in bulk fluid. The periodic boundary can be visualised by taking the box in which the simulation system is enclosed and replicating it in all directions, producing a periodic array and resulting in the simulation box being surrounded by 26 of its neighbours. A simplified two-dimensional representation is shown in **Figure 2.1**.

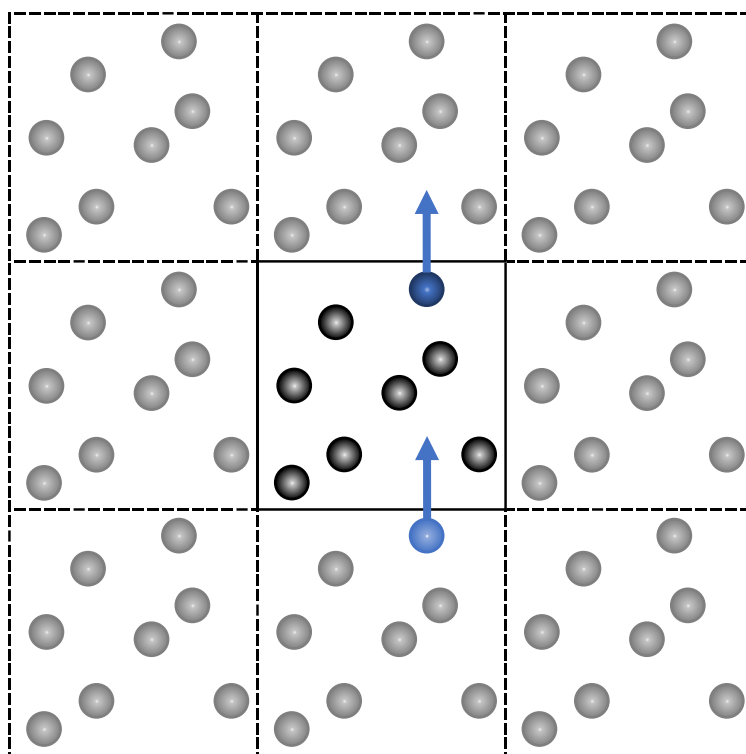


Figure 2.1 – A two-dimensional representation of periodic boundary conditions

The simulation box at the centre is surrounded by eight replications. If a particle diffuses out of the box it reappears on the opposite side of the box. In addition to this, interactions between the particles are calculated across the image boundaries. This abolishes the boundary effects that would otherwise be present if the particle was able to interact with the system edge.

If the system size utilised is too small, larger system components such as molecules and proteins might span the periodic boundary and be able to interact with themselves, resulting in artefacts. The use of a minimum image convention, combined with cut-off schemes, ensures that each particle is only able to interact with other particles which are within a distance of less than half the unit cell length as shown in **Figure 2.2**. In simulation systems where the unit cell is asymmetric, the length of the shortest component of the unit cell is used.

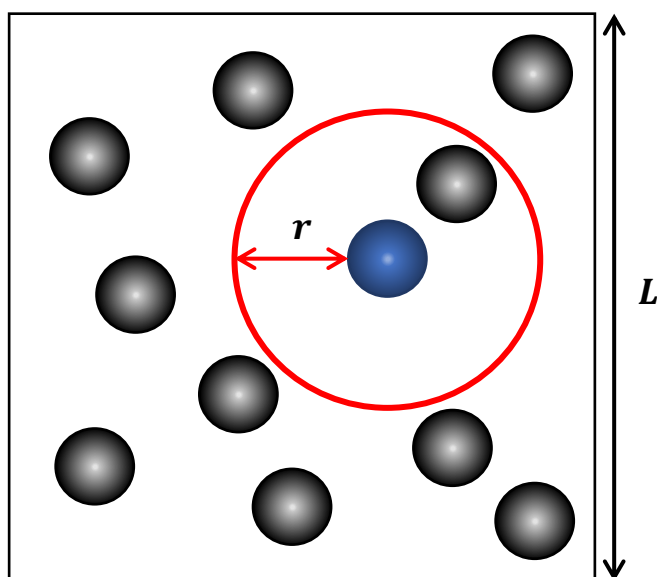


Figure 2.2 – A two-dimensional representation of a cut-off scheme

The blue particle can interact with particles that are within the cut-off distance, r , which is assigned a value of less than half the unit cell, L .

Determination of the correct system size to ensure that system components are unable to self-interact can require experimentation in order to avoid artefacts. Additional drawbacks to the use of periodic boundary conditions include unintentional system ordering as well as influences on the wavelength of fluctuation when membranes are simulated.

2.2.8 Temperature and Pressure Coupling

The procedure as currently described would result in the simulation of an NVE ensemble, where the number of particles (N), system volume (V), and energy (E) remain fixed. However, for biomolecular MD temperature (T) control is typically required, resulting in an NVT or canonical ensemble, enabling the approximate re-creation of the conditions *in vivo* or *in vitro*. In an NVT ensemble the system is not isolated; it is able to exchange energy with a heat bath, though the total energy of the system and the heat bath is constant.

Alternatively, simulations can be conducted in the NPT ensemble, where a barostat maintains the pressure (P) by adjusting the volume of the box. In this research, the simulations in CG detail used the Berendsen thermostat and the Berendsen barostat ¹⁶⁶, while those in atomistic detail used the velocity-rescale thermostat ¹⁶⁷ and the Parrinello-Rahman barostat ¹⁶⁸.

2.2.8.1 Thermostats

The modification of Newtonian MD by the introduction of a thermostat enables the generation of a statistical ensemble at a constant temperature. This aims to reproduce experimental conditions and negate drifts in energy resulting from the accumulation of numerical errors. The kinetic energy of a system is given by the following equation:

$$E_{kinetic} = \frac{1}{2}mv^2 = \frac{3}{2}Nk_B T$$

Where m and v represent the mass and velocity of a particle respectively, N is the number of particles, k_B is the Boltzmann constant, and T is the instantaneous temperature of the system, maintained at a given value by coupling to an external heat bath using thermostat algorithms at a reference temperature, T_{ref} . Deviation of system temperature from T_{ref} is used to scale the particle velocities as follows:

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

Where T is the system temperature at time t , and τ_t is the coupling constant, which controls how closely temperature deviation is coupled to T_{ref} .

This approach is used by the Berendsen thermostat ¹⁶⁶ which was used here for systems simulated in CG detail. This has been extended into the velocity-rescale (or v-rescale) thermostat ¹⁶⁷, which includes a further stochastic term to ensure correct kinetic energy distribution.

2.2.8.2 Barostats

The system pressure can be controlled in a similar manner, by coupling to an external pressure bath as described in the equation below:

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

Where $P(t)$ is the pressure component at time t , τ_p is the coupling constant and P_{ref} is the reference pressure. P_{ref} in all simulations performed here was 1 bar. This is implemented in the Berendsen barostat ¹⁶⁶, which was utilised here for simulations performed in CG detail. The Berendsen barostat rescales the simulation box shape and volume, and can be applied isotropically, as for the nanodisc systems here, semi-isotropically, as for the bilayer systems here, or anisotropically. Implementation of the barostat in an isotropic manner results in the x , y and z box vectors being rescaled uniformly. In contrast, in an anisotropic barostat application, the x , y and z box vectors undergo rescaling independently from each other. In semi-isotropic barostat applications, the x and y box vectors undergo rescaling independently to the z box vector, a technique which is commonly used for simulations of bilayers due to associated improvements in system stability. The Parrinello-Rahman barostat ¹⁶⁸ is an updated barostat which enables the use of cells of arbitrary shapes and allows

variation in simulation box shape. This barostat was used with a cuboid or cubic simulation box for all simulations presented here.

2.2.9 Long-Ranged Interactions

The treatment of the non-bonded interactions in a simulation system has the potential to be the most computationally intensive operation to calculate because, as shown in the equations in **Section 2.2.2**, each particle in the system interacts with every other particle. This means that the number of interactions made in the system increases as the square of the number of particles present.

In order to reduce the computational load, an assumption is made which applies a cut-off distance for the treatment of the van der Waals energy terms. The Lennard-Jones equation, which describes the van der Waals energy terms, decays with a distance dependence of r^{-6} . This results in the potential energy rapidly decaying and tending to 0 as the distance between the particles increases. A cut-off scheme is used to calculate the van der Waals interactions of only those particles within a given proximity, and past this distance the Lennard-Jones energies are assumed to be 0. Several cut-off schemes have been developed including those which make use of a ‘hard’ cut-off, which abruptly terminates the interactions at the specified cut-off distance, and ‘soft’ cut-offs which gradually reduce the size of the interactions as the distance approaches the specified cut-off distance. The tracking of which particles are within the cut-off distance is carried out by a ‘neighbour list’ which is regularly updated as the positions of the particles change.

In comparison to van der Waals forces, Coulombic interactions are longer ranged. This results from the energy of interaction decaying with a distance dependence of r^{-1} , meaning that a cut-off scheme is inappropriate. Instead, the Particle Mesh Ewald (PME) method ¹⁶⁹⁻¹⁷⁰ is typically used in atomistic simulations for the treatment of the Coulombic interactions. In this scheme the interactions are grouped into short and long-ranged potentials, which are calculated in real space and Fourier space respectively. This is done to allow rapid convergence. For simulations in CG detail the reaction field scheme is used to calculate the effects of long ranged interactions. In this scheme the interactions within a certain distance from the particle are calculated as specific Coulombic interactions, while those outside this region are treated as a single uniform dielectric continuum ¹⁷¹⁻¹⁷².

2.3 Coarse-Grained Simulations

The approaches outlined so far for the simulation of biochemical systems have focused mainly on an atomistic description of the system. However, this is not the only level of detail at which simulations are performed. Instead, several different approaches have been developed to allow the study of biological systems in various levels of detail. The resolution used can affect the length of time for which a system can be simulated, as well as the number of particles which may be included. The degree of system abstraction can thus be matched to the biological question being addressed, in order to enable events occurring over larger time scales or involving very large number of particles to be investigated, while balancing computational demands due to system complexity.

One such level of system abstraction is the utilisation of a Newtonian approach in place of a quantum mechanical one, such as is applied in atomistic MD simulations. Due to the increased computational load associated with simulation of quantum mechanical models, typical atomistic forcefields make use of a fixed-charge model which does not permit changes in polarisability, though some polarisable models have been developed, such as the AMOEBA forcefield¹⁷³. However, the complexity of the simulation system can be further reduced to enable longer and larger simulations. Typically, the simulation time and system size accessible to a particular approach scale with the amount of system abstraction, as demonstrated in *Figure 2.3*.

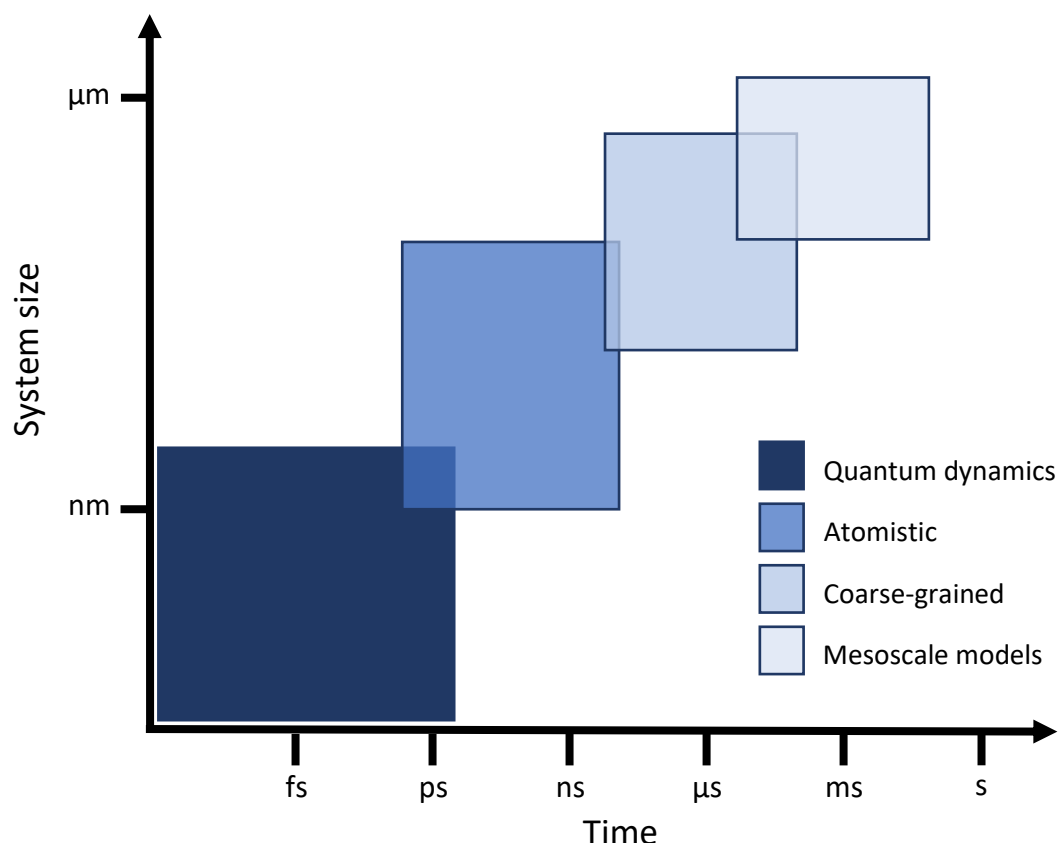


Figure 2.3 – System size and length for four different simulation methodologies

A visual representation of the simulation times and system sizes accessible to four different approaches: quantum dynamics, atomistic, coarse-grained and mesoscale simulation models.

Figure adapted from ¹⁷⁴

Choice of the level of system abstraction is influenced by the nature of the biological question which is being investigated. Quantum mechanical approaches are well suited to the study of processes such as enzymatic catalysis ¹⁷⁵ and the protonation states of lipids ¹⁷⁶. Atomistic simulation approaches enable the study of events which involve larger numbers of atoms over longer time spans, such as the selectivity filter of a channel ¹⁷⁷. Coarse-graining describes the process of approximating groups atoms into a single CG particle. Several CG approaches have been developed including the

SIRAH force field ¹⁷⁸ and the Martini force field ^{85, 87}, which was utilised in this research and will be discussed in more detail in **Section 2.3.1**. Further levels of system abstraction, such as mesoscale modelling ¹⁷⁹ and the ESPResSo package ¹⁸⁰, involve using several particles to represent an entire lipid, in order to study larger scale processes such as membrane curvature and protein diffusion ¹⁸¹.

2.3.1 The Martini Force Field

The Martini force field was originally developed for the simulation of lipids in 2004, by Marrink and co-workers ¹⁸². The force field received the moniker of Martini in the 2007 2.0 update ¹⁸³, with a subsequent further update to version 2.2 ¹⁸⁴. The aim of the Martini force field is to provide the computational advantages of the CG approach in a manner which does not require re-parameterisation in order to study higher-order macromolecular scales, through the use of well-parameterised CG elements. The Martini 3 update ¹⁸⁵, which was published during the late stages of this research, aimed to improve interaction balance by representing specific interactions such as hydrogen bonding and polarisability. However, due to the absence of required parameters, such as those for cholesterol, this update was not appropriate for use in the majority of the research presented here.

2.3.1.1 Martini Structural Mapping

The mapping scheme utilised by the Martini 2.2 force field represents groups of typically four heavy atoms with a single CG particle. Four main types of particles are employed: polar (P), apolar (C), non-polar (N), and charged (Q).

These groups of particles can be further subdivided, as indicated by the insertion of additional characters after the P, C, N or Q. Furthermore, S-type particles have been added so as to better represent the geometries found in some biological molecules such as ring structures, sterols and nucleic acids. These employ higher resolution mapping, where two or three heavy atoms rather than four are represented by one CG particle.

The C and P particles are subdivided according to polarity, as denoted by the addition of a numeral after the particle name: 1 for low polarity, ranging to 5 for highest polarity. The hydrogen bonding ability of N and Q particles is denoted by a letter: donor (d), acceptor (a), donor/acceptor (da), with the exception of those which neither donate nor accept (0).

The residues of a protein are assigned a single backbone particle which represents the amino acid backbone region, with additional sidechain particles added for larger amino acids. The particle type assigned to a residue is dependent upon the local secondary structure of a particle. Typically, the polarity of a particle is reduced when the residue is involved in secondary structures, representing the reduction in hydrogen bonding capacity when compared to particles in a random coil arrangement. The side chain mapping of the 20 naturally occurring amino acids is shown in **Figure 2.4**.

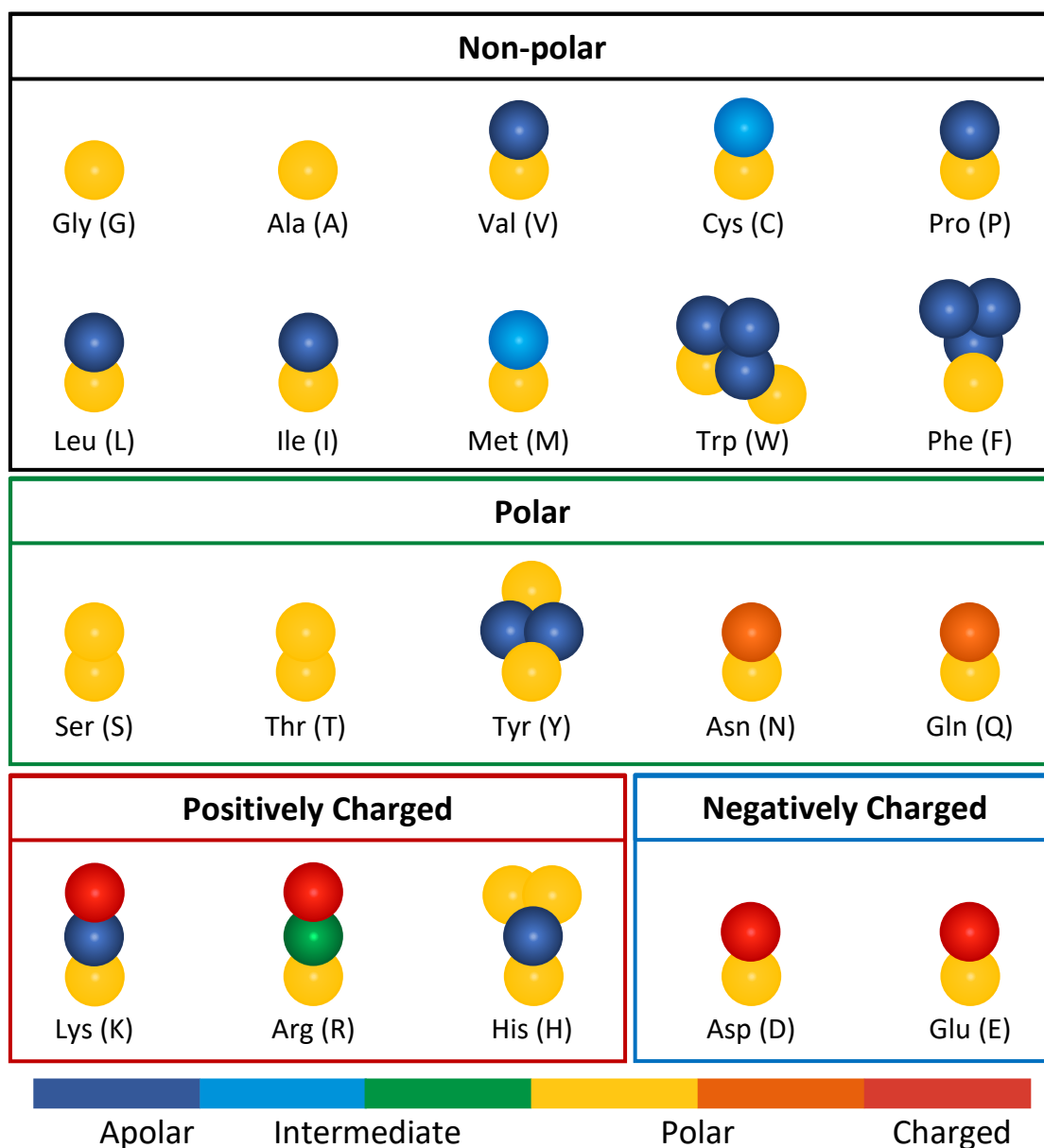


Figure 2.4 – Martini 2 mapping of amino acids

A visual representation of the Martini mapping scheme for the 20 naturally occurring amino acids. The side chains are coloured according to particle subtype, as shown in the colour scale. For ease of reference the amino acids have been displayed in groups according to their properties.

Figure adapted from ¹⁸⁶

2.3.1.2 Martini Particle Interactions

Harmonic potentials are used to mimic the bonds between Martini particles, as was mentioned **Section 2.2.1**. The angles and dihedrals between particles are described using harmonic and periodic potentials respectively. The

non-bonded interactions are primarily modelled by the application of a Lennard-Jones 6-12 potential, and the strength of the interaction undergoes modification by differential well-depth (ϵ) values for each of the pairwise interaction¹⁸³. The value of the σ parameter, which denotes the distance at which the energy is 0, is assigned the value of 0.47 nm for all particles, with the exception of the S-type particles and the interaction of Q type particles with C1 and C2 particles, for which the values of σ are set at 0.43 nm and 0.47 nm respectively. As S-type particles represent a smaller number of atoms, the value of ϵ was scaled to 75% of the corresponding larger particles. A Lennard-Jones potential was used to describe the partial charge interactions and a Coulomb potential was used for the ionic interactions. The Lennard-Jones potentials use a shifted cut-off scheme, which reduces the partial charge interactions to 0 at distance of between 0.9 nm and 1.2 nm. The Coulombic interactions are assigned a value of 0 at a distance of 1.2 nm in order to avoid introduction of noise resulting from energy discontinuity.

The Martini CG force field was parametrised using comparison with experimental data and atomistic simulations. The bond lengths and bond angles within proteins were obtained from a survey of over 2000 protein entries deposited in the PDB¹⁸⁷. The experimental partitioning between aqueous solutions and organic phases of small molecule analogues of the side chains of amino acids was also used for CG parameterisation. The CG values were optimised to within 2 $k_B T$ for all compounds, and to within 1 $k_B T$ for many. The free energy for the partitioning of water in hexane was calculated and

compared to experimental estimates; simulations produced a value of $24.2 \pm 2 \text{ kJ mol}^{-1}$ compared with the experimental estimate of 25 kJ mol^{-1} . Other properties of biological systems have also been used for validation of the CG approach including the APL of lipids assembled into bilayers. The Martini CG force field has been utilised to predict lipid binding sites on membrane proteins which have been in agreement with structural results from techniques including X-ray crystallography and NMR. A more detailed description of the parameterisation protocols utilised for the Martini force field can be found in the original publications ^{87, 182} and a perspective from the developers of the Martini force fields ⁸⁵.

2.3.1.3 Secondary and Tertiary Structural Elements in Martini

The coarse-graining process removes the hydrogen atoms from the structures which are simulated. As this would have an impact on the secondary and tertiary structure of proteins, which often rely on hydrogen bonding, additional considerations are made by the Martini force field. A structure 'type', is assigned to each particle, using the DSSP program ¹⁸⁸. Following this, the secondary structural dynamics of a particle are restrained around an equilibrium based upon the structure type, using harmonic and periodic potentials.

The typical treatment of tertiary structure by Martini involves the application of an Elastic Network Model (ENM) to the protein. Two such ENMs are the Martini ENM and the ElNeDyn ENM ¹⁸⁹. Both apply harmonic potentials between neighbouring backbone particles within a set distance. This

results in a web of bonds which prevent large-scale disruption of the tertiary structure during the simulation, while enabling small-scale changes in backbone dynamics. The side chain particles are not typically restrained.

In the default Martini ENM, harmonic bonds are applied to backbone particles within 0.5 nm - 0.9 nm of each other, with a force constant of 500 kJ mol⁻¹ nm⁻². The ElNeDyn ENM uses a cut off of 0.9 nm with the same force constant applied by default. The two ENMs differ in their treatment of backbone particles; in the former the backbone particles are positioned at the centre of mass of the heavy atoms in the backbone of each residue, whereas in the latter the backbone particles are positioned at the C α of each residue. Due to this, the ElNeDyn ENM model requires adjustment of the non-bonded parameters and underlying mapping, as well as the usual application of harmonic bonds between backbone particles.

The cut-offs and force constants used for the ENMs are designed to be highly tuneable variables and are ideally parameterised *de novo* for each different protein by making comparison with atomistic simulations. However, the default values for the parameters are widely used and in the case of the ElNeDyn ENM have demonstrated good performance across a range of scales and protein structures¹⁸⁹. In the simulations performed here, the Martini ENM has been used unless otherwise specified.

2.3.1.4 The Martini Water Model

In the standard Martini model, each group of four water molecules is treated as a single Martini water particle. The Martini waters use a P4 particle type, and

as a result are not influenced by the electric fields of charged particles. Therefore, like many atomistic water models, the Martini water particles are unable to reproduce polarisation effects.

In order to address this deficiency, a polarisable water model that is compatible with Martini was developed by the introduction of two additional particles types: WP and WM, bearing a charge of $+q$ and $-q$ respectively. Harmonic bonds are used to attach these particles and their orientation is fixed using a harmonic angle potential. The WP and WM particles were parameterised to only interact via Coulombic interactions, and they therefore only interact with charged (Q) type particles. This polarisable water model has been shown to model the dielectric screening properties of bulk water and its freezing point better than alternative approaches. It has also been used to improve modelling of protein-lipid interactions involving glycolipids ¹⁹⁰.

A drawback to the polarisable Martini water model is the uncertainty surrounding its application to existing Martini parameters which were developed in systems using the standard Martini water model. In addition to this, there is uncertainty about the interaction between the polarisable Martini water model and non-polarisable system components, such as lipids and proteins. The addition of more particles to a simulation in this way also partially negates the advantages of the use of a CG simulation approach; for systems containing only polarisable Martini water particles, the speed of the simulation is reduced by a factor of three when compared with the standard Martini water model.

An alternative approach developed for simulation of water particles in CG is the implicit water model, termed 'Dry Martini' ¹⁹¹. The water particles are replaced by a continuous medium which mimics the properties of the solvent. In larger systems, such as the envelopes of virions, a greater proportion of the system is composed of water and the use of an implicit solvent dramatically increases the simulation speed. The 'Dry Martini' approach therefore aims to enable the simulation of these larger systems, though omission of the solvent results in a more abstracted system unable to model protein behaviour as well as the standard Martini water model. It also raises the same parameterisation and compatibility concerns as were applicable to the polarisable model.

For the simulation of CG ions, Martini makes use of a single particle to describe the ion and four water molecules, resulting in a 'hydrated ion'. In the research presented here the standard Martini water model was used for all CG simulations.

2.3.2 The Setup of CG Systems

A python script, `martinize.py` ^{184, 192}, uses coordinates of an atomistic protein structure to produce Martini protein structure and topology files to which further system components can be added prior to CG simulation. During this process `martinize.py` uses information about the secondary structure of the protein to aid assignment of the CG particles. It thereby enables conversion of atomistic protein coordinates into a CG representation, around which a lipid bilayer can be assembled. This process is illustrated for a GPCR, ADRB2, in

Figure 2.5.

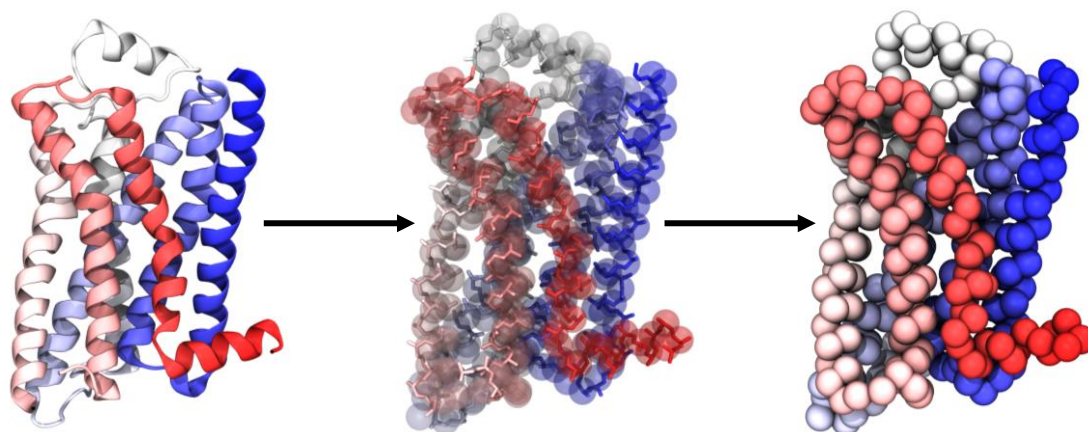


Figure 2.5 – The coarse-graining process of a membrane protein

The atomistic coordinates of a protein, in this case the GPCR ADRB2, are shown on the left in a helical representation, coloured from the C-terminal residue in dark blue to the N-terminal residue in bright red. CG mapping according to the parameterised values of the force field, in this case Martini 2, is illustrated in the central image. The protein backbone is shown in licorice representation and is overlaid by the CG particles, each of which represents on average four heavy atoms. The completed process is illustrated by the CG protein on the right, coloured from blue to red as above.

Following CG conversion of the system components as required, several methodologies are available for the production of CG simulation systems containing membrane proteins in bilayers. These include the self-assembly of lipid bilayers around a membrane protein^{10, 193}, the online webserver CHARMM-GUI¹⁹⁴ and the approach utilised in this research: the *insane.py* script¹⁹⁵. This script inserts desired lipids in a grid like arrangement around a membrane protein and a subsequent short simulation is carried out to allow equilibration of the bilayer.

2.3.3 Multiscale Modelling

Some of the limitations of the CG approach can be mitigated by the use of multiscale modelling, whereby the simulation system undergoes conversion from one level of abstraction to another. The procedure utilised in this research

made use of CG simulations which were followed by conversion to atomistic detail and then further simulation. The conversion from CG to atomistic detail was performed using the CG2AT2 program⁹⁰, as shown in **Figure 2.6**, which uses a fragment-based approach for the conversion of a CG system into atomistic detail. A database of atomistic fragments which correspond to the CG particles present in the system is used to produce an atomistic output from the supplied CG input. This protocol is applicable to proteins, lipids, ions and solvent. Once the system is converted it undergoes minimisation and then equilibration. One of the main advantages of this approach is that it enables the preservation of protein-lipid interactions.

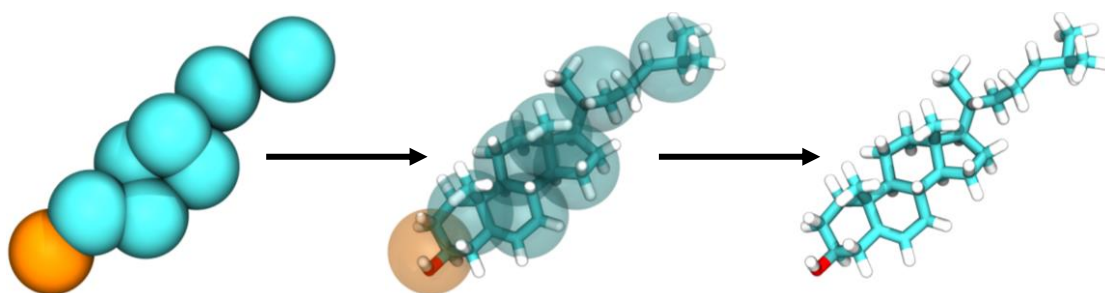


Figure 2.6 – The conversion of CG cholesterol to atomistic detail

The CG particles of cholesterol are shown on the left, with orange representing the hydroxyl head group and the rest of the molecule shown in cyan. The CG2AT2 program⁹⁰ is applied and the fragment-based approach enables conversion of the CG cholesterol to atomistic detail. The atomistic cholesterol is shown on the right in liquorice representation, with the oxygen of the hydroxyl head group shown in red, and the cyclic and hydrocarbon chain regions of the molecule shown in cyan, with hydrogen atoms in white.

3

Chapter 3 - An Automated Nanodisc Pipeline Methodology

Contents:

3.1 Introduction	63
3.1.1 Solubilisation Techniques	63
3.1.2 Nanodiscs Overview	66
3.1.3 Status of Simulation Studies of Nanodiscs	68
3.1.4 Pipeline Approach.....	71
3.2 Methods	72
3.2.1 Coarse-Grained Simulations	72
3.2.2 Atomistic Simulations.....	74
3.2.3 Analysis.....	75
3.2.3.1 SAXS Analysis.....	75
3.2.3.2 Lipid Analysis	76
3.3 Results	78
3.3.1 Alternative Nanodisc Simulation Methodologies	78
3.3.2 A Multiscale Pipeline	80
3.3.3 Elastic Network Models	83
3.3.3.1 Effects of ENM choice on CG Systems: Comparison with SAXS Data.....	86
3.3.3.2 Effects of ENM Choice on Atomistic Systems: Comparison with SAXS Data	88
3.3.4 Analysis of Pipeline Output in Atomistic Detail	89
3.3.4.1 Root Mean Square Fluctuation	90
3.3.4.2 Axial Ratio	92
3.3.4.3 Area Per Lipid	93
3.3.4.4 Order Parameters.....	95
3.4 Conclusions	98
3.4.1 ENM Use in Nanodisc Simulations	99
3.4.2 Lipid Properties	101

3.1 Introduction

3.1.1 Solubilisation Techniques

Structural determination of membrane proteins using techniques such as X-ray crystallography and cryo-EM can be challenging for a variety of reasons. Primarily, these difficulties arise from the hydrophobic properties of the TM region of the protein commonly necessitating the use of detergents for isolation from the bilayer environment, as well as the tendency for integral membrane proteins to be more flexible and unstable than soluble proteins¹⁹⁶. The application of detergents, as was described in **Section 1.2.4**, promotes protein solubilisation by encasing the hydrophobic regions of the protein. Following initial solubilisation, the detergent, for example dodecyl maltoside¹⁹⁷, can be exchanged for a different species before structural studies are performed. As previously discussed, there are drawbacks to using these methodologies involving detergents and as such, several other techniques have since been developed for protein solubilisation.

The use of amphipathic polymers, amphipols⁶³, is one such technique. The constituent subunits form strong interactions with the TM regions of the membrane protein⁶², solubilising the protein by encircling it and shielding the hydrophobic regions from the solution, as shown in **Figure 3.1 A**. In addition, amphipols restrict conformational flexibility and stabilise protein structure⁵⁸.

All of the previously described techniques for structural determination of membrane proteins necessitate removal of the protein from the bilayer environment, causing potential disruptions in conformation, function and

protein-lipid interactions. The use of SMALPs, shown in **Figure 3.1 B**, avoids this requirement. Amphipathic styrene-maleic acid copolymers are used to directly extract membrane proteins from a bilayer in a detergent independent manner, resulting in membrane proteins embedded in discs of approximately 10 nm in diameter ¹⁹⁸. SMALPs have been used to determine the structure of the KimA potassium transporter at a resolution of 3.7 Å ¹⁹⁹. Another recent development is that of saposin-lipoprotein nanoparticles ⁶⁴, which, despite requiring initial extraction of the proteins from the bilayer using detergents, reconstitute them in a lipid environment for further study. These nanoparticles comprise a number of saposin proteins which constrain a lipid patch into which a membrane protein may be embedded, as shown in **Figure 3.1 C**. Despite their theoretical advantages, both SMALPs and saposin-lipoprotein nanoparticles have seen limited adoption for structural determination and will not be discussed further here.

The development of nanodiscs is a similar technique, in which a membrane protein is first isolated using detergents and then reconstituted in a bilayer-like environment in order to minimise disruptions to protein structure and function. In contrast to saposin-lipoprotein nanoparticles, nanodiscs employ two copies of an MSP derived from apolipoprotein A1 to constrain the lipid patch, arranged in either a parallel or an antiparallel configuration. The MSPs here are in an antiparallel arrangement, as shown in **Figure 3.1 D**.

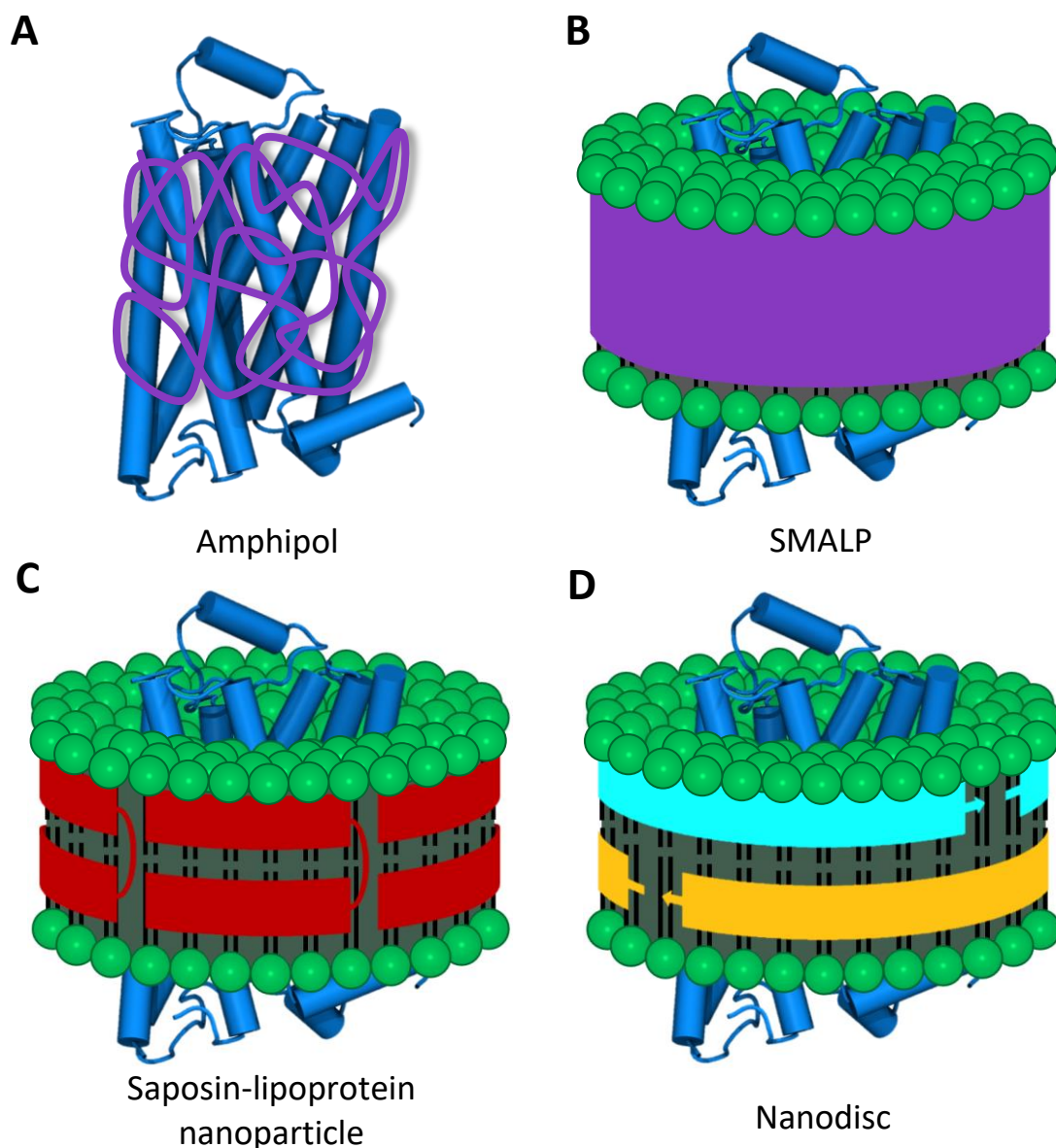


Figure 3.1 – Further techniques for solubilisation of membrane proteins

Membrane proteins (blue) may be reconstituted into bilayer-like environments with lipids (green).

A: Amphipols are amphipathic polymers which adsorb onto the TM region of the protein, shielding the hydrophobic residues from the aqueous environment.

B: SMALPs comprise a patch of lipid bilayer constrained by amphipathic styrene maleic acid co-polymers. The membrane protein can be extracted directly from the physiological bilayer, without the requirement for detergents.

C: Saposin-lipoprotein nanoparticles use saposin proteins to constrain a patch of lipids, into which a membrane protein can be embedded following isolation from the physiological bilayer using detergents.

D: Nanodiscs also contain a patch of lipids in a bilayer-like arrangement, into which a membrane protein can be embedded. However, instead of using saposins, the lipid patch is constrained by two copies of an amphipathic MSP.

Nanodiscs have been widely used and in particular, their use for solubilisation of proteins for structural determination with cryo-EM has produced numerous structures at near atomic resolutions, including over 40 protein structures with a resolution of 3 Å to 4.7 Å. Examples of membrane proteins determined using cryo-EM include TRP channels²⁰⁰ and GPCRs²⁰¹. Another structural determination method which has been used in combination with nanodiscs to produce high resolution atomic structures of membrane proteins is NMR²⁰². However, there is an upper limit to the size of protein which can be structurally determined using NMR, often excluding large membrane proteins such as TRP channels²⁰³. In addition, NMR studies often involve membrane protein solubilisation in ‘harsh’ detergents, such as dodecylphosphorylcholine, which have a higher propensity to denature membrane protein structures than ‘soft’ detergents²⁰⁴. Cryo-EM is therefore used more frequently, often producing protein structures of a lower resolution but with a wider scope and reduced risk of disparity with the physiological state of the protein than is suffered by detergent-only methods.

3.1.2 Nanodiscs Overview

A nanodisc comprises a discoidal arrangement of lipids at the centre, surrounded and constrained by two copies of an amphipathic α -helical belt protein, the MSPs, as shown in **Figure 3.2**. Solubilisation and combination of the constituent parts, followed by removal of the detergent, results in self-assembly of these nanodisc systems⁶⁹. In addition, a membrane protein can be embedded into the central lipid patch if desired.

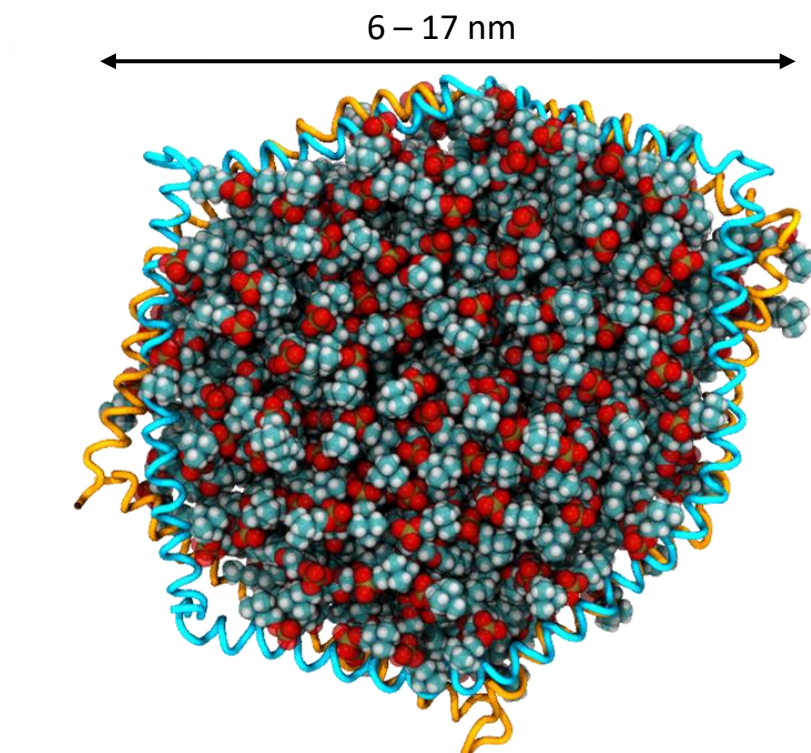


Figure 3.2 – An atomistic representation of a nanodisc

A nanodisc consists of a discoidal arrangement of lipids and proteins. A central patch of lipid bilayer is surrounded by two copies of an amphipathic MSP, shown in orange and cyan. The length of the MSPs can be varied by insertion or deletion of helices in order to constrain a range of different numbers of lipids, thereby allowing a variety of sizes of nanodisc to be created. Membrane proteins can be incorporated into the lipid patch enabling their solubilisation.

The MSPs shield the hydrophobic acyl chains of the phospholipids from the aqueous solution, allowing them to adopt a bilayer conformation. The MSPs are typically derived from human apolipoprotein A-1 following removal of the globular N-terminal domain, though other proteins have been used to produce nanodisc-like structures using proteins based upon ApoE3 ²⁰⁵. The diameter of the nanodisc can be varied by using different MSP constructs, produced by insertion or deletion of helices in order to change the length of the MSPs ⁶⁷. Commonly used diameters vary from 10 nm to 17 nm ⁶⁸, and different diameter nanodiscs allow different sized membrane proteins to be embedded ⁷⁰.

Nanodiscs are increasingly used for the solubilisation of membrane proteins for structural determination. However, their use is not restricted to solubilisation and they have also been utilised for many other purposes, including for characterisation of membrane protein assemblies²⁰⁶ and for studies of structural transitions of α -Synuclein in nanodiscs containing a variety of lipids²⁰⁷.

It has previously been established that the structure and function of membrane proteins can undergo modulation by the lipids present in the membrane environment²⁰⁸⁻²⁰⁹. Lipids have also been shown to bind to specific sites on membrane proteins, such as have been demonstrated with transporters¹⁸⁻¹⁹ and channels²¹⁰. However, it is currently unknown if the properties of small enclosed patches of bilayer, such as those present within a nanodisc, differ compared to those of bulk bilayers, and in addition whether there are differences in specific lipid interactions and binding sites for membrane proteins encapsulated in such a system.

3.1.3 Status of Simulation Studies of Nanodiscs

MD simulations are able to provide valuable insights into the system properties of nanodiscs, in particular the effects of the constraining MSPs on the membrane protein and the resulting discrepancies between nanodisc and bilayer environments. Several different approaches have been previously developed for the simulation of nanodisc systems and CG simulation approaches have been used to successfully simulate nanodiscs which can include membrane proteins¹⁴⁰. However, these approaches require prior

knowledge of the number of lipids to be inserted into the system and this information is often unavailable for systems containing a membrane protein. As a result, many have focused on methods to prepare nanodiscs which only contain a lipid patch and MSPs, without an embedded membrane protein, thereby investigating the properties of the lipids in these systems. A comparison of existing membrane mimetics used for solubilisation of membrane proteins, including nanodiscs and bicelles, has identified variation in lipid properties between different techniques, including variations in bilayer thickness ²¹¹.

Experimental results, including those from NMR structural studies of small diameter nanodiscs, have been used to construct a number of simulated nanodisc systems in the absence of an embedded membrane protein which have shown good agreement with experimentally measured properties ²¹². Other approaches have also integrated the results from scattering experiments in order to aid the production of nanodisc models ¹⁴¹.

As a computational re-creation of experimental techniques, a multiscale simulation approach for the self-assembly of nanodisc systems has been demonstrated using both CG and atomistic simulations ²¹³. This requires an extended CG simulation, up to 42 μ s in length, in order to allow the lipid and MSP components to self-assemble, followed by conversion to atomistic detail using back-mapping ²¹⁴. However, self-assembly methodologies require significant computational resources and are therefore less suitable for large diameter nanodiscs and those with embedded membrane proteins.

Simulations of nanodiscs have also suggested that the properties of the system components are not consistent throughout the nanodisc. Investigation of different sized nanodiscs with a single lipid species suggested that the structural properties, including the bending modulus and thickness of the system components, vary spatially ²¹⁵. Alterations to the saturation of nanodisc lipids have also been demonstrated to affect the structural rearrangements which occur upon the increase in temperature of nanodiscs ²¹⁶. Simulations involving nanodisc-like structures, in which DNA is used to constrain the lipid patch, have shown similar disparities between system lipid properties and bulk lipid properties as have been observed in MSP based nanodiscs ²¹⁷.

Several membrane proteins have now been simulated in nanodiscs, for a variety of purposes. The B cell lymphoma-2-associated X protein was embedded into a nanodisc in order to study its effects on a membrane, where it was noted to cause disruption of the bilayer, leading to the formation of pores ²¹⁸. The perturbation of membranes has also been studied using nanodiscs by embedding the membrane protein β -barrel assembly machinery (BAM), which disrupts the membrane, particularly in the region around the lateral gate. This is proposed to be important in the function of this protein ²¹⁹. Simulations have also been used to study the use of nanodiscs for delivery of lipid soluble drugs to plasma membranes ²²⁰ and this research also suggested the presence of different regions of lipid properties within the nanodisc, dependent upon the location of the lipids in the system. This effect has also been observed in other simulations and experimental work involving EPR ²²¹.

The approach taken here, particularly the utilisation of MD simulations in both CG and atomistic detail ²⁷, overcomes some of the difficulties associated with previous methodologies such as poor system stability, limited convergence and high computational demands, and allows simulation of greater lengths of time with increased complexity.

3.1.4 Pipeline Approach

A multiscale automated pipeline methodology was developed in order to address some of the limitations of previous approaches. In particular, this approach aimed to offer a method for simulation of nanodiscs containing a customisable lipid patch, in an automated manner, which could be adapted to accommodate different sized nanodiscs and include a membrane protein if required. The use of a multiscale simulation strategy allowed CG simulation to be used for system assembly, so as to eliminate the need for extended simulations in the order of 10 μ s and to improve the stability of the resultant system when converted to atomistic detail.

In order to determine the effectiveness of the simulation protocol proposed here, the output from the pipeline was compared with experimental data as determined by SAXS for a nanodisc system containing the same number and lipid species as well as the same MSPs ¹⁴².

3.2 Methods

GROMACS (version 2018.6; www.gromacs.org)¹⁵⁸ was used to perform the MD simulations. The nanodisc systems were composed of 152 molecules of 1,2-dilauroyl-sn-glycero-3-phosphocholine (DLPC) lipids, matching the numbers of reported lipids by scattering experiments¹⁴², and two copies of the amphipathic MSP1D1^{137,222}. Atomic coordinates of the MSP1D1 were obtained from the CHARMM-GUI webserver¹⁹⁴.

3.2.1 Coarse-Grained Simulations

CG simulations used the Martini 2.2 force field^{183, 192}. The MSPs were converted to a CG representation using the `martinize.py` script (version 2.4). An elastic force constant was applied to backbone particles within the MSPs at one of three values (0, 500 or 1000 kJ mol⁻¹ nm⁻²), using either the Martini ENM or ElNeDyn ENM¹⁸⁹ with a 0.7 nm cut-off.

A CG lipid bilayer containing DLPC lipids was prepared using the `insane.py` script¹⁹⁵ and was simulated for 100 ns. Following this, a core was selected containing the same number of lipids per nanodisc (152) as determined by SAXS and SANS¹⁴². Lipids were selected from the CG bilayer within a specified radial cut-off from the centre of the bilayer.

Lipid selection was performed by a python script, which took a concatenated PDB file produced by combining the coordinates of the MSPs, into which a dummy particle had been inserted at the centre of mass of the MSPs to provide the centre for the calculations which followed, and the lipid bilayer produced earlier. This PDB coordinate file was then processed by the script to select all of

the MSP particles and lipids within a defined cut-off distance from the centre of mass dummy particle. The lipid selection was carried out using the distance of the centre of mass of each individual lipid from the dummy particle. In order to aid the selection of the lipids that were closest to the centre of the nanodisc the Cartesian coordinates of the lipids were converted by the script to polar coordinates, with the origin centred on the dummy particle.

The radius used to carry out the selection of the lipids was calibrated using the experimentally reported ratio of DLPC lipids to MSP for this system, as well as the established relationship between the number of residues present in the MSPs and the number of lipids incorporated into one leaflet of the nanodisc lipid core²²³. For systems containing a single lipid species the following relationship has been observed for the number of lipids in one leaflet of the nanodisc, N :

$$M = \frac{2(\pi r + \sqrt{\pi NS})}{L}$$

Where M represents the number of residues present in one leaflet of the nanodisc, r is the mean radius of the MSP α -helix, which is approximated to 0.55 nm, S is the mean surface APL and L the helical pitch per residue in the MSP, which was estimated to be 0.15 nm²²⁴.

The cut-off distance was calculated to return numbers of lipids reported from experimental observations. This core of lipids was then combined with the two MSPs to create a nanodisc. Subsequently, energy minimisation and

equilibration were performed and then the system was simulated for 5 μ s to allow assembly to take place.

This initial system was neutralised using counterions and the final NaCl concentration adjusted to ~ 0.15 M, following which the system was energy minimised for 5000 steps and was then equilibrated. The temperature was maintained at 310 K by the application of the velocity rescaling thermostat. Isotropic pressure coupling was applied with a pressure of 1 atmosphere, using the Berendsen barostat with a time constant of 1.0 ps. Subsequent simulations were performed for 5 μ s with a step size of 20 fs.

3.2.2 Atomistic Simulations

The final frame from each of the CG simulations was converted to atomistic detail using the CG2AT2 program ⁹⁰, producing all atom representations of the species present in the CHARMM-36m force field ¹⁶². Virtual sites were added, following which the system was solvated using the TIP3P water model and neutralised using counterions with the final NaCl concentration adjusted to ~ 0.15 M. A 1000 step energy minimisation simulation was carried out and subsequent equilibration simulations were performed for 0.1 ns with a 4 fs timestep with backbone restraints imposed to maintain the orientation of backbone hydrogen bonding.

The product of the CG2AT2 conversion was simulated for 200 ns using a 4 fs timestep. The neighbour list was updated every 20 fs. PME ¹⁷⁰ was used to calculate long range electrostatic interactions further than 1.0 nm. Van der Waals interactions were switched to 0 at 1.0 nm. The system pressure was

maintained at 1 bar through use of the Parrinello-Rahman barostat using a relaxation time of 2.0 ps and temperature was maintained at 310 K using the velocity rescaling thermostat and a relaxation time of 0.2 ps.

3.2.3 Analysis

Simulation trajectories were processed using GROMACS utilities¹⁵⁸. The trajectories were aligned in order to maintain the bilayer normal with respect to the *z* axis. The 200 ns atomistic output of the method was then used for analysis using locally developed python scripts.

3.2.3.1 SAXS Analysis

The CRY SOL software package²²⁵ was used to calculate the expected SAXS profiles of both the CG and atomistic systems. For the CG analysis, snapshots were taken of the CG trajectory every 50 ns following the initial 1 μ s of simulation. These frames were converted to atomistic detail (using the CG2AT2 program⁹⁰) and used for the calculation of the predicted SAXS profile. To reduce the number of independent fitting parameters, a two-step calculation was performed. For the atomistic analysis, the atomistic simulations were seeded from the final frames of the CG systems with or without and ENM applied. Snapshots were taken from the final 100 ns of the atomistic trajectories and the CRY SOL software package was again used to calculate the predicted scattering profiles.

During the first step, the application of CRY SOL²²⁵ provided values for the contrast of the hydration shell and the optimal atomic group radius for each system. These were averaged and used for the second stage, and during the

second calculation the fitting parameters for the contrast of hydration shell and the optimal atomic group radius were kept constant, using a previously established methodology ²²⁶. In all cases, 99 spherical harmonics were used, and this procedure was used for each of the ENMs. Experimental errors were assessed to be slightly underestimated and were rescaled ²²⁷ to obtain realistic values for χ^2 :

$$\chi^2 = \frac{1}{M-4} \sum_{i=1}^M \frac{I_{SAXS,i} - I_{fit}(q_i)}{\sigma_i},$$

Where M is the number of points, $M-4$ is the estimated degrees of freedom after fitting four free parameters to the data, I_{SAXS} are experimental SAXS intensities, σ denotes the rescaled experimental errors and $I_{fit}(q_i)$ is the model fit evaluated at q_i . q is the momentum transfer and is calculated as follows:

$$q = 4\pi \sin(\theta) / \lambda$$

Where 2θ is the scattering angle and λ is the wavelength.

3.2.3.2 Lipid Analysis

APLs were calculated using the FATSLiM software package ²²⁸ on the output from the 200 ns atomistic simulations, with default settings. Order parameters of the carbon atoms in the acyl chains of the nanodisc lipids were compared to those in bilayer simulations under the same conditions.

Deuterium order parameters were calculated using gmx order on the final 25 ns atomistic output of the pipeline for each carbon-carbon bond in the lipid tails, on a per lipid basis.

This calculation was performed using the following equation:

$$S = \frac{3\cos^2\theta - 1}{2}$$

Where θ denotes the angle between the acyl chain and membrane normal. The order parameters, S , of the two acyl chains were calculated separately and then averaged to provide a value for each lipid.

3.3 Results

3.3.1 Alternative Nanodisc Simulation Methodologies

The production of bilayer simulation systems using CG self-assembly has been well established¹⁰. Investigations were carried out to determine if this method could also be utilised for the production of nanodisc systems. In order to perform this, two copies of MSP1D1 were inserted into a simulation box, followed by the insertion of randomly positioned DLPC lipids. Position restraints were applied to the MSP particles and a short 100 ns simulation was performed to allow the assembly of a CG bilayer which incorporated the MSPs. The final stage of production required the removal of lipids which did not assemble into the lipid patch of the nanodisc. An overview of this approach is shown in *Figure 3.3*.

Once the extraneous lipids had been removed from the simulation system it could be further simulated. However, often the number of lipids incorporated within the MSPs was greater than had been observed experimentally and upon removal of the positional restraints on the MSP particles, deterioration of the structure was observed with loss of large numbers of lipids from the nanodisc and loss of MSP structure. Generally, this approach was too unreliable for the automation of nanodisc simulation production and was deemed unlikely to be able to produce systems which contain the membrane protein embedded within the bilayer core of the nanodisc.

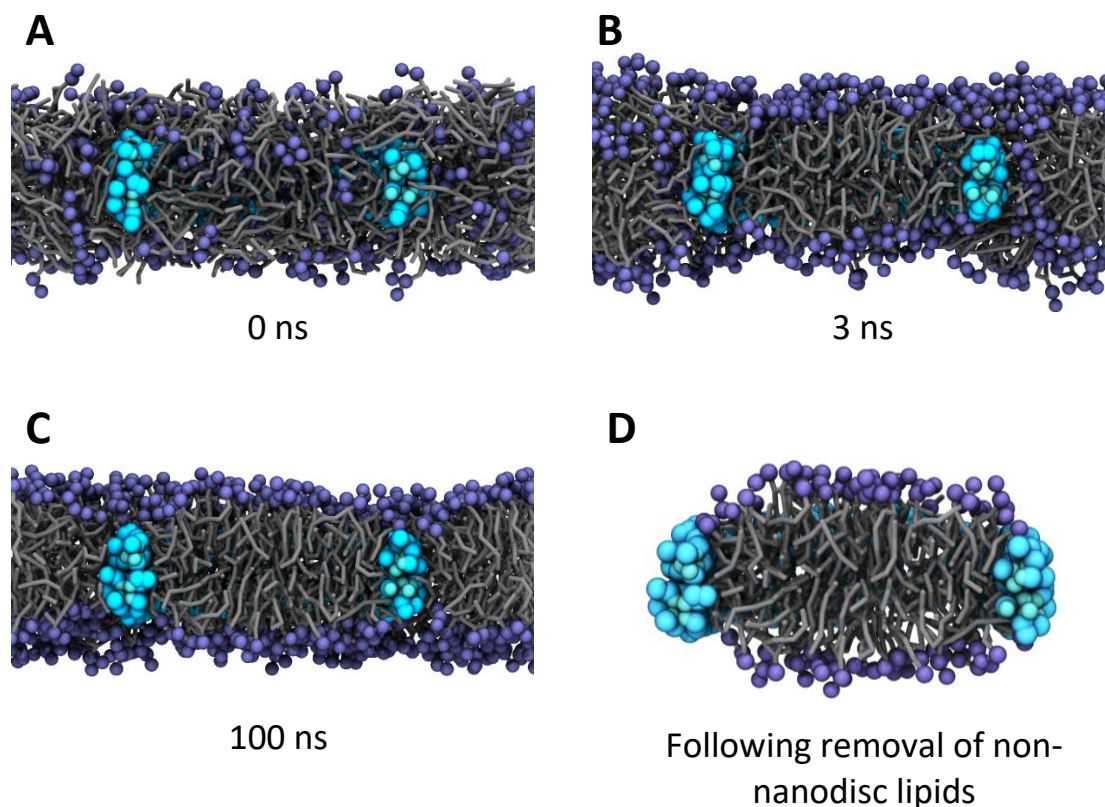


Figure 3.3 – The nanodisc self-assembly methodology

The use of a CG self-assembly methodology was investigated for the production of nanodisc systems. However, this proved to be an unreliable approach due to large variations in the number of lipids incorporated into the nanodisc, often exceeding the experimentally reported numbers of lipids for each system. This approach also had some other drawbacks; firstly, that manual curation of the structures was often required in order to remove incorrectly positioned lipids, and secondly, that the approach allowed limited control over the precise composition of the lipid patch and leaflet symmetry. The images here show a lateral cross-section of the simulation system at each stage of the assembly process.

A: POPC lipids were randomly inserted into a simulation box containing two copies of MSPID1.

B: A CG simulation was performed to allow bilayer assembly, with some organisation of the lipids visible after the first few nanoseconds.

C: Separation of the lipid population into two separate leaflets occurred by the end of the 100 ns simulation.

D: Following this simulation, extraneous lipids which were not within the MSPs were removed, resulting in the formation of a nanodisc.

3.3.2 A Multiscale Pipeline

A pipeline was developed with the aim of enabling the robust multiscale simulation of nanodisc systems in a manner which allows flexibility in the choice of MSP, lipid composition and presence of a membrane protein embedded into the lipid core of the nanodisc. The production of such a methodology allows the investigation of the properties of nanodisc systems and the impact that the constraining MSP has on the lipid and membrane protein properties.

To this end, the pipeline was developed to allow the incorporation of a CG bilayer, produced by previously established methods, with MSPs, thereby extending existing methods for the production of CG lipid bilayers. Such methods include the use of the CHARMM-GUI¹⁹⁴, the insane.py script¹⁹⁵ as well as lipid self-assembly. In this case the insane.py script was used to generate a pure DLPC bilayer. This bilayer was then equilibrated using a 100 ns simulation before being combined with CG MSPs in the subsequent step of the pipeline. The pipeline is flexible in the origin of the CG bilayer as well as in the composition of the bilayer, allowing customisation of the lipid species that are present in the simulation system. The bilayer which is supplied to the pipeline can contain an embedded membrane protein without any changes to the pipeline methodology, allowing simulation of nanodisc systems that contain a membrane protein as well. This enables the use of previously generated lipid bilayers containing membrane proteins, which can be pre-assembled or obtained from by a number of sources such as MemProtMD⁹ or the

CHARMM-GUI ¹⁹⁴. Simulation of nanodiscs containing membrane proteins will be discussed in more detail in **Chapter 4**. An overview of the automated pipeline methodology for simulating nanodiscs without an embedded membrane protein is shown in *Figure 3.4*.

Comparison between the output of the pipeline and experimentally produced nanodiscs was made by performing simulations of a nanodisc system that has been well characterised. In this case, SAXS experiments have been performed on nanodisc systems composed of MSP1D1 with DLPC lipids and so this composition was chosen for re-creation computationally.

A core of lipids was selected from the previously assembled DLPC bilayer and combined with two copies of an MSP, in this case, MSP1D1. The system was solvated, neutralised using counterions and the final NaCl concentration adjusted to ~0.15 M, following which a 5 μ s simulation was performed.

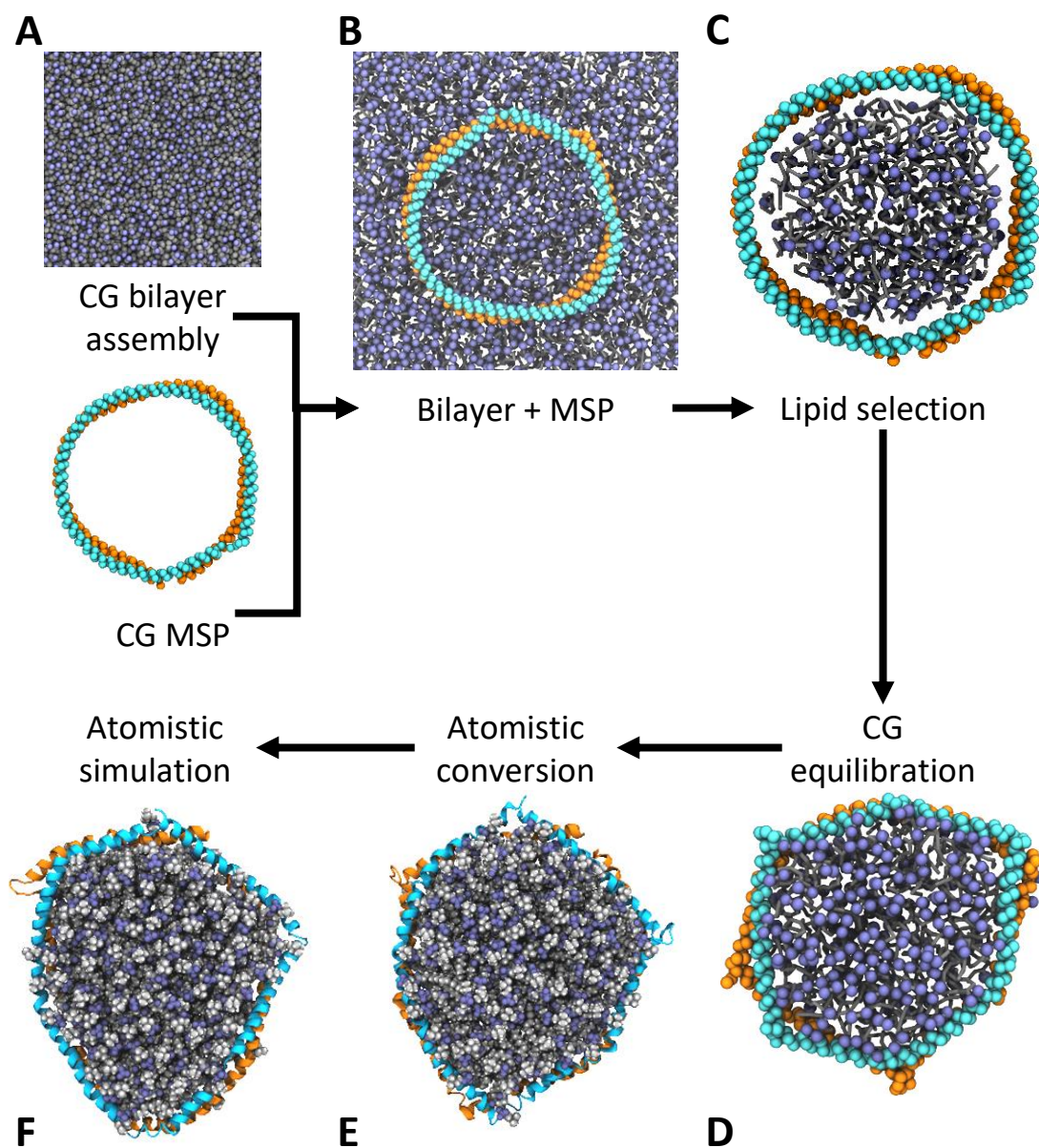


Figure 3.4 – An overview of the automated pipeline methodology

A: A CG lipid bilayer of desired composition (purple and grey) and two CG MSPs (cyan and orange) are generated.

B: The CG lipid bilayer and MSPs are combined.

C: A selection of lipids within the centre of the nanodisc is made.

D: A 5 μ s CG MD simulation is performed to allow equilibration of the system.

E: The system is converted to atomistic detail

F: Further simulation and analysis can now take place.

3.3.3 Elastic Network Models

In CG simulations, in addition to the harmonic restraints used to impose the secondary structural elements of the protein, an ENM is frequently used to maintain the tertiary structure of a protein during extended simulations^{189, 229}. The use of an ENM biases the protein towards its initial conformation. In the case of MSPs, where the conformation of the protein in the nanodisc may vary considerably from the initial model, it was thought that the use of an ENM could also bias the final structural conformation of the nanodisc. Therefore three variations in the overall protocol in terms of the ENM were evaluated: (i) no ENM, with just the Martini secondary structure restraints applied to the MSPs; (ii) Martini ENM, using the `-elastic` option in the `martinize.py` script to generate local harmonic bonds between backbone particles; and (iii) ElNeDyn ENM, using a global ENM between the backbone particles¹⁸⁹. For the latter two options variation of the force constant of the ENM was also explored.

The ElNeDyn ENM, illustrated in **Figure 3.5**, maintained the MSPs in a conformation closer to the original starting model, which was circular. The Martini ENM resulted in the MSPs adopting a more elliptical conformation as was suggested by experimental data¹⁴².

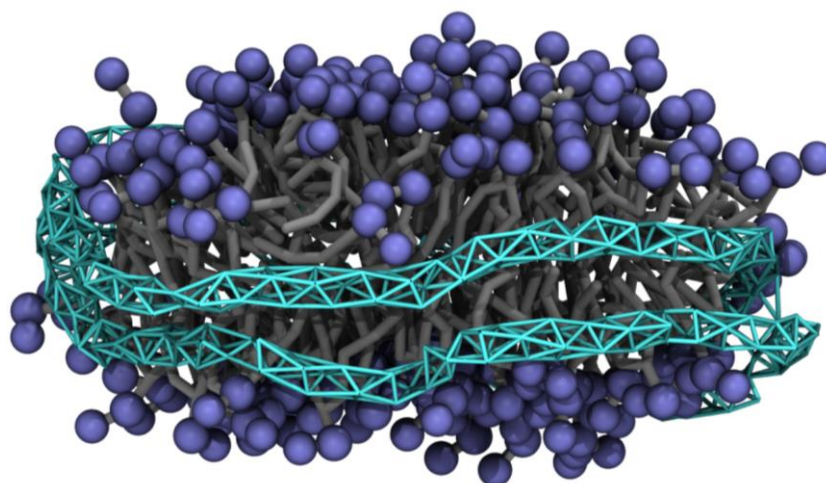


Figure 3.5 – A visual representation of the restraints applied to the MSPs by the ElNeDyn force field

The restraints applied between the backbone particles of the MSPs when using the ElNeDyn force field are shown in cyan (the backbone particles not displayed). Unlike the Martini scheme, where restraints are only applied through existing bonds, the ElNeDyn methodology applies restraints to all neighbouring backbone particles within a given distance, resulting in a mesh-like arrangement.

In CG MD simulations with either the Martini or ElNeDyn ENMs, distortion of the MSPs was observed over the course of the 5 μ s simulations, involving the buckling and intertwining of the two MSP chains as is visible in **Figure 3.6 A**. This distortion was not observed in the No ENM simulations, in which the MSPs remained in a train-track like arrangement with two anti-parallel bands wrapping around the lipid core. In these simulations without an ENM there was much less distortion of the protein backbone and the lipid tails in the nanodisc core were less exposed to the solvent, resulting in a more bilayer-like arrangement. This is demonstrated in **Figure 3.6 B** by comparison of the distributions of lipid headgroup and MSP densities along the normal to the nanodiscs. The No ENM protocol therefore yielded a nanodisc model more consistent with prior expectations of nanodisc organisation¹³⁷.

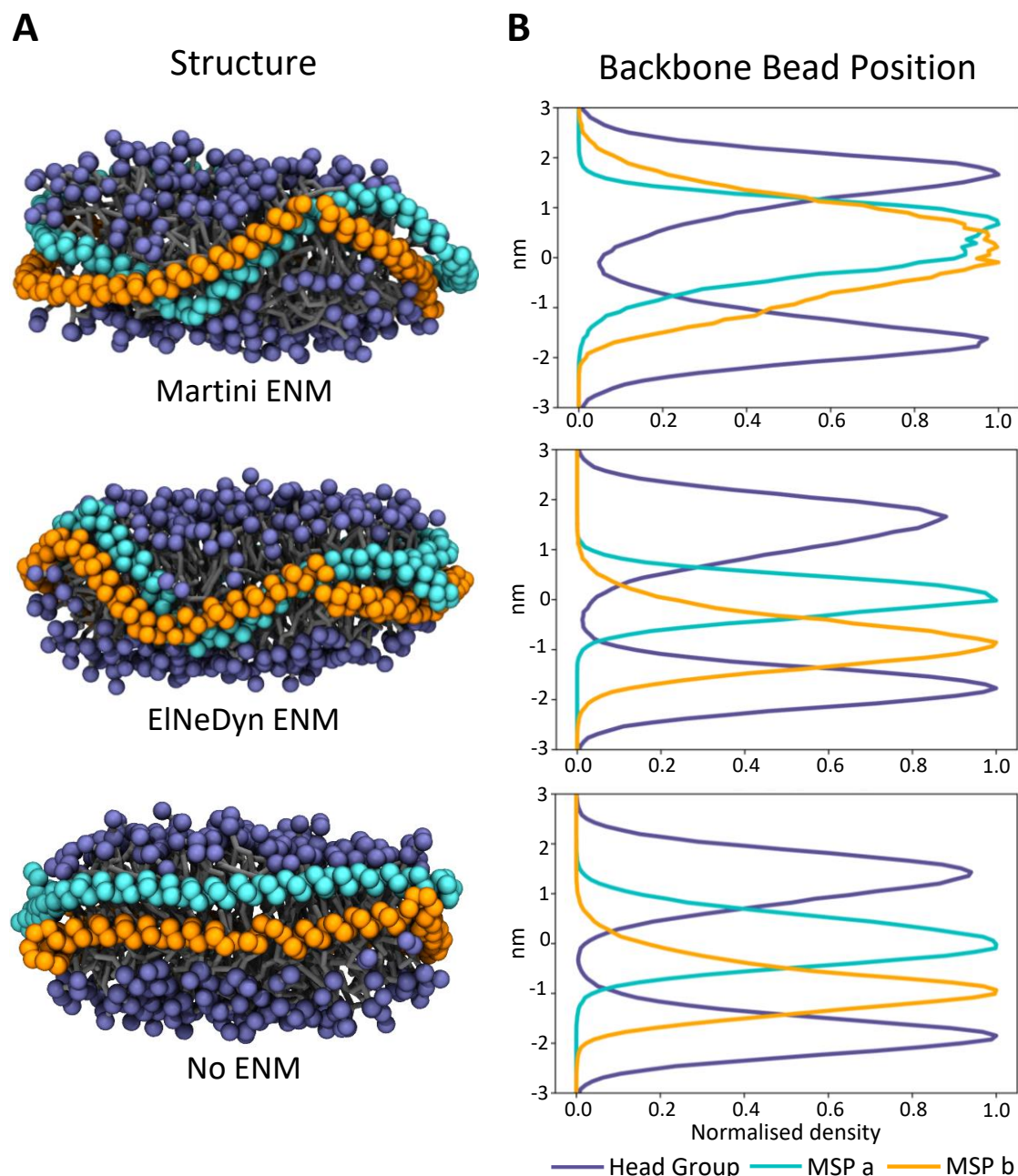


Figure 3.6 – Effects of ENM choice on MSP stability

Comparison of the effects on MSP stability resulting from changes to the ENM used. All systems contain two copies of MSPID1 (orange and cyan), and 152 DLPC lipids (purple). Each system was simulated for three repeats of 5 μ s in CG detail.

A: Intertwining of the MSPs was observed in the systems simulated using elastic networks, but not in the No ENM system.

B: The extent of the intertwining of the MSPs can be measured by plotting the distribution of the backbone MSP particles over the vertical (z) axis of the simulation box, with a higher normalised density indicating more particles at that z axis value. Overlap between the orange and cyan peaks therefore indicates overlap between the particles of the two MSPs, and thus intertwining of the helices. The normalised densities of the lipid head groups are shown in purple for reference.

3.3.3.1 Effects of ENM choice on CG Systems: Comparison with SAXS Data

The nanodisc models generated by the CG simulations were evaluated by comparison with experimental SAXS data¹⁴². For these, SAXS data was referenced for a nanodisc system containing the same lipid, DLPC, and MSP1D1 as was used in the simulations, the only difference being the presence of a purification His-tag in the experimental data. This comparison was made using the CRY SOL package²²⁵ to calculate the expected SAXS profiles following conversion of CG models to atomistic detail using the CG2AT2 program, which was discussed in more detail in **Chapter 2**.

The comparison of the simulation-based and experimental SAXS data shown in **Figure 3.7** indicated that the No ENM protocol for treatment of the MSP restraints resulted in predicted scattering closest to that which was observed experimentally. This was confirmed by both inspection of normalised residuals as a function of momentum transfer (q) and by comparison of overall χ^2 values. In particular, the No ENM protocol gave a better fit to data at q values between 1 nm^{-1} and 2 nm^{-1} than did either of the ENM protocols. This strengthens the conclusion above that the No ENM CG protocol yields a more accurate nanodisc model, as least for the current MSP and lipid combination.

All ENMs produced systems which showed good agreement with the overall size of the nanodisc system seen experimentally, as confirmed at q values between 0.1 nm^{-1} and 0.5 nm^{-1} . Agreement in this region would be expected even with crude models such as a uniform ellipse. Deviation from experimental results was observed at q values above 2 nm^{-1} ; this was likely to

be due to the absence of a His-tag on the MSPs and the effects of coarse-graining on the nanodisc model. In addition, the experimental error in this region was larger than at lower q values. The No ENM protocol gave the best fit to the observed data at q values between 1 nm^{-1} and 2 nm^{-1} , and this was true even for systems where the force constant was altered. This region provides the best test of the nanodisc model produced by the pipeline, showing that it was able to accurately reproduce experimentally observed nanodisc properties including the height of the bilayer, structural oscillations and MSP conformation.

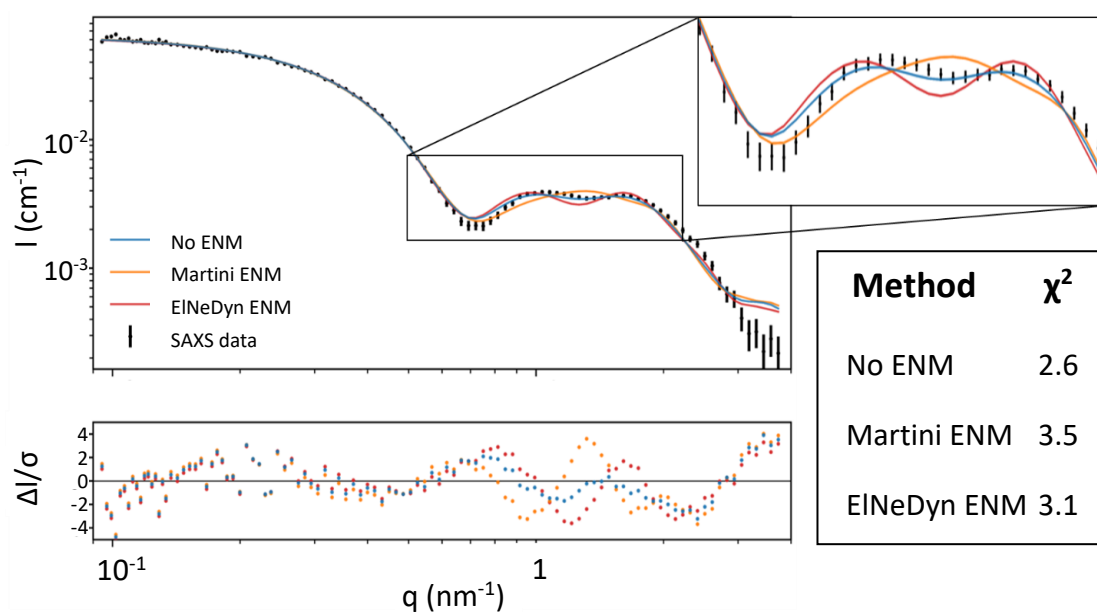


Figure 3.7 – Predicted SAXS profiles of CG systems constructed with different ENMs

The CRY SOL software package was used to compare the experimentally determined SAXS profile¹⁴² of a nanodisc with CG simulation models of the same composition constructed with different ENMs. All systems had the same number of lipids and the same MSP construct. Much better agreement is demonstrated between the experimental SAXS profile (shown in black) and the system without an ENM applied to the MSPs (shown in blue) than with the CG simulations where ENMs were applied. The absence of an ENM results in decreased residuals and hence a reduction in χ^2 values.

3.3.3.2 Effects of ENM Choice on Atomistic Systems: Comparison with SAXS Data

Conversion of each system to atomistic detail and subsequent simulation, followed by application of the SAXS profile prediction software, resulted in the scattering profiles shown in **Figure 3.8**, which were consistent with those demonstrated in the CG simulations. This emphasises the importance of using the correct CG procedures for system initialisation, and in particular how the effects of the ENM must be considered. Varying force constant applied was also investigated; increasing the force constant from $500 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ to $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$, the value that is typically used for CG simulations, resulted in further deviation from the experimentally expected results.

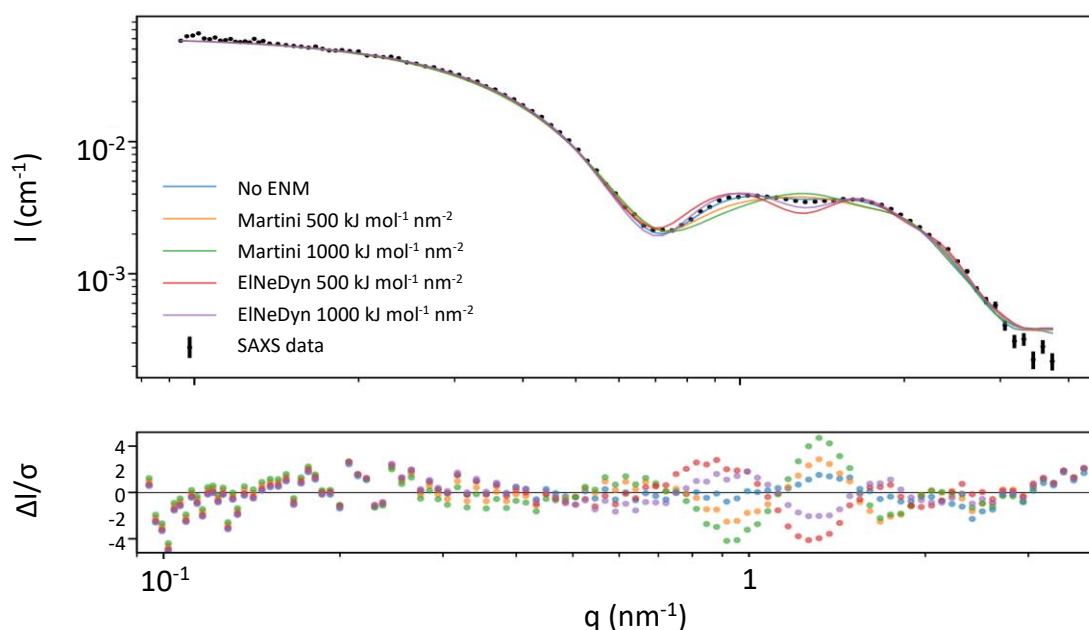


Figure 3.8 – Predicted SAXS profiles of systems constructed with different ENMs after conversion to atomistic detail

The final frames of the CG simulations were converted to atomistic detail and simulated for 200 ns. Predicted scattering profiles of the final 100 ns of the atomistic trajectories was calculated using the CRY SOL software package. These were then compared with the experimentally determined SAXS curve (black), demonstrating better agreement with the No ENM methodology (blue) than those where a force constant was applied to the MSPs during the CG stage of the pipeline.

3.3.4 Analysis of Pipeline Output in Atomistic Detail

Following conversion to atomistic detail using the CG2AT2 program⁹⁰ and a subsequent 200 ns of unrestrained simulations, further investigation of the properties of the pipeline output structures could be performed. In systems where intertwining of the MSP chains was observed during the CG stage, and particularly where this occurred in the terminal regions of the MSPs, large scale rearrangements of the structure were observed over the course of the atomistic simulation as demonstrated in **Figure 3.9**. This was likely to be due to sub-optimal contacts between the nanodisc components (MSPs and lipids) in the respective models at the end of the CG simulations. For example, comparison of the No ENM and Martini ENM models in atomistic MD simulations, as shown in **Figure 3.9 A**, demonstrates that in the No ENM system more substantive contacts are retained between the two MSP polypeptide chains. This is reflected in substantially smaller Root Mean Square Fluctuations (RMSFs) for the No ENM simulation MSPs shown in **Figure 3.9 B** than for the Martini ENMs. Similar behaviour is also observed for the atomistic MD simulations based on the ElNeDyn ENM CG models.

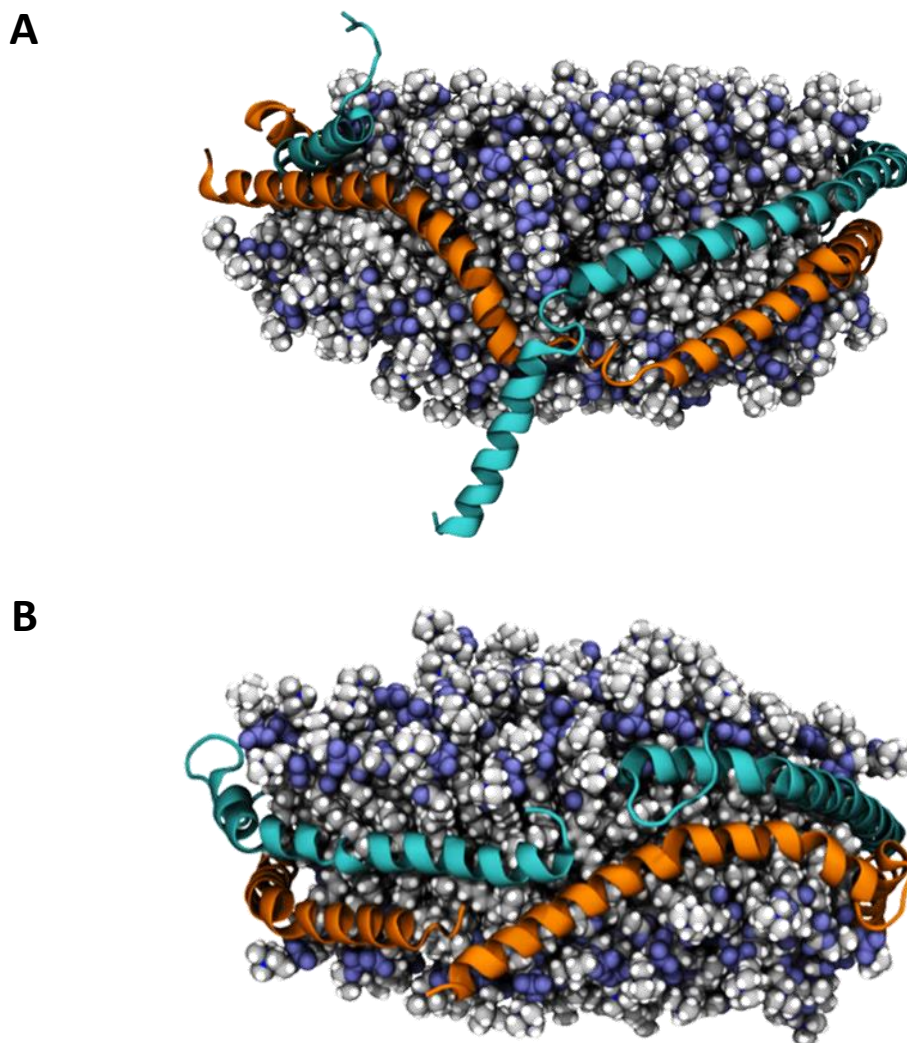


Figure 3.9 – The final structures from two atomistic simulations, with and without an ENM

A: The structure from the final frame of a 200 ns atomistic simulation of a system initially generated in CG using the Martini ENM and a force constant of $500 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. MSP a is shown in cyan and MSP b in orange. There is disruption of the MSPs with loss of contact between the MSP terminal regions and the lipid patch.

B: The output from the final frame of a 200 ns atomistic simulation of a system initially generated in CG without an ENM. The MSPs have not intertwined and are arranged in an antiparallel configuration around the periphery of the lipid patch, with no loss of contact at the terminal regions.

3.3.4.1 Root Mean Square Fluctuation

In the systems created with the Martini ENM a greater loss of secondary structure was observed during the atomistic MD simulations, particularly in

the terminal regions of the MSPs. A common observation was that one of the terminal helices separated completely from the bilayer and only made contact with the solvent. As shown in **Figure 3.10**, terminal residues of the MSPs adopted an unfavourable conformation, with regions of the amphipathic MSP helices no longer in contact with the lipid bilayer, resulting in the terminal regions interacting mainly with the surrounding water. This behaviour was not observed for any of the atomistic MD simulations based on the No ENM models.

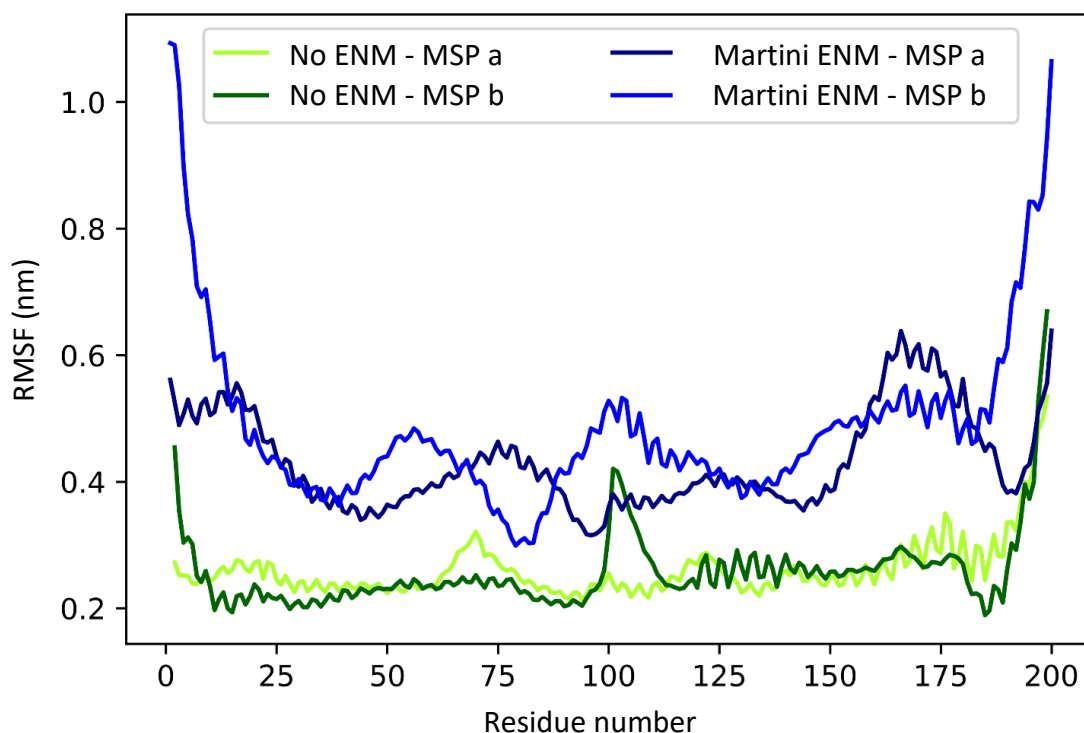


Figure 3.10 – RMSF plots for two atomistic simulations with and without an ENM

The RMSF plots for residues in each MSP for the systems shown in **Figure 3.9** over the duration of a 200 ns atomistic simulation, with MSPs a and b corresponding to those indicated in the previous figure. The atomistic simulations were seeded from the CG simulations with or without an ENM applied. The system initially generated using the Martini ENM exhibits higher RMSFs than the system generated without an ENM, an effect which is particularly pronounced at the terminal regions of the MSPs.

Thus, the No ENM models not only agree better with the SAXS data but are more stable in unrestrained atomistic simulations and therefore will be subject to further analysis in terms of the lipid properties within the nanodisc.

3.3.4.2 Axial Ratio

The axial ratio is calculated by taking the longer axis of a nanodisc ellipse and dividing it by the shorter axis. It can be used to provide insight into nanodisc properties and structure²³⁰. This ratio can be measured experimentally using SAXS, as well as calculated from a simulation system by fitting an ellipse to the MSP backbone particles. SAXS measurements report an axial ratio of 1.29 ± 0.03 , which is in good agreement with the no ENM system results, which showed an axial ratio of 1.27 ± 0.03 . Systems created with the Martini and ElNeDyn ENM had axial ratios of 1.16 ± 0.04 and 1.40 ± 0.07 respectively. The results are summarised in **Table 3.1**.

Table 3.1 – Nanodisc axial ratios for atomistic simulations

Average axial ratios were calculated over the duration of the 200 ns atomistic simulation of each system, by fitting an ellipse to the MSPs. The atomistic systems which were generated in CG without an ENM resulted in axial ratios most similar to those in the experimental SAXS data.

System setup procedure	Atomistic Axis Ratio
Experimental (SAXS)	1.29 ± 0.03
No ENM	1.27 ± 0.03
Martini ENM	1.16 ± 0.04
ElNeDyn ENM	1.40 ± 0.07

The lower axial ratios for the Martini systems suggest that the ENM is maintaining the nanodisc in a conformation which is closer to the initial starting configuration of the MSPs than was noted in the experimental data. The higher

axial ratios for the ElNeDyn ENM are due to the increased eccentricity displayed by these nanodisc systems, and the larger range of axial ratios within this group indicates increased structural variation in these systems. Simulations in which no ENM was used produced the best agreement with the experimental data both in terms of the axial ratio and in terms of the system dynamics.

3.3.4.3 Area Per Lipid

Following the analysis of overall nanodisc properties, the behaviour of lipids in the atomistic MD simulations was investigated further. Properties of the lipid components of the nanodisc were assessed by comparison with the atomistic MD simulations of the same lipid species (DLPC) arranged in a bilayer, as well as with experimentally reported values²³¹. As shown in **Figure 3.11**, the APL increases with distance from the centre of the nanodisc, such that at more than 3 nm from the centre, most lipid molecules occupy a greater area than their equivalent in a lipid bilayer. At the centre of the nanodisc, the APL was in good agreement with experimental observations²³¹ and simulations in a lipid bilayer. This trend in APL away from the centre was observed for lipids in both leaflets of the nanodisc and seemed to correspond to the MSPs disturbing the peripheral lipids resulting in their adoption of conformations less typically encountered in bilayer environments.

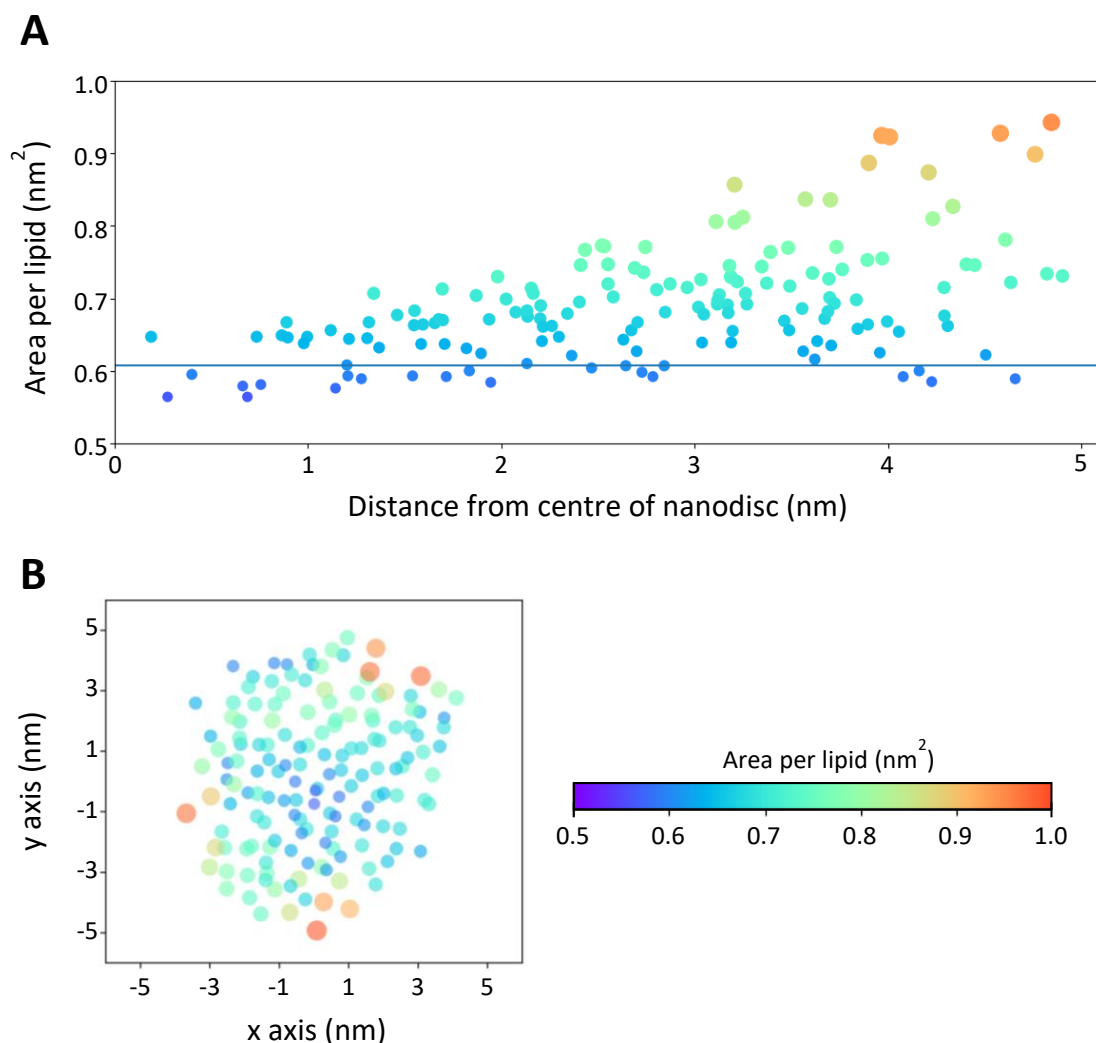


Figure 3.11 – Variation in APL with distance from the centre of a nanodisc

APLs were calculated from the final frame of a 200ns atomistic simulation of a system which had been generated in CG without an ENM.

A: Radial distribution of APL, which increases with distance from the nanodisc centre. The experimentally derived APL value for DLPC in a bilayer is 0.61 nm^2 ²³², which is indicated by the blue line. While some lipids of the expected value are found at all distances, the lipids closer to the MSPs are more perturbed and have correspondingly higher APLs than those at the centre of the nanodisc.

B: The same system when viewed from above the plane of the nanodisc, illustrating the increase in APL with distance from the centre. The axes have been zeroed on the centre of the nanodisc and it is demonstrated that while the majority of lipids exhibit an APL similar to that found experimentally, those around the periphery of the nanodisc tend to occupy a larger area.

For comparison with APLs calculated for a system generated using the Martini ENM, see **Appendix 1**.

3.3.4.4 Order Parameters

Lipid order parameters (S) provide further insights into the orientation and the order of a bilayer membrane. This measure is frequently used for the parameterisation of force fields and can be experimentally measured using solid state deuterium NMR spectra²³³. MD simulations allow the determination of lipid order parameters on a per lipid basis, unlike experimental observations which are limited to determination of average lipid order parameters. This enabled investigation of the influence of the position of a lipid within the nanodisc upon the order parameter, which can be compared with how the same lipid species behaves in a bilayer environment, as shown in **Figure 3.12**.

When the overall average lipid order parameters for the nanodisc were calculated they were comparable to those for bilayer lipids, with the average $S(sn1)$ values a little lower for the nanodisc than the bilayer and the average $S(sn2)$ profiles were nearly identical for the two systems. However, when the nanodisc order parameters were considered on a per lipid basis, different populations of lipid could be identified within the nanodisc. Increased order parameters ($S \sim 0.22$) characterised the lipids at the centre of the nanodisc, forming a region of approximately 3 nm across which was more ordered than would be expected for the same lipid in a bilayer ($S \sim 0.18$). A general decrease in S was observed as the distance from the centre of the nanodisc increased. The peripheral lipids in direct contact with the MSPs were the most perturbed and, in a few cases, lipids displayed very low order parameters due to their

being trapped between the helices of the MSPs. However, most lipids at the bilayer-MSP interface displayed a more moderately reduced order parameter. This is a result of the MSPs causing these lipids to adopt a different orientation with respect to the bilayer normal, so as to enable them to pack against the hydrophobic amino acids of the MSPs. This observed difference in lipid properties in a nanodisc when compared to a bilayer may lead to variation in protein-lipid interactions, and thus affect protein conformation and function.

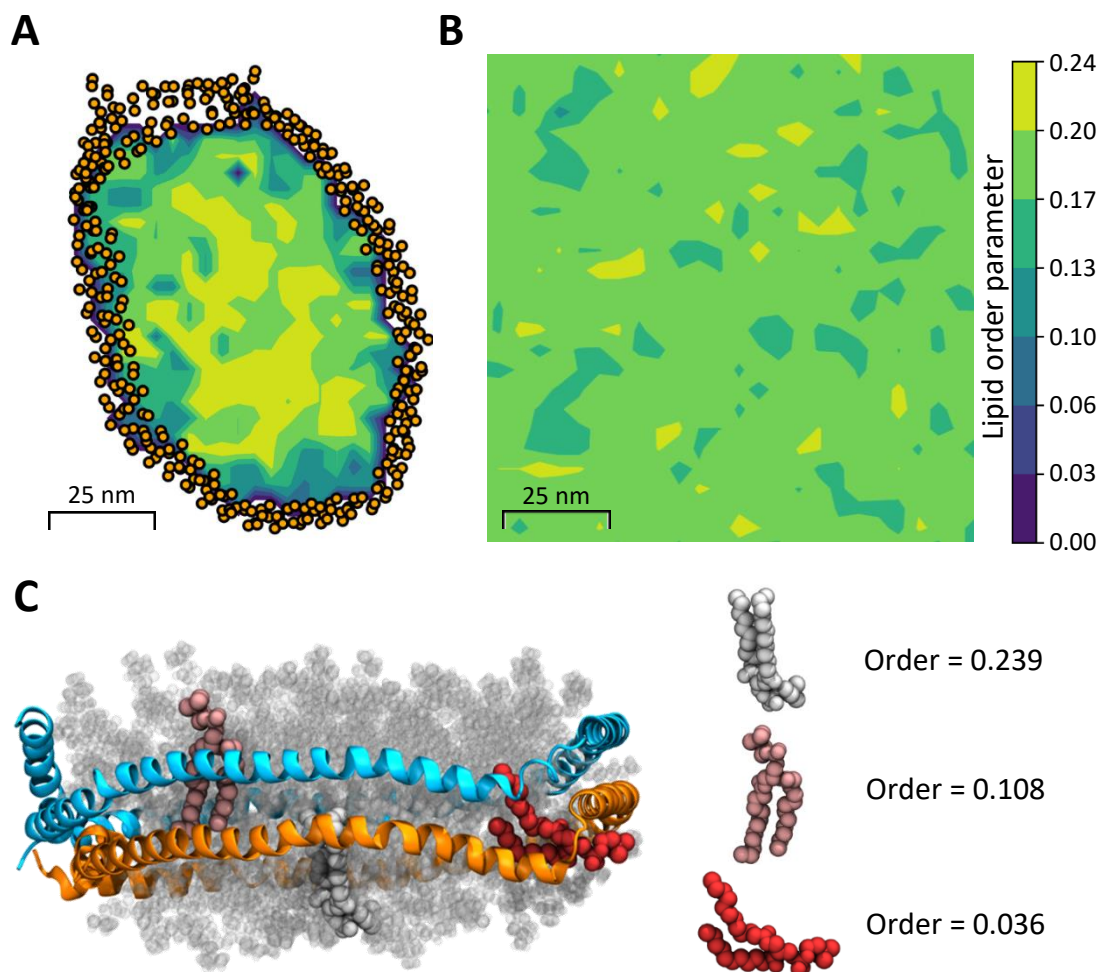


Figure 3.12 – Lipid order variation in nanodisc and bilayer environments

A: Order parameters from the final 25 ns of the atomistic simulation, plotted as a contour map. Higher lipid order parameters are demonstrated at the nanodisc centre, with decreasing order with increased distance from the centre.

B: The same analysis performed on a bilayer system. Lipid order throughout is similar to that at the nanodisc centre, without the perturbation of lipids demonstrated at the periphery of the nanodisc due to interaction with the MSPs.

C: Three representative lipids from different regions of the nanodisc are shown in white, pink and red, with their respective order parameters. MSPs are shown in cyan and orange. The white lipid is located near the nanodisc centre and exhibits order parameters similar to those expected for a lipid in a bilayer. The pink lipid is further from the centre of the nanodisc and has an intermediate order parameter. The red lipid is at the interface between two MSP helices, which has perturbed the positioning of the lipid tails, resulting in a low order parameter.

3.4 Conclusions

Recent rapid improvements in the resolution of protein structures determined using methods such as cryo-EM, coupled with increased utilisation of nanodiscs for the solubilisation of membrane proteins, has resulted in a requirement for further investigation of the properties of nanodisc systems. This includes assessment of the impact of the constraining MSPs on the internal components of the nanodisc and the resulting changes in lipid and system properties. Through the use of MD simulations, properties of individual lipids within the system can be investigated at a level of detail which is difficult to achieve using other methodologies.

Described here is a methodology for the simulation of nanodiscs using a variable approach, allowing customisation of size, lipid composition, and MSPs used for the system. The development of a multiscale pipeline approach enables both rapid simulation and detailed investigation of nanodiscs. By making use of the computational speed increases resulting from the coarse-graining of simulation systems, this pipeline allows faster system initialisation and equilibration, enabling the study of many different systems in rapid succession. Simulations are then converted to atomistic detail, allowing the user to obtain further insights about system properties and behaviour. This multiscale approach was required due to a lack of ability to produce nanodisc systems in a repeatable manner in previous methodologies for the simulation of membrane proteins. Self-assembly of nanodiscs, for example, showed system instability that would have made it unsuitable for the development of

an automated methodology. The procedure of using a selection of lipids taken from an assembled CG bilayer and combining this with CG MSPs, followed by simulation and then conversion to atomistic detail, allowed the nanodisc system to be assembled in a computationally efficient manner. In addition, the output of the pipeline is in good agreement with experimental scattering observations.

3.4.1 ENM Use in Nanodisc Simulations

Analysis during pipeline development indicated that the use of an ENM at the CG stage resulted in large differences in simulation output, as well as divergence from the experimental data in multiple parameters. Comparison of the predicted SAXS profiles of systems where no ENM was applied with experimentally observed scattering profiles provided validation of this protocol. Good agreement between the predicted and experimental scattering patterns was demonstrated for most q values, with the best agreement at q values between 1 nm^{-1} and 2 nm^{-1} . Deviation from experimental scattering patterns at higher q values could result from the lack of His-tags on the MSPs as well as potentially under reported errors from the experimental data in this region.

Upon conversion to atomistic detail, for which the CG2AT program was used, further investigation of system properties and behavioural characteristics of particular nanodisc components could be performed. Visual inspection of the trajectories revealed that systems created without an ENM did not display the intertwining of MSPs and subsequent loss of parallel arrangement which

was present in other systems, nor did these systems show loss of interaction between the terminal regions of the MSPs and the lipid tails of the nanodisc.

Analysis of the axial ratio exhibited by different nanodisc systems provides a useful metric for the evaluation of nanodisc models and was used to investigate the overall structural conformation adopted by the nanodisc systems. Eccentricity of the system is important when considering the distribution of lipids within the nanodisc, as systems with greater eccentricities have a higher proportion of lipids in close proximity to the encircling MSPs, potentially altering the properties and behaviours of these lipids.

The stability of systems created without an ENM was better than that of systems created with an ENM, as demonstrated in the unrestrained atomistic simulations. RMSF values were much greater for the nanodiscs with an ENM applied during the CG phase, and they were particularly high in the terminal regions of the MSPs. This indicated that these regions suffered a loss of structure, increasing their flexibility due to reduced contact with the hydrophilic tails of the lipid patch. This was energetically unfavourable due to the amphipathic nature of the MSP helices and resulted in increased system instability.

The findings from the scattering, axial ratio and stability analyses, combined with observations of the variation in MSP conformation over the course of these simulations, suggest that simulation of nanodiscs should be performed without an ENM imposed on the MSPs. As a result, the No ENM methodology was chosen to be used in this pipeline.

All subsequent analysis was performed on No ENM systems, and aimed to provide insight into the properties of the constituent lipids of the nanodisc and to compare these properties with those exhibited by the same lipids in a bilayer environment.

3.4.2 Lipid Properties

Importantly, the results of the APL analysis demonstrated that for nanodiscs it is not appropriate to make use of the average lipid properties. A nanodisc is an enclosed space with a relatively small number of lipids when compared to a bilayer. As a result, large variations in lipid properties could be observed due to the presence of the encircling MSPs at the nanodisc boundary. Imposition of this protein boundary at the edge of the lipid patch caused perturbation of the lipids, with increased variance in APL observed at the periphery of the lipid patch. The MSP used for this study, MSP1D1, is at the smaller end of the range of MSPs typically used in nanodiscs experimentally. This smaller diameter resulted in a lipid patch that is similarly smaller in size, meaning that changes to lipid properties resulting from proximity to the boundary MSPs affected a greater proportion of the lipids present in the system. In order to reduce this effect a larger MSP could be used, such as MSP2N2, reducing the proportion of lipids in close proximity to the MSPs.

The APL analysis showed that at the central region of the nanodisc lipids are present which exhibit properties that would be typically expected for the same lipid in a bilayer environment; an APL for each DLPC lipid of 0.62 nm². This was in agreement with experimentally observed data as well as bilayer

simulation results. However, the further from the centre of the nanodisc a lipid was positioned, the greater its APL became. This was likely to be caused by the lipid head groups attempting to form favourable interactions with the confining MSP ring. The extent of this effect might also depend upon the choice of lipid used in the patch at the centre of the nanodisc, since lipids with longer hydrophobic tails may have to alter their conformations more at the nanodisc periphery in order to form contacts with the hydrophobic residues of the MSPs. This explanation might be validated by the typical choice of lipids with shorter tails for experimental use in smaller diameter nanodiscs; typically, the 12-carbon tails of DLPC as was utilised here.

A further example of variation in lipid properties with distance from the centre of the nanodisc was demonstrated in the order parameters of the acyl chains of DLPC. The order parameters of lipids found at the centre of the nanodisc were consistent with those of lipids situated within a bilayer. However, with increasing distance from the centre of the nanodisc the level of disorder increased until the lipids that were directly in contact with the MSPs were reached. These boundary lipids displayed the lowest order parameters and therefore the highest level of disorder.

The observation of a spatial dependence of lipid properties as a function of their distance from the centre of the nanodisc has also been observed through other simulation studies¹⁴⁰⁻¹⁴¹ and experimentally²¹⁵. The changes in lipid properties that resulted from the presence of the encircling MSPs are of interest as these changes may result in the lipids forming different interactions with a

membrane protein which could be embedded into the nanodisc. This is of particular significance in reference to the multiple bound lipid species that have been observed in membrane protein structures solved experimentally using nanodiscs. The results presented here could be of interest when considering selection of particular MSPs for use when studying membrane proteins, giving an indication of the size of MSPs required to solubilise a membrane protein in a nanodisc while allowing the protein to interact with lipids whose properties are minimally perturbed by the encircling MSPs.

The observation of a region near the centre of a nanodisc which contains more ordered lipids means that this region might possess properties which are closer to those proposed to occur in lipid-rafts, where an enrichment of cholesterol and other lipids causes changes in the properties of the raft region of the bilayer independent of the rest of the bilayer. This finding suggests the use of simulations to study nanodiscs which have had a membrane protein embedded into their bilayer patch, and the method developed here is able to produce simulation systems of this variety. The perturbation of lipid properties by nanodiscs and the effect that this has on membrane proteins and their interactions with lipids can be studied by comparison of the same protein in a bilayer environment. This is discussed further in **Chapter 4**.

The accuracy of the pipeline is dependent upon the conformation of the initial MSPs and membrane protein structures used for simulation. These limitations may potentially be alleviated by increasing the length of the CG simulations in the pipeline if required. In addition, the full release of

Martini 3¹⁸⁵ may enable improvements in the ability of the simulations to reproduce experimental results.

In conclusion, when a No ENM methodology is used this multiscale customisable pipeline showed good agreement with experimentally derived data, suggesting that it is suitable for the creation and simulation of nanodiscs. Of note, the data revealed a large variation in lipid properties throughout the nanodisc, often with dependence upon distance from the nanodisc centre. This may have implications when simulating membrane proteins in nanodiscs and also when using nanodiscs to study membrane proteins experimentally.

4

Chapter 4 - Simulation Studies of Nanodiscs with Membrane Proteins

Contents:

4.1 Introduction	106
4.1.1 Background	106
4.1.2 Existing Approaches	107
4.1.2.1 Constructing Nanodiscs Using CHARMM-GUI.....	107
4.1.2.2 Constructing Nanodiscs Using Self-Assembly.....	109
4.1.2.3 Constructing Nanodiscs Using insane.py	111
4.1.3 Simulation of Membrane Proteins in Nanodiscs	113
4.2 Methods	114
4.2.1 Coarse-Grained Simulations	114
4.2.2 Atomistic Simulations	116
4.2.3 Analysis.....	116
4.2.3.1 Lipid Analysis	117
4.2.3.2 Protein-Lipid Interaction Analysis.....	117
4.3 Results	120
4.3.1 Membrane Protein Pipeline Overview and Output	120
4.3.2 Membrane Protein Interactions with MSPs	126
4.3.3 Analysis of Membrane Protein System Properties	129
4.3.3.1 Area Per Lipid in Nanodiscs with Membrane Proteins	129
4.3.3.2 Lipid Deformation	131
4.3.3.3 Order Parameters for Bilayer and Nanodisc Systems	132
4.3.3.4 Root Mean Square Fluctuation Analysis for TRPV6	134
4.3.4 Protein-Lipid Interactions	136
4.3.4.1 Protein-Lipid Interactions in Nanodiscs and Bilayers	136
4.3.4.2 The Effect of Saturation on Protein-Lipid Interactions	142
4.4 Conclusions	145
4.4.1 Membrane Protein-MSP Interactions	145
4.4.2 Lipid Properties	147
4.4.3 Protein Properties and Protein-Lipid Interactions	149

4.1 Introduction

4.1.1 Background

Application of nanodiscs for solubilisation of membrane proteins in a bilayer environment has increased in recent years, as demonstrated by the exponential increase in numbers of publications involving nanodiscs over the last two decades ²³⁴ shown in **Figure 4.1**. A primary reason for the increased utilisation of these systems is their size; at between 6 nm and 17 nm in diameter ⁶⁷, nanodiscs are ideal for the study of most membrane proteins using techniques such as cryo-EM ²³⁵.

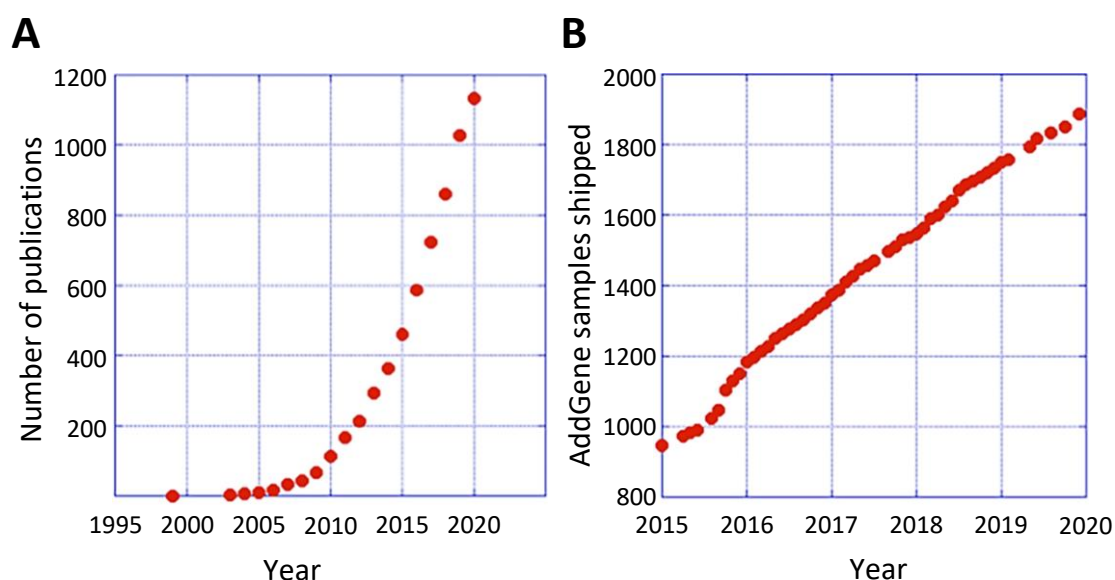


Figure 4.1 – Trends in nanodisc utilisation

A: Utilisation of nanodiscs for the study of membrane proteins has been increasing exponentially over time, a trend which is observed in the cumulative number of publications which make use of nanodiscs and their applications.

B: Another metric that can be used to demonstrate the growing utilisation of nanodiscs for study is the number produced by AddGene per year, as shown here on a cumulative graph.

Figure adapted from ²³⁴

As discussed in **Chapter 3**, the properties exhibited by lipids enclosed within a nanodisc, including APL and order parameters, diverge from those that would be expected for the same lipids in a bilayer environment. These perturbations in properties were demonstrated to not occur uniformly throughout the entirety of the nanodisc structure, instead variations in lipid properties were observed as a function of proximity to the boundary of the nanodisc. It is also understood that since many membrane proteins undergo allosteric regulation by the lipids in their environment ^{120, 236-238}, changes to their regulation and function can occur as a result of changes in lipid properties. However, the impact of these changes on the properties of membrane proteins embedded into a nanodisc environment are currently less well understood.

MD simulations provide a way to investigate these systems and compare the properties of membrane proteins in both bilayer and nanodisc environments. Unfortunately, the *in vitro* scale and complexity of an entire bilayer is currently out of reach of both atomistic and CG simulation-based techniques ²³⁹. However, nanodisc systems, including those containing large multimeric membrane proteins and large MSPs, are well suited for study by MD particularly when a multiscale approach is utilised in order to combine both CG and atomistic simulation techniques ²⁴⁰.

4.1.2 Existing Approaches

4.1.2.1 Constructing Nanodiscs Using CHARMM-GUI

Nanodisc systems with embedded membrane proteins have undergone less investigation than systems lacking a membrane protein, and as a result few

methodologies have been published that are able to produce such systems. One such methodology is the CHARMM-GUI Nanodisc Builder ²⁴¹ which has been incorporated into the CHARMM-GUI Webserver ¹⁹⁴. The Nanodisc Builder utilises a regular grid-like positioning of lipids around a selection of pre-assembled MSPs of varying length, and at the centre of this grid a region of lipids can be removed to allow a membrane protein to be embedded. After assembly, these systems are then simulated to enable the lipids to adopt a bilayer-like arrangement, incorporating the MSPs and membrane protein. This method allows the creation of systems in either atomistic or CG detail. However, a potential drawback of this approach is that it is dependent upon the initial number of lipids inserted into the nanodisc. To calculate the number of lipids to be inserted into the system the Nanodisc Builder allows the input of a desired APL for each lipid species present. As shown previously, this is insufficient for the setup of nanodisc simulations as the APL of an individual lipid varies with its position in the nanodisc. This, along with the Nanodisc Builder not inserting lipids within a certain radius of the protein elements, often results in insufficient lipids being inserted into the nanodisc as a whole. The insufficient number of lipids manifests itself in the appearance of gaps or voids in the lipid patch, particularly adjacent to the MSPs at the boundary of the nanodisc. In these regions, solvent can then come into contact with the hydrophobic regions of the MSPs causing loss of the bilayer arrangement of lipids and exchange of lipids from one leaflet to the other, as demonstrated in **Figure 4.2**. These regions remain present after a microsecond of simulation. An

additional limitation of the CHARMM-GUI approach is that due to its access through an online Webserver, it is not straightforward for it to be automated as was desired for the creation of this pipeline.

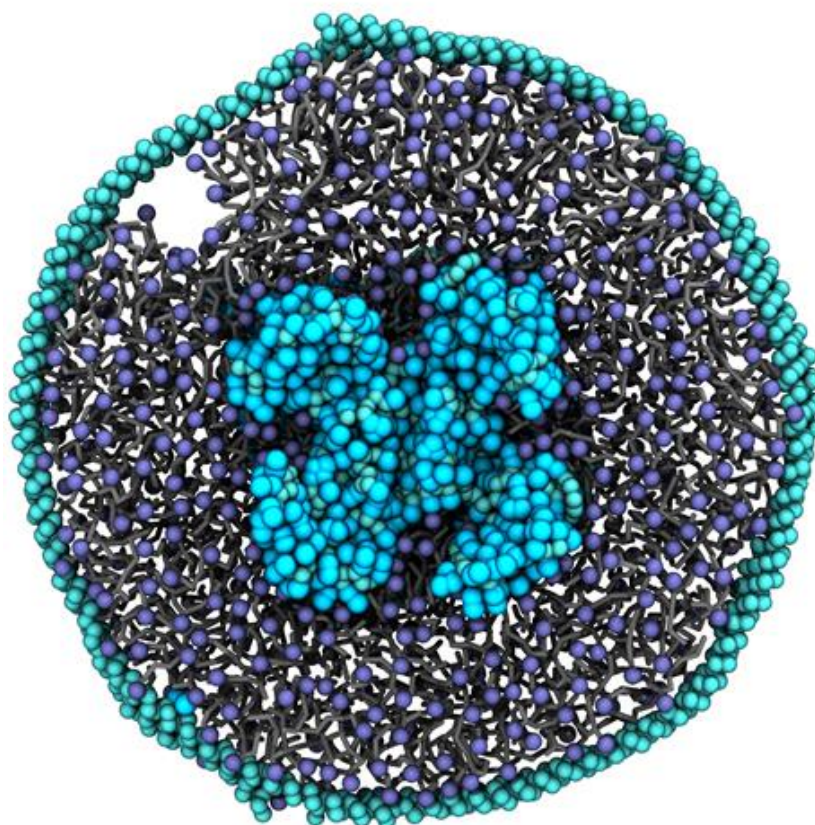


Figure 4.2 – The product of the CHARMM-GUI Nanodisc Builder with an embedded membrane protein

The output of the CHARMM-GUI Nanodisc Builder, with protein in cyan, lipid head groups in purple and lipid tails in grey. Use of this methodology can result in incorporation of an incorrect number of lipids into the system, resulting in voids in the lipid patch where pockets of solvent are found between the lipids and the MSPs.

4.1.2.2 Constructing Nanodiscs Using Self-Assembly

Other methodologies for production of simulation systems with a membrane protein embedded in a lipid bilayer were therefore explored to assess their suitability for adaptation in order to simulate nanodisc systems. One such method is CG self-assembly, as utilised for simulation of membrane

proteins in the MemProtMD database⁹, which attempts to incorporate all published membrane protein structures into a bilayer using initial CG simulations and then conversion to atomistic detail. In order to adapt this approach for nanodisc simulations, randomly positioned lipids were inserted into a region approximately corresponding to where the lipid patch was to form, including both the TM regions of the protein and the MSPs. As a result of this, a patch of lipids formed in a bilayer arrangement both inside the nanodisc and outside the boundary MSPs. Protein components were position-restrained in order to allow the lipid patch to assemble. Following the initial simulation, lipids not incorporated into the nanodisc were removed.

The self-assembly of nanodisc systems that did not contain a membrane protein was unsuccessful, as demonstrated in **Chapter 3**, on account of issues relating to the incorporation of the correct number of lipids, as well as system stability. This was primarily caused by difficulties in allowing the self-assembly of the lipid patch in a way which formed favourable contacts with the interior hydrophobic region of the MSPs. There was often loss of the bilayer arrangement of lipids with movement of lipids from one leaflet to the other, combined with failure to incorporate correctly with the MSPs, as shown in **Figure 4.3**. An additional drawback to this approach was difficulty in determining the criteria used for the removal of lipids that have not been incorporated into the nanodisc system. In particular this methodology would be difficult to automate without user input to select which lipids had not been incorporated correctly into the nanodisc.

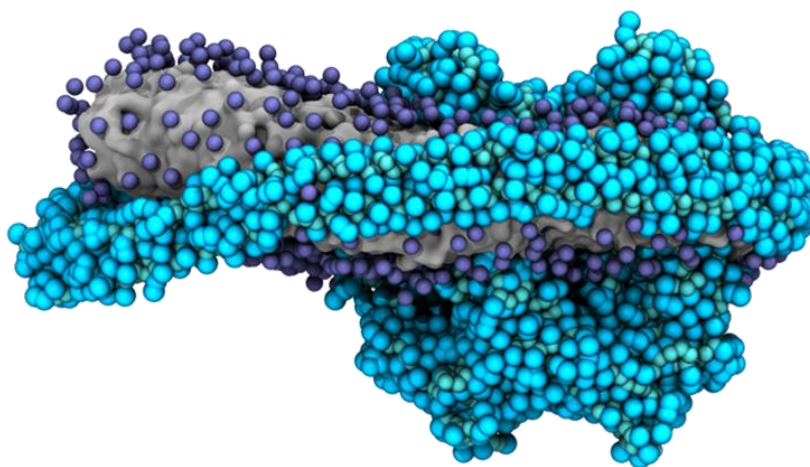


Figure 4.3 – The product of the self-assembly methodology with an embedded membrane protein

The output of the self-assembly methodology, as was described in the previous chapter for nanodiscs not containing membrane proteins. This system is coloured as before with protein in cyan, lipid head groups in purple and lipid tails in grey. The embedded membrane protein displayed instability and there was poor interaction of the lipid patch with the boundary MSPs, resulting in loss of the bilayer arrangement of the lipids in one region. In addition, this methodology did not allow control over the composition and symmetry of the leaflets within the lipid patch.

4.1.2.3 Constructing Nanodiscs Using insane.py

The use of the `insane.py` script in order to generate a CG initial configuration for nanodisc simulation was also investigated as a potential method of nanodisc construction. A copy of a TRPV6 channel (PDB ID 6BO8) was combined with two copies of MSP2N2, within which the system lipids were positioned using the `insane.py` script. The `-disc` option was used in the script in order to produce a circular arrangement of lipids surrounding the membrane protein and extending to the boundary MSPs. The lipids inserted by the script were aligned according to a grid, in a similar manner to that used by the CHARMM-GUI Nanodisc Builder in its initial system configuration. However,

this approach shared similar difficulties to the CHARMM-GUI method, primarily resulting from issues in assigning the correct number of lipids that should be inserted into the system. However, in the case of the `insane.py` script, an excess of lipids incorporated into the lipid patch was more commonly observed as shown in **Figure 4.4**. A possible way to alleviate this drawback would be to change the average APL of the lipids included within the system and the radius of the initial disc of lipids created by the script. However, this approach was undesirable for an automated system because it has not been possible to calculate the required APL for a given mixture of lipid species that would produce the correct number of lipids for a given nanodisc system.

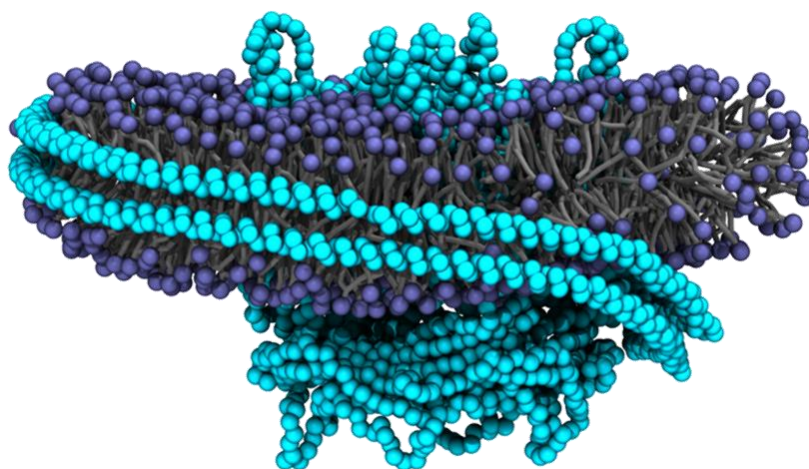


Figure 4.4 – The product of assembly using the `insane.py` script

The output of the `insane.py` script¹⁹⁵ with protein in cyan, lipid head groups in purple and lipid tails in grey. Having been supplied with a CG membrane protein and two copies of MSP2N2, the script inserted a circular patch of lipids in order to generate a nanodisc system. However, this resulted in a larger number of lipids being incorporated into the nanodisc than required, leading to loss of bilayer arrangement in some regions at the periphery.

As a result of these issues, systems that were created using the `insane.py` script suffered from instability in a similar way to previously investigated methodologies. In particular, these systems exhibited an unlikely arrangement whereby there was reduced contact between the lipid patch and the MSPs throughout the structure, resulting in regions where there was complete loss of separation of the two bilayer leaflets and non-incorporation of the lipids into interactions with the hydrophobic regions of the MSPs.

4.1.3 Simulation of Membrane Proteins in Nanodiscs

In response to the increased utilisation of nanodiscs for the characterisation and structural determination of membrane proteins, as well as deficiencies in current methodologies for the simulation of these systems, an automated customisable simulation system was required. The pipeline-based system presented here is intended to enable the simulation of any membrane protein, in a lipid patch of desired composition, bounded by individually selected MSPs, in an automated manner. The approach taken here builds upon the methodology demonstrated in **Chapter 3**, adapted to allow simulation of nanodiscs which also contain an embedded membrane protein. Following initial development, this approach was validated by the study of a selection of different membrane proteins whose structures have been determined using nanodiscs. The lipid environment that used during structural determination was the same as used within the simulated nanodiscs. Membrane protein properties in the nanodisc simulations were compared with those in bilayer simulations with the same membrane protein and lipid compositions used.

4.2 Methods

Simulations were performed using GROMACS (version 2018.6) ¹⁵⁸. Five different membrane proteins were simulated, which had been solved in different lipid compositions and using a variety of MSPs, as detailed in **Table 4.1**. Atomistic coordinates of the membrane proteins were obtained from the MemProtMD database ⁹ and the MSP coordinates were obtained from the CHARMM-GUI webserver ²⁴².

4.2.1 Coarse-Grained Simulations

CG simulations were performed using the Martini 2.2 force field ^{183, 192}. The MSP coordinates were obtained from the CHARMM-GUI webserver ²⁴³ and were converted to a CG representation using the martinize.py script (version 2.4). The coordinates for the membrane proteins used for the simulations were obtained from the MemProtMD database ⁹. Consistent with the findings from **Chapter 3**, no ENM was applied to the CG MSPs. However, the Martini ENM was applied to the backbone beads of the membrane protein using the default values, including applying a force constant of 500 kJ mol⁻¹ nm⁻² with a 0.7 nm cut-off.

The membrane protein of interest was embedded into a CG bilayer created using the insane.py script. The composition of this bilayer and the relative abundance of each lipid species is shown in **Table 4.1**. These bilayers were created to be symmetrical in composition, in order to be consistent with the lipid symmetry that is present in nanodisc systems. Following this, the systems were simulated for 100 ns in order to allow membrane assembly to occur.

Following production of the CG bilayer of desired composition, with the membrane protein of interest *in situ*, the bilayer was then incorporated into the pipeline designed in **Chapter 3**. The dimensions of the bilayer generated by the `insane.py` script depended upon the size of the MSPs to be used, thereby ensuring that sufficient lipids were present in the simulation system to allow incorporation of the nanodisc into the system. Additional sizes of MSP were also incorporated into the pipeline using a different value for the radial cut-off in the lipid selection script. After the script had been applied, a PDB coordinate file was produced containing the membrane protein surrounded by a core of lipids encased by the MSPs.

The subsequent steps of the pipeline as described in **Chapter 3** remained the same. Consistent with the previously described methodology, the combined MSPs, membrane protein and lipids were then solubilised and neutralised using counterions and the final NaCl concentration adjusted to ~0.15 M. Following this the system underwent energy minimisation for 5000 steps. The temperature was maintained at 310 K by the application of the velocity rescaling thermostat. Isotropic pressure coupling was applied for all nanodisc systems with a pressure of 1 atmosphere, using the Berendsen barostat with a time constant of 1.0 ps. Subsequent simulations were performed on the systems for 5 μ s with a step size of 20 fs with three repeats for each simulation system.

In order to facilitate later comparison between bilayer and nanodisc environments, the initial system produced by the `insane.py` script, containing the membrane protein of interest embedded in a lipid bilayer, was extended

after the equilibration simulation to a simulation length of 5 μ s. The same settings were used for the nanodisc and bilayer simulations with the exception of the pressure coupling, which was applied isotropically for the nanodisc systems and semi-isotropically for the bilayer systems.

4.2.2 Atomistic Simulations

The final frame from each of the CG simulations was converted into atomistic detail using the CG2AT2 program⁹⁰, producing all atom representations of the species present in the CHARMM36m force field¹⁶². The system was solvated using the TIP3P water model²⁴⁴ and neutralised using counterions with the final NaCl concentration adjusted to ~0.15 M. A 1000 step simulation was carried out with backbone restraints imposed on the atoms of the MSPs in order to maintain the orientation of backbone hydrogen bonding, followed by a subsequent 0.1 ns minimisation simulation with a 2 fs timestep.

The product of the CG2AT2 conversion was then simulated for 200 ns using a 2 fs timestep, with the neighbour list updated every 20 fs. PME¹⁷⁰ was used to calculate long range electrostatic interactions further than 1.0 nm and van der Waals interactions were switched to 0 at 1.0 nm. The system pressure was maintained at 1 bar through use of the Parrinello-Rahman barostat using a relaxation time of 2.0 ps and temperature was maintained at 310 K using the velocity rescaling thermostat and a relaxation time of 0.2 ps.

4.2.3 Analysis

GROMACS utilities were used to process the simulation trajectories and align the bilayer normal with respect to the z-axis. Analysis was performed on

the CG and atomistic trajectories depending on the analysis method. The GROMACS utility `gmx rmsf` was used to calculate the RMSF of the atomistic membrane protein over the 200 ns trajectory. For the TRPV6 channel the RMSF was averaged over each of the four subunits of the protein.

4.2.3.1 Lipid Analysis

The APL and order parameter analysis was performed on the 200 ns atomistic simulation trajectories using the same procedure as described in **Section 3.2.3.2**. The FATS_{Li}M software package ²²⁸ was used to investigate the APL of the components of the systems. The GROMACS utility, `gmx order` was used to calculate the deuterium order parameter for each carbon-carbon bond in the acyl chains of the lipids.

The deformation of the lipid bilayer was investigated by processing the trajectories with Perl analysis scripts which measured the amount of membrane curvature and deformation in the constituent lipids of a bilayer ^{9, 245}.

4.2.3.2 Protein-Lipid Interaction Analysis

Using the PyLipID software package ²⁴⁶ interactions between the membrane protein and the lipids were analysed in both nanodisc and bilayer simulations, enabling comparison between interactions in the two environments. This analysis was performed on the CG simulations for each lipid species present in the system on a per residue basis. Residence times were calculated using dual cut-offs of 0.55 nm and 0.8 nm. Interactions were calculated from the time of a lipid moving closer to the membrane protein than the first cut-off value, and the interaction was considered finished when the lipid moved further away

than the upper cut-off value. These cut-offs were taken from previously published values ²⁴⁷.

The residence time of a lipid is the average time that an individual lipid is continuously interacting with a particular residue. Residence times were calculated using the normalised survival time correlation function, $\sigma(t)$:

$$\sigma(t) = \frac{1}{N_j} \frac{1}{T-t} \sum_{j=1}^{N_j} \sum_{v=0}^{T-t} \tilde{n}_j(v, v+t)$$

Where T is the length of the simulation trajectory and N_j represents the number of lipids of a particular species with that have an interaction time of greater than 0. The function $\tilde{n}_j(v, v+t)$ takes the value of 1 if lipid j interacted continuously with the protein from time v to time $v+t$ (inclusive) and if this is not the case it takes the value of 0 ²⁴⁸⁻²⁵⁰.

Over the course of the trajectory for each timestep of t from 0 to T ns, the values of $\sigma(t)$ are calculated and then normalised by dividing by $\sigma(0)$, resulting in the survival time-correlation function having a value of 1 at $t = 0$. Following this calculation, a biexponential is fitted to the survival function in order to model the long and short decays of lipid relaxation:

$$\sigma(t) \sim A e^{-k_{off}t} + B e^{-k_{on}t} (k_{off} \leq k_{on})$$

Where k_{off} was the dissociation constant and k_{on} the association constant. The residence time was calculated as the reciprocal of k_{off} .

The PyLipID software package was also used to calculate the lipid occupancy of each lipid species present in a given system. Occupancy was calculated using the percentage of frames of the trajectory in which the lipid of interest was within the distance of the lower cut-off from the residues of the membrane protein.

4.3 Results

4.3.1 Membrane Protein Pipeline Overview and Output

The difficulty of adapting existing techniques for creation of bilayer systems containing membrane proteins to allow the production of nanodisc systems, combined with increasing interest in these systems, has necessitated the production of an automated pipeline for their creation and simulation. The approach developed here is based upon that described in **Chapter 3**, modified to incorporate the input of embedded membrane proteins.

First, the atomistic coordinates of the protein must be obtained, in this case as cleaned PDB coordinates from the MemProtMD database, though any source of atomistic protein coordinates could be utilised. The membrane protein used to demonstrate the functionality of this pipeline was TRPV6²⁵¹, the structure of which was solved using nanodiscs to a resolution of 3.6 Å. It was converted to CG detail using the martinize.py script (version 2.4) and embedded into a CG bilayer of a composition mimicking that used during structural determination. This mixed lipid bilayer was generated using the insane.py script¹⁹⁵, containing POPC, POPE, PIP₂, POPA and cholesterol in the ratio 46:23:16:6:10. An additional TRPV6 system was prepared with POPC alone to allow investigation of membrane proteins in nanodiscs with a single lipid species. For these systems MSP2N2 rather than MSP1D1 was used in order to maintain agreement with the MSPs used experimentally. Subsequent steps were the same as those previously described for systems without a membrane protein. An overview of the pipeline is shown in **Figure 4.5**.

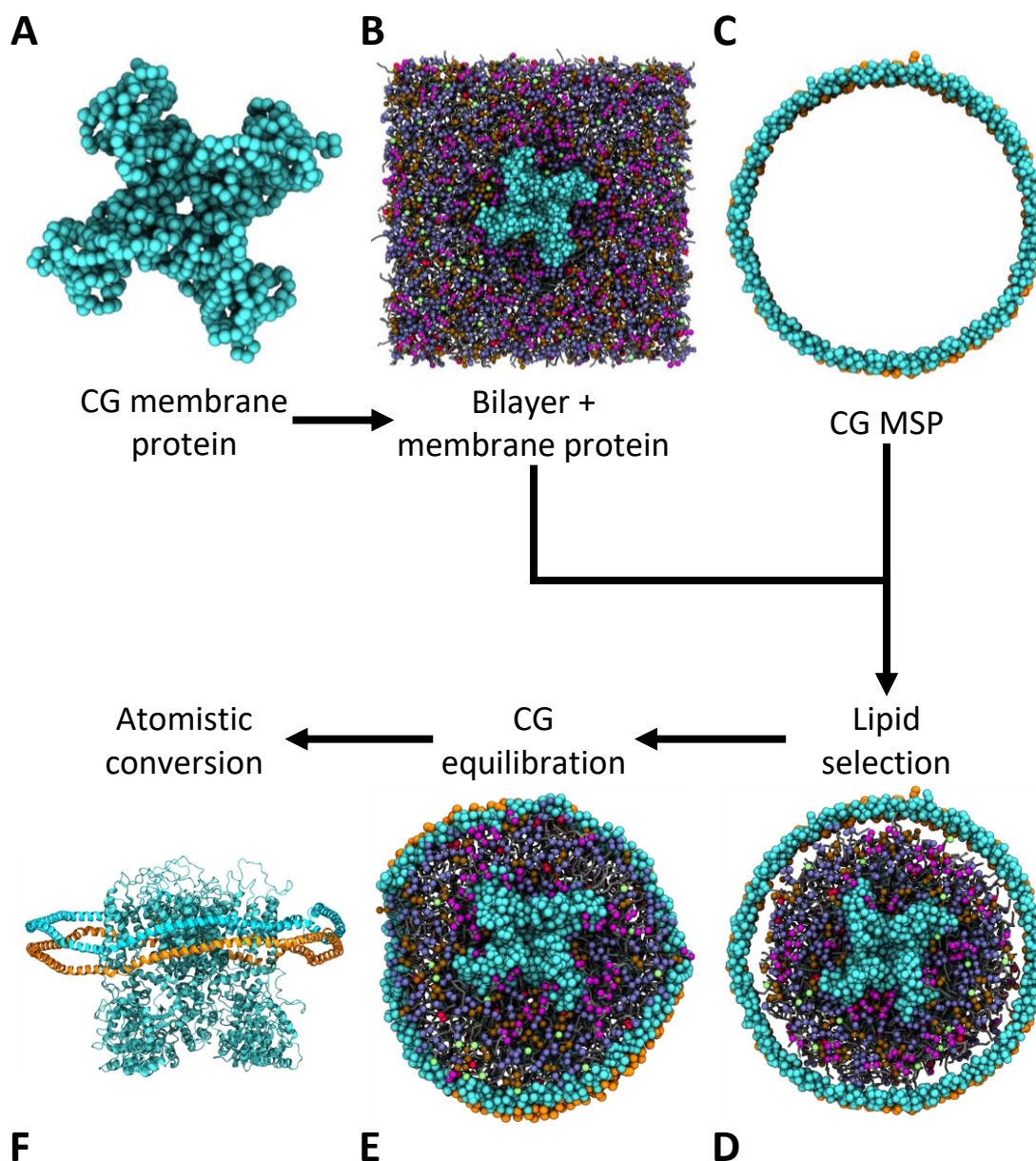


Figure 4.5 – Application of the automated pipeline methodology to an integral membrane protein

A: Atomistic coordinates of a membrane protein are obtained and converted to CG detail, in this case the coordinates of TRPV6 were obtained from MemProtMD⁹⁻¹⁰.

B: The CG protein is embedded into a CG bilayer using the *insane.py* script.

C: The automated pipeline methodology is then applied; the CG MSPs are combined with the bilayer plus membrane protein model.

D: Lipid selection is performed, producing an approximately circular patch of lipids around the membrane protein.

E: CG simulation then takes place to allow interactions between the lipids and the MSPs, resulting in an equilibrated nanodisc system.

F: The system is then converted to atomistic detail prior to further simulation and analysis.

Over the course of the simulation pipeline the CG bilayer containing the membrane protein was combined with the specified MSPs and the system underwent lipid selection to produce a lipid patch bounded by the MSPs at the periphery and with the membrane protein at the centre. The completed CG system then underwent 5 μ s of simulation after which the system was converted to atomistic detail for further simulation and analysis. The output from the atomistic simulations of TRPV6 in a nanodisc shown in **Figure 4.6**.

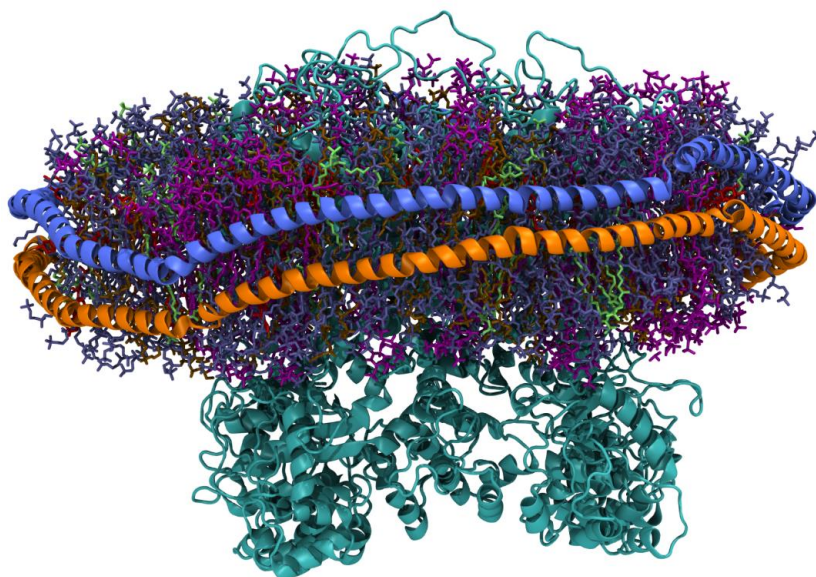


Figure 4.6 – The product of the automated pipeline methodology when applied to TRPV6

The product of the final frame of the 200 ns atomistic simulation on the example pipeline output. This system used MSP2N2 and a mixed lipid bilayer consisting of POPC, POPE, PIP₂, POPA and cholesterol in the ratio 46:23:16:6:10. This lipid composition was chosen to mimic the bilayer in which the structure of this channel was solved experimentally.

This pipeline methodology has been designed to be flexible, with three main elements that can be altered in order to produce different nanodisc systems: firstly, the membrane protein to be incorporated (or absence of a membrane protein); secondly, the MSPs which are used to enclose the system; and thirdly, the composition of the lipid patch. A PDB ID can be used to retrieve the membrane protein structure for input into the pipeline, although the suitability of this approach is dependent upon the completeness of the deposited structure. Alternatively, the input of a set of coordinates can be done manually by the user, allowing structures which are not in public repositories to be studied, or those that have defects in the reported structure to be repaired.

In order to test the ability of the pipeline methodology to simulate nanodisc systems containing membrane proteins of various sizes and MSPs of different lengths, several systems were simulated. Either MSP1D1 or MSP2N2 was used depending upon the required size of the system. The proteins chosen originated from a variety of organisms and had been structurally determined in nanodiscs of different lipid compositions. The lipid compositions for each of the simulations were chosen to mimic those of the bilayers used during structural determination of the membrane proteins and a summary of the system components and lipid compositions used for each of the simulations is shown in **Table 4.1**.

Table 4.1 – Application of the automated pipeline methodology to several membrane proteins with individual lipid patch compositions

Membrane proteins which have been structurally determined using nanodiscs were simulated using the pipeline methodology. The lipid composition was selected in order to mimic the experimental conditions that were utilised for the structural determination of the membrane proteins.

Membrane protein	PDB ID	MSP	Lipid composition	Lipid ratio
TmrAB ABC exporter	6RAF	MSP1D1	POPC POPE POPG CARD	30 46.9 16.2 6.9
Glucosyltransferase ALG6	6SNI	MSP1D1	POPC POPE POPI PIP ₂ POPS POPA CHOL	30 8 15 10 8 2 20
ABCG2 transporter	6HBU	MSP1D1	POPC POPE PIP ₂ POPS POPA CHOL	12.6 33.1 4.1 18.5 0.8 25
PfCRT transporter	6UKJ	MSP1D1	POPG	100
TRPV6 ion channel	6BO8	MSP2N2	POPC POPE PIP ₂ POPA CHOL	46 23 15 6 10
Glutamate transporter homologue GltTk	6XWR	MSP2N2	POPC POPE POPG	25 56 19

Using these components, the automated pipeline was used to generate a selection of simulation systems, as shown in **Figure 4.7**. This demonstrates the ability of the pipeline to simulate both monomeric and multimeric membrane proteins in nanodiscs of various sizes and lipid compositions.

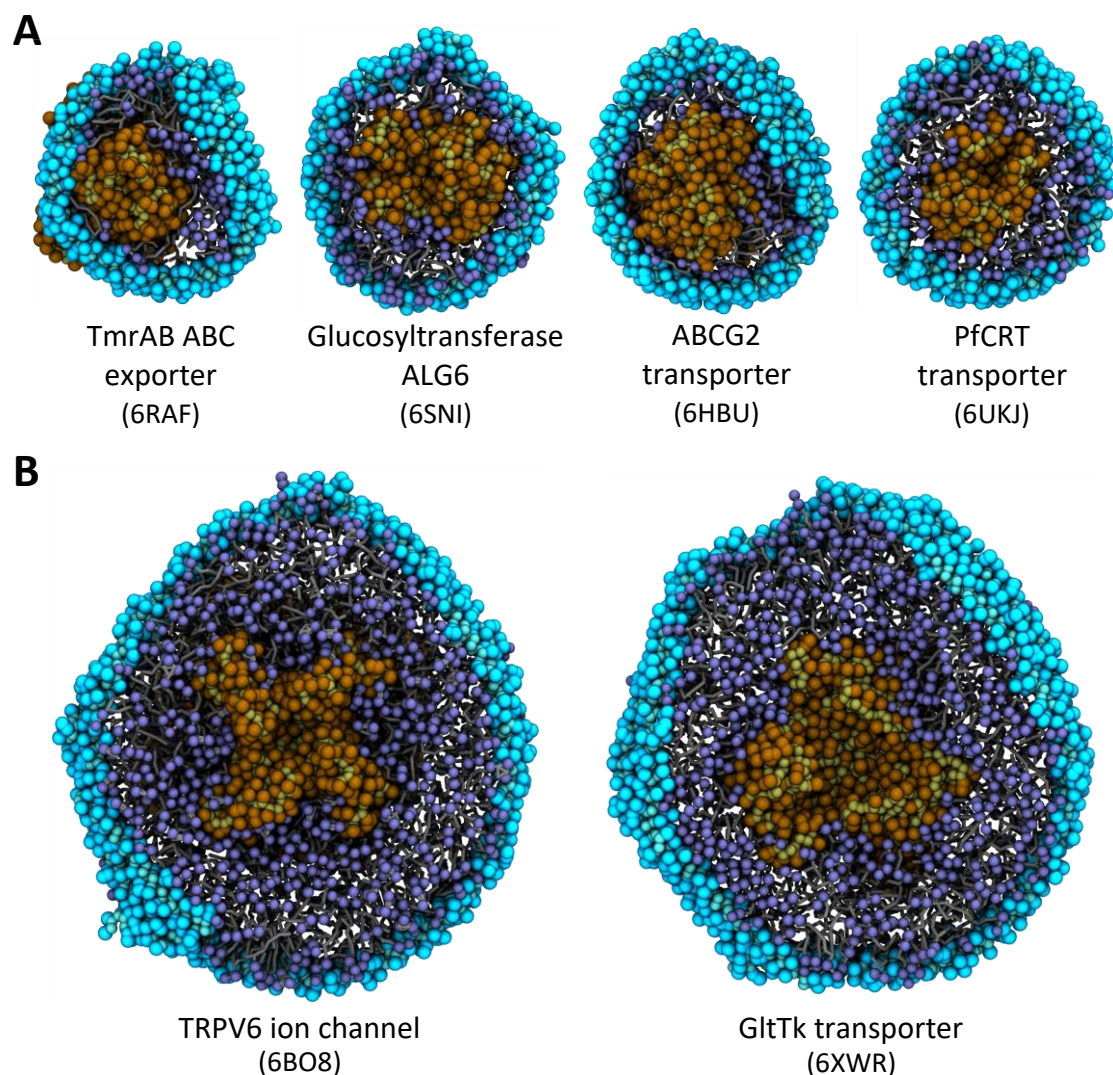


Figure 4.7 – The products of the automated pipeline methodology for multiple systems of varying sizes

The composition of the lipid patch utilised in these simulations mimicked that used during structural determination of these membrane proteins, as detailed in **Table 4.1**. Both monomeric and multimeric proteins were simulated, with a variety of sizes of nanodisc produced by variation of the MSPs used and the number of lipids incorporated.

A: The output of the automated pipeline for four different membrane proteins, using MSP1D1 and a comparatively small number of lipids.

B: The output of the automated pipeline for two larger membrane proteins, simulated in nanodiscs of a larger size using MSP2N2 and incorporating a much larger number of lipids. TRPV6 is a tetrameric membrane protein and GltTk is a trimeric membrane protein, illustrating the ability of the pipeline to simulate non-monomeric proteins.

4.3.2 Membrane Protein Interactions with MSPs

Over the course of the simulations, the membrane protein often did not remain in the centre of the nanodisc, instead diffusing around the lipid patch and in some cases contacting the MSPs. This was particularly observed in systems created using smaller diameter MSPs such as MSP1D1. Embedding into such a system a moderately sized membrane protein, for example the glucosyltransferase ALG6 (PDB ID 6SNI) with twelve membrane-spanning α -helices, results in a layer of lipid patch between the membrane protein and the MSPs of only two or three lipids wide. Movement of the system components results in the membrane protein coming into contact with the MSPs, as is shown in **Figure 4.8 A**, usually after a relatively short time and often within the first 100 ns of the simulation.

Membrane proteins simulated in larger diameter nanodiscs using longer MSPs also exhibited this behaviour to some extent. However, due to the larger system size and the resulting increase in the number of lipid molecules between the membrane protein and the MSP boundary, the collision between the membrane protein and the MSPs occurred later in the simulations.

It was observed that despite further lengthy simulation, up to 5 μ s, once the membrane protein associated with the MSPs it would remain so for the duration of the simulation. In some situations, the membrane protein was observed to slide or roll along the inside of the encircling MSPs, resulting in changes to the specific residues that were in contact with the MSPs over the course the simulation. In multiple repeats of the same system, different regions

of the membrane protein were observed to collide with the MSPs suggesting that this is not a specific interaction occurring with a particular region of the membrane protein sequence and is instead driven by the hydrophobic properties of both the interior of the MSPs and the membrane protein.

To further investigate this phenomenon and assess whether it resulted from the Martini 2 coarse-graining of the simulation system, a simplified system was created using a single lipid species, POPC, using both Martini 2 and the beta release of the Martini 3 force field¹⁸⁵. The membrane protein again collided with the MSPs in the simplified POPC nanodisc system, as shown in **Figure 4.8 A**. It also interacted with the MSPs in a similar manner and over a similar timeframe when simulated with the newer force field, as shown in **Figure 4.8 B**. The composition of the lipid patch used in the nanodisc also did not appear to influence the likelihood of the membrane protein contacting the MSPs.

In order to determine if the contact between the membrane protein and the MSPs was an artefact of coarse-graining in general, the same simulation system was created in atomistic detail with the membrane protein embedded at the centre of the nanodisc. As shown in **Figure 4.8 C**, in atomistic detail the membrane protein again encountered the MSPs and once contact began, the membrane protein did not dissociate from the MSPs over the course of the simulations. This suggests that the observed phenomenon is not a result of the approximations made by the Martini CG model, as it is observed in both CG as well as atomistic detail.

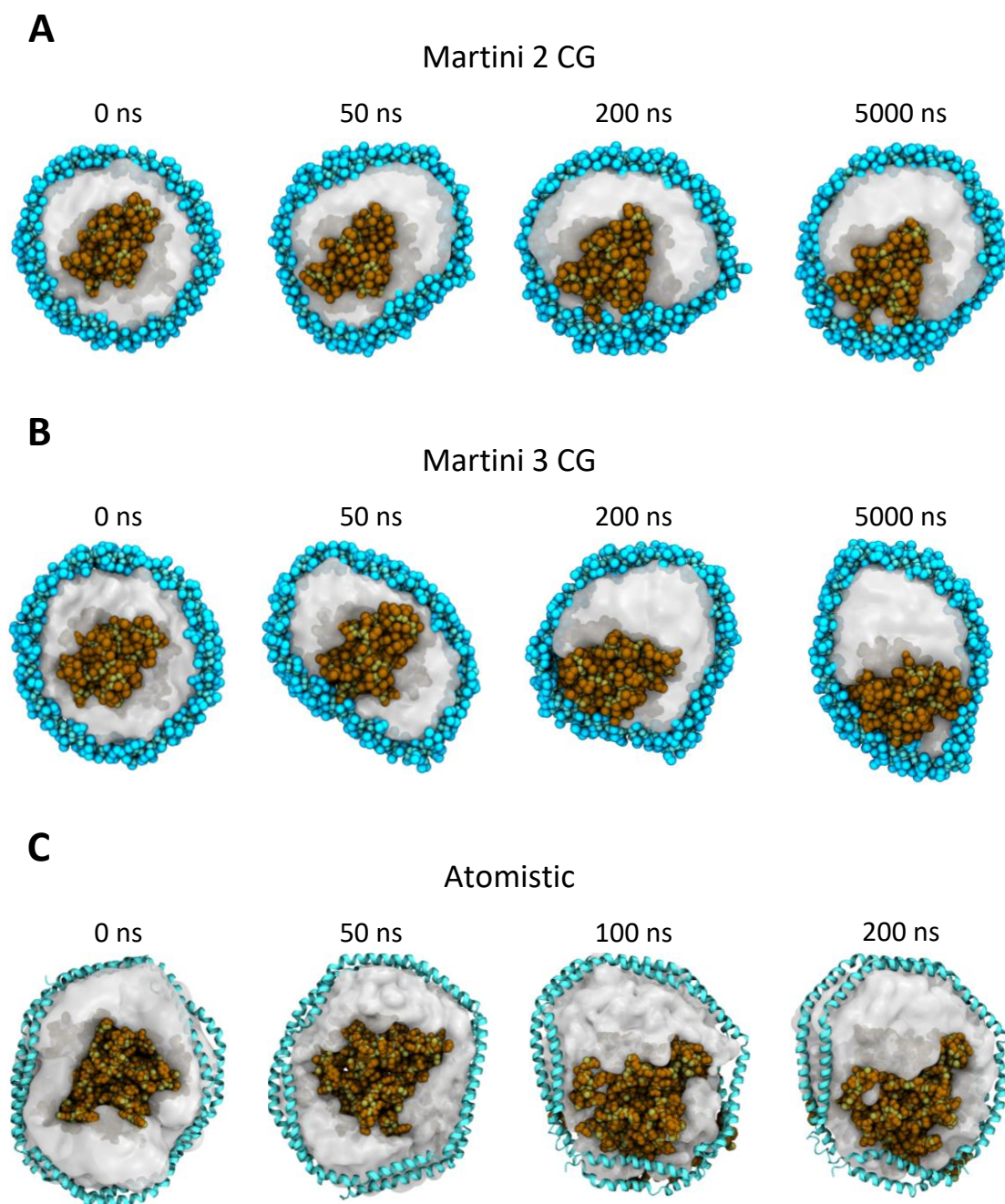


Figure 4.8 – Interactions between a membrane protein and the MSPs

Small diameter nanodisc simulations using MSPID1 and glucosyltransferase ALG6 in a POPC lipid patch.

A: *Martini 2 CG simulations showing contact between the membrane protein and the MSPs early in the simulation, a contact which was then maintained over the course of the 5 μ s simulation.*

B: *Despite the changes in particle interactions with the beta release of the updated Martini 3 force field, contact was still observed between glucosyltransferase ALG6 and the MSPs.*

C: *The same phenomenon was observed in systems in atomistic detail.*

4.3.3 Analysis of Membrane Protein System Properties

4.3.3.1 Area Per Lipid in Nanodiscs with Membrane Proteins

Analysis of systems containing TRPV6 in a POPC nanodisc enabled observation of changes in APL due to the presence of the membrane protein in a nanodisc with a single lipid species. As shown in **Figure 4.9 A**, there is a general trend in the APL of the lipids in the system; an increase in APL is demonstrated with distance from the centre of the nanodisc. This agrees with the observations of nanodiscs without an embedded membrane protein which were shown in **Figure 3.12** and demonstrates that lipids at the periphery of the nanodisc undergo the greatest perturbation in properties as a result of their contact with the MSPs. Bilayer simulations have reported the APL of POPC to be 0.63 nm^2 ²⁵² and, as shown in **Figure 4.9 B**, for a region extending to approximately 6 nm from the centre of the nanodisc the APL of the nanodisc lipids was comparable with the reported values from bilayer simulations. However, outside this region there was an increase in the APL and this effect was greatest at the periphery, where the lipids were in contact with the MSPs.

The properties of different regions of the nanodisc lipid patch can be studied separately in order to investigate changes which occur depending upon location within the system. The constituent lipids of the nanodisc were divided into three groups: those in contact with the MSPs, those in contact with the membrane protein, and the remainder, as illustrated in **Figure 4.9 C** with differently coloured particles. In **Figure 4.9 D**, differences in the APL of these three groups are shown, along with the average APL of each group.

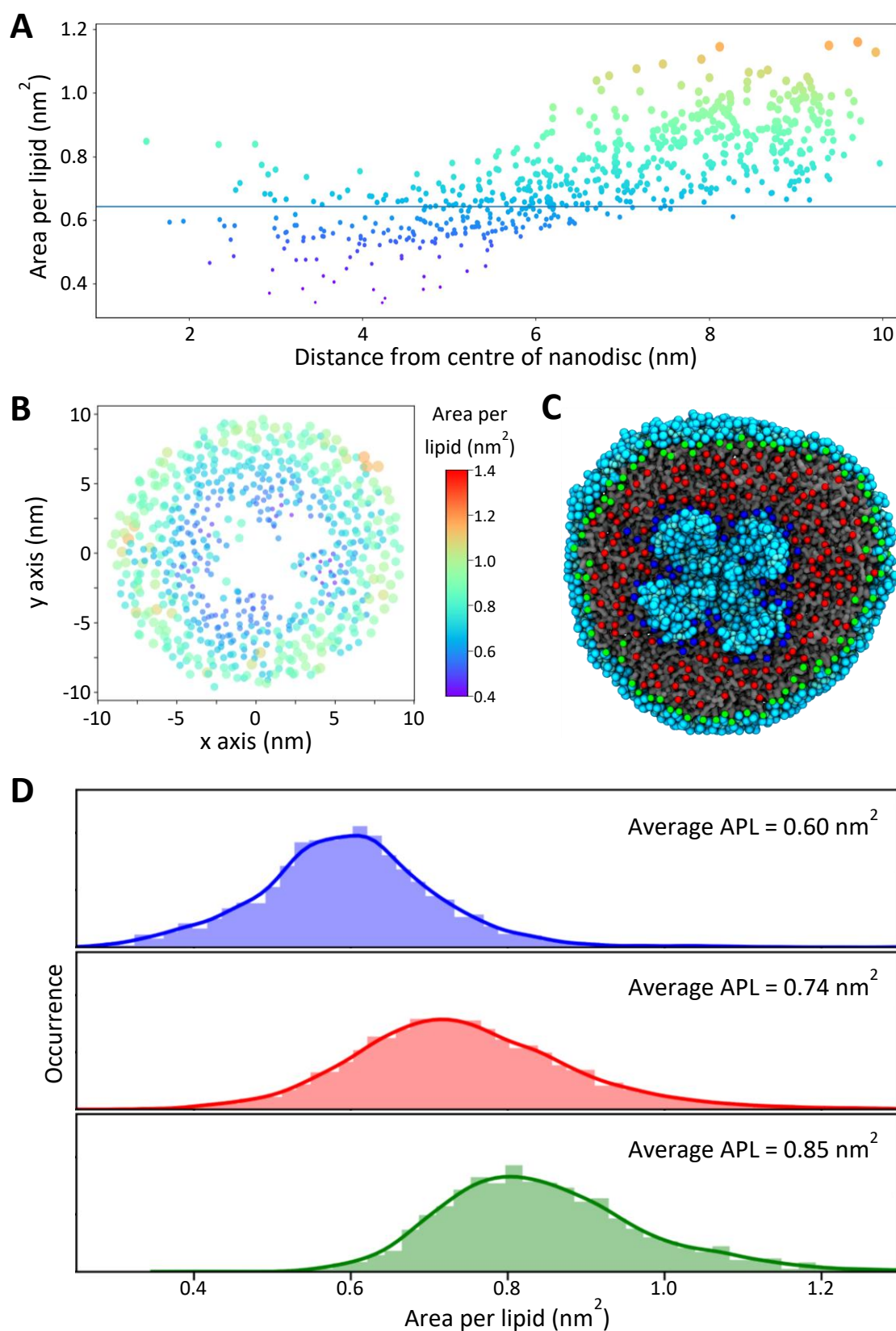


Figure 4.9 – Variation in APL with distance from the centre of a nanodisc with an embedded membrane protein

Caption overleaf.

A: Radial distribution of APL with distance from the nanodisc centre for the final frame of a 200 ns atomistic simulation. The APL value for POPC from bilayer simulations is 0.63 nm^2 , which is indicated by the blue line. While some lipids of the expected value are found at all distances, those nearest the centre exhibit APLs most similar to the expected value and a larger proportion of those at the periphery have greater than expected APL values, indicating a greater degree of perturbation due to their proximity to the MSPs.

B: The same system when viewed from above the plane of the nanodisc, illustrating the increase in APL with distance from the centre. As was demonstrated for a nanodisc system without a membrane protein in **Figure 3.11**, lipids at the periphery of the nanodisc tend to occupy a larger area than those towards the centre.

C: The lipids in the nanodisc grouped according to their location. The lipids nearest the membrane protein are coloured blue, those nearest the MSPs are coloured green and the remaining lipids are coloured red.

D: Histograms of the APLs of each group described in **C** demonstrate the dependence of lipid properties upon position in the lipid patch. Those nearest the membrane protein (lipids in blue) exhibit APLs which are analogous to those of the same lipid in a bilayer, those nearest the MSPs (lipids in green) have APLs which differ most from those in a bilayer, and intermediate lipids (in red) exhibit APLs which lie between the two other groups.

4.3.3.2 Lipid Deformation

Changes to lipid properties due to the proximity of an MSP could also be investigated by measurement of the degree of lipid head group deformation for lipids in different regions of the nanodisc. This was calculated using previously developed protocols^{9, 245} and is shown in **Figure 4.10**. The regions which exhibited greatest lipid deformation were localised to the edges of the nanodisc, an arrangement which resulted from the requirement of lipids in this region to change their orientation and position in order to form interactions with the rest of the bilayer and the hydrophobic regions of the MSPs.

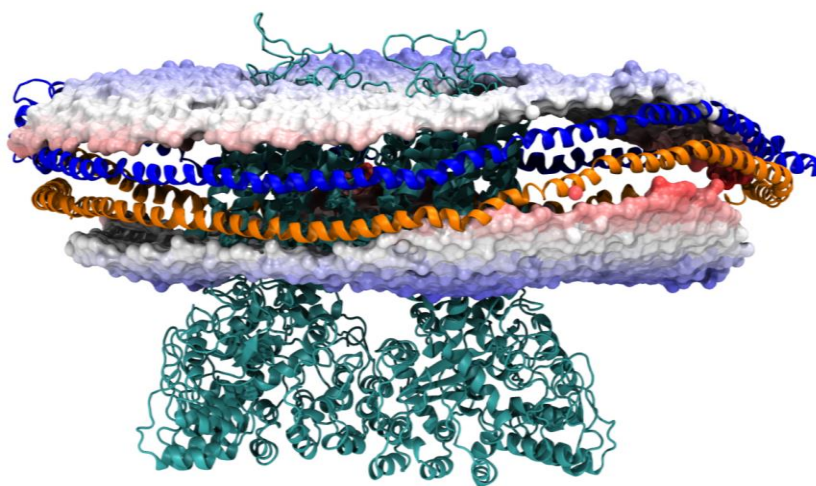


Figure 4.10 – Visual representation of lipid patch deformation

Deformation of the lipids was observed to be greatest at the periphery of the patch, where the headgroups of lipids that are more deformed are coloured in red. This increased deformation results from changes in conformation of peripheral lipids in order to form more favourable interactions between the encircling MSPs and the other nanodisc lipids. This is consistent with the observation that the lipids in closest proximity to the MSPs undergo greatest perturbation of properties relative to the same lipids in a bilayer.

4.3.3.3 Order Parameters for Bilayer and Nanodisc Systems

Simulations of the TRPV6 channel in a mixed lipid patch consisting of POPC, POPE, PIP₂, POPA and cholesterol demonstrated similar trends in the order parameters of lipids as were observed for previously described nanodisc systems containing a single lipid species. The deuterium carbon-carbon bond order parameters for each bond in the acyl chains of the lipids were calculated using `gmx order`¹⁵⁸ on trajectories from the final 25 ns of the atomistic simulations. Contour map analysis was performed using python scripts to group the lipids into brackets containing similar order parameter values.

As was observed for a nanodisc without a membrane protein, shown in **Figure 3.12**, a general decrease in the order parameter of lipids in the system was observed with distance from the centre of the nanodisc, **Figure 4.11 A**.

Results from the order parameter analysis of the same TRPV6 channel in a bilayer environment, **Figure 4.11 B**, demonstrate that the central region of the nanodisc lipid patch exhibited properties which were comparable with those found in a bilayer environment. However, the regions of the periphery of the lipid patch exhibited lower order parameters due to their contact with the surrounding MSPs. The lipids at the periphery of the nanodisc system shown in **Figure 4.11 A** were more disordered than those at the periphery of the nanodisc system without an embedded membrane protein shown in **Figure 3.12 A**. This was possibly a result of the longer acyl chains of the lipids used in this larger system, which have to adopt a more tilted conformation in order to make interactions with the MSP.

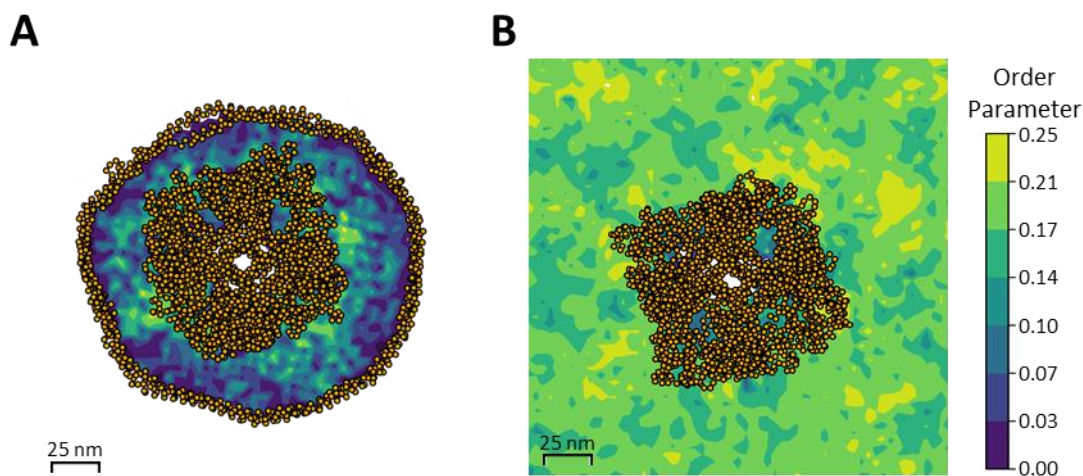


Figure 4.11 – Variations in lipid order in nanodisc and bilayer environments with embedded membrane proteins

A: Average order parameters from the final 25 ns of the atomistic simulation, plotted as a contour map. Higher lipid order parameters are demonstrated close to the nanodisc centre, where the membrane protein is embedded. The order parameter decreases with increased distance from the nanodisc centre.

B: The same analysis performed on a bilayer system. Lipid order throughout is similar to that at the nanodisc centre, without the perturbation of lipids that was noted at the periphery due to interaction with the MSPs.

4.3.3.4 Root Mean Square Fluctuation Analysis for TRPV6

These observations suggest that the trends in lipid properties observed for nanodiscs which do not contain a membrane protein are also found in those systems which do contain membrane proteins, including systems with a range of different sized MSPs. While the lipids in close proximity to the MSPs are perturbed, lipids which are not present at the interface between the lipid patch and the MSPs demonstrate properties which are closer to those expected from bilayer environments. The incorporation of a membrane protein which is able to exist in this region is therefore more likely to enable this protein to experience bulk bilayer properties and lipid interactions that would be observed in bilayer systems.

One metric that can be used to assess if the dynamics of the protein residues are similar in the two environments is the RMSF function. This provides information about the flexibility and dynamics of a given residue over the course of the simulation. These were calculated and averaged over the TRPV6 monomers. The RMSF plot indicated that the properties and dynamics of membrane proteins remain consistent between nanodisc and bilayer simulations. As shown in **Figure 4.12**, residues of the TRPV6 channel retained similar degrees of flexibility in nanodisc and bilayer environments, particularly in the TM regions of the protein. The RMSF of the intracellular regions of the channel was slightly higher in nanodiscs than in the bilayer simulations, suggesting that the non-membrane regions of the channel in a nanodisc have higher degrees of flexibility than the same regions in a membrane.

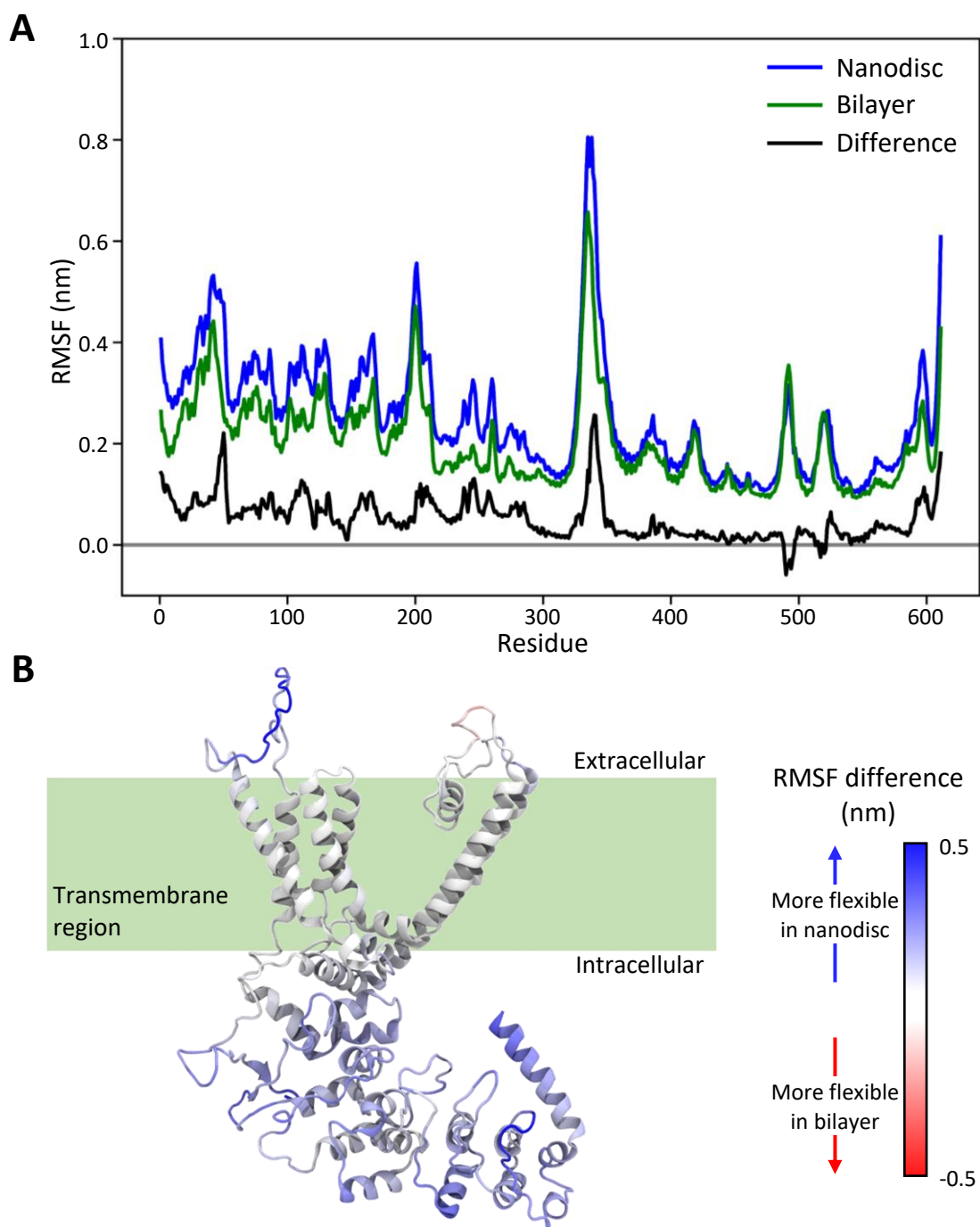


Figure 4.12 – RMSF plots for TRPV6 in nanodisc and bilayer environments

A: The RMSF for the TRPV6 monomers shows that in most regions a generally higher RMSF is observed in nanodiscs when compared to a bilayer environment. However, in TM regions (residues 329 to 611) there are similar RMSF values in the two environments.

B: The residues in the TRPV6 monomer are coloured based on the difference between RMSF values in the two environments. Residues in the intracellular and extracellular regions tend to show slightly higher RMSF values in nanodisc environments, indicating a higher degree of flexibility than the same residues in a bilayer. The TM regions have similar RMSF values in both environments.

4.3.4 Protein-Lipid Interactions

4.3.4.1 Protein-Lipid Interactions in Nanodiscs and Bilayers

The PyLipID software package was utilised to investigate the interactions of the constituent lipids in both nanodisc and bilayer environments, thereby determining if the presence of constraining MSPs affects the interactions of a membrane protein with its lipid environment. The analysis was performed on the 5 μ s CG trajectories on a per residue basis.

The occupancy of lipids, defined as the percentage of frames in which a lipid of a given species was within a specified distance of a binding site, was the same for both nanodisc and bilayer simulations. For example, as shown by the violin plots in **Figure 4.13 A**, in the case of the glutamate transporter homologue GltTk (PDB ID 6XWR) the median values for lipid occupancy for the three lipid species, POPG, POPC and POPE, showed similarity in the two environments; for those in nanodiscs the values were 10.74%, 7.53% and 21.75% respectively, compared to bilayer values of 10.30%, 7.94% and 18.12%.

This consistency in lipid occupancy in the two environments suggests that the change of environment resulting from the presence of the MSPs constraining the lipid patch had little effect on the occupancy of lipid binding to a membrane protein.

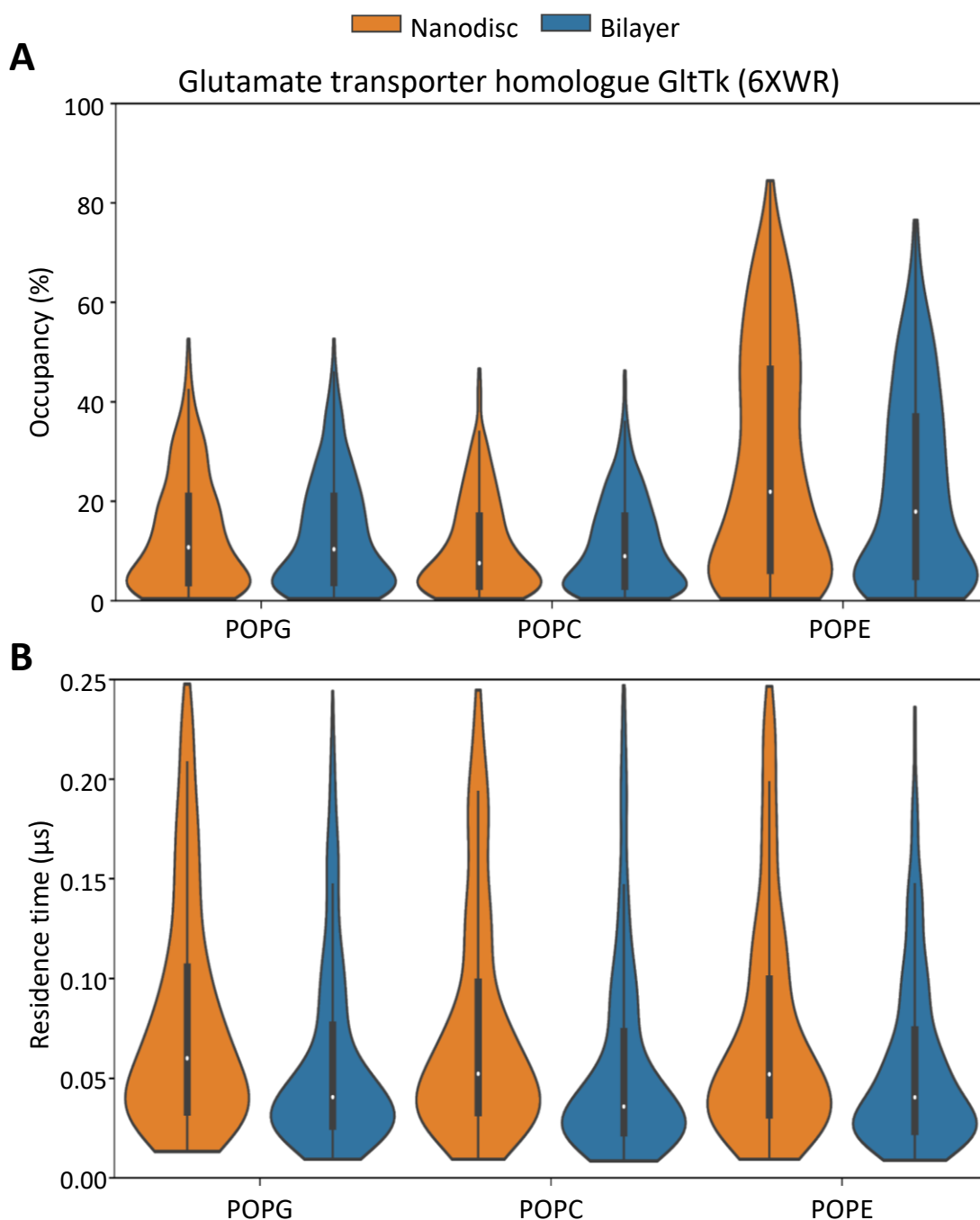


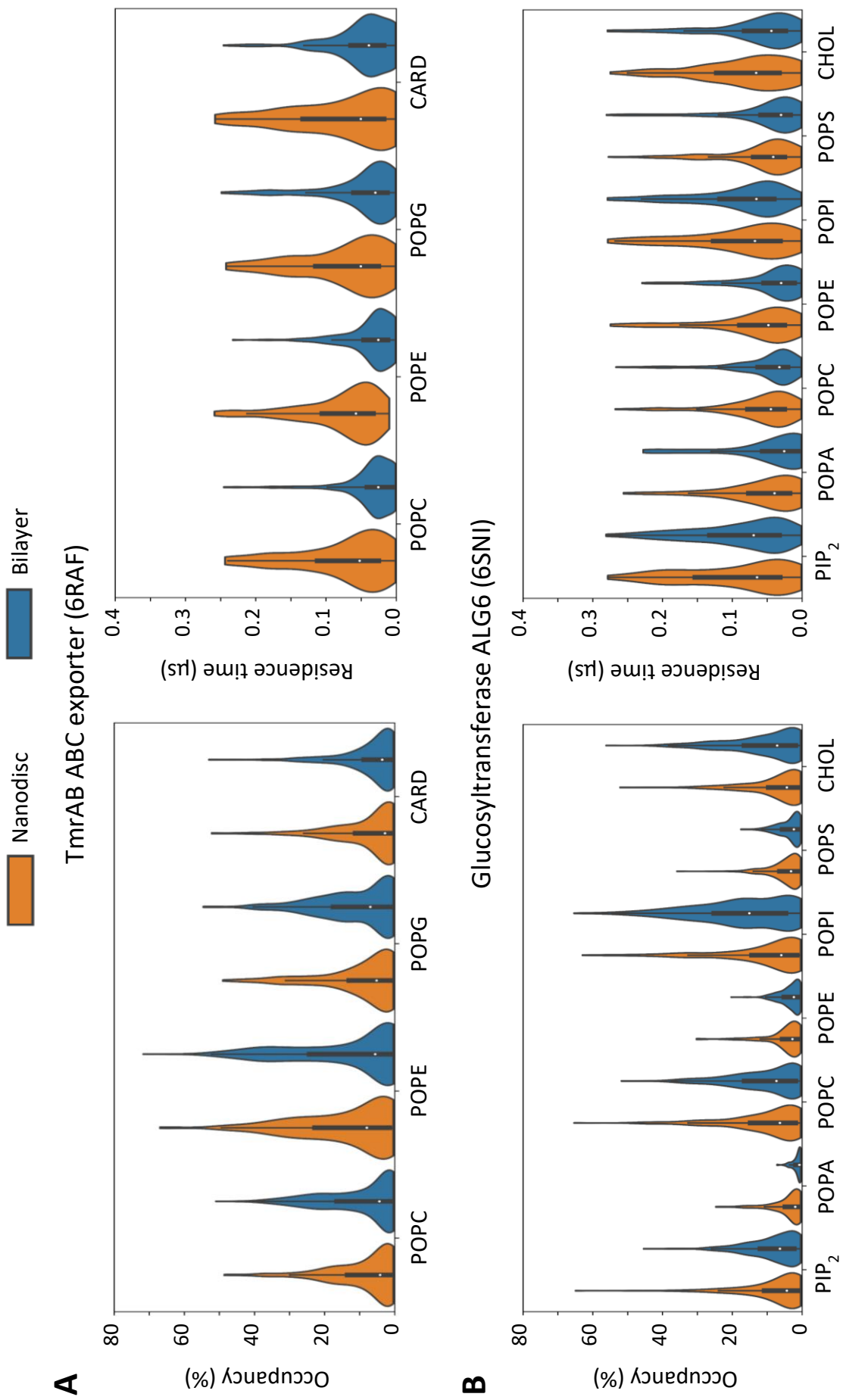
Figure 4.13 – Occupancy and residence times of three lipid species during nanodisc and bilayer simulations with embedded GltTk

The interactions of lipids in nanodisc and bilayer environments with the same membrane protein, the glutamate transporter homologue GltTk, as calculated using the PyLipID software package. The quartiles Q1 and Q3 are the extremes of the black area in the centre of each box and whisker plot, with Q2 shown in white.

A: The occupancy for the three lipid species remains similar despite variation in environment, shown by overlap in the quartiles.

B: The presence of the constraining MSPs reduces exchange of lipids with the bulk population once contact with the membrane protein occurs, increasing the residence times in nanodisc systems.

The residence time can give additional information about the strength of interactions between a lipid and the membrane protein, including the dynamics of bound lipids. While it is not a directly measured parameter, it is calculated using the reciprocal of k_{off} , which is an indication of the tendency of a lipid to dissociate from the membrane protein. As shown in **Figure 4.13 B**, the residence times of lipids in contact with individual residues varied with the type of environment in which the membrane protein was simulated. The median residence time in the nanodisc simulations was greater than that in bilayer simulations for the three lipid species simulated; POPG, POPC and POPE. For nanodiscs the values were 0.060 μs , 0.052 μs and 0.053 μs respectively, compared to bilayer values of 0.041 μs , 0.036 μs and 0.040 μs . This difference in residence times for lipids in a nanodisc compared to those in a bilayer was consistent for a variety of membrane proteins and lipid environments, as shown in **Figure 4.14**.



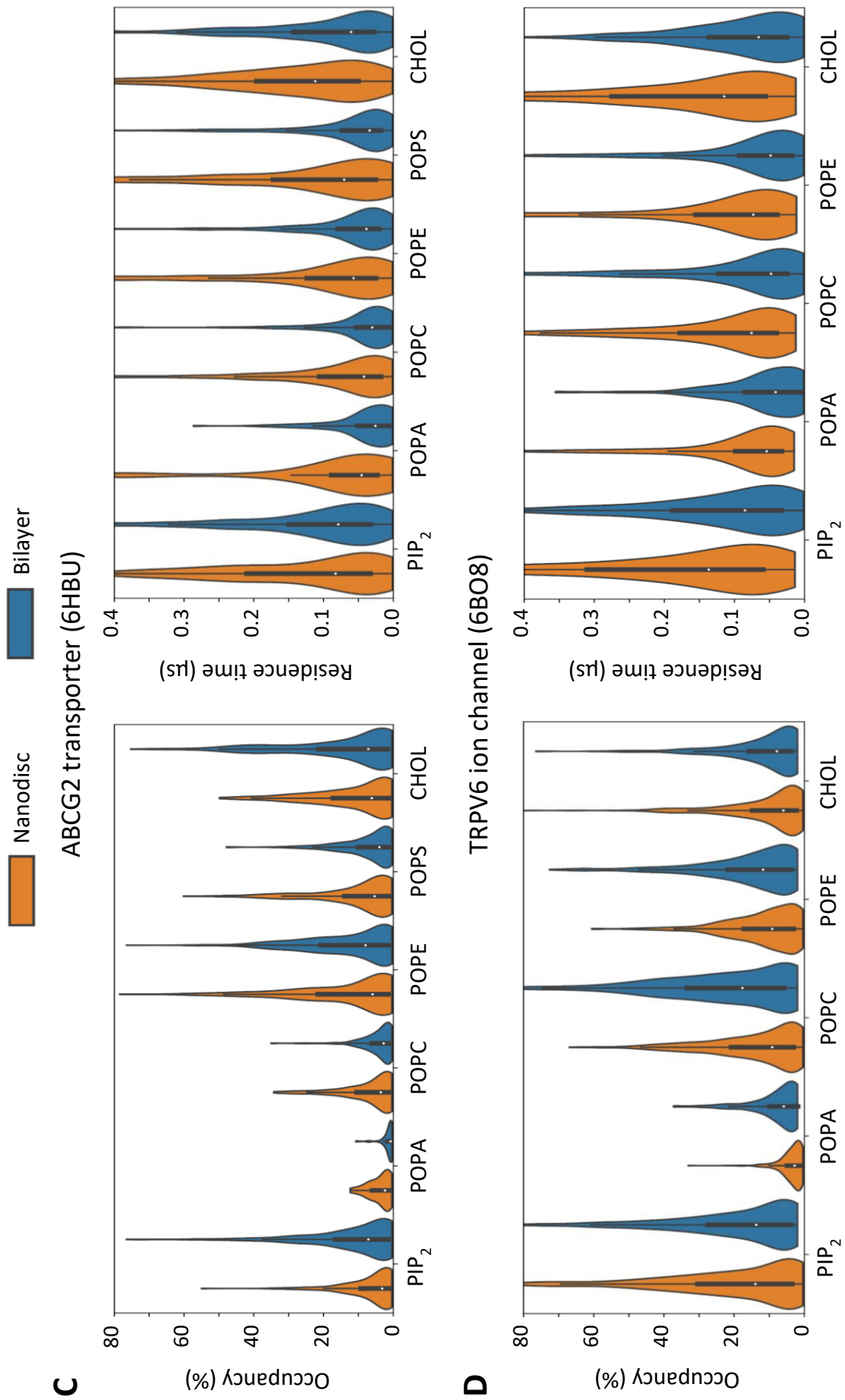


Figure 4.14 – Occupancy and residence times of lipid species during nanodisc and bilayer simulations with embedded membrane proteins

*The PyLipID software package was used to investigate the differences between the protein-lipid interactions that occur in nanodisc and bilayer environments. The lipid occupancy and residence time plots are shown for four different membrane proteins; **A:** the TmrAB ABC transporter, **B:** the glucosyltransferase ALG6, **C:** the ABCG2 transporter, and **D:** the TRPV6 ion channel. These systems were simulated with different lipid compositions designed to mimic those used during their structural determination.*

*The occupancy of each lipid species is similar in both the nanodisc and bilayer environments for all proteins simulated. However, the variation in residence time with environment, which was demonstrated with one protein in **Figure 4.13**, is also present in the four membrane proteins displayed here. This indicates that the presence of encircling MSPs affects the residence time of the lipids in a system, resulting in reduced exchange of lipids with the bulk population once they contact the membrane protein.*

The diameter of the constituent MSPs did not affect the trend in residence times and lipid occupancies. Lipids which did not have PO tails, such as cholesterol, also demonstrated higher residence times in a nanodisc as opposed to a bilayer environment. However, the trend of increased residence time in nanodiscs was not observed in all lipid species. Instead, it was noted that phosphatidylinositol lipids such as POPI and PIP₂ demonstrated less dependence on environment than other lipid species, with residence times that were comparable between the nanodisc and bilayer environments.

The higher residence times of lipids in nanodisc systems indicates that once the lipids contact membrane protein residues, they are slower to dissociate. This would slow the exchange of annular lipids with the bulk lipids surrounding them. As the membrane protein was present in both the nanodisc and bilayer simulations, and this increase in residence times was observed only in the nanodisc systems, it is likely to be a result of the presence of the MSPs

reducing the rate of diffusion within the lipid patch, thus increasing the time required for the annular lipids to dissociate from the membrane protein.

The higher lipid residence times in nanodisc systems could have an impact upon the behaviour of the membrane protein, particularly if a bound lipid is required for the activity or function of a protein. This phenomenon could be of importance for the interpretation of bound lipid species found during structural determination using nanodiscs, as it might promote the binding of lipids to sites of interest on the membrane protein being studied.

4.3.4.2 The Effect of Saturation on Protein-Lipid Interactions

As shown in **Section 4.3.3.3**, the presence of MSPs in a nanodisc results in changes to the order parameters of lipids in the system when compared with the properties of the same lipid species in a bilayer. In order to investigate this effect further, a nanodisc system containing the TmrAB ABC exporter (PDB ID 6RAF) in MSP1D1 was used, and changes were made to the saturation of the acyl chains of the lipids present in the simulation. The same membrane protein and nanodisc were simulated using lipid species which varied by changes to the degree of saturation of their acyl chains, while maintaining the head groups and relative concentrations of these lipids. Three types of lipids were investigated, grouped according to their acyl chains: 1,2-dipalmitoyl (DP) with no double bonds, 1-palmitoyl-2-oleoyl (PO) with one double bond, and 1,2-dioleoyl (DO) with two singly unsaturated acyl chains.

Changes to the degree of lipid saturation affected the ordering of lipids in the bilayer, with an increase in the number of double bonds present in the acyl

chains resulting in greater membrane disorder due to the less efficient packing of the lipids. The results shown in **Figure 4.15** demonstrate that changes to the degree of lipid acyl chain saturation had limited effects on the occupancy of lipid binding sites on the membrane protein, when comparing between a nanodisc and bilayer environment. However, the residence times in these two environments did exhibit a dependence upon the saturation of the lipid tails in the system. An increase in the disorder of the bilayer in a nanodisc promoted an increase in the residence time of lipids present, meaning that in more unsaturated systems the exchange of bound lipids with the rest of the system was reduced.

The observed decrease in the exchange rate of lipids in nanodiscs, as measured by the increase in residence time compared with that in bilayer simulations, could potentially be alleviated through the use of lipids which have more saturated acyl chains.

It can also be observed in **Figure 4.15** that the changes to acyl chain saturation of the lipids also had effects on other constituent components of the bilayer. While the saturation and concentration of cardiolipin was constant in all of the systems, there was variation in cardiolipin residence times with the membrane protein in the nanodisc simulations; it exhibited an increase in residence times in nanodiscs which had more unsaturated lipids.

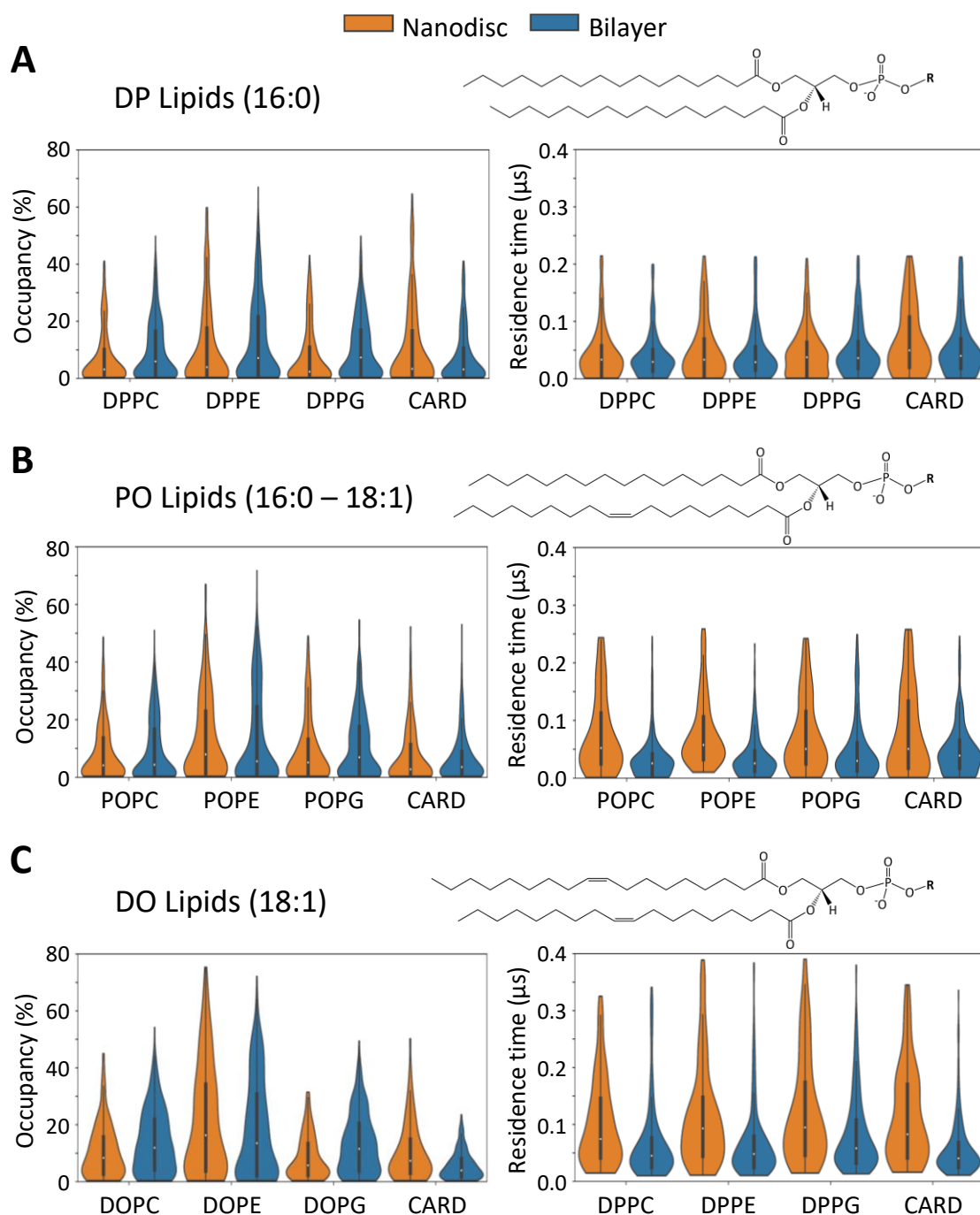


Figure 4.15 – Effect of lipid saturation on occupancy and residence times during nanodisc and bilayer simulations with a TmrAB ABC exporter

The head groups and relative concentrations of the lipids were constant, while the degree of saturation was systematically modified, with the exception of cardiolipin which was kept constant. The occupancies of the three lipid types were similar in the two environments, though there was a trend in residence times with saturation.

A: Saturated lipids had the lowest residence times of the three groups.

B: Lipids with a double bond in one of their chains had increased residence times.

C: Lipids with a double bond in both of their acyl chains showed the highest residence times of the three groups.

4.4 Conclusions

The increasing utilisation of nanodiscs for solubilisation of membrane proteins in preparation for structural determination necessitates an understanding of these systems and the effects that result from the constraining of a small region of lipid bilayer within a nanodisc. An MD approach is perhaps uniquely positioned to be able to provide insights into the properties of the membrane proteins as well as the constituent lipids that make up a nanodisc and how these vary in comparison with membrane proteins and lipids in a bilayer environment.

The pipeline approach described here allows simulation of nanodiscs containing membrane proteins in an automated manner. The composition of the nanodiscs created by this pipeline is customisable, enabling the simulation of different membrane proteins in multiple lipid patch configurations, constrained by a variety of MSPs. This allows replication of experimental systems used to solubilise membrane proteins, which can then be compared to bilayer simulations with the same lipid composition and membrane protein. This facilitates investigation of differences in both membrane protein and lipid properties resulting from the change of environment experienced by membrane proteins when embedded into a nanodisc.

4.4.1 Membrane Protein-MSP Interactions

As a result of the computational approach employed here, the properties of individual lipids can be assessed in a manner which is inaccessible to experimental approaches. Another advantage of the computational approach

is the ability to observe events which may not be detected during the class averaging employed in the cryo-EM imaging of the nanodisc systems, such as the occurrence of the collision between the solubilised membrane protein and the boundary MSPs in some nanodiscs.

This effect was observed in both atomistic detail and CG detail, including both Martini 2 and Martini 3, as well as in systems constructed using differing lengths of MSP, suggesting that this is unlikely to simply be an artefact resulting from a particular simulation methodology. The interaction between the membrane protein and the MSPs did not appear to be transient, since over the course of the simulations in which this behaviour was identified, dissociation of the membrane protein from the MSPs was not observed.

The region of the membrane protein which interacted with the MSPs varied in different repeats suggesting that it was not being driven by a particular region on the membrane protein and instead was a result of the hydrophobic interactions between the two proteins. This interaction between the membrane protein and the MSPs also has implications for membrane protein-lipid interactions, as the presence of the MSPs can shield regions of the membrane protein from interactions with the lipids of the nanodisc bilayer. In addition to this, other lipid interactions can be elevated resulting from small numbers of lipid species being trapped in between the membrane protein and MSPs.

The choice of nanodisc size should also be considered. For example, smaller diameter MSPs, such as MSP1D1, more commonly produced interactions between the MSPs and the membrane protein. The collision of these two

protein components of the nanodisc system would often occur near the beginning of a trajectory and did not display dissociation over the course of the simulation, though this may be shown in longer simulation trajectories.

4.4.2 Lipid Properties

The confined environment resulting from the presence of the surrounding MSPs on a lipid patch in a nanodisc can have an impact upon the properties of the lipids present. One metric used to measure this effect was the APL. As was discussed in **Chapter 3**, the presence of the MSPs altered the conformation and properties of the lipids present in a nanodisc system, an effect which was also observed in nanodiscs which contained a membrane protein. To demonstrate this behaviour, the TRPV6 membrane protein was embedded into a nanodisc containing a single lipid species, POPC, and the APL of the lipids present in the nanodisc was compared with bilayer simulations. Upon investigation of the APL of lipids in the nanodisc system as a function of their distance from the centre, a region was identified near the centre of the nanodisc where lipids exhibited bilayer-like APLs. However, as was observed in the simulations of nanodiscs performed in the absence of a membrane protein, lipids which were located near the periphery of the nanodisc exhibited APLs which differed more from the values that would be expected in a bilayer. A general increase in APL was observed as a function of distance from the nanodisc centre. This suggests that in order for the lipids in these regions to form favourable interactions with the MSPs, their conformations must be different to those of the same lipid species in a bilayer.

As was observed in **Chapter 3** for nanodiscs that do not contain a membrane protein, the perturbation of lipid properties as a result of the proximity to the constraining MSPs can also be observed in the order parameters of lipids in the nanodisc. The lipids present near the centre of a nanodisc exhibited order parameters which were analogous to those observed for lipids in a bilayer system. In both environments minor fluctuations in the order parameters were observed depending upon the lipid make up in a given region of the membrane. However, with increasing distance from the centre of the nanodisc there was a decrease in the order parameters measured. This resulted from the necessity of lipids in this region to adopt an unusual conformation in order to form favourable interactions with the MSPs. Since the order parameter provides information about the orientation of the acyl chains with respect to the membrane normal, this suggests that the lipids which are in close proximity to the MSPs deviate more from the membrane normal than lipids near the centre of the nanodisc. A similar dependence of lipid properties on distance from the centre of the nanodisc was observed for membrane curvature; the lipids at the edges of the nanodisc displayed greater deformation due to their interactions with the surrounding MSPs.

The changes in lipid properties that result from the constraining of the membrane by the MSPs is of interest for the use of nanodiscs in structural determination and the characterisation of membrane proteins. MSPs which have smaller diameters have been preferred for structural studies of membrane proteins, as this allows more of the recorded signal to be produced by the

membrane protein of interest rather than the MSPs and lipids. However, the changes which occur to the lipids in the nanodisc system should be considered, particularly if bound lipids and lipid interactions are to be observed in the structure.

4.4.3 Protein Properties and Protein-Lipid Interactions

Despite the changes which occurred to the lipids present in the nanodisc system, particularly those near the MSPs, the conformation adopted by membrane proteins in both nanodisc in bilayer environments appeared to be consistent as measured by the RMSF of the proteins. In both environments, similar fluctuations were observed along the protein structures and the TM regions in particular demonstrated near identical RMSFs. This suggests that the structural flexibility of a membrane protein that has been embedded into the lipid patch of a nanodisc exhibits similar properties when instead embedded into a bilayer. While the RMSF calculations did not indicate membrane protein instability over the course of the 200 ns simulations, experimental studies often are conducted on significantly greater timescales and longer simulations would therefore be required in order to confirm protein stability in these environments.

Interaction of the membrane protein with the surrounding lipids in the nanodisc was investigated using the PyLipID software package and compared with the same lipid composition in a bilayer environment. The analysis showed that the protein-lipid interactions that occurred in both environment types displayed analogous lipid binding occupancies to the membrane protein,

showing that the percentage of frames during which the membrane protein was in contact with a given lipid species was consistent between the two environments. This behaviour was observed for different membrane proteins in different sized nanodiscs, using a variety of lipid species. However, differences were observed between the two environment types when the residence time of the lipids was investigated. For the majority of lipid species, it was observed that the residence times in nanodiscs were higher than those in a bilayer environment. This difference in residence times between the two environments indicates that the rate of lipid exchange that occurs in nanodiscs is slower than that in a bilayer environment. However, some lipid species did not display this behaviour; in particular, phosphatidylinositol lipids seem to exhibit analogous residence times in both environments.

By making changes to the saturation of the acyl chains of lipids present in a nanodisc system with the membrane protein TmrAB ABC exporter the effect of the ordering of lipids on the membrane protein-lipid interactions could be investigated. The acyl chains used were DP, PO and DO, while the head groups used were the same as in the previous simulations. Cardiolipin was also present in these systems, though its degree of saturation was not altered in order to investigate if changes to the lipid saturation of the other lipids in the bilayer affected its interactions with the membrane protein.

As for the cases where different membrane proteins were simulated, the lipid occupancy with relation to the membrane protein remained constant in the systems created with differing lipid tail saturations. However, an increase

in residence times was observed with increasing unsaturation. This observation demonstrates the role of lipid conformation and ordering with relation to the interaction of the membrane protein with its surrounding lipids in a nanodisc system, and this could be of note for structural analysis. Studies using lipids of differing degrees of acyl chain saturation could therefore assist in the detection of binding sites on membrane proteins.

The observations reported here demonstrate that the utilisation of nanodiscs is a valid approach for the structural determination and study of membrane proteins in environment which is analogous to the physiological bilayer from which the membrane proteins originate. This is of benefit for studies aiming to understand the physiological conformation and interactions that these membrane proteins make within cells; the protein structure appears to be unaffected by incorporation into a nanodisc. However, differences are observed in the interactions which occur between the membrane protein and lipid species present, and these differences could be utilised in order to promote detection of lipid binding sites on membrane proteins. Despite the differences in binding characteristics, the sites at which these membrane protein-lipid interactions occur are unaffected, suggesting that studies using nanodiscs are suitable for the discovery of lipid binding sites which may be utilised for structure-based drug design as well as the identification of allosteric lipid regulators of membrane proteins, as will be discussed in **Chapter 5**.

5

Chapter 5 - Interactions Between Cholesterol and G-Protein-Coupled Receptors

Contents:

5.1 Introduction	153
5.1.1 G-Protein-Coupled Receptors.....	154
5.1.1.1 The Activation of GPCRs.....	156
5.1.2 GPCR Interactions with Lipophilic Ligands	157
5.1.2.1 The TM345 Site on Free Fatty Acid Receptor 1	158
5.1.2.2 GPCR Interactions with Cholesterol	159
5.2 Methods	162
5.2.1 Coarse-Grained Simulations	162
5.2.2 Analysis.....	162
5.3 Results.....	164
5.3.1 Cholesterol Interaction with the TM345 Site on FFAR1	164
5.3.1.1 Residues of Interest in the TM345 Binding Site on FFAR1	167
5.3.2 Investigation of the TM345 Site on Additional GPCRs	170
5.3.2.1 General Properties of the TM345 Binding Site	183
5.3.2.2 The TM345 Binding Site on ADRB2.....	188
5.3.2.3 TM345 Binding Site Specificity	191
5.3.2.4 Comparison of Residence Times and GPCR State.....	192
5.3.2.5 Protein Conformation and Cholesterol Binding	194
5.4 Conclusions	196
5.4.1 Cholesterol Binding Variation within Class A GPCRs	199
5.4.2 Cholesterol Binding Specificity at the TM345 Site.....	201
5.4.3 Cholesterol Binding Variation with Conformational State	201

5.1 Introduction

Membrane proteins perform a wide range of roles within the body as transporters, channels, enzymes and receptors. The lipid bilayer in which they are embedded affects their function in two main ways¹²⁰: firstly through variation in the overall properties of the bilayer, such as thickness and fluidity, and secondly through more specific interactions with individual lipids. The latter form of interaction can enable allosteric regulation of a particular membrane protein at a given site, and thus investigation of these specific interactions can be utilised for the identification of potentially druggable sites.

Cellular membranes contain a wide variety of lipid species; over 47,000 different biologically relevant lipids have been catalogued by the LIPID MAPS Structure Database³. The majority of lipids within a bilayer behave somewhat as a solvent in which the intrinsic membrane proteins reside. These are termed 'annular lipids'⁷¹ and while they do not form specific interactions with the membrane protein, variation in this group of lipids can affect the membrane protein by changing the overall properties of the bilayer. Two examples of this are the changes in membrane fluidity which arise from different lipid compositions and levels of saturation in the lipid hydrocarbon chains, and changes in membrane thickness as a function of hydrocarbon chain length.

In addition to the non-specific effects of the annular lipids on membrane proteins, specific interactions may occur between the protein and a subset of the bilayer lipids. These 'non-annular lipids' are able to function in a manner similar to that of a co-factor or allosteric regulator⁷¹. Specific lipid interaction

sites have been detected on multiple membrane proteins, including ion channels²¹⁰ and receptors²⁵³. Computational approaches have also been utilised to identify further specific lipid interaction sites²⁵⁴⁻²⁵⁵. Typically, the lipids are able to associate between the TM helices of a membrane protein or at the interface of two membrane proteins in a multimeric protein²⁵⁶. Specific protein-lipid interactions are able to affect protein function by promotion of protein oligomerisation¹²⁸, as well as by allosteric regulation^{120, 236}; the regulation of a protein by ligand binding at a location which is not the active site of the protein²⁵⁷. This regulation is usually achieved by a conformational change in the protein caused by the binding of the allosteric regulator.

5.1.1 G-Protein-Coupled Receptors

GPCRs are the largest receptor superfamily found in eukaryotes and the human genome encodes approximately 800 GPCRs²⁵⁸. This family of integral membrane proteins has a highly conserved seven TM α -helical core; the seven TM helices are connected via three extracellular loops (ECL1-3) and three intracellular loops (ICL1-3), as shown in **Figure 5.1**. There is also an additional shorter eighth helix (H8), which is an intracellular amphipathic helix with an orientation perpendicular to that of the other TM helices. This H8 is thought to be present in most GPCR structures and has been resolved in many of them²⁵⁹.

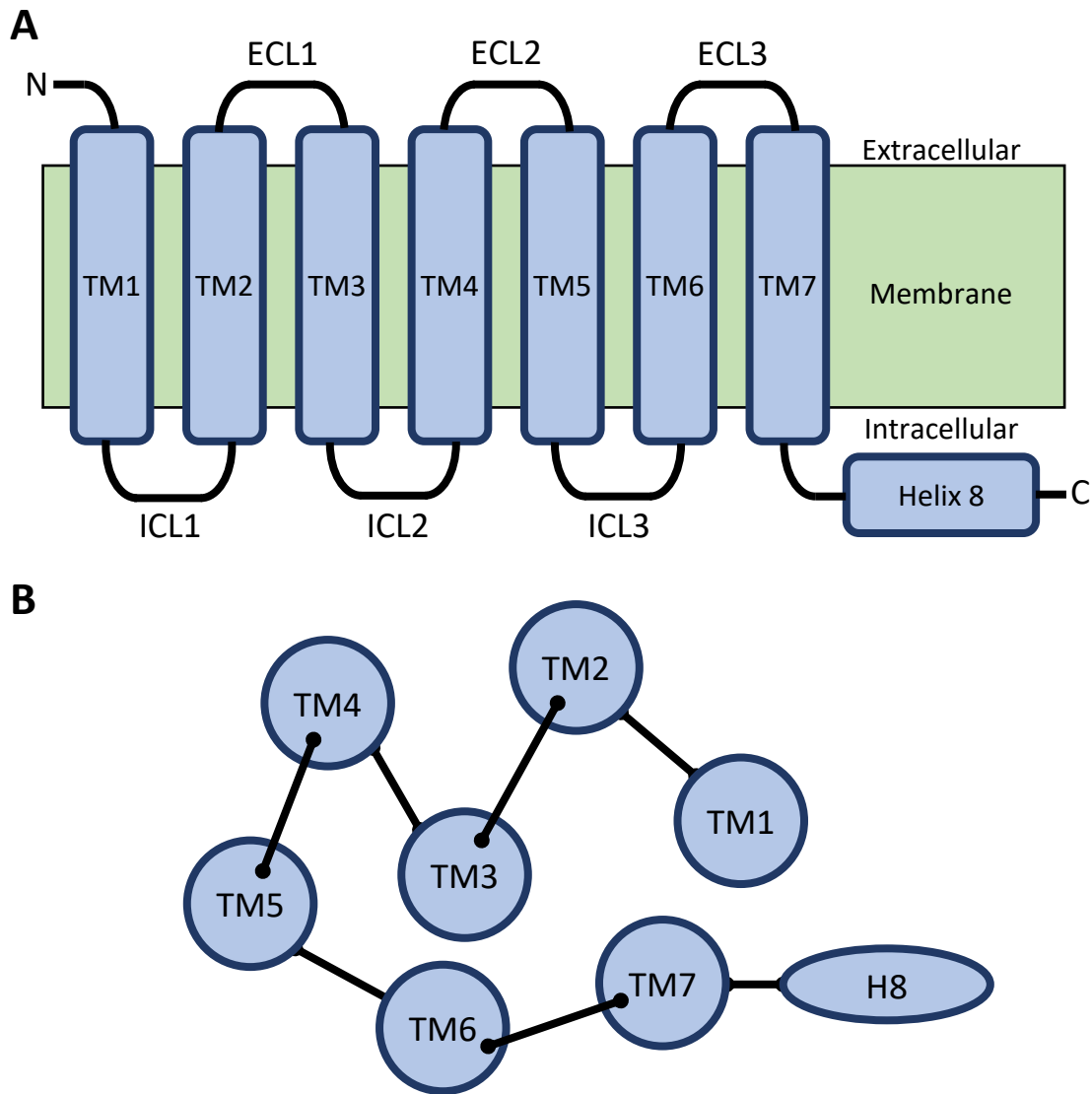


Figure 5.1 – Schematic representation of the structure of a GPCR

A: A lateral view of a GPCR displaying the 7 TM helices, C-terminal Helix 8 (H8), intracellular loops (ICL) and extracellular loops (ECL).

B: A top-down view of a GPCR displaying the relative positions of the TM helices as well as the position of H8, as viewed from the extracellular side of the membrane.

Vertebrate GPCRs can be classified into five main groupings: the Rhodopsin, Secretin, Glutamate, Frizzled and Adhesion groups²⁶⁰⁻²⁶¹. This research is confined to the study of the Rhodopsin group, known as Class A, and the Secretin group, known as Class B²⁶². Approximately 85% of the GPCRs found in humans belong to Class A. GPCRs perform a wide variety of functions

within the body, signalling in response to various stimuli including hormones, ions, photons, neurotransmitters and odorants ²⁶³. As a result, GPCRs have many roles in both physiology and pathology; it has been estimated that 30 to 40% of marketed drugs target GPCRs ²⁶⁴.

The Ballesteros–Weinstein GPCR numbering scheme ²⁶⁵ may be used to refer to residues within the TM helices of Class A GPCRs, by assigning a decimal number to each residue in the TM regions of a GPCR. The integer part denotes number of the TM helix while the fractional part denotes the position of the amino acid along the helix. The most evolutionarily conserved residue is assigned the number 50, with residues either side being assigned numbers sequentially radiating from this highly conserved residue. For example, the tyrosine of the highly conserved DRY motif ²⁶⁶ is referred to as 3.51, with the '3' referring to its position on TM3, and the '.51' indicating its position one residue along from the most conserved residue on that helix. The same is true for the numbering of residues on the ECLs and ICLs, with the integer part of the number indicating which two helices the loop lies between. For example, the most highly conserved residue on ICL2 would be indicated by 34.50. The Ballesteros-Weinstein numbering scheme has been adapted for use in Class B GPCRs as the Wootten numbering scheme ²⁶⁷, which uses the same principle but with unique reference positions.

5.1.1.1 The Activation of GPCRs

Heterotrimeric G-proteins associate with GPCRs and consist of the subunits $G\alpha$, $G\beta$ and $G\gamma$. These G-proteins are GTPases and so are able to catalyse the

hydrolysis of GTP into GDP. The canonical activation of GPCRs involves three main stages: (i) the binding of a ligand to the GPCR results in the stabilisation of the receptor in an active conformation, which in Class A GPCRs causes the breaking of the 'ionic lock', a highly conserved salt bridge formed between the arginine of the TM3 DRY motif and a negatively charged residue at position 6.30²⁶⁸; (ii) GDP dissociates from the $G\alpha$ subunit of the G-protein and is then exchanged for GTP; (iii) the G-protein disassociates into two membrane-anchored subunits $G\alpha$ -GTP and $G\beta\gamma$. Signalling can also occur through β -Arrestin mediated pathways. GPCRs are able to adopt multiple active and inactive conformations which can be stabilised by interaction of the GPCR with various ligands, including the lipids from their environment²³⁶. In some cases, GPCRs can signal in the absence of a ligand and their activity is regulated through binding of inverse agonists to reduce their signalling²⁶⁹.

5.1.2 GPCR Interactions with Lipophilic Ligands

The binding of several lipophilic ligands to GPCRs has been documented, with lipophilic molecules able to bind both within the helix bundle and at the receptor-lipid interface. The most widely studied GPCR, rhodopsin, for example, is covalently bound to a lipophilic 11-cis retinal molecule via Lys296. When rhodopsin is activated, the retinal undergoes photoisomerisation to the trans-configuration and is released from the GPCR, where it is replaced by a new cis-retinal molecule²⁷⁰. Another example is the Sphingosine-1-Phosphate Receptor (S1PR1), which binds lipophilic ligands at an elongated amphipathic

binding pocket between TM1 and TM7²⁷¹ and is able to select for sphingosine lipids.

5.1.2.1 The TM345 Site on Free Fatty Acid Receptor 1

Free Fatty Acid Receptor 1 (FFAR1) plays an important role in glucose dependent insulin secretion and is therefore of interest for studies on type II diabetes. A 3.2 Å crystal structure of FFAR1 has been reported with a bound lipophilic allosteric agonist, AgoPAM compound AP8²⁷², in a cleft between TM3, TM4 and TM5, and also making contact with ICL2. The location of this binding site is shown in **Figure 5.2 A**. AP8 enters this binding site via the membrane²⁷³ and it has been suggested that this TM345 site could be occupied physiologically by cholesterol, with which AP8 shares structural similarities.

Computational approaches have suggested that the TM345 site which has been observed in FFAR1 might also be found in other GPCRs. Pharmacophore screening of the adenosine A_{2A} receptor (AA2AR) for example, using CG probes, suggests the presence of a hydrophobic binding pocket at the intersection of these three helices: TM3, TM4 and TM5. This binding site is shown in **Figure 5.2 B**, and its location would correspond to the AP8 site identified on FFAR1. The same site on AA2AR was subsequently identified using CG simulations¹⁰⁹, as shown in **Figure 5.2 C**.

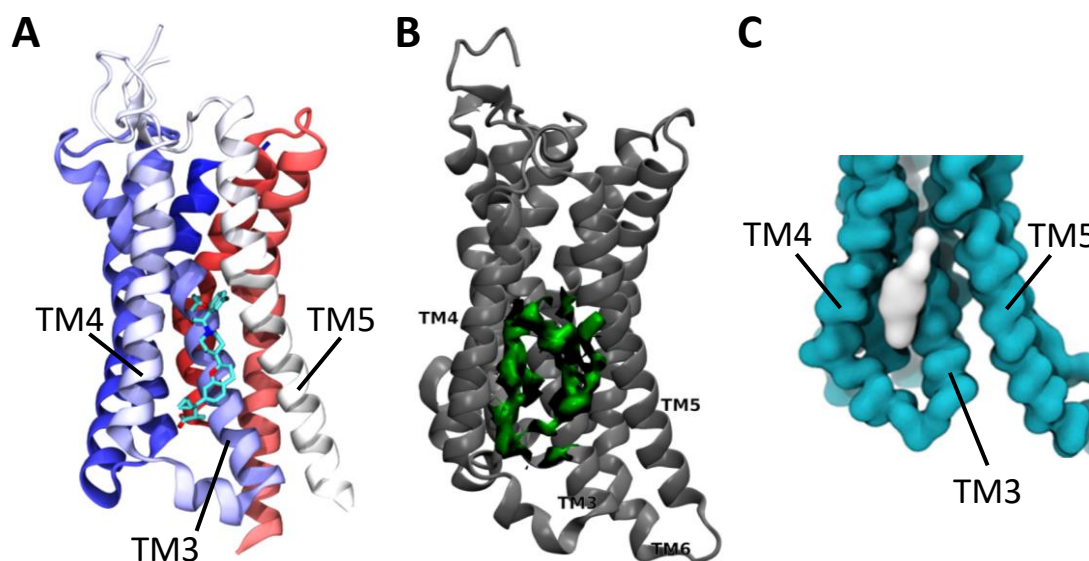


Figure 5.2 – The TM345 binding site

A: The AP8 binding site identified on the crystal structure of FFAR1. The allosteric regulator, AP8, is shown in cyan binding at the TM345 site.

B: The hydrophobic TM345 binding site identified on AA2AR using CG pharmacophore screening. Non-polar pharmacophore binding locations are shown in green.

C: The same TM345 site has since been located using CG simulations of AA2AR, shown here as a close-up view of the binding site with bound cholesterol shown in white.

Figure B reproduced from ²⁷⁴ and Figure C adapted from ¹⁰⁹.

5.1.2.2 GPCR Interactions with Cholesterol

Structural determination of GPCRs has identified bound cholesterol in several protein structures ^{123, 275} and cholesterol has also been implicated in regulation of GPCR conformation, stability and function ^{120, 276}. It stabilises the β_2 adrenergic receptor (ADRB2) ²⁷⁷ and acts as an allosteric regulator for multiple members of the GPCR family including the cannabinoid CB1 receptor ²⁷⁸, the cholecystokinin A (CCK1) ²⁷⁹, and the μ -opioid receptor ²⁸⁰. In addition, simulation studies have been used to study the binding affinity of cholesterol to multiple sites on the 5-HT_{1A} receptor ²⁸¹.

Structural determination of ADRB2 led to the proposal of a cholesterol binding motif: the Cholesterol Consensus Motif (CCM) ¹²¹, comprising the side chains of four residues located on TM helices 2, 3 and 4, non-adjacent in the primary sequence (W158, I154, R151, and Y70) and forming a shallow pocket on the intracellular region of the GPCR. Strong interactions occur between the hydroxyl head group of the bound cholesterol and positively charged arginine or lysine residues located on TM4, while hydrophobic interactions occur between the cholesterol and an aromatic residue located on TM4. Sequence alignments performed on other members of the GPCR family have identified the CCM site in up to 44% of the GPCRs in humans ¹²¹. In addition, the cholesterol recognition amino acids consensus (CRAC) domain has also been identified in multiple membrane protein families including cholesterol-dependent GPCRs ²⁸². Localised to a single TM helix, it binds cholesterol longitudinally. A reversed CRAC domain has also been identified: CARC ²⁸³.

Despite the identification of the CCM, CRAC and CARC sequences, a recent study has shown that sequence conservation is not responsible for the widespread binding of cholesterol in the GPCR family ²⁸⁴, with cholesterol binding sites in GPCRs lacking consensus motifs and binding occurring instead at conserved regions without specific sequence conservation. This study investigated nearly 500 structures of GPCRs, solved using X-ray crystallography and cryo-EM, using Consensus Network Analysis ²⁸⁵⁻²⁸⁶ to identify twelve spatially distinct locations of interaction, termed Cholesterol Network Clusters, to which 92% of cholesterol binding could be mapped ²⁸⁴.

This approach was then used to test the binding of cholesterol to the previously identified motifs, but the presence of CRAC, CARC or CCM motifs was not found to be predictive of cholesterol binding in GPCRs.

5.2 Methods

5.2.1 Coarse-Grained Simulations

CG simulations were performed using GROMACS (version 2020²⁸⁷) with the Martini 2.2 force field^{183, 192}. The membrane protein coordinates were obtained from the MemProtMD database⁹⁻¹⁰, following which they were converted to a CG representation using the `martinize.py` script (version 2.4). The Martini ENM was applied to the membrane protein using the default values, including applying a force constant of 500 kJ mol⁻¹ nm⁻² with a 0.7 nm cut-off.

The `insane.py` script was utilised to embed the membrane protein into a CG bilayer containing 35% cholesterol. The systems were then neutralised using counterions and the final NaCl concentration was adjusted to ~0.15 M. Finally, the systems were simulated for 100 ns to allow membrane assembly to occur.

The temperature was maintained at 310 K by the application of the Berendsen thermostat¹⁶⁶. Semi-isotropic pressure coupling was applied for all systems with a pressure of 1 atmosphere, using the Berendsen barostat¹⁶⁶ with a time constant of 1.0 ps. Simulations were performed on each system for 10 μ s, with a step size of 20 fs and three repeats for each system.

5.2.2 Analysis

The PyLipID software package²⁴⁶ was utilised to analyse the binding interactions between the membrane proteins and the lipid ligands. The residence time and occupancies were calculated as described in **Chapter 4**. Lipid occupancy was calculated using the percentage of frames of the trajectory

in which the lipid of interest was within the distance of the lower cut-off from the residues of the membrane protein. Using dual cut-offs of 0.55 nm and 0.8 nm, taken from previously published values ²⁴⁷, residence times were calculated from the time at which a lipid moved closer to the membrane protein than the lower cut-off value, until it moved further away than the upper cut-off value. Data analysis was performed using python scripts and GROMACS utilities ¹⁵⁸. Sequence alignment and phylogenetic analysis of the GPCRs was performed using GPCRdb.org ²⁸⁸⁻²⁸⁹. The aligned sequences were visualised using the ALINE Protein Sequence Alignment Editor ²⁹⁰.

5.3 Results

5.3.1 Cholesterol Interaction with the TM345 Site on FFAR1

A lipophilic binding site at the TM345 cleft in the structure of FFAR1 has been identified which binds an allosteric regulator, AP8. This prompted further study to determine if this site would also bind cholesterol, due to the structural similarities between the two molecules as shown in **Figure 5.3**. In both molecules one terminus is a polar region, and the other is a hydrophobic domain, with a cyclic region in between.

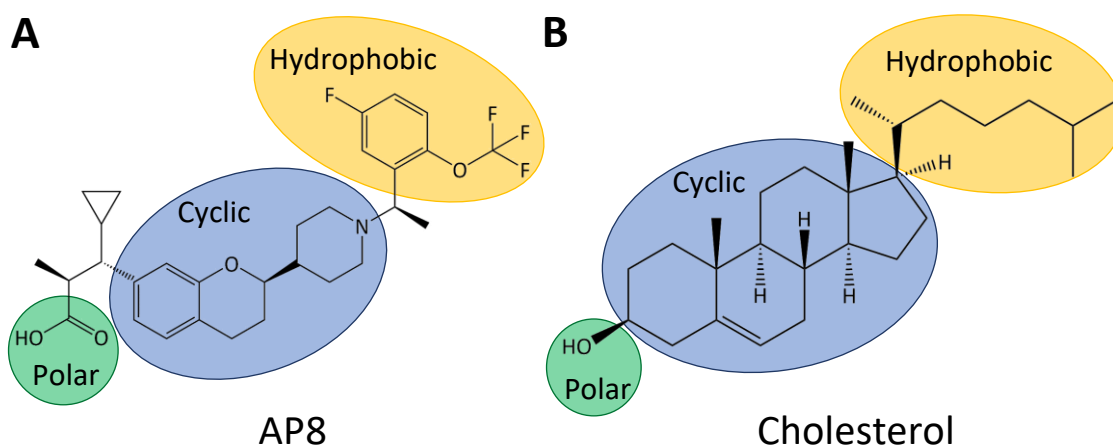


Figure 5.3 – Structure of the allosteric regulator AP8 alongside cholesterol

Similarities can be seen between the structure of AP8 (A) and the structure of cholesterol (B). Both AP8 and cholesterol have a central cyclic region, outlined above in blue. Adjacent to this region in AP8 is a polar carboxylic acid group, while in cholesterol there is a polar hydroxyl group. At the opposite end of the cyclic group there is a hydrophobic region in both molecules. In AP8 this comprises a fluorobenzene to which a trifluoromethyl group is joined by an ether linkage, while in cholesterol it comprises a hydrocarbon chain.

To validate the utilisation of a CG MD approach for identification of cholesterol binding sites, FFAR1 was first investigated to ensure that a lipophilic TM345 binding site could be identified. The FFAR1 structure was converted to a CG representation and simulated in a bilayer containing 35% cholesterol, for three

repeats of 10 μ s each. Following this, analysis was performed using the PyLipID software package and comparison was made to the location of the AP8 binding site shown in **Figure 5.4 A**.

A cholesterol binding site was identified at the TM345 cleft, corresponding to the location of the previously identified AP8 binding site, as shown in **Figure 5.4 B**, where the cholesterol binding site has been overlaid on the experimentally determined structure²⁷². For ease of comparison, the CG cholesterol bound at the TM345 site has been converted to an atomistic representation here. The cholesterol residence time was 2.56 μ s with a surface area of 12.9 nm². In spite of the structural similarities between AP8 and cholesterol, there are some differences in the specifics of the binding site. The position of the hydroxyl group of the cholesterol in the identified binding site occupies a location analogous to that of an oxygen atom in the cyclic ether group in the AP8 molecule, where it makes similar interactions with the GPCR. The cyclic region and the hydrocarbon chain of the bound cholesterol occupy a corresponding position in the binding site to that occupied by the majority of the AP8 molecule, with the exception of the cyclopropane and carboxylic acid groups at its terminal end. This supports the suggestion that the physiological ligand for this FFAR1 binding site is cholesterol. AP8 might be able to displace a bound cholesterol due to an ability to form interactions with regions of FFAR1 which are beyond the reach of cholesterol. In particular, cyclopropane and carboxylic acid groups of the AP8 allow it to form contacts with charged and aromatic residues at the base of TM4 and in ICL2.

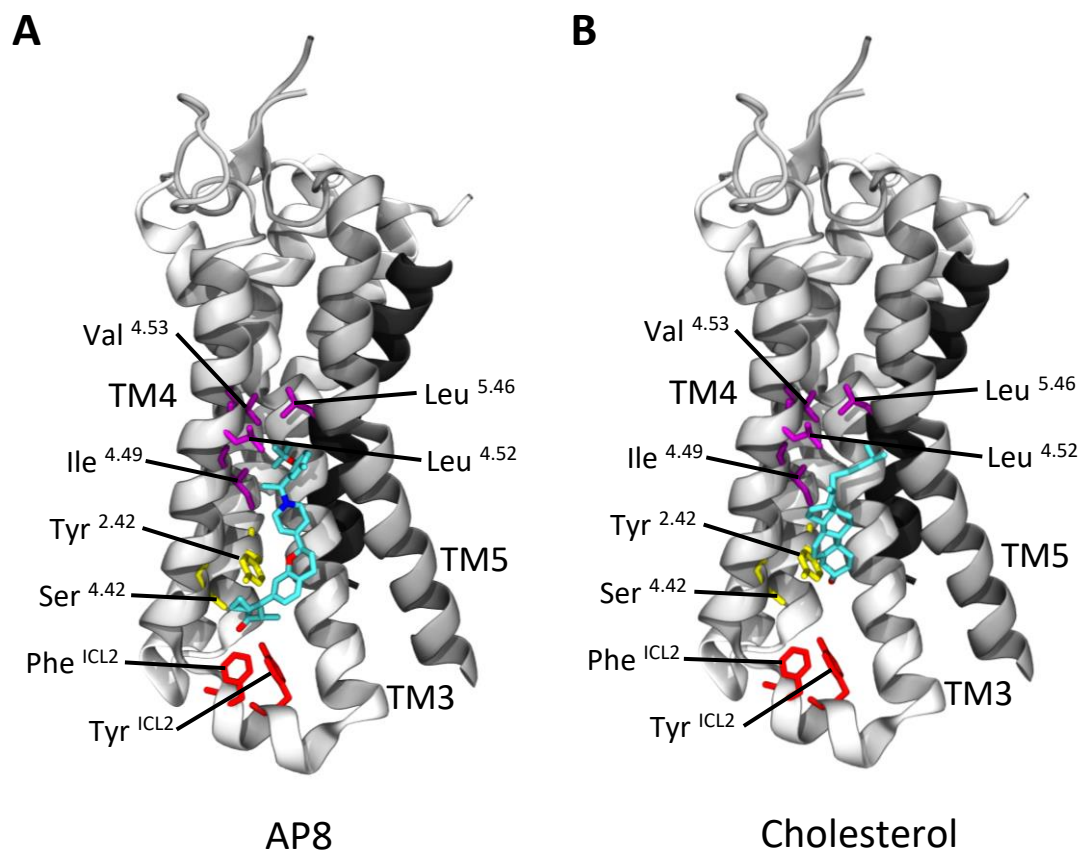


Figure 5.4 – The TM345 binding site of AP8 and cholesterol on FFAR1

AP8 and cholesterol binding to the same TM345 site on FFAR1. The same residues of note have been labelled in both diagrams. They are shown in liquorice representation and are coloured as follows: those forming the hydrophobic pocket are purple, those forming a polar cavity are yellow, and aromatic and charged residues in ICL2 are shown in red.

A: AP8, shown in cyan, bound at the experimentally determined TM345 binding site. The experimentally identified groups of interacting residues are coloured as described above. The hydrophobic pocket accommodates the hydrophobic region of AP8 with its polar cavity accommodating the polar region of the molecule. The AP8 molecule also interacted with aromatic and charged residues in ICL2.

B: An atomistic cholesterol shown in cyan generated from the conversion of the CG cholesterol bound pose identified through simulation studies, at the same location as the AP8 TM345 binding site. Cholesterol is noted to interact with the same hydrophobic pocket as AP8, with its hydroxyl head group accommodated by the same polar cavity as the cyclic ether group of the AP8 molecule. Interactions with the aromatic and charged residues in ICL2 are not noted with cholesterol, due to its smaller structure not reaching these residues at the bottom of the binding cleft.

As described in **Figure 5.4**, in the structure of FFAR1 that was solved in the presence of AP8, it was observed that there were three groups of residues which formed interactions with AP8: a hydrophobic pocket, a polar cavity and aromatic and charged residues from the ICL2 region. The residues of the hydrophobic pocket include: Ile ^{4.49}, Leu ^{4.52}, Val ^{4.53} and Leu ^{5.46}. They form interactions with both AP8 and cholesterol, with the cyclic central region as well as the hydrophobic region. The polar cavity is formed by Tyr ^{2.42} and Ser ^{4.42}, both of which contain a hydroxyl group in their side chains. They were noted to interact with the cyclic ether group on the AP8 molecule, and the same polar cavity is able to accommodate the hydroxyl head group of the cholesterol. However, as noted above, not all of the interactions reported in the AP8 binding site on the FFAR1 structure were observed with cholesterol, in particular the interactions with the aromatic and charged residues of ICL2.

5.3.1.1 Residues of Interest in the TM345 Binding Site on FFAR1

The FFAR1 TM345 cholesterol binding site was investigated further by identification of the specific residues responsible for cholesterol binding. As **Figure 5.5 A** shows, there is a binding pocket surrounding the position of the hydroxyl group of the cholesterol. This extends further than the head group of the cholesterol ligand and provides a location for the binding of the cyclopropane and carboxylic acid groups of AP8.

Cholesterol binding residence times with the individual CG-protein particles can be utilised in order to identify the residues responsible for cholesterol binding, as shown in **Figure 5.5 B** and **Figure 5.5 C**. The greatest

level of contact between the binding site and the cholesterol occurs at the residues surrounding the hydroxyl head group and involves interactions between the cholesterol and two tyrosine residues: Tyr^{2.49} and Tyr^{4.41}. A hydrophobic cleft formed by residues of TM3 is responsible for interactions with the cyclic region of the cholesterol. This region, shown in **Figure 5.5 D**, is formed by the residues Gly^{3.41}, Phe^{3.42}, Ala^{3.44} and Ala^{3.45}. These residues are in close proximity to the DRY motif and could be a site of possible allosteric regulation of the GPCR. One of these residues, Gly^{3.41}, has also been identified to affect the stability of several GPCRs including ADRB2²⁹¹. The residues of the DRY motif in FFAR1 are not the canonical DRY residues, instead the Asp^{3.49} is replaced by Ala^{3.49}. This substitution has been correlated with an increase in receptor activity in other members of the GPCR family²⁹², while in some others the Asp^{3.49} is replaced by a Glu^{3.49}, leading to this motif also being termed the E/DRY motif²⁶⁶.

The tail region of the cholesterol is in proximity to a histidine residue, coloured blue in **Figure 5.5 D**. This residue, His^{3.45}, has been identified as a site of allosteric regulation in some other members of the free fatty acid receptor family²⁹³.

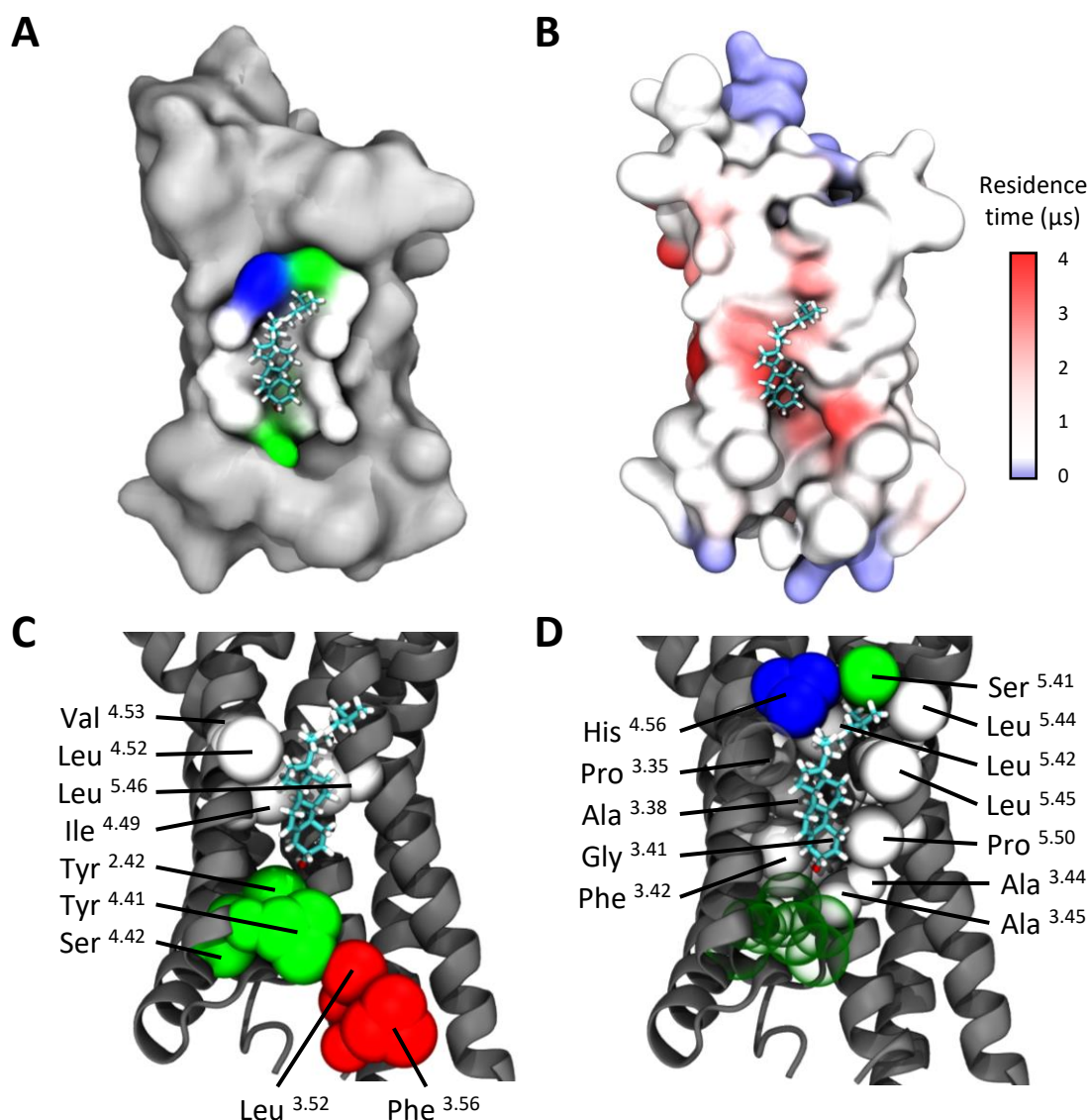


Figure 5.5 – The TM345 binding site of cholesterol on FFAR1 with residence times and residues of interest

A: The TM345 cholesterol binding pocket on FFAR1. Cholesterol is shown in atomistic representation and residues are coloured according to their properties: white for non-polar and green for polar, with blue representing histidine.

B: Residence times of the residues of FFAR1 with cholesterol, with highest values in bright red and blue indicating no interaction with cholesterol at that residue.

C: Residues which were noted to be of interest in the AP8 binding site²⁷² and were also noted to be in close proximity to the bound cholesterol at the TM345 binding site. The exception is the two residues in red, with which the cholesterol did not interact due to not extending that far into the binding cleft.

D: The TM345 cholesterol binding site comprises all the residues from the AP8 binding site, with the exception of Leu^{3.52} and Phe^{3.56} (as previously discussed), plus several additional residues which have been labelled. The hydroxyl head group is in proximity to two tyrosine residues, Tyr^{2.49} and Tyr^{4.41} which are labelled in C. A non-polar cleft is formed by residues Gly^{3.41}, Phe^{3.42}, Ala^{3.44} and Ala^{3.45}, which accommodates the cyclic region of the cholesterol.

5.3.2 Investigation of the TM345 Site on Additional GPCRs

The confirmation that a CG approach is able to detect cholesterol binding at the same TM345 binding site as AP8, and the detection of cholesterol binding at the same region in simulation studies of AA2AR, as well as pharmacophore screening, prompted further simulations to investigate if this site is present within other GPCRs. The approach that was used for investigation of FFAR1 was expanded to study whether this binding site was unique to this receptor or if it was present on other members of the GPCR family.

In order to investigate the binding of cholesterol to GPCRs at the TM345 site, fourteen GPCRs were selected for simulation. These proteins were chosen to be representative of both the Class A and Class B GPCR families. Consideration was also given to variation in sequence identity in order to enable the investigation of conservation of cholesterol binding at this location. The evolutionary relationships between the GPCRs simulated are shown in the phylogenetic tree in **Figure 5.6**. This analysis was performed by the GPCRdb.org database ²⁹⁴. Additionally, the properties of the ligands of each of the GPCRs is shown. The GPCRs were selected for study based upon several characteristics including: pharmacological relevance, structural completeness and resolution.

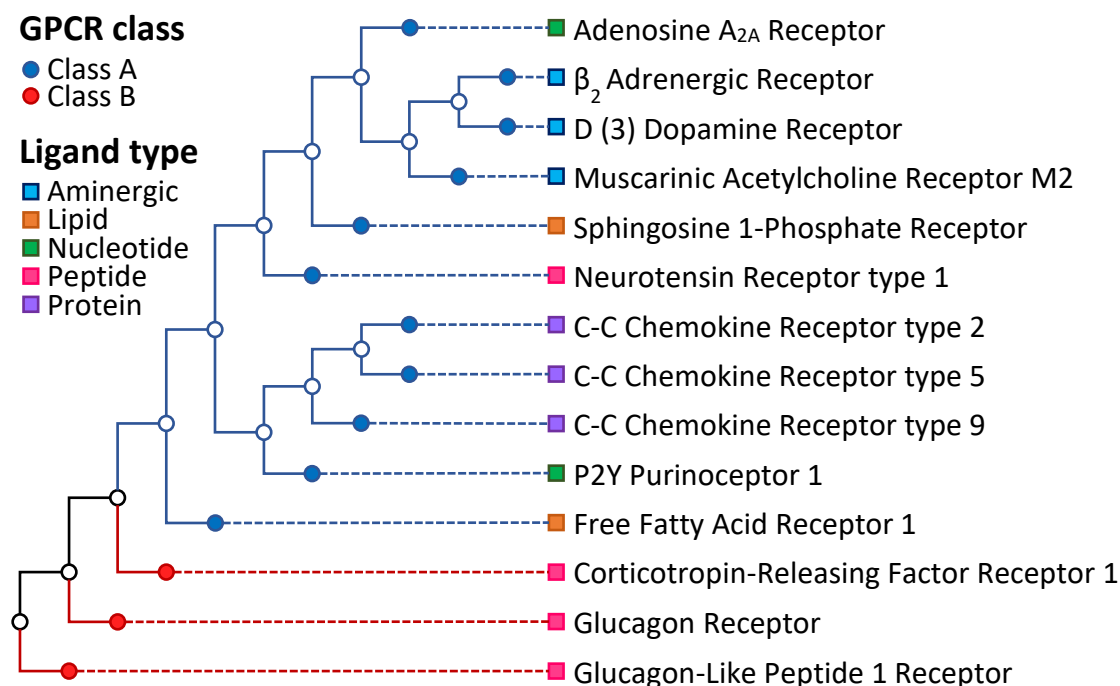


Figure 5.6 – Phylogenetic analysis of the GPCRs used in this research, with GPCR class and ligand type

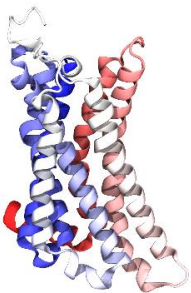
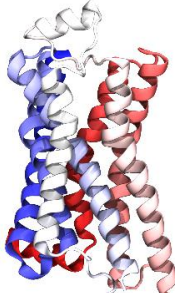
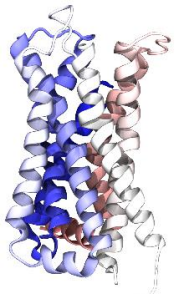
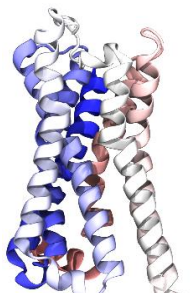
The results of phylogenetic analysis of the GPCRs simulated show the evolutionary relationships between members of the GPCR family and illustrate the distance between the Class A and Class B families. The type of ligand bound by each receptor is also shown.

Figure adapted from GPCRdb.org²⁹⁴

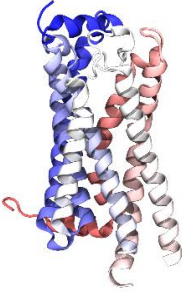
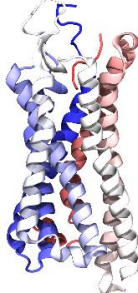
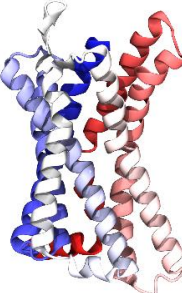
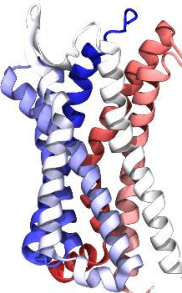
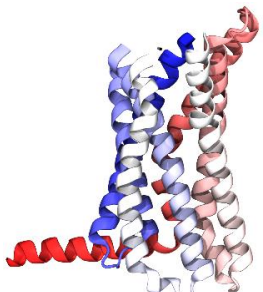
The binding of cholesterol to the GPCRs under investigation was studied using the same procedure as was utilised for FFAR1. The simulations were then analysed in the same manner using the PyLipID software package to investigate the location and dynamics of cholesterol binding sites. The GPCRs simulated are shown in the following table, **Table 5.1**.

Table 5.1 – Structures of the GPCRs used for investigation of cholesterol binding at the TM345 site

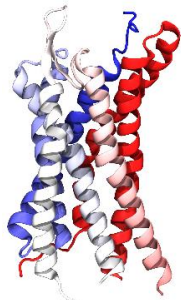
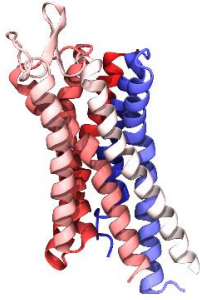
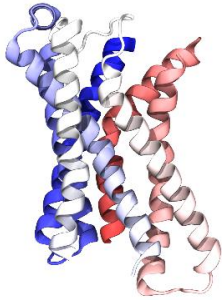
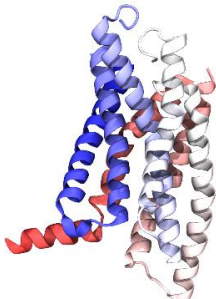
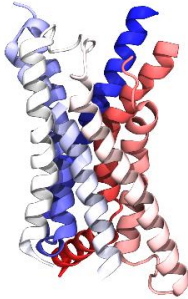
The GPCR structures are identified by the UniProt name and PDB ID, and are listed with respective resolutions, GPCR families (Class A or Class B) and ligand nature. The proteins are coloured according to residue number, starting from TM1 in dark blue and progressing to H8 in dark red.

Name (PDB ID)	Resolution (Å)	Structure	Class	Ligand
AA2AR (3EML ²⁹⁵)	2.60		A	Nucleotide
ADRB2 (2RH1 ¹²³)	2.40		A	Aminergic
DRD3 (3PBL ¹²³)	2.89		A	Aminergic
ACM2 (4MQT ²⁹⁶)	3.70		A	Aminergic

Interactions Between Cholesterol and G-Protein-Coupled Receptors

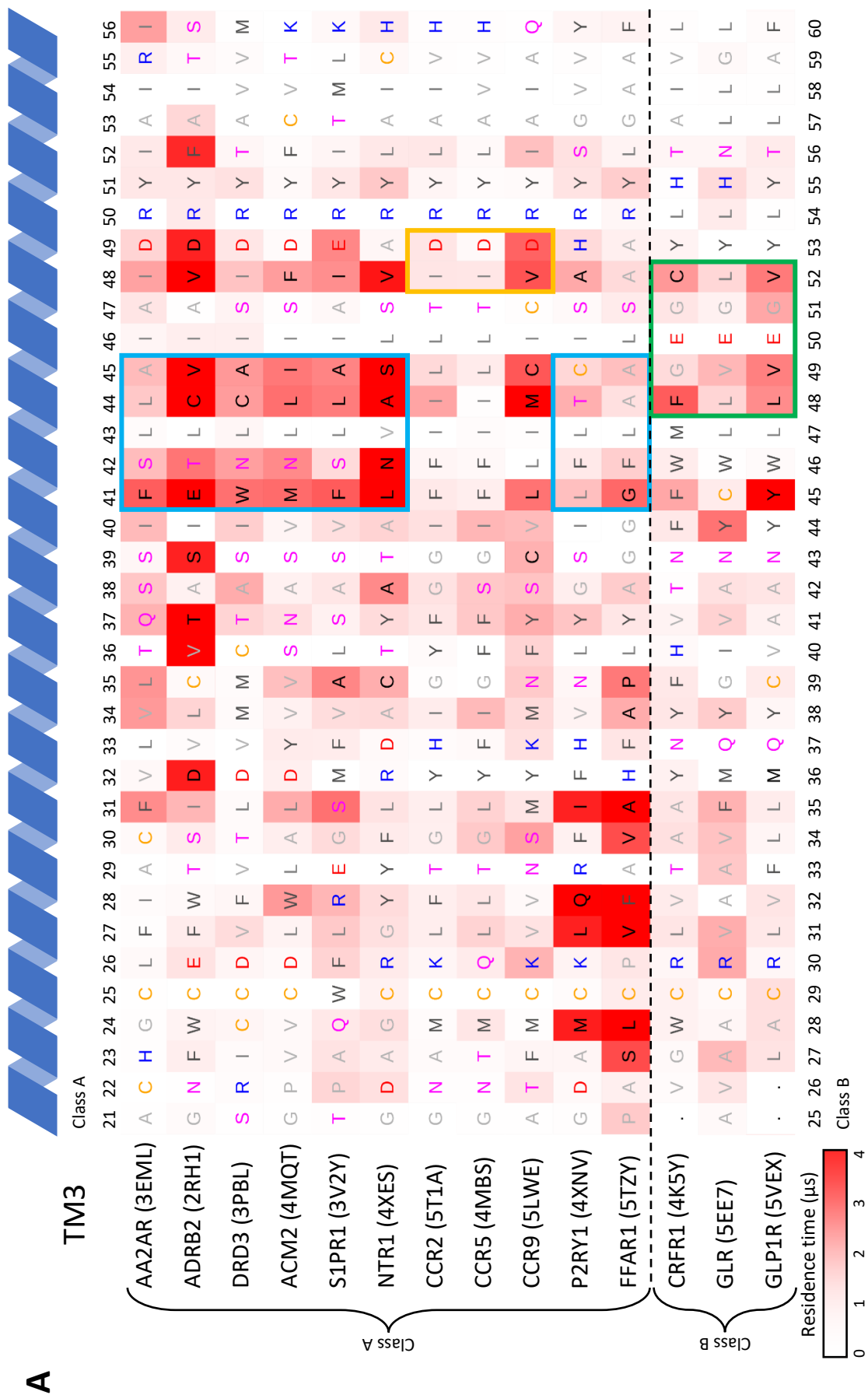
S1PR1 (3V2Y ²⁷¹)	2.80		A	Lipid
NTR1 (4XES ²⁹⁷)	2.60		A	Peptide
CCR2 (5T1A ²⁹⁸)	2.81		A	Protein
CCR5 (4MBS ²⁹⁹)	2.71		A	Protein
CCR9 (5LWE ³⁰⁰)	2.80		A	Protein

Interactions Between Cholesterol and G-Protein-Coupled Receptors

P2RY1 (4XNV ³⁰¹)	2.20		A	Nucleotide
FFAR1 (5TZY ²⁷²)	2.20		A	Lipid
CRFR1 (4K5Y ³⁰²)	2.98		B	Peptide
GLR (5EE7 ³⁰³)	2.50		B	Peptide
GLP1R (5VEX ³⁰⁴)	3.00		B	Peptide

The sequences of the GPCRs under investigation were then aligned using GPCRdb.org²⁹⁴ according to the grouping provided by the phylogenetic analysis. The residues in the aligned GPCR sequences were coloured by the duration of their residence times with cholesterol, allowing the strength of interaction with cholesterol to be observed, as shown in **Figure 5.7**.

As had been previously identified from simulation studies and pharmacophore screening, the binding of cholesterol to the TM345 site on AA2AR was confirmed here. The binding of cholesterol to AA2AR at this site occurs in a location which is equivalent in the GPCR sequence to the position of the TM345 cholesterol binding site on FFAR1, involving the residues which formed a hydrophobic cleft including the residues at positions 3.41, 3.43 and 3.44. The binding of cholesterol was not restricted to these two GPCRs and cholesterol binding in this region appeared to be common to many of the GPCRs investigated. In six of the GPCRs simulated a polar residue was found at position 3.42, typically a serine or a threonine. This residue also made interactions with the bound cholesterol, often at the end of the cyclic region nearest the hydroxyl headgroup with which it also interacted.



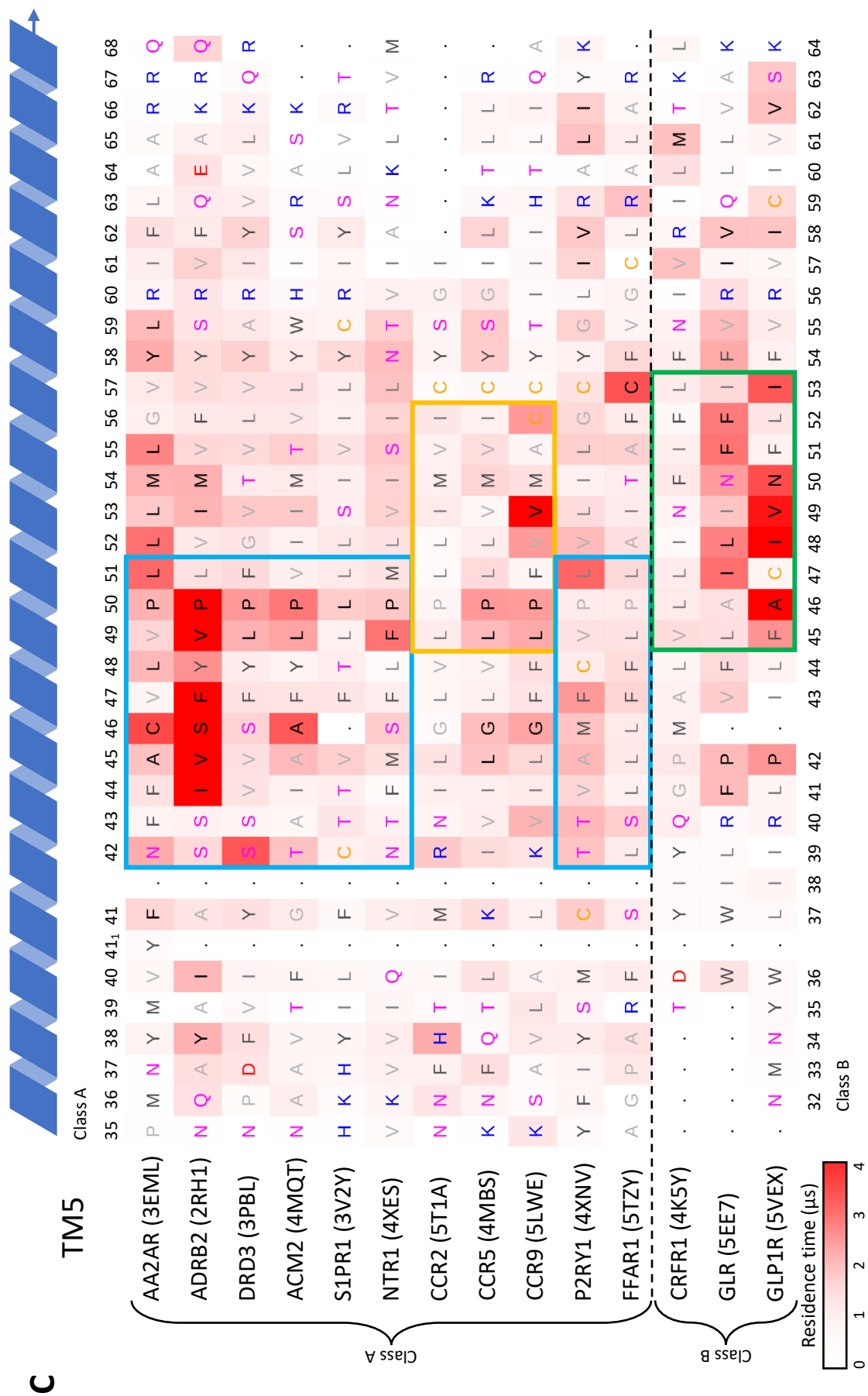


Figure 5.7 – Sequence alignments of TM 3, 4 and 5, with residues coloured according to residence time

The TM sequences of the GPCRs simulated were aligned using GPCRdb.org²⁸⁸⁻²⁸⁹ and visualised in phylogenetic order using the ALINE Protein Sequence Alignment Editor²⁹⁰. The sequences are coloured by residence time of cholesterol, as calculated by the PyLipID software package²⁴⁶. Regions which show similar cholesterol binding patterns in multiple GPCRs have been indicated with boxes on the sequences. These have been coloured according to subgroups: blue for Class A GPCRs, with the exception of the three C-C Chemokine Receptors (CCRs) for which yellow boxes were used; and green for the Class B GPCRs.

A: The TM3 sequence alignment shows the region where the majority of interactions between the cholesterol hydroxyl head group and the proteins occur. In the non-CCR Class A GPCRs, interactions often occurred between residues 3.41 to 3.45 and the hydroxyl head group of the cholesterol, along with some neighbouring parts of the cyclic region of the lipid. A polar residue was often present at position 3.42, which would interact with the polar hydroxyl group. In the CCRs, the slightly lower binding position of cholesterol resulted in interactions between its head group and residues slightly further along the sequence, particularly at positions 3.48 and 3.49, immediately adjacent to the DRY motif. The Class B GPCRs had a similar pattern of interaction to the CCRs, with cholesterol head group interactions occurring predominantly from residue 3.48 to residue 3.52.

B: The TM4 sequence alignment shows a mixed picture, with some regions of interaction with the cholesterol head group and some with the cyclic and hydrophobic regions of the lipid. In the Class A GPCRs, interactions occurred primarily with the cyclic and hydrocarbon chain regions of the cholesterol. This was due to these non-polar regions of the lipid interacting with non-polar residues on the GPCR, particularly from position 4.44 to 4.50. In the CCRs and Class B GPCRs, the lower cholesterol binding position resulted in interactions between the hydroxyl head group of the cholesterol and the N-terminal residues of TM4. In addition, interactions between the cholesterol head group and the residues of ICL2 were also noted, though the ICL and ECL sequences have not been shown here.

C: The TM5 sequence alignment primarily shows interactions between the proteins and the hydrocarbon chain region of the cholesterol. In the Class A GPCRs, an extended region encompassing non-polar residues from 5.42 to 5.51 was responsible for the majority of contact in TM5, including at Pro^{5.50} which has been demonstrated to be involved in GPCR activation³⁰⁵. In the CCRs and Class B GPCRs, the lower cholesterol binding position resulted in interactions with residues further along the sequence, as was noted in the other TM helices.

Sequence alignments for the TM3, TM4 and TM5 regions of the GPCRs show that the cholesterol bound at the TM345 site makes contact with residues from all three of these helices. In particular, the intracellular and central regions of these helices formed the majority of the interactions with cholesterol. These regions which encompassed positions 3.41 to 3.45 on TM3, 4.44 to 4.50 on TM4 and 5.42 to 5.51 on TM5 contributed to the hydrophobic cleft which bound cholesterol. The residues of this region lacked sequence conservation. However, as has been previously noted ¹²¹, the absence of a highly conserved binding sequence does not exclude the possibility of a conserved cholesterol binding site, as typically cholesterol binding sites in GPCRs lack sequence conservation.

The sequence alignment of the GPCRs in **Figure 5.7** shows that the TM345 binding site is situated near to the highly conserved DRY motif of Class A GPCRs. The DRY motif is highly conserved due to the importance of this region for GPCR signalling. The binding of cholesterol at a region that is separated by approximately ten residues from this important region might provide a mechanism for the ability of cholesterol to modulate the GPCR activity by binding to the site.

The binding of cholesterol at the TM345 binding site is not completely conserved throughout all the GPCRs investigated. However, cholesterol has been found to bind to some members of both Class A and Class B GPCRs at this region. For members of Class A, the binding of cholesterol at the TM345 site appears to be correlated with the nature of the ligand which the GPCR binds.

Of the GPCRs simulated, the GPCRs which bind peptides and proteins have a lower affinity for cholesterol at the TM345 site than other ligands such as the aminergic, nucleotide and lipid binding GPCRs. This means that despite the evolutionary distance between FFAR1 and some of the other GPCRs simulated the interaction with cholesterol is still preserved. In GPCRs that bind protein ligands, such as the members of the C-C Chemokine Receptor (CCR) family, lower cholesterol residence times were detected. For example, while the majority of Class A GPCRs show high interaction at residues 3.41 and 3.42, as well as in the 4.44 to 4.50 region, members of the CCR family show much less cholesterol interaction at these sites.

The Class B GPCRs overall exhibited less interaction with cholesterol at the TM345 site than the Class A GPCRs, with the exception of the GLP-1 receptor which had high residence times for cholesterol binding at this site. While CRFR1 and GLR had low residence times of 2.11 μ s and 1.18 μ s respectively, the TM345 site of GLP1R has a cholesterol residence time of 3.96 μ s. Sequence conservation of the residues in the TM345 region of the Class B GPCRs simulated was more consistent than the members of Class A family, with a greater proportion of the residues in this region being charged or polar. The conserved Glu^{3.50} in the Class B GPCRs occupies the location of the non-polar cleft found in the Class A GPCRs, which consists of the residues at positions 3.41, 3.42, 3.44 and 3.45. This charged residue might disrupt the binding of cholesterol at this location. It was observed that when binding to Class B GPCRs, the cholesterol binding poses differed to those with Class A GPCRs.

The position of the cholesterol was closer to the intracellular side of the membrane, resulting in the residues of the ICL2 of the Class B GPCRs making contact with the cholesterol head group. Present in this loop in the Class B GPCRs were several charged residues including glutamate and arginine.

A shifted cholesterol binding position was also observed in one sub-family of GPCRs; the CCRs. The position was closer to the intracellular region of the membrane protein and, like in the Class B GPCRs, this enabled the bound cholesterol to form contacts with the residues of ICL2. In addition, the hydroxyl head group was therefore positioned closer to the arginine of the DRY motif with which it could interact. However, the overall binding of cholesterol to these CCRs was weaker than that with other members of the Class A GPCRs.

The particularly high residence time of cholesterol binding to GLP1R in comparison to the other Class B GPCRs can be explained by the difference in binding pose adopted by the cholesterol to this protein. As shown in **Table 5.2**, the orientation of the cholesterol at this location is different to the other GPCRs. When binding to GLP1R the cholesterol adopts an orientation with its cyclic and tail regions perpendicular to the membrane normal, unlike in the other GPCRs simulated where these regions are parallel to the membrane normal. This binding is primarily comprised of interactions between the cholesterol and Asn^{5.50}.

An additional cholesterol binding site was identified at the extracellular end of TM3 for FFAR1 and P2RY1, but not in the other GPCRs simulated. This site involved residues at positions 3.23 to 3.31, but is distinct from the TM345 site

due to the position of the TM4 helix, which separates ligands binding at this site from those binding at the TM345 site. This higher site encompasses the highly conserved Cys^{3.25} residue which forms a disulphide bridge with a cysteine in ECL2, and disruption of this residue has been shown to affect the function of members of the GPCR family³⁰⁶⁻³⁰⁷. During structural studies of FFAR1 a partial agonist, MK-8666, was noted to bind at this location²⁷², raising the question of whether this ligand may also bind to the analogous site on P2RY1 which was identified here.

5.3.2.1 General Properties of the TM345 Binding Site

To further investigate the nature of cholesterol binding at the TM345 binding site, the binding cleft on the GPCRs can be visualised, as is shown in **Table 5.2**, where the individual residues of the GPCR structure are coloured according to the residence time of bound cholesterol. As described previously, the binding of cholesterol to GPCRs that have protein or peptide physiological ligands is lower than that of GPCRs with other ligands. It should also be noted that cholesterol binding at the TM345 site on FFAR1 does not have the highest residence time of all the GPCRs investigated, though it is higher than the majority; only one other Class A GPCR, ADRB2, has a higher residence time than FFAR1. This variation in binding strength appears to be correlated with the properties of the amino acids in the binding cleft which form interactions with the hydroxyl head group of the cholesterol, with higher residence times observed when the amino acids in this region are polar or charged.

Overleaf:**Table 5.2 – Cholesterol binding at the TM345 site on a variety of GPCRs, with residence times and residues of particular interest**

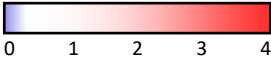
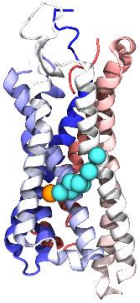
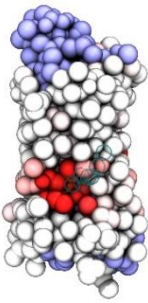
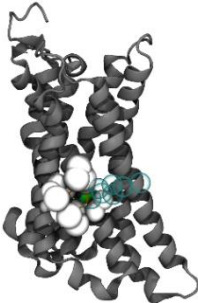
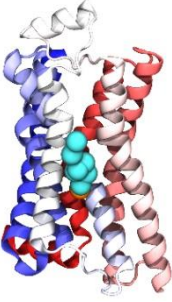
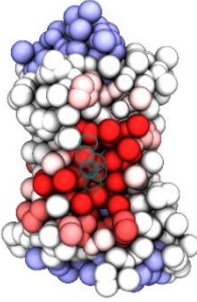
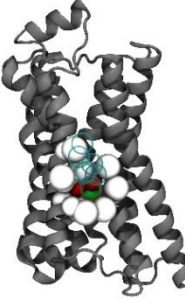
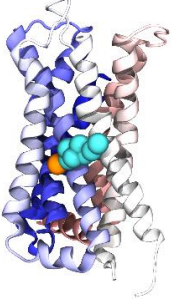
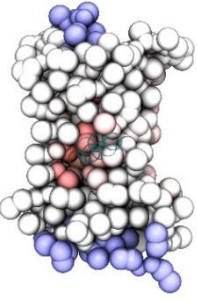
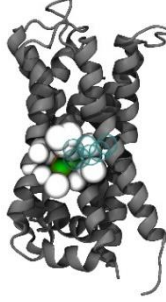
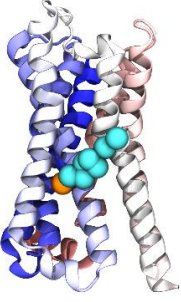
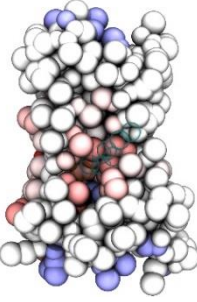
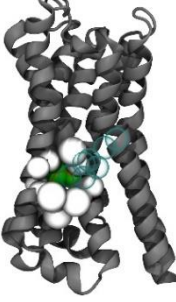
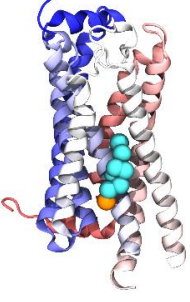
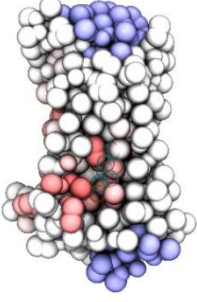
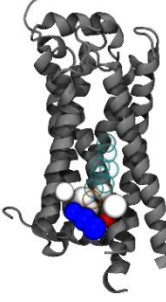
Column 1: The UniProt name³⁰⁸⁻³⁰⁹ and PDB ID¹⁴⁷ of each GPCR investigated for cholesterol binding. Note that CCR5 (PDB ID 4MBS) is not included as it did not show a detectable TM345 binding site for cholesterol.

Column 2: Cholesterol bound at the TM345 binding site on each of the GPCRs. The proteins are coloured according to residue number, starting from TM1 in dark blue and progressing to H8 in dark red. The bound cholesterol is in cyan with the hydroxyl head group in orange.

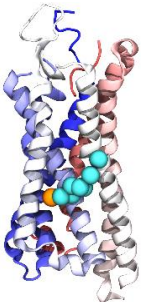
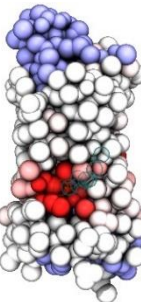
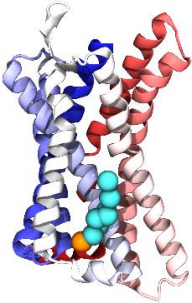
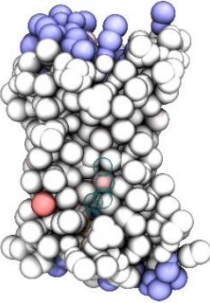
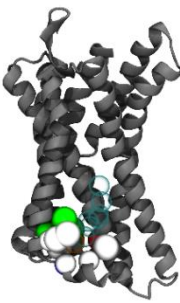
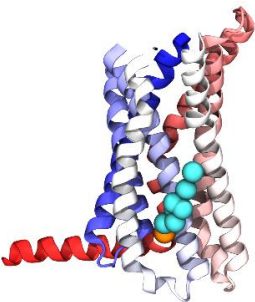
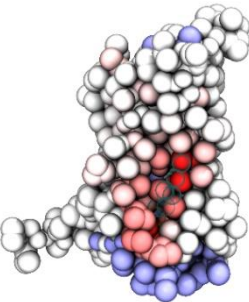
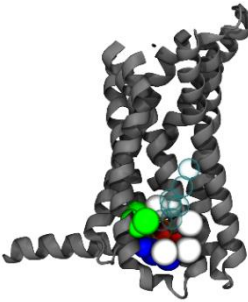
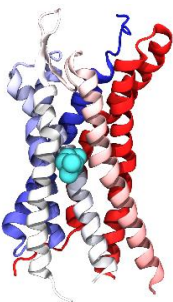
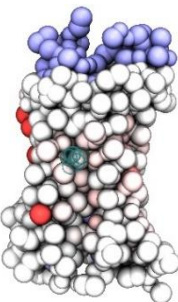
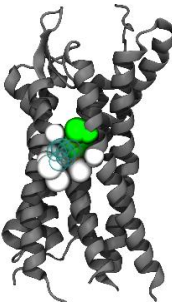
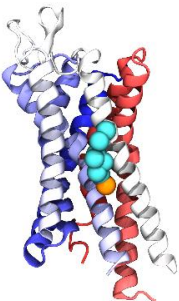
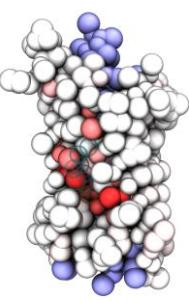
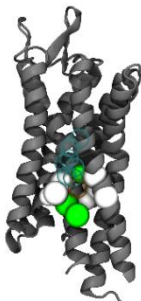
Column 3: Residence times of cholesterol in the same binding sites, depicted using a vdW representation of the GPCR. The bound cholesterol is displayed as transparent in order to allow the residues beneath to be seen. Residues are coloured according to cholesterol residence time; increasing intensity of red indicating higher residence time, and blue indicating that there was no interaction between that residue and cholesterol.

Column 4: The residues surrounding the cholesterol head group when in the TM345 binding site, coloured according to their properties: green for hydrophobic residues, blue for negatively charged residues, red for positively charged residues and white for non-polar residues. In addition, residues of interest have been listed to the right, using the Ballesteros–Weinstein²⁶⁵ and Wootten³¹⁰ GPCR numbering schemes.

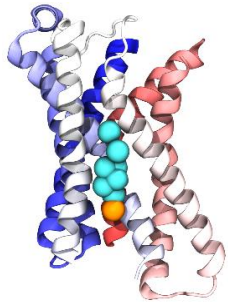
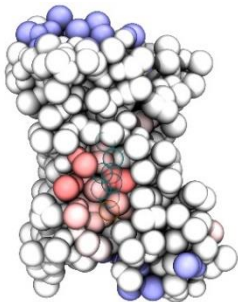
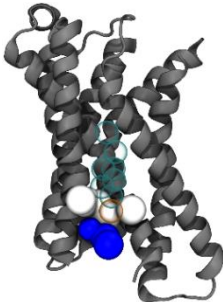
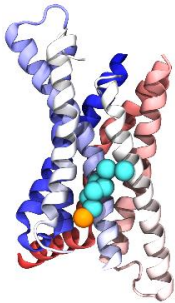
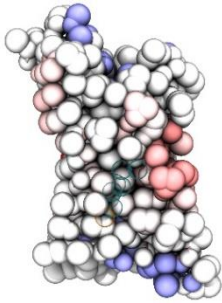
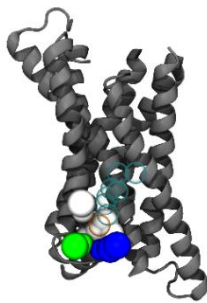
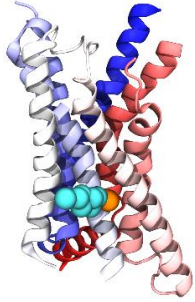
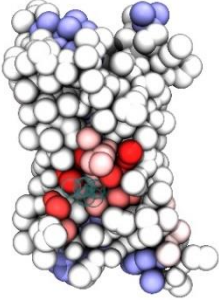
Interactions Between Cholesterol and G-Protein-Coupled Receptors

Name (PDB ID)	Cholesterol binding site	Residence time (μ s) 	Selected residues in proximity to the cholesterol hydroxyl head group
AA2AR (3EML)			 Ser ^{3.42}
ADRB2 (2RH1)			 Glu ^{3.41} Thr ^{3.42}
DRD3 (3PBL)			 Asn ^{3.42}
ACM2 (4MQT)			 Asn ^{3.42}
S1PR1 (3V2Y)			 Glu ^{3.49} Arg ^{4.41}

Interactions Between Cholesterol and G-Protein-Coupled Receptors

NTR1 (4XES)			 Asn ^{3.42} Ser ^{3.45}
CCR2 (5T1A)			 Asp ^{3.49} Arg ^{34.57}
CCR9 (5LWE)			 Asp ^{3.49} Arg ^{4.38} Tyr ^{4.41} Ser ^{4.42}
P2RY1 (4XNV)			 Tyr ^{3.37} Thr ^{5.42}
FFAR1 (5TZY)			 Gly ^{3.41} Tyr ^{4.41}

Interactions Between Cholesterol and G-Protein-Coupled Receptors

CRF1 (4K5Y)			 Arg ^{4.41}
GLR (5EE7)			 Arg ^{ICL2} Ser ^{4.40}
GLP1R (5VEX)			 Glu ^{3.50} Gly ^{3.51} Asn ^{5.50}

5.3.2.2 The TM345 Binding Site on ADRB2

The residence time of the cholesterol TM345 binding site on ADRB2 was the one of highest of the Class A GPCRs simulated, 2.40 μ s. However, as shown in **Table 5.2**, the binding orientation of the cholesterol is different to the other cholesterol molecules bound at the TM345 cleft. The cholesterol at the TM345 cleft in ADRB2 binds in an orientation which positions its tail and cyclic regions to be more perpendicular to the membrane normal, as is shown in **Figure 5.8 A**. This is in contrast to its orientation at other binding sites at the TM345 location, where the tail of the cholesterol is positioned parallel to the membrane normal and the hydroxyl group is orientated towards the intracellular region. In spite of this change in binding pose, there are strong interactions formed between the cholesterol ligand and ADRB2, as measured by residue residence times.

Many of the surrounding residues at the TM345 cleft make contact with the cholesterol molecule at this binding site, as shown in **Figure 5.8 B**. This observation combined with the particularly high residence time measured between the headgroup of the cholesterol and the Glu^{3.41} residue, suggests that the head group is anchored at this binding position while the rest of the cholesterol is able to rotate to contact multiple residues. The charged residue, Glu^{3.41}, near the centre of the TM helix, is not found in the other Class A GPCRs simulated.

Rotation of the cholesterol molecule at the TM345 cleft also allows its hydrophobic region to make greater contact with the residues immediately adjacent to the DRY motif, which is comprised of residues 3.49 to 3.51. This

rotation of the cholesterol also allows contact with Val^{3.48} and Phe^{3.52}, which show higher residence times in ADRB2 than in the other Class A GPCRs simulated. It can also be observed despite the different binding pose of cholesterol, the binding pocket which was identified in FFAR1 and other GPCRs at the TM345 binding site is still present in ADRB2.

The different orientation of the cholesterol in ADRB2 might be explained by the presence of several polar residue side chains in this region. These polar serine and threonine residues, are found in close proximity to Glu^{3.41}, as shown in **Figure 5.8 C**. They appear to prevent the non-polar cyclic region of the cholesterol adopting the membrane parallel conformation seen in other Class A GPCRs.

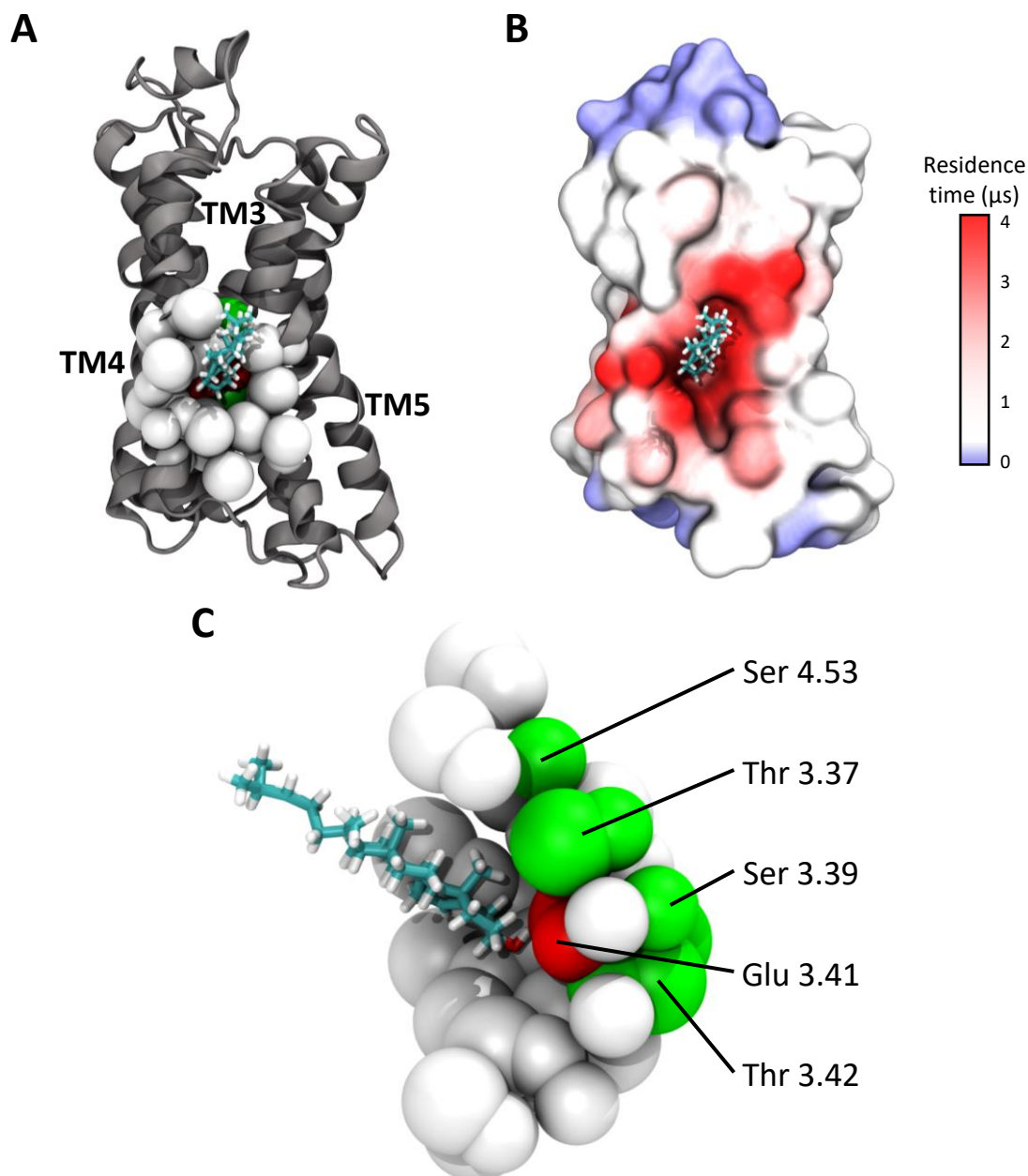


Figure 5.8 – The binding of cholesterol to the TM345 site on ADRB2

A: The TM345 binding site on ADRB2. Cholesterol is shown in cyan, with the residues surrounding its head group shown as spheres coloured according to their properties; green for polar and red for negatively charged.

B: A surface representation of ADRB2 coloured according to residue residence time of interaction with cholesterol.

C: An expanded lateral view of the binding site shown in A. The positioning of the bound cholesterol almost perpendicular to the TM helices may be due to the strong interaction between its polar head group and the charged residues of the binding site, shown in red, which act as an anchor point. The hydrophobic part of the cholesterol is prevented from adopting an orientation parallel to the TM helices by the polar residues above.

5.3.2.3 TM345 Binding Site Specificity

A comparison between the ligand residence time and its occupancy at a binding site can provide information about the binding specificity. Ligand binding sites which are specific can be identified by the ligand having a high residence time and a high occupancy, whereas sites which have lower occupancy and residence times are indicative of non-specific binding. The GPCR analysis that was performed here identified other cholesterol binding sites in addition to the TM345 site. By comparing the binding properties of these sites, the degree of the specificity of a site can be determined.

Consideration of residence times and occupancies of cholesterol binding sites, both at TM345 and at other locations, provides information about the specificity of the cholesterol interactions that occur on the GPCRs. Both the cholesterol interactions at the TM345 site and the other interactions that were detected between cholesterol and the GPCRs are shown in **Figure 5.9**. The binding at the TM345 site exhibits very high occupancies with some interactions nearing 100% occupancy over the course of the simulations. This, combined with the relatively elevated residence times of these binding locations in comparison to other cholesterol binding sites on GPCRs, suggests that cholesterol binding at the TM345 site is a specific interaction and is likely to not be shared with other lipids.

It can also be observed in **Figure 5.9** that despite the published structure of FFAR1 including a cholesterol-like ligand bound at the TM345 site, FFAR1 did not form the strongest interactions with cholesterol at that site. Several other

GPCRs were observed to bind cholesterol more tightly in this region than FFAR1; in particular, ADRB2, which has a residence time of 2.40 μ s in comparison a residence time of 2.56 μ s for FFAR1. This analysis suggests that the specific binding of cholesterol at the TM345 location is a property conserved within several members of the GPCR family.

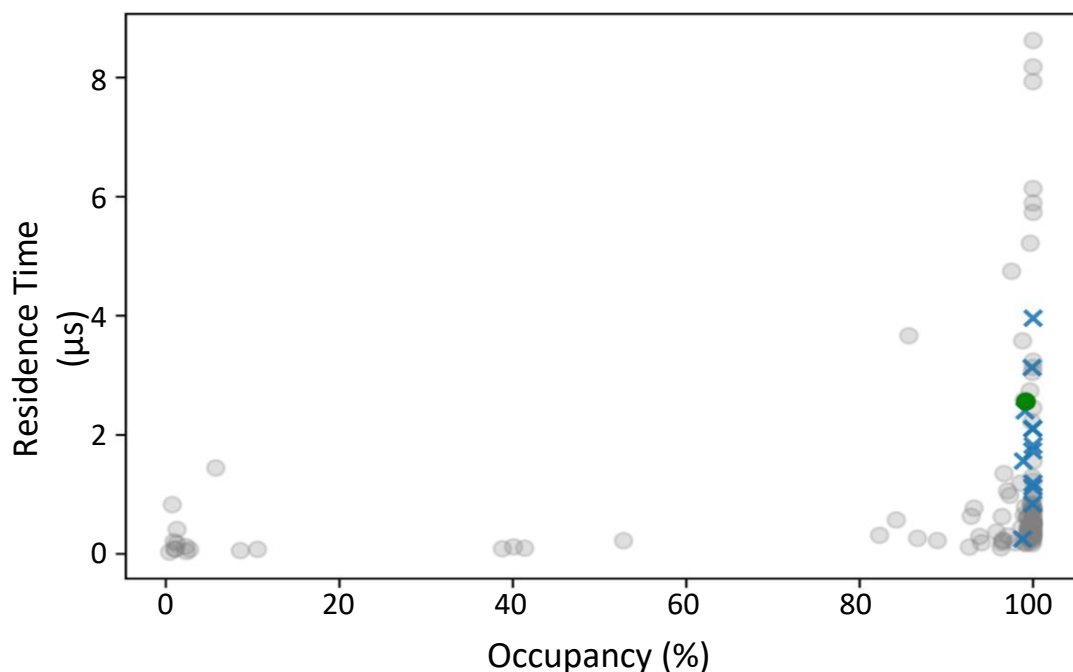


Figure 5.9 – Specificity of cholesterol binding at the TM345 site

The specificity of a cholesterol binding site can be visualised on a graph of the residence time and occupancy of a bound cholesterol at that site. Non-specific interactions display low residence times and occupancies. Conversely, the residence time and occupancy of specific interactions are greater. The binding sites identified by the analysis for all the GPCRs simulated are shown above; those which correspond to the TM345 location are displayed in blue and those which do not are in grey. The FFAR1's TM345 binding site is in green, showing both high occupancy and residence time; this is therefore a specific interaction. The same is true for most of the other GPCRs where cholesterol binding at the TM345 site was identified.

5.3.2.4 Comparison of Residence Times and GPCR State

During their activation GPCRs undergo conformational changes. As part of this process several of the TM helices undergo rearrangements, often including the intracellular side of TM6, which moves outwards from TM3. In ADRB2 the

intracellular end of TM6 has been measured to move by 1.4 nm³¹¹. Additionally, the presence of cholesterol has been suggested to stabilise AA2AR³¹².

The activation state of the GPCRs, as determined by GPCRdb.org²⁹⁴, was compared to the strength of binding interaction, as measured by the binding site residence time at the TM345 site, as shown in **Table 5.3**. The binding of cholesterol at the TM345 site does not appear to be indicative of the activation state of the GPCR, although this might require simulation of more GPCRs in the active state to evaluate this observation. Alternatively, simulations of both the active and the inactive state of a GPCR might provide more information about the effect of GPCR state on cholesterol binding.

Table 5.3 – Activation state of the GPCRs simulated

Binding of cholesterol at the TM345 site does not appear to be correlated with the activation state of the GPCR: inactive, intermediate, active or active-like. Active-like states are typically produced through mutation of the protein to produce a conformation analogous to its active form but in the absence of a bound G-protein

UniProt	PDB ID	Residence time (μs)	Activation State
CCR5	4MBS	0.00	Inactive
CCR2	5T1A	0.25	Inactive
P2RY1	4XNV	0.84	Intermediate
DRD3	3PBL	1.12	Inactive
GLR	5EE7	1.18	Inactive
AA2AR	3EML	1.19	Inactive
ACM2	4MQT	1.73	Active
CCR9	5LWE	1.82	Inactive
S1PR1	3V2Y	2.10	Inactive
CRFR1	4K5Y	2.11	Inactive
ADRB2	2RH1	2.40	Inactive
FFAR1	5TZY	2.56	Intermediate
NTR1	4XES	3.13	Active-like
GLP1R	5VEX	3.96	Inactive

5.3.2.5 Protein Conformation and Cholesterol Binding

The structures of several GPCRs have been solved in multiple conformations, such as NTR1. In order to further investigate the effect of protein conformation on the binding of cholesterol, simulations were prepared using both the active-like and inactive conformations of NTR1, shown in **Figure 5.10 A** and **Figure 5.10 B** respectively. These simulations were performed using the same procedure as described for the other GPCRs simulated, including the use of a bilayer containing 35% cholesterol. As shown previously, in **Table 5.2**, a binding site for cholesterol was observed at the TM345 site on the active-like conformation of NTR1, shown in **Figure 5.10 C**, with a residence time of 3.13 μ s. However, in the inactive conformation this binding site was not present, as indicated by the lower residence times shown in **Figure 5.10 D**. The structural rearrangements that occur during the shift from the inactive to active state might be responsible for this change in cholesterol binding. As shown by comparison of the two conformations, the cleft where the cholesterol's head group binds in the active conformation is not present in the inactive conformation. The alteration in cholesterol binding properties during the change from the inactive to active conformations might be explained by the presence of an allosteric binding site at the TM345 location.

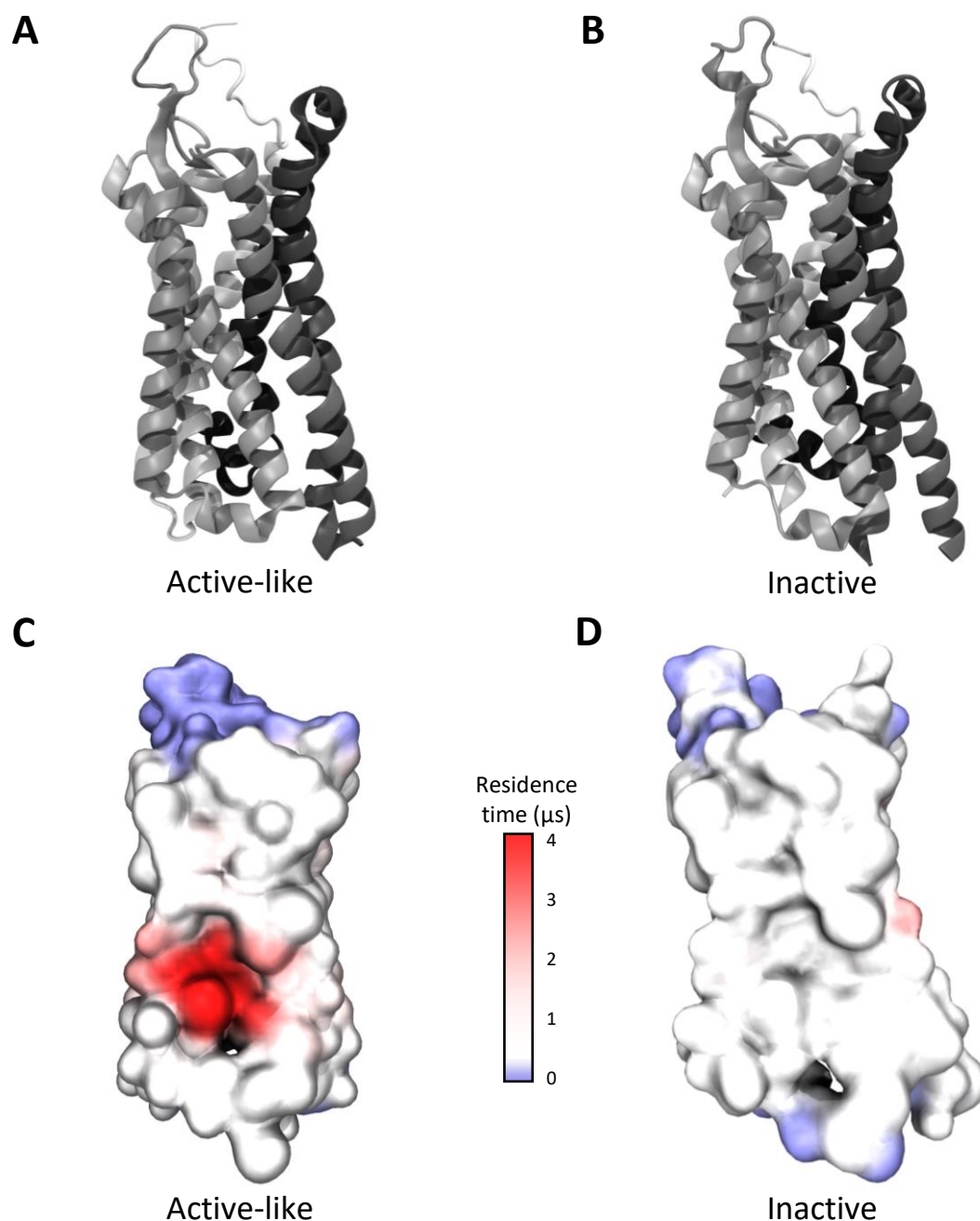


Figure 5.10 – The effect of conformational change on cholesterol binding to the TM345 site on NTR1

A: The structure of NTR1 in the active-like conformation, as was used for the previous simulations described above.

B: The structure of NTR1 in the inactive conformation; the conformational change has obscured the binding pocket that was present in the active-like structure.

C: The same structure as **A** shown in surface representation and coloured according to residence times; a cholesterol binding site at TM345 is indicated by high residence times, with more highly interacting residues coloured a brighter red.

D: The same structure as **B** shown in surface representation and coloured as above; the low residence times demonstrate the absence of a TM345 binding site in this structure.

5.4 Conclusions

The approach taken here aims to investigate further the interaction of membrane proteins with their lipid environment. During structural determination it had been found that a lipophilic allosteric modulator, AP8, can bind to a cleft present on the TM345 location of FFAR1. Due to structural similarities between AP8 and cholesterol, it has been proposed that this binding site could potentially be physiologically occupied by a cholesterol ligand. The binding of cholesterol to FFAR1 was therefore simulated and compared to these experimental findings in order to validate this computational approach.

Simulations were performed using three 10 μ s repeats in CG detail, which were then analysed using the PyLipID software package to identify and characterise lipid binding sites on FFAR1. A binding site for cholesterol was identified at the TM345 position which corresponded to the location of the binding site of AP8 during structural experiments. The residence time was 2.56 μ s, with a surface area of 12.9 nm². The binding cleft appeared to extend slightly beyond the length of the cholesterol head group towards the intracellular region. This corresponds to experimental findings: in the structure produced with AP8 bound at the TM345 position, the rest of the binding pocket was occupied by a cyclopropane and carboxylic acid groups, allowing interaction with aromatic and charged residues in the ICL2 region of the GPCR. This additional interaction with the GPCR could explain how AP8 is able to displace cholesterol at the site.

The experimentally identified residues which were important for the binding of AP8 were also observed to form strong interactions with cholesterol binding at the same site. The exceptions to this were the aromatic and charged residues of ICL2 which formed interactions with the carboxylic acids group of the AP8 molecule, but not with cholesterol. This was as a result of the cholesterol head group not extending as far into the binding pocket as that of AP8. Instead, the hydroxyl group of the cholesterol made contact with the polar residues which bind the cyclic ether moiety of AP8. The analysis also identified other residues involved in binding with the ligand at the TM345 site, which resulted in a complete binding pocket forming around the ligand.

The successful identification of a cholesterol binding site at the TM345 cleft on FFAR1 and experimental identification of AP8 binding at this location confirmed that this CG approach is able to identify lipophilic binding sites on membrane proteins. The identification of cholesterol binding sites on members of the GPCR family has been impeded by the lack of conserved sequences for the binding of cholesterol. Despite the observation of motifs such as the CRAC and CCM domains, these sites are not reliable indicators of cholesterol on GPCRs.

The ability of the simulation approach undertaken here to detect cholesterol binding sites, combined with the lack of conserved sequences for the binding of cholesterol, led to investigations of other members of the GPCR family for cholesterol binding at the TM345 site. Fourteen members of the GPCR family were simulated using the same procedure as was utilised for FFAR1, including

representatives from both Class A and Class B. These GPCRs were selected based upon several characteristics including pharmacological relevance, structural completeness, ligand nature and structural resolution.

The GPCRs simulated were sorted according to phylogenetic analysis performed by GPCRdb.org, and their sequences were then aligned using the same web server. In order to aid identification of the residues responsible for the interaction with cholesterol, the residues were coloured according to residence time of the interaction with cholesterol. Consistent with previous findings, the analysis performed here was able to correctly identify the TM345 binding site of cholesterol on AA2AR, as had been identified during previous structural experiments, pharmacophore screening and simulation studies.

In the majority of GPCRs simulated a hydrophobic pocket was observed to bind the cyclic region of the cholesterol; this was formed from several residues from TM3 including those at positions 3.41, 3.42, 3.44 and 3.45. Additionally, residues from the central regions of TM4 and TM5 also contributed to interactions with cholesterol in this region. Significant sequence conservation was not observed in this region. Residence times demonstrated that the binding of cholesterol at this site is not a universally conserved interaction between all members of the GPCR family, instead being prominent in members of Class A, but only noted in one member of the Class B: GLP1R. In addition, the binding pose of the cholesterol at the TM345 site was different in GLP1R compared to the other GPCRs simulated. The cyclic and hydrophobic regions of the cholesterol were positioned perpendicular to the membrane normal, reducing

contact between these regions of the cholesterol and the GPCR. However, the head group of the bound cholesterol was able to make strong interactions with Tyr^{3.45} and Asn^{5.50}, resulting in high residence times.

The comparatively low level of cholesterol binding with members of Class B can be explained by the presence of the conserved Glu^{3.50} in the Class B GPCRs. This occupies the location of the non-polar cleft which was common to the Class A GPCRs; consisting of residues 3.41, 3.42, 3.44 and 3.45, potentially disrupting the binding of cholesterol. It was observed that the binding poses of the Class B GPCRs differed to the Class A GPCRs. The position of the cholesterol was closer to the intracellular side of the membrane, resulting in the residues of ICL2 in the Class B GPCRs making contact with the cholesterol head group. In particular, this loop has several charged residues including glutamate and arginine, which would interact with the polar head group of the cholesterol.

5.4.1 Cholesterol Binding Variation within Class A GPCRs

Within Class A, a difference in cholesterol binding was observed between the chemokine receptors and the other Class A GPCRs, with chemokine receptors CCR2, CCR5 and CCR9 showing less cholesterol binding at the TM345 site than the rest of the class. It was observed that both Class B GPCRs and the chemokine receptors can bind different types of ligands to the other GPCRs simulated, peptides or proteins respectively, and this difference might be responsible for the reduction in cholesterol binding in these two groups. However, further simulations of GPCRs which have peptide or protein ligands would be required to investigate this further.

One of the strongest binding sites of cholesterol at the TM345 location was observed in ADRB2. Unlike in some of the other members of Class A, such as FFAR1 and AA2AR, the residue at the 3.41 position in ADRB2 is not hydrophobic. Instead, a glutamate is present, which has a negative charge. Interactions between the charged glutamate and the hydroxyl moiety of the cholesterol's head group result in stronger binding, evident in the increased residence times at this location. Despite this increased interaction with the head group of the cholesterol, the binding pocket does not extend to hydrocarbon chain of the cholesterol. This is visualised in the surface representation of GPCRs with bound cholesterol, as well as elevated residence times for multiple residues surrounding the binding site as a result of the tail region rotating as the head group remains in place.

The physiological state in which GPCRs exist in the bilayer must be considered when interpreting the results presented here. Though the GPCRs simulated here were in a monomeric form, GPCRs have been demonstrated to undergo oligomerisation³¹³. Cholesterol is often found at the interface between multimeric membrane proteins^{53, 314}, and this absence of additional protein subunits might reduce the residence time of certain cholesterol binding sites which are involved in protein oligomerisation. This effect may be alleviated by the inclusion of additional GPCR monomers into the simulations. Additionally, the presence of a G-protein might affect the cholesterol binding strength at the TM345 site.

5.4.2 Cholesterol Binding Specificity at the TM345 Site

A comparison of the properties of the TM345 site with those of all other cholesterol binding sites on the GPCRs simulated was performed in order to provide information about the specificity of cholesterol binding at this location. The specificity of a binding site can be understood in terms of the residence times and occupancy of bound ligands at this location. Binding sites which are specific for a particular ligand will possess high occupancy along with elevated residence times. In contrast, sites which have lower occupancy and residence times are indicative of non-specific binding. In GPCRs where binding of cholesterol is identified at the TM345 site, the occupancy is typically close to 100% over the duration of the simulations. This, combined with the relatively elevated residence times of these binding locations in comparison to other cholesterol binding sites on GPCRs, suggests that the binding of cholesterol at the TM345 site is a specific interaction.

The prevalence of specific cholesterol binding at the TM345 site, particularly within Class A, suggests that this binding location for cholesterol is a conserved property of some members of the GPCR family. This binding location is in proximity to the highly conserved DRY motif on TM3, and since the DRY motif is essential for the activation and function of Class A, the binding of a cholesterol at this location could have a key role in modulation of these GPCRs.

5.4.3 Cholesterol Binding Variation with Conformational State

During the activation of GPCRs, large scale rearrangements of the TM helices occur. Often this involves the tilting and kinking of helices at conserved

proline residues in the structure. TM6 undergoes the largest rearrangement upon activation, moving away from the rest of the GPCR.

The binding strength of cholesterol interactions with GPCRs, as measured by the residence time of the interactions, was poorly correlated with the activation state of the GPCR. This could result from the methods used to solve the GPCR in an active-like state, as often thermostabilising mutations are made to the protein in order to produce a conformation which is thought to be analogous to the active form of the protein but in the absence of a bound G-protein³¹⁵. These mutations might affect the binding of cholesterol at this site or fail to cause the rearrangements required at this location for cholesterol to interact. In addition to this there is currently a lack of consensus on what constitutes an active, intermediate or active GPCR state, which may obscure a correlation between residence times of cholesterol and protein conformational state.

Mini G-proteins³¹⁶⁻³¹⁷ are engineered GTPase domains of the $G\alpha$ subunits of G-proteins which are reduced in size and can be used during the structural determination of GPCRs. The simulation of active state GPCRs which have been solved in the presence of these proteins could provide a more detailed understanding of the relationship between cholesterol and the TM345 site.

Some members of the GPCR family have been solved in multiple conformational states, for example NTR1. This enables simulation and study of cholesterol binding in both inactive and active-like conformations to investigate whether changes to cholesterol binding could be measured in

different states. In the inactive state, NTR1 demonstrated negligible TM345 cholesterol binding. However, when simulated in the active-like conformation an increase in cholesterol binding was observed as evidenced by the increase in residence times at the TM345 location. If this trend was also present in FFAR1, this binding preference of cholesterol for the active conformation of the receptor could suggest a possible mechanism for the function of the allosteric regulator AP8: perhaps AP8 is able to displace a cholesterol which would typically be found at this location, thereby stabilising the active conformation.

6

Chapter 6 - Summaries and Future Directions

Contents:

6.1 Chapter Summaries	205
6.1.1 Simulation Models of Nanodiscs	205
6.1.2 Protein-Lipid Interactions in Nanodiscs	206
6.1.3 Interaction of Cholesterol with GPCRs	208
6.2 Future Directions.....	211
6.2.1 General Remarks	211
6.2.2 Nanodiscs	211
6.2.3 Membrane Protein-Lipid Interactions.....	214
6.3 Closing Remarks.....	216

6.1 Chapter Summaries

6.1.1 Simulation Models of Nanodiscs

The exponential increase in the deposited number of membrane protein structures is testament to the advances in this field. However, despite the wealth of structures now available for study, the assumptions and limitations of structural determination techniques necessitate careful interpretation of these results. In particular, utilisation of nanodiscs for structural determination of membrane proteins has seen recent rapid growth; and therefore, better understanding is required of the impact of this solubilisation method on the properties of the embedded proteins. MD simulations offer one avenue for this investigation and can provide benefits over *in vitro* studies.

Before the properties of nanodisc systems could be investigated using MD simulations, the development and validation of a simulation protocol was required. The protocol described in **Chapter 3** enables the creation and simulation of nanodisc systems in an automated customisable manner which utilises a multiscaled approach, involving simulations in both CG and atomistic detail. This approach was validated by good agreement with experimental results from techniques including SAXS ¹⁴² and NMR spectroscopy ¹³⁷.

The findings in this chapter demonstrated the advantages provided by a multiscaled approach: the increased speed accessible to CG simulations and the finer resolution available to atomistic simulations. Observations made during development of the pipeline methodology demonstrated the importance of protein stability and the detrimental effects of an incorrectly applied ENM over

the course of the simulation. This was also demonstrated following conversion to atomistic detail, where simulations created from the final coordinates of the CG systems with an ENM were less able to reproduce the expected X-ray scattering than those without an ENM applied to the MSPs.

The simulations also demonstrated a dependence of the nanodisc lipid properties on distance from the centre of the nanodisc. An overall increase in APL and order parameters was observed with increasing distance from the centre of the nanodisc, a finding which was consistent with observations from other simulation studies¹⁴⁰⁻¹⁴¹. This change in properties was likely to be a result of the MSPs constricting the lipid patch and causing perturbation, particularly at the periphery of the nanodisc system.

Validation of this pipeline methodology developed for simulation of nanodisc systems consisting only of a lipid patch and MSPs preceded extension of the protocol to allow simulation of integral membrane proteins embedded into a nanodisc.

6.1.2 Protein-Lipid Interactions in Nanodiscs

Once the ability of the pipeline methodology to simulate nanodisc systems without a membrane protein was established, the methodology was extended to allow the simulation of nanodiscs containing membrane proteins. A multiscale simulation approach, as described in **Chapter 4**, enabled creation and simulation of different diameters of nanodisc, including the large diameter MSP2N2 with a multimeric membrane protein such as TRPV6 incorporated into the system.

The ability to simulate nanodiscs containing integral membrane proteins facilitated investigation of changes to the properties of these proteins when embedded into a nanodisc. This was achieved by comparison of protein and lipid properties in two different environments: nanodiscs and bilayers. Five membrane proteins which had been structurally determined using nanodiscs for solubilisation were selected for further study. In order to mitigate disruption of protein properties, they were embedded into nanodiscs and bilayers of a lipid composition mimicking that used during experimental structural determination.

Confinement of the lipid patch by the MSPs was observed to perturb the properties of lipids in nanodiscs with an embedded membrane protein, in a similar manner to that previously observed for systems without a membrane protein. Changes to the APL and order parameters of the lipids had an observed dependence upon distance from the centre of the nanodisc, again suggesting that the confining MSPs perturb the properties of the nanodisc components.

Comparison of the structure of membrane proteins in nanodiscs and bilayers, as measured by the RMSF of the protein, did not indicate any significant changes in the two environments. However, the protein-lipid interactions were different; lipid residence times were consistently increased in the nanodiscs compared to those in the bilayer. Despite this, lipid occupancy of binding sites on the membrane proteins were consistent between the two environments. This suggests that the constraining MSPs cause reduced

exchange of lipids with the bulk population once the lipids make contact with the membrane protein.

It was also observed that protein-lipid interactions in nanodiscs could be altered depending upon the level of saturation of the hydrocarbon chains of the lipids; lipid residence times increased with higher levels of hydrocarbon chain saturation. This observation has particular significance for the utilisation of nanodiscs for structural determination of membrane proteins, where bound lipids are often observed in the structure. Additionally, the findings presented here will assist selection of MSP constructs with a particular diameter for experimental solubilisation membrane proteins, allowing them to experience an environment analogous to that of the physiological bilayer.

6.1.3 Interaction of Cholesterol with GPCRs

Cellular membranes are composed of a vast number of components, which can be divided into two main groups: membrane proteins and lipids. There is significant diversity in both these groups; over 47,000 different biologically relevant lipids have been catalogued and an estimated quarter of protein coding genes in humans encode membrane proteins. Understanding the interactions between these membrane components is critical due to the wide range of roles performed by membrane proteins and their regulation by lipids. The interaction between one family of membrane proteins, the GPCRs, and a membrane lipid, cholesterol, was therefore investigated in **Chapter 5**.

The identification of a lipophilic ligand binding site on FFAR1 at the TM345 cleft prompted speculation that cholesterol might also bind at this site, due to

general structural similarities between this ligand and cholesterol. Simulations of FFAR1 in a bilayer containing 35% cholesterol demonstrated that the TM345 cleft was indeed able to bind cholesterol and a specific cholesterol binding site was identified with a residence time of 2.56 μ s. This supports the hypothesis that the physiological ligand of FFAR1 at the TM345 cleft is cholesterol.

Other members of GPCR family were also investigated for interactions with cholesterol at the TM345 site in order to determine whether this binding location was a conserved feature of GPCRs or if it was isolated to FFAR1. Simulations were performed on members of both Class A and Class B GPCRs, with systems analysed in the same manner as that of FFAR1. Interactions were observed between the majority of the GPCRs simulated and cholesterol at the TM345 region. However, there was disparity in cholesterol binding in different groups within the GPCR family. Cholesterol binding at the TM345 site was reduced in the Class B GPCRs compared to that in the Class A GPCRs. Within Class A, there was a lower level of cholesterol binding at the TM345 site in the CCR family, the only protein ligand binding GPCRs simulated. In contrast, some of the highest cholesterol binding at the TM345 site was demonstrated in ADRB2, but the mode of binding was different to that in the other Class A GPCRs; the cholesterol adopted an orientation more perpendicular to the membrane normal than at other TM345 cholesterol binding sites.

The specificity of the interactions of cholesterol at the TM345 region was then investigated by comparison with other cholesterol binding sites which had been detected on the simulated GPCRs. Analysis showed that compared to the

non-TM345 binding sites, cholesterol binding at the TM345 site showed greater specificity of interaction with the GPCRs.

Further simulations provided insights about the effect of GPCR conformation on the degree of cholesterol binding at the TM345 site. In the case of NTR1, where structures have been determined in both the inactive and active-like conformations, there were differences in cholesterol binding between the two states; much higher residence times were measured for the binding of cholesterol with the active-like state of NTR1 than with the inactive state.

6.2 Future Directions

6.2.1 General Remarks

The field of computational MD simulation is under constant development and regular advances are being made, with releases of more efficient and accurate approaches for the study of biological systems. One such development is an update to the Martini CG force field: Martini 3¹⁸⁵. This update introduces new particle types and a new mapping scheme, aiming to facilitate replication of hydrogen bonding and electronic polarisability. The parametrisation of molecules in Martini 3 is ongoing which prevented the use of Martini 3 for the majority of studies undertaken in this research. Once parameterisation is complete, its application to many of the approaches described here might provide more information and a deeper understanding of these biological systems.

6.2.2 Nanodiscs

Development of membrane mimetics to facilitate the study of membrane protein dynamics, structures and functions has resulted in a number of approaches for protein solubilisation. Nanodiscs are one such method which have seen rapid uptake in recent years. However, due to their relatively recent development, some aspects of their properties and the effects on their components are yet to be fully understood. The research presented here aimed to improve our understanding of these systems and it could be expanded upon to further improve our knowledge and enable the optimisation of nanodisc systems.

Further validation of the pipeline methodology could be undertaken by comparison of output structures with additional experimental data such as that from SAXS experiments of a nanodisc containing a membrane protein. However, the pipeline developed here is already well suited to the study of nanodisc systems in several ways.

The pipeline approach could be used to investigate membrane proteins which undergo large scale conformational changes during their function, such as some transporters with outward-facing and inward-facing conformations. When simulated in both nanodisc and bilayer environments, these transporters could be investigated using free energy calculations³¹⁸ to assess whether the kinetic rates between the two states vary depending on environment.

The methodology presented here could also be utilised to aid experimental design, investigating the impact of variation in acyl chain length and saturation on the embedded membrane protein in a mixed lipid bilayer. Investigation of the order parameters of lipids interacting with the MSPs, indicated that DLPC lipids (12:0) were less perturbed than those with longer tails such as POPC (16:0-18:1). This observation could be used to aid the selection of lipids for membrane protein solubilisation in nanodiscs. If some shorter tail-length lipids were included in the lipid mixture they could potentially form an insulating ring at the interface between the MSPs and the rest of the lipid patch, thereby reducing membrane protein interactions with more highly perturbed lipids. The pipeline developed here would be well suited to investigating this hypothesis, along with the impacts of variation in lipid tail saturation, aiming

to develop systems which better mimic the physiological bilayer from which a membrane protein originated.

The selection of optimal nanodisc diameter used experimentally during membrane protein solubilisation is currently empirical. The pipeline approach described here could be utilised to systematically study the effect of MSP length on protein stability, properties, and interactions with lipids. This information could aid selection of the length of MSPs used experimentally. Further study of membrane proteins with different numbers of TM helices and subunits could also facilitate development of a methodology for the prediction of nanodisc diameter required for solubilisation of a protein of a particular size, thereby ensuring that protein-lipid interactions can occur as they would in a physiological bilayer.

Adjustments could be made to the pipeline to allow the simulation of a more diverse selection of nanodiscs, including those with non-apolipoprotein A1 derived MSPs ^{205, 319} and covalently circularised nanodiscs ³²⁰. The pipeline could be further developed to allow simulation of other recently developed solubilisation methodologies, including SMALPs ³²¹, saposin-lipoprotein nanoparticles ³²² and DNA-corralled nanodiscs ³²³. It could also be combined with the results from the recently developed AlphaFold ³²⁴⁻³²⁵ or RoseTTAFold ³²⁶ approaches, which predict the structures of proteins that have not yet been structurally determined. The combination of these techniques with the pipeline methodology described here would enable embedding of predicted protein structures into nanodiscs, with subsequent simulations

providing information about the length of MSPs and lipid compositions most suited to aiding protein stability during experimental structural determination.

6.2.3 Membrane Protein-Lipid Interactions

The wide-ranging roles and functions performed by the members of the GPCR family, combined with their widespread potential as drug targets, has led to their regulation being the focus of much research and investigation. The extensive roles and effects of cholesterol on biological membranes and their constituent proteins has also prompted study of this lipid. The results presented here detailed the binding of cholesterol at one site on several members of the GPCR family. However, our understanding of protein-lipid interactions and the role of membrane lipids, including cholesterol, in regulation of GPCRs is still incomplete.

While the investigations of the TM345 cholesterol binding site were performed here on simulations of GPCRs in bilayers, further studies could be undertaken to assess for variation in cholesterol binding with simulation in a nanodisc environment. The alterations in lipid properties which have been demonstrated in nanodisc environments could have an impact on protein-lipid interactions, and knowledge of this could aid the identification of lipid binding sites on membrane proteins during structural determination.

Simulations in atomistic detail might provide finer resolution understanding of the interactions which occur between members of the GPCR family and cholesterol at the TM345 cleft. Comparison of atomistic simulations of GPCRs with and without cholesterol could be used to confirm the side chain positions

of the residues in the TM345 binding site. In addition, mutations could be introduced to the residues of the TM345 cleft to either improve or abolish the cholesterol binding site. Comparison with known disease-causing mutations in this region could aid our understanding of these pathological effects.

Further information could also be gained about the role of cholesterol binding at other locations and on other GPCRs, elucidating the conformational changes which occur during binding to affect protein function. GPCRs have been demonstrated to form multimeric complexes when signalling³²⁷⁻³²⁸. The MD approach described here could be adapted to study the interaction between these GPCRs, as well as the role of cholesterol in the interface between GPCRs, as has been observed in dimerisation of ADRB2⁵³.

The presence of a G-protein or mini-G-protein during the simulation would be of interest to assess the effects of binding site conformation on cholesterol binding. The presence of an arrestin bound to the GPCR could also alter the interactions between a GPCR and cholesterol, and investigations of this hypothesis could be performed using the simulation methods and techniques presented here.

Furthermore, the approaches described in this research for the investigation of interactions between cholesterol and GPCRs could be adapted to allow the study of other lipids, for example PIP₂, which has also been proposed to regulate GPCRs¹⁰⁸. This research could therefore provide a basis for detailed investigation of the structure and function of a wide range of membrane proteins, along with their regulation by the lipids of their environment.

6.3 Closing Remarks

Computational approaches such as MD are well suited to the investigation of the interactions between membrane proteins and lipids. This research has provided a new methodology for the simulation of membrane proteins in nanodiscs, investigated protein-lipid interactions in these systems and provided new insights of the TM345 cholesterol binding site on GPCRs. However, the research presented here represents only a fraction of what might be achieved through the use of these methods. Henceforth, MD simulations, together with experimental investigations, will continue to enhance our understanding of membrane proteins and the lipids in their environment which play such a vital role in the cells of all living organisms.

Bibliography

1. Zagnoni, M., Miniaturised technologies for the development of artificial lipid bilayer systems. *Lab on a Chip* **2012**, 12 (6), 1026.
2. Mitra, K.; Ubarretxena-Belandia, I.; Taguchi, T.; Warren, G.; Engelman, D. M., Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol. *Proc. Natl. Acad. Sci. USA* **2004**, 101, 4083–4088.
3. Sud, M.; Fahy, E.; Cotter, D.; Brown, A.; Dennis, E. A.; Glass, C. K.; Merrill, A. H.; Murphy, R. C.; Raetz, C. R. H.; Russell, D. W.; Subramaniam, S., LMSD: LIPID MAPS structure database. *Nucleic Acids Res.* **2007**, 35 (Database), D527-D532.
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.; Dennis, E. A., Update of the LIPID MAPS comprehensive classification system for lipids. *J. Lipid Res.* **2009**, 50 (Supplement), S9-S14.
5. Pertea, M.; Salzberg, S. L., Between a chicken and a grape: estimating the number of human genes. *Genome Biol.* **2010**, 11 (5), 206.
6. Uhlen, M.; Fagerberg, L.; Hallstrom, B. M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, A.; Kampf, C.; Sjostedt, E.; Asplund, A.; Olsson, I.; Edlund, K.; Lundberg, E.; Navani, S.; Szigyanto, C. A. K.; Odeberg, J.; Djureinovic, D.; Takanen, J. O.; Hober, S.; Alm, T.; Edqvist, P. H.; Berling, H.; Tegel, H.; Mulder, J.; Rockberg, J.; Nilsson, P.; Schwenk, J. M.; Hamsten, M.; Von Feilitzen, K.; Forsberg, M.; Persson, L.; Johansson, F.; Zwahlen, M.; Von Heijne, G.; Nielsen, J.; Ponten, F., Tissue-based map of the human proteome. *Science* **2015**, 347 (6220), 1260419-1260419.
7. Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L., Opinion - How many drug targets are there? *Nat. Rev. Drug Discov.* **2006**, 5 (12), 993-996.
8. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucl. Acids Res.* **2000**, 28 (1), 235-242.
9. Newport, T. D.; Sansom, M. S. P.; Stansfeld, P. J., The MemProtMD database: a resource for membrane-embedded protein structures and their lipid interactions. *Nucl. Acids Res.* **2019**, 47 (D1), D390-D397.
10. Stansfeld, P. J.; Goose, J. E.; Caffrey, M.; Carpenter, E. P.; Parker, J. L.; Newstead, N.; Sansom, M. S. P., MemProtMD: automated insertion of membrane protein structures into explicit lipid membranes. *Structure* **2015**, 23, 1350-1361.
11. Yin, H.; Flynn, A. D., Drugging Membrane Protein Interactions. *Annu. Rev. Biomed. Eng.* **2016**, 18 (1), 51-76.
12. Boes, D. M.; Godoy-Hernandez, A.; Mcmillan, D. G. G., Peripheral Membrane Proteins: Promising Therapeutic Targets across Domains of Life. *Membranes* **2021**, 11 (5), 346.

13. Arinaminpathy, Y.; Khurana, E.; Engelman, D. M.; Gerstein, M. B., Computational analysis of membrane proteins: the largest class of drug targets. *Drug Discov. Today* **2009**, *14* (23), 1130-1135.
14. Sriram, K.; Insel, P. A., G protein-coupled receptors as targets for approved drugs: how many targets and how many drugs? *Molec. Pharmacol.* **2018**, *93* (4), 251-258.
15. Da Silva, C. S. B.; Thaler, M.; Tas, A.; Ogando, N. S.; Bredenbeek, P. J.; Ninaber, D. K.; Wang, Y.; Hiemstra, P. S.; Snijder, E. J.; Van Hemert, M. J., Suramin inhibits SARS-CoV-2 infection in cell culture by interfering with early steps of the replication cycle. *Antimicrob. Agents Chemother.* **2020**, *64* (8), e00900-20.
16. Moore, D. T.; Berger, B. W.; Degrado, W. F., Protein-Protein Interactions in the Membrane: Sequence, Structural, and Biological Motifs. *Structure* **2008**, *16* (7), 991-1001.
17. Hansen, S. B.; Tao, X.; MacKinnon, R., Structural basis of PIP₂ activation of the classical inward rectifier K⁺ channel Kir2.2. *Nature* **2011**, *477* (7365), 495-498.
18. Khelashvili, G.; Stanley, N.; Sahai, M. A.; Medina, J.; LeVine, M. V.; Shi, L.; De Fabritiis, G.; Weinstein, H., Spontaneous inward opening of the dopamine transporter is triggered by PIP₂-regulated dynamics of the N-terminus. *ACS Chem. Neurosci.* **2015**, *6* (11), 1825-1837.
19. Parker, J. L.; Corey, R. A.; Stansfeld, P. J.; Newstead, S., Structural basis for substrate specificity and regulation of nucleotide sugar transporters in the lipid bilayer. *Nature Comms.* **2019**, *10*, 4657.
20. Durrant, J.; McCammon, J. A., Molecular dynamics simulations and drug discovery. *BMC Biol.* **2011**, *9* (1), 71.
21. Yeagle, P. L., Non-covalent binding of membrane lipids to membrane proteins. *Biochim. Biophys. Acta* **2014**, *1838* (6), 1548-1559.
22. Wacker, D.; Wang, C.; Katritch, V.; Han, G. W.; Huang, X.-P.; Vardy, E.; McCorvy, J. D.; Jiang, Y.; Chu, M.; Siu, F. Y.; Liu, W.; Xu, H. E.; Cherezov, V.; Roth, B. L.; Stevens, R. C., Structural Features for Functional Selectivity at Serotonin Receptors. *Science* **2013**, *340* (6132), 615-619.
23. Gupta, K.; Li, J.; Liko, I.; Gault, J.; Bechara, C.; Wu, D.; Hopper, J. T. S.; Giles, K.; Benesch, J. L. P.; Robinson, C. V., Identifying key membrane protein lipid interactions using mass spectrometry. *Nat. Protoc.* **2018**, *13*, 1106.
24. Marsh, D., Electron spin resonance in membrane research: Protein-lipid interactions. *Methods* **2008**, *46* (2), 83-96.
25. Marsh, D., Electron spin resonance in membrane research: protein-lipid interactions from challenging beginnings to state of the art. *Eur. Biophys. J.* **2010**, *39* (4), 513-525.
26. Huster, D., Solid-state NMR spectroscopy to study protein-lipid interactions. *Biochim. Biophys. Acta* **2014**, *1841* (8), 1146-1160.

27. Stansfeld, P. J.; Sansom, M. S. P., Molecular simulation approaches to membrane proteins. *Structure* **2011**, *19* (11), 1562-1572.
28. Stansfeld, P. J.; Hopkinson, R.; Ashcroft, F. M.; Sansom, M. S. P., PIP₂-binding site in Kir channels: definition by multiscale biomolecular simulations. *Biochem.* **2009**, *48* (46), 10926-10933.
29. Henderson, R.; Unwin, P. N. T., Three-dimensional model of a purple membrane obtained by electron microscopy. *Nature* **1975**, *257*, 28-32.
30. Iwata, S.; Ostermeier, C.; Ludwig, B.; Michel, H., Structure at 2.8Å resolution of cytochrome c oxidase from *Paracoccus denitrificans*. *Nature* **1995**, *376*, 660-669.
31. Schiltz, E.; Kreusch, A.; Nestel, U.; Schulz, G. E., Primary structure of porin from *Rhodobacter capsulatus*. *Eur. J. Biochem.* **1991**, *199* (3), 587-594.
32. Henderson, R.; Baldwin, J. M.; Ceska, T. A.; Zemlin, F.; Beckmann, E.; Downing, K. H., Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* **1990**, *213* (4), 899-929.
33. PebayPeyroula, E.; Rummel, G.; Rosenbusch, J. P.; Landau, E. M., X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases. *Science* **1997**, *277*, 1676-1681.
34. Opella, S. J.; Marassi, F. M., Structure determination of membrane proteins by NMR spectroscopy. *Chem. Rev.* **2004**, *104* (8), 3587-3606.
35. Hiller, S.; Wagner, G., The role of solution NMR in the structure determinations of VDAC-1 and other membrane proteins. *Curr. Opin. Struct. Biol.* **2009**, *19* (4), 396-401.
36. Guo, Y., Be Cautious with Crystal Structures of Membrane Proteins or Complexes Prepared in Detergents. *Crystals* **2020**, *10* (2), 86.
37. Landau, E. M.; Rosenbusch, J. P., Lipidic cubic phases: A novel concept for the crystallization of membrane proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93* (25), 14532-14535.
38. Cherezov, V., Lipidic cubic phase technologies for membrane protein structural studies. *Curr. Opin. Struct. Biol.* **2011**, *21* (4), 559-566.
39. Rosenbaum Daniel, M.; Cherezov, V.; Hanson Michael, A.; Rasmussen Søren, G. F.; Thian Foon, S.; Kobilka Tong, S.; Choi, H.-J.; Yao, X.-J.; Weis William, I.; Stevens Raymond, C.; Kobilka Brian, K., GPCR Engineering Yields High-Resolution Structural Insights into β 2-Adrenergic Receptor Function. *Science* **2007**, *318* (5854), 1266-1273.
40. Doyle, D. A.; Cabral, J. M.; Pfuetzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Cahit, B. T.; MacKinnon, R., The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* **1998**, *280*, 69-77.
41. de Oliveira, T. M.; van Beek, L.; Shilliday, F.; Debreczeni, J. É.; Phillips, C., Cryo-EM: The Resolution Revolution and Drug Discovery. *SLAS Discov.* **2020**, *26* (1), 17-31.

42. Kühlbrandt, W., The Resolution Revolution. *Science* **2014**, *343* (6178), 1443-1444.
43. McMullan, G.; Faruqi, A. R.; Clare, D.; Henderson, R., Comparison of optimal performance at 300keV of three direct electron detectors for use in low dose electron microscopy. *Ultramicroscopy* **2014**, *147*, 156-163.
44. Rivera-Calzada, A.; Carroni, M., Editorial: Technical Advances in Cryo-Electron Microscopy. *Front. Mol. Biosci.* **2019**, *6* (72).
45. Brilot, A. F.; Chen, J. Z.; Cheng, A.; Pan, J.; Harrison, S. C.; Potter, C. S.; Carragher, B.; Henderson, R.; Grigorieff, N., Beam-induced motion of vitrified specimen on holey carbon film. *J. Struct. Biol.* **2012**, *177* (3), 630-637.
46. Li, X.; Mooney, P.; Zheng, S.; Booth, C. R.; Braunfeld, M. B.; Gubbens, S.; Agard, D. A.; Cheng, Y., Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **2013**, *10* (6), 584-590.
47. Nogales, E.; Scheres, H. W., Sjors, Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Mol. Cell* **2015**, *58* (4), 677-689.
48. Musselman, C. A.; Kutateladze, T. G., Characterization of functional disordered regions within chromatin-associated proteins. *iScience* **2021**, *24* (2), 102070.
49. García-Nafría, J.; Tate, C. G., Structure determination of GPCRs: cryo-EM compared with X-ray crystallography. *Biochem. Soc. Trans.* **2021**, *49* (5), 2345-2355.
50. Liao, M. F.; Cao, E. H.; Julius, D.; Cheng, Y. F., Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **2013**, *504* (7478), 107-112.
51. Du, J.; Lu, W.; Wu, S. P.; Cheng, Y. F.; Gouaux, E., Glycine receptor mechanism elucidated by electron cryo-microscopy. *Nature* **2015**, *526* (7572), 224-229.
52. Matthies, D.; Dalmas, O.; Borgnia, M. J.; Dominik, P. K.; Merk, A.; Rao, P.; Reddy, B. G.; Islam, S.; Bartesaghi, A.; Perozo, E.; Subramaniam, S., Cryo-EM Structures of the Magnesium Channel CorA Reveal Symmetry Break upon Gating. *Cell* **2016**, *164* (4), 747-756.
53. Prasanna, X.; Chattopadhyay, A.; Sengupta, D., Cholesterol modulates the dimer interface of the β 2-adrenergic receptor via cholesterol occupancy sites. *Biophys. J.* **2014**, *106* (6), 1290-1300.
54. Zhang, H.; Kurisu, G.; Smith, J. L.; Cramer, W. A., A defined protein-detergent-lipid complex for crystallization of integral membrane proteins: The cytochrome b6f complex of oxygenic photosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (9), 5160-5163.
55. Rosenbusch, J. P.; Lustig, A.; Grabo, M.; Zulauf, M.; Regenass, M., Approaches to determining membrane protein structures to high resolution: do selections of subpopulations occur? *Micron* **2001**, *32* (1), 75-90.

56. Dürr, U. H. N.; Gildenberg, M.; Ramamoorthy, A., The Magic of Bicelles Lights Up Membrane Protein Structure. *Chem. Rev.* **2012**, *112* (11), 6054-6074.
57. Akbarzadeh, A.; Rezaei-Sadabady, R.; Davaran, S.; Joo, S. W.; Zarghami, N.; Hanifehpour, Y.; Samiei, M.; Kouhi, M.; Nejati-Koshki, K., Liposome: classification, preparation, and applications. *Nanoscale Res. Lett.* **2013**, *8* (1), 102.
58. Mio, K.; Sato, C., Lipid environment of membrane proteins in cryo-EM based structural analysis. *Biophys Rev* **2018**, *10* (2), 307-316.
59. Valiyaveetil, F. I.; Zhou, Y.; MacKinnon, R., Lipids in the Structure, Folding, and Function of the KcsA K⁺ Channel. *Biochemistry* **2002**, *41* (35), 10771-10777.
60. Marius, P.; de Planque, M. R. R.; Williamson, P. T. F., Probing the interaction of lipids with the non-annular binding sites of the potassium channel KcsA by magic-angle spinning NMR. *Biochim. Biophys. Acta* **2012**, *1818* (1), 90-96.
61. Gimpl, G.; Reitz, J.; Brauer, S.; Trossen, C., Oxytocin receptors: ligand binding, signalling and cholesterol dependence. *Prog. Brain Res.* **2008**, *170*, 193-204.
62. Popot, J. L., Amphipols, nanodiscs, and fluorinated surfactants: three nonconventional approaches to studying membrane proteins in aqueous solutions. *Annu. Rev. Biochem.* **2010**, *79*, 737.
63. Tribet, C.; Audebert, R.; Popot, J.-L., Amphipols: Polymers that keep membrane proteins soluble in aqueous solutions. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93* (26), 15047-15050.
64. Lyons, J. A.; Bøggild, A.; Nissen, P.; Frauenfeld, J., Saposin-Lipoprotein Scaffolds for Structure Determination of Membrane Transporters. *Meth. Enzymol.* **2017**, 85-99.
65. Jamshad, M.; Lin, Y.-P.; Knowles, J., Timothy; Parslow, A., Rosemary; Harris, C.; Wheatley, M.; Poyner, R., David; Bill, M., Roslyn; Thomas, R. T., Owen; Overduin, M.; Dafforn, R., Tim, Surfactant-free purification of membrane proteins with intact native membrane environment. *Biochem. Soc. Trans.* **2011**, *39* (3), 813-818.
66. Denisov, I. G.; Grinkova, Y. V.; Lazarides, A. A.; Sligar, S. G., Directed self-assembly of monodisperse phospholipid bilayer nanodiscs with controlled size. *J. Amer. Chem. Soc.* **2004**, *126* (11), 3477-3487.
67. Bayburt, T. H.; Grinkova, Y. V.; Sligar, S. G., Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins. *Nano Lett.* **2002**, *2* (8), 853-856.
68. Ritchie, T. K.; Grinkova, Y. V.; Bayburt, T. H.; Denisov, I. G.; Zolnerciks, J. K.; Atkins, W. M.; Sligar, S. G., Reconstitution of membrane proteins in phospholipid bilayer nanodiscs. *Meth. Enzymol.* **2009**, *464*, 211-231.

69. Denisov, I. G.; Sligar, S. G., Nanodiscs for structural and functional studies of membrane proteins. *Nat. Struct. Mol. Biol.* **2016**, 23 (6), 481-486.
70. Efremov, R. G.; Gatsogiannis, C.; Raunser, S., Lipid nanodiscs as a tool for high- resolution structure determination of membrane proteins by single-particle cryo-EM. *Meth. Enzymol.* **2017**, 594 (Structure-Function Toolbox for Membrane Transporter and Channels), 1-30.
71. Lee, A. G., How lipids affect the activities of integral membrane proteins. *Biochim. Biophys. Acta* **2004**, 1666, 62-87.
72. Ackerman, D. G.; Feigenson, G. W., Lipid bilayers: clusters, domains and phases. *Essays Biochem.* **2015**, 57, 33-42.
73. Bleecker, J. V.; Cox, P. A.; Foster, R. N.; Litz, J. P.; Blosser, M. C.; Castner, D. G.; Keller, S. L., Thickness Mismatch of Coexisting Liquid Phases in Noncanonical Lipid Bilayers. *J. Phys. Chem. B* **2016**, 120 (10), 2761-2770.
74. Powl, A. M.; East, J. M.; Lee, A. G., Importance of Direct Interactions with Lipids for the Function of the Mechanosensitive Channel MscL. *Biochemistry* **2008**, 47 (46), 12175-12184.
75. van Meer, G.; Voelker, D. R.; Feigenson, G. W., Membrane lipids: where they are and how they behave. *Nature Rev. Mol. Cell Biol.* **2008**, 9, 112-124.
76. Fabri, J. H. T. M.; De Sá, N. P.; Malavazi, I.; Del Poeta, M., The dynamics and role of sphingolipids in eukaryotic organisms upon thermal adaptation. *Prog. Lipid Res.* **2020**, 80, 101063.
77. Breslow, D. K.; Weissman, J. S., Membranes in Balance: Mechanisms of Sphingolipid Homeostasis. *Mol. Cell* **2010**, 40 (2), 267-279.
78. Dufourc, E. J., Sterols and membrane dynamics. *J. Chem. Biol.* **2008**, 1 (1-4), 63-77.
79. Appala, K.; Bimpeh, K.; Freeman, C.; Hines, K. M., Recent applications of mass spectrometry in bacterial lipidomics. *Anal. Bioanal. Chem.* **2020**, 412 (24), 5935-5943.
80. Roger, A. J.; Muñoz-Gómez, S. A.; Kamikawa, R., The Origin and Diversification of Mitochondria. *Curr. Biol.* **2017**, 27 (21), R1177-R1192.
81. Pomorski, T. G.; Nylander, T.; Cárdenas, M., Model cell membranes: Discerning lipid and protein contributions in shaping the cell. *Adv. Colloid Interface Sci.* **2014**, 205, 207-220.
82. Funai, K.; Summers, S. A.; Rutter, J., Reign in the membrane: How common lipids govern mitochondrial function. *Curr. Opin. Cell Biol.* **2020**, 63, 162-173.
83. Szule, J. A.; Fuller, N. L.; Peter Rand, R., The Effects of Acyl Chain Length and Saturation of Diacylglycerols and Phosphatidylcholines on Membrane Monolayer Curvature. *Biophys. J.* **2002**, 83 (2), 977-984.
84. Berman, H.; Henrick, K.; Nakamura, H., Announcing the worldwide Protein Data Bank. *Nature Struct. Biol.* **2003**, 10, 980.

85. Marrink, S. J.; Tieleman, D. P., Perspective on the Martini model. *Chem. Soc. Rev.* **2013**, 42 (16), 6801-6822.
86. Voth, G. A., *Coarse-Graining of Condensed Phase and Biomolecular Systems*. CRC Press: 2008; p 456.
87. Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H., The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, 111, 7812.
88. Stansfeld, P. J.; Sansom, M. S. P., From coarse-grained to atomistic: a serial multi-scale approach to membrane protein simulations. *J. Chem. Theor. Comp.* **2011**, 7, 1157–1166.
89. Wassenaar, T. A.; Pluhackova, K.; Böckmann, R. A.; Marrink, S. J.; Tieleman, D. P., Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models. *J. Chem. Theor. Comput.* **2014**, 10 (2), 676-690.
90. Vickery, O. N.; Stansfeld, P. J., CG2AT2: an Enhanced Fragment-Based Approach for Serial Multi-scale Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2021**, 17 (10), 6472-6482.
91. Lee, A. G., Lipid–protein interactions in biological membranes: a structural perspective. *Biochim. Biophys. Acta* **2003**, 1612 (1), 1-40.
92. Marsh, D.; Horváth, L. I., Structure, dynamics and composition of the lipid-protein interface. Perspectives from spin-labelling. *Biochim. Biophys. Acta* **1998**, 1376 (3), 267-296.
93. Marsh, D.; Páli, T., The protein–lipid interface: perspectives from magnetic resonance and crystal structures. *Biochim. Biophys. Acta* **2004**, 1666 (1-2), 118-141.
94. Koldsø, H.; Sansom, M. S. P., Local lipid reorganization by a transmembrane protein domain. *J. Phys. Chem. Lett.* **2012**, 3, 3498-3502.
95. Hedger, G.; Sansom, M. S. P.; Koldsø, H., The juxtamembrane regions of human receptor tyrosine kinases exhibit conserved interaction sites with anionic lipids. *Sci. Reports* **2015**, 5, 9198.
96. Zhou, H.-X.; Cross, T. A., Influences of Membrane Mimetic Environments on Membrane Protein Structures. *Annu Rev Biophys* **2013**, 42 (1), 361-392.
97. Kinnun, J. J.; Mallikarjunaiah, K. J.; Petrache, H. I.; Brown, M. F., Elastic deformation and area per lipid of membranes: Atomistic view from solid-state deuterium NMR spectroscopy. *Biochim. Biophys. Acta* **2015**, 1848 (1), 246-259.
98. Sheetz, M. P.; Singer, S. J., Biological Membranes as Bilayer Couples. A Molecular Mechanism of Drug-Erythrocyte Interactions. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, 71 (11), 4457-4461.
99. Gruner, S. M., Intrinsic curvature hypothesis for biomembrane lipid composition: a role for nonbilayer lipids. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, 82 (11), 3665-3669.

100. Gullingsrud, J.; Schulten, K., Lipid bilayer pressure profiles and mechanosensitive channel gating. *Biophys. J.* **2004**, *86* (6), 3496-3509.
101. Poveda, J. A.; Giudici, A. M.; Renart, M. L.; Molina, M. L.; Montoya, E.; Fernández-Carvajal, A.; Fernández-Ballester, G.; Encinar, J. A.; González-Ros, J. M., Lipid modulation of ion channels through specific binding sites. *Biochim. Biophys. Acta* **2014**, *1838* (6), 1560-1567.
102. Valiyaveetil, F. I.; Zhou, Y.; MacKinnon, R., Lipids in the structure, folding and function of the KcsA channel. *Biochem.* **2002**, *41*, 10771-10777.
103. Deol, S. S.; Domene, C.; Bond, P. J.; Sansom, M. S. P., Anionic phospholipids interactions with the potassium channel KcsA: simulation studies. *Biophys. J.* **2006**, *90*, 822-830.
104. Neale, C.; Herce, Henry D.; Pomès, R.; García, Angel E., Can specific protein-lipid interactions stabilize an active state of the beta 2 adrenergic receptor? *Biophys. J.* **2015**, *109* (8), 1652-1662.
105. Hansen, S. B., Lipid agonism: The PIP₂ paradigm of ligand-gated ion channels. *Biochim. Biophys. Acta* **2015**, *1851* (5), 620-628.
106. Lopes, C. M. B.; Zhang, H. L.; Rohacs, T.; Jin, T. H.; Yang, J.; Logothetis, D. E., Alterations in conserved Kir channel-PIP₂ interactions underlie channelopathies. *Neuron* **2002**, *34* (6), 933-944.
107. Suh, B. C.; Hille, B., PIP₂ is a necessary cofactor for ion channel function: How and why? *Ann. Rev. Biophys.* **2008**, *37*, 175-195.
108. Yen, H. Y.; Hoi, K. K.; Liko, I.; Hedger, G.; Horrell, M. R.; Song, W. L.; Wu, D.; Heine, P.; Warne, T.; Lee, Y.; Carpenter, B.; Pluckthun, A.; Tate, C. G.; Sansom, M. S. P.; Robinson, C. V., PtdIns(4,5)P₂ stabilizes active states of GPCRs and enhances selectivity of G-protein coupling. *Nature* **2018**, *559* (7714), 424-427.
109. Song, W. L.; Yen, H. Y.; Robinson, C. V.; Sansom, M. S. P., State-dependent lipid interactions with the A_{2a} receptor revealed by MD simulations using in vivo-mimetic membranes. *Structure* **2019**, *27* (2), 392-403.
110. Komolov, K. E.; Du, Y.; Duc, N. M.; Betz, R. M.; Rodrigues, J.; Leib, R. D.; Patra, D.; Skiniotis, G.; Adams, C. M.; Dror, R. O.; Chung, K. Y.; Kobilka, B. K.; Benovic, J. L., Structural and functional analysis of a β 2-adrenergic receptor complex with GRK5. *Cell* **2017**, *169* (3), 407-421.
111. Song, W.; Duncan, A. L.; Sansom, M. S. P., Modulation of adenosine A_{2a} receptor oligomerization by receptor activation and PIP₂ interactions. *Structure* **2021**, *29* (11), 1312-1325.e3.
112. Mouritsen, O. G.; Zuckermann, M. J., What's so special about cholesterol? *Lipids* **2004**, *39* (11), 1101-1113.
113. Peters, C.; Wolf, A.; Wagner, M.; Kuhlmann, J.; Waldmann, H., The cholesterol membrane anchor of the Hedgehog protein confers stable membrane association to lipid-modified proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (23), 8531-8536.

114. Silvius, J. R., Role of cholesterol in lipid raft formation: lessons from lipid model systems. *Biochim. Biophys. Acta* **2003**, *1610* (2), 174-183.
115. Edidin, M., The state of lipid rafts: from model membranes to cells. *Ann. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 257-283.
116. Saslowsky, D. E.; Lawrence, J.; Ren, X.; Brown, D. A.; Henderson, R. M.; Edwardson, J. M., Placental Alkaline Phosphatase Is Efficiently Targeted to Rafts in Supported Lipid Bilayers. *J. Biol. Chem.* **2002**, *277* (30), 26966-26970.
117. Gandhavadi, M.; Allende, D.; Vidal, A.; Simon, S. A.; McIntosh, T. J., Structure, Composition, and Peptide Binding Properties of Detergent Soluble Bilayers and Detergent Resistant Rafts. *Biophys. J.* **2002**, *82* (3), 1469-1482.
118. Song, Y.; Kenworthy, A. K.; Sanders, C. R., Cholesterol as a co-solvent and a ligand for membrane proteins. *Protein Sci.* **2014**, *23* (1), 1-22.
119. Marsh, D., Protein modulation of lipids, and vice-versa, in membranes. *Biochim. Biophys. Acta Biomembranes* **2008**, *1778* (7-8), 1545-1575.
120. Duncan, A. L.; Song, W.; Sansom, M. S. P., Lipid-dependent regulation of ion channels and GPCRs: insights from structures and simulations. *Ann. Rev. Pharmacol. Toxicol.* **2020**, *60*, 31-50.
121. Hanson, M. A.; Cherezov, V.; Griffith, M. T.; Roth, C. B.; Jaakola, V.-P.; Chien, E. Y. T.; Velasquez, J.; Kuhn, P.; Stevens, R. C., A specific cholesterol binding site is established by the 2.8 Å structure of the human β 2-adrenergic receptor. *Structure* **2008**, *16* (6), 897-905.
122. Manna, M.; Niemelä, M.; Tynkkynen, J.; Javanainen, M.; Kulig, W.; Müller, D. J.; Rog, T.; Vattulainen, I., Mechanism of allosteric regulation of β 2-adrenergic receptor by cholesterol. *eLife* **2016**, *5*, e18432.
123. 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.; Stevens, R. C., High-resolution crystal structure of an engineered human β 2-adrenergic G protein-coupled receptor. *Science* **2007**, *318* (5854), 1258-1265.
124. Prasanna, X.; Sengupta, D.; Chattopadhyay, A., Cholesterol-dependent conformational plasticity in GPCR dimers. *Sci. Reports* **2016**, *6*, 31858.
125. Schlame, M., Thematic Review Series: Glycerolipids. Cardiolipin synthesis for the assembly of bacterial and mitochondrial membranes. *J. Lipid Res.* **2008**, *49* (8), 1607-1620.
126. Kawai, F.; Shoda, M.; Harashima, R.; Sadaie, Y.; Hara, H.; Matsumoto, K., Cardiolipin Domains in *Bacillus subtilis* Marburg Membranes. *J. Bacteriol.* **2004**, *186* (5), 1475-1483.
127. Uphoff, A.; Hermansson, M.; Haimi, P.; Somerharju, P., Analysis of complex lipidomes. In *Medical Applications of Mass Spectrometry*, Elsevier: 2008; pp 223-249.
128. Patrick, J. W.; Boone, C. D.; Liu, W.; Conover, G. M.; Liu, Y.; Cong, X.; Laganowsky, A., Allostery revealed within lipid binding events to membrane proteins. *Proc. Natl. Acad. Sci. USA* **2018**, *115* (12), 2976-2981.

129. Khalid, S.; Rouse, S. L., Simulation of subcellular structures. *Curr. Opin. Struct. Biol.* **2020**, *61*, 167-172.
130. Hedger, G.; Rouse, S. L.; Domański, J.; Chavent, H.; Koldsø, H.; Sansom, M. S. P., Lipid loving ANTs: molecular simulations of cardiolipin interactions and the organization of the adenine nucleotide translocase in model mitochondrial membranes. *Biochem.* **2016**, *55*, 6238–6249.
131. Nolte, R. T.; Atkinson, D., Conformational analysis of apolipoprotein A-I and E-3 based on primary sequence and circular dichroism. *Biophys. J.* **1992**, *63* (5), 1221-1239.
132. Phillips, J. C.; Wriggers, W.; Li, Z.; Jonas, A.; Schulten, K., Predicting the structure of apolipoprotein A-I in reconstituted high-density lipoprotein disks. *Biophys. J.* **1997**, *73* (5), 2337-2346.
133. Tricerri, M. A.; Behling Agree, A. K.; Sanchez, S. A.; Bronski, J.; Jonas, A., Arrangement of Apolipoprotein A-I in Reconstituted High-Density Lipoprotein Disks: An Alternative Model Based on Fluorescence Resonance Energy Transfer Experiments. *Biochemistry* **2001**, *40* (16), 5065-5074.
134. Segrest, J. P.; Jones, M. K.; Klon, A. E.; Sheldahl, C. J.; Hellinger, M.; De Loof, H.; Harvey, S. C., A detailed molecular belt model for apolipoprotein A-I in discoidal high density lipoprotein. *J. Biol. Chem.* **1999**, *274*, 31755.
135. Carney, C. E.; Lenov, I. L.; Baker, C. J.; Macrenaris, K. W.; Eckermann, A. L.; Sligar, S. G.; Meade, T. J., Nanodiscs as a Modular Platform for Multimodal MR-Optical Imaging. *Bioconj. Chem.* **2015**, *26* (5), 899-905.
136. Kijac, A.; Shih, A. Y.; Nieuwkoop, A. J.; Schulten, K.; Sligar, S. G.; Rienstra, C. M., Lipid-protein correlations in nanoscale phospholipid bilayers determined by solid-state nuclear magnetic resonance. *Biochemistry* **2010**, *49*, 9190.
137. Bibow, S.; Polyhach, Y.; Eichmann, C.; Chi, C. N.; Kowal, J.; Albiez, S.; Mcleod, R. A.; Stahlberg, H.; Jeschke, G.; Güntert, P.; Riek, R., Solution structure of discoidal high-density lipoprotein particles with a shortened apolipoprotein A-I. *Nat. Struct. Mol. Biol.* **2017**, *24* (2), 187-193.
138. Nath, A.; Atkins, W. M.; Sligar, S. G., Applications of phospholipid bilayer nanodiscs in the study of membranes and membrane proteins. *Biochemistry* **2007**, *46*, 2059.
139. Pourmousa, M.; Pastor, R. W., Molecular dynamics simulations of lipid nanodiscs. *Biochim. Biophys. Acta-Biomembranes* **2018**, *1860* (10), 2094-2107.
140. Siuda, I.; Tieleman, D. P., Molecular models of nanodiscs. *J. Chem. Theor. Comput.* **2015**, *11* (10), 4923-4932.
141. Bengtsen, T.; Holm, V. L.; Kjolbye, L. R.; Midtgaard, S. R.; Johansen, N. T.; Tesei, G.; Bottaro, S.; Schiott, B.; Arleth, L.; Lindorff-Larsen, K., Structure and dynamics of a nanodisc by integrating NMR, SAXS and SANS experiments with molecular dynamics simulations. *eLife* **2020**, *9*, e56518

142. Skar-Gislinge, N.; Simonsen, J. B.; Mortensen, K.; Feidenhans'L, R.; Sligar, S. G.; Lindberg Møller, B.; Bjørnholm, T.; Arleth, L., Elliptical Structure of Phospholipid Bilayer Nanodiscs Encapsulated by Scaffold Proteins: Casting the Roles of the Lipids and the Protein. *J. Am. Chem. Soc.* **2010**, *132* (39), 13713-13722.
143. Alder, B. J.; Wainwright, T. E., Phase Transition for a Hard Sphere System. *J. Chem. Phys.* **1957**, *27* (5), 1208-1209.
144. Mccammon, J. A.; Gelin, B. R.; Karplus, M., Dynamics of folded proteins. *Nature* **1977**, *267* (5612), 585-590.
145. Lifson, S.; Warshel, A., Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkane andn- Alkane Molecules. *J. Chem. Phys.* **1968**, *49* (11), 5116-5129.
146. Levitt, M.; Lifson, S., Refinement of protein conformations using a macromolecular energy minimization procedure. *J. Mol. Biol.* **1969**, *46* (2), 269-279.
147. Berman, H.; Henrick, K.; Nakamura, H.; Markley, J. L., The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.* **2007**, *35*, D301-D303.
148. Lonsdale, R.; Harvey, J. N.; Mulholland, A. J., A practical guide to modelling enzyme-catalysed reactions. *Chem. Soc. Rev.* **2012**, *41* (8), 3025.
149. Gundelach, L.; Fox, T.; Tautermann, C. S.; Skylaris, C.-K., Protein–ligand free energies of binding from full-protein DFT calculations: convergence and choice of exchange–correlation functional. *Phys. Chem. Chem. Phys.* **2021**, *23* (15), 9381-9393.
150. Prentice, J. C. A.; Aarons, J.; Womack, J. C.; Allen, A. E. A.; Andrinopoulos, L.; Anton, L.; Bell, R. A.; Bhandari, A.; Bramley, G. A.; Charlton, R. J.; Clements, R. J.; Cole, D. J.; Constantinescu, G.; Corsetti, F.; Dubois, S. M.-M.; Duff, K. K. B.; Escartín, J. M.; Greco, A.; Hill, Q.; Lee, L. P.; Linscott, E.; O'Regan, D. D.; Phipps, M. J. S.; Ratcliff, L. E.; Serrano, Á. R.; Tait, E. W.; Teobaldi, G.; Vitale, V.; Yeung, N.; Zuehlsdorff, T. J.; Dziedzic, J.; Haynes, P. D.; Hine, N. D. M.; Mostofi, A. A.; Payne, M. C.; Skylaris, C.-K., The ONETEP linear-scaling density functional theory program. *J. Chem. Phys.* **2020**, *152* (17), 174111.
151. Leach, A. R., *Molecular modelling. Principles and applications*. 2nd ed.; Longman: 2001.
152. González, M. A., Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique* **2011**, *12*, 169-200.
153. Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu,

- X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: The biomolecular simulation program. *J Comput Chem* **2009**, 30 (10), 1545-1614.
154. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J Comput Chem* **2005**, 26 (16), 1781-1802.
155. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *J Comput Chem* **2005**, 26 (16), 1668-1688.
156. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., GROMACS: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.* **1995**, 95, 43-56.
157. van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J., GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, 26 (16), 1701-1718.
158. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, 1–2, 19-25.
159. Chang, C.-E.; Huang, Y.-M.; Mueller, L.; You, W., Investigation of Structural Dynamics of Enzymes and Protonation States of Substrates Using Computational Tools. *Catalysts* **2016**, 6 (6), 82.
160. Hockney, R. W.; Goel, S. P.; Eastwood, J. W., Quiet high-resolution computer models of a plasma. *J. Comput. Phys.* **1974**, 14 (2), 148-158.
161. Verlet, L., Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard Jones molecules. *Phys. Rev.* **1967**, 159, 98-103.
162. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D., CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, 14 (1), 71-73.
163. Loubet, B.; Kopec, W.; Khandelia, H., Accelerating All-Atom MD Simulations of Lipids Using a Modified Virtual-Sites Technique. *J. Chem. Theory Comput.* **2014**, 10 (12), 5690-5695.
164. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.* **1997**, 18, 1463-1472.
165. Hess, B., P-LINCS: a parallel linear constraint solver for molecular simulation. *J. Chem. Theor. Comp.* **2007**, 4 (1), 116-122.
166. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, 81, 3684-3690.
167. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, 126 (1), 014101.

168. Parrinello, M.; Giaquinta, P. V., Analytical solution of a new integral equation for triplet correlations in hard sphere fluids. *J. Chem. Phys.* **1981**, *74* (3), 1990-1997.
169. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103* (19), 8577-8593.
170. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald - an $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98* (12), 10089-10092.
171. Barker, J. A.; Watts, R. O., Monte Carlo studies of the dielectric properties of water-like models. *Mol. Phys.* **1973**, *26* (3), 789-792.
172. Watts, R. O., Monte Carlo studies of liquid water. *Mol. Phys.* **1974**, *28* (4), 1069-1083.
173. Lindert, S.; Bucher, D.; Eastman, P.; Pande, V.; Mccammon, J. A., Accelerated Molecular Dynamics Simulations with the AMOEBA Polarizable Force Field on Graphics Processing Units. *J. Chem. Theory Comput.* **2013**, *9* (11), 4684-4691.
174. Ozboyaci, M.; Kokh, D. B.; Corni, S.; Wade, R. C., Modeling and simulation of protein-surface interactions: achievements and challenges. *Quart. Rev. Biophys.* **2016**, *49*, 1-45.
175. Van Der Kamp, M. W.; Mulholland, A. J., Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **2013**, *52* (16), 2708-2728.
176. Slochow, D. R.; Huwe, P. J.; Radhakrishnan, R.; Janmey, P. A., Quantum and all-atom molecular dynamics simulations of protonation and divalent ion binding to phosphatidylinositol 4,5-bisphosphate (PIP₂). *J. Phys.Chem. B* **2013**, *117* (28), 8322-8329.
177. Domene, C.; Sansom, M. S. P., A potassium channel, ions and water: simulation studies based on the high resolution X-ray structure of KcsA. *Biophys. J.* **2003**, *85*, 2787-2800.
178. Machado, M. R.; Barrera, E. E.; Klein, F.; Sónora, M.; Silva, S.; Pantano, S., The SIRAH 2.0 Force Field: Altius, Fortius, Citius. *J. Chem. Theory Comput.* **2019**, *15* (4), 2719-2733.
179. Chavent, M.; Duncan, A. L.; Rassam, P.; Birholz, O.; Hélie, J.; Reddy, T.; Belaeyv, D.; Hambly, B.; Piehler, J.; Kleanthous, C.; Sansom, M. S. P., How nanoscale protein interactions determine the mesoscale dynamic organisation of bacterial outer membrane proteins. *Nature Comms.* **2018**, *9*, 2846.
180. Limbach, H. J.; Arnold, A.; Mann, B. A.; Holm, C., ESPResSo - An extensible simulation package for research on soft matter systems. *Comput. Phys. Commun.* **2006**, *174*, 704-727.

181. Chavent, M.; Duncan, A. L.; Sansom, M. S. P., Molecular dynamics simulations of membrane proteins and their interactions: from nanoscale to mesoscale. *Curr. Opin. Struct. Biol.* **2016**, *40*, 8-16.
182. Marrink, S. J.; de Vries, A. H.; Mark, A. E., Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B* **2004**, *108*, 750-760.
183. Marrink, S. J.; Risselada, J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H., The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, *111*, 7812-7824.
184. de Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J., Improved parameters for the Martini coarse-grained protein force field. *J. Chem. Theor. Comput.* **2013**, *9* (1), 687-697.
185. Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A.; Kroon, P. C.; Melcr, J.; Nieto, V.; Corradi, V.; Khan, H. M.; Domański, J.; Javanainen, M.; Martinez-Seara, H.; Reuter, N.; Best, R. B.; Vattulainen, I.; Monticelli, L.; Periole, X.; Tieleman, D. P.; de Vries, A. H.; Marrink, S. J., Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **2021**, *18* (4), 382-388.
186. Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J., The MARTINI coarse-grained force field: extension to proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819.
187. Bernstein, F.; Koetzle, T.; Williams, G.; Meyer, E.; Brice, M.; Rodgers, J.; Kennard, O.; Shimanouchi, T.; Tasumi, M., The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535-542.
188. Kabsch, W.; Sander, C., Dictionary of protein secondary structure: pattern-recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577-2637.
189. Periole, X.; Cavalli, M.; Marrink, S. J.; Ceruso, M. A., Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. *J. Chem. Theory Comput.* **2009**, *5* (9), 2531-2543.
190. Gu, R. X.; Ingolfsson, H. I.; de Vries, A. H.; Marrink, S. J.; Tieleman, D. P., Ganglioside-lipid and ganglioside-protein interactions revealed by coarse-grained and atomistic molecular dynamics simulations. *J. Phys. Chem. B* **2017**, *121* (15), 3262-3275.
191. Arnarez, C.; Uusitalo, J. J.; Masman, M. F.; Ingólfsson, H. I.; De Jong, D. H.; Melo, M. N.; Periole, X.; De Vries, A. H.; Marrink, S. J., Dry Martini, a Coarse-Grained Force Field for Lipid Membrane Simulations with Implicit Solvent. *J. Chem. Theory Comput.* **2015**, *11* (1), 260-275.

192. Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J., The MARTINI coarse grained force field: extension to proteins. *J. Chem. Theor. Comp.* **2008**, *4*, 819-834.
193. Scott, K. A.; Bond, P. J.; Ivetac, A.; Chetwynd, A. P.; Khalid, S.; Sansom, M. S. P., Coarse-grained MD simulations of membrane protein-bilayer self-assembly *Structure* **2008**, *16*, 621-630.
194. Jo, S.; Cheng, X.; Lee, J.; Kim, S.; Park, S. J.; Patel, D. S.; Beaven, A. H.; Lee, K. I.; Rui, H.; Park, S.; Lee, H. S.; Roux, B.; MacKerell, A. D.; Klauda, J. B.; Qi, Y. F.; Im, W., CHARMM-GUI 10 years for biomolecular modeling and simulation. *J. Comp. Chem.* **2017**, *38* (15), 1114-1124.
195. Wassenaar, T. A.; Ingólfsson, H. I.; Böckmann, R. A.; Tieleman, D. P.; Marrink, S. J., Computational lipidomics with *insane*: a versatile tool for generating custom membranes for molecular simulations. *J. Chem. Theor. Comput.* **2015**, *11* (5), 2144-2155.
196. Carpenter, E. P.; Beis, K.; Cameron, A. D.; Iwata, S., Overcoming the challenges of membrane protein crystallography. *Curr. Opin. Struct. Biol.* **2008**, *18* (5), 581-586.
197. Privé, G. G., Detergents for the stabilization and crystallization of membrane proteins. *Methods* **2007**, *41* (4), 388-397.
198. Simon, K. S.; Pollock, N. L.; Lee, S. C., Membrane protein nanoparticles: the shape of things to come. *Biochem. Soc. Trans.* **2018**, *46* (6), 1495-1504.
199. Tascón, I.; Sousa, J. S.; Corey, R. A.; Mills, D. J.; Griwatz, D.; Aumüller, N.; Mikusevic, V.; Stansfeld, P. J.; Vonck, J.; Hänelt, I., Structural basis of proton-coupled potassium transport in the KUP family. *Nat. Commun.* **2020**, *11* (1).
200. Chen, Q. F.; She, J.; Zeng, W. Z.; Guo, J. T.; Xu, H. X.; Bai, X. C.; Jiang, Y. X., Structure of mammalian endolysosomal TRPML1 channel in nanodiscs. *Nature* **2017**, *550* (7676), 415-418.
201. Liang, Y. L.; Khoshouei, M.; Radjainia, M.; Zhang, Y.; Glukhova, A.; Tarrasch, J.; Thal, D. M.; Furness, S. G. B.; Christopoulos, G.; Coudrat, T.; Danev, R.; Baumeister, W.; Miller, L. J.; Christopoulos, A.; Kobilka, B. K.; Wootten, D.; Skiniotis, G.; Sexton, P. M., Phase-plate cryo-EM structure of a class B GPCR-G-protein complex. *Nature* **2017**, *546* (7656), 118-123.
202. Fang, Z.; Marshall, C. B.; Nishikawa, T.; Gossert, A. D.; Jansen, J. M.; Jahnke, W.; Ikura, M., Inhibition of K-RAS4B by a Unique Mechanism of Action: Stabilizing Membrane-Dependent Occlusion of the Effector-Binding Site. *Cell Chem. Biol.* **2018**, *25* (11), 1327-1336.e4.
203. Liu, S.; Li, S.; Krezel, A. M.; Li, W., Stabilization and structure determination of integral membrane proteins by termini restraining. *Nat. Protoc.* **2022**.
204. Nadeau, V. G.; Rath, A.; Deber, C. M., Biochemical Definition of 'harsh' Vs. 'mild' Detergents For Membrane Protein Solubilization. *Biophys. J.* **2010**, *98* (3), 50a-51a.

205. Larsen, A. H.; Johansen, N. T.; Gajhede, M.; Arleth, L.; Midtgaard, S. R., Lipid-bound ApoE3 self-assemble into elliptical disc-shaped particles. *Biochim. Biophys. Acta* **2021**, 1863 (1), 183495.
206. Henrich, E.; Peetz, O.; Hein, C.; Laguerre, A.; Hoffmann, B.; Hoffmann, J.; Dotsch, V.; Berhard, F.; Morgner, N., Analyzing native membrane protein assembly in nanodiscs by combined non-covalent mass spectrometry and synthetic biology. *eLife* **2017**, 6, e20954
207. Oganessian, I.; Lento, C.; Tandon, A.; Wilson, D. J., Conformational Dynamics of α -Synuclein during the Interaction with Phospholipid Nanodiscs by Millisecond Hydrogen–Deuterium Exchange Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2021**, 32 (5), 1169-1179.
208. Laganowsky, A.; Reading, E.; Allison, T. M.; Ulmschneider, M. B.; Degiacomi, M. T.; Baldwin, A. J.; Robinson, C. V., Membrane proteins bind lipids selectively to modulate their structure and function. *Nature* **2014**, 510 (7503), 172-175.
209. Stace, C.; Ktistakis, N., Phosphatidic acid- and phosphatidylserine-binding proteins. *Biochim. Biophys. Acta* **2006**, 1761 (8), 913-926.
210. Long, S. B.; Tao, X.; Campbell, E. B.; MacKinnon, R., Atomic structure of a voltage-dependent K^+ channel in a lipid membrane-like environment. *Nature* **2007**, 450, 376-382.
211. Vestergaard, M.; Kraft, J. F.; Vosegaard, T.; Thøgersen, L.; Schiøtt, B., Bicelles and Other Membrane Mimics: Comparison of Structure, Properties, and Dynamics from MD Simulations. *J. Phys. Chem. B* **2015**, 119 (52), 15831-15843.
212. Kjølbye, L. R.; De Maria, L.; Wassenaar, T. A.; Abdizadeh, H.; Marrink, S. J.; Ferkinghoff-Borg, J.; Schiøtt, B., General Protocol for Constructing Molecular Models of Nanodiscs. *J Chem Inf Model* **2021**, 61 (6), 2869-2883.
213. Debnath, A.; Schäfer, L. V., Structure and Dynamics of Phospholipid Nanodiscs from All-Atom and Coarse-Grained Simulations. *J. Phys. Chem. B* **2015**, 119 (23), 6991-7002.
214. Rzepiela, A. J.; Schäfer, L. V.; Goga, N.; Risselada, H. J.; de Vries, A. H.; Marrink, S. J., Reconstruction of atomistic details from coarse grained structures. *J. Comp. Chem.* **2010**, 31, 1333-1343.
215. Schachter, I.; Allolio, C.; Khelashvili, G.; Harries, D., Confinement in Nanodiscs Anisotropically Modifies Lipid Bilayer Elastic Properties. *J. Phys. Chem. B* **2020**, 124 (33), 7166-7175.
216. Rangubpit, W.; Paritanon, P.; Pandey, R. B.; Sompornpisut, P., Thermally induced structural organization of nanodiscs by coarse-grained molecular dynamics simulations. *Biophys. Chem.* **2020**, 267, 106464.
217. Maingi, V.; Rothmund, P. W. K., Properties of DNA- and Protein-Scaffolded Lipid Nanodiscs. *ACS Nano* **2021**, 15 (1), 751-764.

218. López, C. A.; Swift, M. F.; Xu, X.-P.; Hanein, D.; Volkmann, N.; Gnanakaran, S., Biophysical Characterization of a Nanodisc with and without BAX: An Integrative Study Using Molecular Dynamics Simulations and Cryo-EM. *Structure* **2019**, *27* (6), 988-999.e4.
219. Iadanza, M. G.; Schiffrin, B.; White, P.; Watson, M. A.; Horne, J. E.; Higgins, A. J.; Calabrese, A. N.; Brockwell, D. J.; Tuma, R.; Kalli, A. C.; Radford, S. E.; Ranson, N. A., Distortion of the bilayer and dynamics of the BAM complex in lipid nanodiscs. *Commun. Biol.* **2020**, *3* (1).
220. Li, S.; Luo, Z.; Xu, Y.; Ren, H.; Deng, L.; Zhang, X.; Huang, F.; Yue, T., Interaction pathways between soft lipid nanodiscs and plasma membranes: A molecular modeling study. *Biochim. Biophys. Acta* **2017**, *1859* (10), 2096-2105.
221. Stepien, P.; Augustyn, B.; Poojari, C.; Galan, W.; Polit, A.; Vattulainen, I.; Wisniewska-Becker, A.; Rog, T., Complexity of seemingly simple lipid nanodiscs. *Biochim. Biophys. Acta* **2020**, *1862* (11), 183420.
222. Hagn, F.; Nasr, M. L.; Wagner, G., Assembly of phospholipid nanodiscs of controlled size for structural studies of membrane proteins by NMR. *Nat. Protoc.* **2018**, *13* (1), 79-98.
223. Denisov, I. G.; Sligari, S. G., Nanodiscs in membrane biochemistry and biophysics. *Chem. Rev.* **2017**, *117* (6), 4669-4713.
224. Benedetti, E.; Di Blasio, B.; Pavone, V.; Pedone, C.; Toniolo, C.; Crisma, M., Characterization at atomic resolution of peptide helical structures. *Biopolymers* **1992**, *32* (4), 453-456.
225. Franke, D.; Petoukhov, M. V.; Konarev, P. V.; Panjkovich, A.; Tuukkanen, A.; Mertens, H. D. T.; Kikhney, A. G.; Hajizadeh, N. R.; Franklin, J. M.; Jeffries, C. M.; Svergun, D. I., ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr* **2017**, *50* (4), 1212-1225.
226. Larsen, A. H.; Wang, Y.; Bottaro, S.; Grudinin, S.; Arleth, L.; Lindorff-Larsen, K., Combining molecular dynamics simulations with small-angle X-ray and neutron scattering data to study multi-domain proteins in solution. *PLoS Comp. Biol.* **2020**, *16* (4), e1007870.
227. Larsen, A. H.; Pedersen, M. C., Experimental noise in small-angle scattering can be assessed and corrected using the Bayesian Indirect Fourier Transformation. *J Appl Crystallogr* **2020**, *54* (5), 1281--1289.
228. Buchoux, S., FATSLiM: a fast and robust software to analyze MD simulations of membranes. *Bioinformatics* **2017**, *33* (1), 133-134.
229. Bond, P. J.; Holyoake, J.; Ivetac, A.; Khalid, S.; Sansom, M. S. P., Coarse-grained molecular dynamics simulations of membrane proteins and peptides. *J. Struct. Biol.* **2007**, *157*, 593-605.
230. Camp, T.; Mclean, M.; Kato, M.; Cheruzel, L.; Sligar, S., The hydrodynamic motion of Nanodiscs. *Chem. Phys. Lipids* **2019**, *220*, 28-35.

231. Gambin, Y.; Lopez-Esparza, R.; Reffay, M.; Siernecki, E.; Gov, N. S.; Genest, M.; Hodges, R. S.; Urbach, W., Lateral mobility of proteins in liquid membranes revisited. *Proc. Natl. Acad. Sci. USA* **2006**, *103* (7), 2098-2102.
232. Kucerka, N.; Liu, Y.; Chu, N.; Petrache, H. I.; Tristram-Nagle, S.; Nagle, J. F., Structure of fully hydrated fluid phase DMPC and DLPC lipid bilayers using X-ray scattering from oriented multilamellar arrays and from unilamellar vesicles. *Biophys. J.* **2005**, *88*, 2626.
233. Melcr, J.; Ferreira, T. M.; Jungwirth, P.; Ollila, O. H. S., Improved cation binding to lipid bilayers with negatively charged POPS by effective inclusion of electronic polarization. *J. Chem. Theor. Comput.* **2020**, *16* (1), 738-748.
234. Sligar, S. G.; Denisov, I. G., Nanodiscs: A toolkit for membrane protein science. *Protein Sci.* **2021**, *30* (2), 297-315.
235. Kalienkova, V.; Alvadia, C.; Clerico Mosina, V.; Paulino, C., Single-Particle Cryo-EM of Membrane Proteins in Lipid Nanodiscs. *Methods Mol. Biol.* **2020**, 245-273.
236. Dawaliby, R.; Trubbia, C.; Delporte, C.; Masureel, M.; Van Antwerpen, P.; Kobilka, B. K.; Govaerts, C., Allosteric regulation of G protein-coupled receptor activity by phospholipids. *Nature Chem. Biol.* **2016**, *12* (1), 35-39.
237. Manna, M.; Nieminen, T.; Vattulainen, I., Understanding the role of lipids in signaling through atomistic and multiscale simulations of cell membranes. *Ann. Rev. Biophys.* **2019**, *48*, 421-439.
238. Damian, M.; Louet, M.; Gomes, A. A. S.; M'Kadmi, C.; Denoyelle, S.; Cantel, S.; Mary, S.; Bisch, P. M.; Fehrentz, J.-A.; Catoire, L. J.; Floquet, N.; Banères, J.-L., Allosteric modulation of ghrelin receptor signaling by lipids. *Nat. Commun.* **2021**, *12* (1).
239. Marrink, S. J.; Corradi, V.; Souza, P. C. T.; Ingólfsson, H. I.; Tieleman, D. P.; Sansom, M. S. P., Computational modeling of realistic cell membranes. *Chem. Rev.* **2019**, *119*, 6184-6226.
240. Lee, S. C.; Khalid, S.; Pollock, N. L.; Knowles, T. J.; Edler, K.; Rothnie, A. J.; R.T.Thomas, O.; Dafforn, T. R., Encapsulated membrane proteins: A simplified system for molecular simulation. *Biochim. Biophys. Acta* **2016**, *1858* (10), 2549-2557.
241. Qi, Y.; Lee, J.; Klauda, J. B.; Im, W., CHARMM-GUI Nanodisc Builder for modeling and simulation of various nanodisc systems. *J Comput Chem* **2019**, *40* (7), 893-899.
242. Hsu, P. C.; Bruininks, B. M. H.; Jefferies, D.; de Souza, P. C. T.; Lee, J.; Patel, D. S.; Marrink, S. J.; Qi, Y. F.; Khalid, S.; Im, W., CHARMM-GUI Martini Maker for modeling and simulation of complex bacterial membranes with lipopolysaccharides. *J. Comp. Chem.* **2017**, *38* (27), 2354-2363.
243. Lee, J.; Patel, D. S.; Stahle, J.; Park, S. J.; Kern, N. R.; Kim, S.; Lee, J.; Cheng, X.; Valvano, M. A.; Holst, O.; Knirel, Y. A.; Qi, Y. F.; Jo, S.; Klauda, J. B.; Widmalm, G.; Im, W., CHARMM-GUI Membrane Builder for Complex

Biological Membrane Simulations with Glycolipids and Lipoglycans. *J. Chem. Theor. Comput.* **2019**, *15* (1), 775-786.

244. Price, D. J.; Brooks, C. L., A modified TIP3P water potential for simulation with Ewald summation. *J. Chem. Phys.* **2004**, *121* (20), 10096-10103.

245. Newport, T. Tools and resources for molecular simulations of integral membrane proteins. University of Oxford, 2017.

246. Song, W.; Corey, R. A.; Ansell, T. B.; Cassidy, C. K.; Horrell, M. R.; Duncan, A. L.; Stansfeld, P. J.; Sansom, M. S. P., PyLipID: A Python package for analysis of protein-lipid interactions from MD simulations. *J. Chem. Theory Comput.* **2021**.

247. Duncan, A. L.; Robinson, A. J.; Walker, J. E., Cardiolipin binds selectively but transiently to conserved lysine residues in the rotor of metazoan ATP synthases. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 8687-8692.

248. Rocchi, C.; Bizzarri, A. R.; Cannistraro, S., Water dynamical anomalies evidenced by molecular-dynamics simulations at the solvent-protein interface. *Phys. Rev. E* **1998**, *57* (3), 3315-3325.

249. Arnarez, C.; Marrink, S. J.; Periole, X., Identification of cardiolipin binding sites on cytochrome c oxidase at the entrance of proton channels. *Sci. Rep.* **2013**, *3*, 1263.

250. Arnarez, C.; Mazat, J.-P.; Elezgaray, J.; Marrink, S.-J.; Periole, X., Evidence for cardiolipin binding sites on the membrane-exposed surface of the cytochrome bc1. *J. Amer. Chem. Soc.* **2013**, *135*, 3112–3120.

251. McGoldrick, L. L.; Singh, A. K.; Saotome, K.; Yelshanskaya, M. V.; Twomey, E. C.; Grassucci, R. A.; Sobolevsky, A. I., Opening of the human epithelial calcium channel TRPV6. *Nature* **2018**, *553* (7687), 233-237.

252. Shahane, G.; Ding, W.; Palaokostas, M.; Orsi, M., Physical properties of model biological lipid bilayers: insights from all-atom molecular dynamics simulations. *J. Mol. Model.* **2019**, *25* (3).

253. Salas-Estrada, L. A.; Leioatts, N.; Romo, T. D.; Grossfield, A., Lipids alter rhodopsin function via ligand-like and solvent-like interactions. *Biophys. J.* **2018**, *114* (2), 355-367.

254. Corradi, V.; Mendez-Villuendas, E.; Ingólfsson, H. I.; Gu, R.-X.; Siuda, I.; Melo, M. N.; Moussatova, A.; DeGagné, L. J.; Sejdiu, B. I.; Singh, G.; Wassenaar, T. A.; Delgado Magnero, K.; Marrink, S. J.; Tieleman, D. P., Lipid–protein interactions are unique fingerprints for membrane proteins. *ACS Central Sci.* **2018**, *4* (6), 709-717.

255. Sejdiu, B. I.; Tieleman, D. P., ProLint: a web-based framework for the automated data analysis and visualization of lipid–protein interactions. *Nucleic Acids Res.* **2021**, *49* (W1), W544-W550.

256. Poveda, J. A.; Giudici, A. M.; Renart, M. L.; Millet, O.; Morales, A.; González-Ros, J. M.; Oakes, V.; Furini, S.; Domene, C., Modulation of the

potassium channel KcsA by anionic phospholipids: Role of arginines at the non-annular lipid binding sites. *Biochim. Biophys. Acta* **2019**, 1861 (10), 183029.

257. Monod, J.; Wyman, J.; Changeux, J.-P., On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **1965**, 12 (1), 88-118.

258. Fredriksson, R.; Lagerström, M. C.; Lundin, L.-G.; Schiöth, H. B., The G-protein-coupled receptors in the human genome form five main families. phylogenetic analysis, paralogon groups, and fingerprints. *Molec. Pharmacol.* **2003**, 63 (6), 1256-1272.

259. Gimpl, G., Interaction of G protein coupled receptors and cholesterol. *Chem. Phys. Lipids* **2016**, 199, 61-73.

260. Krishnan, A.; Almén, M. S.; Fredriksson, R.; Schiöth, H. B., The Origin of GPCRs: Identification of Mammalian like Rhodopsin, Adhesion, Glutamate and Frizzled GPCRs in Fungi. *PLoS ONE* **2012**, 7 (1), e29817.

261. Bjarnadóttir, T. K.; Gloriam, D. E.; Hellstrand, S. H.; Kristiansson, H.; Fredriksson, R.; Schiöth, H. B., Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. *Genomics* **2006**, 88 (3), 263-273.

262. Hu, G.-M.; Mai, T.-L.; Chen, C.-M., Visualizing the GPCR Network: Classification and Evolution. *Scientific Reports* **2017**, 7 (1).

263. Hilger, D.; Masureel, M.; Kobilka, B. K., Structure and dynamics of GPCR signaling complexes. *Nat. Struct. Mol. Biol.* **2018**, 25 (1), 4-12.

264. Miao, Y. L.; Nichols, S. E.; McCammon, J. A., Mapping of Allosteric Druggable Sites in Activation-Associated Conformers of the M2 Muscarinic Receptor. *Chem Biol Drug Des* **2014**, 83 (2), 237-246.

265. Ballesteros, J. A.; Weinstein, H., Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Meth. Neurosci.* **1995**, 25, 366-428.

266. Rovati, G. E.; Capra, V.; Neubig, R. R., The Highly Conserved DRY Motif of Class A G Protein-Coupled Receptors: Beyond the Ground State. *Mol. Pharmacol.* **2007**, 71 (4), 959-964.

267. Isberg, V.; De Graaf, C.; Bortolato, A.; Cherezov, V.; Katritch, V.; Marshall, F. H.; Mordalski, S.; Pin, J.-P.; Stevens, R. C.; Vriend, G.; Gloriam, D. E., Generic GPCR residue numbers – aligning topology maps while minding the gaps. *Trends Pharmacol. Sci.* **2015**, 36 (1), 22-31.

268. Schneider, E. H.; Schnell, D.; Strasser, A.; Dove, S.; Seifert, R., Impact of the DRY Motif and the Missing “Ionic Lock” on Constitutive Activity and G-Protein Coupling of the Human Histamine H4Receptor. *J. Pharmacol. Exp. Ther.* **2010**, 333 (2), 382-392.

269. Nygaard, R.; Zou, Y.; Dror, O.; Ron; Mildorf, J., Thomas; Arlow, H., Daniel; Manglik, A.; Pan, C., Albert; Liu, W., Corey; Fung, J., Juan; Bokoch, P., Michael; Thian, S., Foon; Kobilka, S., Tong; Shaw, E., David; Mueller, L.;

Prosser, S., R.; Kobilka, K., Brian, The Dynamic Process of β 2-Adrenergic Receptor Activation. *Cell* **2013**, *152* (3), 532-542.

270. Kong, Y.; Karplus, M., The Signaling Pathway of Rhodopsin. *Structure* **2007**, *15* (5), 611-623.

271. Hanson, M. A.; Roth, C. B.; Jo, E. J.; Griffith, M. T.; Scott, F. L.; Reinhart, G.; Desale, H.; Clemons, B.; Cahalan, S. M.; Schuerer, S. C.; Sanna, M. G.; Han, G. W.; Kuhn, P.; Rosen, H.; Stevens, R. C., Crystal structure of a lipid G Protein-Coupled Receptor. *Science* **2012**, *335* (6070), 851-855.

272. Lu, J.; Byrne, N.; Wang, J.; Bricogne, G.; Brown, F. K.; Chobanian, H. R.; Colletti, S. L.; Di Salvo, J.; Thomas-Fowlkes, B.; Guo, Y.; Hall, D. L.; Hadix, J.; Hastings, N. B.; Hermes, J. D.; Ho, T.; Howard, A. D.; Josien, H.; Kornienko, M.; Lumb, K. J.; Miller, M. W.; Patel, S. B.; Pio, B.; Plummer, C. W.; Sherborne, B. S.; Sheth, P.; Souza, S.; Tummala, S.; Vonnrhein, C.; Webb, M.; Allen, S. J.; Johnston, J. M.; Weinglass, A. B.; Sharma, S.; Soisson, S. M., Structural basis for the cooperative allosteric activation of the free fatty acid receptor GPR40. *Nature Struct. Molec. Biol.* **2017**, *24*, 570.

273. Srivastava, A.; Yano, J.; Hirozane, Y.; Kefala, G.; Gruswitz, F.; Snell, G.; Lane, W.; Ivetac, A.; Aertgeerts, K.; Nguyen, J.; Jennings, A.; Okada, K., High-resolution structure of the human GPR40 receptor bound to allosteric agonist TAK-875. *Nature* **2014**, *513* (7516), 124-127.

274. Michelarakis, N. Towards dynamic pharmacophore models through the use of coarse grained molecular dynamic simulations. University of Oxford, 2019.

275. Zhang, D.; Gao, Z.-G.; Zhang, K.; Kiselev, E.; Crane, S.; Wang, J.; Paoletta, S.; Yi, C.; Ma, L.; Zhang, W.; Han, G. W.; Liu, H.; Cherezov, V.; Katritch, V.; Jiang, H.; Stevens, R. C.; Jacobson, K. A.; Zhao, Q.; Wu, B., Two disparate ligand-binding sites in the human P2Y1 receptor. *Nature* **2015**, *520*, 317.

276. Genheden, S.; Essex, J. W.; Lee, A. G., G protein coupled receptor interactions with cholesterol deep in the membrane. *Biochim. Biophys. Acta* **2017**, *1859* (2), 268-281.

277. Yao, Z.; Kobilka, B., Using synthetic lipids to stabilize purified β 2 adrenoceptor in detergent micelles. *Anal. Biochem.* **2005**, *343* (2), 344-346.

278. Oddi, S.; Dainese, E.; Fezza, F.; Lanuti, M.; Barcaroli, D.; De Laurenzi, V.; Centonze, D.; Maccarrone, M., Functional characterization of putative cholesterol binding sequence (CRAC) in human type-1 cannabinoid receptor. *J. Neurochem.* **2011**, *116* (5), 858-865.

279. Potter, R. M.; Harikumar, K. G.; Wu, S. V.; Miller, L. J., Differential sensitivity of types 1 and 2 cholecystokinin receptors to membrane cholesterol. *J. Lipid Res.* **2012**, *53* (1), 137-148.

280. Muth, S.; Fries, A.; Gimpl, G., Cholesterol-induced conformational changes in the oxytocin receptor. *Biochem. J.* **2011**, *437* (3), 541-553.

281. Ansell, T. B.; Curran, L.; Horrell, M. R.; Pipatpolkai, T.; Letham, S. C.; Song, W.; Siebold, C.; Stansfeld, P. J.; Sansom, M. S. P.; Corey, R. A., Relative Affinities of Protein–Cholesterol Interactions from Equilibrium Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2021**, *17* (10), 6548-6558.
282. Paila, Y. D.; Tiwari, S.; Chattopadhyay, A., Are specific nonannular cholesterol binding sites present in G-protein coupled receptors? *Biochim. Biophys. Acta* **2009**, *1788* (2), 295-302.
283. Tripathy, R.; Mishra, D.; Konkimalla, V. B., A novel fuzzy C-means approach for uncovering cholesterol consensus motif from human G-protein coupled receptors (GPCR). *Karbala International Journal of Modern Science* **2015**, *1* (4), 212-224.
284. Taghon, G. J.; Rowe, J. B.; Kapolka, N. J.; Isom, D. G., Predictable cholesterol binding sites in GPCRs lack consensus motifs. *Structure* **2021**, *29* (5), 499-506.e3.
285. Isom, D. G.; Sridharan, V.; Dohlman, H. G., Regulation of Ras Paralog Thermostability by Networks of Buried Ionizable Groups. *Biochemistry* **2016**, *55* (3), 534-542.
286. Isom, D. G.; Dohlman, H. G., Buried ionizable networks are an ancient hallmark of G protein-coupled receptor activation. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112* (18), 5702-5707.
287. *GROMACS 2020.6 Source code*, 2020.6; Zenodo: 2021.
288. Gloriam, D.; Kooistra, A.; Munk, C.; Hauser, A., An online GPCR structure analysis platform. Research Square Platform LLC: 2021.
289. Munk, C.; Mutt, E.; Isberg, V.; Nikolajsen, L. F.; Bibbe, J. M.; Flock, T.; Hanson, M. A.; Stevens, R. C.; Deupi, X.; Gloriam, D. E., An online resource for GPCR structure determination and analysis. *Nat. Methods* **2019**, *16* (2), 151-162.
290. Bond, C. S.; Schuttelkopf, A. W., ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr. D* **2009**, *65*, 510-512.
291. Bradley, S. J.; Wiegman, C. H.; Iglesias, M. M.; Kong, K. C.; Butcher, A. J.; Plouffe, B.; Goupil, E.; Bourgognon, J.-M.; Macedo-Hatch, T.; Legouill, C.; Russell, K.; Laporte, S. A.; König, G. M.; Kostenis, E.; Bouvier, M.; Chung, K. F.; Amrani, Y.; Tobin, A. B., Mapping physiological G protein-coupled receptor signaling pathways reveals a role for receptor phosphorylation in airway contraction. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113* (16), 4524-4529.
292. Römpler, H.; Yu, H.-T.; Arnold, A.; Orth, A.; Schöneberg, T., Functional consequences of naturally occurring DRY motif variants in the mammalian chemoattractant receptor GPR33. *Genomics* **2006**, *87* (6), 724-732.
293. Milligan, G.; Stoddart, L. A.; Smith, N. J., Agonism and allosterism: the pharmacology of the free fatty acid receptors FFA2 and FFA3. *Br. J. Pharmacol.* **2009**, *158* (1), 146-153.

294. Kooistra, A. J.; Mordalski, S.; Pándy-Szekeres, G.; Esguerra, M.; Mamyrbekov, A.; Munk, C.; Keserű, G. M.; Gloriam, E., David, GPCRdb in 2021: integrating GPCR sequence, structure and function. *Nucleic Acids Res.* **2021**, *49* (D1), D335-D343.
295. Jaakola, V. P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y.; Lane, J. R.; Ijzerman, A. P.; Stevens, R. C., The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* **2008**, *322* (5905), 1211-7.
296. Kruse, A. C.; Ring, A. M.; Manglik, A.; Hu, J.; Hu, K.; Eitel, K.; Hubner, H.; Pardon, E.; Valant, C.; Sexton, P. M.; Christopoulos, A.; Felder, C. C.; Gmeiner, P.; Steyaert, J.; Weis, W. I.; Garcia, K. C.; Wess, J.; Kobilka, B. K., Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* **2013**, *504* (7478), 101-6.
297. Krumm, B. E.; White, J. F.; Shah, P.; Grisshammer, R., Structural prerequisites for G-protein activation by the neurotensin receptor. *Nat. Commun.* **2015**, *6* (1), 7895.
298. Zheng, Y.; Qin, L.; Zacarias, N. V. O.; De Vries, H.; Han, G. W.; Gustavsson, M.; Dabros, M.; Zhao, C.; Cherney, R. J.; Carter, P.; Stamos, D.; Abagyan, R.; Cherezov, V.; Stevens, R. C.; Ijzerman, A. P.; Heitman, L. H.; Tebben, A.; Kufareva, I.; Handel, T. M., Structure of CC chemokine receptor 2 with orthosteric and allosteric antagonists. *Nature* **2016**, *540* (7633), 458-461.
299. Tan, Q.; Zhu, Y.; Li, J.; Chen, Z.; Han, G. W.; Kufareva, I.; Li, T.; Ma, L.; Fenalti, G.; Li, J.; Zhang, W.; Xie, X.; Yang, H.; Jiang, H.; Cherezov, V.; Liu, H.; Stevens, R. C.; Zhao, Q.; Wu, B., Structure of the CCR5 Chemokine Receptor–HIV Entry Inhibitor Maraviroc Complex. *Science* **2013**, *341* (6152), 1387-1390.
300. Oswald, C.; Rappas, M.; Kean, J.; Doré, A. S.; Errey, J. C.; Bennett, K.; Deflorian, F.; Christopher, J. A.; Jazayeri, A.; Mason, J. S.; Congreve, M.; Cooke, R. M.; Marshall, F. H., Intracellular allosteric antagonism of the CCR9 receptor. *Nature* **2016**, *540* (7633), 462-465.
301. Zhang, D.; Gao, Z. G.; Zhang, K.; Kiselev, E.; Crane, S.; Wang, J.; Paoletta, S.; Yi, C.; Ma, L.; Zhang, W.; Han, G. W.; Liu, H.; Cherezov, V.; Katritch, V.; Jiang, H.; Stevens, R. C.; Jacobson, K. A.; Zhao, Q.; Wu, B., Two disparate ligand-binding sites in the human P2Y1 receptor. *Nature* **2015**, *520* (7547), 317-21.
302. Hollenstein, K.; Kean, J.; Bortolato, A.; Cheng, R. K. Y.; Doré, A. S.; Jazayeri, A.; Cooke, R. M.; Weir, M.; Marshall, F. H., Structure of class B GPCR corticotropin-releasing factor receptor 1. *Nature* **2013**, *499* (7459), 438-443.
303. Jazayeri, A.; Dore, A. S.; Lamb, D.; Krishnamurthy, H.; Southall, S. M.; Baig, A. H.; Bortolato, A.; Koglin, M.; Robertson, N. J.; Errey, J. C.; Andrews, S. P.; Teobald, I.; Brown, A. J.; Cooke, R. M.; Weir, M.; Marshall, F. H., Extra-helical binding site of a glucagon receptor antagonist. *Nature* **2016**, *533* (7602), 274-7.
304. Song, G.; Yang, D.; Wang, Y.; de Graaf, C.; Zhou, Q.; Jiang, S.; Liu, K.; Cai, X.; Dai, A.; Lin, G.; Liu, D.; Wu, F.; Wu, Y.; Zhao, S.; Ye, L.; Han, G. W.;

- Lau, J.; Wu, B.; Hanson, M. A.; Liu, Z. J.; Wang, M. W.; Stevens, R. C., Human GLP-1 receptor transmembrane domain structure in complex with allosteric modulators. *Nature* **2017**, *546* (7657), 312-315.
305. Sansuk, K.; Deupi, X.; Torrecillas, I. R.; Jongejan, A.; Nijmeijer, S.; Bakker, R. A.; Pardo, L.; Leurs, R., A Structural Insight into the Reorientation of Transmembrane Domains 3 and 5 during Family A G Protein-Coupled Receptor Activation. *Mol. Pharmacol.* **2011**, *79* (2), 262-269.
306. Davidson, F. F.; Loewen, P. C.; Khorana, H. G., Structure and function in rhodopsin: replacement by alanine of cysteine residues 110 and 187, components of a conserved disulfide bond in rhodopsin, affects the light-activated metarhodopsin II state. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91* (9), 4029-4033.
307. Conner, M.; Hawtin, S. R.; Simms, J.; Wootten, D.; Lawson, Z.; Conner, A. C.; Parslow, R. A.; Wheatley, M., Systematic Analysis of the Entire Second Extracellular Loop of the V1a Vasopressin Receptor. *J. Biol. Chem.* **2007**, *282* (24), 17405-17412.
308. Bairoch, A.; Apweiler, R.; Wu, C. H.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M. J.; Mazumder, R.; O'Donovan, C.; Redaschi, N.; Suzek, B., The universal protein resource (UniProt). *Nucleic Acids Res.* **2005**, *33*, D154 - D159.
309. Consortium, T. U., UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2020**, *49* (D1), D480-D489.
310. Wootten, D.; Simms, J.; Miller, L. J.; Christopoulos, A.; Sexton, P. M., Polar transmembrane interactions drive formation of ligand-specific and signal pathway-biased family B G protein-coupled receptor conformations. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110* (13), 5211-5216.
311. Weis, W. I.; Kobilka, B. K., The Molecular Basis of G Protein–Coupled Receptor Activation. *Annu. Rev. Biochem.* **2018**, *87* (1), 897-919.
312. Lyman, E.; Higgs, C.; Kim, B.; Lupyan, D.; Shelley, J. C.; Farid, R.; Voth, G. A., A role for a specific cholesterol interaction in stabilizing the Apo configuration of the human A(2A) adenosine receptor. *Structure* **2009**, *17* (12), 1660-8.
313. Milligan, G.; Ramsay, D.; Pascal, G.; Carrillo, J. J., GPCR dimerisation. *Life Sci.* **2003**, *74* (2), 181-188.
314. Prasanna, X.; Mohole, M.; Chattopadhyay, A.; Sengupta, D., Role of cholesterol-mediated effects in GPCR heterodimers. *Chem. Phys. Lipids* **2020**, *227*, 104852.
315. Vaidehi, N.; Grisshammer, R.; Tate, C. G., How Can Mutations Thermostabilize G-Protein-Coupled Receptors? *Trends Pharmacol. Sci.* **2016**, *37* (1), 37-46.
316. Nehmé, R.; Carpenter, B.; Singhal, A.; Strege, A.; Edwards, P. C.; White, C. F.; Du, H.; Grisshammer, R.; Tate, C. G., Mini-G proteins: Novel tools for

studying GPCRs in their active conformation. *PLOS ONE* **2017**, *12* (4), e0175642.

317. Wan, Q.; Okashah, N.; Inoue, A.; Nehmé, R.; Carpenter, B.; Tate, C. G.; Lambert, N. A., Mini G protein probes for active G protein-coupled receptors (GPCRs) in live cells. *J. Biol. Chem.* **2018**, *293* (19), 7466-7473.

318. Meshkin, H.; Zhu, F., Toward Convergence in Free Energy Calculations for Protein Conformational Changes: A Case Study on the Thin Gate of Mhp1 Transporter. *J. Chem. Theory Comput.* **2021**, *17* (10), 6583-6596.

319. Waldie, S.; Sebastiani, F.; Moulin, M.; Del Giudice, R.; Paracini, N.; Roosen-Runge, F.; Gerelli, Y.; Prevost, S.; Voss, J. C.; Darwish, T. A.; Yepuri, N.; Pichler, H.; Maric, S.; Forsyth, V. T.; Haertlein, M.; Cárdenas, M., ApoE and ApoE Nascent-Like HDL Particles at Model Cellular Membranes: Effect of Protein Isoform and Membrane Composition. *Front. Chem.* **2021**, *9*.

320. Nasr, M. L.; Wagner, G., Covalently circularized nanodiscs; challenges and applications. *Curr. Opin. Struct. Biol.* **2018**, *51*, 129-134.

321. Juarez, J. F. B.; Harper, A. J.; Judge, P. J.; Tonge, S. R.; Watts, A., From polymer chemistry to structural biology: The development of SMA and related amphipathic polymers for membrane protein extraction and solubilisation. *Chem. Phys. Lipids* **2019**, *221*, 167-175.

322. Frauenfeld, J.; Löving, R.; Armache, J.-P.; Sonnen, A. F.-P.; Guettou, F.; Moberg, P.; Zhu, L.; Jegerschöld, C.; Flayhan, A.; Briggs, J. A. G.; Garoff, H.; Löw, C.; Cheng, Y.; Nordlund, P., A saposin-lipoprotein nanoparticle system for membrane proteins. *Nat. Methods* **2016**, *13* (4), 345-351.

323. Zhao, Z.; Zhang, M.; Hogle, J. M.; Shih, W. M.; Wagner, G.; Nasr, M. L., DNA-Corralled Nanodiscs for the Structural and Functional Characterization of Membrane Proteins and Viral Entry. *J. Am. Chem. Soc.* **2018**, *140* (34), 10639-10643.

324. AlQuraishi, M., AlphaFold at CASP13. *Bioinformatics* **2019**, *35* (22), 4862-4865.

325. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D., Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583-589.

326. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; Millán, C.; Park, H.; Adams, C.; Glassman, C. R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues, A. V.; van Dijk, A. A.; Ebrecht, A. C.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D., Accurate

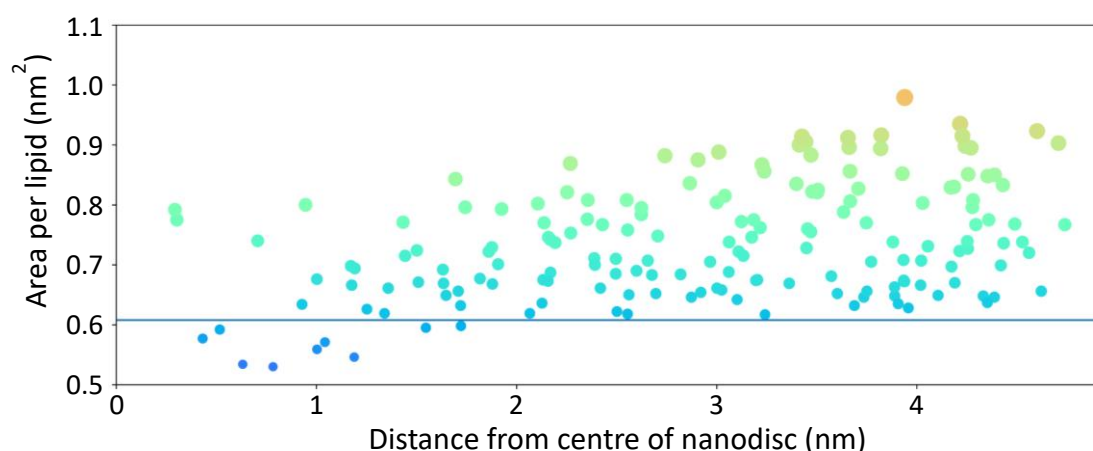
prediction of protein structures and interactions using a 3-track network. *bioRxiv* **2021**, 2021.06.14.448402.

327. Vischer, H. F.; Castro, M.; Pin, J.-P., G Protein–Coupled Receptor Multimers: A Question Still Open Despite the Use of Novel Approaches. *Mol. Pharmacol.* **2015**, *88* (3), 561-571.

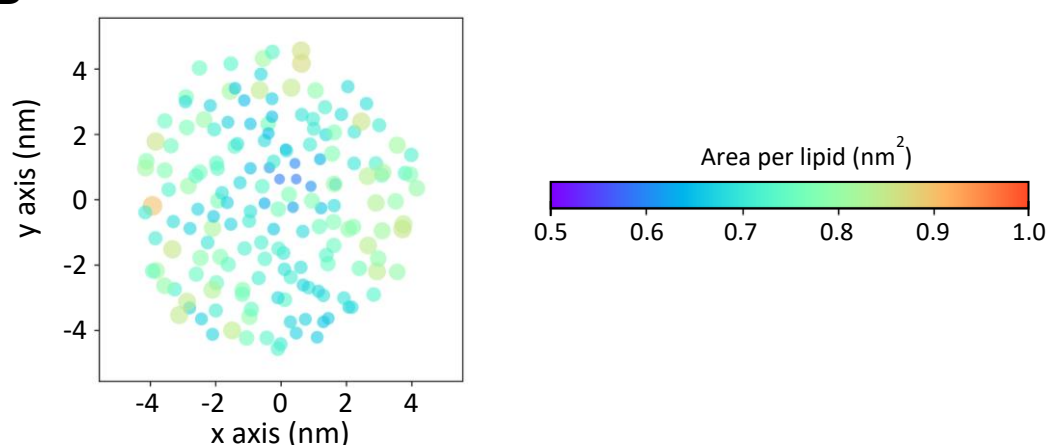
328. Haack, K. K. V.; Mccarty, N. A., Functional Consequences of GPCR Heterodimerization: GPCRs as Allosteric Modulators. *Pharmaceuticals* **2011**, *4* (3), 509-523.

Appendix

A



B



Appendix 1: Variation in APL with distance from the centre of a nanodisc in a system created using the Martini ENM

APLs were calculated from the final frame of a 200ns atomistic simulation of a system which had been generated in CG with the Martini ENM applied.

A: The radial distribution of APL, which increases with distance from the nanodisc centre. The experimentally derived APL value for DLPC in a bilayer is 0.61 nm^2 ²³² (indicated by the blue line). In contrast to the systems generated in CG without an ENM which were shown in **Figure 3.11**, a greater proportion of lipids at all distances from the centre of the nanodisc displayed higher than expected APLs; this is particularly the case towards the centre of the nanodisc.

B: The same system when viewed from above the plane of the nanodisc, illustrating the increase in APL with distance from the centre. The proportion of lipids with an APL exceeding that reported for DLPC in a bilayer is greater in these systems generated in CG using the Martini ENM than those without an ENM which were shown in **Figure 3.11**.