Investigating the Role of Lipids and Small-Molecule Ligands in Hedgehog Signalling using Molecular Dynamics Simulations

Tillson Bertie Geoffrey Ansell

Wadham College
Department of Biochemistry
University of Oxford



A thesis submitted for the degree of
Doctor of Philosophy

Michaelmas 2022

Acknowledgements

I will be indebted to the education I have received over the past eight years at Oxford for the rest of my life. For one moment aged 17, hard work, privilege and luck aligned, granting me access to the education of my dreams. This is something I think about a lot, and I can only hope that in my time here I have been of credit to those gracious enough to let me into their worlds.

My journey to Oxford was propelled by the friendships of Tuli Saha and Amy Parkes; fiercely bold (and sometimes terrifying) in all that they do. Their healthy competition pulled me through my teenage years (often literally). Luckily, I can still hear them coming a mile off. Also, to Kate Nichols, my A level chemistry teacher, for lending me her personal copies of Chemistry World and for believing in me. State schooled and proud.

For guidance through my first four years at Oxford, I must thank Prof. Ben Berks who remains the smartest person I've ever met. I hope one day to ask compelling and uncompromising questions half as well as Ben does.

My extended gratitude to Prof. Mark Sansom and Prof. Christian Siebold for their supreme supervision throughout my PhD. Their humour, expert knowledge and relentless belief (just about) prevailed the length of my time in Oxford (mean residence time ~3.2 years). A true pleasure. Thanks also to SBCB members Patrick Simcock, Laura John, Robin Corey, Andreas Larsen, Wanling Song, Keith Cassidy and Anna Duncan for their company and conversation on this journey. Special acknowledgment goes to my long-suffering desk buddies Rocco Meli and Linda Phan for putting up with my eccentricities.

To our experimental collaborators upon a large hill yonder - I must thank Amalie Rudolf, Claire Coupland, Rachel Woolley, Lucrezia Viti, Megan Healey and all those within the STRUBI community for their expertise. Thanks also to Prof. Raj Rohatgi and Maia Kinnebrew at Stanford for their long-term collaboration on Hedgehog biochemistry.

Appreciation to the hum of human and non-human communities at Bartlemas Close Allotment, OxGrow and Oxford Mutual Aid, who have taught me so much about convivial existence, reciprocity and collective regeneration.

I am grateful for the many beautiful, unbounded friendships nurtured outside the lab. Firstly, to the Heathens of Cumbria/Uno Fandom Massive (Emma Flint, Claire H. Evans, Xander Sykes and Rian Harris) for collective weirdness, divergent journeys and never being in the same place twice. Also, unending support. To Tai Lee for shared tea, politics and eye-rolls. To Jack Harberd for knowing ourselves before we knew ourselves. Lastly, to the sweetest of Clementines (Lucrezia Zarcone, Emily Graham, Emile Roberts, Violaine André, Arnau Bertran and Yang Pei) for teaching me how to be myself again, for belly aching laughs and for the benign calmness of mid-20s life. Also, for pondering over life's true mysteries like '*where does potato salad come from if you can't eat raw potatoes?*'

I must thank all those in the walking, triathlon and swimming teams past and present, as well as the lifeguards of Iffley Road. The pool has been a second home for me at Oxford, ever forgiving and ever reviving.

To my wild little sister Jess for never being afraid to knock my ego down a peg or two, and for being the best adventure otter a sibling could ask for.

Finally, to Mum and Dad, for all of the everything in between. Love always. 11 stars. x

Publications

Published (* denotes joint first authorship):

*Rudolf, A.F., Kinnebrew, M., Kowatsch, C., **Ansell, T.B**., El Omari, K., Bishop, B., Pardon, E., Schwab, R.A., Malinauskas, T., Qian, M., Duman, R., Covey, D.F., Steyaert, J., Wagner, A., Sansom, M.S.P., Rohatgi, R., Siebold, C. (2019). The morphogen Sonic hedgehog inhibits its receptor Patched by a pincer grasp mechanism. *Nat. Chem. Biology,* 15, 975-982

Ansell, T.B., Song, W., Sansom, M.S.P. (2020). The Glycosphingolipid GM3 Modulates Conformational Dynamics of the Glucagon Receptor. *Biophys. J.,* 119, 300-313

Corey, R.A., Song, W., Duncan, A.L., **Ansell, T.B**., Sansom, M.S.P., Stansfeld, P.J. (2021). Identification and assessment of cardiolipin interactions with *E. coli* inner membrane proteins. *Sci. Advances,* 7, eabh2217

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. (2021). Relative Affinities of Protein-Cholesterol Interactions from Equilibrium Molecular Dynamics Simulations. *J. Chem. Theory Comput.*, 17, 10, 6548-6558

*Coupland, C.E., Andrei, S.A., **Ansell, T.B**., Carrique, L., Kumar, P., Sefer, L., Schwab, R.A., Byrne, E.F.X., Pardon, E., Steyaert, J., Magee, A. I., Lanyon-Hogg, T., Sansom, M.S.P., Tate, E.W., Siebold, C. (2021). Structure, mechanism, and inhibition of Hedgehog acyltransferase. *Mol. Cell*., 81, 24, 5025-5038

Song, W., Corey R.A., **Ansell, T.B**., Cassidy, C.K., Horrell, M.R., Duncan, A.L., Stansfeld, P.J., Sansom, M.S.P. (2022). PyLipID: A Python Package for Analysis of Protein-Lipid Interactions from Molecular Dynamics Simulations. *J. Chem. Theory Comput.*, 18, 2, 1188-1201

Kinnebrew, M., Woolley, R.E., **Ansell, T.B.**, Byrne, E.F.X., Frigui, S., Luchetti, G., Sircar, R., Nachtergaele, S., Mydock-McGrane, L., Krishnan, K., Newstead, S., Sansom, M.S.P., Covey, D.F., Siebold, C. and Rohatgi, R. (2022). Patched1 regulates Smoothened by controlling sterol binding to its extracellular cysteine-rich domain. *Sci. Advances*, 8, eabm5563

McGlone, E.R., **Ansell, T.B.**, Dunsterville, C., Song, W., Carling, D., Tomas, A., Bloom, S.R., Sansom, M.S.P., Tan, T., Jones, B. (2022). Hepatocyte cholesterol content modulates glucagon receptor signalling. *Molecular Metabolism*, 63, 101530

In preparation:

Ansell, T.B., Song, W., Coupland, C.E., Carrique, L., Corey, R.A., Duncan, A.L., Cassidy, C.K., Geurtz, M.M.G., Rassmussen, T., Ward, A.B., Siebold, C., Stansfeld, P.J., Sansom, M.S.P. (2022). LipIDens: Simulation Assisted Interpretation of Lipid Densities in Cryo-EM Structures of Membrane Proteins. *Nat. Comms., under review*

Ansell, T.B., Corey, R.A., Viti, L.V., Kinnebrew, M., Rohatgi, R., Siebold, C., Sansom, M.S.P. (2022). The Energetics and Ion Coupling of Cholesterol Transport through Patched1. *Sci. Advances, Manuscript in preparation*

Ansell, T.B., Healey, M., Coupland, C., Sansom, M.S.P., Siebold, C. (2022). Investigation of Structural and Dynamic Divergence Across the MBOAT family. *Manuscript in preparation*

*Coupland, C., **Ansell, T.B.**, Sansom, M.S.P., Siebold, C. (2022). Rocking the MBOAT: Structural Insights into the Membrane bound *O*-acyltransferase family. *Curr. Opin. Struc. Biol., under review*

Conflitti, P., Lyman, E., Sansom, M.S.P., Gutiérrez-de-Terán, Carloni, P., **Ansell, T.B,** H., Yuan, S., Barth, P., Robinson, A.S., Hildebrand, P., Eddy, M.T., Prosser, S., Grzesiek, S., Gloriam, D., Tate, C.G., Robinson, C.V., Kobilka, B., Limongelli, V. (2022). Future directions in GPCR drug discovery enlightened by biophysical and structural studies. *Nat. Rev. Drug. Discov., Manuscript in preparation*

Ashraf, K.U., **Ansell, T.B.**, Stansfeld, P.J., Petrou, V.I. (2022). Structure and dynamics of undecaprenyl-phosphate transferase protein ArnC. *Nat. Struc. Mol. Biol., Manuscript in preparation*

Abstract

The Hedgehog (HH) signalling pathway orchestrates cellular patterning during embryonic development and adult stem cell programming. Sonic hedgehog (SHH) is the principle morphogenic ligand responsible for paracrine induction of HH signalling. Pathway dysregulation results in multiple cancer pathologies and ineffective tissue partitioning, leading to severe developmental diseases. SHH is directly conjugated to two lipids; a N-terminal palmitate and a C-terminal cholesterol moiety. Unique bilipidation of SHH occurs within the endoplasmic reticulum via autocatalytic cholesterylation and Hedgehog acyl-transferase (HHAT) catalysed palmitoylation steps, after which SHH is released from cells. SHH binds to and inhibits the tumour suppressor protein Patched1 (PTCH1) to relieve PTCH1 mediated inhibition of the Class F G-protein coupled receptor (GPCR) Smoothened (SMO).

Small-molecule ligands, notably cholesterol and palmitate, are inextricably linked to HH signalling, inducing a perplexing range of signalling outputs. This thesis outlines a conveyor-belt tour of HH membrane proteins for SHH processing, engagement, and pathway initiation. Molecular dynamics (MD) simulations of PTCH1, SMO and HHAT are employed to rationalise the, often counter-inductive, roles of small-molecules in HH signalling. Simulations are coupled with structural and biochemical data, demonstrating how the SHH-cholesterol contributes to PTCH1 inhibition. Simulations and free energy calculations are used to provide energetic contextualisation of PTCH1 cholesterol transport models and predict ion-coupling stoichiometries. Additionally, we find simulation and biochemical evidence for cooperativity between cholesterol binding sites on SMO. Finally, simulations are used to assist HHAT structural interpretation, elucidating roles for residues in solvent gating, product exit and priming HHAT catalysis. Bioinformatics and simulations were also employed to assess structural divergence between HHAT and related homologues. Biological insights were assisted by concurrent development of computational methodologies for characterisation of protein-lipid interactions, their affinities and integration with experimental data. Thus, this thesis spotlights the diverse roles lipids and small-molecules play in HH pathway regulation – from activators to allosteric modulators and inhibitors, illuminating new avenues for therapeutic intervention.

Contents

Abbreviations ..... **ix**

List of Figures ..... **x**

List of Tables ..... **xii**

1. Introduction **1**

1.1 Hedgehog Signalling and the Biology of the Ciliary Membrane **2**
1.1.1 Hedgehog signalling and disease 2
1.1.2 The primary cilium 4
1.1.3 Lipids in Hedgehog signalling 6

1.2 Protein-lipid Interactions as Viewed by MD Simulations **8**
1.2.1 Protein-lipid interactions 8
1.2.2 Simulations of Hedgehog pathway proteins 10

1.3 Research Novelty and Objectives **16**

2. Methods for Membrane Protein Simulations **18**

2.1 Molecular Dynamics Simulations of Membrane Proteins **20**
2.1.1 Forcefields and parameterisation 20
2.1.2 Running MD simulations of membrane proteins 24

2.2 Methods for Determining Protein-Lipid Interactions **24**
2.2.1 Identification of interaction events 24
2.2.2 Selecting cut-off values 25
2.2.3 Quantification of lipid binding kinetics 27
2.2.4 Binding site identification 30
2.2.5 Obtaining representative lipid binding poses 34
2.2.6 Assessing binding site quality 37
2.2.7 Integrating simulation and structural data 40

2.3 Affinities of Protein-Lipid Interactions **45**
2.3.1 Biased sampling methods 45
2.3.2 Unbiased sampling methods 48
2.3.3 Comparing methods for obtaining protein-lipid affinities 51

3. Patched **56**

3.1 Introduction **57**

3.2 Methods **60**
3.2.1 Coarse-grained simulations 61
3.2.2 Constructs used in atomistic simulations 62
3.2.3 Atomistic simulations 63
3.2.4 Application of a membrane potential 63
3.2.5 *In silico* mutation screen 64
3.2.6 Atomistic simulations of Dispatched1 64
3.2.7 Potential of mean force calculations 64
3.2.8 ABFE Calculations 67
3.2.9 FEP Calculations 68

3.3 Results and Discussion 69
3.3.1 SHH-cholesterol binds the PTCH1 ECD 69
3.3.2 Collapse of tunnels throughout PTCH1 by SHH-cholesterol 74
3.3.3 Probing the energetics of cholesterol transport pathways 76
3.3.4 Investigating the behaviour of ions in PTCH1 81
3.3.5 The hydrophobic cap 84
3.3.6 Insights from Dispatched1 86
3.4 Conclusion 89
3.4.1 The SHH-cholesterol inhibits PTCH1 89
3.4.2 Models for PTCH1 function 90
4. SMO 94
4.1 Introduction 95
4.2 Methods 98
4.2.1 Constructs used in atomistic simulations 98
4.2.2 Atomistic simulations 99
4.2.3 Allosteric pathways 99
4.2.4 Targeted MD simulations 99
4.3 Results and Discussion 100
4.3.1 Investigating SMO conformational stability 100
4.3.2 Allostery between cholesterol binding sites 102
4.3.3 Inspecting the proposed ‘mid’ site cholesterol 105
4.3.4 ECL3 displacement during activation 107
4.3.5 Conformational clustering 109
4.4 Conclusion 113
4.4.1 PTCH1 inhibits SMO by regulating cholesterol binding to the CRD 113
4.4.2 SMO dynamics upon activation 114
5. HHAT 116
5.1 Introduction 117
5.2 Methods 121
5.2.1 Coarse-grained simulations of HHAT 121
5.2.2 Atomistic simulations of HHAT 122
5.2.3 Steered MD simulations of HHAT 124
5.2.4 Simulations of other MBOATs 124
5.3 Results and Discussion 125
5.3.1 The structure of Hedgehog acyl-transferase promotes membrane deformation 125
5.3.2 An extended water network forms throughout HHAT 129
5.3.3 HHAT gating by W335 132
5.3.4 Reaction centre dynamics 134
5.3.5 Assessing lipid-like densities around HHAT 135
5.3.6 A luminal exit gate for SHH 139
5.3.7 HHAT and other MBOATs 142
5.4 Conclusion 144
6. Anti-Ending 147

6.1 Context is key 149
6.2 Acknowledge the complexity of the results – an eye to the future 149
7. References 152
8. Appendices 168

Abbreviations

CDL	Cardiolipin
CG	Coarse-grained
CHS	Cholesterol-hemisuccinate
CRD	Cysteine-rich domain
cryo-EM	Cryogenic electron microscopy
CTD	C-terminal domain
DISP1	Dispatched1
DOPC	1,2-dioleoyl-sn-glycero-3-phosphocholine
DOPE	1,2-dioleoyl-sn-glycero-3-phosphoethanolamine
DOPS	1,2-dioleoyl-sn-glycero-3-phosphoserine
ECD	Extracellular domain
GM3	N-stearoyl-D-erythro-monosialodihexosylganglioside
GOAT	Ghrelin O-acyl-transferase
GPCR	G-protein coupled receptor
HH	Hedgehog
HHAT	Hedgehog acyl-transferase
MBOAT	Membrane bound O-acyltransferase
MD	Molecular dynamics
NPC1	Niemann-Pick disease type-C1
Palm-CoA	Palmitoyl-coenzyme A
PBC	Periodic boundary conditions
PDB	Protein data bank
PIP_2	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylinositol-4,5-bisphosphate
PMF	Potential of mean force
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPE	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
POPG	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol
POPS	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoserine
PORCN	Porcupine
PTCH1	Patched1
RND	Resistance-Nodulation-Division
SHH	Sonic Hedgehog
SMO	Smoothened
TMD	Transmembrane domain
WHAM	Weighted histogram analysis method

List of Figures

Figure 1.1: The mammalian Hedgehog (HH) signalling pathway. 3
Figure 1.2: Architecture of the primary cilium. 5
Figure 1.3: Steroids in HH signalling. 7
Figure 1.4: Computational studies of HH pathway proteins and homologues. 11

Figure 2.1: The dynamic nature of protein-lipid contacts. 25
Figure 2.2: Testing dual cut-off schemes. 26
Figure 2.3: Calculating interaction residence times. 29
Figure 2.4: Lipid interaction kinetics. 30
Figure 2.5: Determining lipid binding sites. 32
Figure 2.6: Exploring binding site metrics. 33
Figure 2.7: Binding site analyses and lipid pose detection. 35
Figure 2.8: High throughput application of PyLipID. 36
Figure 2.9: Screening PyLipID binding site outputs. 38
Figure 2.10: Diverse applications of PyLipID. 39
Figure 2.11: LipIDens: a pipeline for simulation assisted interpretation of lipid-like densities. 40
Figure 2.12: Comparison of lipid poses from simulations with cryo-EM densities. 42
Figure 2.13: Interpreting lipid-like densities surrounding OTOP1. 44
Figure 2.14: Cholesterol binding sites on P-gp. 52
Figure 2.15: Comparison of apparent free energies for cholesterol from biased and equilibrium simulations. 53
Figure 2.16: The binding saturation method applied to multiple sites on 5-HT_{1A}. 55
Figure 2.17: Convergence of site affinities from binding saturation curves. 55

Figure 3.1: Patched1 (PTCH1) function. 58
Figure 3.2: Structural and biochemical studies of PTCH1. 70
Figure 3.3: The SHH-cholesterol moiety binds and inhibits PTCH1. 72
Figure 3.4: ‘Free cholesterol’ stabilises the PTCH1 SBD. 73
Figure 3.5: Preferential stabilisation of the ‘free cholesterol’ orientation. 74
Figure 3.6: Analysis of tunnels through PTCH1. 75
Figure 3.7: The indirect pathway. 77
Figure 3.8: The direct pathway. 79
Figure 3.9: Cholesterol export is associated with a gain in free energy. 81
Figure 3.10: Cation binding within the PTCH1 TMD. 83
Figure 3.11: Site 1 cation free energy perturbation. 84
Figure 3.12: The hydrophobic cap. 85
Figure 3.13: Water within Dispatched1 (DISP1). 87
Figure 3.14: Expansion of the intracellular TMD helices. 88
Figure 3.15: Resistance-nodulation-division (RND) family structures. 89
Figure 3.16: Models for PTCH1 function. 92
Figure 3.17: Free energy and coupling stoichiometry of Na^+/K^+. 93

Figure 4.1: Smoothened (SMO) structures. 96
Figure 4.2: SMO stability in simulations. 101
Figure 4.3: Investigating SMO allostery. 103

Figure 4.4: Information flow through SMO. 104
Figure 4.5: Debunking the ‘mid’ cholesterol site. 106
Figure 4.6: Outward tilt of ECL3. 108
Figure 4.7: Targeted MD simulations of hSMO activation and inactivation. 108
Figure 4.8: Selecting a clustering threshold. 110
Figure 4.9: Clustering hSMO conformations during activation. 111
Figure 4.10: Highly dynamic CRD behaviour between clusters. 112
Figure 4.11: SMO activation: cholesterol binding and conformational transitions. . 114

Figure 5.1: Sonic Hedgehog (SHH) morphogen processing. 118
Figure 5.2: Structures of MBOAT family members. 119
Figure 5.3: Structure of Hedgehog acyl-transferase (HHAT). 126
Figure 5.4: Re-arrangement of the termini between MBOAT subtypes. 128
Figure 5.5: HHAT induces membrane deformation. 129
Figure 5.6: Formation of water filled cavity within HHAT. 130
Figure 5.7: Water within HHAT. 131
Figure 5.8: Water flux through HHAT. 131
Figure 5.9: Structure of HHAT bound to inhibitor IMP-1575. 132
Figure 5.10: Gating of Palm-CoA binding pocket by W335. 133
Figure 5.11: Reaction centre dynamics. 135
Figure 5.12: Probing a proposed cholesterol binding site. 136
Figure 5.13: Ambiguous lipid-like density within HHAT. 137
Figure 5.14: Probing the ‘archway’ Palm-CoA. 138
Figure 5.15: Insertion of lipids between HHAT transmembrane helices in CG simulations. 139
Figure 5.16: A luminal exit gate for palmitoylated-SHH. 141
Figure 5.17: Comparison of HHAT with other MBOATs. 143
Figure 5.18: A potential catalytic mechanism for HHAT. 145

Figure 6.1: Lipids as ligands, modulators and inhibitors in Hedgehog signalling. .. 148

List of Tables

Table T3.1: Summary of equilibrium PTCH1/DISP1 simulations. 60
Table T3.2: Predicted pKa of PTCH1 residues. 63
Table T3.3: Summary of PMF calculations. 67
Table T3.4: Summary of ABFE calculations. 68

1. Introduction

"'My house was once an acorn."
Roger Deakin

1. Introduction 1
1.1 Hedgehog Signalling and the Biology of the Ciliary Membrane 2
1.1.1 Hedgehog signalling and disease 2
1.1.2 The primary cilium 4
1.1.3 Lipids in Hedgehog signalling 6
1.2 Protein-lipid Interactions as Viewed by MD Simulations 8
1.2.1 Protein-lipid interactions 8
1.2.2 Simulations of Hedgehog pathway proteins 10
1.3 Research Novelty and Objectives 16

1.1 Hedgehog Signalling and the Biology of the Ciliary Membrane

1.1.1 Hedgehog signalling and disease

The Hedgehog (HH) signalling pathway is conserved across metazoans and is implicated in cellular patterning during embryogenesis and tissue homeostasis in adults[1,2]. HH signalling is orchestrated by the morphogenic HH ligand, discovered in *Drosophila melanogaster,* whereby striped embryonic cellular patterns instead form disordered spikes upon HH gene deletion[3]. This thesis focusses on the mammalian HH pathway which differs in complexity compared to that of invertebrates. In mammals there are three HH morphogen paralogs; Sonic hedgehog (SHH) which is widely expressed in the notochord, neuronal floor plate and limb buds, Indian hedgehog (IHH) which localises to skeletal structures and Desert hedgehog (DHH), expressed in germ-cells and neural Schwan cells[4].

The HH pathway can be decomposed into two stages corresponding to HH processing and release from cells, and HH engagement and initiation of signalling (Fig. 1.1). In HH-releasing cells the SHH morphogen precursor undergoes autocatalytic C-terminal cholesterylation, releasing the C-terminal protein segment for degradation[5–7]. SHH is further modified by N-terminal palmitoylation catalysed by the ER resident enzyme Hedgehog acyl-transferase (HHAT)[8–10]. Post trafficking to the plasma membrane, SHH is released into the surrounding media via Dispatched1 (DISP1) and SCUBE2[11–13]. The mechanism of SHH transport to target cells is still unclear and could involve, protein multimer formation, vesicular association, cytonemes or SHH-SCUBE2 complexes, and may also depend on the distance to the target cell[14].

In the absence of SHH, Patched1 (PTCH1) inhibits the G-protein coupled receptor (GPCR) Smoothened (SMO) activation and downstream signalling[15]. Relief of SMO suppression occurs when SHH binds to and inhibits PTCH1[15–17]. A second PTCH protein, PTCH2, is also present in mammals however *in vivo* signalling is primarily under control of PTCH1[18]. SHH engages with the cellular surface via a sequence of molecular handover events on HH-receiving cells. The CDON/BOC co-receptors engage SHH and promote unloading from SCUBE2. SHH is relayed from CDON/BOC

to PTCH1 via the GPI-anchored protein GAS1[19,20]. Transfer of SHH via this complex sequence of intermediary scaffolds ensures the correct SHH interface engages PTCH1.

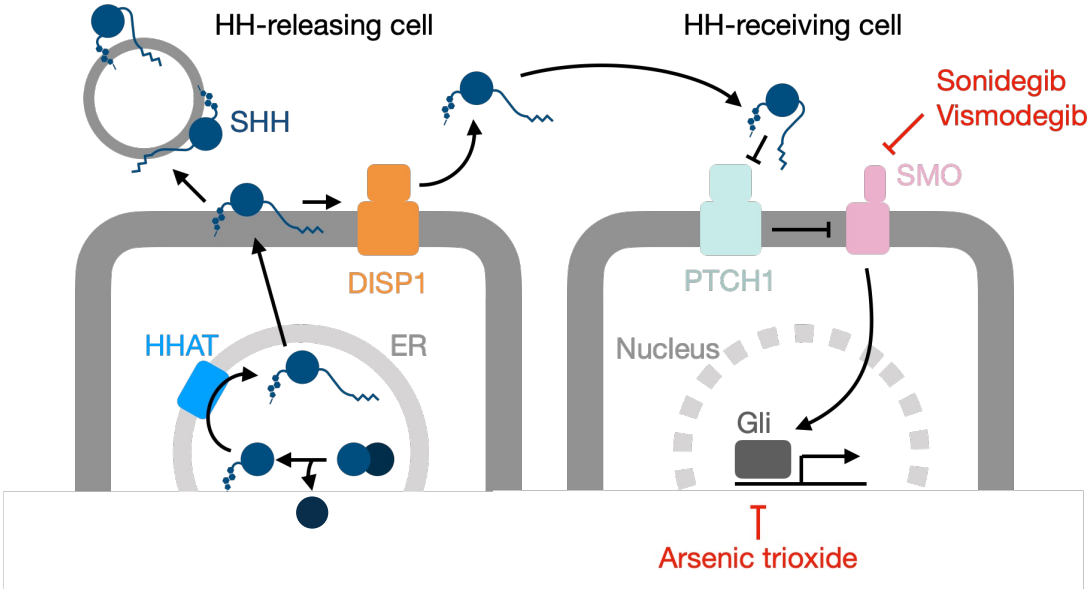


Figure 1.1: The mammalian Hedgehog (HH) signalling pathway.

Sonic Hedgehog (SHH, dark blue) is processed via autocatalytic cleavage of the C-terminal protein lobe and C-terminal cholesterylation within the endoplasmic reticulum (ER). Hedgehog acyl-transferase (HHAT, blue) palmitoylates the SHH N-terminus for release from cells via Dispatched1 (DISP1, orange). In the absence of SHH, Patched1 (PTCH1, light blue) inhibits Smoothened (SMO, pink). SHH binds to and inhibits PTCH1 to initiate SMO mediated HH signalling. FDA approved pathway inhibitors are labelled in red.

While the exact mechanism of PTCH1 action is still uncertain, it is thought to involve repartitioning of cholesterol, the sterol which is both necessary and sufficient for SMO activation[21,22]. SMO activation leads to inhibition of PKA activity, suggested to result from direct interaction of the PKA-C domain with the SMO C-terminal tail[23,24]. PKA inhibition, relieves phosphorylation of zinc-finger GLI2/3 transcription factors which enter the nucleus and activate HH regulated genes such as GLI1[25]. PKA activity may also be under control of G-protein coupled SMO signalling and/or other GPCRs (reviewed in [2]).

Dysregulation of HH signalling is linked to development of aggressive cancers including basal cell carcinoma, medulloblastoma, rhabdomyosarcoma, lung, colon, pancreatic and liver cancer[26]. Formation of multiple solid tumours, such as present in

Gorlins syndrome, is also associated with HH pathway mutations[27,28]. Embryos with HH mutations often have ineffective tissue partitioning resulting in holoprosencephaly and developmental malformations[29]. Two FDA approved drugs are available for treatment of HH associated diseases (Vismodegib and Sonidegib) which both target SMO[30,31]. However, development of drug resistance in HH associated cancers is common, resulting from mutations with the SMO transmembrane domain (TMD) which reduce drug binding[32]. In addition, arsenic trioxide targets nuclear HH events.

1.1.2 The primary cilium

HH signalling is further complicated by state dependant partitioning of proteins within the primary cilium[33]. Primary cilia are hair-like protrusions of the plasma membrane, extending 1-10 μm into the extracellular space[34]. Unlike motile cilia (responsible for movement of mucus across cell surfaces), approximately one primary cilium is present on the surface of quiescent cells during the G phase[35]. Primary cilia have a unique axoneme structure compared to other cilia (Fig. 1.2A). In motile cilia the axoneme is composed of 9 microtubule doublets surrounding 2 microtubules (so called '9+2' architecture)[36]. Each doublet is composed of a complete protofilament ring (A tubule) and a semi-circular adjunction (B tubule). By contrast cryogenic electron tomography (cryo-ET) structures of the primary cilium reveal 9 microtubule triplets at the base of the cilium which convert to 9 microtubule doublets ('9+0' architecture) above the transition zone (the network of protein scaffolds which anchor the axoneme to the ciliary pocket membrane)[37,38]. The '9+0' structure quickly converts to a '7+2' arrangement after which microtubules sequentially terminate, resulting in a conical axoneme with ~3 microtubules at the ciliary tip (Fig. 1.2B)[37,39].

In SHH deficient cells, PTCH1 localises to the ciliary base and SMO resides within the plasma membrane. SHH binding to PTCH1 restricts its diffusion, triggering its internalisation and exclusion from the cilium and permitting SMO entry for activation of downstream proteins (Fig. 1.2C)[15,20,40]. Primary cilia are proposed to constitute specialised organelles which function as 'signalling hubs' for a range of pathways in addition to HH signalling. These include Wnt and Notch signalling pathways, GPCRs such as GPR161, the ion channels polycystin-1/2 (PC1/2) and signalling mediated by

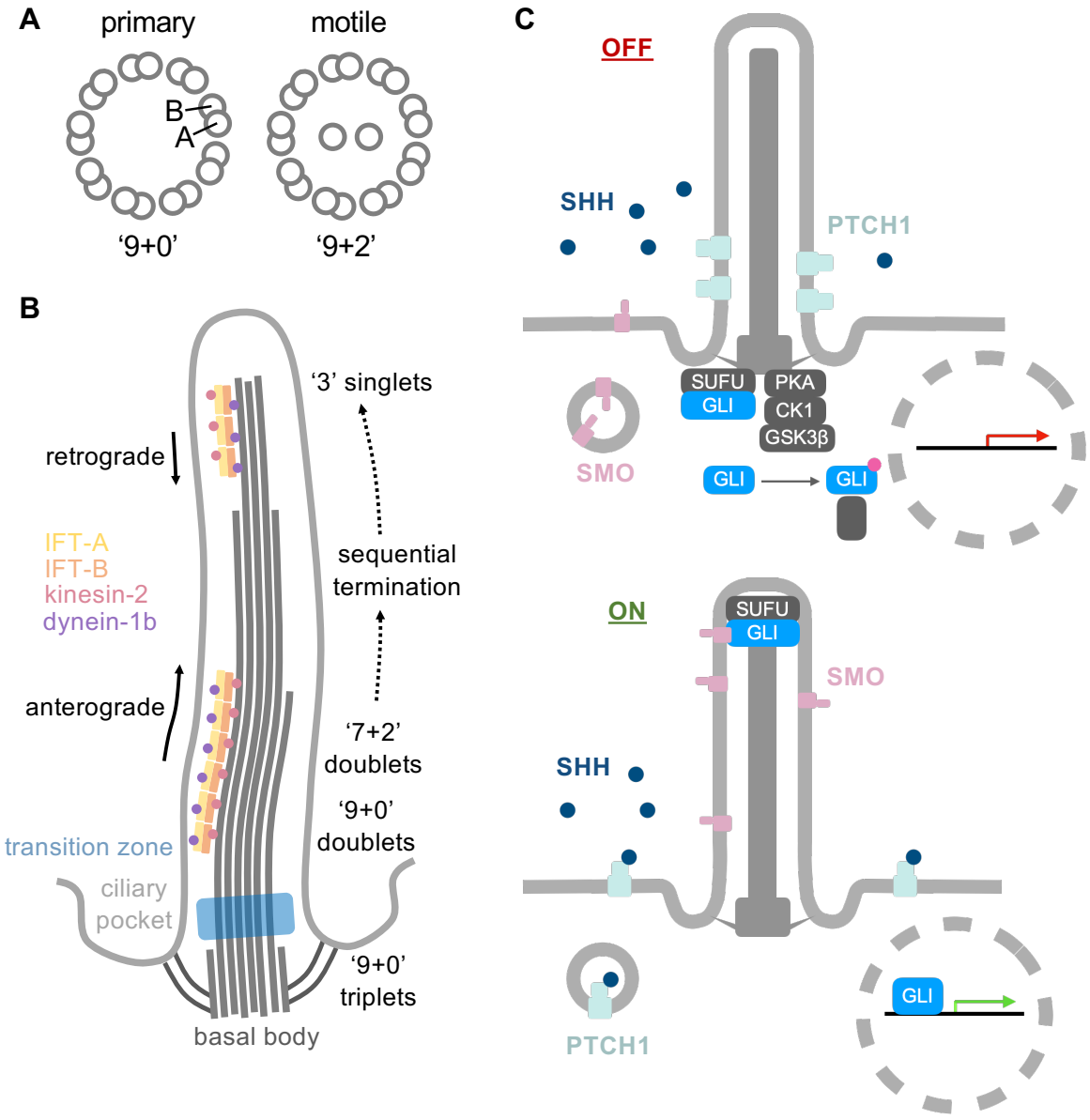


Figure 1.2: Architecture of the primary cilium.

A) Comparison of axoneme microtubule arrangements between primary and motile cilia. **B)** Schematic representation the primary cilia. Changes in axoneme structure along the ciliary length are highlighted. Additional structural protein (transition zone, basal body attachments) and membrane (ciliary pocket) components are indicated. Transport of ciliary localised proteins is mediated via intraflagellar transport (IFT) trains. **C)** Changes in distribution of HH pathway proteins during signalling.

the insulin-like growth factor 1 among others[41]. Many of these proteins also repartition between the ciliary and plasma membrane, for example binding of agonist to Dopamine receptor 1 triggers its exclusion from the cilium[42]. Movement of proteins within the cilium, including microtubules for axoneme assembly, is mediated by intraflagellar transport (IFT) trains bound dynein and kinesin[43] (Fig. 1.2B). Recent

cryo-ET structures and expansion microscopy data show IFT-B forms the basal protein scaffold onto which IFT-A, dynein-1b and kinesin-2 bind sequentially[38,39]. Kinesin-2 mediates cargo transport in the anterograde direction whereas dynein-1b mediates retrograde transport along microtubule tracks, hence coordinating localisation of ciliary proteins[44].

Why do cells invest energy in ciliary partitioning of signalling components? There are several advantages conferred by evolution of a semi-occluded signalling organelle. These include rapid and dynamic repartitioning of proteins between the cellular and ciliary compartments in response to extracellular conditions, with minimised cross-talk during dwell times[41]. Additionally, the high surface area to ciliary volume allows rapid detection of extracellular signals and fold changes in cilia localised secondary messenger levels such as cAMP. Hence, fewer downstream molecules are required to induce the same fold change in signal compared to in the cellular cytosol, increasing the speed and sensitivity of signal transduction. The probability of molecular encounters are also increased within the reduced pocket volume and lipid budding (e.g. for endo/exocytosis of signalling molecules) is assisted by the high membrane curvature. Thus, the energetic cost of ciliary production and molecule partitioning is offset by gains in signalling efficiency and molecule biogenesis.

1.1.3 Lipids in Hedgehog signalling

The discovery of HH morphogen bilipidation provided an early clue for the involvement of lipids in HH signalling[7,8,45]. SHH, DHH and IHH are the only proteins uniquely modified by both cholesterol and palmitate. At the molecular level HH signalling is also activated by sterols and oxysterols[46] (Fig. 1.3A). Dihydroxylated oxysterols, notably 20(S)-dihydroxycholesterol (20(S)-OHC) strereoselectively activate HH signalling by binding SMO[47]. Cholesterol binding to SMO is also both necessary and sufficient for SMO activation[22,48,49]. The plant steroidal alkaloid cyclopamine also inhibits SMO[50], a discovery arising from the birth of cyclops lambs when ewes had been feeding on corn lilies during gestation[51].

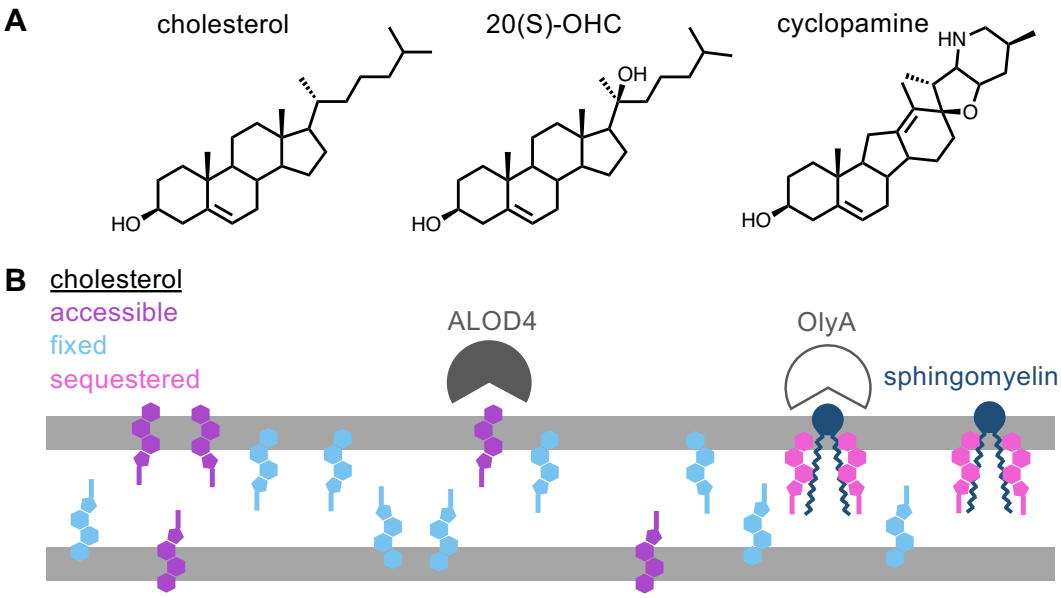


Figure 1.3: Steroids in HH signalling.

A) Structures of cholesterol, 20(S)-dihydroxycholesterol (20(S)-OHC) and cyclopamine. **B)** Membrane cholesterol occupies distinct pools; a fixed pool (blue) responsible for membrane integrity and biophysical properties, an active or accessible pool (purple) and a sequestered pool such as in complex with sphingomyelin (pink). PFO-based probes (ALOD4) are able to bind the accessible pool whereas fungal toxins (OlyA) can interact with the sphingomyelin sequestered pool.

Sterol involvement is further evidenced by differential partitioning between the plasma and ciliary membrane whereby the ciliary transition zone is sterol depleted[52]. PTCH1 also localises to this region[15], leading to the proposal that sterol binding to SMO is directly controlled by PTCH1 activity[53]. Recent biochemical analyses support this hypothesis and suggest PTCH1 controls cholesterol accessibility either by intermembrane re-partitioning or transfer to an extra/intracellular donor[21,54]. To understand this, we must first re-consider our understanding of membrane cholesterol, not as a homogenous population, but as residing in distinct membrane pools (Fig. 1.3B). Technological advancement of cholesterol-bindings probes, co-opted from bacteria and fungi, illuminate the presence of at least three cholesterol pools; a sequestered pool (such as in complex with sphingomyelin), a fixed pool (for maintenance of membrane integrity) and an accessible cholesterol pool[55,56]. Combined with CRISPR analyses, probe-based studies show PTCH1 depletes levels of accessible cholesterol. SMO activation is further increased when sphingomyelin levels are depleted and cholesterol shifts between the sequestered and accessible

pools[54]. Fluorescence based imaging methods diverge as to whether levels of PTCH1 regulated accessible cholesterol differ between inner and outer membrane leaflets and therefore how cholesterol binds to specific sites on SMO[21,57]. In addition, these studies use overexpressed PTCH1 within the plasma membrane. Hence, it remains unclear which cholesterol binding sites on SMO responsible for *in vivo* pathway activation in a ciliary context.

In addition, other lipids such as phosphatidyl inositols are heterogeneously partitioned. PI(4,5)P_2 localises to the base of the ciliary pocket and plasma membrane whereas the ciliary membrane is enriched in PI(4)P due to specific ciliary localisation of INPP5E, the enzyme which catalyses dephosphorylation of PI(4,5)P_2[58,59]. The biological implications of this remain understudied but may regulate ciliary length[60] or protein trafficking[61].

1.2 Protein-lipid Interactions as Viewed by MD Simulations

1.2.1 Protein-lipid interactions

Membrane proteins can be grouped into three subcategories; integral membrane proteins containing transmembrane segments which bisect both membrane leaflets, peripheral membrane proteins which bind to a single leaflet and anchored membrane proteins bound to the membrane via non-protein attachments[62]. All subtypes (by definition), comprise regions that are exposed to the lipid bilayer. The physical and chemical properties of these segments are uniquely suited to compliment the membrane environment in which they reside and distinct from those which govern interactions of connecting linkers with the surrounding water. For example, the TMDs of integral membrane proteins are predominantly hydrophobic whereas peripheral membrane proteins may have positive residues which anchor the protein to anionic lipids[63].

While broadly applicable biophysical ‘rules’ are useful for identifying generalisable features governing membrane protein behaviour, work over the last two decades suggests protein-lipid interactions are diverse[64] and depend on e.g. protein structure

or membrane localisation[65,66]. Hence, the nature and strength of lipid interactions varies across the membrane protein surface and between different proteins resulting in an interaction continuum from lipids as solvent to lipids as specific ligands.

This continuum of lipid interaction types and lipid chemical diversity facilitate roles in a range of biological processes[64]. These include: inhibitors such as for sphingomyelin and the Na^+/K^+-ATPase[67], activators as with PIP_2 binding to G-protein gated inwardly rectifying K^+ (GIRK) channels[68] and protein regulation such as cholesterol binding to rhodopsin[69]. In addition, membrane proteins are further regulated by membrane biophysical properties like tension, thickness, curvature and rigidity such as occurs for the mechanosensitive ion channel Flycatcher1[70].

Pioneering work using molecular dynamics (MD) simulations to examine protein-lipid interactions has focused on understanding the latter end of this spectrum; that is those specific binding events which exhibit the strongest or longest interactions with proteins. This is, in part, due to *in silico* assessment of protein-lipid interactions observed via structural methods which must survive the relatively harsh purification and solubilisation steps and are therefore likely high affinity interactions[71]. Simulations have also been used to predict the lipid binding sites where direct biochemical evidence was present but the site location was unknown. For example, the location of activating PIP_2 binding sites on each subunit of Kir tetrameric channels was predicted from coarse-grained (CG) simulations[72] two years before structural determination[73]. Similar simulation protocols have been applied to rationalise data derived from mass spectrometry which revealed PIP_2 binding to several Class A GPCRs[74]. These interactions prevail on experimental timescales, likely due to the strong electrostatically driven nature of the protein-PIP_2 contacts.

Recently studies have begun to assess protein-lipid interactions from a more encompassing perspective[64]. For example, MD simulations increasingly report lipid interactions over the protein surface, rather than being limited to a subset of experimentally observed sites[75,76]. This runs parallel to use of more complex or biologically relevant membrane compositions[77,78] or tissue/disease specific markers[79,80]. Experimentally, NMR studies or the GPCR leukotriene B_4 receptor BLT1 show lipid dependant modulation of protein equilibria between functionally relevant

states that are further modulated by ligand binding[81]. Hence, we are poised to understand the diversity of lipid interactions across the proteome, as exemplified by a recent study of lipid interactions with all *E. coli* inner membrane proteins[82], and begin to assess the functional implications of interactional variety. Understanding environment specific site dynamism may be particularly relevant to 'weaker' lipid interactions which have K_ds within the range of physiologically relevant lipid concentrations, and which are more amenable to assessment by computational studies.

1.2.2 Simulations of Hedgehog pathway proteins

SMO:

MD simulations have been applied to a subset of HH pathway proteins although, like many other aspects of HH signalling, protein dynamics remain understudied (Fig. 1.4). Almost all simulation studies have focussed on SMO, likely due to the availability of earlier single-domain crystal structures[83–86]. SMO simulations can be broadly separated into i) ligand/drug binding studies, ii) analysis of SMO activation mechanisms and iii) interactions with the surrounding membrane (Fig. 1.4A).

Early atomistic MD simulations of the SMO TMD investigated binding of ligands to the upper TMD site. These included investigation of ligand stability and contacts when antagonists LY2940680[87], vismodegib, cyclopamine or taladegib were bound, or with molecules identified from virtual screening such as ZINC12368305[88]. Docking and simulations were also used to demonstrate TMD pocket volume expansion within the inhibitor resistant SMO D473H mutant[89] and free energy calculations were applied to predict ligand binding affinities[87,89]. A combined docking and simulation study investigated oxysterol binding to a homology model of the extracellular SMO cysteine-rich domain (CRD), suggesting orientation of oxysterol tails may be important for ligand specificity[90]. Docking studies were also used to investigate binding of alternative pharmacophoric precursors such as flavonoids to the SMO TMD[91]. There are, however, some caveats with these simulations such as omission of a membrane bilayer[92], short simulations and a reliance on ligand binding poses generated by docking[88,89].

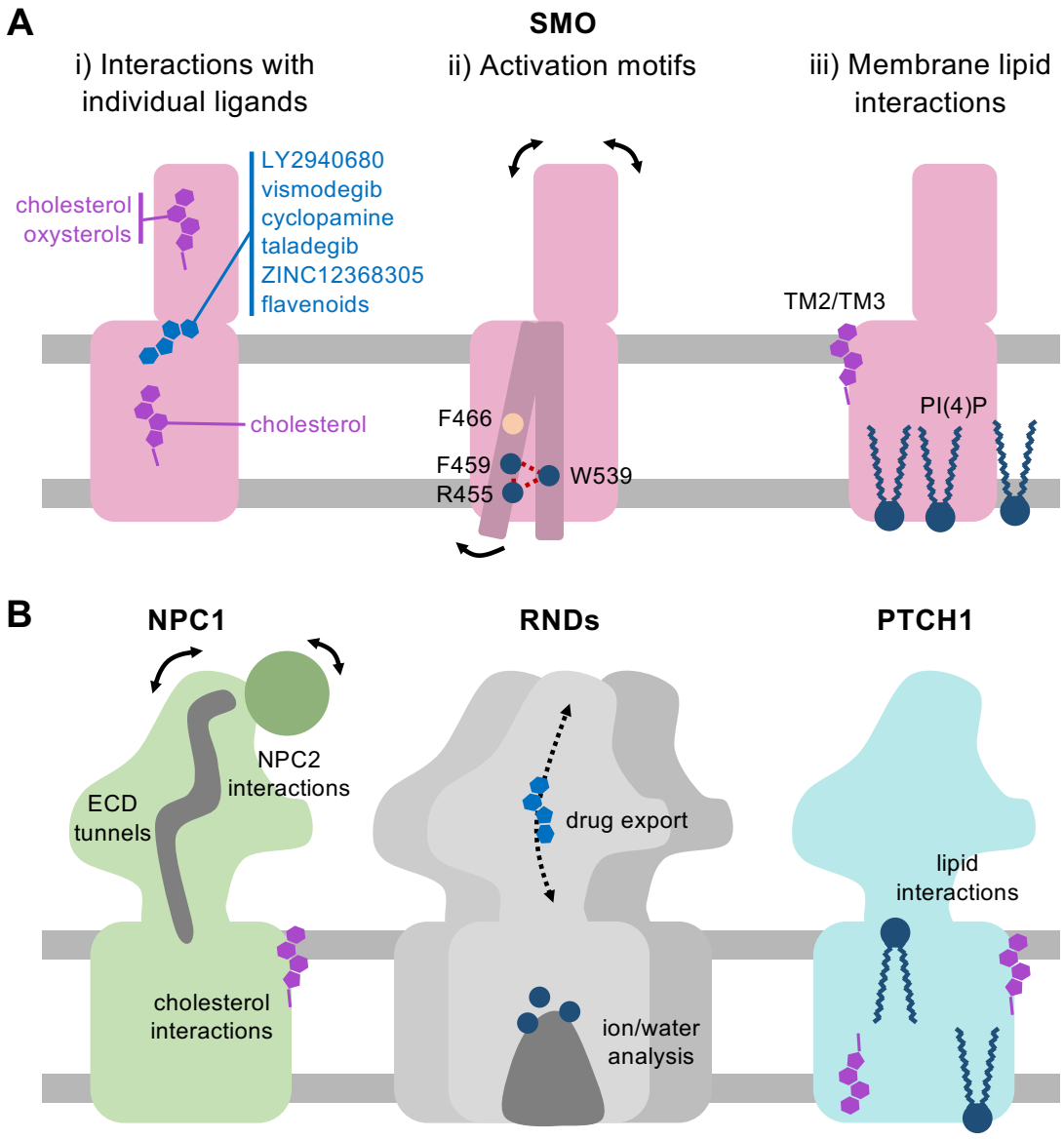


Figure 1.4: Computational studies of HH pathway proteins and homologues.

Summary of current simulation/docking studies on HH pathway membrane proteins or their homologues. **A)** SMO studies can be grouped into i) those investigating interaction of individual ligands (labelled) to specific binding sites, ii) investigation of SMO activation mechanisms including CRD dynamics, strait helical displacement of TM6 (orange) and proposed 'molecular switch' residues (blue), and iii) interactions with the surrounding membrane. Residue numbering for mSMO is shown. **B)** PTCH1 insights have been obtained from simulations of Niemann-Pick disease type-C1 (NPC1) or related Resistance-Nodulation-Division (RND) homologues in the absence of PTCH1 simulation studies. Recently PTCH1 simulations investigating interaction with the surrounding membrane have become available.

When the full-length SMO structure was determined with cholesterol bound to the CRD, simulations were used to show cholesterol induced CRD stabilisation compared to the apo conformation[48]. These were subsequently validated using a different SMO structure whereby authors proposed CRD stabilisation upon cholesterol binding may be important for activation[93]. Only one study investigates cholesterol binding to the TMD, whereby authors show reduced RMSD of cholesterol bound with the β_3-hydroxyl group coordinated by Y398 compared to a flipped conformation[94].

Later simulations of SMO applied biasing methods to investigate SMO activation mechanisms and/or the influence of bound ligands. For example, metadynamics simulations were used to investigate SMO-vismodegib and SMO-CRD-cholesterol complexes, suggesting vismodegib promotes movement of TM5/6/7 compared to when cholesterol is bound. Authors also propose the CRD sterol binding groove may expand to contact ECL3, proposed to represent an 'auto-inhibited' conformation in the absence of cholesterol[95]. A complementary study applied adaptive biasing and umbrella sampling simulations to show formation of an activation barrier occurs earlier along the reaction coordinate when vismodegib was bound compared to CRD-cholesterol or when apo[96]. Both studies also suggest CRD motions are coordinated to those of ECL3 and/or the linker domain[95,96].

An alternative approach to investigating SMO activation has focussed on integrated sequencing, biochemical and simulation data to probe activation motifs within the TMD. For example, authors show the TM6 and TM7 residues R455 and W539 have a low amount of basal ('natural') sequence variation but are frequently mutated in cancer. In simulations cancer mutants R455H and W539L fluctuate more and TM6/TM7 contacts are decreased compared to in the WT, leading to a proposed 'molecular switch' required for receptor activation and G-protein engagement[97]. Comparison of SMO with related Frizzled receptors suggests loss of a key TM6 proline residue (phenylalanine in SMO) prevents receptor activation proceeding via formation of a bent helix. Instead, SMO activation proceeds via concerted and strait TM6 displacement which authors proposed may aid discrimination of downstream signalling partners compared to Frizzled receptors[98].

Finally, CG simulations have been used to investigate SMO interactions with the surrounding bilayer. Initially, these focussed on cholesterol interactions whereby a prominent cholesterol site between extracellular TM2/TM3 segments emerged[99]. This site was subsequently validated in two other CG studies[100,101] and has a ~10 kJ mol^{-1} predicted binding affinity[99]. Binding of the phosphatidyl inositol PIP_2 was also investigated[99] however other studies suggest phosphatidyl inositol interactions with SMO are reduced compared to other GPCRs[101] which may reflect differences in cilia localised PIPs. To address this, a recent study investigated PI(4)P binding to SMO, including a modelled cytosolic C-terminal domain (CTD). They suggest PI(4)P bridges interactions between the CTD and the membrane which may affect partner protein recruitment[102].

PTCH1, NPC1 and RNDs:

Prior to the onset of this thesis there were no simulations of PTCH1. Subsequently, simulation insight was acquired from studies of PTCH1 homologues Niemann-Pick disease type-C1 (NPC1) and prokaryotic Resistance-Nodulation-Division (RND) transporters (Fig. 1.4B).

As with SMO, simulation studies of NPC1 focus on atomistic resolution of ligand binding and the effect of disease associated point mutants. Simulations suggest the upper portion of the extracellular domain (ECD), termed the N-terminal domain (NTD) is more dynamic in simulations than the middle and lower counterparts[103] an effect exacerbated by the I1061T mutant[104]. Docking studies have been used to identify putative cholesterol binding sites throughout the ECD[105] and several studies report identification of ECD localised tunnels[103,106]. The identified tunnels are broadly S shaped, with two kinks located at the base of the NTD and the interface of lower ECD regions. The transport compromised L472P mutant leads to tunnel break in NPC1 simulations[106] however a phenotypically similar R518W mutant did not affect tunnel formation[103]. Intriguingly, one study reports a 10 Å shift to the position of a TMD bound cholesterol, induced by the transport defective P691S mutant[107]. Polar repulsion of the cholesterol tail in the P691S mutant, positions cholesterol within the lower ECD region at the end of the simulation, overlapping approximately with the position of a structurally observed itraconazole NPC1 inhibitor[108].

In addition, a subset of studies are dedicated to understanding the mode of NPC1 engagement with NPC2, the lysosomal shuttle responsible for cholesterol transfer during biogenesis. Simulations and free energy calculations suggest one binding mode (termed the 'Texas' pose) is more stable when cholesterol is bound to NPC2[109] but an alternative 'Californian' pose is favoured when cholesterol is localised within NPC1[110]. These simulations have been used to rationalise a model for cholesterol hand-off and NPC2 re-orientation during interface engagement, however, there is no known NPC2 equivalent in HH signalling.

RND transporters are a large, functionally diverse protein family, for which a number of simulation studies are available[111]. As such, this summary is limited to a few key papers which implement similar methodology to described within Chapter 3. AcrB is the most well studied prokaryotic RND transporter. Structure-derived models have been used to suggest substrate export by the AcrB trimer is coupled to sequential transition between open (O), loose (L) and tight (T) binding states across protein subunits[112,113]. In simulations, conformation specific water networks are observed within the TMD whereby the L- and O-states permit formation of a cytosolic water cavity which evolves into a thin membrane-spanning water wire in the T-state[114,115]. Targeted MD simulations have been used to drive conformational transition between the AcrB T- to O-states coupled to steered MD and umbrella sampling simulations of the doxorubicin substrate through the ECD[116]. The free energy at the top and base of the ECD are similar however several undulations are observed along the doxorubicin transport pathway ranging 10-20 kJ mol^{-1}. Surprisingly, doxorubicin is surrounded by a layer of water molecules throughout the transport tunnel, despite the relative hydrophobicity of the substrate, which authors suggest may lubricate smooth transition along the pathway. A simulation and structural study of the trimeric heavy-metal exporter CusA identified lateral movements of the ECDs between 'Binding' and 'Extrusion' states when Cu(I) ions were added or removed[117]. Authors also use pKa calculations to predict protonated residues within the TMD which may be involved in formation of a water wire. While there is evidence for a sequential three-state conformational cycle in multidrug exporters like AcrB, uncoordinated transitions have been proposed in other trimeric RNDs such as CusA and CmeB[117,118] and it remains

unclear how monomeric RND members such as PTCH1, MmpL3 and NPC1 may coordinate their functions.

A couple of PTCH1 simulation studies have recently become available. In one study, authors use docking to investigate binding of the putative PTCH1 inhibitor panicein A hydroquinone (PAH) and simulate PAH stereoisomers in water. They find PAH is more elongated during simulation in water compared to oxidised quinone-like counterparts which form π-π stacking interactions[119]. They also suggest PAH may bind to a pocket at the base of the PTCH1 ECD, in a similar location to itraconazole observed within a structure of NPC1[108]. In a separate study, extensive CG simulations were used to explore lipid interactions with PTCH1 monomers and dimers in plasma membrane-like bilayers whereby cholesterol interaction durations were prolonged in the monomeric state. Authors also found enrichment of PE, PS, PA, PI and GM3 lipids around PTCH1 and suggest GM3/cholesterol interactions colocalise to similar regions[120].

Other HH pathway simulations:

Some simulations of HH morphogens have been performed. This included SHH binding to a small loop of PTCH1 ECD2 modelled based on a SHH-HHIP structure. Authors suggest ion binding is integral to complex formation and use docking to pharmacophorically characterise the interface[121]. However atomistic simulations were extremely short (2-4 ns) and later structures call into question the relevance of ECD2 for SHH engagement[122]. A combined QM/MM study was used to investigate Zn(II)-SHH hydrogen bonding interactions[123] and MD simulations were used to assess DHH-BOC interface stability, decreased by the DHH W173C mutant[124].

Prior to this study, there were no simulations of HHAT however combined modelling and simulation studies of Porcupine (PORCN) and ghrelin O-acyl-transferase (GOAT) homologues have been performed in the absence of experimental structural data. Both studies use simulations to validate the topology of predicted models. For PORCN models were combined with antibody labelling of the protein and N-linked glycosylation sites at specified ER exposed loops[125]. For GOAT site-directed mutagenesis and functional assays were used to justify structural predictions[126]. Both studies predict 11 TM helices which form a trapezoid arrangement around a central

cavity, proposed to bind acyl-CoA substrates. It is worth noting that both models differ from more recent AlphaFold2[127] predictions implemented within simulations reported in Chapter 5.

1.3 Research Novelty and Objectives

Full-length structures of HH pathway membrane proteins are just beginning to emerge, aided by advances in cryogenic electron microscopy (cryo-EM). Current HH pathway simulations are limited and mostly focussed on understanding how individual ligands affect SMO behaviour and/or on putative drug candidate screening. This thesis seeks to address two fundamental gaps in our understanding of HH membrane proteins. Firstly, improved understanding of the basic biological mechanisms governing protein function and regulation. Secondly, how the underlying conformational landscape dictating protein dynamics is altered by the presence of one or more ligands.

In parallel, this thesis outlines the development of simulation methodologies for the standardisation of protein-lipid interactions. These methodologies were designed to be encompassing, minimally disruptive (i.e. unbiased and/or suitable for high-throughput applications), replicable and include metrics for assessing data quality. Crucially, the methods explored within Chapter 2 were guided by key biological questions and existing experimental protocols, with a view toward integrated structural biology.

A tripartite study of HH pathway membrane proteins is detailed forthwith, including candidates previously unstudied by MD simulations:

- In Chapter 3, extensive atomistic simulations of PTCH1 and other RND homologues are combined with free energy calculations to investigate cholesterol transport through PTCH1. Simulations provide energetic and stoichiometric contextualisation to PTCH1 transport models designed to bridge the gap between current structural and cellular data. In addition, we show how the SHH-cholesterol moiety contributes to PTCH1 inhibition by blocking formation of tunnels throughout the PTCH1 ECD.

- Chapter 4 explores the relationship between multiple ligand binding events on the SMO CRD and TMD. We identify a mechanism of allosteric communication between ligand binding sites and explore the molecular rearrangements which occur during SMO activation.
- Chapter 5 details the structure and dynamics of HHAT. We detail the dynamic behaviour of catalytic residues, identify a mechanism of solvent gating, and explore interactions with the surrounding membrane environment. In addition, simulations are used to identify a putative product exit gate and investigate the molecular determinants of divergent protein dynamics across the membrane bound O-acyltransferase (MBOAT) family.

This work was conducted in close collaboration with members of the Siebold (STRUBI Oxford) and Rohatgi (Stanford, CA) labs. Collaborations also engaged members of the Tate (Imperial), Ward (Scripps, CA), Jones (Imperial), Rasmussen (Würzburg, Germany) and Stansfeld (Warwick) groups.

2. Methods for Membrane Protein Simulations

"Nature does not name itself. Granite does not self-identify as igneous. Light has no grammar. Language is always late for its subject. Sometimes on the top of a mountain I just say, 'Wow'."
Robert Macfarlane

2. Methods for Membrane Protein Simulations 18

2.1 Molecular Dynamics Simulations of Membrane Proteins 20
2.1.1 Forcefields and parameterisation 20
2.1.2 Running MD simulations of membrane proteins 24

2.2 Methods for Determining Protein-Lipid Interactions 24
2.2.1 Identification of interaction events 24
2.2.2 Selecting cut-off values 25
2.2.3 Quantification of lipid binding kinetics 27
2.2.4 Binding site identification 30
2.2.5 Obtaining representative lipid binding poses 34
2.2.6 Assessing binding site quality 37
2.2.7 Integrating simulation and structural data 40

2.3 Affinities of Protein-Lipid Interactions 45
2.3.1 Biased sampling methods 45
2.3.2 Unbiased sampling methods 48
2.3.3 Comparing methods for obtaining protein-lipid affinities 51

This chapter contains data also published in:

Ansell, T.B., Song, W., Sansom, M.S.P. (2020). The Glycosphingolipid GM3 Modulates Conformational Dynamics of the Glucagon Receptor. *Biophys. J.,* 119, 300-313

Corey, R.A., Song, W., Duncan, A.L., **Ansell, T.B**., Sansom, M.S.P., Stansfeld, P.J. (2021). Identification and assessment of cardiolipin interactions with *E. coli* inner membrane proteins. *Sci. Advances,* 7, eabh2217

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. (2021). Relative Affinities of Protein-Cholesterol Interactions from Equilibrium Molecular Dynamics Simulations. *J. Chem. Theory Comput*., 17, 10, 6548-6558

Song, W., Corey R.A., **Ansell, T.B**., Cassidy, C.K., Horrell, M.R., Duncan, A.L., Stansfeld, P.J., Sansom, M.S.P. (2022). PyLipID: A Python Package for Analysis of Protein-Lipid Interactions from Molecular Dynamics Simulations. *J. Chem. Theory Comput.*, 18, 2, 1188-1201

Ansell, T.B., Song, W., Coupland, C.E., Carrique, L., Corey, R.A., Duncan, A.L., Cassidy, C.K., Geurtz, M.M.G., Rassmussen, T., Ward, A.B., Siebold, C., Stansfeld, P.J., Sansom, M.S.P. (2022). LipIDens: Simulation Assisted Interpretation of Lipid Densities in Cryo-EM Structures of Membrane Proteins, *Nat. Comms., under review*

Methodological practices for MD simulations of proteins are now well established. These methods are detailed briefly below with a focus on application to membrane proteins and are covered in extensive detail elsewhere[128–130]. Accordingly, this chapter describes how best to integrate MD simulations to study membrane protein-lipid or -ligand interactions. This chapter details recent methodological advances in characterising lipid/ligand interactions, designed to standardise protocols used to determine binding sites and their associated kinetics. This chapter also details how

protein-lipid binding affinities can be determined using biased sampling and equilibrium methods.

2.1 Molecular Dynamics Simulations of Membrane Proteins

Computational simulations can be used to study the motions of molecules, using a method known as molecular dynamics (MD). In classical or molecular mechanics simulations, the electronic configuration of electrons around atoms are ignored (Born-Oppenheimer approximation). This enables the simulation of much larger systems, because the potential energy of the system at each timepoint is calculated from the position of only the atomic nuclei. The resultant trajectory, i.e. a description of the behaviour of atomic nuclei with time, is obtained by iterating Newton's equations of motion for all atoms in the system[128,129].

2.1.1 Forcefields and parameterisation

In molecular mechanics a forcefield refers to the collection of equations which describe the inter- and intra- molecular behaviour of atoms and a set of parameters for atomic interaction types in a given environment. The potential energy function underlying a forcefield can be decomposed into harmonic descriptions of bond stretching and angle bending, a torsional potential for dihedral motions and the Coulombic equation and Lennard-Jones potential for non-bonded electrostatic and van der Waals interactions respectively (Equation 2.1).

$$[2.1] \quad U(r) = \sum_{bonds} \frac{k_b}{2}(b - b_0)^2 + \sum_{angles} \frac{k_\theta}{2}(\theta - \theta_0)^2 + \sum_{dihedrals} \frac{k_\phi}{2}(1 + \cos(n\phi - \phi_0)) + \sum_{impropers} \frac{k_\xi}{2}(\xi - \xi_0)^2 + \sum_{\substack{non-bonded \\ pairs}} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{\substack{non-bonded \\ pairs}} \frac{q_i q_j}{4\pi\varepsilon_o r_{ij}}$$

The harmonic potential for bond stretching describes deviation in bond length b from the equilibrium length b_0 and k_b indicates the force constant of the bond. A similar potential is used to describe the energy associated with angle (θ) changes from their

equilibrium value (θ_0) where k_θ is generally less than k_b. The dihedral potential describes the torsional motion of two atoms whereby the number (n) and relative position/phase (ϕ_0) of energetic minima compared to the current bond phase (ϕ) is described. k_ϕ indicates the barrier height of the cosine function describing dihedral rotation. The improper dihedral is a harmonic potential for out of plane deviation (ξ) of certain flat groups of atoms from the reference plane (ξ_0) and constitutes the final bonded term of the potential.

Non-bonded interactions between two atoms (i, j) separated by distance r_{ij} are described by the Lennard-Jones and electrostatic potentials. The Lennard-Jones potential describes the repulsion of atoms in close proximity due to the Pauli exclusion principle $\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}$ and attraction at extended distances resulting from dispersion forces $\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}$. σ_{ij} is the value of r_{ij} at which the potential is zero and ε_{ij} is the potential well depth. Electrostatic interactions are described by the coulombic equation whereby q is the charge of the atom and ε_0 is the vacuum permittivity.

Forcefields are parameterised using biophysical data and quantum mechanical calculations[128]. A forcefield may for example include experimentally observed bond vibration frequencies from IR spectroscopy or bond lengths derived from small-molecule crystallographic data[129]. Forcefields are generally parasitised for a specific system type such as for application to biological macromolecules. The below sections describe commonly used forcefields for the simulation of protein and membrane components.

Atomistic forcefields:

Several atomistic forcefields are available for the simulation of membrane protein systems. These include variants of the widely used CHARMM, AMBER, OPLS and GROMOS forcefields[131–134]. In general, those most suitable include lipid parameters which were intentionally developed to be compatible with the protein forcefield components and which accurately recapitulate lipid properties such as gel phase transition temperatures[135]. This thesis implements the CHARMM-36 forcefield[131] (for

protein-lipid systems) and Amber-ff14SB[132] in combination with the Lipid17[136], GAFF2[137] and GLYCAM-06j forcefield[138] (for a glycosylated-protein-lipid-ligand system).

The CHARMM-36 forcefield was developed to correct the α-helical bias[139] of earlier CHARMM versions and to improve parameters for protein backbones and dihedrals[131]. CHARMM-36 was re-parameterised using QM energy surfaces computed with a higher level or theory, which were manually adjusted based on NMR data for unfolded peptides[140]. The parameters were further adjusted using NMR derived sidechain *J* couplings from folded proteins and were able to recapitulate a range of protein NMR properties[131].

Earlier versions of the CHARMM forcefield (22/27) for lipids were developed to reproduce QM calculations of smaller lipid substructures which, when combined, would match experimental geometries of the whole molecule[141,142]. The forcefield was somewhat successful at recapitulating experimental findings however lipids readily condensed into ordered phases above gel phase transition temperatures[135]. The CHARMM-36 lipid forcefield was parameterised to tackle these limitations, incorporating NMR data on lipid headgroups, improved treatment of non-bonded interactions and adjusted dihedral parameters at the headgroup-tail interface[131]. Today, a wide variety of lipid types are available for the CHARMM-36 forcefield which can be conveniently packaged together using the CHARMM-GUI bilayer builder[143].

The AMBER-ff14SB forcefield was developed to tackle problems with helical bias and preferential sidechain rotamers present in AMBER prerequisites. AMBER-ff14SB employs many of the same potentials as previous versions (ff94, ff99, ff99SB) but uses improved QM calculations to make empirical adjustments to bond rotation and dihedral parameters[132]. The General Amber Forcefield (GAFF) was designed for compatibility with other AMBER forcefields and is used extensively for drugs and ligands[137]. Unlike other forcefields where parameters are developed for each unique molecular entity, GAFF employs general atomic descriptions for organic chemical space. Thus, parameters for whole ligands can be developed by a process of pattern matching chemical units. There are obvious limitations with this approach which may result in discrepancies from QM predictions of e.g dihedral angles, however they can provide

a useful starting point for investigating basic ligand properties and/or improving ligand parameters including e.g. recent machine learning optimisation methods for matching to QM data[144].

The first AMBER forcefield for lipids, Lipid11[145] was predominantly parameterised using GAFF. Subsequent forcefields, sought to integrate experimental data to improve the treatment of dihedral angles and Lennard-Jones interactions (Lipid14[136]), or expand the repertoire of lipids (Lipid17). The resultant forcefield could recapitulate key biophysical properties of membranes such as bilayer thicknesses, areas per lipid and NMR order parameters[136].

Coarse-grained forcefields:

Further exploration of membrane protein behaviour has been enabled by use of coarse-grained (CG) simulations. In CG simulations atoms are grouped together into beads, particles[146,147] or building-blocks[148], resulting in a smoothened energetic landscape and a reduced system size. The most commonly used CG forcefield for simulation of biological macromolecules is the Martini forcefield[146] which maps approximately every 4 atoms into one bead. Martini version 2[149] has been extensively applied to study the biophysical processes of membrane and/or protein systems. These include the determination of protein-lipid interactions or lipid fingerprints, membrane-protein oligomerisation and complex formation, lipid partitioning and phase behaviour, lipid dynamics and synergy, membrane curvature, vesicle/micelle formation and fusion and viral particle simulations[66,150–153]. A recent development to this forcefield, Martini 3, expands upon the repertoire of CG beads available for simulation by re-parameterising tiny (2 to 1 mapping), small (3 to 1) and regular (4 to 1) beads for cross compatibility[154]. Preliminary results suggest Martini 3 has close to ideal mixing behaviour of bead types, a reduced propensity for protein-protein aggregation and more closely matches some experimental free energies compared to Martini 2. However, development of Martini 3 lipid parameters is ongoing and therefore lipid species variation is somewhat limited. As such, Martini 2 is used most frequently throughout this thesis.

2.1.2 Running MD simulations of membrane proteins

When running MD simulations, one must define a number of parameters pertaining to e.g. the treatment simulation box boundaries, temperature/pressure coupling, electrostatics, Lennard-Jones interactions and restraints (detailed extensively elsewhere[128,129]). These parameters can be implemented using one of many packages such as the GROMACS simulation software[155] which is used in this thesis.

There are a number of additional considerations for running simulations of membrane proteins including alignment of the bilayer with respect to the box orientation and independent treatment of pressure changes in xy (membrane plane) *vs* z (membrane normal). The methods for membrane protein simulations are now well established[130,156], examples of which can also be found within the results chapters of this thesis.

The following section details how to obtain meaningful analyses from membrane protein simulations.

2.2 Methods for Determining Protein-Lipid Interactions

Details underlying the PyLipID[157] toolkit and LipIDens[158] pipeline are described within this section.

2.2.1 Identification of interaction events

A central problem within simulations of membrane proteins relates to selection of a metric which best defines a protein-lipid interaction event. Traditionally a single distance cut-off has been used to differentiate bound *vs* unbound lipid. The cut-off is usually selected based on observed minimum distances between VDW radii; ~0.3 nm in atomistic or ~0.6 nm in CG simulations[159]. An additional complexity arises from which protein or lipid groups are used in the distance calculations e.g. centre of mass of an atom/bead *vs* the whole residue. One disadvantage of using a single distance cut-off is that it cannot account for transient increases in the protein-lipid distance due

to Brownian motion, resulting in premature truncation of a continuous binding event (Fig. 2.1). This is particularly relevant to CG simulations where the increased timestep can result in distance jumps of ~0.2-0.3 nm between steps[157]. Thus, an alternative metric is required which can sufficiently capture transient fluctuations in protein-lipid distances.

A solution to this problem is offered by introduction of a dual distance cut-off scheme (i.e. lower cut-off: 0.55/upper cut-off: 1.0 nm). A continuous interaction duration is initiated when the protein-lipid distance come within the lower cut-off and is ended only when the protein-lipid distance surpasses the upper cut-off. Early applications of dual cut-off schemes were applied to simulations of the Class A GPCR adenosine receptor (A_{2A})[160] and the Class B glucagon receptor (GCGR)[161], capturing key lipid binding events for PIP_2 and the glycosphingolipid GM3. This dual cut-off scheme underpins the protein-ligand analysis toolkit PyLipID we developed to robustly capture and quantify ligand binding events[157].

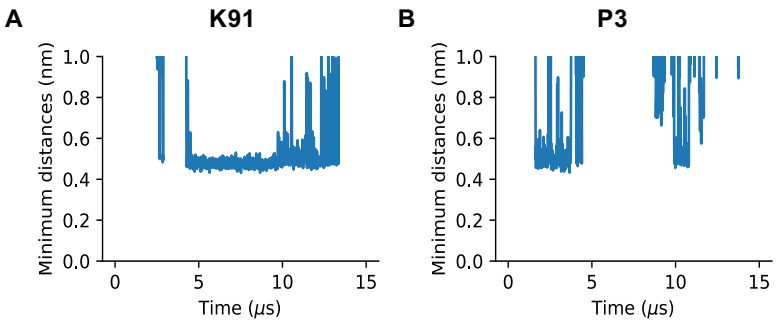


Figure 2.1: The dynamic nature of protein-lipid contacts.

Fluctuations to the minimum distance between HHAT residues **A)** K91 and **B)** P3 and a PIP_2 lipid across a 15 μs CG trajectory. The minimum distance is defined between any bead of the lipid and residue.

2.2.2 Selecting cut-off values

Various analysis can be performed to test different dual cut-off values. Initially, cut-offs can be selected from the probability distribution of minimum protein-lipid distances. This is exemplified for interactions of PIP_2 with the enzyme Hedgehog acyl-transferase (HHAT) derived from 10x 15 μs CG simulations within a mixed lipid environment (Fig. 2.2A-B). Here, and in all subsequent PyLipID analysis, the minimum

distance is defined between the centre of mass of any atom/bead of the lipid and any atom/bead on the protein. The lower cut-off is selected based on the location of the first solvation peak. The upper cut-off corresponds to the trough proceeding the first interaction shell.

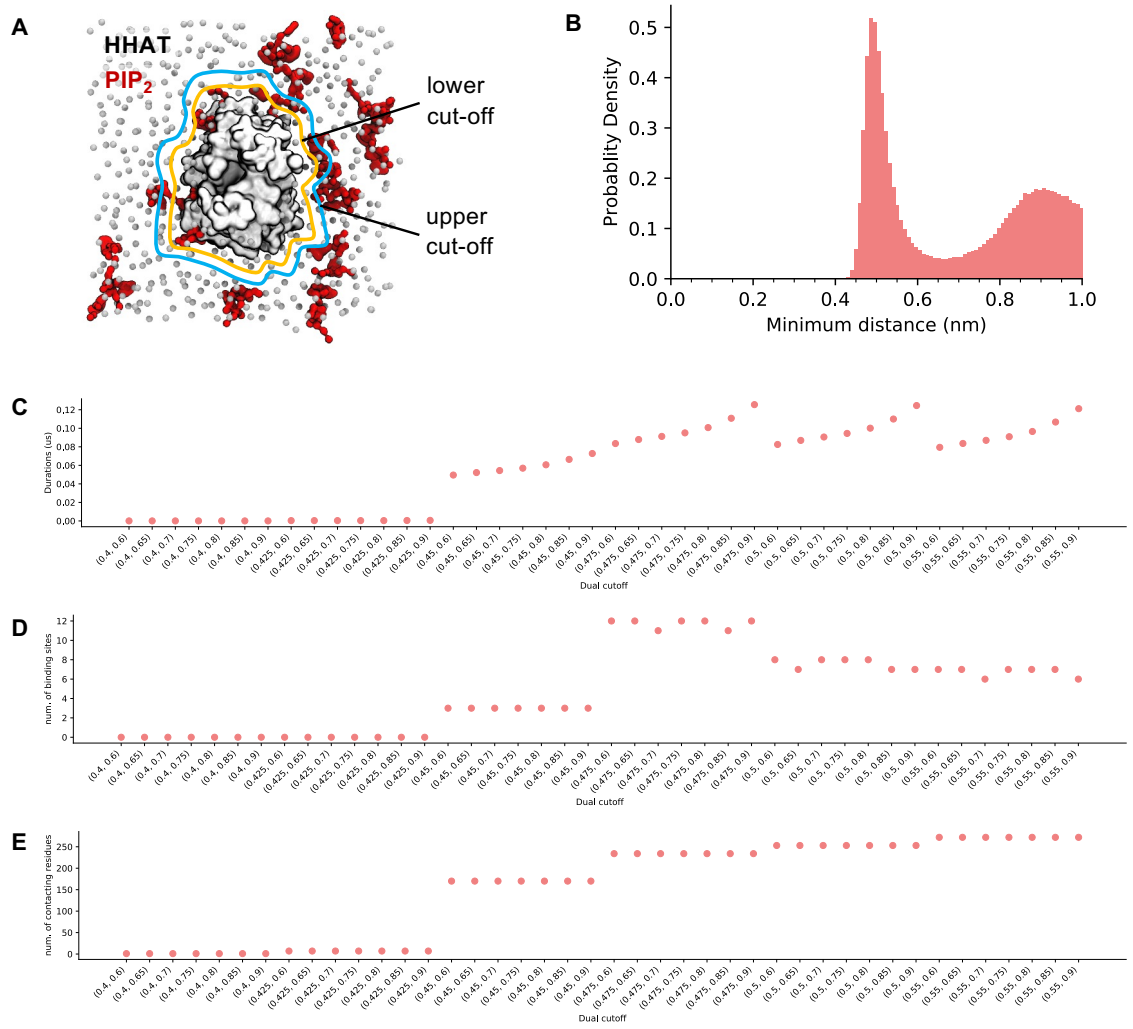


Figure 2.2: Testing dual cut-off schemes.

A) CG representation of HHAT (white) as viewed from the ER luminal face. Lipid phosphate beads are shown as grey spheres and PIP_2 is coloured red. **B)** Probability distribution of HHAT-PIP_2 minimum distances provided the lipid came within 0.65 nm of the protein for at least 30 frames. **C-E)** Exhaustive testing of different dual cut-off pairs using PyLipID. The mean interaction duration **(C)**, total number of binding sites **(D)** and total number of contacting residues **(E)** for HHAT-PIP_2 interactions are indicated for each cut-off pair.

Fine tuning of cut-off values can be achieved through exhaustive testing of interaction durations and numbers of contacting residues using different lower/upper cut-off value combinations (Fig. 2.2C-E). A clear jump in interaction durations and no. interacting residues is observed at 0.45 nm for HHAT-PIP_2 interactions, below which the lower cut-off is too small to detect interaction events. Interaction durations reach a plateau when the upper cut-off is greater than ~0.7 nm while the no. contacting residues are invariant. Interaction durations increase again at very long upper cut-off values (~0.9 nm) as these capture PIP_2 dissociation events into the second lipid solvation shell (Fig. 2.2B). This suggests there is a medial range of upper cut-off values between 0.65-0.8 nm which are sufficiently large to capture lipid jumps resulting from Brownian motion but small enough to exclude PIP_2 molecules which have completely dissociated form the first solvation shell.

Based on these data, we used a lower cut-off of 0.475 nm and an upper cut-off of 0.7 nm for assessment of HHAT-PIP_2 interactions. The early studies of A_{2A} and GCGR used a 0.55/1.0 nm dual cut-off scheme for all lipids[160,161]. While the lower cut-off is reasonable, subsequent testing suggests an upper cut-off of 1.0 nm may be too generous and could potentially result in artifactually prolonged interaction durations by capturing lipids within the second solvation shell. Additionally, analysis of recent systems suggests there may be subtle differences in cut-off choice between lipid types and/or proteins[157]. This emphasises the importance of system specific testing of different dual cut-off schemes.

2.2.3 Quantification of lipid binding kinetics

To obtain a comprehensive picture of protein-lipid interactions we sought to develop methods for robust quantification of interaction kinetics. Thus, the PyLipID toolkit was extended to standardise kinetic derivations within a modular python package.

The primary interaction metrics are defined as follows. The interaction *duration* is the average duration of lipid contacts, as calculated using the dual cut-off scheme described above. The *occupancy* is the percentage of frames where a lipid contact is within the lower cut-off with respect to the total number of frames in a trajectory. The

lipid *count* is the number of lipid molecules surrounding the contact site, also defined using the lower cut-off. Finally, we define an interaction *residence time*, which is used to reduce down-shifted bias in interaction durations resulting from a large number of transient interactions. Plotting the sorted index of interactions (Fig. 2.3A) shows many transient contacts and a subset of more prolonged interaction events. These data are used to plot the survival time correlation function σ(t) (Fig. 2.3B, purple dots), defined in Equation 2.2 where N_j is the total number of contacts, T is the length of the trajectory, t is a given time between 0 and T and $\sum_{v=0}^{T-t} \tilde{n}_j(v, v+t)$ is equal to 1 if the contact j lasts from v to $v+t$ or equal to 0 otherwise. σ(t) is calculated at each timepoint in the trajectory between 0 and *T* and normalised to 1 at $t = 0$ by dividing through by σ(0).

$$[2.2] \qquad \sigma(t) = \frac{1}{N_j}\frac{1}{T-t}\sum_{j=1}^{N_j}\sum_{v=0}^{T-t}\tilde{n}_j(v, v+t)$$

A biexponential curve is fitted to the normalised survival time correlation function (Fig. 2.3B, red line) from which the short (k_{off2}) and long (k_{off1}) decay constants are obtained (Equation 2.3). Hereafter the dissociation constant k_{off} corresponds to the long decay value k_{off1} which is inversely proportional to the residence time τ (Equation 2.4).

$$[2.3] \qquad \sigma(t) \sim Ae^{-k_{off1}t} + Be^{-k_{off2}t} \quad (k_{off1} \leq k_{off2})$$

$$[2.4] \qquad \tau = \frac{1}{k_{off}} \quad : \quad k_{off} = k_{off1}$$

The k_{off} value can also be calculated via Bayesian bootstrapping[162] of the contact durations and recalculating the k_{off} /residence times (Fig. 2.3B, grey lines). Comparison of the k_{off} derived from curve fitting with the bootstrapped value can provide a useful method for assessing the sampling of interaction events. In addition, R^2 values are provided for biexponential fitting of the normalised survival time correlation function which may also be used to assess the quality of derived k_{off}s. All four interaction metrics (duration, occupancy, count and residence time) can be applied on a per residue or per binding site basis (see section 2.2.4).

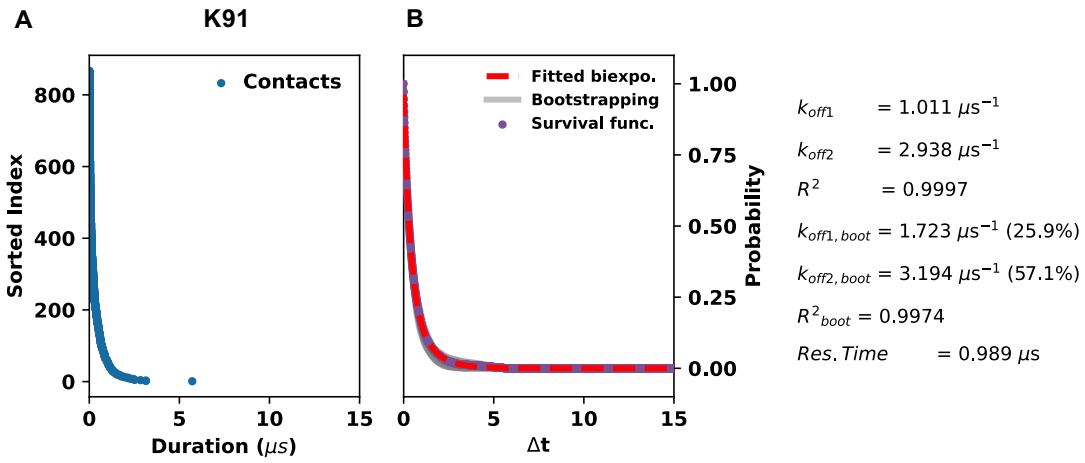


Figure 2.3: Calculating interaction residence times.

Per residue lipid interaction kinetics as obtained using PyLipID. **A)** Sorted interaction durations for HHAT residue K91 with PIP_2 across CG simulations. **B)** The short (k_{off2}) and long (k_{off1}) interaction decay constants were calculated by bi-exponential fitting (red) to the survival time correlation function (purple, Equation 2.2). Per residue k_{off}s ($k_{off1,boot}$/$k_{off2,boot}$) were also calculated via bootstrapping the interaction durations (10 times). The residence time is defined as $1/k_{off1}$.

To demonstrate the calculation of lipid binding kinetics we applied PyLipID to the glucagon receptor (GCGR), a Class B GPCR. GCGR was simulated in a mixed lipid environment comprising 10 lipid species, of which PIP_2 and the glycosphingolipid GM3 showed preferential localisation around the receptor. We mapped interaction durations, occupancies, lipid count and residence times for GM3 and PIP_2 headgroups onto the structure of GCGR (Fig. 2.4). Thus, PyLipID can be used to untangle the diversity of protein-lipid interactions and robustly quantify differences in their interaction kinetics.

While here we apply PyLipID to assess lipid interactions, the technicalities of the code are such that the software can be applied to obtain interaction kinetics of any ligand type. The PyLipID code implements MDTraj[163] for handling of trajectories and can therefore be used for all standard simulation software's and both atomistic and CG forcefields. In addition, the code is open source (github.com/wlsong/PyLipID) and structured in a highly readable and optimised modular format. Thus, implementation

of the package can be tuned and adapted to the users' requirements to e.g. obtain the residence times of protein oligomerisation events[150].

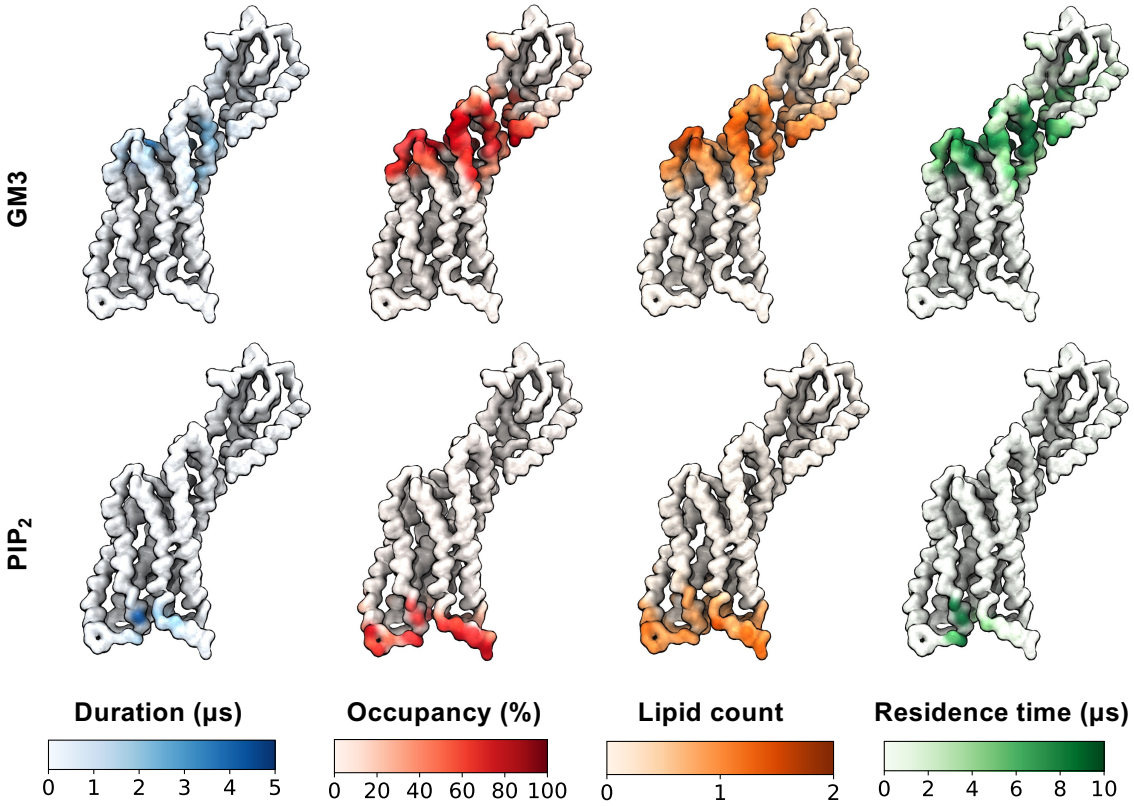


Figure 2.4: Lipid interaction kinetics.

GM3 and PIP_2 headgroup interactions with the glucagon receptor (GCGR) from CG simulations in a mixed lipid environment[161,164]. Lipid headgroup interaction durations, occupancies, counts and residence times are shown on the structure of GCGR, as calculated using PyLipID with a 0.55/1.0 nm dual cut-off scheme.

2.2.4 Binding site identification

Methods for determining lipid binding sites have developed in sophistication over recent years. Binding sites can be crudely identified from visualisation of trajectories or assessment of simulation averaged lipid densities around the protein[159]. While these methods are successful in identifying the most prominent binding sites, they may be subject to user biased interpretation of e.g. the exact residues comprising a site and/or selection of representative binding poses. This rudimentary method of site identification is exemplified for GCGR where PIP_2 binding sites were identified from

average PIP_2 densities across the simulations (Fig. 2.5A). Charged residues in proximity to sites were selected by the user based on a) proximity to the lipid density and b) assessment of interaction durations. For strong lipid interaction events driven by relatively 'simple' electrostatic proximity effects this method is sufficient. However more stringent methods of determining binding sites may be needed for lipids which form transient interactions or those which may be governed by the relative topological arrangement of residues such as e.g. for cholesterol.

PyLipID implements a community analysis approach to determine binding sites. This method, originally applied to examine cholesterol interactions with the Kir2.2 channel[165], groups together residues which simultaneously interact with a lipid into 'communities' based on a correlation matrix of lipid interactions. Each residue corresponds to a node. A distance vector is calculated between each residue node and each lipid across the trajectory. Edges between nodes are weighted according to the Pearson's correlation coefficient of the distance vectors for each residue pair. Communities of connected residues are calculated using the Louvain algorithm[166] whereby highly modular (densely connected) nodes are grouped together (Equation 2.5).

$$[2.5] \qquad Q = \frac{1}{2m}\sum_{i,j}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j)$$

$$[2.6] \qquad m = \frac{1}{2}\sum_{i,j} A_{ij}$$

Q indicates the degree of modularity, i and j are two adjoining nodes weighted by A_{ij}, k_i and k_j and the sum of the weighted nodes attached to nodes i and j respectively, m is the number of edges (Equation 2.6) c_i and c_j and the communities to which each node belongs and $\delta(c_i, c_j)$ is a binary function which equals 1 if $i = j$ or 0 otherwise.

During optimisation of the community network the nodes are sequentially ordered and assigned to a given community. Each node is then removed and reinserted into a different community until the network modularity stabilises (network optimization). After assignment of communities to the per residue nodes, each community is merged into a single node (i.e. a binding site) and the optimisation process is repeated until

only a single node is present. This is called an optimisation-aggregation loop and is a commonly applied method of node filtering and assignment during community analysis[166].

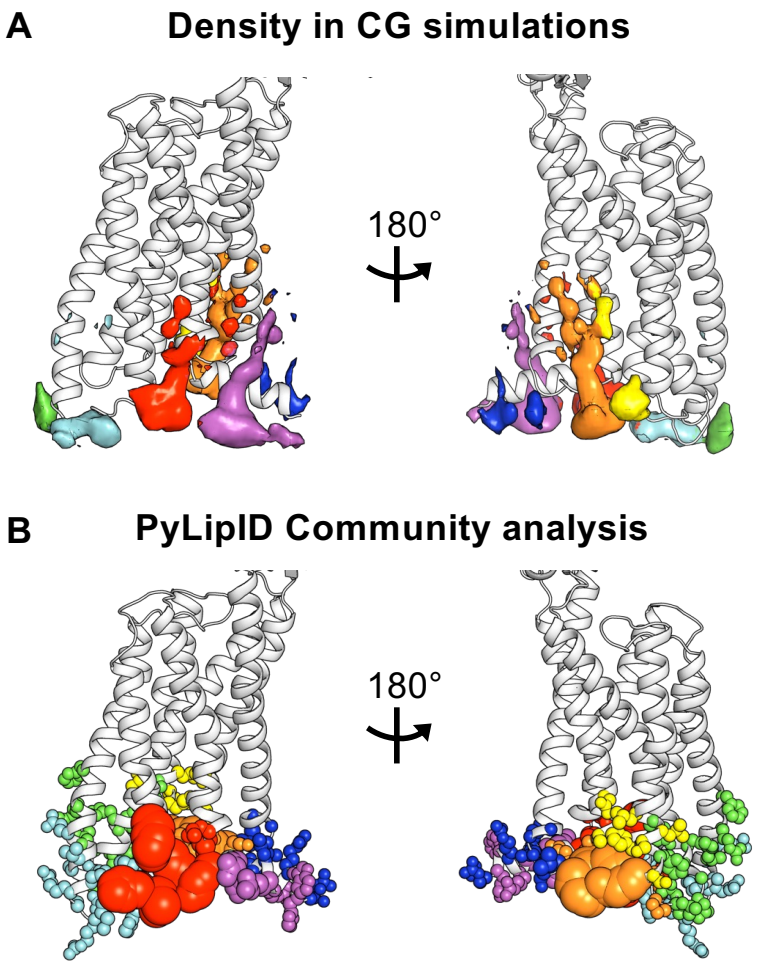


Figure 2.5: Determining lipid binding sites.

A) Time averaged density of PIP_2 lipids across CG simulations of the glucagon receptor (GCGR) in a mixed lipid environment[161,164]. Densities were calculated using the VMD VolMap plugin (σ=0.1)[167]. **B)** PIP_2 headgroup binding sites on GCGR, detected using the community analysis module implemented in PyLipID (cut-off scheme: 0.55/1.0 nm). Residues in each binding site are shown as spheres, scaled by residence time. Corresponding densities and binding sites are coloured accordingly.

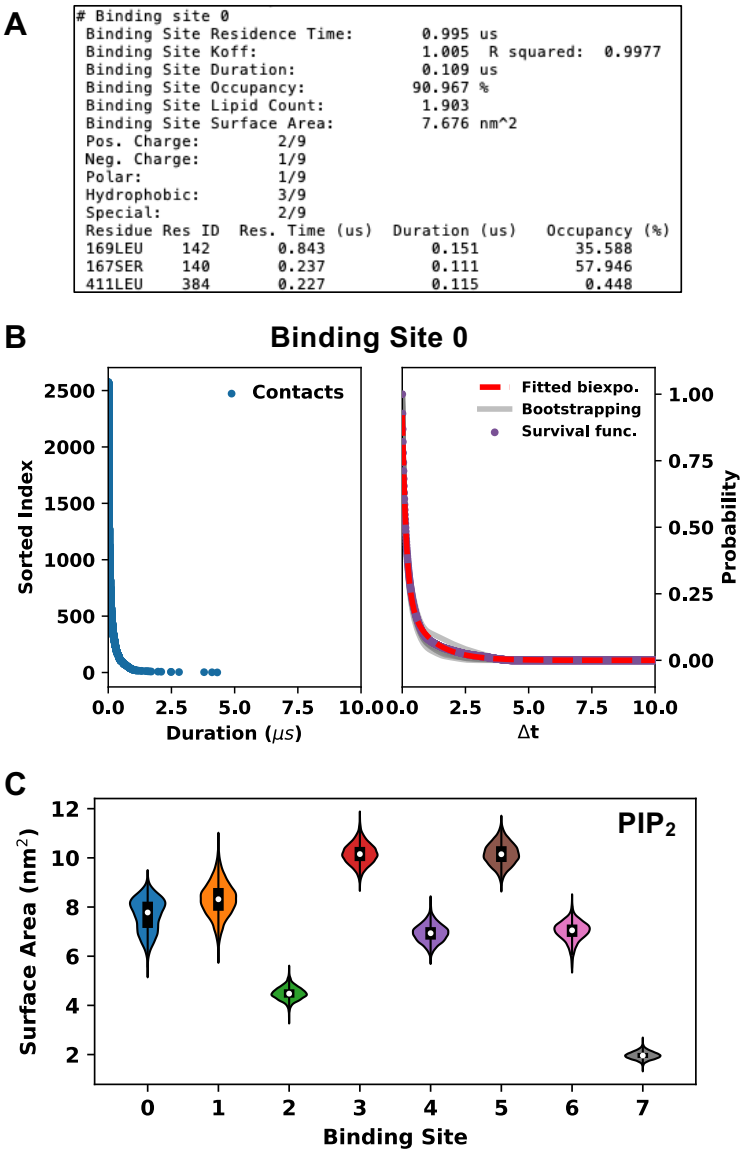


Figure 2.6: Exploring binding site metrics.

Detailed binding sites analyses for PIP_2 headgroup interactions with the glucagon receptor. **A)** A snapshot from the BindingSites_Info file outputted by PyLipID, summarising the kinetics of binding site 0 and its comprising residues. **B)** Residence time/k_{off} plot for all interactions at binding site 0 (see Figure 2.3 legend for details)**. C)** Total binding site surface areas, calculated across each frame in the CG trajectories using MDTraj[163].

PyLipID returns several outputs for visualisation of binding sites and their kinetics. Binding sites can be mapped onto an atomistic structure whereby residues are scaled according to a kinetic parameter (in this case residence time) (Fig. 2.5B). A summary .csv file is also produced which contains a list of the binding sites, the interaction metrics for each binding site (site residence time, site occupancy, site k_{off} etc) and a

list of the comprising residues and their associated kinetics (Fig. 2.6A). By default, the minimum number of residues which must comprise an identified site is 4 however this can be adjusted by the user. The interaction duration and survival time correlation function plot for each binding site can be used to assess the quality of derived site residence times/k_{off}s, akin to described on a per residue basis in section 2.2.3 (Fig. 2.6B). Finally, the surface area of each site is calculated using the Shrake-Rupley algorithm[168] implemented in MDTraj[163] whereby the accessible surface area of residues comprising each site are summed over each frame of the trajectory (Fig. 2.6C).

2.2.5 Obtaining representative lipid binding poses

Proceeding binding site detection, representative lipid binding poses are written to coordinate files using an empirical density-based scoring function (Equation 2.7), where W_i is the weight assigned to an atom i in the lipid, H is the bandwidth (default 0.15) and $\hat{f}_{i,H}(D)$ is a multivariate kernel density estimation of the atom position with respect to all lipid poses at a site. PyLipID uses as Gaussian kernel function for the kernel density estimation, as implemented in *statsmodels*[169]. The position of i is defined using a vector $[D_{i1}, D_{i2}, \ldots, D_{ip}]$ where D_{ip} indicates the minimum distance between the atom i and the residue p in the binding site. The user can specify whether to reweight specific atoms of the lipid to a higher weighting (by default all atoms/beads are weighted equally) during pose calculations. This can be useful for up-weighting of e.g. lipid headgroups or charged atoms/beads in electrostatically driven poses. The user can specify the number of top ranked lipid poses, which are written out as coordinate files (using the simulations snapshot from which the pose was derived), for each binding site. This is exemplified for the top ranked PIP_2 binding pose for a binding site on HHAT (Fig. 2.7A-B).

$$[2.7] \qquad score = \sum_i W_i \cdot \hat{f}_{i,H}(D)$$

To assess the diversity of lipid binding modes, poses are clustered using the KMeans function from *scikit-learn*[170] into n poses provided by the user. If n is not provided, poses are clustered in an unsupervised fashion using the DBSCAN module of *scikit-*

learn[170]. A random pose from each cluster (KMeans) or from one of the core samples of each cluster (DBSCAN) is written to a coordinate file (as with the ranked poses). Clustered PIP_2 poses for a binding site on HHAT are shown in Fig. 2.7C.

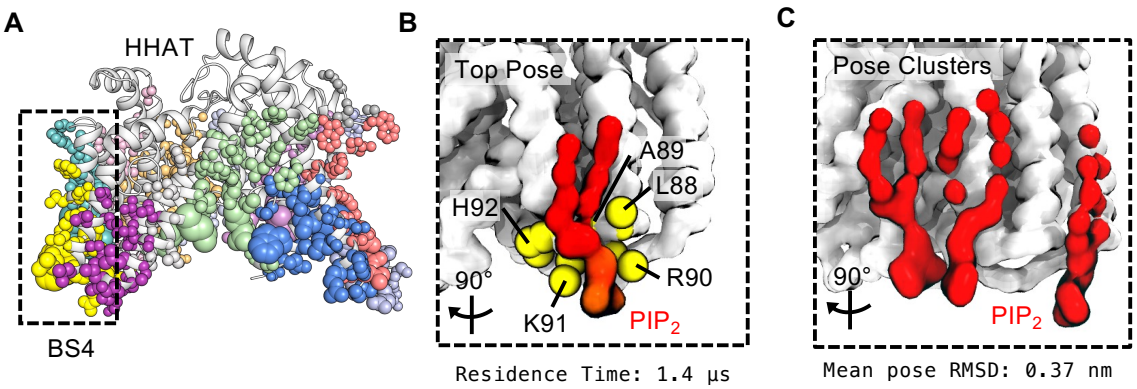


Figure 2.7: Binding site analyses and lipid pose detection.

A) PIP_2 binding sites on HHAT as calculated using the community analysis method implemented in PyLipID. Binding sites are coloured individually and residues in each site are scaled via residence times. The site with the longest residence time, binding site (BS) 4 is boxed. **B)** Top ranked PIP_2 binding pose at BS4. The top 5 residues in BS4 with the longest residence times are shown as yellow spheres and PIP_2 is coloured red. **C)** Three distinct PIP_2 pose clusters within BS4. The mean RMSD of all PIP_2 poses at BS4 is indicated.

The relative similarity of lipid poses at a site can be assessed using the RMSD of bound poses (Equation 2.8). N indicates the number of atoms in a lipid molecule, D_i is a distance vector for atom i to residues in the binding site and $\overline{D}$ is the average of distance vectors of atom i for all bound poses at a site. Thus, the RMSD indicates the relative deviation of poses with respect to the average position of bound poses.

$$[2.8] \qquad RMSD = \sqrt{\frac{\sum_{i=1}^{N}(D_i - \overline{D})^2}{N}}$$

The standardisation of binding site detection and pose calculation by PyLipID has enabled high throughput analysis of lipid binding sites. For example, we assessed cardiolipin (CDL) interactions with 42 inner membrane proteins from *E. coli*[82] (Fig. 2.8A). Using PyLipID we identified ~700 CDL binding sites which were asymmetrically distributed across the membrane. An increase in the prevalence of positive Arg and

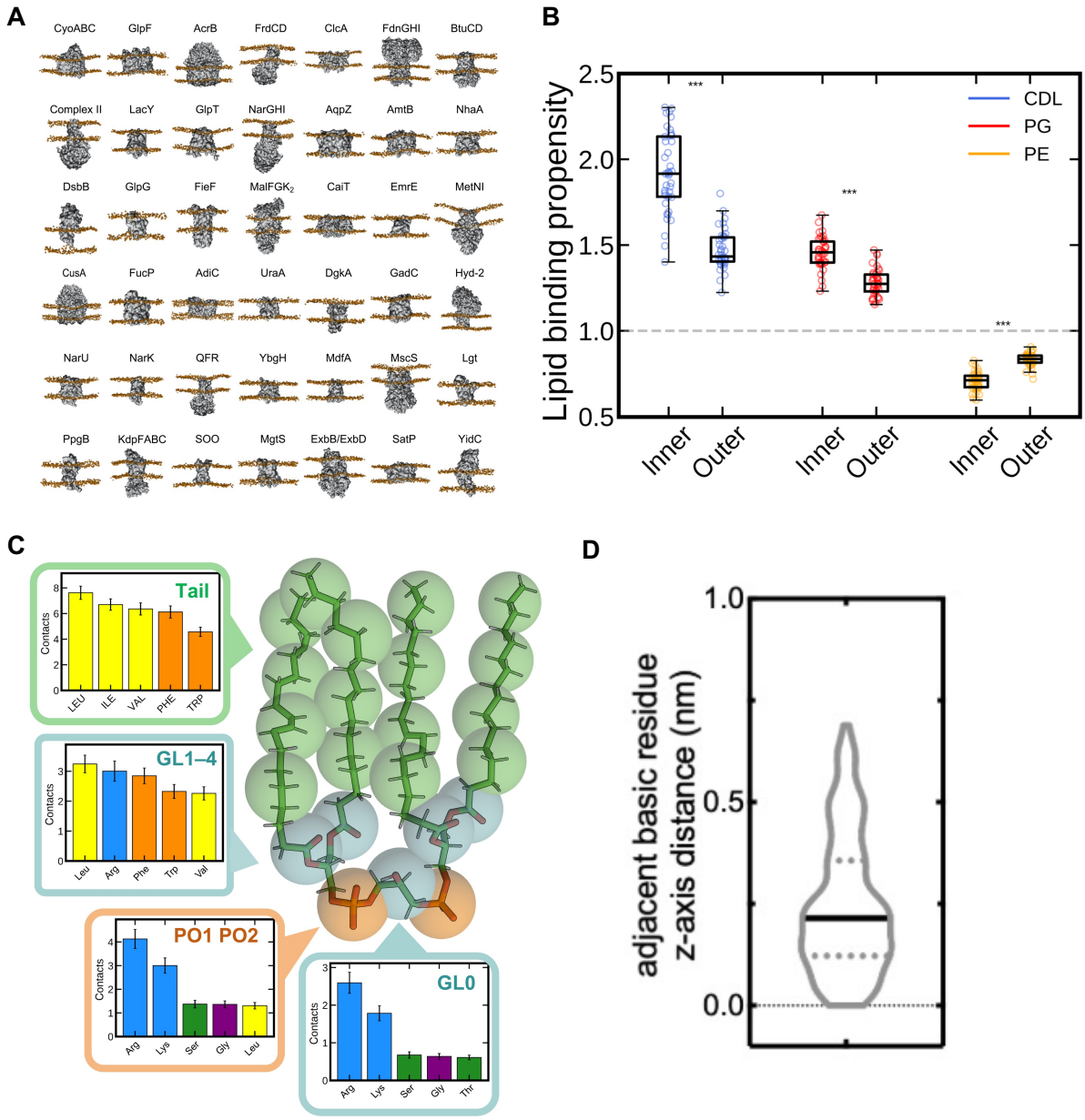


Figure 2.8: High throughput application of PyLipID.

Figure adapted from Corey *et* al., [82]. Simulations and analysis performed by Dr Robin Corey. **A)** CG representation of 42 integral inner membrane proteins from *E. coli* (grey). Lipid phosphate beads are shown in brown. **B)** Lipid binding propensity (defined as the lipid as a % of bound lipids/ lipid as % of total lipid) within each leaflet for bilayers composed of 67% POPE, 23% POPG and 10% cardiolipin (CDL). Bound lipid was defined using a 0.6 nm cut-off. **C)** Prototypical CDL binding site defined based on the number of contacts each martini bead type made to surrounding residues. PyLipID analysis was performed using a 0.55/1.0 nm cut-off scheme. **D)** Z axial distance of positive residues identified as being within 0.8 nm within detected CDL binding sites.

Lys residues on the intracellular transmembrane region of proteins favoured binding of anionic lipids, including CDL, to intracellular binding sites (Fig. 2.8B). Analysis of CDL binding sites also revealed the presence of a consensus CDL binding mechanism formed by Arg/Lys residues which coordinate the CDL phosphates and are accompanied by polar Ser/Thr residues or glycine, as well as an aromatic residue embedded more deeply in the membrane which may coordinate the glycerol groups (Fig. 2.8C). Furthermore, CDL binding was favoured by the presence of two or more Arg/Lys residues in close z axial proximity (0.2-0.3 nm) within the binding site (Fig. 2.8D). Thus, PyLipID can be used to identify the features which form a specific lipid binding site that can be used to predict interaction sites in the absence of simulation data. This highlights the potential applicability of PyLipID aided binding site detection to high throughput methods and/or machine learning protocols. The later may be particularly useful for characterising enigmatic lipid binding sites such as for cholesterol whereby sequence led site detection (e.g. through CRAC and CARC motifs[171]) still prevails, despite a body of evidence which suggests the composition of cholesterol sites to be much more nuanced[75,99,172].

Thus sections 2.2.1 to 2.2.5 detail how simulation methodologies have advanced from user biased manual lipid pose selection and binding site analysis to rigorous and robust metrics for capturing protein-lipid interaction kinetics, standardised binding site detection and pose determination. These methodologies have transformed the analysis of protein-lipid interactions described in subsequent chapters and represent a significant advance in computer aided interaction discovery.

2.2.6 Assessing binding site quality

To adequately utilise the kinetic and binding site information obtained from PyLipID, it is essential to assess the quality of derived interaction metrics. This can be achieved via ranking site residence times, occupancies, surface areas and Δk_{off} values (that is the difference in k_{off} obtained from biexponential fitting of the survival time correlation function and the k_{off} obtained from bootstrapping). We ranked all PIP_2 binding sites on HHAT from lowest to highest residence time/occupancy/surface area or closest to zero for Δk_{off} (Fig. 2.9A). Binding site 4 was ranked highest for occupancy and

residence time and had a Δk_{off} close to 0, indicative of a well fitted survival function (Fig. 2.9B). In contrast, binding site 12 was ranked last by all metrics (Fig. 2.9A). Assessment of site 12 interaction durations reveals low duration times that are infrequently observed, resulting in a poorly fitted interaction survival function (Fig. 2.9C). We note that while the presence of a very low site surface area is a hallmark of a poorly defined binding site, the site surface area does not scale linearly with residence time, as demonstrated also for cholesterol interactions with ten GPCRs[157]. Thus, the relative ranking of site surface areas (provided these are not very small) is less important than the other metrics.

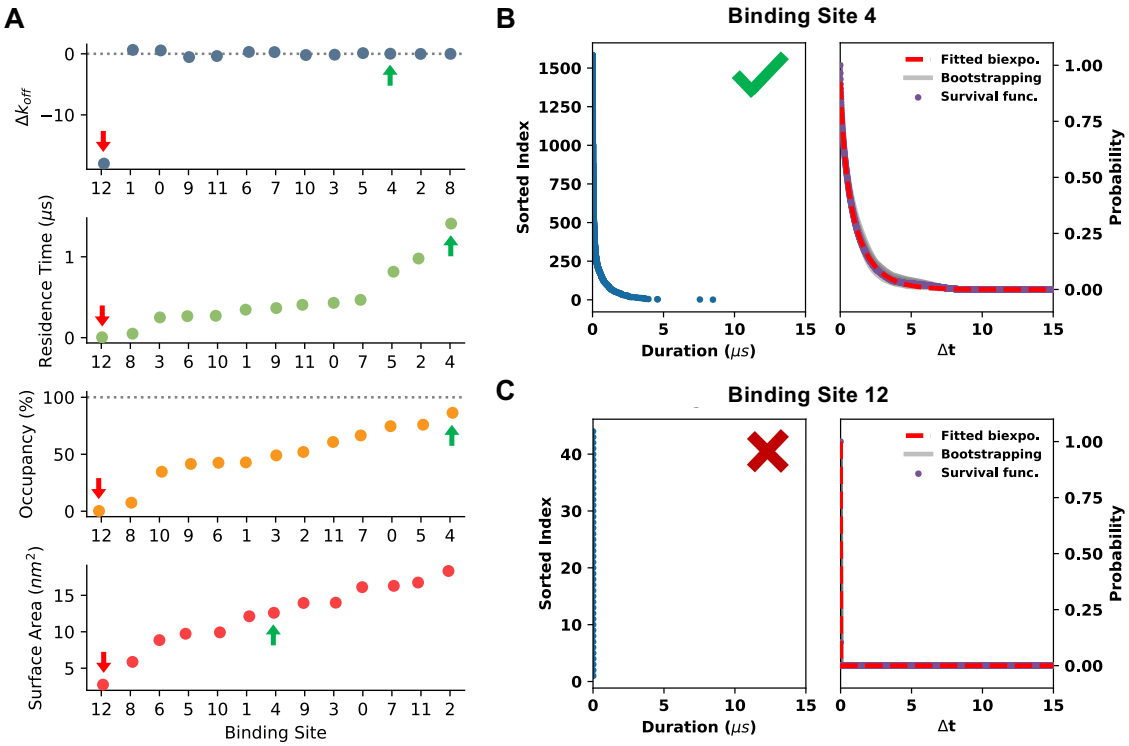


Figure 2.9: Screening PyLipID binding site outputs.

A) Ranked binding sites from order of lowest to highest site residence time (green), occupancy (yellow), surface area (red) or by Δk_{off} value (blue) closest to zero. Δk_{off} indicates the difference in site k_{off} derived from biexponential curve fitting minus that derived via bootstrapping the site interaction durations. Arrows denote examples of a well defined (green) or poorly defined (red) binding sites, also highlighted in **B/C**. Example binding site k_{off} plots for **B)** a well defined site or **C)** a poorly defined site. The left panel corresponds to the ranked site interaction durations. The right panel corresponds to the survival time correlation function used to obtain site k_{off}s (described in Figure 2.3).

Screening binding site data via this method is essential for identifying whether simulations have been run for long enough to adequately sample interactions the calculated binding sites and remove poorly defined/infrequently observed sites. Site ranking can also be used to infer which sites may be the most relevant/prevalent in a biological context, easing interpretation of PyLipID data.

An additional metric for assessing the quality of obtained sites may be possible via comparison to available structural data. Comparison of lipid poses from simulations of HHAT with structural densities revealed exceptional agreement with the predicted lipid binding sites (Fig. 2.12, see below for discussion). The PyLipID analysis toolkit has been applied successfully to a number of other protein examples[76,130,173,174] (Fig. 2.10) which also match structural lipid and ligand binding poses. Thus, kinetic ranking of PyLipID data and comparison to known experimental sites both serve as suitable methods for assessing the quality of calculated binding sites.

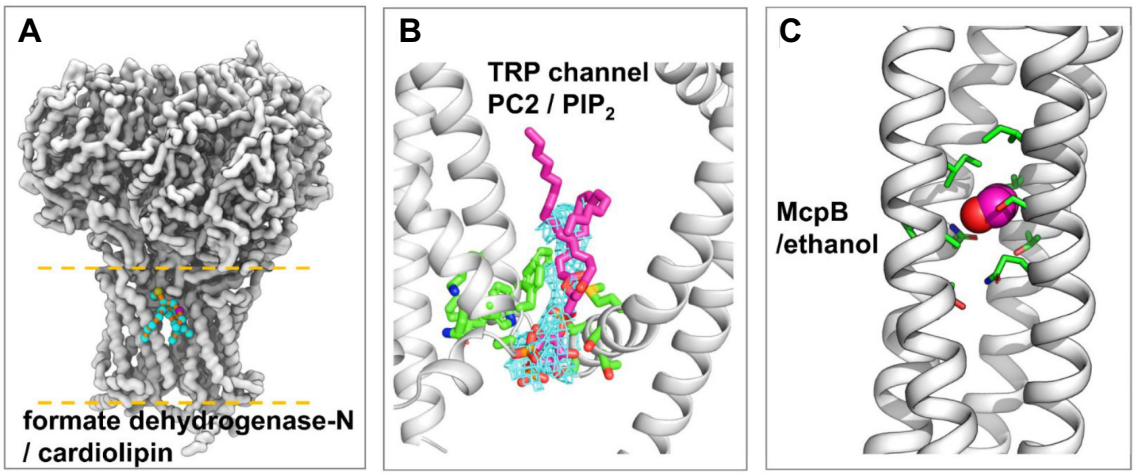


Figure 2.10: Diverse applications of PyLipID.

Figure adapted from Song *et al* [157]. Simulations and analysis were performed by Dr Robin Corey **(A)**, Dr Keith Cassidy **(C)** and myself **(B)**. **A)** Cardiolipin (cyan) binding site with the longest residence time bound to formate dehydrogenase-N. **B)** PIP_2 (pink) binding site on the TRP channel polycystin-2 (PC2). Binding site residues calculated by PyLipID are shown as green sticks surrounding the PIP_2 density in the cryo-EM map (cyan). **A** and **B** were derived from CG simulations and match structurally observed sites. **C)** Ethanol (pink) binding site on McpB, as determined from atomistic simulations.

2.2.7 Integrating simulation and structural data

Methods for determining structures of membrane proteins by cryo-EM, continue to sophisticate and, often reveal additional non-protein densities resulting from the detergent or nanodisc environment[175,176]. However, assignment of these additional densities is challenging, owing to the moderate local resolution at more extended differences from the transmembrane core and the chemical similarity of lipid/detergent species[176–178].

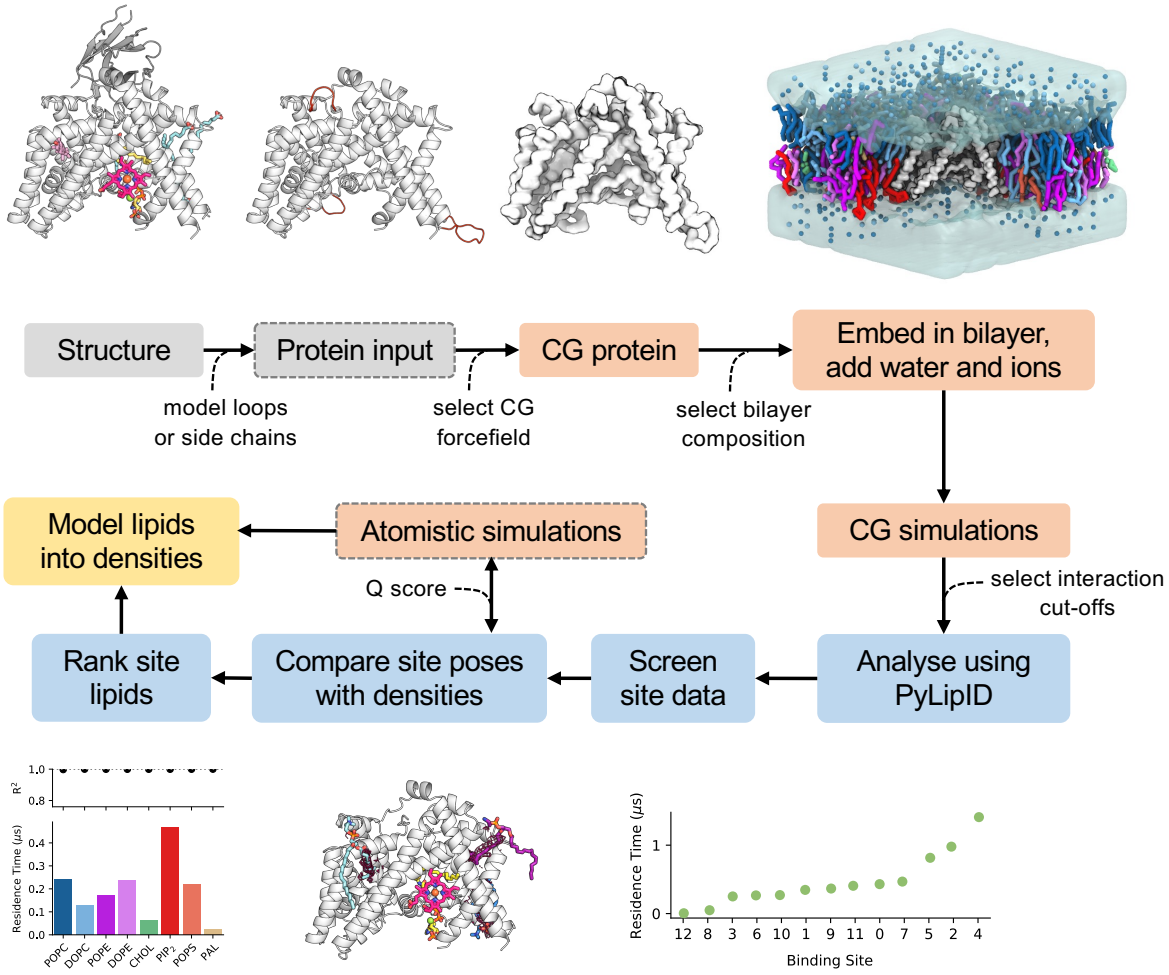


Figure 2.11: LipIDens: a pipeline for simulation assisted interpretation of lipid-like densities.

A workflow for assisted interpretation of lipid densities using simulations, exemplified using HHAT[179]. Steps involving structure processing (grey), setup and performing MD simulations (orange), analysis of lipid sites/densities (blue) and modelling (yellow) are indicated. Optional steps are boxed by grey dashed lines.

MD simulations can be used to complement structural methods by discerning which lipid species are most likely to bind in a corresponding location to lipid-like densities. We extended the methodologies described above into a pipeline termed LipIDens (Fig. 2.11); first implementing PyLipID to identify lipid binding sites in CG simulations then ranking the relative residence times of all lipids which bind to a site. Identified lipid binding poses could then be compared against structural densities and/or refined using atomistic simulations to assist in cryo-EM model building and map refinement. For a detailed breakdown of the pipeline see [158].

We applied LipIDens to assist interpretation of five lipid-like densities surrounding HHAT (Fig. 2.12). Identified lipid binding sites matched the location of all densities, with CG lipid poses closely mimicking the shapes of observed densities despite the reduced resolution (Fig. 2.12, left panel). The lipid binding poses were refined using atomistic simulations (Fig. 2.12, middle left) revealing surprising good agreement with cryo-EM densities (Fig. 2.12, middle right). The match between the simulation derived lipid poses and the lipid-like densities were assessed using Q scores[180]. For HHAT we observe reasonable Q scores for lipid poses, considering these were derived *ab initio* from the simulations in the absence of any density guided restraints (Q_{avg} = ~0.4 compared to ~0.7 for structurally modelled palmitate moieties and HHAT heavy atoms at 2.7 Å). This included a kinked lipid binding site extending into the HHAT core which could be captured in CG simulations initiated from a preliminary 5 Å map, before the lipid density was revealed when the map resolution was improved to ~2.7 Å (Fig. 2.12A). Thus, LipIDens can be applied at multiple points in the cryo-EM process to assist model building and refinement as the resolution improves. Finally, we ranked the relative residence times of all lipids at a given site and used these to suggest the most likely identity of the lipid-like densities within a bilayer context (Fig. 2.12, right). Sites exhibit variable residence times and relative orders of ranked lipids. For examples, the kinked lipid density in Fig. 2.12A can be occupied by the tails of PE and PC lipids for extended periods (residence time ≥ 15 μs) in addition to cholesterol entering the enclosed lipid pocket. By contrast the site in Fig. 2.12D is occupied by a number of lipids with PIP_2 exhibiting the longest site residence time (~0.5 μs). This highlights the non-binary nature of lipid binding to a given site and the power of LipIDens to assist structure interpretation.

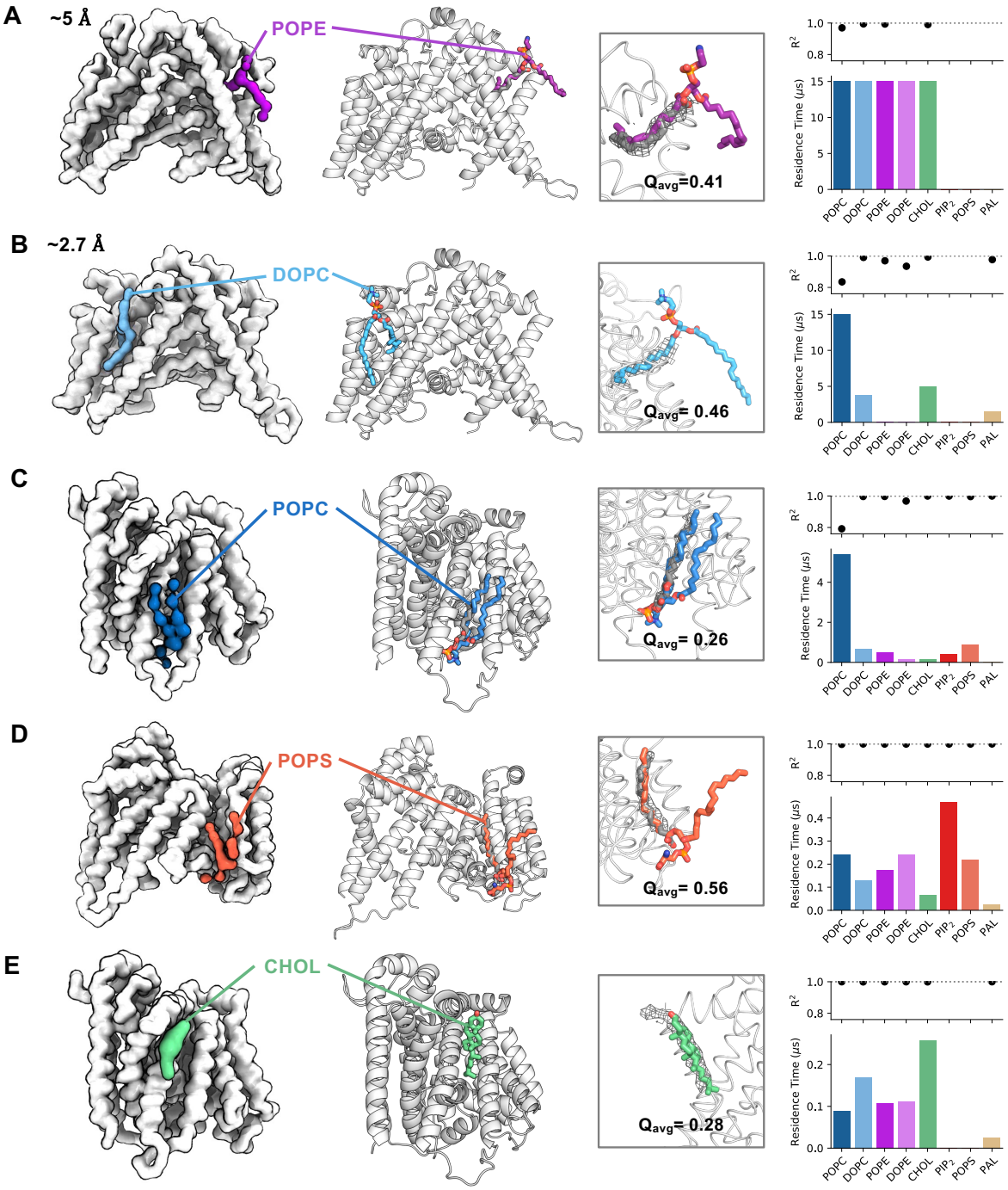


Figure 2.12: Comparison of lipid poses from simulations with cryo-EM densities.

Left: CG binding pose for a specified lipid bound at a site on HHAT. CG simulations (10x 15 μs) were initiated using coordinates from 5 Å **(A)** or 2.7 Å **(B-E)** cryo-EM maps indicating how PyLipID can be applied an multiple stages in map refinement and processing. Middle left: lipid pose from atomistic simulations of HHAT initiated via backmapping from CG simulations. Middle right: comparison of the simulation derived lipid binding pose with lipid-like density in the cryo-EM map. Average Q-scores (Q_{avg}) for the match between lipids poses and densities are indicated[180]. Right: comparison of binding site residence times and R^2 values for all lipids which bind to a given site, used to assess preferential binding of a lipid species to the site.

In addition to the applications described for HHAT, the pipeline can also be used to:

- Assess whether adjacent tail-like densities are likely to belong to the same or different binding sites.
- Assess how the properties of a site might promote preferential binding of one lipid type over another by examining the relative residence times of distinct lipids binding to the same site.
- Obtain a more complete picture of lipid interactions within the context of a native-like membranes. This may reveal transient lipid interaction sites which are less likely to survive the relatively harsh purification strategies used in cryo-EM.
- Assess differences in lipid binding properties compared with related detergent densities e.g for cholesterol-hemisuccinate (CHS), a widely used detergent.
- Assist differentiation of sterol-like *vs* lipid-like densities.
- Assess the relative contribution of a lipid head *vs* tail group to the interaction kinetics at a site.

To exemplify these additional applications, we applied LipIDens to three membrane proteins; the eukaryotic proton channel Otopetrin1 (OTOP1)[181], the *Escherichia coli* pentameric ligand-gated ion channel ELIC[182] and mechanosensitive channel of small conductance (MscS) also from *E. coli*[183]. All three proteins revealed putative bound lipids within the cryo-EM maps and were simulated in bilayers approximating the native membrane. Here we describe the results for OTOP1. Details of ELIC and MscS can be found within Ansell *et al.*[158].

Assignment of densities surrounding the OTOP1 homodimer was challenging owing to a resolution of 3.1-3.4 Å around the TMD. Three CHS molecules per subunit were modelled within a central pore formed by the dimer interface. The only peripheral density modelled in the deposited structure was a cholesterol molecule localised between the N- and C-domain of each monomer[181]. We used the LipIDens pipeline to assess which of the remaining 17 densities per OTOP1 monomer may also correspond to cholesterol and to rank the lipid binding preference at the remaining sites. Cholesterol binding poses matched the location of 4/17 of the additional lipid-like densities including, for which cholesterol was one of the highest ranked lipids (Fig.

2.13A). We observe exclusive binding of cholesterol to the site between OPOT1 N- and C-domains, validating modelling within the structure (Fig. 2.13B). Three sites showed preferential binding of anionic lipids (PIP_2/PS), including one site with a distinctive hooked density. The top ranked PIP_2 binding pose at this site was similarly curved, suggesting this density may correspond to PIP_2 extracted from the native bilayer (Fig. 2.13C-D). There were several smaller densities which we could not identify using the protocol. These may result from differences between the binding properties of lipids *vs* detergent or as a result of map processing.

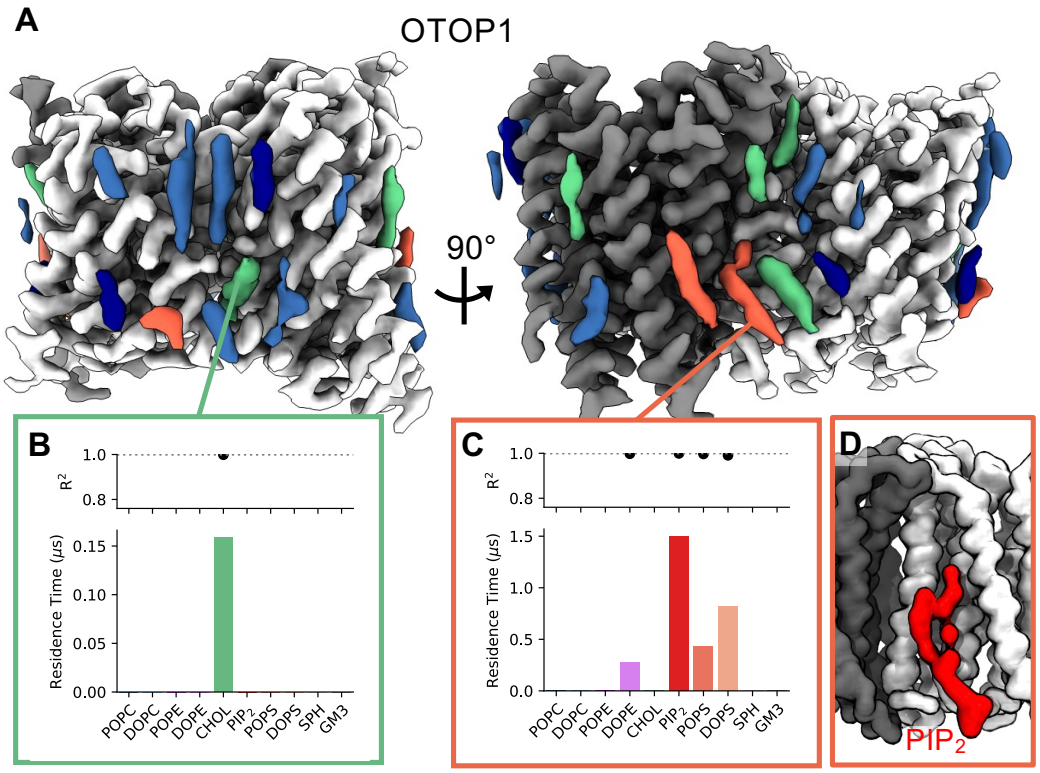


Figure 2.13: Interpreting lipid-like densities surrounding OTOP1.

The LipIDens protocol was used to characterise lipid-like densities surrounding the eukaryotic proton channel Otopetrin1 (OTOP1)[181]. **A)** Cryo-EM density map of OTOP1 dimer at ~3.1 Å (PDB: 6NF4). Dimer subunits are coloured white/grey. Lipid-like densities coloured according to whether cholesterol (green) or PIP_2/PS (red) were identified as binding to the site among the highest lipid residence times. Sites where other phospholipids bind are coloured light blue and sites that could not be identified by PyLipID are coloured dark blue. Residence times ard R^2 values for lipids binding to **B)** a cholesterol exclusive site between OTOP1 lobes and **C)** a kinked density at the dimer interface which preferentially binds anionic lipids. **D)** Top ranked CG PIP_2 binding pose for the density indicated in **C**.

The LipIDens method described here has been formalised into a pipeline for the wider structural community. Code to setup simulations, run and process lipid site analyses is freely available at github.com/TBGAnsell/LipIDens. Notebooks are provided to assist protocol implementation and automation. It is hoped that this pipeline will provide a framework for systematic integration of simulation and structural methodologies for membrane protein structure interpretation.

2.3 Affinities of Protein-Lipid Interactions

Protein-lipid binding events can be quantified via kinetic (as described in the previous section) or thermodynamic measures. Current approaches for obtaining the free energy or dissociation constant (K_d) of a protein-lipid interaction can be grouped into biased or unbiased sampling methods[184]. Biased sampling methods can be used to obtain rare events more readily (such as protein conformation changes or ligand unbinding) that are unlikely to be captured via long equilibrium simulations[185]. They rely on careful selection of appropriate reaction coordinates and force constants and thus may lead to simulation falsehoods if not handled competently. Unbiased simulations offer an alternate way to explore the energetic landscape[186,187] but are limited by the ability to sample a sufficient number of events of interest, which may not always be possible for a given system and/or finite computational resource. Thus, biased sampling methods have been used more frequently to obtain protein-lipid interaction free energies however, as computational technologies advance a renewed-interest in equilibrium methods is justified. The following sections summarise the work described in Ansell *et al*., *JCTC*.

2.3.1 Biased sampling methods

Steered MD simulations:

In steered MD simulations a force constant is applied to an atom selection (such as a lipid) along a reaction coordinate. Reaction coordinates are often linear, such as a distance vector between two groups or along an axis[188], but may take other forms,

such as the RMSD between two conformations[189]. Steered MD may be performed using constant force or constant velocity pulls. In the former, the centre of mass of the atom selection is pulled in the direction of the reaction coordinate via application of a constant force. In constant velocity steered MD a harmonic potential is applied between the centre of mass of the atom selection and a dummy atom which it iterated progressively along the reaction coordinate with time.

Potential of mean force calculations:

The potential of mean force (PMF) $W(r)$ refers to the free energy landscape along a reaction coordinate (r) (Equation 2.9). The PMF can be obtained using umbrella-sampling simulations initiated from snapshots obtained from e.g. a steered MD simulation. The potential energy ($U(r)$) is described by deviation from the unbiased state ($U(r')$) by a biasing potential ($U_b(r)$) (Equation 2.10). An umbrella biasing potential ($U_b(r)$) (Equation 2.11) is used to restrain the centre of mass of the atom selection within a 'window' i along the reaction coordinate during the umbrella-sampling simulation. The force constant k_i is applied uniformly across windows and r_0 is the minimum value in window i.

[2.9] $$W(r) = -RTln(p(r)) + C$$

[2.10] $$U(r) = U(r') + U_b(r)$$

[2.11] $$U_b(r_i) = \frac{k_i}{2}\,(r_i - r_0)^2$$

Windows should be spaced to sufficiently overlap such that the PMF is not influenced by gaps along the profile. This thesis employs a window spacing of ~0.05 nm, previously shown to produce overlap between the histograms of at least 3 windows for protein-lipid CG PMF calculations[188]. Umbrella sampling windows were simulated for 1-3 μs until convergence, and the *pmf.py* tool was used to assist setup and analysis of PMF calculations[188].

After calculating the local unbiased potential for each window, the global PMF profile can be obtained via the weighted histogram analysis method (WHAM)[190]. WHAM can be used to estimate the statistical uncertainty associated with the probability

distribution of states in the histogram of umbrella windows (Equation 2.12), where n_i is the number of bins in a histogram h_i and β is given by the inverse boltzman constant k_B at a temperature T ($\beta = 1/k_B T$). g_i is the statistical inefficiency ($g_i = 1 + 2\tau_i$) of each window in term of the integrated autocorrelation time τ_i.

$$[2.12] \qquad p(r) = \frac{\sum_i \ g_i^{-1} \, h_i(r)}{\sum_i \ n_i g_i^{-1} \, e^{\beta f - \beta U(r)}}$$

In practice the PMF free energy and statistical error are obtained via Bayesian bootstrapping to resample the probability distributions of windows[162]. This thesis bootstraps PMF data 200 or 2000 times.

Free energy perturbation and absolute binding free energy calculations:

In a free energy perturbation (FEP) pathway state i is gradually exchanged for state j through a series of alchemical λ windows to increase the phase space overlap[191]. The free energy difference between states is given by the Helmholtz function (Equation 2.13), where k_B is the Boltzmann constant, V is the phase space volume and $U(\vec{q})$ is the potential energy with respect to the coordinates and momenta ($\vec{q}$). When intermediate states are present $\Delta A_{i,j}$ is given by summing the free energy difference between substates (K) (Equation 2.14). For convenience, the transition across phase states is often described in terms of λ whereby $\lambda = 0$ and $\lambda = 1$ correspond to states i and j along a reaction coordinate[129].

$$[2.13] \qquad \Delta A_{i,j} = \ -k_B T ln \frac{\int V_j e^{\frac{-U_j(\vec{q})}{k_B T}} d\vec{q}}{\int V_i e^{\frac{-U_i(\vec{q})}{k_B T}} d\vec{q}}$$

$$[2.14] \qquad \Delta A_{i,j} = \ \sum_{i=1}^{K-1} \Delta A_{i,i+1}$$

For protein-lipid systems the free energy difference can be computed by e.g. comparing the free energy associated with gradual modulation of PIP_2 (state i) to PC (state j) when bound to the protein with that in the bilayer (or water for aqueous ligands)[188] (Equation 2.15).

$$[2.15] \qquad \Delta\Delta G = \Delta G^{i}_{free \rightarrow bound} - \ \Delta G^{j}_{free \rightarrow bound} = \Delta G^{free}_{i \rightarrow j} - \Delta G^{bound}_{i \rightarrow j}$$

To transform the lipid through chemical space the electrostatic and van der Waals interactions are scaled between those in state i and j. Scaling of each interaction type is often done sequentially, and a soft-core potential is used for the Lennard Jones potential[192].

Absolute binding free energy (ABFE) calculations are a variant of FEP calculations whereby a molecule is completely alchemically decoupled from a site. In both FEP and ABFE calculations, additional restraints may be needed to prevent the diffusion of ligands away from sites at high λ values[193].

Metadynamics simulations:

In metadynamics simulations the free energy landscape is explored by addition of a history-dependant bias potential to gradually 'fill up' the energy landscape. The biasing potential (V_G) at time t is described by Equation 2.16. Here ω and δs are the height and width of the applied gaussian respectively, $s(x)$ is the collective variable space, s_t is the value of the collective variable at t and t_G is the time frequency at which gaussians are added[185,194]. Thus, addition of a history-dependant bias enables escape from local minima and exploration of the underlying free energy landscape.

$$[2.16] \qquad V_G(s(x),t) = \sum_{\substack{t'=t_G,t_{2G},t_{3G}\ldots \\ t'<t}} \omega e^{\frac{-(s(x)-s_{t'})^2}{2\delta s^2}}$$

At extended time periods the free energy ($F(s)$) surface should be approximated by the applied bias potential i.e. $\lim_{t\to\infty} V_G(s,t) \sim -F(s)$. In well-tempered metadynamics, the height of the applied gaussian is scaled such that progressively smaller biasing perturbations are made as energetic wells fill. This reduces fluctuation in the biasing potential when $V_G \sim -F(s)$, smoothing landscape convergence[195].

2.3.2 Unbiased sampling methods

A key challenge in the calculation of protein-ligand affinities from equilibrium simulations relates to sampling enough binding and unbinding events. Equilibrium MD

simulations have been applied in a subset of examples to characterise protein-lipid interactions from both atomistic and CG simulations[196–198]. In addition, recent extensions to the Martini 3 forcefield have enabled the study of protein-drug affinities from equilibrium simulations[154,199].

Density based methods:

All published equilibrium methods for obtaining protein-lipid binding affinities employ a density-based approach[196–198]. This method has been used to study cholesterol interactions with Class A GPCRs (A_{2A} and β_2-AR), whereby the average 2D density of cholesterol in the xy plane is calculated across the simulation[196,197]. The free energy can be calculated directly from the probability of lipid binding at a given site relative to in bulk (Equation 2.17).

$$[2.17] \qquad \Delta G = -RTln\left(\frac{p_{site}}{p_{bulk}}\right)$$

This method can also be employed without dimensional averaging across z to obtain 3D lipid densities and their associated free energies[71]. This is useful for cases when two or more lipid binding sites are located in close xy proximity but in opposing leaflets, to avoid amalgamation of multiple site densities in 2D.

A recently developed variation of this method, the 'density-threshold' approach has been used to study interactions of the nicotinic acetyl-choline receptor (nAChR) with the neuronal membrane environment[198]. The probability distribution of finding n lipid beads within the bulk ($p_{bulk}(n)$) is calculated for each lipid type. The peak of the $p_{bulk}(n)$ distribution, n_{peak}, is compared with the probability distribution of finding n lipids beads within the site ($p_{site}(n)$). By integrating $p_{site}(n)$ either side of n_{peak}, the free energy can be calculated (Equation 2.18) from the ratio of the integrals ($P_<$ and $P_>$).

$$[2.18] \quad P_< = \sum_{n \leq n_{peak}} p_{site}(n) \quad : \quad P_> = \sum_{n > n_{peak}} p_{site}(n) \quad : \quad \Delta G = -RTln\left(\frac{P_>}{P_<}\right)$$

Density-based methods are usually applied to study bilayers of a given % lipid composition. While, in theory, the p_{site}/p_{bulk} ratio should be independent of % lipid

composition we found the method to be somewhat subject to the overall lipid %. We attribute this sensitivity to the quality of site sampling at lower % cholesterols and therefore care should be taken to select a suitable % lipid composition and to run simulations until converged.

Binding saturation method:

Experimental determination of protein-lipid affinities remains challenging and there is a paucity of quantitative biophysical data for such interactions. Successful approaches have used labelled lipid analogues to monitor binding to labelled membrane proteins via fluorescent or bioluminescence approaches[200,201]. Dissociation constants (K_ds) can be obtained from the binding saturation curves (Equation 2.19) produced by monitoring the change in FRET/BRET signal over increasing concentrations of the lipid-analogue (L)[200].

$$[2.19] \qquad Saturation\ (\%) = \frac{B_{max}[L]_{free}}{K_d + [L]_{free}} \quad : \quad [L]_{free} \sim [L]_{total}$$

We sought to assess whether a similar approach could be applied computationally to derive protein-cholesterol affinities. An outline of this novel method follows (for further details see Ansell *et al.*, *JCTC*). Equilibrium CG MD simulations are performed in bilayers composed of increasing concentrations of cholesterol (1, 2.5, 5, 10, 15, 30 and 40%). Binding sites are identified using PyLipID and the mean occupancy of six site residues conserved over all cholesterol concentrations is calculated for each site. Binding saturation curves are constructed by plotting the mean site occupancy as a function of free cholesterol concentration ($[CHOL]_{free}$) and K_d^{app}s are derived by fitting to Equation 2.20.

$$[2.20] \qquad Occupancy\ (\%) = \frac{B_{max}^{app}[CHOL]_{free}}{K_d^{app} + [CHOL]_{free}}$$

$$[2.21] \qquad [CHOL]_{free} = \frac{nCHOL_{free}}{nLipids}$$

Unlike in experimental approaches where the concentration of free ligand is approximated to the concentration of total ligand, in simulations it is possible to

calculate the free ligand concentration, defined as the mean number of cholesterols > 0.8 nm from the protein surface as a fraction of the total number of lipids (Equation 2.21).

This method is exemplified by application to the ATP-dependant pump P-glycoprotein (P-gp) (see Ansell *et al*., *JCTC* for other protein examples). Equilibrium MD simulations in bilayers of increasing cholesterol concentration were performed and PyLipID was used to identify two cholesterol sites between TM10/TM12 (Site A) and TM7/8 (Site B) which had been previously observed in CG simulations[202] (Fig. 2.14A). Saturable, non-linear binding curves were obtained for both Sites A and B (Fig. 2.14B) which had distinct affinities (Site A K_d^{app}: 0.8 ± 0.2 %, Site B K_d^{app}: 16.0 ± 0.8 %). In contract a linear increase in occupancy (gradient ~1) was observed for non-specific interactions (defined as the mean occupancy of residues which had an occupancy of 30-50% in bilayers containing 40% cholesterol). The relative order of site affinities obtained from binding saturation curves agreed with those reported in other studies[202]. Thus, computational binding saturation curves derived from equilibrium MD simulations can be used as an alternative method to obtain site affinities. The novel method mimics protocols applied experimentally and can be readily adapted to other ligands provided a sufficient number of binding/unbinding events can be observed. The method also circumvents approximations of micro and macro dissociation constants which can be challenging to disentangle experimentally.

2.3.3 Comparing methods for obtaining protein-lipid affinities

Selecting which of the available free energy methods to apply is dependent on the protein and/or lipid under study. In general, electrostatically driven interactions such as for PIP_2 tend to have longer residence times and thus equilibrium methods are less suitable due to a reduced frequency of association/dissociation events within the simulation timescale. Thus, biased sampling methods are more frequently applied to study charge driven lipid affinities. A comparison of PMF, FEP and metadynamics approaches to obtain affinities of PIP_2 interactions with the ion channel Kir2.2 or CDL interactions with the transporter proteins AAC or LeuT, revealed these methods to be in broad quantitative agreement[188]. However, the computational cost was greatest for

the metadynamics simulations which required approximately 3-4x more simulation time than PMF calculations and 4-5x more time compared to FEP/ABFE calculations to reach convergence[188]. Thus, FEB/ABFE or PMF calculations may be a more suitable and cost-effective choice of biased sampling method for obtaining protein-lipids affinities[184].

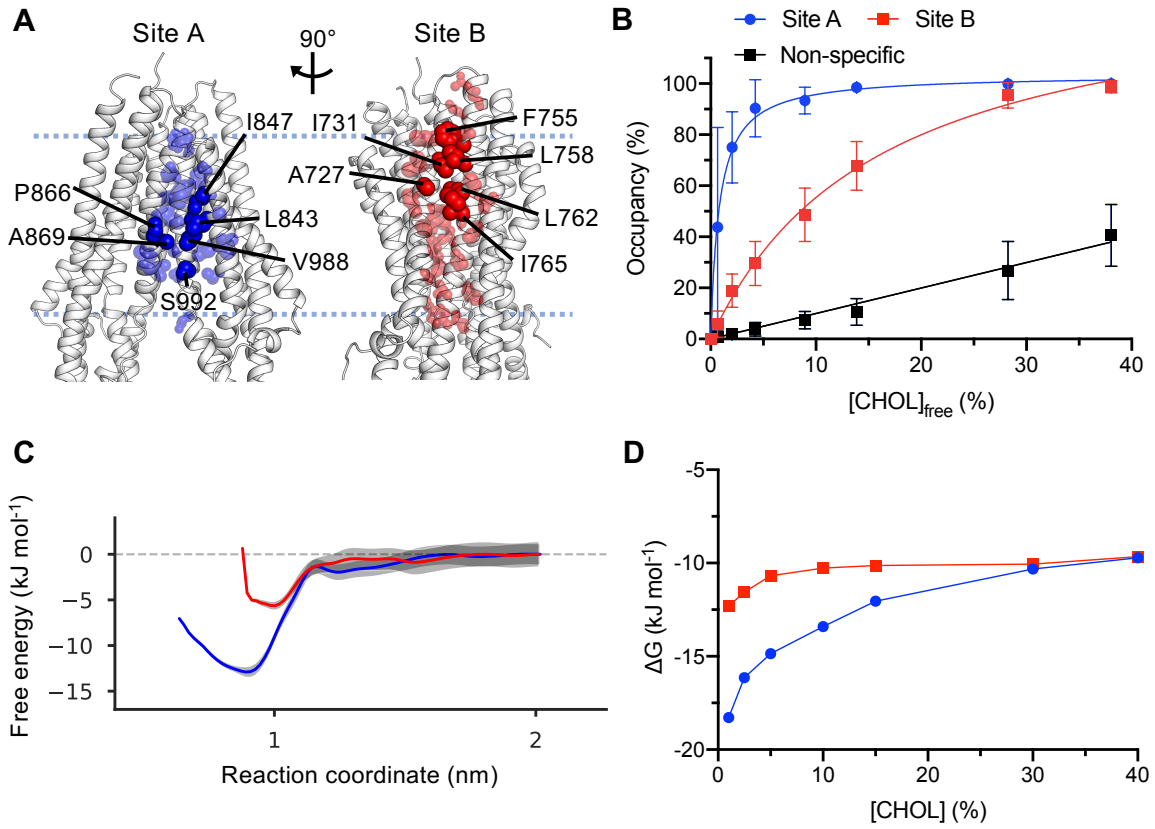


Figure 2.14: Cholesterol binding sites on P-gp.

A) Cholesterol binding site A (blue) and site B (red) on P-glycoprotein (P-gp), identified using PyLipID from equilibrium CG simulations (5 x 5 μs) in bilayers containing 40% cholesterol. The 6 residues in each site used to calculate site occupancies are opaque. Other residues in each site are transparent. **B)** Binding saturation curves for cholesterol binding to sites A/B. Site occupancy is defined as the mean occupancy of the 6 site residues **(A)** across simulations in bilayers comprised of increasing free cholesterol concentrations (Equation 2.20). Non-specific interactions (black) were derived from the mean occupancies of those residues which had an occupancy of 30-50% in bilayers containing 40% cholesterol. Error bars indicate standard deviations. **C)** PMF profile for cholesterol at sites A/B in bilayers containing 30% cholesterol. Bootstrapping errors are shown in grey. **D)** Free energies of binding derived from 3D densities for cholesterol binding to a site compared to in the bulk membrane (Equation 2.17). Figure adapted from Ansell *et al* [71].

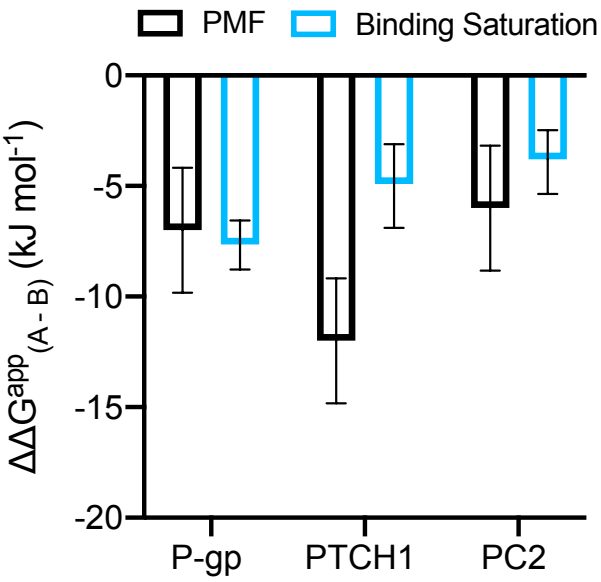


Figure 2.15: Comparison of apparent free energies for cholesterol from biased and equilibrium simulations.

Comparison of apparent free energy differences ($\Delta\Delta G^{app} = \Delta G^{app}{}_{SiteA} - \Delta G^{app}{}_{SiteB}$, Equation 2.22) for cholesterol binding to P-glycoprotein (P-gp), Patched1 (PTCH1) and polycystin-2 (PC2) derived from PMF calculations (black) or via the binding saturation method (blue). For the binding saturation method, apparent free energies were calculated from the obtained apparent dissociation constants ($\Delta G^{app} = -RTln(K_d{}^{app})$). Since the PMFs were performed independently, errors were calculated in quadrature ($total\ error = \sqrt{A_{err}^2 + B_{err}^2}$). Figure reproduced from Ansell *et al* [71].

An additional consideration arises from the fact that biased sampling methods can require more system specific parameters and user input than equilibrium methods, which may lead to simulation falsehoods. For example, in metadynamics simulations of Kir2.2 the protein was restrained in a conformation which was tilted by a mere 5° with respect to the bilayer normal. This resulted in an uneven binding free energy for PIP_2 across the four chains and a shift in PIP_2 binding to a secondary site in one of the subunits, highlighting the sensitivity of the method to input bias[188]. Moreover, in PMF calculations the window spacing must be carefully selected (and often refilled/supplemented after post-processing) to maximise histogram overlap and avoid under-sampled gaps in the PMF profile. Thus, the level of 'fine tuning' required by biased sampling methods limits their applicability to high throughput pipelines[203].

Sampling of more transient events, as can be observed for cholesterol binding in CG simulations, are suitable for study using equilibrium methods. We compared affinities obtained from binding saturation and density-based equilibrium methods with PMF calculations for cholesterol binding to P-gp (Fig. 2.14B-D) (see Ansell *et al*, *JCTC* for other examples). The relative order of site affinities was in agreement for all methods. The density-based method showed some sensitivity to the overall lipid concentration and/or the degree of site sampling at low concentrations. The binding saturation method circumvents this limitation because K_d^{app}s are calculated using data from all lipid concentrations. Furthermore, comparison of the free energy difference between sites for P-gp and other proteins (the transporter/receptor Patched1 (PTCH1) and the ion channel polycystin-2 (PC2)) revealed similar $\Delta\Delta G^{app}$s (Equation 2.22, Fig. 2.15). The magnitude of $\Delta\Delta G^{app}$ agreed for P-gp and PC2 with $\Delta\Delta G^{app}$s between -4 to -8 kJ mol^{-1}. The magnitude of $\Delta\Delta G^{app}$ for PTCH1 was larger for the PMFs, which we attribute to a PMF free energy value for Site A, much larger than has been observed for cholesterol binding to other membrane proteins via PMF calculations[184].

$$[2.22] \qquad \Delta\Delta G^{app} = \Delta G^{app}_{SiteA} - \Delta G^{app}_{SiteB} \quad : \quad \Delta G^{app}_{Site} = -RTln(K_d^{app})$$

An additional advantage of the binding saturation method is the ability to calculate the affinities of multiple sites from the same simulation dataset for no additional cost. To demonstrate this, we simultaneously calculated the affinities of 10 sites on the serotonin receptor (5-HT$_{1A}$) guided by the location of densities within the cryo-EM structure[204] (Fig. 2.16). Binding saturation curves were observed for all sites, validating modelling within the structure. The total computational cost of obtaining all site affinities was 175 μs (5 x 5 μs x 7 lipid conditions). The cost of performing an equivalent number of PMF calculations would be ~4x (~65 x 1 μs x 10 sites). Furthermore, we found the binding saturation method to require as few as 2 replicates and 3 μs per simulation required to reach K_d^{app} convergence (Fig. 2.17). Thus, the total simulation time could be reduced to 21-42 μs and still yield quantitative differences between site affinities. As the availability and speed of compute resources continues to advance, determination of protein-lipid affinities by equilibrium methods may see a revival, particularly given the recent interest in highly automated structural pipelines[127,158,203].

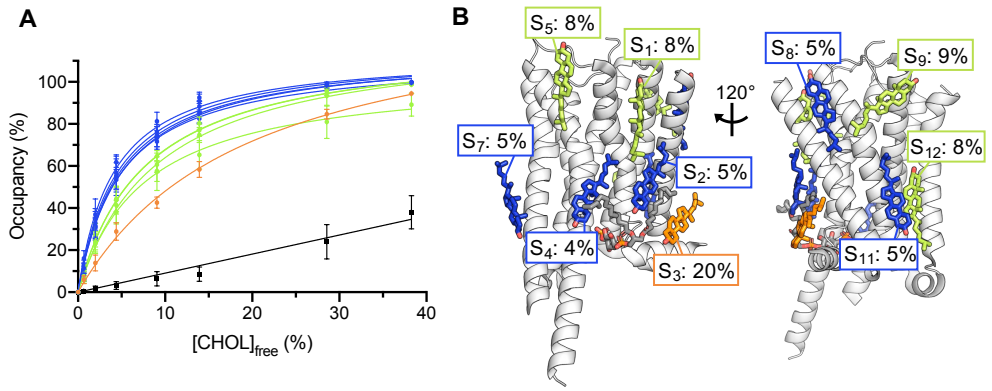


Figure 2.16: The binding saturation method applied to multiple sites on 5-HT$_{1A}$.

A) Binding saturation curves for cholesterol binding to 10 sites on 5-HT$_{1A}$[204]. Site occupancies were obtained from the mean occupancy of 6 sites residues in proximity to the modelled cholesterols in PDB: 7E2X. Error bars indicate the standard deviation. Sites are coloured according to the calculated apparent dissociation constants (K_d^{app}) ('strong', blue; 'moderate', lime; 'medium', orange). **B)** Location of cholesterol sires on 5-HT1A, coloured by calculated K_d^{app} values. Numbers correspond to densities in the structural paper[204]. Figure reproduced from Ansell *et al.,* [71].

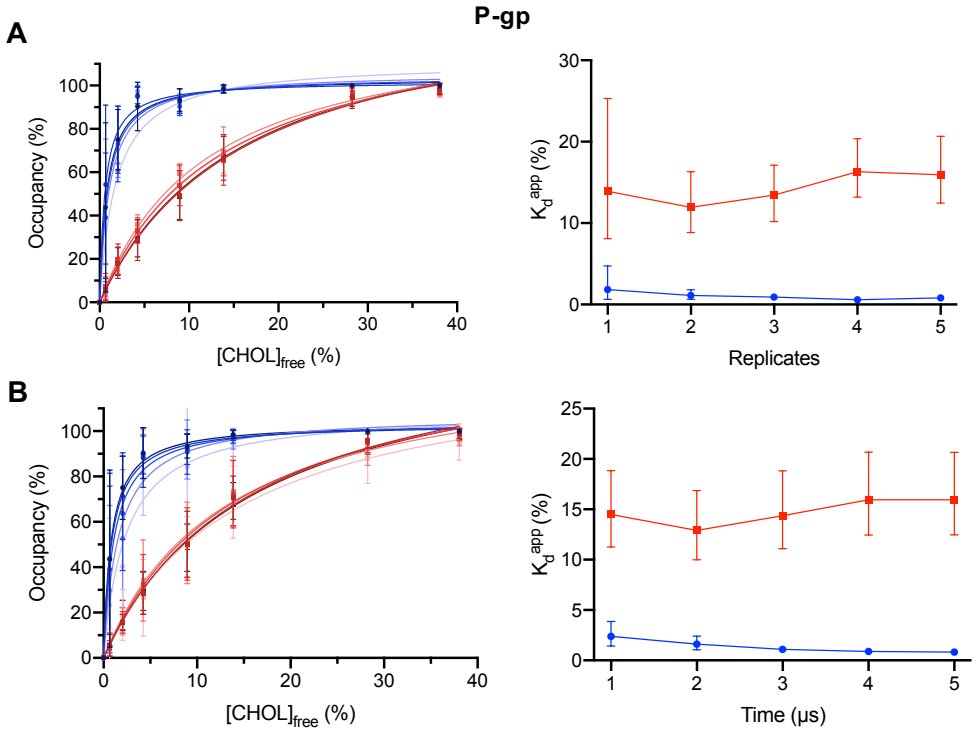


Figure 2.17: Convergence of site affinities from binding saturation curves.

Binding saturation curves (left) and K_d^{app} plots (right) as a function of the number of replicate simulations **(A)** and simulation length **(B)**. Site A (blue) and B (red) occupancies for cholesterol interactions with P-glycoprotein (P-gp) are coloured from light (fewer replicates/shorter simulation length) to dark (more replicates/longer simulation length analysed). Figure adapted from Ansell *et al* [71].

3. Patched

"'And I was sensible of nothing but the vigorous poetry of action."
D. H. Lawrence

3. Patched 56

3.1 Introduction 57

3.2 Methods 60
3.2.1 Coarse-grained simulations 61
3.2.2 Constructs used in atomistic simulations 62
3.2.3 Atomistic simulations 63
3.2.4 Application of a membrane potential 63
3.2.5 *In silico* mutation screen 64
3.2.6 Atomistic simulations of Dispatched1 64
3.2.7 Potential of mean force calculations 64
3.2.8 ABFE Calculations 67
3.2.9 FEP Calculations 68

3.3 Results and Discussion 69
3.3.1 SHH-cholesterol binds the PTCH1 ECD 69
3.3.2 Collapse of tunnels throughout PTCH1 by SHH-cholesterol 74
3.3.3 Probing the energetics of cholesterol transport pathways 76
3.3.4 Investigating the behaviour of ions in PTCH1 81
3.3.5 The hydrophobic cap 84
3.3.6 Insights from Dispatched1 86

3.4 Conclusion 89
3.4.1 The SHH-cholesterol inhibits PTCH1 89
3.4.2 Models for PTCH1 function 90

This chapter contains data also published in:

Rudolf, A.F., Kinnebrew, M., Kowatsch, C., **Ansell, T.B**., El Omari, K., Bishop, B., Pardon, E., Schwab, R.A., Malinauskas, T., Qian, M., Duman, R., Covey, D.F., Steyaert, J., Wagner, A., Sansom, M.S.P., Rohatgi, R., Siebold, C. (2019). The morphogen Sonic hedgehog inhibits its receptor Patched by a pincer grasp mechanism. *Nat. Chem. Biology,* 15, 975-982

Ansell, T.B., Corey, R.A., Viti, L.V., Kinnebrew, M., Rohatgi, R., Siebold, C., Sansom, M.S.P. (2022). The Energetics and Ion Coupling of Cholesterol Transport through Patched1. *Sci. Advances, Manuscript in preparation*

3.1 Introduction

Patched1 (PTCH1) is the primary tumour suppressor protein in the eukaryotic HH signalling pathway[16,17]. In the absence of Sonic HH (SHH) ligand, PTCH1 inhibits the downstream Class F GPCR Smoothened (SMO) via an uncertain mechanism. SHH binds to and inhibits PTCH1, enabling SMO activation and initiation of downstream signalling events (Fig. 3.1)[2]. Thus, PTCH1 is the linchpin protein between two inhibitory reaction steps, whereby the respective rates of each step are inversely related and control the level of overall HH signalling.

PTCH1 is a pseudo-symmetric membrane protein, which shares architecture with members of the resistance-nodulation-division (RND) family, whose members include prokaryotic multidrug efflux proteins such as AcrB and eukaryotic cholesterol transporters such as Niemann-Pick disease type-C1 (NPC1)[205]. PTCH1 also shares TMD homology with Dispatched1 (DISP1) which functions upstream in HH signalling to release SHH from cells. PTCH1 has 12 transmembrane helices (TM1-12) and two large extracellular domains (ECD1/2) positioned between TM1/2 and TM7/8 respectively[122,206]. PTCH1 also has an intracellular domain (ICD) between pseudo-symmetric units (TM6-ICD-TM7) however this has not been observed structurally and removal does not affect PTCH1 function[206].

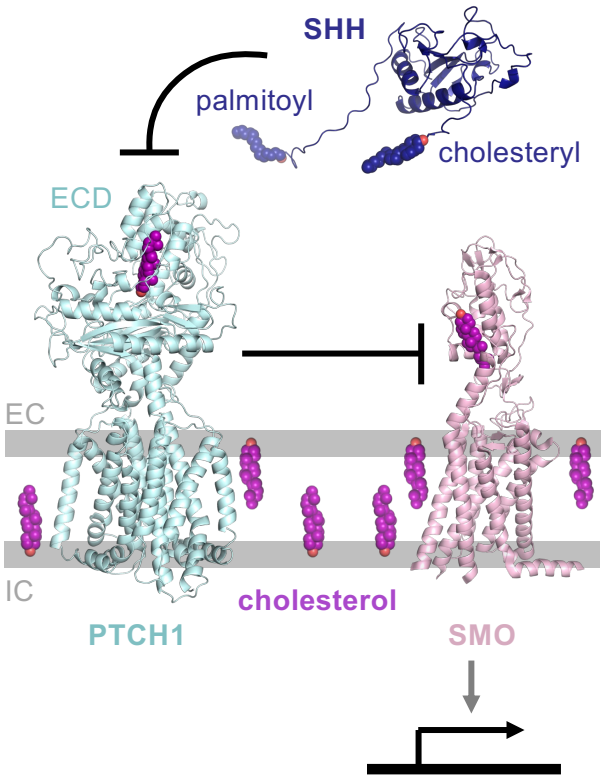


Figure 3.1: Patched1 (PTCH1) function.

In the absence of Sonic Hedgehog (SHH), PTCH1 inhibits Smoothened (SMO) to prevent downstream signalling. SHH binds and inhibits PTCH1, enabling SMO activation.

The mechanism of PTCH1 mediated SMO inhibition is not known, however biochemical and structural studies offer tantalising clues as to how PTCH1 may function. Over two decades ago a mechanism whereby PTCH1 inhibits SMO via direct interaction was revoked with the discovery that sub-stoichiometric quantities of PTCH1 could inhibit SMO, suggestive of a molecular relay mechanism[53]. This proposal is further supported by observations of distinct subcellular localisations of PTCH1 and SMO[15,40]. When SHH in not present PTCH1 localises to the base of the primary cilium in a skirt-like region formed by invagination of the membrane[207]. Conversely SMO localises within the plasma membrane and/or intracellular vesicles. SHH binding to PTCH1 leads to PTCH1 exclusion from the primary cilium, permitting SMO entry and propagation to the cilia tip where downstream signalling partners are localised[15,40].

One question resulting from these observations, is what molecule(s) PTCH1 sub-stoichiometrically regulates to inhibit SMO activation? Recent insights suggest the most likely identify of this species is the membrane sterol cholesterol[208]. The role of sterols in HH signalling has been demonstrated structurally by observations of cholesterol (and other oxysterols such as 24(S),25-expoxycholesterol) binding to the SMO cysteine-rich domain (CRD)[86,209,210] and TMD[94]. Further, cholesterol binding to

SMO is sufficient for activation even at low concentrations[22,47,49]. In contrast, low doses of other oxysterols could not activate SMO[22,47,49]. Additionally, CRISPR based assays show that disruption of cholesterol biosynthesis (but not other oxysterols) negatively regulates HH signalling[54] implicating cholesterol as the principal ligand under PTCH1 mediated HH pathway inhibition. Authors also suggest PTCH1 may regulate the abundance of accessible cholesterol available within the membrane to bind to and activate SMO[54]. Sterol-like densities have also been observed in PTCH1 structures[20,57,122,175,206,211–214]. These were located either surrounding the TMD[20,122,175,206,211,213,214] or interspersed along a tunnel between ECD1/2[20,57,122,212–214].

Several models have been proposed for how PTCH1 may regulate levels of accessible membrane cholesterol[208]. A proposed cholesterol transport function is supported by structural densities and homology to RND transporters. Prokaryotic RNDs are proposed to use e.g. the proton-motive-force to power antiport[215]. Three conserved anionic residues within the PTCH1 TMD are proposed to form a similar ion-coupling site, supported by mutational studies which disrupt PTCH1 function[53,57]. The homologue DISP1 uses Na^+ to drive SHH export[216,217]. PTCH1 has also been suggested to be regulated by Na^+ [218], although recent studies call the molecular identity of this cation into question[21,216]. The direction, mechanism and energetic requirements of a proposed transport mechanism are unclear. Alternatively, PTCH1 may promote re-distribution of accessible cholesterol to a sphingomyelin sequestered pool or intracellular protein/vesicular acceptor[21]. This model is supported by a long-standing conundrum in HH biology that cleavage of PTCH1 ECD2 fails to disrupt PTCH1 function[53], and by modulation of sphingomyelin levels which altered HH signalling[54].

Intriguingly the SHH ligand is also cholesterylated at its C-terminus in addition to an N-terminal palmitoyl moiety that constitutes the native bi-lipidated SHH morphogen[7,8]. The N-terminal palmitate moiety has been shown to contribute to PTCH1 inhibition[219] and was observed to insert into the PTCH1 ECD in structures[206,211]. The role of the C-terminal SHH-cholesterol was unknown until the work described within this thesis and in Rudolf *et al*.

This study describes experiments relating to two key aspects of PTCH1 function:

1) What is the role of the SHH C-terminal cholesterol molecule in PTCH1 inhibition?
2) How could PTCH1 function to inhibit SMO?

Here I detail how MD simulations contributed to understanding these queries, in collaboration with structural and biochemical studies from the Siebold and Rohatgi labs.

3.2 Methods

Equilibrium simulations within this chapter are summarised in Table T3.1.

Table T3.1: Summary of equilibrium PTCH1/DISP1 simulations.

Construct	PDB	CG/AT	Replicates x Time	Initial ligands/conditions
PTCH1-molA	6RVD	CG	5 x 10 µs	*apo*
PTCH1-molB	6RVD	CG	5 x 10 µs	*apo*
$PTCH1_{ECD1}$	6RTW	AT	5 x 300 ns	cholesterol
$PTCH1_{ECD1}$	6RTY	AT	5 x 300 ns	*apo*
$PTCH1_{ECD}$-molA	6RVD	AT	5 x 300 ns	'SHH-cholesterol'
$PTCH1_{ECD}$-molA	6RVD	AT	5 x 300 ns	*apo*
$PTCH1_{ECD}$-molB	6RVD	AT	5 x 300 ns	'free cholesterol'
$PTCH1_{ECD}$-molB	6RVD	AT	5 x 300 ns	*apo*
PTCH1	6DMY	AT	3 x 100 ns	Na^+ (Site 1)
PTCH1	6DMY	AT	3 x 50 ns	*apo* (in 0.15 M NaCl)
PTCH1	6DMY	AT	3 x 100 ns	*apo* (in 0.15 M KCl)
PTCH1	6DMY	AT	3 x 100 ns	Na^+ (Site 1), -100 mV membrane potential
PTCH1	6DMY	AT	3 x 100 ns	Na^+ (Site 1), -200 mV membrane potential
PTCH1-mutants	6DMY	AT	3 x 50 ns x 7 mutants	*apo* (in 0.15 M NaCl)
DISP1	7RPH	AT	3 x 100 ns	'3x Na^+'
DISP1	7RPH	AT	3 x 100 ns	*apo*
DISP1	7RPH	AT	3 x 100 ns	'2x Na^+'
DISP1	7RPH	AT	3 x 100 ns	'1x Na^+'
DISP1	7RPH	AT	3 x 100 ns	'0x Na^+'

3.2.1 Coarse-grained simulations

Simulations were performed using the GROMACS 5.1.4 simulation package (http://manual.gromacs.org/)[155]. PTCH1-molA and PTCH1-molB proteins were extracted from the rebuilt PTCH1–pShhNc (2:1) model (PDB: 6RVD)[220]. The missing TM6–TM7 loop (residues 613–730) is too long to be modelled accurately using conventional modelling software and so L608–C613 and P725–W730 were modelled and residues 613 and 725 were directly connected using Modeller9.20[221] to give an overall sequence (TM6)RRL-DIFCC//TKWT-LSS(TM7). Ten models were produced and the model with the lowest DOPE score was taken forward for each PTCH1 molecule. PTCH1-molA and PTCH1-molB were converted to CG resolution using the MARTINI 2.2 forcefield[149] with the ElNeDyn elastic network[222], a spring force constant of 500 kJ mol^{-1} nm^{-2} and a cut-off of 0.9 nm. CG PTCH1 structures were embedded in a symmetric POPC:cholesterol (3:1) membrane and solvated with MARTINI water[146] using *insane.py*[223] in a 15 × 15 × 17 nm^3 box. The transmembrane region was centred in the bilayer and positioned along the *z* axis according to the atomistic bilayer position determined by MemProtMD[203,224]. The system was adjusted to approximately 0.15 M NaCl and two rounds of steepest-descent energy minimization were performed. The system was equilibrated for 25 ns using NPT ensemble with all protein beads restrained. A second 100 ns equilibration was subsequently performed using NPT ensemble and restrained protein backbone beads. The PTCH1-molA and PTCH1-molB systems were simulated for 5 x 10 µs each, with a 20 fs timestep. The MARTINI 2.2 forcefield was used to describe all components[149]. A V-rescale thermostat[225] with a coupling constant τ_t = 1 ps was used to maintain temperature at 310 K. Pressure was maintained at 1 bar using the Parrinello–Rahman barostat[226], a coupling constant τ_p = 12 ps and compressibility of 3×10^{-4} bar^{-1}. Long-range interactions were switched off at 1.1 nm and modelled using the reaction-field method. Lennard–Jones interactions were cut off at 1.1 nm. The LINCS algorithm was used to constrain bonds to their equilibrium values[227].

3.2.2 Constructs used in atomistic simulations

$PTCH1_{ECD1}$ crystal structures with cholesterol-hemisuccinate (CHS) bound (1.9 Å) or *apo* (2.1 Å) were used as initial structures for simulation (PDB: 6RTW/6RTY)[220]. Nanobody-64 was removed, and cholesterol was precisely positioned in the ECD1 sterol-binding pocket to replace CHS where applicable. $PTCH1_{ECD1}$ was centred in a 7 nm^3 box.

In simulations of the full-length $PTCH1_{ECD}$, residues A119–D436 and R772–G1023 were extracted from the rebuilt PTCH1–pShhNc (2:1) model (PDB: 6RVD) with either the C-terminal 'SHH-cholesterol' moiety bound ($PTCH1_{ECD}$-molA), 'free cholesterol' bound ($PTCH1_{ECD}$-molB) or without cholesterol ($PTCH1_{ECD}$-molA/$PTCH1_{ECD}$-molB *apo*). $PTCH1_{ECD}$ systems were centred in a 9 nm^3 box.

The cryo-EM structure of PTCH1 with a putative ion density visible in the TMD (PDB: 6DMY)[122] was selected for use in atomistic simulations. SHH, ligands and metal ions were removed. A short linker was modelled between TM6 and TM7 using Modeller9.20[221] to give an overall sequence of (TM6)DRR-LDIFCC//TKWTLSSFAE-KHY(TM7), in accordance with the linker used in CG simulations of PTCH1 (PDB: 6RVD). A Na^+ ion was positioned according to the density in the centre of the TMD. The H++ server and propKa[228] were used to predict the pKa of titratable groups (Table T3.2), revealing at least two residues of the anionic triad to be protonated at pH 7 in the absence of bound Na^+, decreasing to one protonated residue when Na^+ was present. The central residue, E1095, was therefore protonated using the CHARMM-GUI PDB generator[229]. PTCH1 was embedded in a symmetric POPC:CHOL (3:1) membrane generated using the CHARMM-GUI bilayer builder[143].

For all systems, the CHARMM-GUI PDB generator[229] was used to rename atoms to match those compatible with the CHARMM-36 forcefield[131] and model disulphide bonds. Systems were solvated using TIP4P water[230] and neutralized using 0.15 M NaCl before steepest-descent energy minimization was performed. Systems were equilibrated in 2 x 5 ns NVT and NPT steps with restraints applied to the protein backbone.

Table T3.2: Predicted pKa of PTCH1 residues.

Residue	Predicted pKa		
	H++	H++ (Na^+ bound)	PropKa
D513	>12.00	>12.00	8.45
D514	6.26	<0.00	8.24
E1095	11.29	0.90	10.95

3.2.3 Atomistic simulations

Simulations were performed using the GROMACS 2018 simulation package[155]. $PTCH1_{ECD1}$ and $PTCH1_{ECD}$ simulations were run for 5 x 300 ns. Simulations of full length PTCH1 were run for 3 x 100 ns with Na^+ positioned at Site 1. 3 x 50 ns simulations of PTCH1 were also run without Na^+ positioned within the TMD and 3 x 100 ns simulations were performed where 0.15 M NaCl was switched to KCl. A 2 fs timestep was used. The CHARMM-36 forcefield was used to describe all components[131]. Periodic boundary conditions were applied. A Nosé–Hoover thermostat[231,232] with a coupling constant $\tau_t = 0.5$ ps was used to maintain temperature at 310 K. Pressure was maintained at 1 bar using the Parrinello–Rahman barostat[226], a coupling constant $\tau_p = 2$ ps and a compressibility of 4.5×10^{-5} bar^{-1}. The PME method was used to model long-range electrostatic interactions[233]. Van der Waals interactions were cut off at 1.2 nm. Bond lengths were constrained to the equilibrium values using the LINCS algorithm[227]. A dispersion correction was not applied. In simulations of the $PTCH1_{ECD}$ restraints were applied to residues A119–N120, S435–D436, R772–D773 and I1022–G1023 to mimic the effect of the TMD.

3.2.4 Application of a membrane potential

Atomistic simulations of the full length PTCH1 were also run in the presence of a -100 mV or -200 mV membrane voltage ($\Delta V = E.L_z$, E=-7.1429 or –14.286 mV, L_z=14 nm) for 3 x 100 ns each. Voltage is reported as inside relative to outside and was achieved using the constant electric field method[234].

3.2.5 *In silico* mutation screen

In silico point mutations of the PTCH1 construct used in atomistic simulations (PDB: 6DMY) were generated using *mutate_model.py* in Modeller9.20[221]. The following mutants were made: V510G, V510A, V510T, V510F, I1092A, L517C/P1125C and H1099C/S547C. CHARMM-GUI[229] was used to regenerate PTCH1 parameters, including the additional disulphide bonds (where applicable). Atomistic simulations (3 x 50 ns) of each construct were run identically to described above.

3.2.6 Atomistic simulations of Dispatched1

The DISP1 structure with 3x Na^+ bound was used in simulations (PDB: 7RPH, 'R conformation')[217]. Detergent ligands were removed, and the missing loop modelled using Modeller9.20[221]. CHARMM-GUI[229] was used to generate DISP1 parameters (separated into two chains to account for the Furin cleavage site) and the anionic triad residue D1049 was protonated for consistency with the PTCH1 simulations. DISP1 was simulated in the '3x Na^+' bound and *apo* states, identically to described above, for 3 x 100 ns each. The '2x Na^+' state was generated from the end snapshots of the '3x Na^+' state by removing the ion missing in an alternative 'T conformation' of DISP1 (PDB: 7RPI)[217] bound to two Na^+ ions. The '1x Na^+' and '0x Na^+' states were generated by an identical process of sequential ion removal from the end snapshots of previous simulations. The ion retained in the '1x Na^+' state was chosen based on the *apo* state simulations whereby an ion spontaneously bound to D572 in all replicates. Each subsequent ion state was also simulated for 3 x 100 ns.

3.2.7 Potential of mean force calculations

PMF-1: between the SSD and SBD and PMF-2: between the SSD and bulk membrane

Snapshots from CG simulations of PTCH1-molA and PTCH1-molB were selected for use in PMF calculations with the ROH bead (equivalent to the β_3-OH group) of the

SSD bound cholesterol either pointing towards the extracellular leaflet headgroups ('OH-up') or towards the bilayer midplane ('OH-down'). The PTCH1 SSD was aligned to the x axis to assist generation of the reaction coordinate and the surrounding membrane and solvent expanded to a box size of 20 x 20 x 19 nm^3 using gmx genconf and VMD[167]. Two round of steepest decent energy minimisation were performed to relax the equilibrated bilayer around PTCH1. For PMF-1a the 'SHH-cholesterol' or 'free cholesterol' molecules were positioned in PTCH1-molA and PTCH1-molB ECDs respectively according to the cryo-EM densities. PMF-1b was initiated with cholesterol bound at the base of the ECD, continuing from the final window in PMF-1a from PTCH1-molB. Since the 'free cholesterol' bound within PTCH1-molB was positioned with the ROH bead orientated towards the bilayer, a bash script was used to invert the cholesterol orientation in frames generated from the steered MD simulations to generate the PTCH1 'SHH-cholesterol' conformations. This was due to difficulties obtaining a reasonable steered MD simulation initiated from the final window of PMF-1a from PTCH1-molA and to standardise the reaction coordinate. For PMF-2 cholesterol was not present in the PTCH1 ECD.

PMF-3: between the bulk membrane and solvent

A 10 x 10 x 18 nm^3 POPC:CHOL (3:1) bilayer patch without protein was built using *insane.py*[223]. The bilayer was solvated using MARTINI water[146] and 0.15 M NaCl before two round of steepest decent energy minimisation. A cholesterol molecule at the centre of the bilayer was selected for use in PMF calculations.

PMF-4: between the SBD and solvent

The ECD of PTCH1-molA and PTCH1-molB (residues A119–D436 and R772–G1023) from the PTCH1-SHH (2:1) structure (PDB ID: 6RVD)[220] was extracted and coarse-grained. The 'SHH-cholesterol' and 'free cholesterol' molecules were positioned in the SBD of PTCH1-molA and PTCH1-molB respectively accordioning to their positions in the cryo-EM density. The PTCH1 ECDs were positioned at one end of a 9 x 9 x 18 nm^3 box with the SBD aligned to the z axis. The system was solvated with MARTINI water[146] and approximately 0.15 M NaCl followed by two rounds of steepest decent energy minimisation.

Execution and analysis of coarse-grained potential of mean force calculations

PMF calculations were carried out in accordance with a previously published method[188]. Steered MD simulations were used to pull the cholesterol of interest through the protein (PMF-1), away from the protein (PMF-2, PMF-4) or out of the membrane (PMF-3). The 1D reaction coordinate was generated by application of a distance dependant pulling force along specified axes between the COM of the cholesterol and the following selections: PMF-1a (I1092 BB bead), PMF-1b (F800 BB bead), PMF-2 (P504 BB bead), PMF-3 (COM bilayer) and PMF-4 (S331 BB bead). An umbrella pulling force of 1000 kJ mol^{-1} nm^{-2} and a pulling rate of 0.1 nm ns^{-1} were used in all cases with the exception of PMF-1b where the pulling rate was 10 nm ns^{-1} to prevent extraneous pulling of cholesterol into the aqueous solvent after egression from the ECD. Position restraints of 1000 kJ mol^{-1} nm^{-2} were applied to the following backbone beads to prevent protein rotation: PMF-1a/b, (A1088, A1157), PMF-2 (A1088, A1157) and PMF-4 (A182, A239). Frames were taken with spacing 0.05 nm along the reaction coordinate and subjected to 1-3 μs of simulation. For each window an umbrella pulling force of 1000 kJ mol^{-1} nm^{-2} was used to confine the cholesterol position along the reaction coordinate. Further details of the number of windows, simulations times and PMF convergence are given in Table T3.3 and Appendix Fig. S3.1. PMF profiles were constructed using the weighted histogram analysis method (WHAM) implemented in *gmx wham*[190]. The first 200 ns of each window was discarded as equilibration time and 2000 Bayesian bootstraps[162] were used to estimate the PMF error. Since PMFs 2-4 are independent, the total error of the indirect pathway was calculated as in Equation 3.1. Setup and analysis of PMFs was aided by use of the *pmf.py* tool developed by Dr Owen Vickery[188] (DOI:10.5281/zenodo.3592318).

$$[3.1] \qquad Error = \sqrt{\sum_{x=2-4} (PMF_x^{error})^2}$$

Table T3.3: Summary of PMF calculations.

PMF	Construct	Ligand	No. windows	Window length (μs)	Total time (μs)
1a	PTCH1-molA	'SHH cholesterol'	92	3	276
	PTCH1-molB	'free cholesterol'	99	2	198
1b	PTCH1-molA	'SHH cholesterol'	98	2	196
2	PTCH1-molA	'OH-up' cholesterol	65	1	65
	PTCH1-molB	'OH-up' cholesterol	56	1	56
	PTCH1-molB	'OH-down' cholesterol	65	1	65
3	None (bilayer only)	Membrane cholesterol	76	1	76
4	$PTCH1_{ECD}$-molA	'SHH cholesterol'	112	2	224
	$PTCH1_{ECD}$-molB	'free cholesterol'	101	2	202

3.2.8 ABFE Calculations

Absolute binding free energy (ABFE) calculations were performed by Dr. Robin Corey on the CG PTCH1 complex with a cholesterol molecule bound to the TMD SSD (just PTCH1-molA), the ECD SBD and a mid-point along PMF-1a at approximately 2.4 nm on the reaction coordinate. Initial poses reflect those used to seed the steered MD simulations as described above. In all cases, the systems were trimmed to *ca.* 15 x 15 x 15 nm^3 and for the TMD and mid-point systems all cholesterol apart from the target molecule were removed from the system, followed by 200 ns of equilibration.

ABFE calculations were essentially run as previously described[188] (Table T3.4). In brief, the target cholesterol is fully decoupled from the system along an alchemical coordinate, λ. The Lennard-Jones interactions of the cholesterol particles were switched off over 29 windows, using a step size of 0.05 for λ=0-0.6 and a step size of 0.25 for λ=0.6-1. A soft-core potential was used, with an α of 0.5 and a σ of 0.3. For each window, the systems were minimised using steepest descents before 3 x 150 ns production simulations, using a stochastic integrator at 323 K, with pressure held at 1

bar using the Parrinello-Rahman barostat[226], with a τ_p of 4 ps and a compressibility of 5 x 10^{-5} bar^{-1}.

Table T3.4: Summary of ABFE calculations.

Construct	Ligand	Replicates x No. windows	Window length (μs)	Total time (μs)
$PTCH1_{ECD}$-molA	'SHH cholesterol'	3 x 29	0.15	13.05
$PTCH1_{ECD}$-molB	'free cholesterol'	3 x 29	0.15	13.05
$PTCH1_{ECD}$-molA	'OH-up' cholesterol	3 x 29	0.15	13.05
$PTCH1_{ECD}$-molA	'SHH cholesterol' mid	3 x 29	0.15	13.05
$PTCH1_{ECD}$-molB	'free cholesterol' mid	3 x 29	0.15	13.05

Throughout the λ windows, flat-bottomed distance restraints of 10 kJ mol^{-1} nm^{-2} were applied between the cholesterol and its bound state to keep it in the binding site at high values of λ. The bound state was defined using two dummy beads whose position updated in relation to reference beads in the binding site: E274+S331 in the SBD site for both states, A936+Y1013/Y801 in the mid-point site for PTCH1-molA/B, and A478+V481 for the SSD pose. At low values of λ, the binding site itself restricts the cholesterol molecule, and the restraints have a negligible effect.

The computed energies were constructed into energy landscapes along λ using Multistate Bennett Acceptance Ratio (MBAR)[235] as implemented in alchemical analysis[191].

3.2.9 FEP Calculations

The end snapshot from one 100 ns atomistic simulations of PTCH1 bound Na^+ (above) was used as input for FEP calculations. Over 31 windows Na^+ was alchemically changed to K^+ by independently modifying the mass (λ windows 0-10) and VDW radii (λ windows 11-30) of the bound ion using a dual topology approach. Mass modification did not contribute significantly to the overall free energy and therefore a λ step size of 0.1 was used in windows 0-10 and a step size of 0.05 in windows 11-30. A soft-core parameter (σ = 0.3) was used for Lennard-Jones interactions. Each window was simulated for 50 ns using the protocol described above for atomistic simulations of PTCH1 with the exception that a stochastic dynamics integrator was used. Perturbation of Na^+ to K^+ in free solution was achieved using a single Na^+ in a 3 x 3 x

3 nm^3 TIP4P water box[230]. A single Cl^- counter ion was present. The solvent box was equilibrated for 20 ns before use in FEP calculations to allow water to equilibrate around the ion. Each window was simulated for 5 ns using the same protocol described for the PTCH1 bound Na^+ systems. 5 repeats of each FEP calculation were performed. 5 repeats of the inverse perturbations (K^+ to Na^+) in the bound and free states were also performed which gave free energy values of the same magnitude but inverse signs (Appendix Fig. S3.2A-B). Alchemical analysis[191] was used in calculation of the overall free energy values from λ windows. There was good agreement between analysis methods and so values calculated using MBAR[235] are reported as the free energy mean and standard deviation between repeats (Appendix Fig. S3.2C-H). The first 200 ps of each window were discarded to allow for equilibration.

3.3 Results and Discussion

3.3.1 SHH-cholesterol binds the PTCH1 ECD

Cryo-EM structures of PTCH1 reveal elongated densities within the ECD[20,57,122,212–214]. The highest resolution PTCH1 structure is at ~3.4 Å, hence limiting the unambiguous assignment of the sterol-like ECD densities. To improve structural insight, members of the Siebold group solved the structure of $PTCH1_{ECD1}$ in *apo* and cholesterol-hemisuccinate (CHS) bound states to 2.1 Å and 1.9 Å respectively and *apo* $PTCH1_{ECD2}$ to 2.2 Å using X-ray crystallography[220] (Fig. 3.2A). CHS binds within a hydrophobic pocket of the $PTCH1_{ECD1}$ structure we termed the sterol-binding domain (SBD) (Fig. 3.2A-B). The CHS ester linkage (at an equivalent position to the β_3-hydroxyl group of cholesterol) is located at the pocket entrance, coordinated by Y215 on a loop of the pocket. To probe whether cholesterol could be accommodated at the position of the bound CHS, we performed atomistic simulations of $PTCH1_{ECD1}$ bound to cholesterol or *apo*. Cholesterol binding stabilised the SBD as indicated by a reduced RMSD of ECD1 (Fig. 3.2C). In support of this, density for ECD1 loop regions were missing in the crystal structure of ECD1 in an *apo* state (PDB: 6RTX)[220]. This suggests cholesterol is favourably accommodated within the SBD at an equivalent location to the crystallographic CHS position.

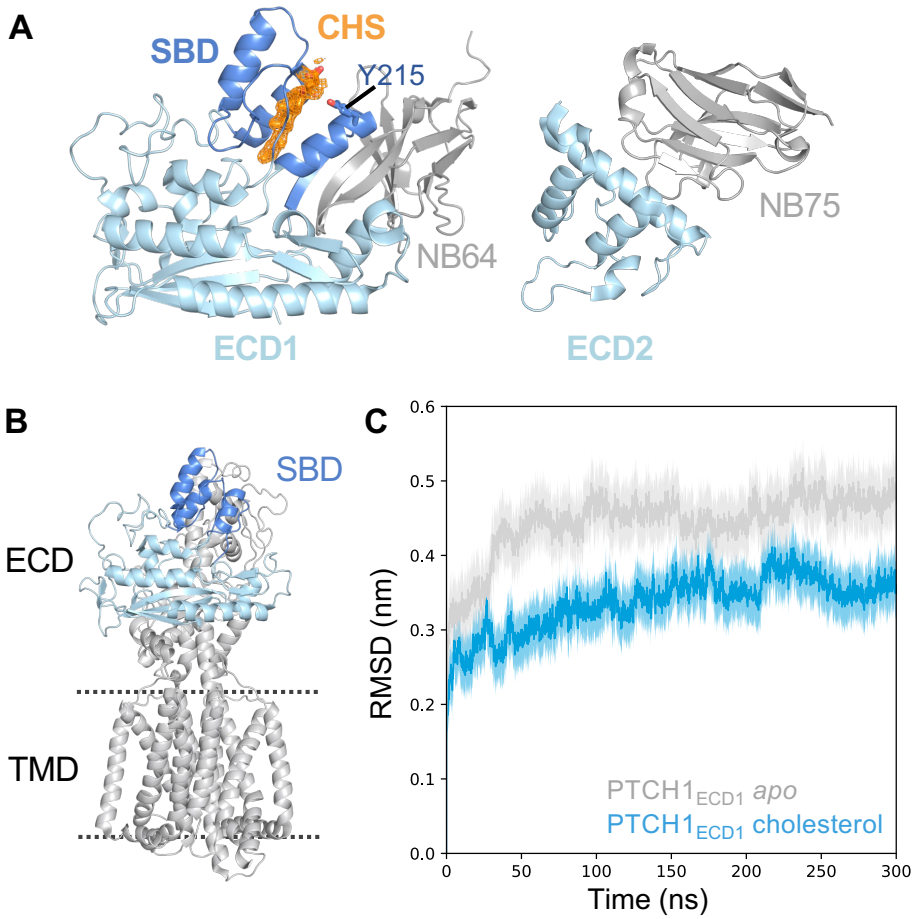


Figure 3.2: Structural and biochemical studies of PTCH1.

A) Crystal structures of the PTCH1 extracellular domains (ECD1 PDB: 6RTW, ECD2 PDB: 6RVC, light blue) bound to nanobodies (NB64 and NB75, grey), solved by members of the Siebold group[220]. Density corresponding to cholesterol hemisuccinate (CHS, orange) is shown within the sterol binding domain (SBD, blue) of ECD1. **B)** ECD1 overlayed with the structure of full length PTCH1 (PDB: 6DMY). **C)** Root mean square deviation (RMSD) of the SBD across 5 x 300 ns atomistic simulations of $PTCH1_{ECD1}$ in the presence (blue) or absence (grey) of cholesterol (defined as the square root of the average MSD across repeats, error bars indicate the standard deviation).

To further characterise the observed sterol binding pocket, members of the Siebold group performed isothermal titration calorimetry (ITC) whereby polyethylene glycol conjugated cholesterol (PEGylated-cholesterol) was added to a cyclised ECD1-ECD2 construct. PEGylated-cholesterol was used to improve the solubility of the lipid for use in ITC. The ITC data indicated binding of PEGylated-cholesterol to the $PTCH1_{ECD}$ with nanomolar affinity (K_d = 370 nM) (Appendix Fig. S3.3A). Mutation of residues lining

the SBD ablated PEGylated-cholesterol binding (M281Q: K_d not determinable) or reduced affinity to the micromolar range (A246M: K_d = 4.9 μM), thus supporting the crystallographically observed cholesterol binding pocket.

The high resolution $PTCH1_{ECD1}$ and $PTCH1_{ECD2}$ crystal structures were used to re-refine a recently published cryo-EM structure of two PTCH1 molecules bound to SHH^{211} (Fig. 3.3A). In the PTCH1-SHH (2:1) complex SHH engages with PTCH1 via distinct interfaces formed either by insertion of the N-terminal palmitate into the PTCH1 ECD (termed PTCH1-molA) or via a metal binding interface (termed PTCH1-molB). Additional sterol-like densities could be observed within the SBD of both PTCH1-molA and PTCH1-molB which were modelled as cholesterol. Comparison of the bound $PTCH1_{ECD1}$ CHS with the modelled cholesterols showed good agreement with the pose in the PTCH1-molA conformation. In PTCH1-molB the cholesterol was located deeper in the pocket and rotated by 180° to position the β_3-hydroxyl towards the centre of the ECD (Fig. 3.3B). Surprisingly additional density could be observed extending between the globular core of SHH and cholesterol bound within the SBD of PTCH1-molA, which we modelled as the SHH C-terminus (Fig. 3.3C). Thus, the refined cryo-EM structure suggests the SHH-cholesterol moiety participates in direct interaction with PTCH1 via insertion into the SBD and may also contribute to PTCH1 inhibition.

To examine whether the SHH-cholesterol moiety contributes to PTCH1 inhibition, members of the Rohatgi group performed HH pathway signalling assays in embryonic mouse cells. Reverse transcriptase quantitative-polymerase chain reaction (RT-qPCR) was used to measure the abundance of *Gli1* mRNA, a direct HH gene target as the concentration of various SHH constructs was increased[220] (Appendix Fig. S3.3B-D). Compared to dually lipidated WT SHH (pSHHc), PTCH1 inhibition was reduced when the C-terminal SHH linker (bound to cholesterol) was truncated by 5 residues (pSHHcΔ189-193) (Appendix Fig. S3.3B). Interestingly the length of this linker region is highly conserved. Additionally, removal of the cholesterol moiety (SHH) from SHH already lacking the N-terminal palmitate (SHHc) reduced levels HH signalling (Appendix Fig. S3.3C). Finally, addition of a palmitoylated N-terminal SHH peptide to systems with/without a cholesterylated C-terminal SHH peptide resulted in enhanced HH pathway activation when the cholesterylated-SHH peptide was present (Appendix Fig. S3.3D). This suggests the C-terminal cholesterol attachment of SHH

contributes to PTCH1 inhibition in cells (in addition to the palmitoyl moiety) and may indicate cooperativity between the lipid binding sites that does not rely on the presence of the globular domain of SHH. Thus, both the structural analysis and functional assays support a mechanism of dual inhibition of PTCH1 by SHH via pincer like insertion of the SHH palmitoyl and cholesteryl attachments into PTCH1.

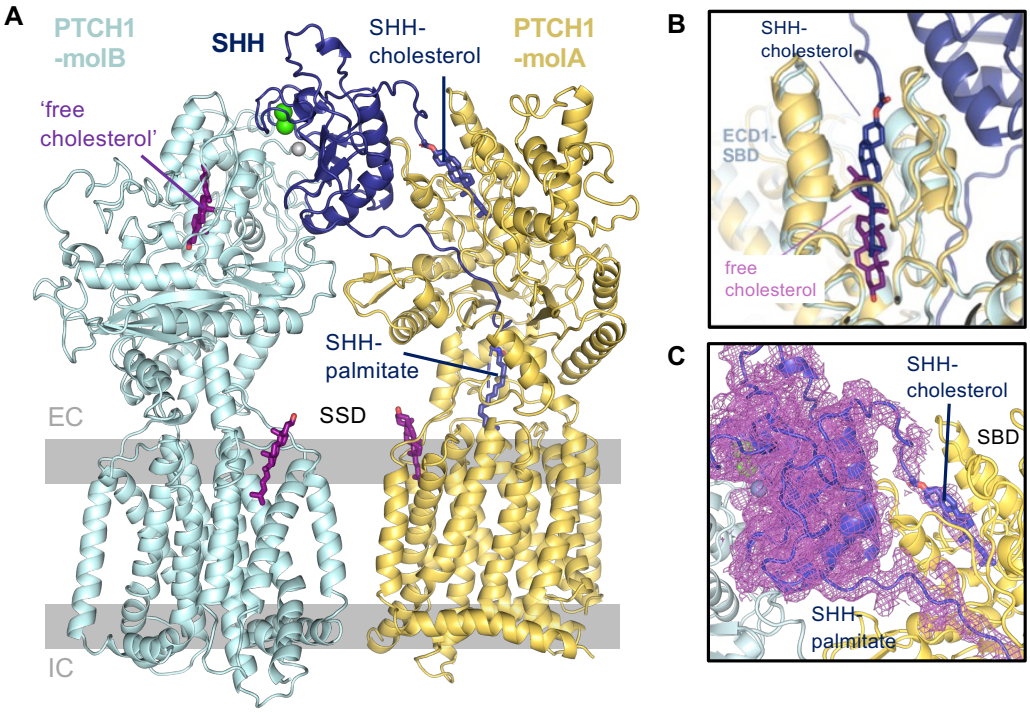


Figure 3.3: The SHH-cholesterol moiety binds and inhibits PTCH1.

A) Re-refined cryo-EM structure of a 2:1 PTCH1:SHH complex (original PDB: 6E1H, revised PDB: 6RVD)[220]. SHH (dark blue) binds to PTCH1-molA (yellow) via insertion of terminal palmitoyl/cholesteroyl attachments or to PTCH1-molB (light blue) via a metal ion binding interface. 'Free cholesterol' bound within PTCH1-molB or to the sterol sensing domains (SSD) is shown in purple. **B)** Comparison of the 'free cholesterol' and 'SHH-cholesterol' orientations within the SBD of PTCH1-molA/B. **C)** Density corresponding to the SHH C-terminus and cholesterol attachment within the SBD of PTCH1-molA. Figure adapted from Rudolf *et al* [220].

To further probe the differences between cholesterol binding poses within the SBD of PTCH1 we performed atomistic simulations of the PTCH1-molA and PTCH1-molB ECDs in the presence or absence of the bound SBD cholesterols. In $PTCH1_{ECD}$-molA the presence of the 'SHH-cholesterol' did not stabilise the upper lobe of the ECD when compared to the *apo* simulations. Conversely, the upper lobe of $PTCH1_{ECD}$-molB was stabilised by the presence of the 'free cholesterol' molecule (Fig. 3.4A). This difference was also reflected in the RMSD of the cholesterol molecules which were lower for the 'free cholesterol' (bound to $PTCH1_{ECD}$-molB) than the 'SHH-cholesterol' orientation

(bound to PTCH1$_{ECD}$-molA or in the PTCH1$_{ECD1}$ crystal structure) (Fig. 3.5). Comparison of the end snapshots from atomistic simulations indicated rotation of the SBD of PTCH1$_{ECD}$-molA around a disulphide linked hinge region (Fig. 3.4B). In PTCH1$_{ECD}$-molB hinge-like rotation is prevented by the deeper position of the cholesterol molecule, accounting for the reduced RMSD. Thus, differences in the position of the 'SHH-cholesterol' and the 'free cholesterol' molecules result in distinct behaviours of the PTCH1-molA and PTCH1-molB ECDs.

We note that in simulations of PTCH1$_{ECD1}$ and PTCH1$_{ECD}$-molA, cholesterol moved laterally within the SBD, suggestive of a tunnel-like pocket able to accommodate both cholesterol orientations, rather than a well-defined binding site.

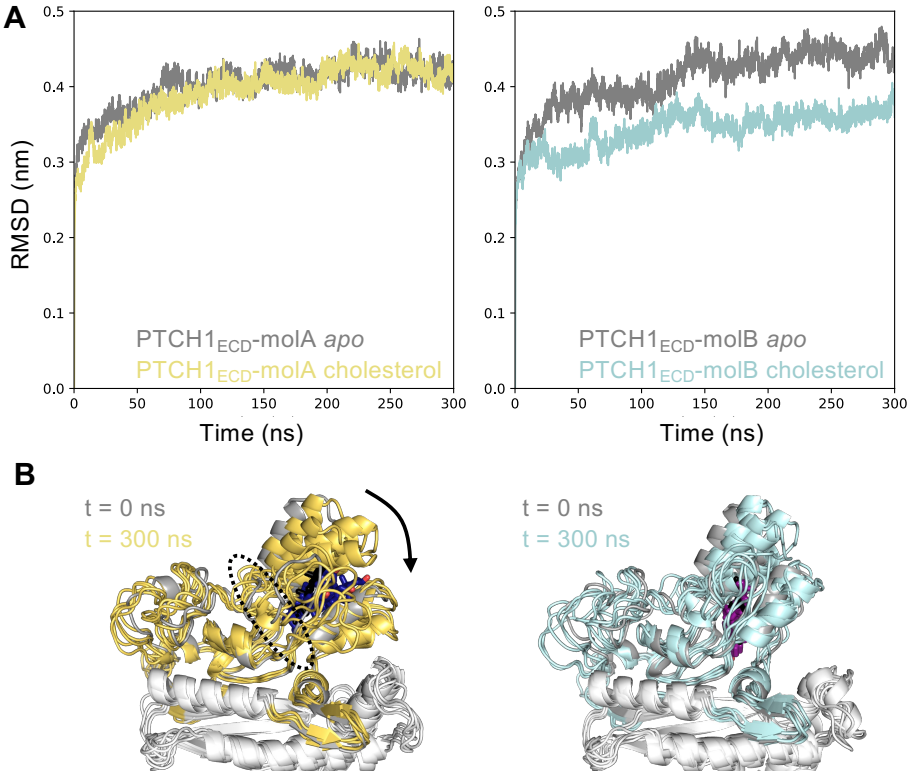


Figure 3.4: 'Free cholesterol' stabilises the PTCH1 SBD.

A) RMSD of the PTCH1 upper lobe in the presence of 'SHH-cholesterol' (in PTCH1$_{ECD}$-molA, yellow), 'free cholesterol' (in PTCH1$_{ECD}$-molB, light blue) or when *apo* (grey) across 5 x 300 ns simulations. RMSD was calculated as in Fig. 3.2C. Figure adapted from Rudolf *et al* [220]. **B)** Comparison of the start and end snapshots of PTCH1$_{ECD1}$ across replicates. Rotation of the SBD around a disulphide hinge region is circled.

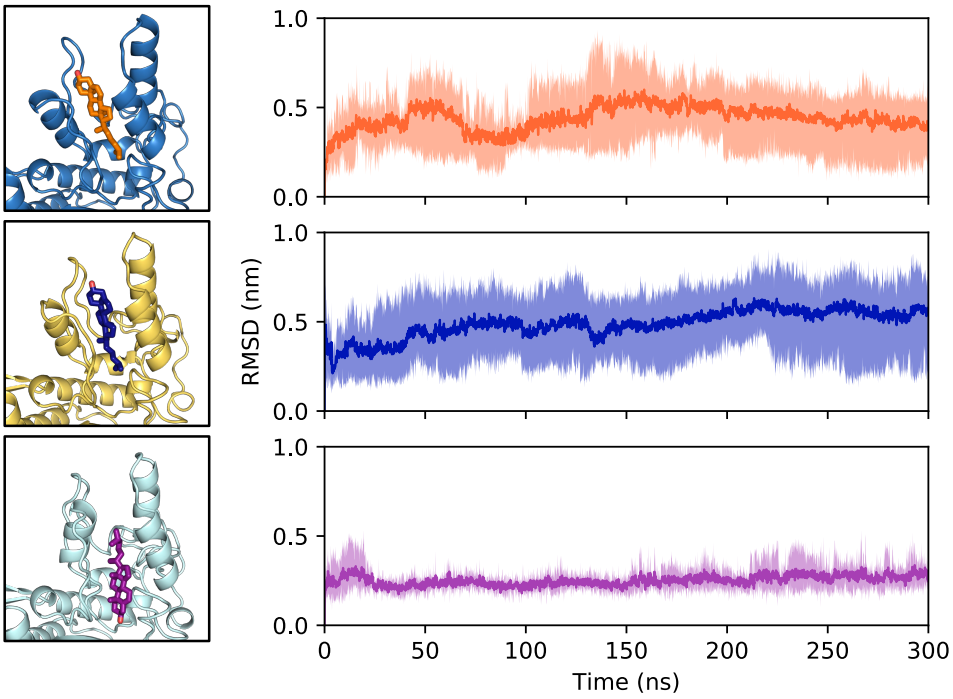


Figure 3.5: Preferential stabilisation of the 'free cholesterol' orientation.

Comparison of the cholesterol RMSD (calculated as in Fig. 3.2C) across 5 x 300 ns simulations of $PTCH1_{ECD1}$ (orange), $PTCH1_{ECD}$-molA ('SHH-cholesterol' pose, dark blue) and $PTCH1_{ECD}$-molB ('free cholesterol' pose, purple). Snapshots indicate relative cholesterol orientations within the SBD. Figure adapted from Rudolf *et al* [220].

3.3.2 Collapse of tunnels throughout PTCH1 by SHH-cholesterol

Previous structural analysis of PTCH1 identified a 150 Å tunnel sandwiched between the PTCH1 ECDs, extending from the SBD to openings near the transmembrane region[122,211,213]. This tunnel was proposed to be a conduit for cholesterol transport although the exact pathway remains uncertain. One tunnel exit, located near the membrane is in proximity to a structural motif formed by TM2-6 called the sterol sensing domain (SSD). SSDs are also found in other proteins implicated in e.g. cholesterol biosynthesis and transport[236]. This, combined with our observations of lateral cholesterol movement within the ECD, led us to further investigate formation of dynamic tunnels during atomistic simulations of the $PTCH1_{ECD}$ (Fig. 3.6). Analysis of tunnel formation in simulations poses two advantages compared to equivalent analysis of static structures: 1) since simulations were performed in the presence of SBD cholesterols, the behaviour of the sterols on tunnel formation can be probed and

2) the relative occurrence of distinct tunnel localities can be assessed by monitoring tunnel formation with time.

For PTCH1$_{ECD}$-molA 88% of tunnel localised to the SBD, extending from the position of the cholesterol tail towards the mouth of the SBD. In contrast for PTCH1$_{ECD}$-molB an extensive network of tunnels protracted throughout the ECD and no tunnels were observed to egress through the SBD mouth. Tunnels within the ECD of PTCH1$_{ECD}$-molB moved either towards two openings in proximity to the TMD, located near the SSD (10%) or near TM7/12 (21%), or via side exits between ECD1/ECD2 (Side Exit 1: 55%, Side Exit 2: 10%). Intriguingly Side Exit 2 overlays with the location where the palmitoyl moiety of SHH inserts into PTCH1[206]. Thus, our analysis of tunnel formation throughout PTCH1 further supports the proposal that PTCH1 may function as a cholesterol transporter protein and that SHH inhibits PTCH1 function by insertion of both lipid attachments into the PTCH1 ECD, leading to collapse of the proposed sterol transport tunnels.

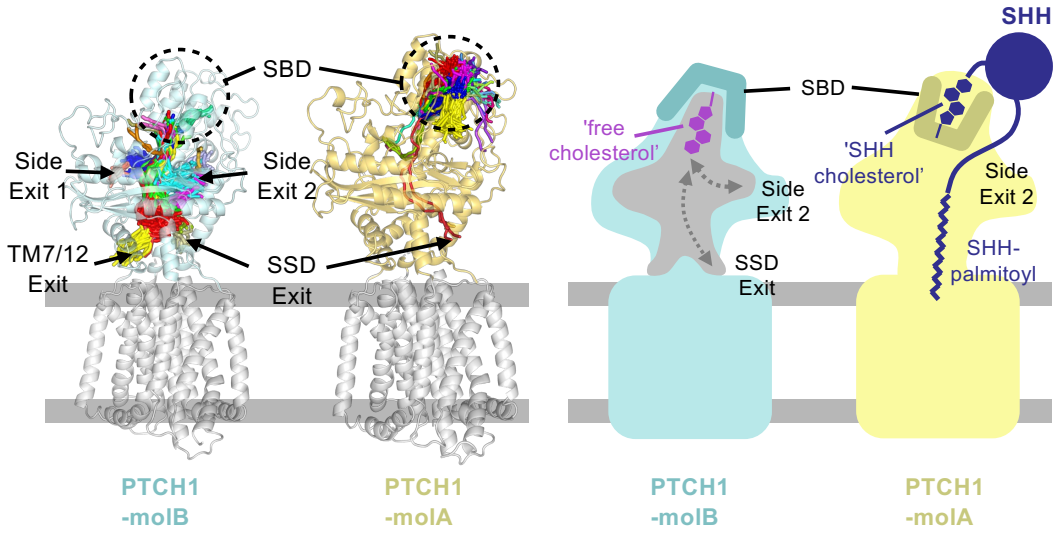


Figure 3.6: Analysis of tunnels through PTCH1.

Formation of dynamic tunnels throughout atomistic simulations of PTCH1$_{ECD}$-molA and PTCH1$_{ECD}$-molB (5 x 300 ns each). Tunnels were calculated using Caver Analyst and are displayed as the centroid through the tunnel (coloured by tunnel cluster). A model for SHH-cholesterol induced collapse of putative sterol transport tunnels through PTCH1 is displayed alongside.

3.3.3 Probing the energetics of cholesterol transport pathways

There is currently no direct *in vitro* assay for PTCH1, owing to difficulties in labelling and measuring changes in the abundance/distribution of the proposed ligand; cholesterol. We sought to assess whether computational simulations could be used to probe the energetics and pathway of the proposed cholesterol transport mechanism. We constructed free energy cycles corresponding to cholesterol movement through PTCH1-molA and PTCH1-molB (Fig. 3.7A). Each cycle is composed of a pathway directly linking the SSD and SBD (ΔG_1), and an indirect pathway between the same sites (composed of ΔG_2, ΔG_3 and ΔG_4). Thus, free energy cycles were constructed to compare how the energetic landscape changes between the PTCH1-molA and PTCH1-molB conformations and between the 'free cholesterol' and 'SHH-cholesterol' orientations.

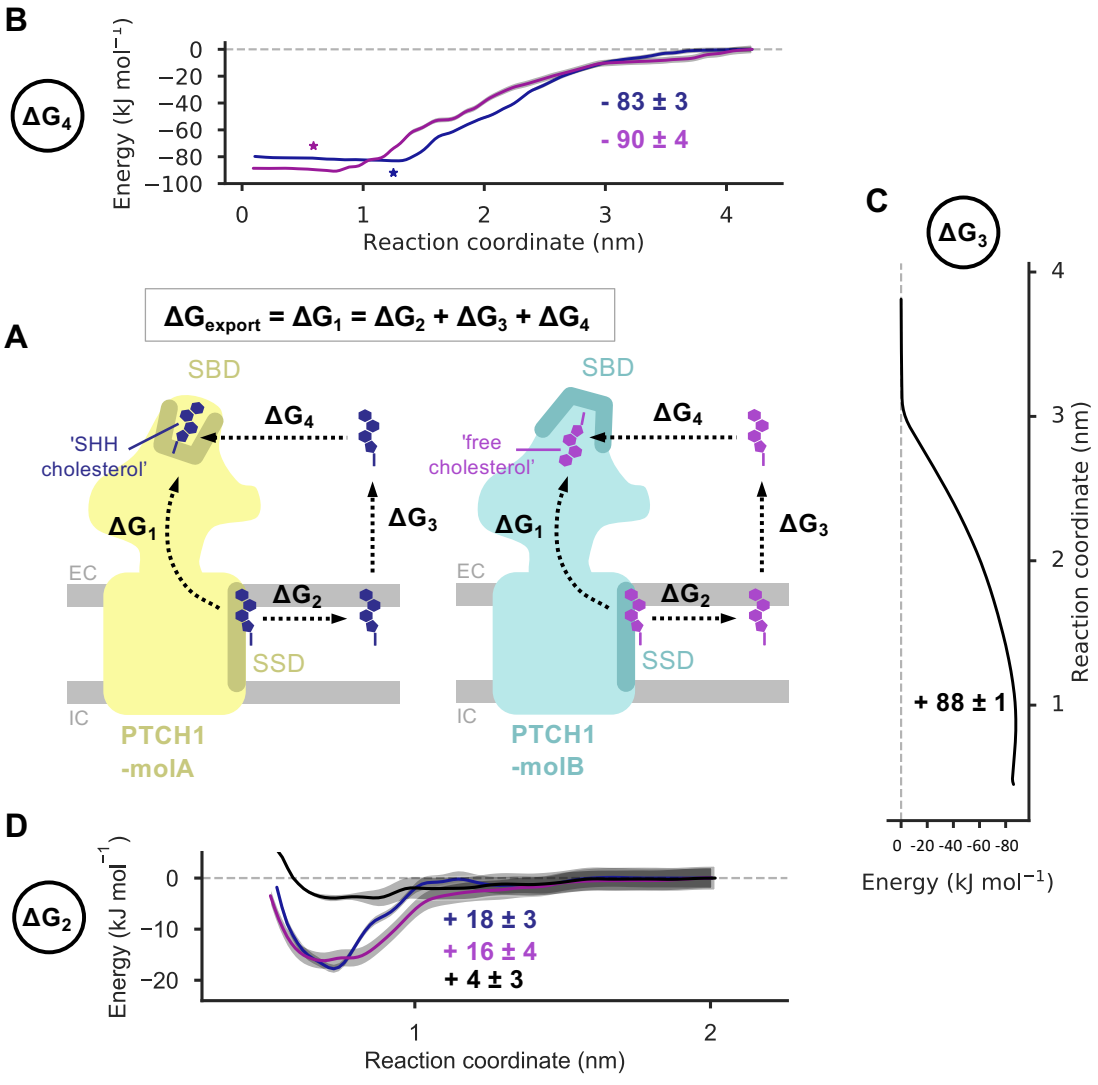


Figure 3.7: The indirect pathway.

A) Schematic of free energy cycles for cholesterol movement through PTCH1 composed of a direct pathway (ΔG_1) through the protein or an indirect pathway (ΔG_{2-4}). Potential of mean force (PMF) profiles for **B)** cholesterol movement between the solvent and SBD (ΔG_4), **C)** extraction of cholesterol from the bilayer (ΔG_3) and **D)** cholesterol movement from the SSD into the membrane (ΔG_2). Traces for the 'SHH-cholesterol' (dark blue) and 'free cholesterol' (purple) orientations are indicated where relevant. Stars correspond to the position of cholesterol molecules within the SBD in the cryo-EM structure (PDB: 6RVD). The black trace in **D** indicates cholesterol bound to the SSD in an unconventional 'OH-down' orientation with the hydroxyl group towards the bilayer midplane.

The indirect pathway:

We used CG potential of mean force (PMF) calculations to probe the energetics of cholesterol movement via the indirect pathway. The indirect pathway is formed by three PMFs corresponding to cholesterol movement between the PTCH1 SSD and the bulk membrane (PMF-2: ΔG_2), between the membrane and solvent (PMF-3: ΔG_3) and between the PTCH1 SBD and the solvent (PMF-4: ΔG_4). Cholesterol binding within the SBD is energetically favourable relative to the solvent (PMF-4) for both PTCH1 conformations (Fig. 3.7B, PTCH1-molA: ΔG_4 = -83 ± 3 kJ mol^{-1}, PTCH1-molB: ΔG_4 = -90 ± 4 kJ mol^{-1}). We observe a flattened landscape within the SBD, rather than a localised energetic well at the cholesterol position observed in the cryo-EM structure (Fig. 3.7B, stared). This is consistent with the tunnel-like nature of the pocket observed in atomistic simulations of PTCH1$_{ECD}$s and the lateral movement of cholesterol within the pocket. The energetic cost associated with extraction of cholesterol from the bilayer into the surrounding solvent (PMF-3), was, expectedly, sizeable and positive (Fig. 3.7C, ΔG_3 = +88 ± 1 kJ mol^{-1}). ΔG values for PMF-3 and PMF-4 are comparable, suggesting the PTCH1 SBD pocket hydrophobicity is well matched to mimic the native bilayer environment of cholesterol.

Given the observations of divergent cholesterol orientations within the PTCH1 ECDs and the moderate resolution surrounding the TMD in PTCH1 structures, we were unsure whether the SSD cholesterol orientation would match the conventional orientation expected within the bilayer. To address this uncertainty, we calculated the free energy change for cholesterol movement away from the SSD when the ROH bead (equivalent to the β_3-hydroxyl group) pointed towards the extracellular leaflet ('OH-up')

or towards the bilayer midplane ('OH-down'). Only the 'OH-up' orientation produced the energetic well expected for a lipid binding site (Fig. 3.7D, PTCH1-molA: ΔG_2 = +18 ± 3 kJ mol^{-1}, PTCH1-molB: ΔG_2 = +16 ± 4). The energetic landscape for the 'OH-down' orientation remained approximately flat (Fig. 3.7D, PTCH1: ΔG_2 = +4 ± 3 kJ mol^{-1}) over the reaction coordinate. Thus, cholesterol most likely binds within the SSD in an 'OH-up' orientation with a moderate free energy change.

By rearranging the free energy values from the indirect pathway (Equation 3.2) we estimate the free energy associated with cholesterol movement from the PTCH1 SSD to the SBD (ΔG_1) to be +23 ± 4 kJ mol^{-1} for PTCH1-molA and +14 ± 6 kJ mol^{-1} for PTCH1-molB. Thus, calculations suggest cholesterol export by PTCH1 between the SSD to the SBD would likely require coupling to an energetic input to overcome the transport barrier. Alternatively, these data are consistent with spontaneous cholesterol import into the membrane.

[3.2] $$\Delta G_{export} = \Delta G_1 = \Delta G_2 + \Delta G_3 + \Delta G_4$$

The direct pathway:

The reaction coordinates used in the indirect pathway were topologically simple – composed of a relatively linear interpolation between the start and end windows. A direct pathway through PTCH1 between the SSD and SBD is likely non-linear. The tunnels identified in atomistic simulations of the $PTCH1_{ECD}$ between the base of the ECD and the SBD, provide insights into a possible cholesterol path. In contrast, it is not obvious how cholesterol transitions between the membrane localised SSD and the base of the ECD. Previous predictions suggest cholesterol may move either between TM helices or outside of the protein[122,211,213]. Thus, defining a complete physical pathway to probe with PMFs is challenging.

To circumvent these limitations, we used absolute binding free energy (ABFE) calculations to obtain the free energy for cholesterol movement along the direct pathway through alchemical space (performed by Dr Robin Corey). The difference in decoupling cholesterol from the SSD and SBD sites can be used to estimate ΔG_1 (or $\Delta\Delta G_1$ as is the case here, where $\Delta\Delta G_1 = \Delta G_{1\text{-SSD}} - \Delta G_{1\text{-SBD}}$) for both the 'SHH-

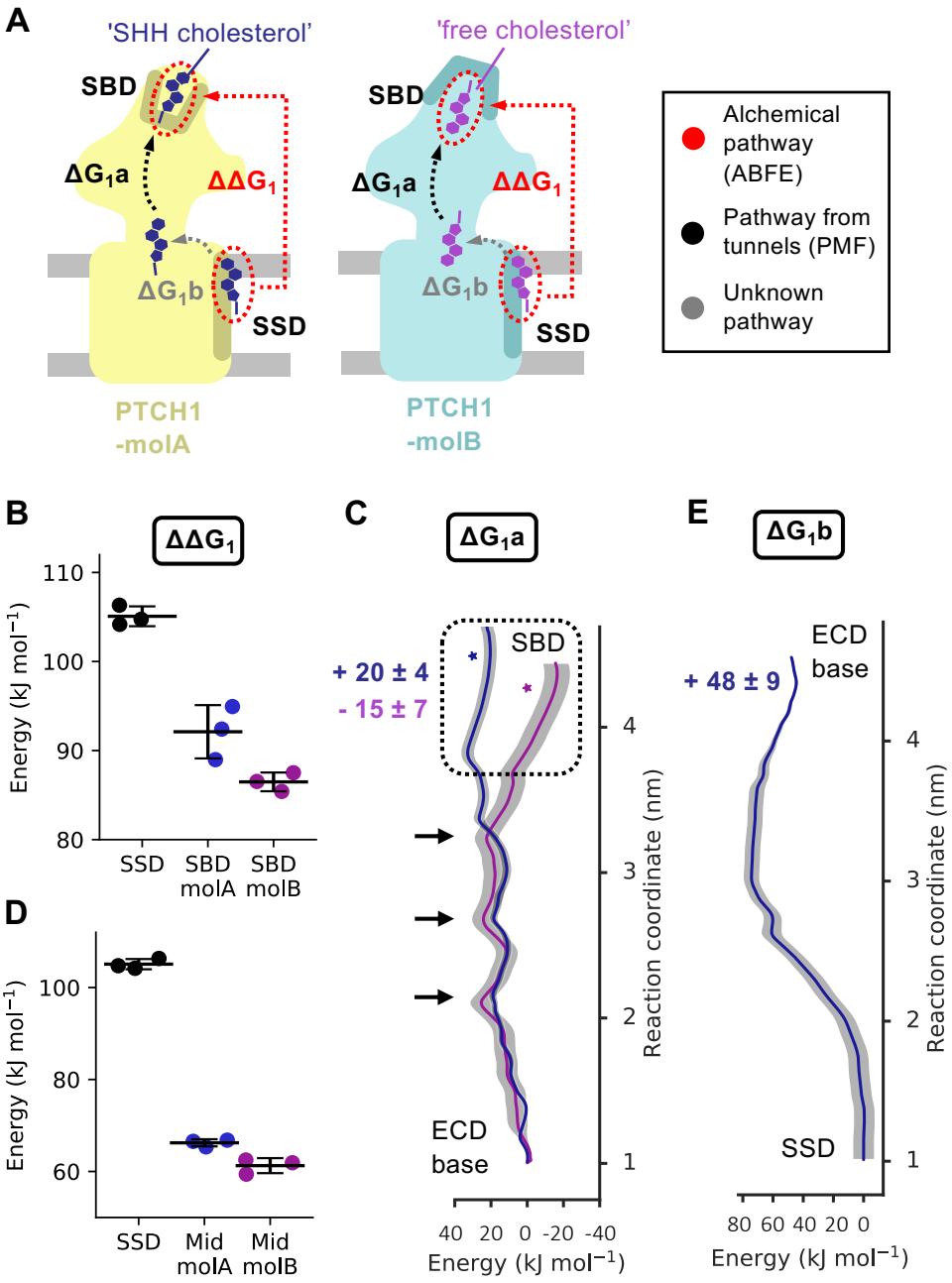


Figure 3.8: The direct pathway.

A) Schematic of alchemical (red) or physical (black/grey) cholesterol transport pathways through PTCH1. **B)** Absolute binding free energy (ABFE) calculations for cholesterol decoupling from the SSD or SBD. **C)** PMF profiles for cholesterol movement between the ECD base and the SBD (ΔG_1a, black arrow in **A**). Arrows indicate conserved energetic peaks within the ECD. Stars correspond to the position of cholesterol molecules within the SBD in the cryo-EM structure (PDB: 6RVD). **D)** ABFE for cholesterol bound at ~2.4 nm along the reaction coordinate in **C** at a 'mid' position along the ECD. **E)** PMF profile for cholesterol movement between the SSD and the ECD base (ΔG_1b, grey arrow in **A**). Plots corresponding to the 'SHH-cholesterol' and 'free cholesterol' orientations are indicated in dark blue/purple respectively. ABFE calculations were run by Dr Robin Corey.

cholesterol' and 'free cholesterol' orientations (Fig. 3.8A). We predict a positive free energy change between the SSD and SBD for both PTCH1 molecules (Fig. 3.8B, PTCH1-molA: $\Delta\Delta G_1$ = +13 ± 3 kJ mol^{-1}, PTCH1-molB: $\Delta\Delta G_1$ = +19 ± 1 kJ mol^{-1}). Thus, the magnitude of ΔG_1 predicted by the indirect PMF pathway and direct ABFE pathways are in agreement, albeit with reversed relative orders for PTCH1-molA and PTCH1-molB. Both methods of free energy calculation therefore support the proposed hypothesis that cholesterol export through PTCH1 would require coupling to an energy source.

We also calculated the free energy associated with cholesterol movement between the ECD base and the SBD (PMF-1a: ΔG_1a), to understand the energetic landscape over this subsection of the direct pathway using PMFs (Fig. 3.8C). Three energetic peaks (+18 to +27 kJ mol-1) are observed for both PTCH1-molA and PTCH1-molB over the lower and mid regions of the ECD (arrowed). The similarity of the profiles in this region was confirmed using ABFE calculations for cholesterol at ~2.4 nm along the reaction coordinate (Fig. 3.8D, PTCH1-molA: + 39 ± 2 kJ mol^{-1}, PTCH1-molB: + 44 ± 1 kJ mol^{-1}). Strikingly, as cholesterol moves towards the SBD the paths for PTCH1-molA and PTCH1-molB diverge (stars in Fig. 3.8C indicate the position of SBD cholesterol modelled in the cryo-EM structure). Movement of cholesterol into the SBD of PTCH1-molB is associated with stabilisation of -15 ± 7 kJ mol^{-1} compared to the free energy at the ECD base (1 nm in Fig. 3.8C). In contrast cholesterol within the SBD of PTCH1-molA is destabilised by +20 ± 4 kJ mol^{-1}. This supports our observed stabilisation of the PTCH1-molB cholesterol and ECD upper lobe in atomistic simulations (Fig. 3.4/3.5), and the slight stabilisation of cholesterol binding to the PTCH1-molB SBD in PMF-4 (Fig. 3.7B).

We also tried to calculate the free energy for cholesterol movement between the ECD base and the SSD (PMF-1b: ΔG_1b) however this was highly non-trivial using PMF calculations. There were challenges with both obtaining a reasonable reaction coordinate, and in separating cholesterol translocation and flipping energies as it enters the bilayer (Fig. 3.8E). The large energetic barriers observed results from cholesterol solvation as the molecule exits the ECD, which may be reduced by e.g. the glycocalyx or association with CDON/BOC partner proteins in cells[19,237]. An alternative (and probably more physiologically relevant) pathway through the TM

helices may circumvent unfavourable solvent exposure, however we were unable to probe this in CG PMFs due to imposition of the elastic network with prevents PTCH1 conformational changes.

In summary, our ΔG_1 values calculated from the indirect (PMF) and direct (ABFE) pathways both suggest a gain in free energy for cholesterol movement from the SSD to the SBD of approximate magnitude +13 to +23 kJ mol^{-1} (Fig. 3.9).

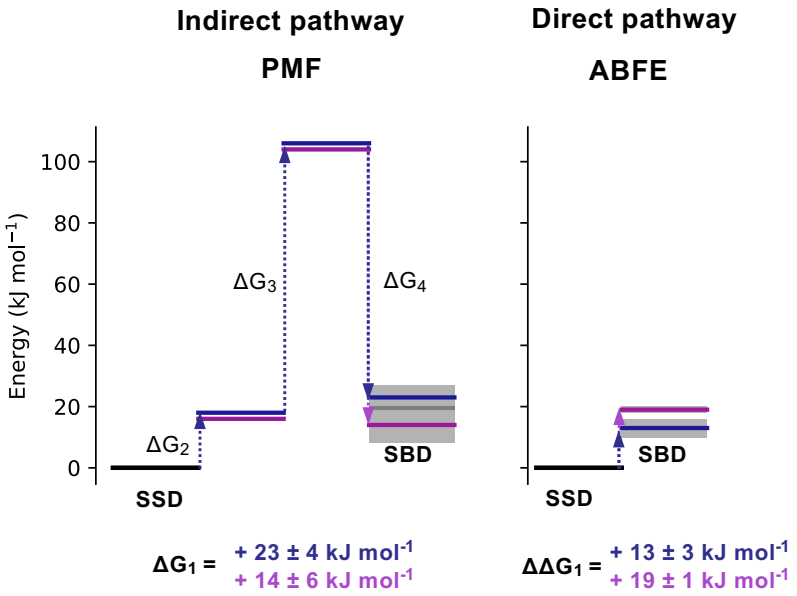


Figure 3.9: Cholesterol export is associated with a gain in free energy.

Comparison of 'SHH-cholesterol' (dark blue) and 'free cholesterol' (purple) export free energies as calculated from the indirect (PMF) and direct (ABFE) pathways. For the indirect pathway the error on ΔG_1 was calculated in quadrature since PMFs 2-4 were independent.

3.3.4 Investigating the behaviour of ions in PTCH1

PTCH1 shares a conserved structural architecture with members of the prokaryotic RND transporter superfamily, which can utilise the proton gradient to export small molecules, drugs, peptides and metal ions[238,239]. In PTCH1, two aspartate residues (D513 and D514) on TM4 and a glutamate residue (E1095) on TM10 localise to equivalent positions to the proposed proton binding site in RND transporters. These

residues also comprise the conserved GXXXDD and GXXX(E/D) motifs at the centre of the TMD, proposed to facilitate RND transport function[57,122].

Visualisation of deposited cryo-EM maps for PTCH1 revealed a putative ion-like density coordinated by the anionic triad residues (Fig. 3.10A, inset). We reasoned that PTCH1 may use cation gradients across the membrane to make cholesterol movement between the SSD and SBD favourable, akin to RND proton coupled export.

We performed simulations of PTCH1 with Na^+ positioned within the TMD according to the cryo-EM density (termed Site 1) to test the viability of the proposed binding site. Na^+ remained bound at Site 1 over 3 x 100 ns replicates with an RMSD of 0.15 ± 0.02 nm compared to the initial ion position (Fig. 3.10B). Similarly in simulations initiated with an *apo* TMD (in 0.15 M NaCl or KCl), Na^+/K^+ cations spontaneously and stably occupy Site 1 in all replicates (Fig. 3.10C-D), coordinated octahedrally by the carboxy groups of D513, D514 and E1095, the backbone carbonyl of V510 and two water molecules (Fig. 3.10E). Na^+ binding to Site 1 could not be disrupted by application of a physiologically relevant membrane voltage (Appendix Fig. S3.4). A second site (Site 2) was observed on the intracellular portion of TM4/11, whereby cations are coordinated by H520, D1128 and four water molecules (Fig. 3.10F). Site 2 localises to a small, solvated hollow on PTCH1, formed by local bilayer deformation (visible also in MemProtMD[203]). Occasionally cations exchange between Site 1 and Site 2, mediated by D1128 flipping, thus forming a putative recruitment/exchange pathway between the solvent and Site 1 within the TMD.

Our reasons for exploring PTCH1 interactions with both Na^+ and K^+ reflect surprising recent findings in PTCH1 biochemistry. One study suggested PTCH1 could be regulated by levels of extracellular Na^+ based on elevated cAMP levels as a reporter for HH signalling[218]. However, a later study reporting on SMO localisation to the cilia, suggested PTCH1 activity is dependent on the K^+ gradient[216]. Both these studies have limitations, notably use of cells which do not undergo HH signalling physiologically and using SMO as an indirect readout for PTCH1 activity. To address these limitations members of the Rohatgi lab used a PFO analogue to monitor changes in levels of accessible membrane cholesterol using TIRF microscopy[21]. They also suggest the presence of a K^+ gradient, and not Na^+, is required for PTCH1 activity.

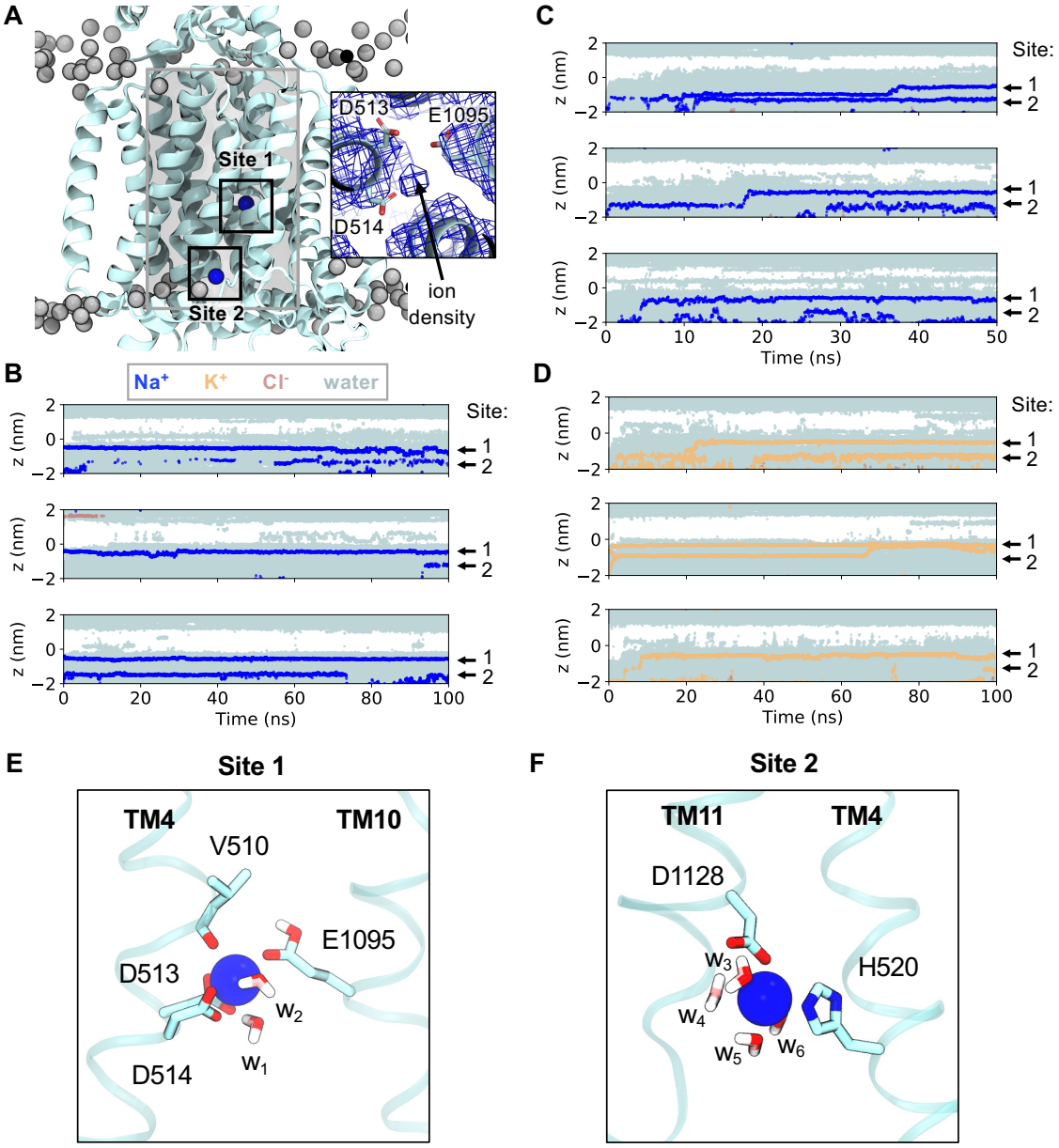


Figure 3.10: Cation binding within the PTCH1 TMD.

A) Snapshot from atomistic simulations of PTCH1 (PDB: 6DMY)[122] showing the location of cation binding sites within the TMD. The location of a cylinder (length: 4 nm, radius: 1.3 nm) centred on the midpoint of the V510/I1092 Cα atoms (see **B-D)** is marked in grey. The inset shows ion-like density coordinated by charged residues visible within the cryo-EM structure. **B-D)** Water and ions within the TMD cylinder across 3 replicate simulations with Na^+ initially positioned at Site 1 according to the density (**B**) or without cations initially bound in 0.15 M NaCl (**C**) or KCl (**D**). **E-F)** Coordinating residues/water at Site 1 and Site 2.

The involvement of the K^+ gradient is still controversial given there are no known eukaryotic K^+ coupled transporters and the resting membrane potential is theoretically closest to the equilibrium K^+ potential. To further probe PTCH1 cation selectivity we performed free energy perturbation (FEP) calculations whereby Na^+ was gradually exchanged for K^+ (or *vice versa*) at Site 1. Compared to ion perturbation in solvent, perturbation of Na^+ to K^+ at Site 1 was 2.7 ± 0.1 kJ mol^{-1} more favourable (Fig. 3.11, Appendix Fig. S3.2). The reverse perturbation agreed with this value, suggesting K^+ binding at Site 1 is fairly similar to Na^+, hence biochemical implication of both Na^+ and K^+ is perhaps less surprising[21,216,218].

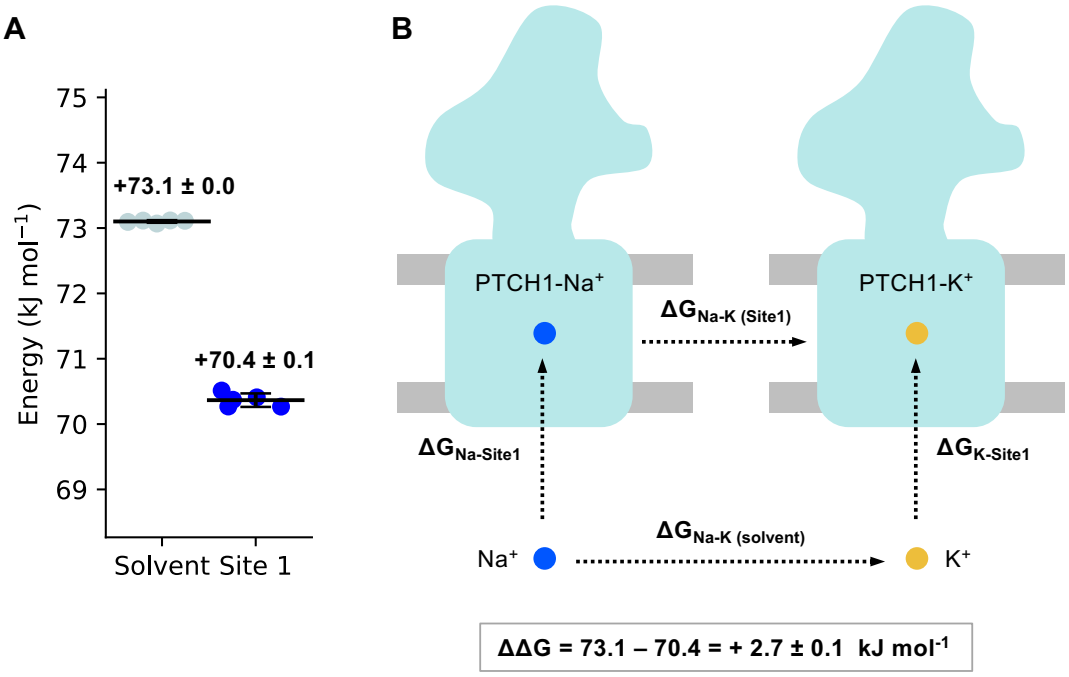


Figure 3.11: Site 1 cation free energy perturbation.

A) Atomistic free energy perturbation of Na^+ to K^+ in solvent and at Site 1. **B)** Schematic representation of free energy cycle used to calculate the difference in Na^+ binding to Site 1 compared to K^+ (ΔΔG).

3.3.5 The hydrophobic cap

Our ion analyses also hinted to the presence of a solvated region within the intracellular portion of the PTCH1 TMD (Fig. 3.10B-D). We calculated the time averaged water density across simulations of PTCH1 (Fig. 3.12A), revealing a hydrated cavity encompassing the anionic triad residues and opening to the cytosol,

similar to those proposed from simulations of the O- and L- states of AcrB[114,115]. Pathways for water entry could be seen between the intracellular regions of TM4/TM10 and via the water pocket surrounding Site 2 (Fig. 3.12A, see arrows). We observe infrequent water crossing events (and no ion crossings) within the extracellular half of the TMD which remained de-wetted.

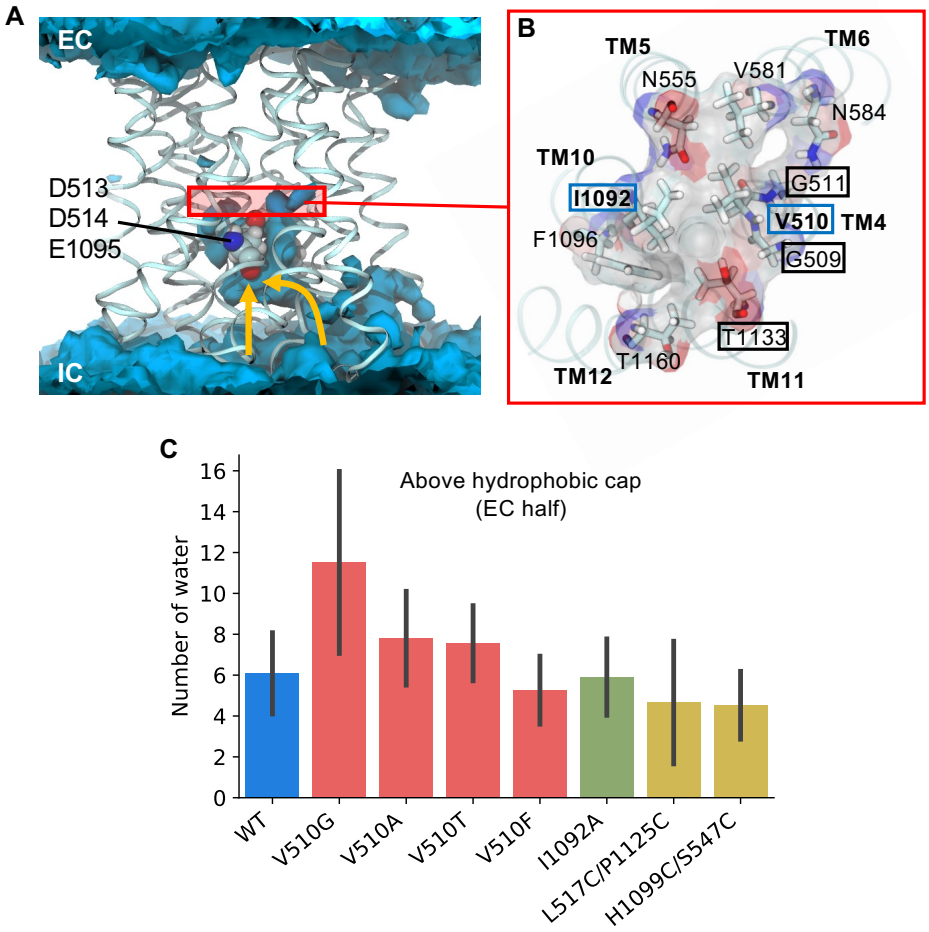


Figure 3.12: The hydrophobic cap.

A) Time averaged solvent density (blue) with the PTCH1 TMD across a 100 ns simulation with Na^+ initially bound at Site 1. Arrows mark pathways for water entry and the anionic triad residues are shown as spheres. **B)** Residues comprising the proposed hydrophobic cap displayed as sticks and also in surface representation. Glycine flanked V510/I1092 residues of the conserved GXXXDD/GXXX(E/D) motifs are boxed in blue and residues mutated in disease phenotypes in black. **C)** Number of water within the extracellular half of the PTCH1 TMD cylinder (see Fig. 3.10) across the final 10 ns of 3 x 50 ns simulations of WT PTCH1 or *in silico* PTCH1 mutants.

Trajectory visualisation revealed a cluster of residues above the anionic triad which may form a hydrophobic cap to restrict water entry (Fig. 3.12B). The sidechains of two of these residues (V510 and I1092) position between TM4/TM10, directly above the cation binding site. V510 and I1092 form part of the conserved GXXXDD and GXXX(D/E) motifs (first X in each sequence), and each are flanked by two glycine residues which could e.g. enable helical rearrangements to permit solvent permeation. To probe whether disruption of the hydrophobic cap is sufficient for solvent permeation we performed multiple *in silico* mutations of PTCH1 and ran atomistic simulations of each (3 x 50 ns). We calculated the average number of waters within the extracellular half of a cylinder centred on V510/I1092 across the last 10 ns of each simulation (Fig. 3.12C). Mutation of V510 to glycine increased the number of waters within the extracellular TMD region (waters per frame: 12 ± 5) compared to WT (waters: 6 ± 2). Mutation of V510 to a smaller hydrophobic (V510A, waters: 8 ± 2) or polar (V510T, waters: 8 ± 1) residue had no effect on the average water number per frame, neither did the I1092A mutation (waters: 6 ± 1). Attempts to disrupt the packing of the cap (V510F, waters: 5 ± 1) or to lock the intracellular TM segments by cross-linking TM4-TM11 (L517C-P1125C, waters: 4 ± 3) or TM5-TM10 (H1099C-S547C, waters: 4 ± 1) had a minimal effect, although the disulphide mutants did prevent Na^+ entry to Site 1 in all but one trajectory. Thus, the driving forces for solvent permeation are more complex than relief of a hydrophobic restriction at the TMD midplane. While V510 may sterically impede water movement (as evidenced by the V510G mutant), the cross-linking mutants suggest breathing-like motions of the intracellular TMD regions may also regulate water/ion entry in PTCH1. These mutants are currently being tested biochemically.

3.3.6 Insights from Dispatched1

Whilst investigating PTCH1 ion interactions a structure of Dispatched1 (DISP1) was published[217] (Fig. 3.13A). DISP1 catalyses the Na^+ coupled export of SHH from the membrane into the extracellular space (upstream of binding to PTCH1) and is also a member of the RND family[12]. Three Na^+ ions were modelled within the TMD, coordinated by equivalent residues to the anionic triad in PTCH1 (Fig. 3.13B). Given the challenges with interpreting ions *vs* water *vs* noise in cryo-EM structures, we

sought to assess whether a) 3 x Na^+ could be stably coordinated within this small, charge-dense region and b) whether DISP1 could inform our understanding of PTCH1 function.

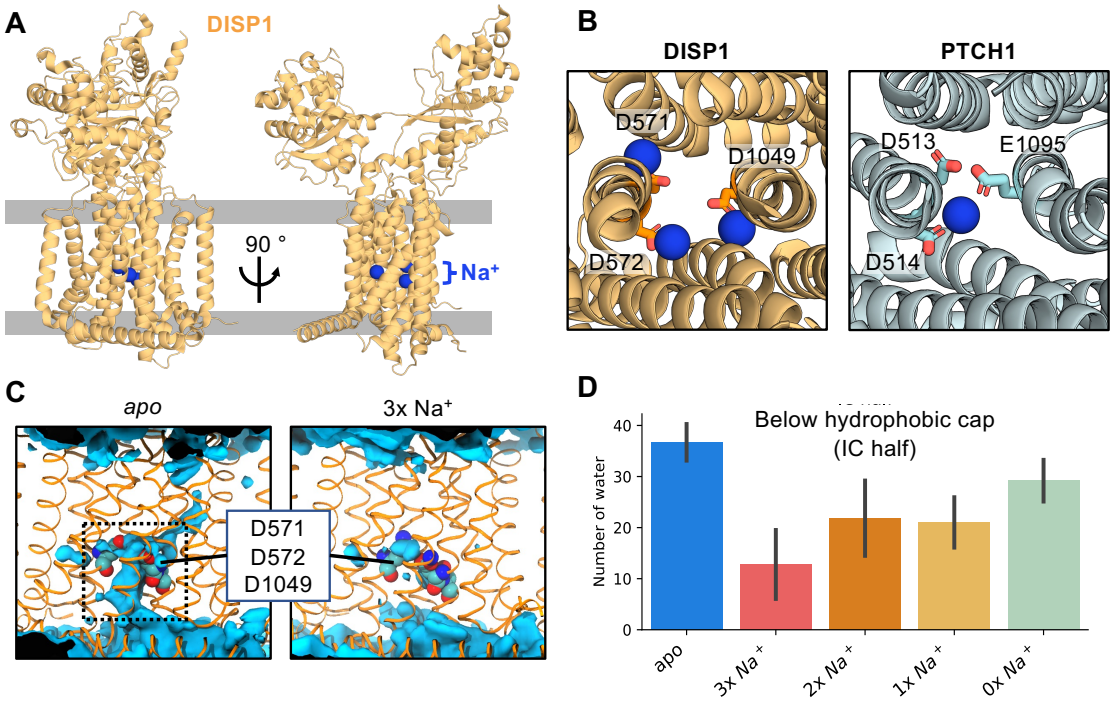


Figure 3.13: Water within Dispatched1 (DISP1).

A) Cryo-EM structure of DISP1 (PDB: 7RPH)[217] showing the location of three Na^+ ions (blue spheres) within the TMD. **B)** Comparison of Na^+ binding sites and anionic triad residues of DISP1 and PTCH1. **C)** Time averaged solvent density across 100 ns atomistic simulations of DISP1 with the TMD initially *apo* or with 3x Na^+ bound. Anionic triad residues are shown as spheres. **D)** Water within the intracellular half of a cylinder (length: 4 nm, radius: 1.3 nm) centred on the midpoint of I568/L1046 Cα residues across the final 10 ns of 3 x 100 ns simulations of DISP1 in distinct ion coupled states. The '2x Na^+', '1x Na^+' and '0x Na^+' states were generated by sequential ion removal from the end snapshot of the previous simulation.

We performed atomistic simulations of DISP1 with all ions bound ('3x Na^+') or removed (*apo*). In the *apo* state DISP1 exhibited a similar water density profile to PTCH1; an intracellular water cavity was clearly visible. Unexpectedly, in the '3x Na^+' state ion binding was extremely stable (RMSD: 0.16 ± 0.01 nm per Na^+) and the intracellular water cavity disappeared (Fig. 3.13C). We assign these conformations putatively as 'inward-open' and 'occluded' states. We reasoned Na^+ binding may restrict water entry by interlocking neighbouring helices of the TMD. We sequentially removed ions from

the end snapshots of simulations ('2x Na^+', '1x Na^+' and '0x Na^+') and calculated the mean number of waters within the intracellular cavity across the last 10 ns of simulations (Fig. 3.13D). Strikingly, sequential ion removal increased the average number of waters per frame within the intracellular half of the TMD, indicating all three Na^+ ions contribute to TMD stabilisation and to restriction of water entry.

To dissect this further we calculated minimum distances between the Cα atoms of residues at the ends of TM helices. All distances between pseudo-symmetric TM helices on the extracellular side were within ~0.1 nm of each other between distinct Na^+ bound states (Appendix Fig. S3.5). Conversely, the intracellular region of TM helices were displaced inwards by ~0.4 nm in the '3x Na^+' compared to *apo* states (Fig. 3.14) and ion removal shifted intracellular distances towards the *apo* state.

In conclusion, DISP1 simulations suggest transition from an 'inward-open' to 'occluded' state involves cation induced stabilisation and inward movement of intracellular helices, which excludes water from the intracellular cavity.

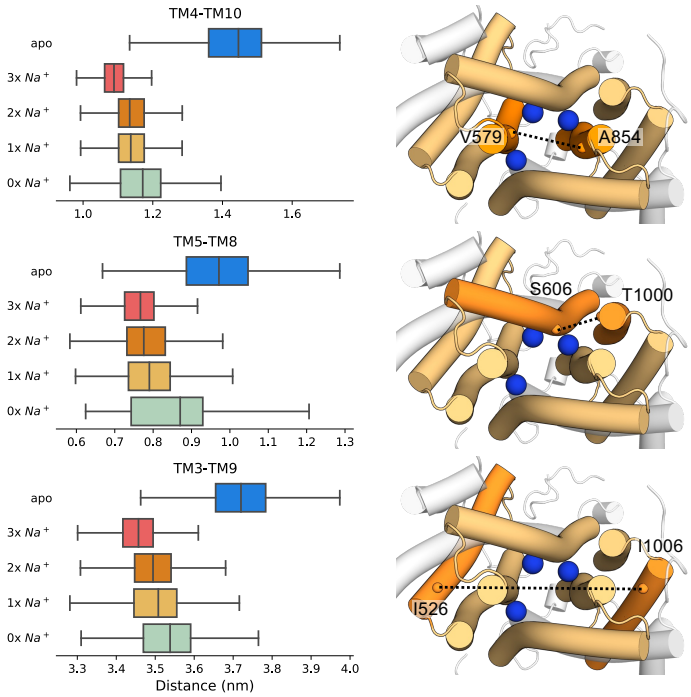


Figure 3.14: Expansion of the intracellular TMD helices.

Minimum distances between the intracellular portions of DISP1 transmembrane helices across 3 x 100 ns atomistic simulations of DISP1 in distinct ion bound states. Intracellular view of Cα atoms (spheres) on labelled residues of TM4-TM10, TM5-TM8 and TM3-TM9 (orange helices) used in the distance calculations.

3.4 Conclusion

3.4.1 The SHH-cholesterol inhibits PTCH1

Using a combination of structural, biochemical and simulation methodologies we demonstrate that the SHH-cholesterol moiety contributes to PTCH1 inhibition by binding to the SBD[220]. The upper portion of the ECD is structurally divergent across the RND family, compared to the conserved lower ECD fold (Fig. 3.15). This suggests RND members have adapted for distinct functions by forming unique modular domains, such as the SBD, at the ECD apex. The distinct SBD conformations between PTCH1-molA and PTCH1-molB result in divergent pathway formation throughout the ECD, as suggested from analysis of dynamic tunnel formation. In all current cryo-EM structures of PTCH1, the SBD conformation can be grouped into two subcategories depending on whether the top of the SBD pocket is open (as in the PTCH1-molA bound 'SHH-cholesterol' conformer) or closed (as in the PTCH1-molB bound 'free cholesterol' form)[213]. Thus, the SHH-cholesterol inhibits PTCH1 by restricting clam-shell-like rocking between the two SBD conformations, which collapses the ECD tunnels. These data also support CRISPR-based biochemical experiments suggesting cholesterol is the principal ligand under PTCH1 regulation[54]; SHH may have evolved to inhibit PTCH1 by directly conjugating the PTCH1 substrate to its C-terminus.

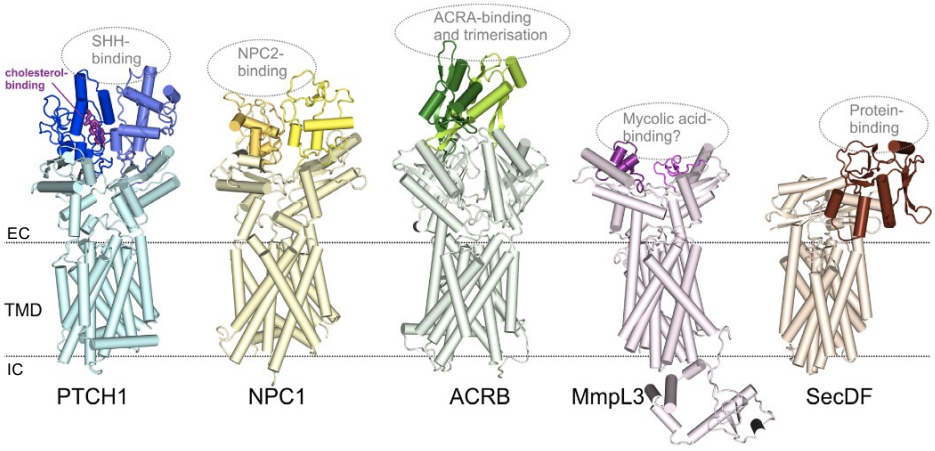


Figure 3.15: Resistance-nodulation-division (RND) family structures.

Comparison of RND family structures (PDBs: PTCH1 6DMY, NPC1 5U73, AcrB 1IWG, MmpL3 6AJF, SecDF 3AQP), highlighting the conserved fold of the TMD and lower ECD region and divergence of the upper ECD. Figure reproduced from Rudolf *et al* [220].

It is possible that the SHH lipids may differentially contribute to HH signalling across the SHH gradient[240,241]. For example, HH signalling closer to the site of SHH release may involve PTCH1 inhibition by both SHH lipids more frequently, since the concentration of SHH is higher. This could e.g. result in prolonged pathway activation compared to more distant cells which may only engage one SHH lipid moiety.

3.4.2 Models for PTCH1 function

Multiple models have been proposed for PTCH1 function and inhibition of SMO (Fig. 3.16). All models integrate current biochemical data suggesting PTCH1 functions by reducing levels of accessible membrane cholesterol available to bind and activate SMO[21,54,57]. These include: removal of cholesterol from the SMO CRD, and import through PTCH1 to a sequestered pool of cholesterol (e.g. in complex with sphingomyelin) or intracellular protein/vesicular acceptor (model-1), extraction of accessible cholesterol from the membrane via export through PTCH1 (model-2), or an ECD-independent PTCH1 function to partition cholesterol (model-3).

Our free energy calculations suggest cholesterol movement in the import direction (model-1) is energetically favourable whereas export (model-2) is associated with an energetic barrier that would require i.e. ion coupling for favourability. Model-1 is not consistent with mutational studies implicating the anionic triad residues in PTCH1 function[53,57] or PTCH1 regulation by cation gradients[21,216,218]. Moreover, it is unclear how PTCH1 would extract cholesterol from SMO given their distinct localisations[15,40] and an absence of known extracellular cholesterol shuttles. NPC1 however, functions without ion-coupling to spontaneously import cholesterol across the lysosomal membrane[242]. Thus, our free energy calculations can be used to rationalise the energetics of NPC1 mediated transport.

Our atomistic simulations suggest the barrier for cholesterol transport may be overcome by coupling to cation translocation through the TMD. To estimate the cholesterol:cation transport stoichiometry we calculated the free energy across the membrane as a function of membrane potential (Fig. 3.17A, Equation 3.3). Standard cellular Na^+/K^+ concentrations are assumed, and a cholesterol export ΔG_1=20 kJ mol$^-$

[1] was used to estimate the number of coupling ions (Fig. 3.17B). Accordingly, 1-3 Na^+ ions (model-2a) or 2-3 K^+ ions (model-2b) may be required for cholesterol export although K^+ coupled transport could only occur if the cilia membrane is more depolarised than predicted (i.e. ~-20mV). The cilia membrane potential has been suggested to be ~30 mV less electronegative[243] however how this is established independently of the neighbouring plasma membrane remains ambiguous.

$$[3.3] \qquad \Delta G = RTln\left(\frac{[X^+]_{in}}{[X^+]_{out}}\right) + zF\Delta V$$

$$[Na^+]_{in} = 12\,mM,\ [Na^+]_{out} = 145\,mM,\ [K^+]_{in} = 150\,mM,\ [K^+]_{out} = 4\,mM$$

Sterol-like densities within the ECD of PTCH1 structures[20,57,122,212–214] support the proposed export model however, this model is juxtaposed to one long-standing observation in the HH field; cleavage of PTCH1 ECD2 renders PTCH1 SHH unresponsive but does not affect PTCH1 function[53]. Model-3 proposes the PTCH1 TMD could function independently to e.g. partition accessible cholesterol towards an intracellular acceptor or into sphingomyelin-sequestered regions[21]. Given cholesterol flip-flop occurs rapidly and spontaneously in cells[244], the free energy barrier for inter-leaflet cholesterol repartitioning would presumably be much lower, and could be more readily coupled to the reduced free energy across the K^+ gradient at ~60 mV. Hence the incompatibility of the cholesterol export and K^+ potential free energies (at -60 mV) could be used as further evidence against an ECD-dependant PTCH1 mechanism.

Mechanistically, our insights from simulations of DISP1 suggest ion coupling may proceed through an occluded intermediate, formed by movement of the intracellular TM portions to exclude solvent entry into the TMD. Further work is needed to elucidate PTCH1/DISP1 intermediates such as potential outward-facing conformations. A further possibility raised by the experimentally observed K^+ dependence[21,216], is that K^+/Na^+ may exchange within the TMD to facilitate ion release, analogous to H^+ mediated Na^+ release from the Na^+/K^+ ATPase[245].

Finally, I touch upon one intriguing observation across the RND family. Why are prokaryotic RND small-molecule exporters dimeric/trimeric whereas lipid/sterol transporters (MmpL3, PTCH1, NPC1, DISP1[122,246–248]) are monomeric? Did lipid

transport activity evolve from a common, monomeric RND ancestor or is there an unexplored facet of lipid transport that is more permissible in the monomeric form? Clever use of chimeric proteins, cross-linking and bioinformatics are needed to unpick this evolutionary puzzle.

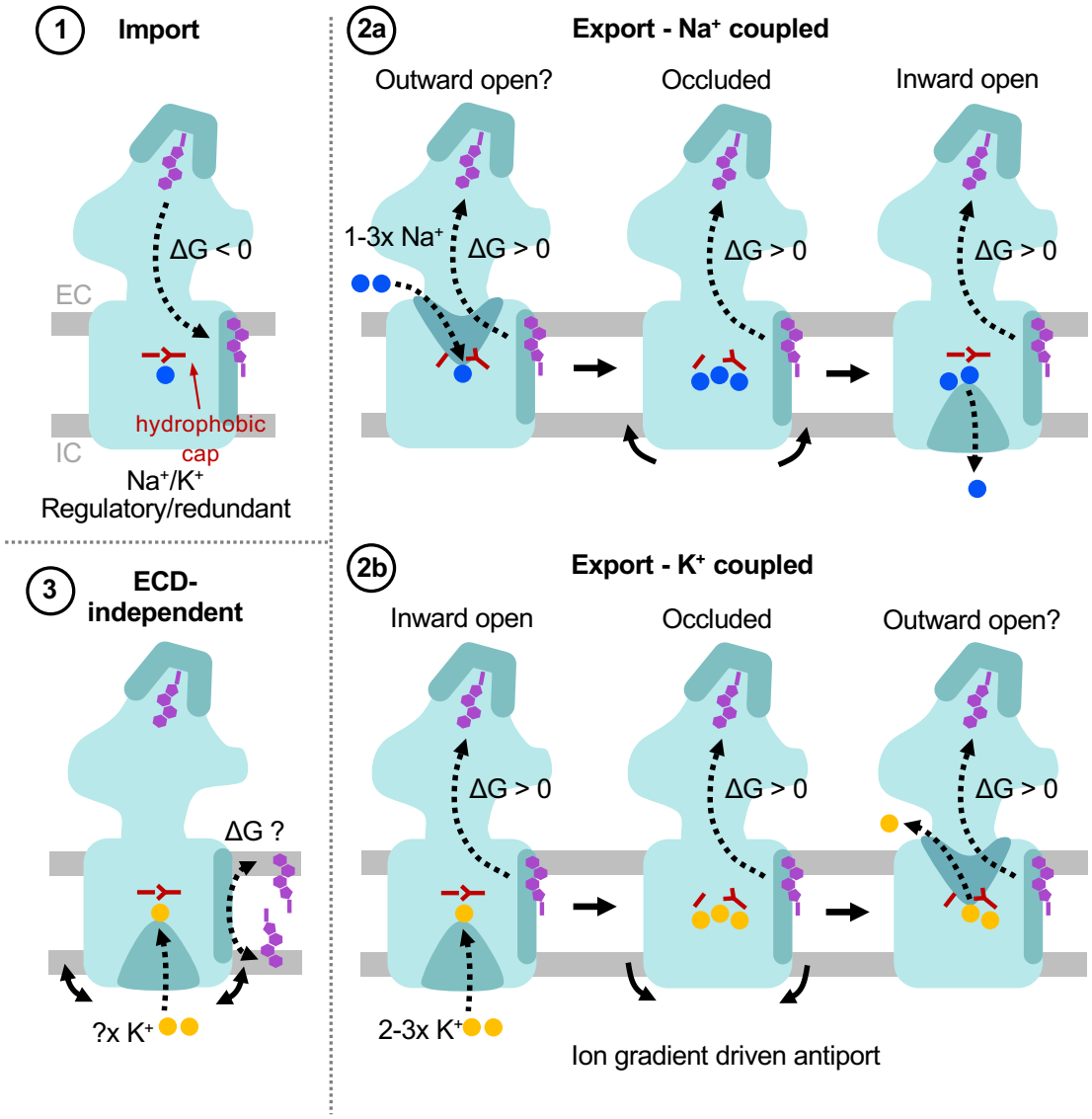


Figure 3.16: Models for PTCH1 function.

PTCH1 may modulate the levels of accessible membrane cholesterol available to bind and activate SMO via one of the proposed models: energetically favourable cholesterol import through PTCH1 (model-1), cation coupled cholesterol export (model-2) or an ECD-independent mechanism for cholesterol re-partitioning (model-3). Predicted free energies and cation coupling stoichiometries for each model are labelled.

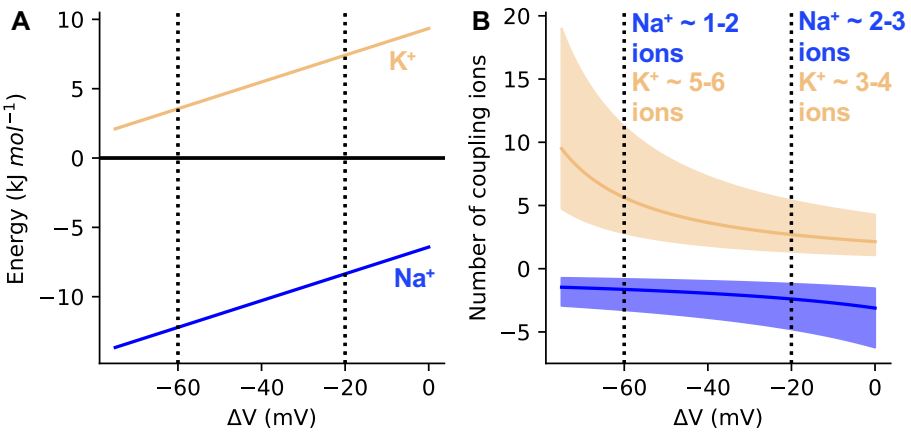


Figure 3.17: Free energy and coupling stoichiometry of Na^+/K^+.

A) Free energy stored across the Na^+ and K^+ gradients as a function of membrane potential (ΔV) (Equation 3.3). Standard cellular ion concentrations are assumed ($[Na^+]_{in}$: 12 mM, $[Na^+]_{out}$: 145 mM, $[K^+]_{in}$: 150 mM, $[K^+]_{out}$: 4 mM). **B)** Number of Na^+/K^+ ions required to drive cholesterol export (defined as 'cholesterol export free energy'/'free energy across Na^+/K^+ gradient') as a function of membrane potential. The cholesterol export free energy was taken as +20 kJ mol^{-1} (solid line) with values between +10 to +40 kJ mol^{-1} in transparent.

4. SMO

"One thing I've learn in the woods is that there is no such thing as random. Everything is steeped in meaning, coloured by relationships, one thing with another."
"We spill over into the world and the world spills over into us."
Robin Wall Kimmerer

4. SMO 94

4.1 Introduction 95

4.2 Methods 98
4.2.1 Constructs used in atomistic simulations 98
4.2.2 Atomistic simulations 99
4.2.3 Allosteric pathways 99
4.2.4 Targeted MD simulations 99

4.3 Results and Discussion 100
4.3.1 Investigating SMO conformational stability 100
4.3.2 Allostery between cholesterol binding sites 102
4.3.3 Inspecting the proposed 'mid' site cholesterol 105
4.3.4 ECL3 displacement during activation 107
4.3.5 Conformational clustering 109

4.4 Conclusion 113
4.4.1 PTCH1 inhibits SMO by regulating cholesterol binding to the CRD 113
4.4.2 SMO dynamics upon activation 114

This chapter contains data also published in:

Kinnebrew, M., Woolley, R.E., **Ansell, T.B.**, Byrne, E.F.X., Frigui, S., Luchetti, G., Sircar, R., Nachtergaele, S., Mydock-McGrane, L., Krishnan, K., Newstead, S., Sansom, M.S.P., Covey, D.F., Siebold, C. and Rohatgi, R. (2022). Patched1 regulates Smoothened by controlling sterol binding to its extracellular cysteine-rich domain. *Sci. Advances*, 8, eabm5563

4.1 Introduction

Smoothened (SMO) is a Class F GPCR responsible for transmission of HH mediated signalling into the intracellular environment[2]. SMO is comprised of a ~140 residue N-terminal extracellular cysteine-rich domain (CRD), a linker domain which connects the CRD to the canonical GPCR 7 TM helix bundle (TMD) and an intracellular cytosolic domain of indeterminate structure[48] (Fig. 4.1). Structures of SMO have been determined in both inactive[48,83–85,93] and active conformations[94,210,249,250] however the cellular signalling mechanism remains largely enigmatic. Interpreting the biological relevance of GPCR conformations is complex and thus 'active'/'inactive' are used tentatively throughout, based on consensus GPCR literature. Active state structures of SMO are characterised by the outward movement of TM6 and inward displacement of TM5 to form a cytoplasmic cavity in which partner proteins bind[94,210,249,250]. Unlike other GPCRs, including other Class F Frizzled receptors, SMO TM6 movement proceeds without formation of an interhelical kink, suggested to result from loss of a proline residue ($P^{6.43}F$, Ballesteros-Weinstein numbering[251]) present in other Class F GPCRs[98].

In structures of SMO resolved using cryo-EM the CRD is often not visible[210,250], suggestive of inherent flexibility also seen in MD simulations[48,93]. By contrast, relatively minor CRD rearrangements are observed within X-ray crystallography structures[46] except for one study which resolved CRD tilt[249] by ~90°, however this conformation has been suggested to be an artifact resulting from mutation of a key glycosylation site on ECL3 and/or crystal contacts[94]. Arguably the most intriguing feature observed

in SMO structures is the presence of a cholesterol molecule bound within the CRD[48]. The CRD-cholesterol is positioned ~12 Å beyond the membrane plane housed by predominantly hydrophobic residues surrounding the tetracyclic rings and Y130 coordinating the β_3-hydroxyl group. Structures of SMO observe binding of other small molecules such as oxysterols and cyclopamine to the CRD[49,249,250] (Fig. 4.1A). ECL3 of SMO is extended by comparison to other GPCR classes and packs against the CRD sterol binding site leading to the proposal that the CRD site might be implicated in TM6 activation and signal transduction via ECL3[48,93].

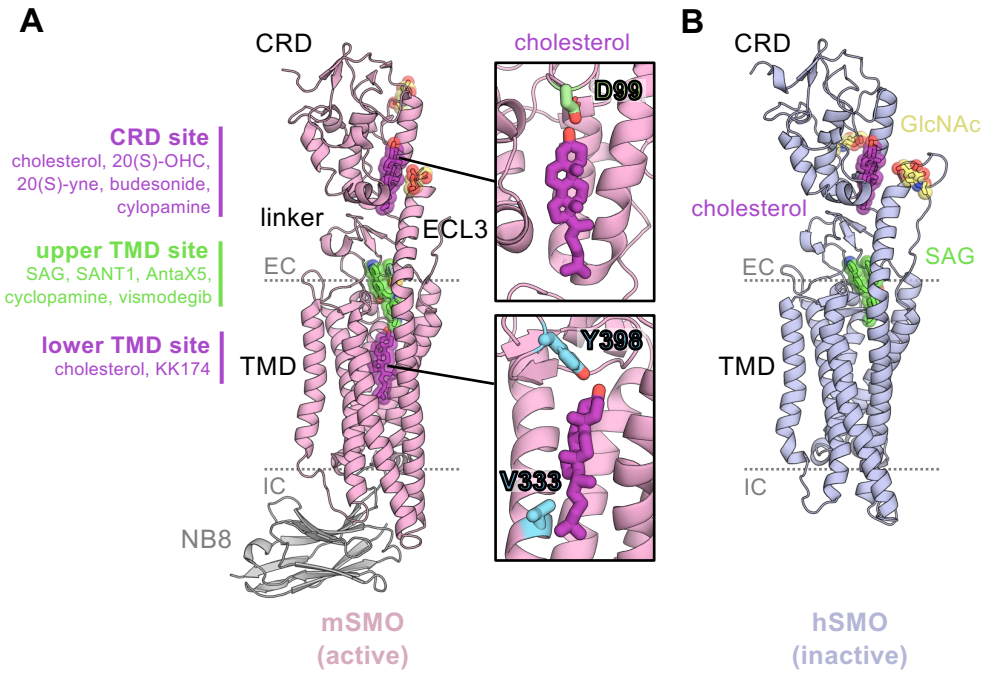


Figure 4.1: Smoothened (SMO) structures.

Structures of **A)** mouse SMO (mSMO, PDB: 6O3C)[94] in an active conformation and **B)** human SMO (hSMO, PDB: 7ZI0)[252] in an inactive conformation. The location of ligand binding pockets within the cysteine-rich domain (CRD) and transmembrane domain (TMD) are marked. Cholesterol (purple), SMO agonist (SAG, green) and glycosylation sites (yellow) are shown as sticks. Inset show coordination of cholesterol by D99 and Y398 within the CRD and lower TMD sites.

In addition, SMO has several other ligand binding sites. On the extracellular side, the TMD bundle encloses a binding site for anti-cancer drugs (Vismodegib), SMO agonists (SAG), antagonists and oxysterols (equivalent to the orthosteric binding pocket in other GPCRs)[48,83–85,93,94]. A second site, located deeper with the TMD has been

observed to bind cholesterol[94] and an intermediary oxysterol binding site has been proposed between the upper TMD site and the CRD site, in proximity to the linker[250]. In addition, simulations predict a peripheral cholesterol binding site between the extracellular portion of TM2/TM3[99–101]. A docking study also suggested cholesterol may bind within the cytoplasmic TMD cavity[209]. These structural studies provide a rich platform for understanding SMO ligand binding properties, however positioning these data within the cellular context is complicated. Which ligands binding to which sites are implicated in SMO signalling *in vivo*? Does SMO engage different sites for basal *vs* Hedgehog mediated signalling or for transmission to different downstream partner proteins? The latter is complicated by the recent emergence of G-protein independent SMO pathways transduced by binding of Protein Kinase A (PKA) to the SMO C-terminal tail[23,24].

Site directed mutagenesis and ligand affinity assays have confirmed binding of sterols to the CRD and show that cholesterol is both necessary and sufficient for SMO signalling in cells[22,49,253]. In CRISPR-based genetic screens cholesterol biosynthetic genes, and not other oxysterols or cholesterol precursors, emerge as regulators of pathway activation[54], further implicating cholesterol as the endogenous SMO activator. Conversely, mutations of the upper TMD ligand binding site are found in drug-resistant cancer patients[32], suggesting this site is less important for cellular SMO signalling. Mutagenesis of residues surrounding the lower TMD site also diminish SMO activity in cells[48,94,98]. Together these data implicate cholesterol as the most probable ligand for SMO activation, by binding to the CRD and/or lower TMD sites. This study focusses on untangling the interplay between these sites and addressing their respective roles in cellular HH signalling.

This work was conducted in collaboration with members of the Siebold and Rohatgi groups and aims to address the following questions:

1) Which ligand binding site(s) on SMO facilitate pathway activation in cells?
2) How are signalling changes transduced through SMO?

4.2 Methods

4.2.1 Constructs used in atomistic simulations

A structure of mouse SMO (mSMO) in an active state bound to CRD and TMD cholesterols, SAG and nanobody (NB8) was obtained from the PDB (PDB ID: 6O3C)[94]. ECL3 was modelled using the mouse SMO sequence based on an existing human SMO (hSMO) structure (PDB ID: 5L7D)[48] using Modeller9.20[221]. PropKa[228] was used to identify three residues with non-standard pKa's (E485, E522 and H73) which were modelled accordingly. The following systems were constructed for simulation: 1) mSMO bound to the nanobody, both cholesterols and SAG, 2) mSMO bound CRD-cholesterol and SAG, 3) mSMO with both CRD- and TMD-cholesterols bound, 4) mSMO bound CRD-cholesterol, 5) mSMO bound TMD-cholesterol an 6) mSMO in its *apo* form. An additional system of mSMO bound to three cholesterols (with the third cholesterol occupying a 'mid' position in proximity to the linker region) was constructed based on the position of the upper 24(S),25-epoxycholesterol in PDB ID: 6XBM[250]. Systems of inactive human SMO (hSMO) bound to CRD-cholesterol and SAG and hSMO bound two cyclopamine molecules were constructed using crystal structures from the Siebold group (PDB ID: 7ZI0[252], and unpublished). The V329F mutants were reverted to wild-type and ECL3 was built based on the position in an existing structure (PDB ID: 4JKV)[83] using Modeller9.20[221].

Each system was embedded in a 9 x 9 nm^2 POPC:cholesterol (3:1) bilayer using CHARMM-GUI[254]. Protein components were described using the AMBER-ff14SB forcefield[132]. Glycan GlcNAc attachments were modelled on N80 and N497 of mSMO and N188 and N493 of hSMO. Glycans were described using the GLYCAM-06j forcefield with associated amino acid attachment parameters[138]. SAG and cyclopamine parameters were derived using antechamber and the generalised AMBER forcefield (GAFF2[137]) with AM1-BCC used to optimise charge interactions. Lipids (including bound cholesterols) were described using the AMBER-lipid17 forcefield[136]. The system was solved using TIP4P-Ew water[255] and approximately 0.15 M NaCl.

4.2.2 Atomistic simulations

Simulations were performed using GROMACS 2019[155]. Each system was minimised using a steepest descent method before NVT and NPT equilibration steps of 5 ns each with restraints applied to protein, glycan and ligand coordinates when present. 5 x 300 ns atomistic simulations were performed for each system. A 2 fs timestep was used and periodic boundary conditions applied. Temperature was maintained at 310 K using the Nosé-Hoover thermostat[231,232] and a τ_T = 0.5 ps coupling constant. The Parrinello-Rahman barostat[226] with a τ_P = 2.0 ps coupling constant and a compressibility of 4.5 x 10^{-5} bar^{-1} was used to maintain pressure at 1 bar. Electrostatics were described using the particle-mesh Ewald[233] (PME) method with a 1.2 nm cut-off. Van der Waals interactions were modelled with a 1.2 nm cut-off using the Verlet method. Long range dispersion corrections were applied for energy and pressure. All bonds were constrained to the equilibrium lengths using the LINCS algorithm[227].

4.2.3 Allosteric pathways

Allosteric pathways were calculated using *allopath*[256] (github.com/delemottelab/allosteric-pathways). Each residue in mSMO was assigned to a node. Source indices were defined as the Cα atoms of residues lining the CRD (D99, L112, W113, Y134, I160, V161, I500). Sink indices were defined as the Cα atoms of the proposed ionic-lock residues on α6/α7 (R455, F459, W539). The heavy atoms of each cholesterol molecule were considered to comprise a lipid ('interactor') node. A cut-off value of c=0.45 nm and a standard deviation of σ=0.138 nm was used to define heavy atom contacts during construction of the contact map, in accordance a previously published method[256]. Information flow was calculated in simulations of mSMO initiated from a ligand free *apo* state, with cholesterol bound to either the CRD or TMD or with both sites occupied.

4.2.4 Targeted MD simulations

Two sets of targeted MD simulations were performed to drive hSMO conformations from an inactive to active-like state or vice versa. The inactive hSMO structure bound

to SAG and CRD-cholesterol was used as input for TMD simulations towards an active conformation (PDB ID: 7ZI0[252]). Systems were constructed, minimised and equilibrated identically to described above for equilibrium atomistic simulations. PLUMED[257,258] was used to implement targeted MD in GROMACS[155]. The collective variable was the RMSD of TM5-TM6 (mSMO residues N400-F479) Cα atoms between the hSMO structure and the active mSMO structure (PDB ID: 6O3C[94]). The reference region to fit the structures was the Cα atoms of TM1-TM4 (mSMO residues E228-K399). A moving restraint was applied over 20 ns to increase the force constant to k=10000 kJ mol^{-1}, after which restraints were maintained for a further 20 ns. In total 16x 40 ns replicates were performed for the inactive to active targeted MD simulations.

For the active to inactive simulations a homology model of hSMO in an active conformation was generated using the mSMO structure and Modeller9.20[221]. SAG and CRD-cholesterol were added based on their positions in the inactive state and systems constructed identically to described above. The collective variable was defined as the RMSD of TM5-TM6 Cα atoms between the active hSMO model and the inactive hSMO structure. Targeted MD simulations were performed as above for the inactive to active runs. A total of 8x 40 ns replicates were run for the active to inactive simulations.

4.3 Results and Discussion

4.3.1 Investigating SMO conformational stability

V333 sits at the base of the mSMO lower TMD cholesterol binding site and mutation (V333F) prevents mSMO activation[94] (Fig. 4.1A). Authors suggest this implicates the TMD-cholesterol in SMO signalling, however the V333F mutant could simply stabilise SMO in an inactive state. Consistent with this, mutation of the equivalent hSMO residue (V329F) has been used in crystallography to stabilise inactive conformations[48]. Members of the Siebold group solved the structure of hSMO-V329F bound to CRD-cholesterol and SAG in the upper TMD site to 3.0 Å (Fig. 4.1B). hSMO-V329F adopts an inactive conformation, similar to previous structures despite the presence of SAG, a potent SMO agonist[85].

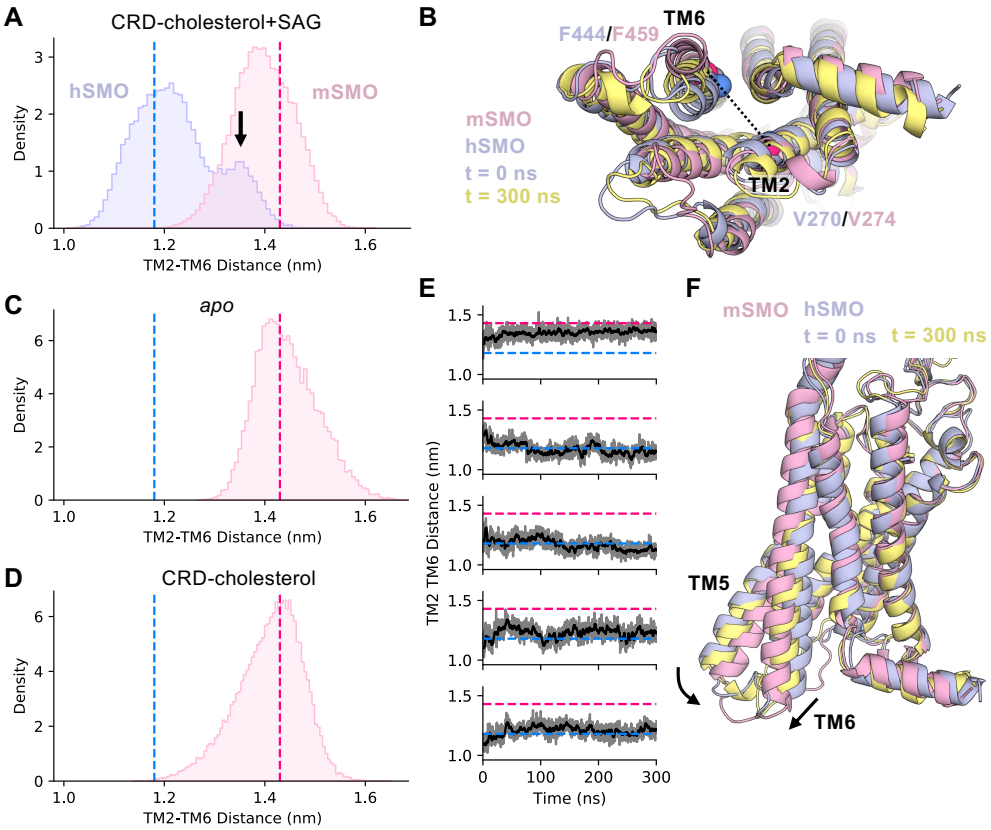


Figure 4.2: SMO stability in simulations.

TM2-TM6 minimum distances between Cα atoms of residues shown in **B** across atomistic simulations (5 x 300 ns) of **A)** mSMO (pink) and hSMO (blue) bound CRD-cholesterol and SAG, **C)** *apo* mSMO and **D)** mSMO bound CRD-cholesterol. **E)** TM2-TM6 distances with time across hSMO simulations bound CRD-cholesterol and SAG. TM2-TM6 distances in the structure are indicated by lines in pink (mSMO, active) and blue (hSMO, inactive). **B)** and **F)** Comparison of hSMO TM helices from the start (blue) and end (yellow) of replicate 1 in **E** with the mSMO structure (pink), showing partial activation of hSMO.

To further assess the role of the TMD-cholesterol site in SMO conformational equilibria we performed atomistic simulations of glycosylated mSMO (active) and hSMO (inactive) bound to CRD-cholesterol and SAG but with the TMD-cholesterol site unoccupied and the hSMO-V329F mutation reverted to WT. mSMO remained in an active conformation even when the TMD-cholesterol was absent (Fig. 4.2A), as probed using the TM2-TM6 distance (Fig. 4.2B). Similar population distributions were observed for mSMO simulations with just the CRD-cholesterol bound or when *apo* (Fig. 4.2C-D). In contrast in hSMO simulations a bimodal TM2-TM6 distance distribution was observed, resulting from partial activation of hSMO in one replicate (possibly induced by SAG) (Fig. 4.2A,E). To check this sub-population did not result from erroneous movement of TM2/TM6 segments and/or adjoining loops, we

compared the start/end snapshots from this replicate with the mSMO structure. Overlay shows concerted outward movement of the whole TM6 helix and inward movement of TM5, consistent with an intermediary conformation (Fig. 4.2B,F). Together, this suggests the TMD-cholesterol is not required to maintain the active mSMO conformation and that the V329F/V333F mutant may function by shifting the conformational equilibria towards inactive states.

4.3.2 Allostery between cholesterol binding sites

To better understand the role of the CRD- and TMD-cholesterol sites in mSMO activation, members of the Rohatgi lab generated D99A and Y398F mutants to disrupt cholesterol coordination (Fig. 4.1A). By evaluating the fold-increase in Gli1 mRNA upon addition of SHH to cells expressing mSMO mutants, they found the increase in mSMO-D99A signalling was reduced compared to SHH induced signalling in WT cells whereas the mSMO-Y398F mutant was unaffected[252] (Appendix Fig. S4.1A). Addition of increasing concentrations of SHH confirmed similar maximal changes to Gli1 mRNA levels for WT and Y398F constructs whereas the D99A mutant had a reduced maximal fold (Appendix Fig. S4.1B). There was some residual signalling in the D99A mutant at high concentration of SHH, which was abrogated in the double D99A/Y398F mutant (for additional controls see Kinnebrew *et al.*). This suggests that cholesterol binding to the CRD (disrupted by the D99A mutant) is the primary site under regulation by PTCH1 but the TMD-cholesterol site may mediate some basal signalling (also under SHH regulation). The non-additive decrease in SHH induced signalling in the D99A/Y398F mutant is also suggestive of allosteric communication between CRD- and TMD-cholesterol sites.

To further probe the suggested allostery we calculated the root mean square fluctuation (RMSF) of mSMO bound to either the CRD- or TMD-cholesterols, with both sites occupied or when *apo* across atomistic simulations (Fig. 4.3). The RMSF of the CRD was similar in the *apo* and singly occupied states but reduced when both cholesterols were bound, supporting the suggested CRD/TMD site cooperativity.

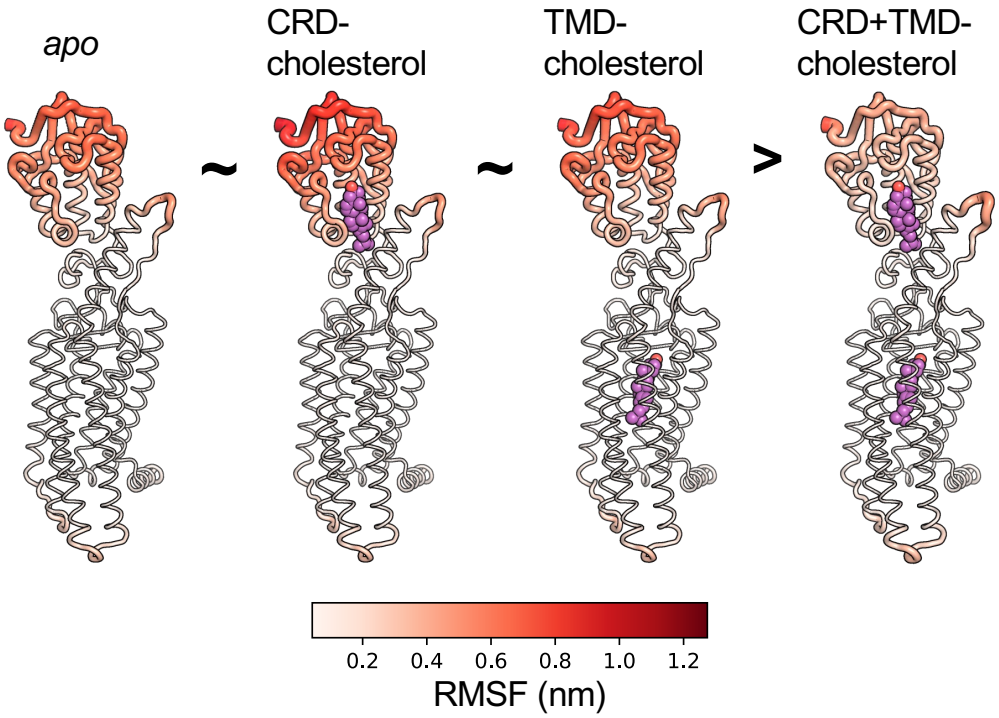


Figure 4.3: Investigating SMO allostery.

Per residue root mean square fluctuation (RMSF) of mSMO across (5 x 300 ns) atomistic simulations in *apo*, CRD-cholesterol, TMD-cholesterol and CRD+TMD-cholesterol bound states. Trajectories were fitted to the TMD.

To further probe communication between CRD- and TMD-cholesterol sites we calculated the flow of information through mSMO using *allopath*[256]. In the allosteric network, information flow is calculated between neighbouring nodes (protein residues) based on the degree of correlated entropic fluctuation around the node centroid (subsequently converted to 'mutual information' based on the fluctuation density). *Allopath* extends this network by explicit inclusion of ligands as nodes, thus computed information pathways encompass both protein and ligand components[256]. Information flow between source nodes surrounding the CRD-cholesterol site and sink nodes on TM6/7 bifurcated through either i) ECL3 or ii) a ladder of residues on the opposing CRD face and the linker region (Fig. 4.4A). When bound, the CRD-cholesterol contributes markedly to information flow, with one of the highest flow values amongst all (protein or lipid) nodes (Fig. 4.4B). The information flow value of the TMD-cholesterol is comparatively moderate. When both cholesterols are bound information flow through the CRD-cholesterol is reduced whereas the TMD-cholesterol is

unaffected, consistent with tuning of CRD-cholesterol mediated signalling by TMD site occupancy (Fig. 4.4B). To determine critical residues responsible for cholesterol induced changes in information flow we calculated the information flow difference

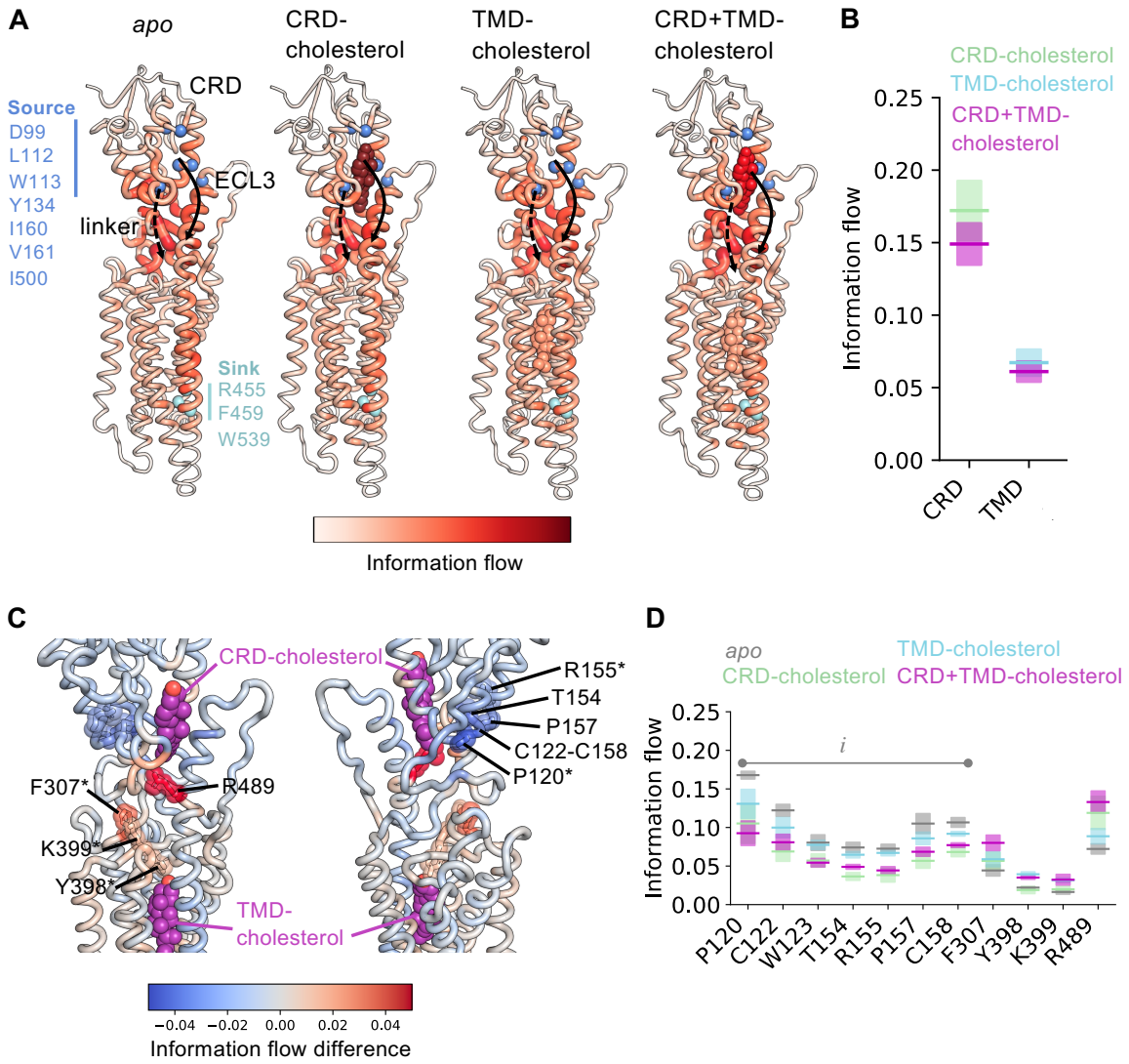


Figure 4.4: Information flow through SMO.

A) Information flow through mSMO across atomistic simulations in *apo*, CRD-cholesterol, TMD-cholesterol and CRD+TMD-cholesterol bound states. Information flow was calculated between source residues (blue) surrounding the CRD-cholesterol site and sink residues (cyan) on TM6/7 using *allopath*[256]. Paths via ECL3 (solid line) or the reverse face of the CRD and linker region (dashed line) are shown and cholesterol is shown as spheres. **B)** Information flow value of CRD- and TMD-cholesterols in single and dually bound states. **C)** Information flow difference between the CRD+TMD-cholesterol bound and *apo* simulations. Residues which affect SMO signalling when mutated are stared. **D)** Information flow values for residues highlighted in **C** in mSMO simulations in distinct states. Residues on the reverse face of the CRD are marked by *i*. Figure adapted from Kinnebrew *et al* [252].

between the CRD+TMD-cholesterol bound and *apo* states (Fig. 4.4C). CRD-cholesterol binding reduces information flow via the opposing CRD face/linker route (Fig. 4.4D, marked as i) and favours flow through a subset of residues on ECL3 and the TMD interface (R489, F307). TMD-cholesterol binding promotes information flow through Y398 and K399 (Fig. 4.4D). The switch in Y398 information flow, further justifies interpretation of mSMO-Y398F biochemical assays, given it is difficult to directly measure whether Y398F disrupts cholesterol binding *in vivo*.

Together the biochemical and computational analyses suggest the CRD-cholesterol site is under principal regulation of SHH (via PTCH1 inhibition), and the TMD-cholesterol tunes signalling by regulating the degree of information flow propagating from the CRD site. This is conducive with allosteric cooperatively between the cholesterol binding sites.

4.3.3 Inspecting the proposed 'mid' site cholesterol

In addition to the TMD- and CRD-cholesterol binding sites, a third sterol site has been proposed in proximity to the linker region (Fig. 4.5A). This 'mid' site was modelled as 24(S),25-epoxycholesterol (24(S),25-EC) in a structure of active hSMO[250], although the cryo-EM density is nebulous (Fig. 4.5B). Authors propose this site may comprise a conduit for cholesterol movement between CRD and TMD sites.

To test this site, we performed atomistic simulations of mSMO with the mid-cholesterol bound in addition to the CRD- and TMD-cholesterols. The RMSD of the mid-cholesterol is considerably higher than cholesterol at either of the other sites (Fig. 4.5C). The RMSD of the TMD-cholesterol is also higher than in all other simulations with TMD-cholesterol bound (Fig. 4.5D), suggesting the instability of the mid-cholesterol propagates to the lower TMD site. We calculated the prevalence of hydrogen bond formation between Y398 and the TMD-cholesterol across simulations (Fig. 4.5E). Binding of the stabilising nanobody (NB8) or CRD-cholesterol to mSMO increased the degree of hydrogen bonding compared to when only the TMD-cholesterol was bound. Conversely, the hydrogen bonding prevalence decreased when the 'mid' site was also occupied, likely induced by the mid-cholesterol instability.

Finally, calculating the number of waters within 0.4 nm of cholesterols, reveals high solvent exposure for the mid-cholesterol, which would presumably be unfavourable given its hydrophobicity (Fig. 4.5F). These data do not strongly support the presence of a third cholesterol binding site between the CRD and TMD and cast some doubt on the interpretation of modelling in the hSMO-24(S),25-EC structure[250].

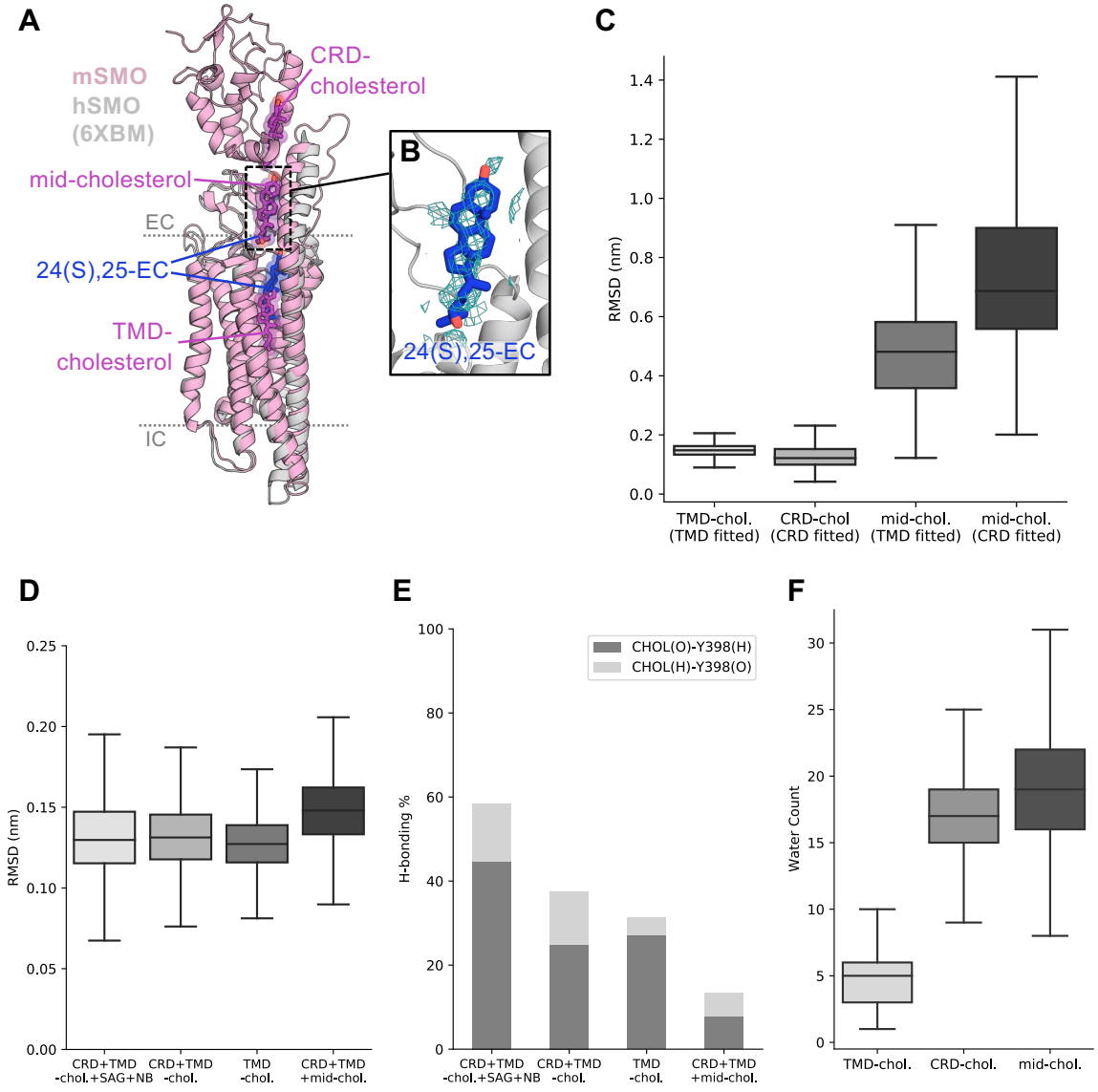


Figure 4.5: Debunking the 'mid' cholesterol site.

A) Structural overlay of hSMO (grey) bound to two 24(S),25-epoxycholesterol (24(S),25-EC, blue) molecules (PDB: 6XBM)[250], positioned within the upper TMD and a proposed 'mid' site, with mSMO (pink) bound to three cholesterol molecules (CRD-, TMD- and mid- sites, purple). **B)** Close up of density at the proposed 'mid' site in the hSMO structure. **C)** RMSD of the CRD-, TMD- and mid-cholesterol molecules in mSMO simulations (5 x 300 ns) bound to all three cholesterols. **D)** RMSD of the TMD-

cholesterol in mSMO simulations (5 x 300 ns each) in distinct ligand bound states. **E)** Hydrogen bonding prevalence between the hydroxyl group of Y398 and the β_3-hydroxyl group of the TMD-cholesterol across mSMO simulations in distinct ligand bound states, as calculated using MDAnalysis[259]. **F)** Number of water within 0.4 nm of TMD-, CRD- and mid-cholesterols across mSMO simulations with all sites occupied.

4.3.4 ECL3 displacement during activation

While comparing the inactive hSMO and active mSMO structures we noted outward displacement of the extracellular portion of ECL3 from the CRD (Fig. 4.6A). Given ECL3 connects to TM6, and the involvement of this region in information flow (Fig. 4.4) we hypothesised ECL3 displacement may be a SMO activation marker, like motifs in other GPCR classes distinct from TM5/6 movement[260]. We calculated the angle of the ECL3 relative to the TMD plane across mSMO and hSMO simulations (Fig. 4.6B). The ECL3 angle was ~13° greater in all mSMO simulations compared to hSMO. To confirm this feature was not a quirk of the hSMO-CRD-cholesterol-SAG structure, we compared to simulations of hSMO bound to two inhibitory cyclopamine molecules, solved by members of the Siebold group. The ECL3 angle was similarly reduced in the cyclopamine inhibited hSMO simulations, further implicating ECL3 displacement as an activation motif (Fig. 4.6B).

To probe whether ECL3 displacement was concurrent with receptor activation we performed targeted MD simulations of hSMO to induce receptor activation. The collective variable was selected as the RMSD of TM5/6-Cα atoms compared to the mSMO structure (aligning on TM1-4) and therefore excluded the ECL3 region. In the inactive to active-like hSMO targeted MD simulations, TM5/6 were successfully driven towards their positions in the mSMO structure, also monitored by the TM2-TM6 distance with time. Strikingly the ECL3 angle was consistently displaced outward during activation (Fig. 4.7A). We also tried driving conformations in the reverse direction although these were less successful (Fig. 4.7B). Issues may have resulted from a) differences in the relative stability of inactive *vs* active conformations, b) the reliance on an active hSMO homology model as starting structure, c) an incompatible binding pose for SAG/CRD-cholesterol or d) some combination of the above. For this reason, latter results refer to targeted MD towards the active conformation only.

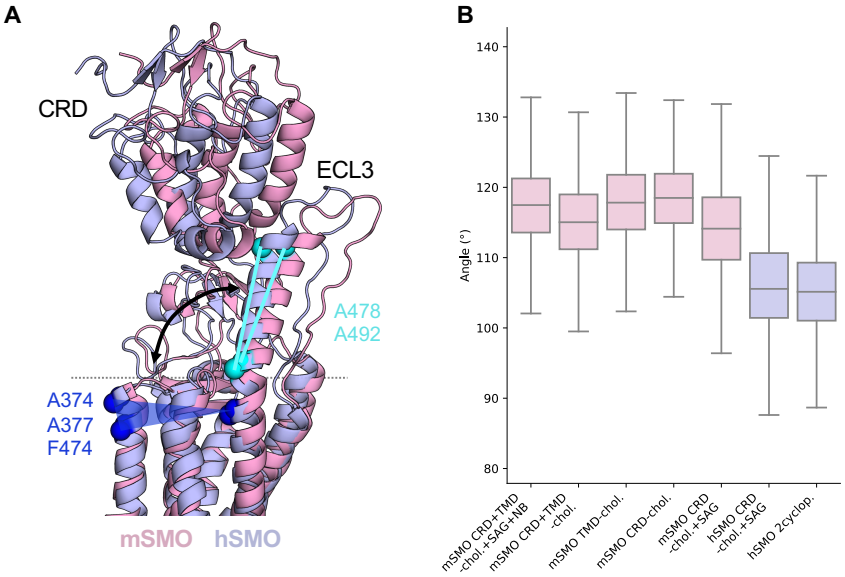


Figure 4.6: Outward tilt of ECL3.

A) Angle between a plane on the TMD (blue) and a vector formed by residue-Cα atoms on ECL3 (cyan) projected on the hSMO and mSMO structures. **B)** ECL3 angle with respect to the TMD across mSMO (pink) and hSMO (blue) simulations (5 x 300 ns of each) in distinct ligand coupled states.

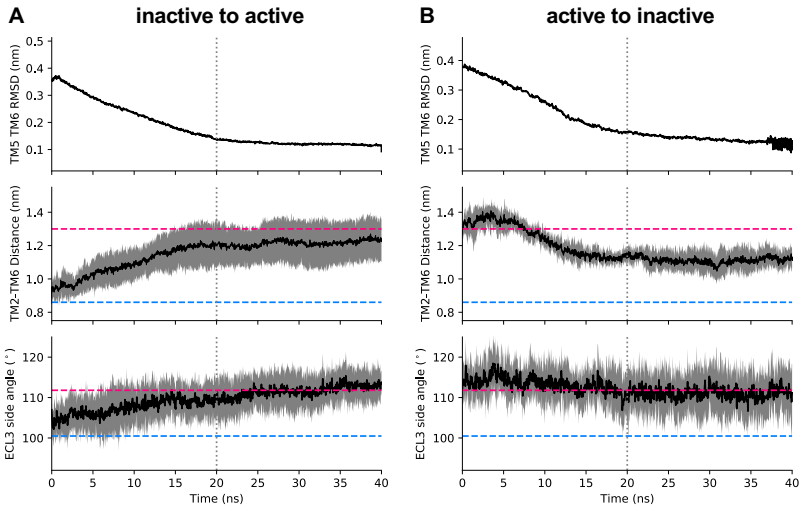


Figure 4.7: Targeted MD simulations of hSMO activation and inactivation.

Targeted MD simulations of hSMO bound CRD-cholesterol and SAG in the **A)** inactive to active and **B)** active to inactive directions. The RMSD of TM5/TM6 Cα atoms was used as the collective variable after aligning on TM1-4. In **A** hSMO was targeted towards the mSMO active structure. In **B** an active homology model of hSMO was targeted towards the inactive hSMO structure. Over the first 20 ns (dashed line) a moving restraint was applied to increase the force constant to 10000 kJ mol^{-1}. Restraints were maintained over the remaining 20 ns of each replicate. The TM2-TM6 distance (Fig. 4.2) and ECL3 angle (Fig. 4.6) were monitored as independent markers of SMO activation/inactivation. Error bars correspond to the standard deviation between the **A)** 16 x 40 ns and **B)** 8 x 40 ns replicates.

4.3.5 Conformational clustering

We sought to further untangle molecular transitions during SMO activation. In equilibrium simulations we tried to identify differences in the CRD behaviour between systems, by exhaustively testing distinct CRD angles relative to the TMD, CRD RMSDs, CRD/ECL3 contacts or CRD-cholesterol/ECL3 contacts. However, CRD motions proved too complex to unpick by relatively simple metric analyses. Thus, we chose to conformationally cluster the targeted MD trajectories and retrospectively assess whether molecular patterns emerge. Specifically, does SMO activation always sample the same conformational transitions during activation?

Clusters were defined using the RMSD of hSMO (after aligning on TM1-4 Cα atoms) at each timepoint whereby snapshots with an RMSD < 0.13 nm were assigned to the same cluster[155] (Fig. 4.8A). The RMSD cut-off was selected to give a moderate number of total clusters (i.e. < 50) (Fig. 4.8B), with 80% of the trajectory accounted for by the top 5 clusters and 90% within the top 8 (Fig. 4.8C). We plotted the cluster number *vs* time for clusters which accounted for > 4% conformations across the trajectory (clusters-1 to -5) and combined for those accounting for 2-4% of conformations (clusters-6 to -8) (Fig. 4.9). Across each of the targeted MD simulations hSMO transitioned through cluster-1, which had an intermediate TM5/TM6 position compared to hSMO and mSMO structures (Fig. 4.9). Overtime, the cluster IDs changed as hSMO moved towards active-like conformations (cluster-2 to -5), validating use of the clustering method on the targeted MD trajectories. Clusters-2 and -3 emerged earlier and were mostly mutually exclusive. Interestingly, clusters-4 and -5 exchanged with cluster-2 but not with cluster-3.

To understand how cluster conformations differ the average/central conformation from each cluster was compared with the mSMO/hSMO structures (Fig. 4.10A). The CRD position differed between active-like clusters, likely accounting for their distinct classification. We also plotted the CRD RMSD compared to hSMO/mSMO structures (Fig. 4.10B), and 2D principal component analyses (PCA) of the TMD (Fig. 4.10C) and CRD (Fig. 4.10D). In the average conformations the CRD of clusters-1 and -3 best matched the hSMO structure. In clusters-2, -4 and -5 the CRD was shifted towards

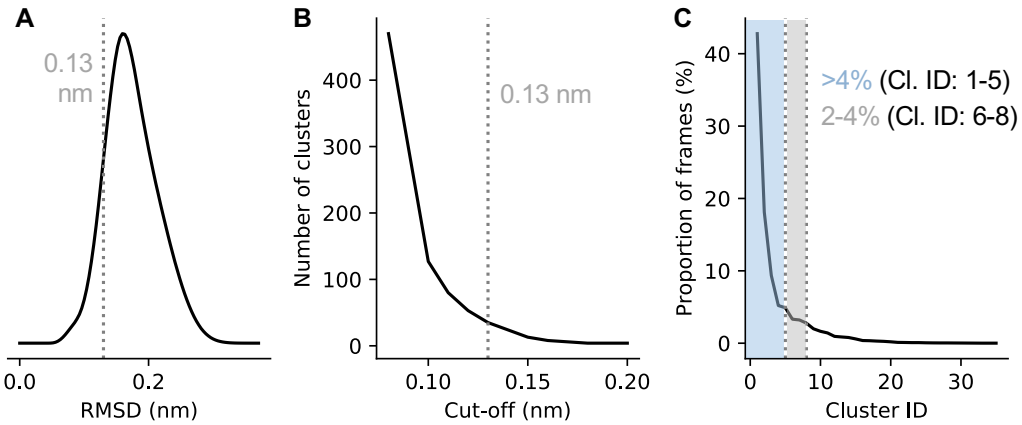


Figure 4.8: Selecting a clustering threshold.

A) RMSD distribution of hSMO after fitting to TM1-4 Cα atoms across the targeted MD simulations in the inactive to active direction. **B)** Number of unique clusters *vs* RMSD cut-off value. A cut-off of 0.13 nm was selected (dashed grey line in **A/B**). **C)** Proportion of trajectory frames attributed to each cluster when clustering was performed using a 0.13 nm cut-off. Clusters accounting for >4% of frames (Cluster ID: 1-5, blue) or 2-4% (Cluster ID: 6-8, grey) are marked.

ECL3 compared to in mSMO/hSMO. However, comparison of the CRD RMSDs and principal components shows the CRD to be highly dynamic and sample a large conformational space across trajectories (Fig. 4.10B/D). Compared to the TMD which presented separable principal component space between the hSMO and mSMO structures or between the intermediate and active-like clusters (Fig. 4.10C), the CRD cluster populations overlapped substantially. Despite this, some trends emerge, notably that in cluster-1 intermediates the CRD was more dynamic than active-like conformations and that cluster-3 overlapped less with clusters-2, -4 and -5, in support of conclusions reached from assessing cluster IDs with time (Fig. 4.9).

In conclusion, targeted activation of hSMO proceeded through a conserved intermediary cluster with a highly dynamic CRD. Trajectories transitioned to two active-like pools with CRDs best matching either the inactive hSMO structure (cluster-3) or shifted towards ECL3 (cluster-2, -4, -5). We also find that the number of CRD-ECL3 contacts is increased at the end of targeted MD simulations, compared to the start (Fig. 4.10E). Combined with our observations of ECL3 outward tilt in active mSMO simulations (Fig. 4.6) and allosteric flow through ECL3 (Fig. 4.4) it is tempting

to speculate whether the shifted CRD position may 'push' against ECL3 to promote TM6 tilt and SMO activation.

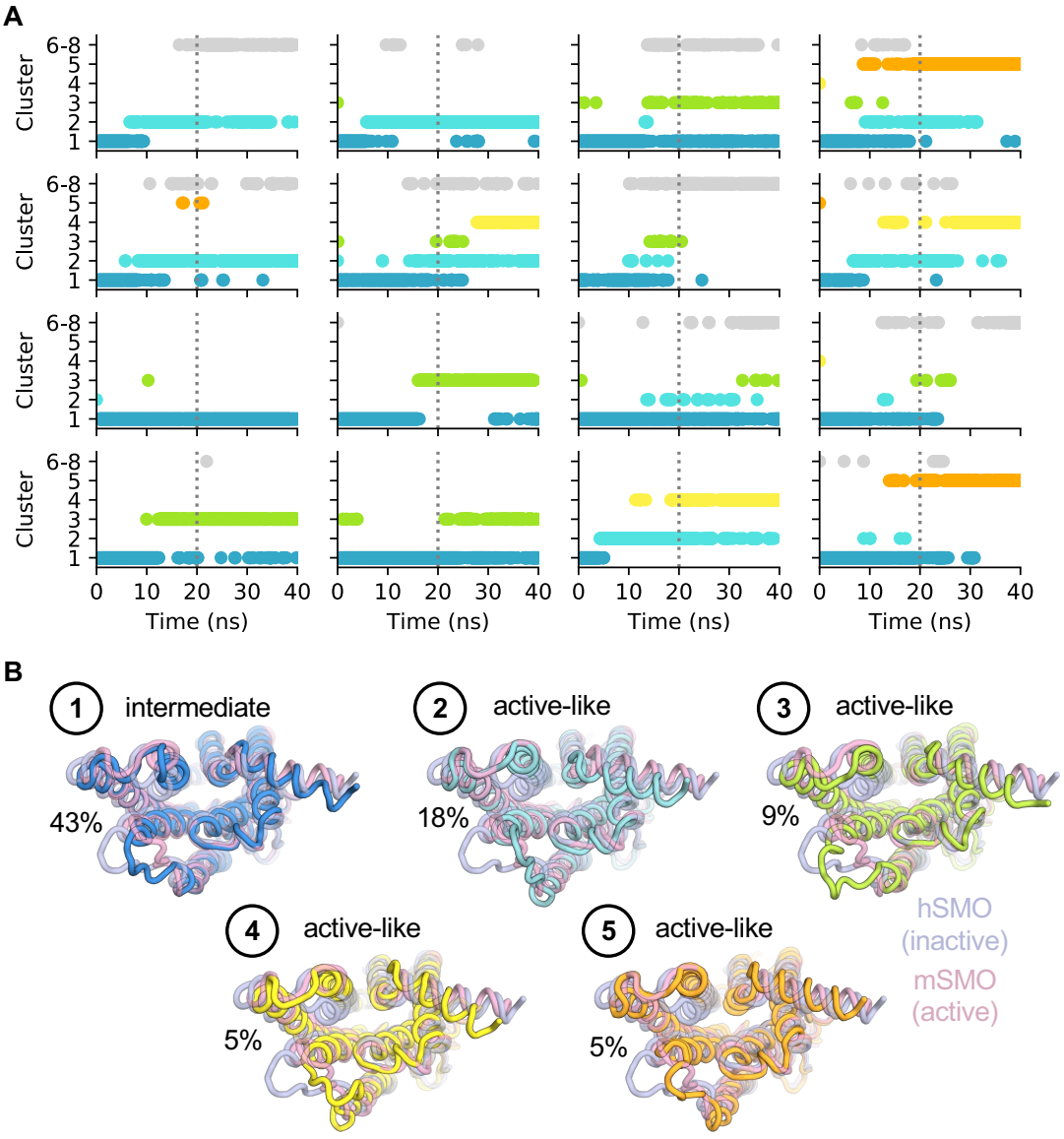


Figure 4.9: Clustering hSMO conformations during activation.

A) Cluster ID *vs* time across 16 x 40 ns targeted MD simulations of hSMO in the inactive to active direction. Cluster IDs 1-5 are coloured individually, cluster IDs 6-8 are combined and indicated in grey. **B)** The central conformation from clusters 1-5 overlayed with the hSMO (blue) and mSMO (pink) structures after aligning on TM1-4 Cα atoms. Clusters were categorised based on the positions of TM5/6 with respect to hSMO/mSMO structures.

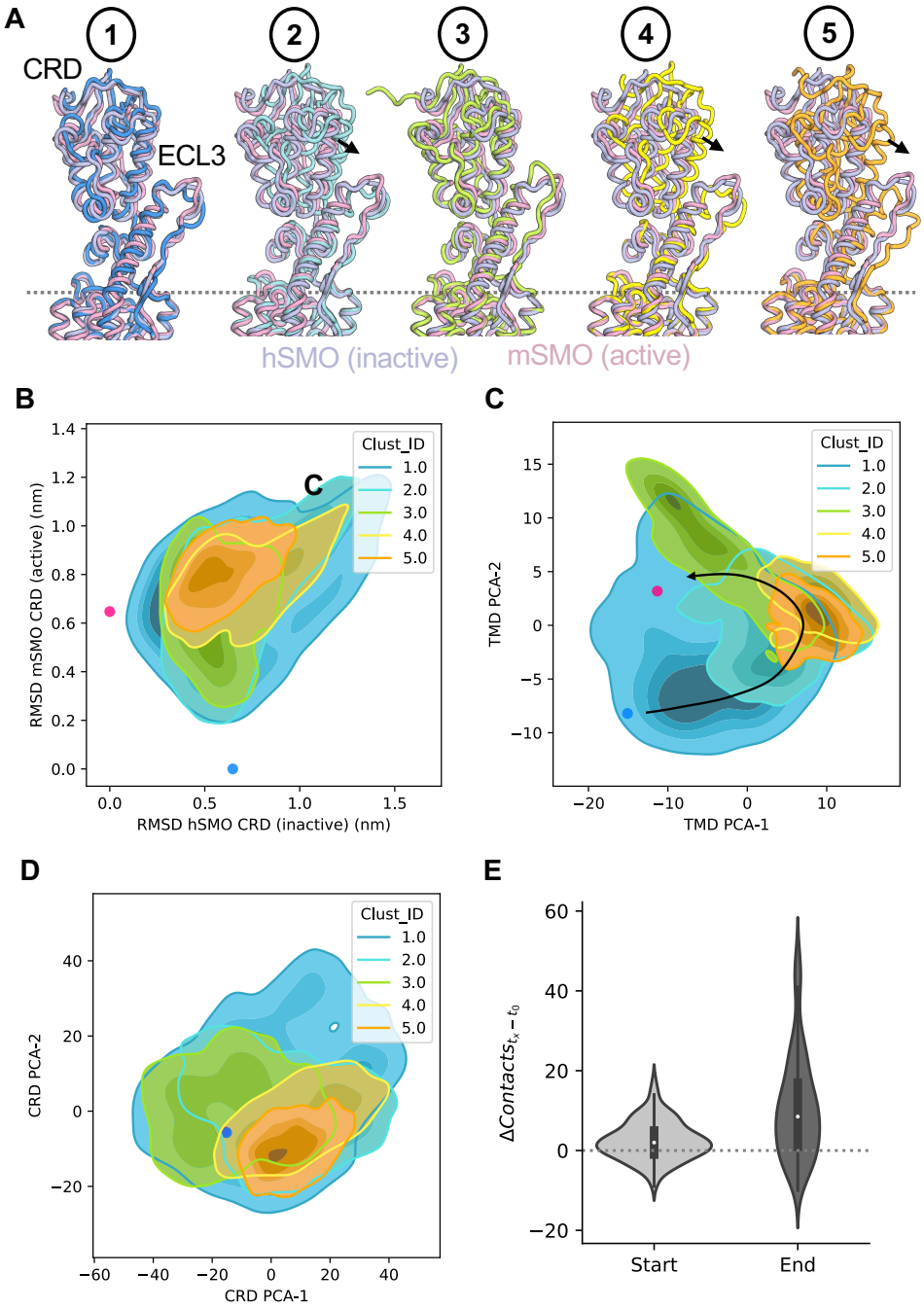


Figure 4.10: Highly dynamic CRD behaviour between clusters.

A) Overlay of the central conformation from clusters 1-5 with the hSMO and mSMO structures, indicating the relative position of the CRD. **B-D)** 2D population distributions of all conformations belonging to clusters 1-5 across targeted MD simulations. **B)** RMSD of the CRD Cα atoms (P58-Q192) to the hSMO and mSMO structures after aligning on TM1-4 Cα atoms (E226-D382). **C-D)** 2D principal component analysis (PCA) of the **C)** TMD and **D)** CRD calculated by projecting the targeted MD trajectory on eigenvectors 1 and 2. Pink and blue spheres represent values in the mSMO and hSMO structures respectively. **E)** Change in the number of contacts between the CRD and ECL3 over the first (Start) or final (End) 10 frame of targeted MD trajectories (t_x) compared to the number of contacts in the hSMO crystal structure (t_0).

4.4 Conclusion

4.4.1 PTCH1 inhibits SMO by regulating cholesterol binding to the CRD

Our combined biochemical, structural and simulation studies lend support to the growing consensus that GPCRs function not as simple switches but via modulation of a complex conformational landscape in response to the cellular environment[81]. Our data suggest HH induced SMO signalling is primarily mediated by cholesterol binding to the CRD site. In contrast the TMD-cholesterol site functions allosterically to tune CRD induced signalling, and is responsible for basal SMO signalling when the CRD site is unoccupied[252]. Thus, unusually for GPCRs, SMO engages the same ligand for orthosteric and allosteric regulation of a gradational signalling response. Not considered here, is additional regulation resulting from e.g. cholesterol binding to sites on the TMD periphery[99,101], as occurs for the Class A serotonin receptor 5-HT_{1A}[204].

These observations raise two questions: how does cholesterol access the TMD/CRD sites and how is SMO activation prevented against the high background levels of membrane cholesterol?

Biochemical studies suggest PTCH1 regulates the abundance of accessible membrane cholesterol by promoting repartitioning to a sphingomyelin sequestered pool or by extrusion from the membrane[21,54]. In addition, cilia localised total cholesterol levels are further reduced in comparison to the plasma membrane by PTCH1 and protein barriers at the cilia base[59]. Thus, the levels of PTCH1 activity (altered by under/over expression or SHH inhibition) may shift the abundance of accessible cholesterol to bind the CRD/TMD (Fig. 4.11A). Furthermore, the relative inaccessibility of the TMD/CRD sites may be a deliberate method of restricting SMO signalling. Cholesterol could bind the TMD via movement through the TM helices, as occasionally observed in GPCR simulations[261], or via a circular SMO TM1-5 precursor/shuttle[262]. How cholesterol binds to the CRD remains an open question however both binding events are likely to have a higher energetic barrier than classical modes of GPCR ligand engagement. The orthosteric TMD ligand pocket in other GPCRs forms a

relatively accessible funnel for ligand binding. In contrast the lower TMD site of SMO is highly buried and has a reduced pocket volume, which may increase the energetic barrier for cholesterol binding. Similarly, for cholesterol to bind to the CRD, it must be liberated from the membrane environment, usually associated with a sizable energetic penalty[263,264]. SMO signalling consequently transpires from the paradoxical entanglement of cholesterol accessibility and inaccessibility.

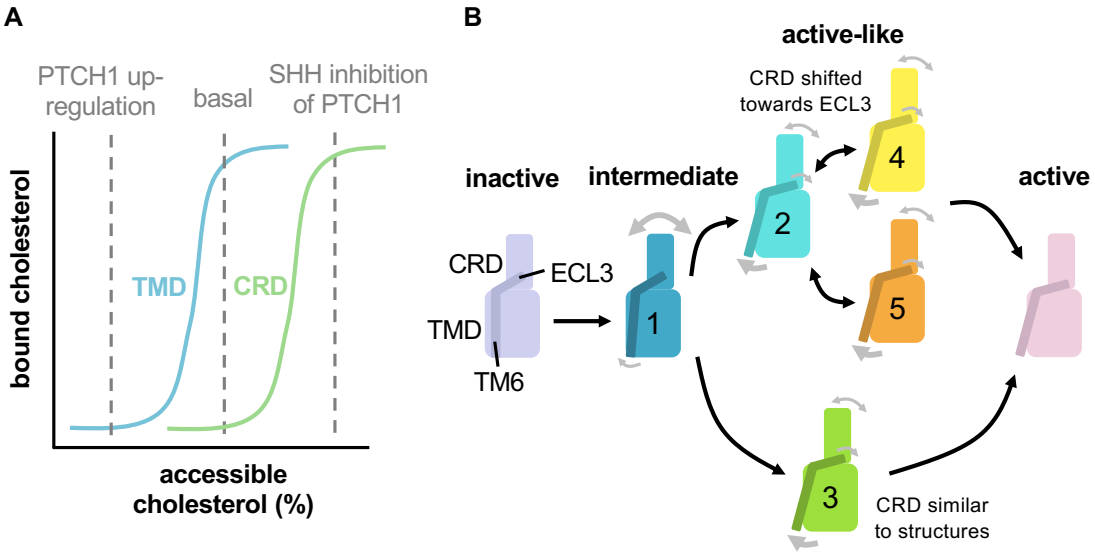


Figure 4.11: SMO activation: cholesterol binding and conformational transitions.

A) Schematic saturation curves for cholesterol binding to the TMD (blue) and CRD (green) sites on SMO *vs* accessible cholesterol. **B)** Schematic pathway for conformational transitions during SMO activation. SMO proceeds through an intermediate (cluster-1) with a dynamic CRD towards two active-like pools with reduced CRD dynamism. Clusters-2, -4 and -5 have a CRD shifted towards ECL3 whereas in cluster-3 the CRD is similar to observed in structures. Activation is accompanied by extracellular outward tilt of ECL3 and intracellular displacement of TM5/6.

4.4.2 SMO dynamics upon activation

We used the change in TM5/6 position between inactive and active SMO conformations to assess whether alternative activation motifs emerged. Equilibrium simulations suggested outward tilt of ECL3 may separate SMO states and outward

movement occurred in targeted MD simulations, strengthening this hypothesis. The number of CRD-ECL3 contacts also increased (Fig. 4.10E). ECL3 emerges as a pathway for information flow in allosteric calculations, which is enhanced when CRD-cholesterol is bound (Fig. 4.4). Simulation and mutational studies also support a role for ECL3 in bridging CRD initiated signalling and TMD activation[93,96].

Clustering of CRD conformations from targeted MD simulations suggested SMO proceeds through a highly dynamic CRD intermediate before occupying one of two active-like conformations whereby the CRD was i) similar to the structurally observed conformation or ii) shifted towards ECL3 (Fig. 4.11B). We note the SMO CRD has only been observed in structures which seem to occupy a subset of similar meta-stable CRD states[2,46]. In cryo-EM structures the CRD is often not observed, supporting CRD dynamics observed in simulations[48,93] (Fig. 4.3). Thus, the true CRD conformational landscape is likely much greater than observed in structures and may be necessary for extraction of accessible cholesterol from the bilayer. It is difficult to rationalise why the CRD-cholesterol site is half open to the solvent in comparison to the enclosed PTCH1 sterol binding domain if this interface does not play an addition role in e.g. interacting with the membrane to liberate cholesterol from the bilayer or a similar protein-handover interface. Our simulations were not in support of an alternative 'tunnel' model between the TMD and CRD sites[250] (Fig. 4.5). An interesting question remains how the CRD dynamics are altered between deglycosylated *vs* GlcNAc (as in our simulations) *vs* 'fully' glycosylated states, particularly considering one of the glycan sites resides on ECL3 and the traditional use of glycosylation inhibitors like kifunensine in structural approaches. Further work is needed to clarify how cholesterol binding alters the SMO activation landscape and to contextualise the conformational predications presented here within the signalling equilibria.

5. HHAT

"I was also missing the sense of discovery from my earlier days on the road; the feeling of striking out into darkness; the nebulous and fearful saviour of direction without destination."
Emily Chappell

5. HHAT 116

5.1 Introduction 117

5.2 Methods 121
5.2.1 Coarse-grained simulations of HHAT 121
5.2.2 Atomistic simulations of HHAT 122
5.2.3 Steered MD simulations of HHAT 124
5.2.4 Simulations of other MBOATs 124

5.3 Results and Discussion 125
5.3.1 The structure of Hedgehog acyl-transferase promotes membrane deformation 125
5.3.2 An extended water network forms throughout HHAT 129
5.3.3 HHAT gating by W335 132
5.3.4 Reaction centre dynamics 134
5.3.5 Assessing lipid-like densities around HHAT 135
5.3.6 A luminal exit gate for SHH 139
5.3.7 HHAT and other MBOATs 142

5.4 Conclusion 144

This chapter contains data also published in:

Coupland, C.E., Andrei, S.A., **Ansell, T.B**., Carrique, L., Kumar, P., Sefer, L., Schwab, R.A., Byrne, E.F.X., Pardon, E., Steyaert, J., Magee, A. I., Lanyon-Hogg, T., Sansom, M.S.P., Tate, E.W., Siebold, C. (2021). Structure, mechanism, and inhibition of Hedgehog acyltransferase. *Mol. Cell*., 81, 24, 5025-5038

Ansell, T.B., Song, W., Coupland, C.E., Carrique, L., Corey, R.A., Duncan, A.L., Cassidy, C.K., Geurtz, M.M.G., Rassmussen, T., Ward, A.B., Siebold, C., Stansfeld, P.J., Sansom, M.S.P. (2022). LipIDens: Simulation Assisted Interpretation of Lipid Densities in Cryo-EM Structures of Membrane Proteins, *Nat. Comms, under review*

Ansell, T.B., Healey, M., Coupland, C., Sansom, M.S.P., Siebold, C. (2022). Investigation of Structural and Dynamic Divergence Across the MBOAT family. *Manuscript in preparation*

5.1 Introduction

Hedgehog acyl-transferase (HHAT) is an ER resident membrane protein involved in processing of the mature hedgehog morphogen[8–10]. The immature hedgehog ligand is composed of N- and C-terminal domains, the latter of which is cleaved via an autocatalytic mechanism involving formation of an internal thioester intermediate at a highly conserved cysteine residue on the C-lobe. The C-terminus of the N-domain of SHH is specifically cholesterylated, releasing the C-lobe for degradation[5–7] (Fig. 5.1). HHAT catalyses the N-terminal palmitoylation of the SHH N-lobe to produce the mature bi-lipidated SHH product[9]. The order of SHH palmitoylation and cholesterylation steps is not known. HHAT functions upstream of PTCH1 and SMO, which reside within the membrane of hedgehog receiving cells, and upstream of morphogen release into the surrounding cellular environment[10]. HHAT is also specific for palmitoylation of the SHH morphogen[9]. For these reasons HHAT, is an attractive target for the development of alternative HH signalling pathway drugs.

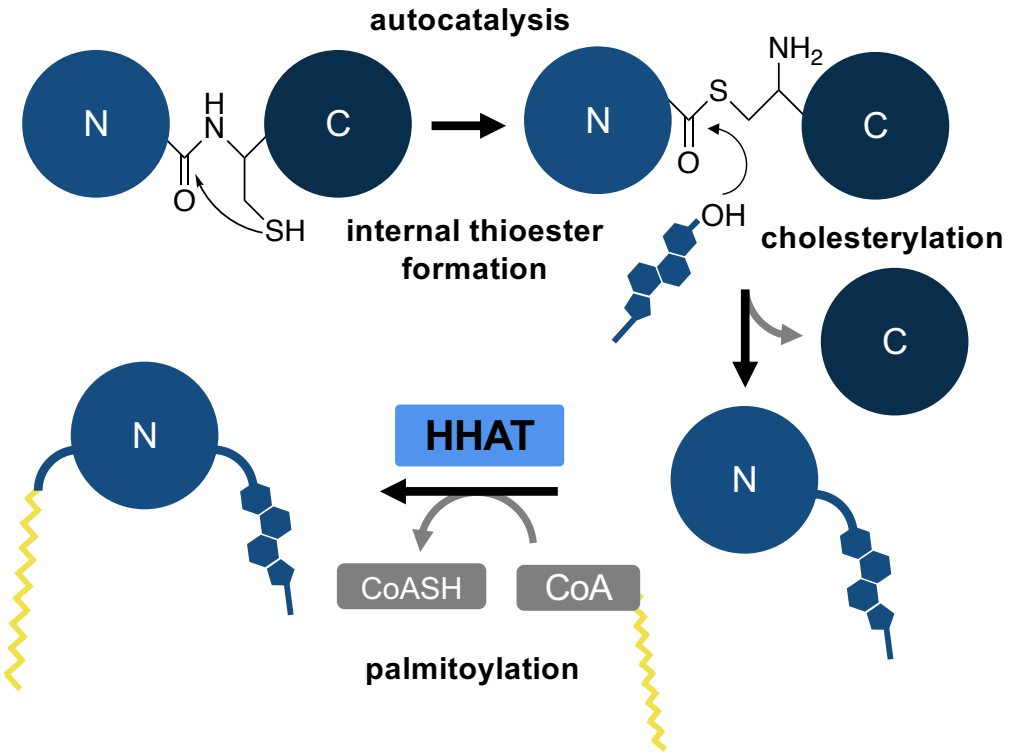


Figure 5.1: Sonic Hedgehog (SHH) morphogen processing.

The premature SHH protein is comprised of an N- and C-terminal lobe. A conserved cysteine residue is implicated in formation of an internal thioester bond via N,S-acyl shift. The β_3-hydroxyl group of cholesterol engages in an autocatalytic nucleophilic exchange reaction, releasing the SHH C-lobe for proteolytic degradation. Palmitoylation of the N-terminus of the SHH N-lobe is catalysed by Hedgehog acyl-transferase (HHAT), whereby the palmitoyl group handover is energised by coenzyme A (reaction mechanism uncertain) to product the mature, bi-lipidated SHH product.

HHAT is a member of the membrane bound O-acyl-transferase (MBOAT) protein family, despite functioning as an N-acylase to palmitoylate SHH-C24[265]. MBOAT members can be categorised into two subfamilies; those involved in modification of small molecules such as lipids and those which catalyse protein acylation. Small-molecule MBOATs include the prokaryotic teichoic acid D-alanylase, DltB[266], as well as eukaryotic acylases such as acyl-coenzyme A cholesterol acyl-transferase[267–269] (ACAT1/SOAT1) and acyl-coenzyme A diacylglycerol acyl-transferase[270,271] (DGAT1). In addition to HHAT, other protein MBOATs include the protein-serine O-acyl-transferase Porcupine[272] (PORCN), involved in Wnt morphogen palmitoleolyation, and ghrelin O-acyl-transferase[273] (GOAT). In all eukaryotic MBOATs handover of the acyl

moiety to the small-molecule/protein substrate is mediated by coenzyme A (hereafter CoA)[274]. For DltB, the protein DltC performs an equivalent carrier function[266].

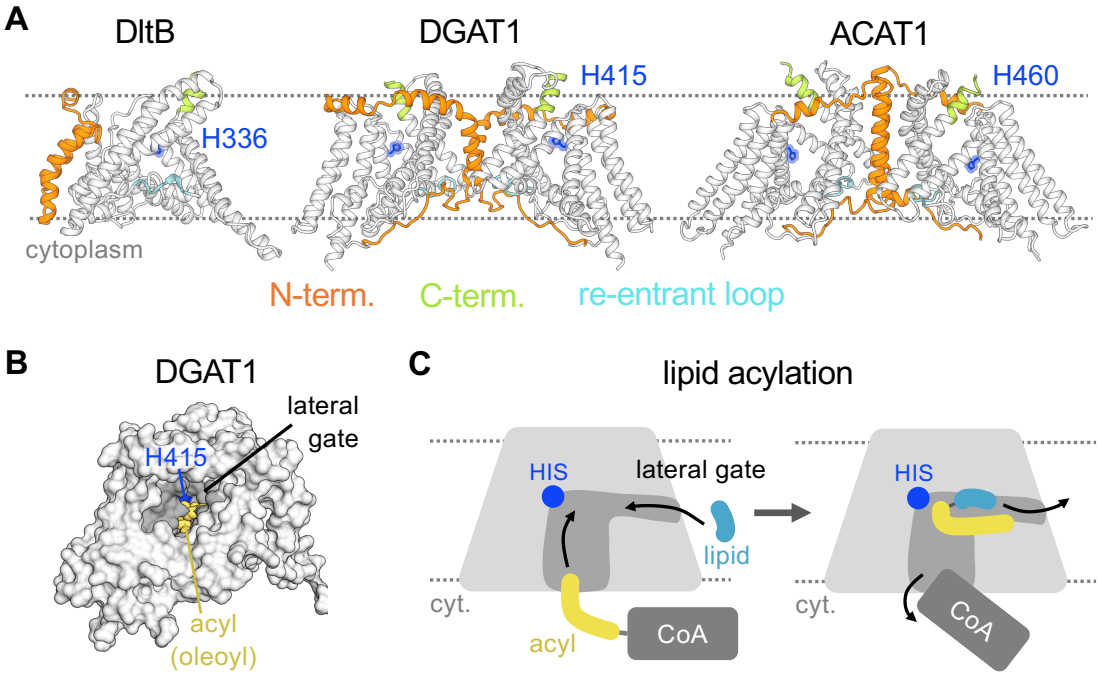


Figure 5.2: Structures of MBOAT family members.

A) Structures of MBOAT family members: the prokaryotic monomer DltB (PDB: 6BUG) and eukaryotic dimers DGAT1 (PDB: 6VP0) and ACAT1 (PDB: 6P2P). The N- and C-terminal regions are coloured orange and lime respectively, the re-entrant loop is coloured cyan and conserved catalytic histidine residues blue. **B)** Proposed tunnel ('lateral gate') for exchange of reactants between the enzymatic reaction centre of DGAT1 and the membrane. The position of H415 (blue) and oleoyl-CoA (yellow) with respect to the lateral gate (grey) are indicated when viewed from the membrane. **C)** Model for the acylation of lipid substrates by small molecule MBOATs. Acyl-CoA binds within a cytoplasmic cavity of the MBOAT whereby the thioester group positions within the reaction centre. Lipid substrates enter through the lateral gate, acylation occurs and the acylated-lipid is released into the membrane.

Prior to the onset of this thesis there were no available structures of MBOAT members, impeding an understanding of MBOAT family mechanisms. Structural insights into the MBOAT fold were provided by the determination of DltB structures[266] by X-ray crystallography and, later, cryo-EM structures of ACAT1 and DGAT1[267–271] (Fig. 5.2A). These structures revealed the presence of a conserved MBOAT fold, despite very low sequence identity (~12%)[126]. DltB is monomeric and comprised of 11 tilted TM helices and two intracellular re-entrant loops, forming an unusual trapezoid shape. On the

extracellular surface (the shortest trapezoid vertex) a funnel is formed which extends towards a conserved histidine residue, proposed to form the MBOAT catalytic reaction centre[266]. Surprisingly both ACAT1 and DGAT1 function as dimers with a protein-protein interface formed by their N-terminal helices[267–271]. In contrast to DltB, ACAT1 and DGAT1 structures were solved with oleoyl-CoA bound within the MBOAT core[267,270,271]. This positions the thioester oleoyl-CoA group within proximity of the catalytic histidines. The ACAT1 and DGAT1 structures also revealed the presence of a tunnel extending between the enzymatic reaction centre and the membrane accessible protein surface, termed the 'lateral gate' (Fig. 5.2B-C). The lateral gate is proposed to represent the portal by which lipid substrates and acylated products diffuse into and out of the membrane[267–271]. This is supported by an observed but unidentifiable density within the lateral gate of DGAT1[271] and by docking of cholesterol within ACAT1[269,270].

Despite these advances within the small-molecule MBOAT subfamily, prediction of protein MBOAT structures remains challenging, attributed to low sequence conservation across the family (<15%)[126]. Homology models of HHAT failed to produce a consensus topological arrangement, with predictions ranging between 10 and 13 transmembrane helices[275,276]. Multiple sequence alignments and mutational assays predict two residues of HHAT to be important for catalysis; H379 and D339. H379 is invariant across HHAT homologues and mutation to alanine reduces HHAT catalysis to ~50% of WT activity[9]. D339A mutations reduced HHAT activity to ~10% of WT activity[277]. In other MBOATs the equivalent residue to D339 is more frequently an asparagine residue[265], challenging the proposed role D339 may play in HHAT catalysis. One caveat with these *in vitro* mutational assays pertains to the use of a truncated SHH peptide rather than the mature SHH morphogen. An additional complexity is presented by some HHAT homology models which predict D339 and H379 to residue on opposite sides of the membrane[275,276]. Thus, a structural understanding of HHAT is necessary for untangling this perplexing array of biochemical and bioinformatic analyses and formulating mechanistic differences between small-molecule and protein MBOATs.

Structural studies of HHAT are needed to help elucidate the following key functional questions:

1. What is the topological arrangement of protein MBOATs?
2. Where does the protein substrate (SHH) bind on HHAT?
3. How do the reactant entry and exit paths differ between eukaryotic small-molecule MBOATs (where the acyl chain is transferred from the cytosolic face into the membrane) and protein MBOATs (which must transfer the acyl moity across both membrane leaflets)?
4. How is palmitoylated-SHH released from HHAT?

This study details a collaborative approach to understanding HHAT mechanisms using structures from the Siebold lab, biochemical assays from the Tate lab at Imperial and simulations.

5.2 Methods

5.2.1 Coarse-grained simulations of HHAT

Two sets of CG simulations were performed using HHAT coordinates from the HHAT-MB177-nhPalm-CoA-Heme structure, initially at 5 Å resolution and later refined to 2.7 Å[179]. In both cases ligands and the bound megabody were removed from the HHAT-MB177-nhPalm-CoA-Heme structure and loops between D45-L51, W189-T197 and P409-Q422 modelled using Modeller9.20[221]. The model with the lowest DOPE score was selected and converted to CG resolution using *matinize.py*[146]. The MARTINI 2.2 forcefield[149] was used to describe all components and the ElNeDyn[222] elastic network applied to HHAT (force constant: 1000 kJ mol^{-1} nm^{-2}, cut-off: 0.9 nm). HHAT was positioned in a bilayer composed of POPC (35%), DOPC (35%), POPE (8%), DOPE (7%), cholesterol (10%) and palmitate (5%) in the luminal leaflet and POPC (15%), DOPC (15%), POPE (19%), DOPE (18%), POPS (8%), PIP_2 (10%), cholesterol (10%) and palmitate (5%) in the cytoplasmic leaflet using *insane.*py[223]. This bilayer composition was selected to recapitulate the main features of ER membranes e.g., low sphingolipid content and a cholesterol:phospholipid ratio of approximately 0.15[278]. Palmitate was modelled with a charged headgroup (named PCN in the MARTINI topology) and cholesterol was modelled with inclusion of virtual sites[279]. The membrane position around DltB (PDB: 6BUG) on the MemProtMD[203] server

(memprotmd.bioch.ox.ac.uk) was used to guide HHAT insertion into the bilayer. The system was solvated using *insane.py*[223] and approximately 0.15 M NaCl added to yield a total box size of 14 x 14 x 9 nm^3. Prior to the production run, each replicate was independently energy minimised using a steepest decent method and equilibrated in two steps (step 1: 25 ns with all protein beads restrained, step 2: 100 ns with backbone beads restrained).

CG MD simulations (10x 15 μs at each HHAT resolution) were performed using GROMACS 2019.4[155]. The timestep was 20 fs and periodic boundary conditions were applied. Coulombic interactions were described using the reaction-field method and a 1.1 nm cut-off. Van der Waals interactions were described with the potential-shift Verlet method and a 1.1 nm cut-off. The V-rescale thermostat[225] was used to maintain temperature as 310 K with a 1.0 ps coupling time constant. Pressure was maintained at 1 bar using the Parrinello-Rahman barostat[226], a 12.0 ps coupling time and a compressibility of 3x 10^{-4} bar^{-1}.

5.2.2 Atomistic simulations of HHAT

A snapshot from the CG simulations was selected for backmapping to atomistic resolution using a revised CG2AT protocol[280] (github.com/owenvickery/cg2at). Prior to backmapping, palmitate was removed from the bilayer and a short equilibration performed to close gaps in the bilayer since topology files for backmapping palmitate were not available. The protein conformation was backmapped to the structural coordinates such that only the lipid arrangement reflected those of the CG snapshot. TIP3P water[230] was readded to the system along with approximately 0.15 M NaCl.

The protein coordinates from the HHAT-MB177-nhPalm-CoA-Heme structure[179] were used to build the following systems: HHAT *apo*, HHAT with Palm-CoA bound, HHAT with Palm-CoA bound and Heme covalently bound to C324 and HHAT with the proposed cholesterol moiety of SHH bound or with the same pocket occupied by CHS or palmitate. The protein coordinates from the HHAT-IMP-1575 structure[179] were used to build the following system: HHAT in an inhibited conformation (but without inhibitor bound) with Heme covalently bound to C324. The protonation state of residue

sidechains in the input structures were assessed using propKa[228] (github.com/jensengroup/propka-3.0) which did not identify any unusual pKas, therefore D339 was modelled as deprotonated and H379 was neutral. The CHARMM-36 forcefield[131] was used to describe all components. Palm-CoA parameters were derived from Stix *et al*[281]. Standard CHARMM-36[131] parameters were used for the Mg ion bound to Palm-CoA (which remained bound without the need for additional restraints) and for the Heme/Fe complex. To describe Heme-C324 bonding C324 was modelled as a thiolate (named CYM in the CHARMM-36 forcefield) and S-Fe bond parameters described according to existing parameters for Heme coordinated by methionine. Control simulations of Heme without bonding to C324 were also performed, yielding similar results to the bonded system (data not shown). Each replicate was energy minimised using a steepest decent method then subject to 2x 5 ns NVT and NPT equilibration steps with position restraint applied to the protein backbone and the heavy atoms of bound ligands.

Each system was simulated for 5x 200 ns with a 2 fs timestep using GROMACS 2019.4[155].The Particle-Mesh Ewald (PME) method[233] was used for long-range electrostatic interactions with a 1.2 nm cut-off. Van der Waals interactions were smoothly switched between 1.0 and 1.2 nm using the force-shift modifier. Temperature was maintained at 310 K using the Nosé-Hoover thermostat[231,232] with a 0.5 ps coupling time. Pressure was maintained at 1 bar using the Parrinello-Rahman barostat[226], a 2.0 ps coupling time and a compressibility of 4.5×10^{-5} bar^{-1}. A dispersion correction was not applied. The LINCS algorithm was used to constrain bonds to their equilibrium values[227].

In simulations initiated from the HHAT *apo* conformation a DOPC lipid bound within the luminal gate was removed after 200 ns in each of the 5 replicates. After lipid removal, each replicate was simulated for an additional 200 ns.

In addition, a further 8x 200 ns simulations of the HHAT *apo* conformation were performed, backmapped from different CG frames. The CG frames selected encapsulated the top ranked lipid binding poses for a couple of sites surrounding HHAT, as identified using PyLipID[157]. These simulations were used to refine the pose

of site lipids and assess how well the lipid binding poses matched lipid-like densities present within the cryo-EM map.

During the study an additional structure of HHAT was published (PDB: 7MHY) in complex with two Fabs, Heme and two Palm-CoA molecules[282]. 5x 100 ns atomistic simulations of this HHAT conformation were performed in complex with the Heme and Palm-CoA molecules, in accordance with those described above.

5.2.3 Steered MD simulations of HHAT

In addition to removal of the DOPC lipid (at t = 200 ns) from the luminal gate (described above) atomistic simulations were also performed where the lipid was pulled laterally into the membrane over a period of 10 ns before 40 ns of additional unbiased simulation time for each of the five replicates. Steered MD simulations were assisted by use of PLUMED[257,258]. The collective variable used during steered MD was defined as the distance between the centre of mass of the DOPC lipid and the Cα atom of Arg176. A moving restraint was applied over 10 ns to gradually increase the force constant to k=1000 kJ mol^{-1} nm^{-2}. This restraint was switched back to k=0 kJ mol^{-1} nm^{-2} over the following 0.5 ns before a further 39.5 ns of unbiased simulation. All other simulations conditions are identical to those described above.

5.2.4 Simulations of other MBOATs

These simulations were performed during (or extended from) supervision of Megan Healey (Part II student 2020-2021). CG and atomistic simulations of other MBOAT proteins were run identically to described above for HHAT in *apo* conformations. Simulations were initiated from structures of DltB (PDB: 6BUG[266]), ACAT1 (PDB: 6P2P[267]) and DGAT1 (PDB: 6VP0[270]) or from AlphaFold2 models of GOAT and PORCN[127]. CG simulations were initiated using ER mimetic bilayers (see above for HHAT) except for DltB which was embedded in a bilayer representative of the Gram-positive membrane (POPG: 69% POPE: 23%, CDL: 8%). In total 10 x 15 μs CG simulations and 5 x 200 ns atomistic simulations were run for each protein.

Bioinformatics analysis was performed by constructing multisequence alignments (MSAs) of each MBOAT using PSI-BLAST[283] with an E-value of 1 x 10^{-3}. PSI-BLAST results were passed to TCoffee[284] to construct the MSAs through the MPI Bioinformatics toolkit[285]. ConSurf[286] was used to map sequence conservation onto the structures.

5.3 Results and Discussion

5.3.1 The structure of Hedgehog acyl-transferase promotes membrane deformation

The structure of Hedgehog acyl-transferase (HHAT) was solved by members of the Siebold group to 2.7 Å resolution in complex with a megabody (MB177) bound to the ER luminal face[179]. HHAT has 12 TM helices and two cytosolic re-entrant loops arranged around a conserved catalytic reaction centre (Fig. 5.3A-B). The MBOAT structural core (comprised of 8 TM helices) is maintained between HHAT and small-molecule MBOATs, however the N- and C-termini form a unique conformation and wrap over the luminal face of HHAT (Fig. 5.4). This re-arrangement of the termini occludes the lateral gate of small-molecule MBOATs proposed to function as a conduit for reactants into the surrounding membrane[267–271]. Additionally, the distinct arrangement of the HHAT termini forms a trigonal pocket, encapsulating a bound Heme molecule. The Heme Fe ion is axially coordinated by a conserved C324 residue while the two carboxylate groups of the Heme porphyrin ring position towards the cytoplasmic leaflet and are each coordinated by an arginine and tryptophan residue (R250/W254 and R323/W486) (Fig. 5.3C). C324A mutation abolishes the typical UV/VIS spectrum observed for Heme binding proteins[287], confirming that HHAT is a Heme binding protein (Fig. 5.3D).

Density corresponding to a non-hydrolysable Palmitoyl-CoA (nhPalm-CoA, Appendix Fig. S5.1) molecule (added during cryo-EM sample preparation) could be observed within HHAT (Fig. 5.3E). The cytosolic location of the CoA adenine ring was similar to that of small-molecule MBOATs[267,270,271]. The nhPalm-CoA molecule is shifted along the membrane normal compared to its simulated position within a membrane[281] to

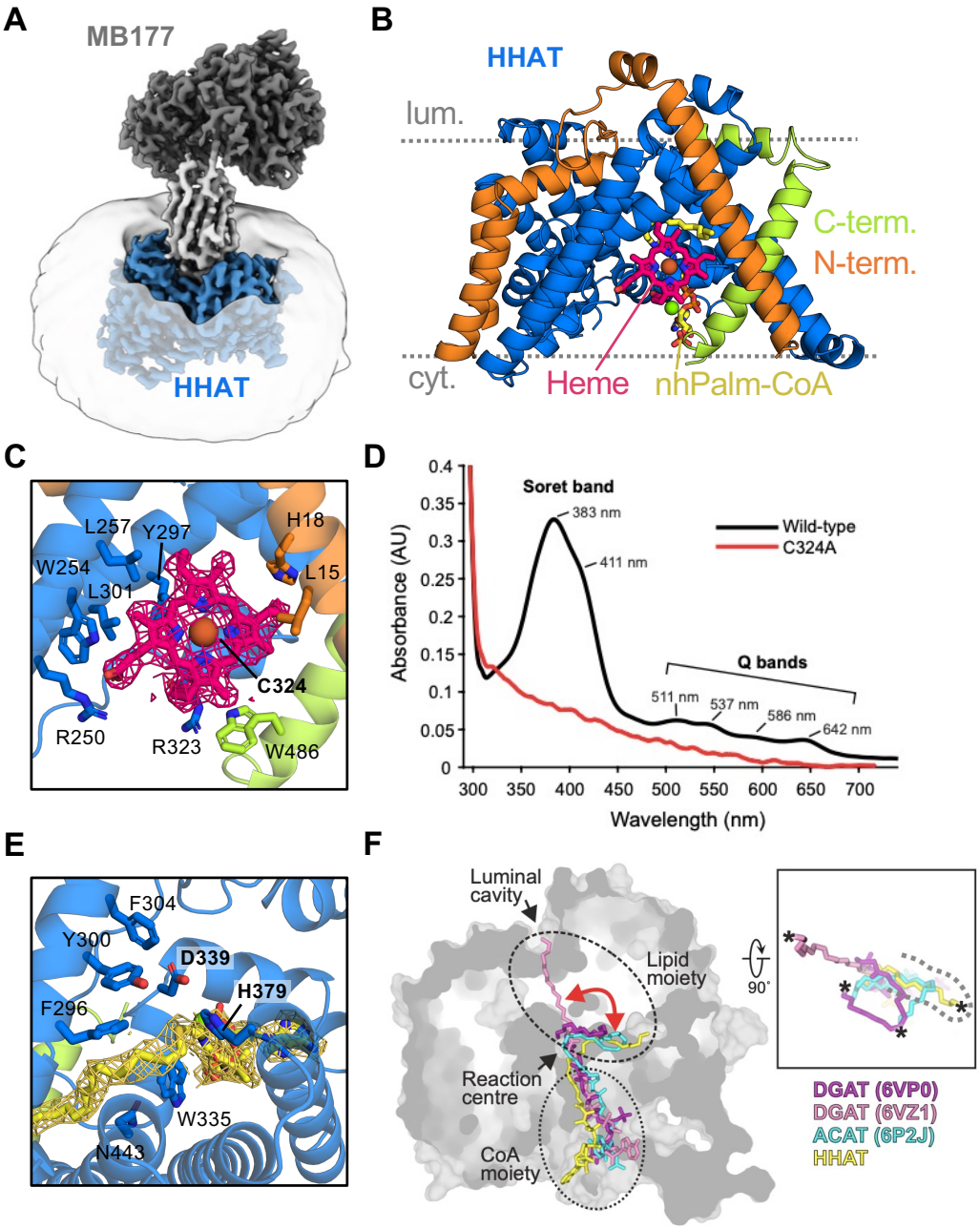


Figure 5.3: Structure of Hedgehog acyl-transferase (HHAT).

A) The cryo-EM structure of HHAT (blue) was solved by Claire Coupland and Dr Loic Carrique to 2.7 Å bound to the megabody (MB177) scaffold (grey). **B)** Cartoon representation of HHAT in complex with a B-Heme (magenta) and non-hydrolysable palmitoyl coenzyme A (nhPalm-CoA, yellow). N- and C- terminal helices are coloured orange and lime respectively. **C)** Close up of Heme coordinating residues within the pocket formed by rearrangement of HHAT termini. **D)** UV-vis spectra of WT HHAT (black) and the Heme coordinating C324A mutant (red), generated by Claire Coupland. **E)** Close up view of nhPalm-CoA coordination within the HHAT reaction centre. Proposed catalytic residues (H379 and D339) are marked in bold. **F)** Comparison of the position of nhPalm-CoA within HHAT (yellow/grey) with respect to acyl-CoA molecules resolved in structures of DGAT1/ACAT1 (purple/cyan), adapted from Coupland *et al* [179].

position the thioester bond close to the proposed reaction centre residues, H379 and D339 (Fig. 5.3E). In contrast, the palmitoyl chain is uniquely positioned within HHAT compared to in other MBOATs (Fig. 5.3F). The terminal C16 palmitoyl carbon is located within proximity of the Heme and an unusual, kinked lipid-like density, modelled as palmitate. Thus, restriction of the tail conformation within this pocket provides rational for previous analyses revealing preferential acylation of SHH by shorter acyl chains compared to chains of >16 carbons[8,288].

The structure of HHAT raises a number of questions – most notably how does HHAT position within the bilayer given its unusual helical arrangement and what is the role of the Heme moiety? To probe these further, we performed CG simulations of HHAT embedded within an ER-like membrane environment which lacked sphingolipids and had a lower cholesterol concentration (10%) compared to plasma membrane models[278]. CG simulations were initially performed using a lower resolution structure of HHAT (~5 Å) in which non-protein components (Heme, Palm-CoA, lipid/detergent densities) were not visible. These simulations were subsequently rerun when the map resolution was improved (2.7 Å) however the 5 Å structure proved useful for predicting the location of lipid binding sites which were subsequently revealed when the resolution was improved (see section 2.2.7).

We calculated the time averaged density of lipid phosphate beads surrounding HHAT across the CG simulations. Two pronounced membrane deformations were observed in opposing leaflets formed by L321-T327 (near the Heme binding site) and around TM5-TM9 in the luminal leaflet (Fig. 5.5). The luminal deformation funnels towards the catalytic reaction centre and H379 on TM9. The net result is a reduction in bilayer width (1.7 nm) by up to ~50% compared to at extended distances from the protein (3.8 nm). This may reduce the energetic cost of palmitoyl transfer across the membrane and/or aid entry/exit of bound ligands.

Alternatively, HHAT induced bilayer thinning may reduce the membrane diffusion limit of the protein to increase the probability of substrate encounter, as has been proposed for the Rhomboid protease family[289]. An enhanced diffusion rate in thinner membranes

has also been observed using high-speed atomic force microscopy of aquaporin (Yining Jiang, BPS 2022, unpublished).

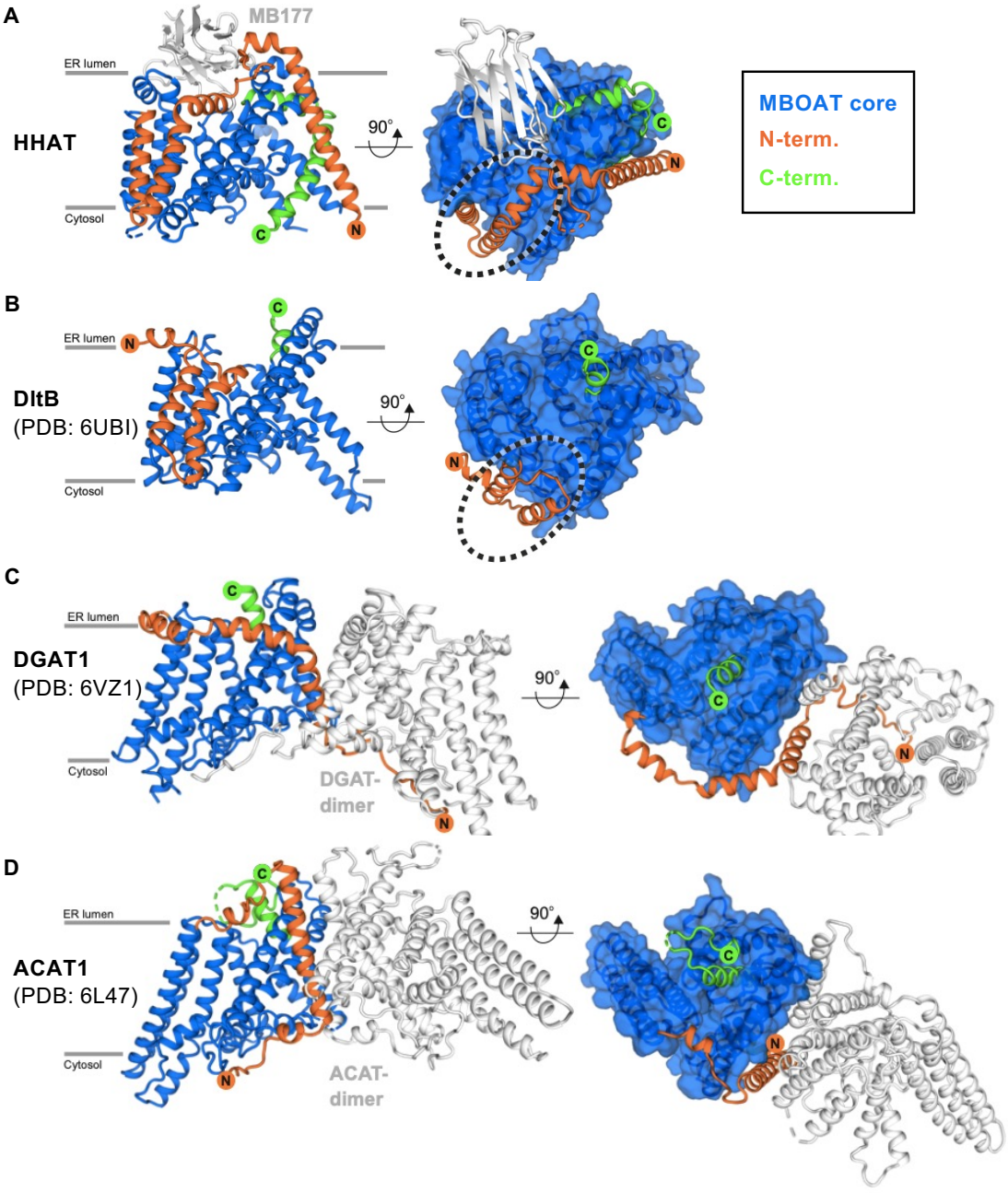


Figure 5.4: Re-arrangement of the termini between MBOAT subtypes.

Figure adapted from Coupland *et al*[179]. The MBOAT core is coloured blue and N-/C-termini are coloured orange and lime respectively. The dashed circle indicates the position of the lateral gate in DGAT1/ACAT1 which is occluded in MBOATs that transfer substrate across both membrane leaflets (DltB/HHAT).

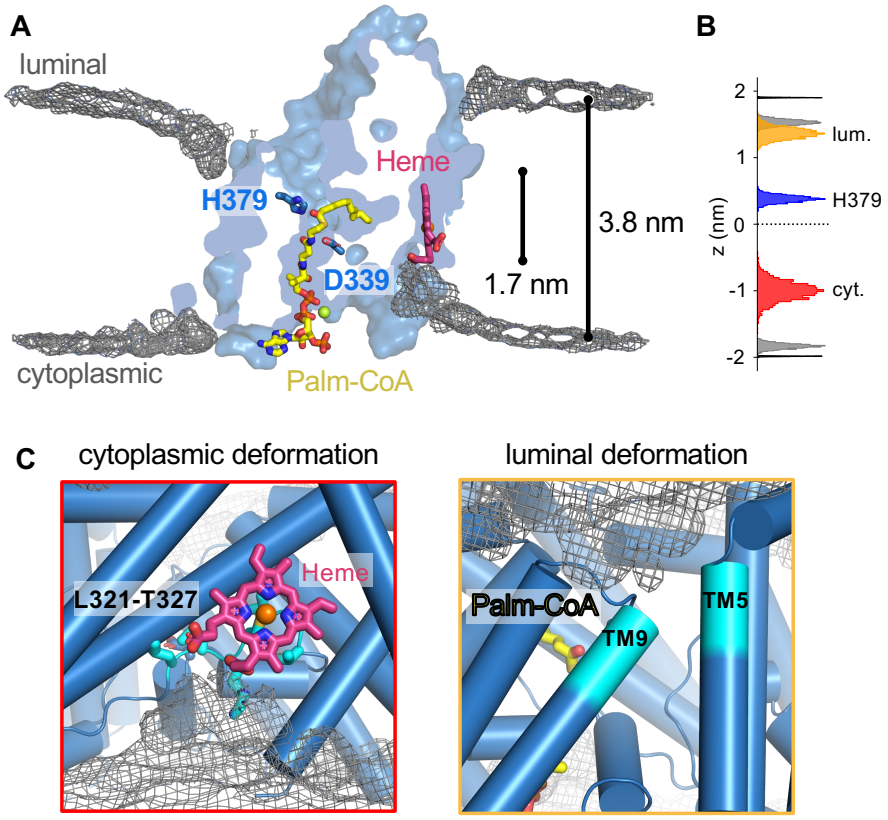


Figure 5.5: HHAT induces membrane deformation.

A) Time averaged density of lipids phosphate beads (isomesh) across 10 x 15 μs CG simulations of HHAT, overlayed with the atomistic structure of HHAT (blue surface). **B)** Histogram of lipid phosphate bead z axial position across CG simulations for all beads (black) or any beads within 0.8 nm of protein backbone (BB) (grey), or within 0.8 nm of residues in proximity to the cytoplasmic (red) or luminal (yellow) deformations (see **C** for residues in cyan). The z coordinates of the H379 BB bead (blue) are shown for reference. **C)** Close up of regions of cytoplasmic (A320-S326) and luminal (T146-G151, W378-G381) membrane deformation (isomesh).

5.3.2 An extended water network forms throughout HHAT

Intrigued by our observations of membrane deformation we sought to assess the solvent environment near HHAT. Atomistic simulations of HHAT in the presence of distinct ligand combinations (*apo*, Palm-CoA, Palm-CoA plus Heme, inhibited (Heme only)) revealed the formation of a solvated water network through HHAT. The solvent accessible region encompasses a luminal funnel (into which CDR3 of the nanobody scaffolded to MB177 binds) the reaction centre and the cytosolic cavity formed by deformation of the membrane (Fig. 5.6). Thus, while visualisation of the HHAT

structure alone might suggest the Heme to be fully embedded within the hydrophobic membrane region, MD simulations indicate that it is situated at the membrane-solvent interface with the Heme-carboxyl groups puncturing the wetted cytoplasmic cavity. Thus, Heme binding could be regulated by either membrane or cytoplasmic components.

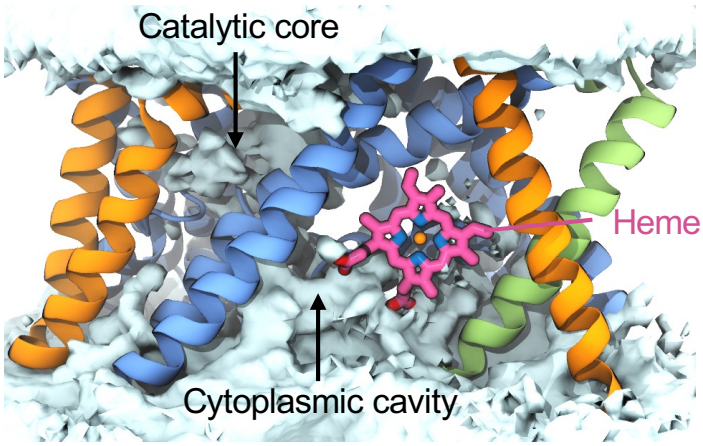


Figure 5.6: Formation of water filled cavity within HHAT.

Time averaged solvent density across 5 x 200 ns atomistic simulations of HHAT. Figure adapted from Coupland *et al*[179].

To investigate this further, we used Caver[290] to calculate tunnels within HHAT and then assessed the number of water within 0.4 nm of the centroid of each tunnel. across the final 50 ns of simulations (Fig. 5.7). There was no difference to the number of water within the reaction centre (blue) or extending towards the Heme (light/dark green) in the presence of different ligands (Fig. 5.7B-F). When Palm-CoA was present water was displaced from the tunnel occupied by the pantothenic moiety (red). This may have implications for future therapeutic targeting of HHAT whereby drugs must be sufficiently polar to enter a water occupied pocket but suitably hydrophobic to displace water from a subsection of the tunnel network. Additionally, this observation forms a conceptual bridge between the well-studied mechanisms of hydrophobic gating and lipid binding.

Curiously, despite the differences in HHAT water networks induced by Palm-CoA binding we observed no change to the net flux of water through HHAT between *apo* or liganded states (Fig. 5.8). In all cases flux of water through HHAT was minimal, prompting further investigation into HHAT sidechain dynamics.

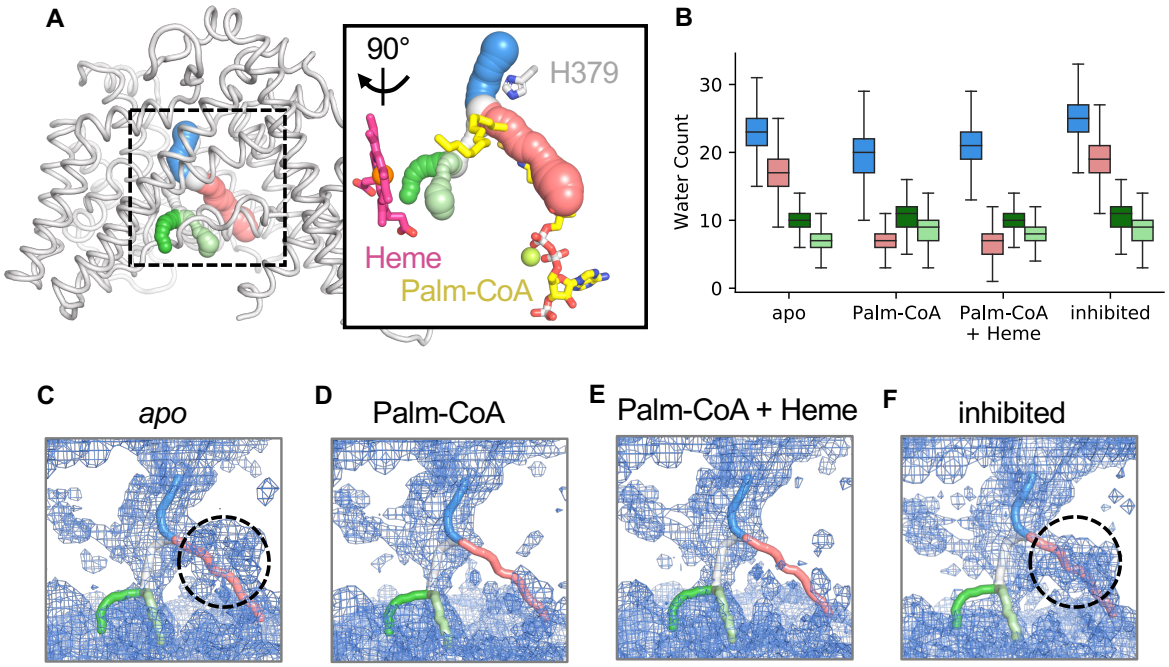


Figure 5.7: Water within HHAT.

A) Tunnels through the HHAT core as calculated using Caver and displayed as spheres. The inset indicates the position of tunnels with respect to Heme (pink), Palm-CoA (yellow) and His379 (grey). **B)** Quantification of the number of water within 4 Å of the centroid coordinates through the tunnels indicated in **A** across the final 50 ns of atomistic MD simulations of HHAT in the *apo*, Palm-CoA, Palm-CoA plus Heme and inhibited states. Boxplots are coloured according to the tunnel colours in **A**. **(C-F)** Time averaged solvent density within the HHAT core across atomistic simulations of HHAT in the *apo* (**C**), Palm-CoA bound (**D**), Palm-CoA plus Heme (**E**) and inhibited (**F**) conformations. Trajectories were combined and fitted before the density calculations which were performed using MDAnalysis. The centroid through the tunnels are shown as sticks and the circles indicates solvent within the Palm-CoA binding pocket. Figure adapted from Coupland *et al*[179].

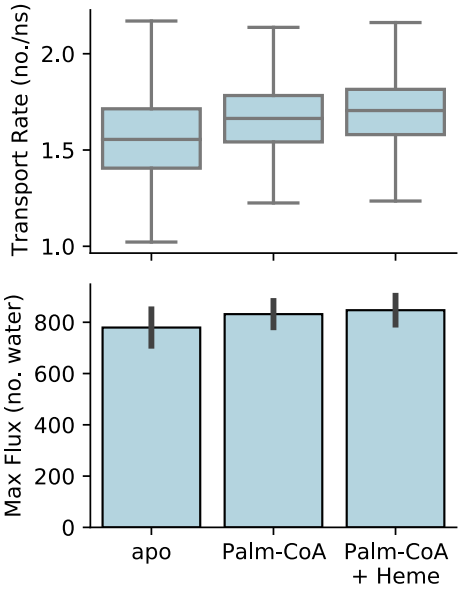


Figure 5.8: Water flux through HHAT.

Transport rate and maximum flux of water through HHAT across atomistic simulations. Flux is defined as the difference in movement of a water from the lumen to the cytosol compared to the reverse direction.

5.3.3 HHAT gating by W335

A second structure of HHAT was solved by members of the Siebold group in complex with the inhibitor IMP-1575 to 3.7 Å. IMP-1575 is a potent HHAT inhibitor which occludes the Palm-CoA binding site within the reaction centre of HHAT[291]. IMP-1575 binding leads to re-orientation of several sidechains including D339, N443 and W335 (Fig. 5.9). The IMP-1575 molecule forms a hydrogen bond with H379, hence IMP-1575 inhibits HHAT by altering the intermolecular landscape of both catalytic residues (D339/H379) and directly blocking Palm-CoA binding.

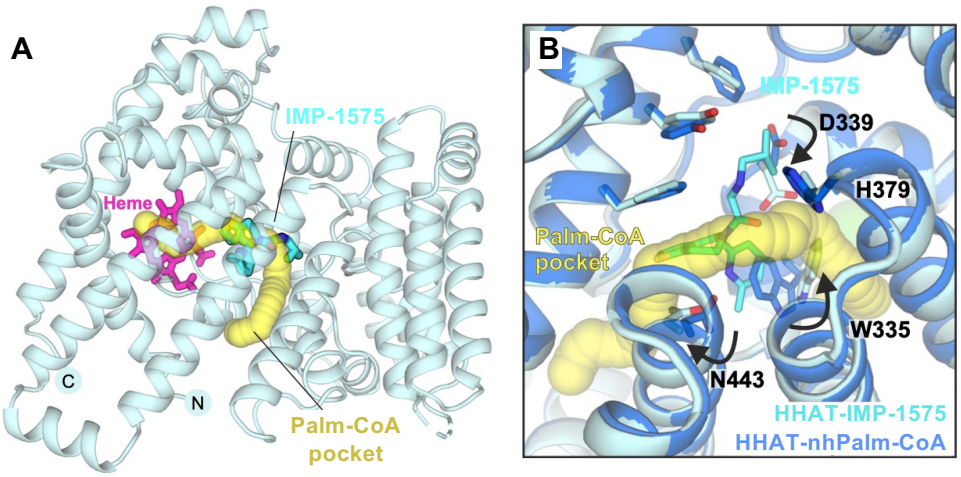


Figure 5.9: Structure of HHAT bound to inhibitor IMP-1575.

A) Cryo-EM structure of HHAT bound to the potent inhibitor IMP-1575 and Heme, solved by Claire Coupland and Loic Carrique to 3.6 Å (adapted from [179]). **B)** Overlay of the HHAT bound nhPalm-CoA (blue, Fig. 5.3) and HHAT bound IMP-1575 (cyan) reaction centres as viewed from the luminal face. IMP-1575 is shown as cyan sticks overlayed with the tunnel occupied by nhPalm-CoA (yellow spheres). Residues involved in side-chain rearrangements are indicated.

Motional dynamics of W335 could also be observed in atomistic simulations. We calculated the distance between the W335-Cη2 sidechain atom and the F372-Cα atom located on the opposing side of the Palm-CoA binding pocket (Fig. 5.10A). When the Palm-CoA site was unoccupied (starting from both the HHAT-MB177-nhPalm-CoA-Heme or HHAT-IMP-1575 structural conformations) the indole ring of W335 moved towards F372 to block the Palm-CoA tunnel (Fig. 5.10C-D). When Palm-CoA was bound the tunnel remained open with W335 in an upright conformation (Fig. 5.10C,E).

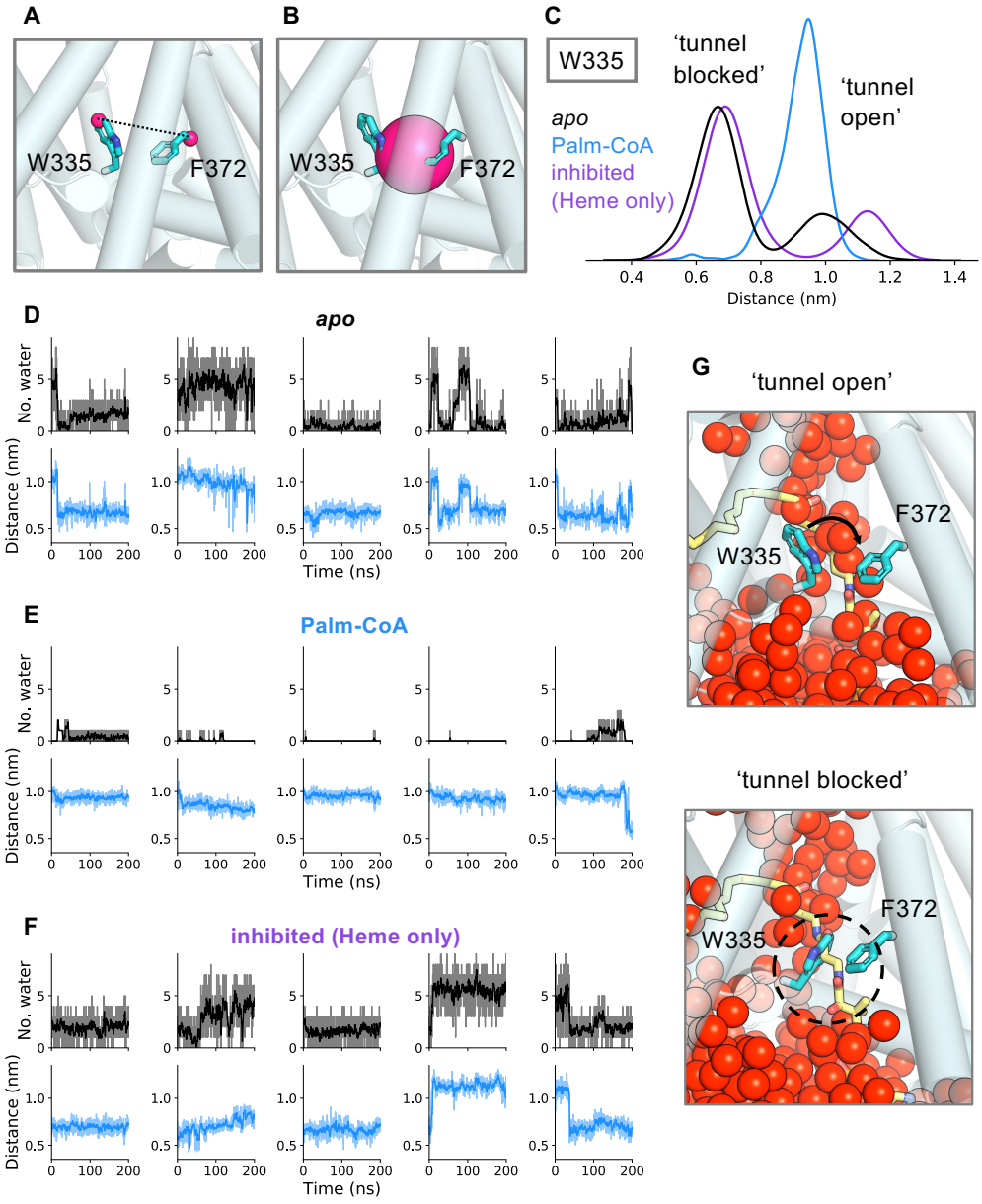


Figure 5.10: Gating of Palm-CoA binding pocket by W335.

A) Distance between W335-Cη$_2$ and F372-Cα atoms (pink) used to assess W335 sidechain motions. **B)** A sphere of radius 4 Å positioned at the centre of geometry of W335/F372 Cα atoms, used to assess gating of water in **D-G**. MDAnalysis[259] was used to calculate the updated position of the sphere at each simulation timepoint. **C)** Distribution of W335-F372 distances (see **A**) across 5 x 200 ns atomistic simulations of HHAT in *apo*, Palm-CoA and inhibited conformations. **D-F)** Correlation between number of water within the sphere (black) and W335/F372 distances (blue) across HHAT simulations in *apo* (**D**), Palm-CoA bound (**E**) and inhibited (**F**) states. The rolling mean is coloured opaque. **G)** Snapshots from the start ('tunnel open') and end ('tunnel blocked') of an *apo* HHAT simulation indicating gating of water (red spheres) movement through HHAT by W335 motions. The position of the Palm-CoA molecule (yellow sticks) is overlayed for reference.

Pocket gating by W335 formed a hydrophobic restriction in the Palm-CoA binding pocket (Fig. 5.10G). We calculated the number of water within a sphere (0.4 nm radius) positioned at the centre of geometry of the W335/F372 Cα atoms, updated each frame of the trajectory (Fig. 5.10B). A clear correlation is observed between the W335/F372 distance and the number of waters within the sphere (Fig. 5.10D-F). Together with the HHAT-IMP-1575 structure this suggests water flux through HHAT is prevented by movement of W335 or Palm-CoA binding.

5.3.4 Reaction centre dynamics

To further examine the reorientation of D339 in the HHAT-IMP-1575 structure we calculated the D339 sidechain angle across simulations (Fig. 5.11). D339 sampled two conformations when Palm-CoA was not bound, flicking between a H379 facing orientation and towards a triangular pocket lined by G341/L342 (Fig. 5.11A-B). Interestingly, Palm-CoA binding restricted the D339 sidechain to the 'pocket facing' orientation.

It is intriguing that a residue proposed to function in HHAT catalysis points away from the reaction centre when the substrate is bound. We calculated the pKas of D339 and H379 across simulations using pKaTraj [228,292] (Fig. 5.11C). The pKa of D339 was more basic (~5.6) and H379 more acidic (~4.6) than expected compared to standard pKa values for these residues (red dashed lines). The pKas of D339/H379 are likely shifted due to a number of proximal hydrophobic residues and, crucially, the order of acid-base pairing is reversed. More standard pKa values for these residues are observed infrequently and only when D339 faces H379 and comes within hydrogen bonding distance (Fig. 5.11D). However, the D339-H379 distance is generally too large (average 0.6 ± 0.1 nm across simulations) for a direct hydrogen bond, thus communication between residues is likely bridged by water. Combined with observations that Palm-CoA binding restricts D339 to the pocket facing orientation we suggest D339 may play a role in priming H379 for catalysis in the absence of substrates rather than directly participating in the catalytic mechanism.

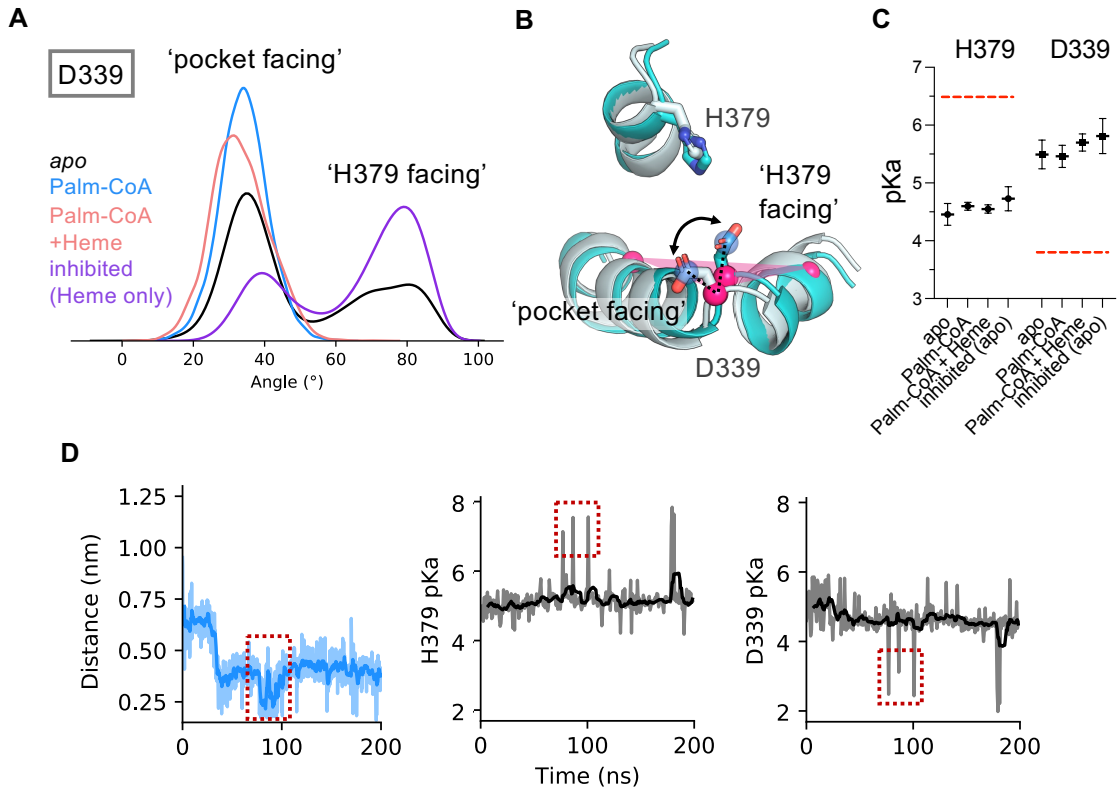


Figure 5.11: Reaction centre dynamics.

A) Distribution of D339 sidechain angles (defined as the angle between a vector formed by the Cα and Cγ atoms of D339 and a plane formed by the Cα atoms of M334, D339 and L346) across atomistic simulations of HHAT in *apo*, Palm-CoA bound, Palm-CoA plus Heme bound and inhibited states. **B)** Overlay of the start (light cyan) and end (cyan) snapshots from a simulation of *apo* HHAT. The position of the sidechain vector (black line) with respect to the plane (pink) is indicated. **C)** pKa values for D339 and H379 across HHAT simulations in distinct ligand bound states. pKa values were calculated using pKaTraj and are shown as the mean ± S.E.M. Red dashed lines indicate standard pKa values for histidine/aspartate residues. **D)** Minimum distance between D339/H379 (blue) and calculated pKa values (black) with time. The red box indicates reversion of pKas to their expected values when H379/D339 directly interact.

5.3.5 Assessing lipid-like densities around HHAT

The 2.7 Å HHAT-MB177-nhPalm-CoA-Heme structure revealed several additional lipid-like densities surrounding HHAT, including two densities protruding into the protein core. A comparison of simulation derived lipid poses with structural densities is described in section 2.2.7, including quantitative ranking of different lipids binding to the same site and how well poses/densities matched using Q scores[180].

One site was modelled as a bound cholesterol molecule in the deposited PDB structure, which could form the site of SHH-cholesterol engagement[179]. To probe this further we performed atomistic simulations of HHAT with cholesterol bound within the pocket. In all replicates, the plane of the cholesterol A ring becomes distorted within the pocket compared to unbound bilayer cholesterol (Fig. 5.12). A number of cholesterol and sidechain poses within the pocket were tested, which gave similar results. This suggests ring distortion is not an artifact of the selected pose or the cholesterol parameters (bilayer cholesterol behaves normally) but the geometry of the binding site.

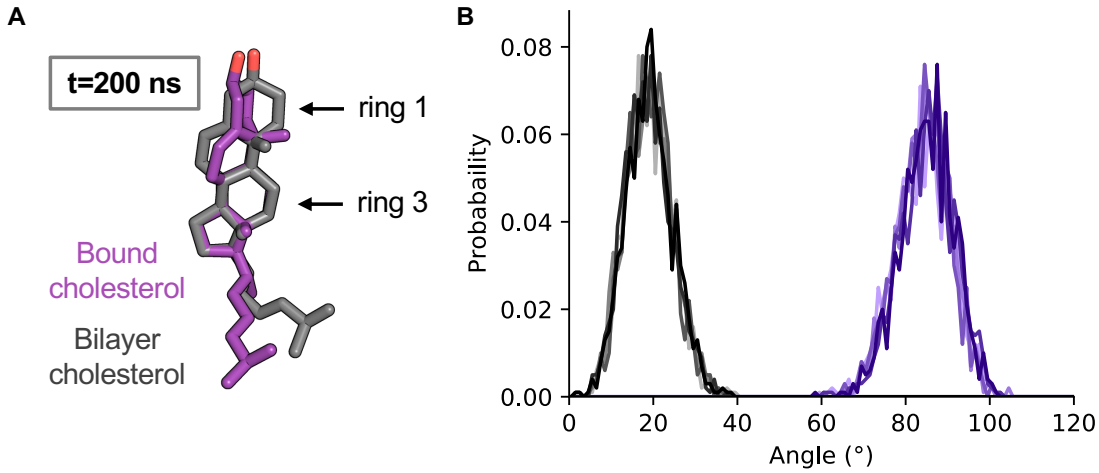


Figure 5.12: Probing a proposed cholesterol binding site.

A) Structural overlay of cholesterol bound within the proposed HHAT binding site (purple) with a free bilayer cholesterol (grey) at the end of a 200 ns atomistic simulation. **B)** Angle of ring 1 compared to ring 3 across 5 x 200 ns simulations of HHAT with cholesterol bound within the proposed binding site (purple) compared to that of bilayer cholesterol (grey). Angles were calculated between planes composed of three alternating carbon atoms within each six-membered ring. Each colour corresponds to an independent replicate.

The cryo-EM density at this site is broad and could be accounted for by an ensemble of sterol, phospholipid or detergent bound states (Fig. 5.13A). In simulations palmitate, cholesterol and CHS bound with similar RMSDs (Fig. 5.13B). Indeed, lipid exchange at this site may be necessary to permit displacement by the SHH-cholesterol upon binding. It is also possible that cholesterol conjugation to SHH could alter the lipid properties to resist distortion in a way that is not adequately captured by the

unconjugated cholesterol parameters used in simulations. Consistent with this proposal, CHS rings were not distorted within the pocket suggesting the additional chemical groups may stabilise sterol accommodation at this site. Structures of a HHAT-SHH complex are needed to elucidate the potential role of this pocket.

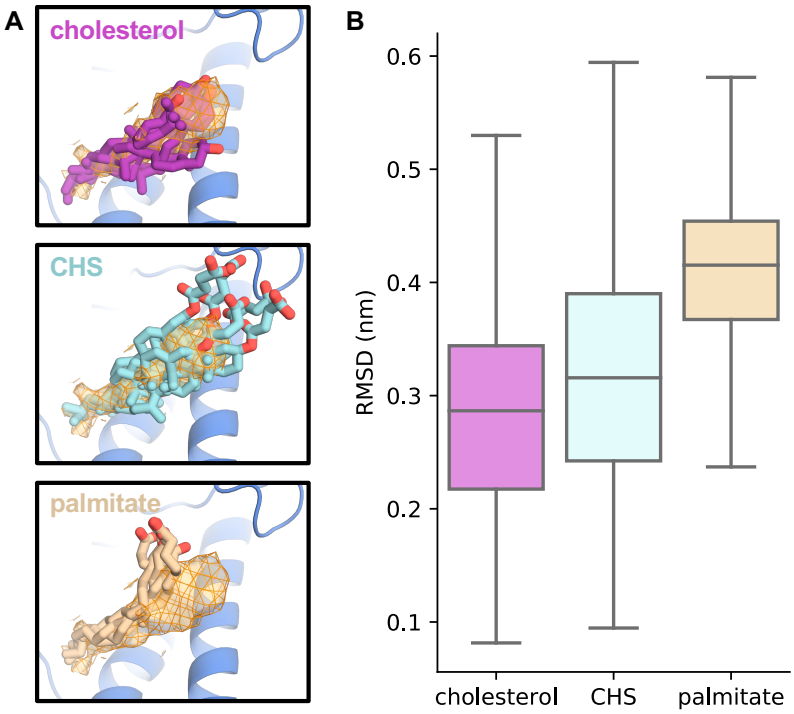


Figure 5.13: Ambiguous lipid-like density within HHAT.

A) Comparison of end snapshots of cholesterol, CHS and palmitate (sticks) from 5 x 200 ns simulations with the density (orange) within the proposed sterol binding site of HHAT. **B)** RMSD of the lipids across 5 x 200 ns simulations bound to HHAT.

While the study was being performed a second structure of HHAT was published in complex with two Fabs, Heme and two Palm-CoA molecules[282]. While one Palm-CoA molecule occupies a similar location to our structure the second Palm-CoA is proposed to bind within an “archway” formed between TM10-TM11 (Fig. 5.14A). In atomistic simulations of this HHAT structure bound to Heme and both Palm-CoAs the tail of the archway Palm-CoA was consistently displaced outwards by 0.9 nm from the proposed binding site (Fig. 5.14B). In addition, the CoA adenine base was shifted towards the cytoplasmic leaflet, in a manner that was not consistent with the observed density. We also observe an additional density at this site however in both cryo-EM maps the density is discontinuous across the membrane leaflets (Fig. 5.14C). We therefore

suggest the cytoplasmic portion of this density is occupied by a membrane phospholipid, evidenced also by our CG simulations (Fig. 2.12C), while the luminal portion may be accounted for by C387-palmitoylation, for which there is experimental evidence[275], or the GDN detergent belt (Fig. 5.14C).

These examples demonstrate how challenging modelling of ancillary cryo-EM densities can be and highlight a need for systematic methods to validate modelled ligands.

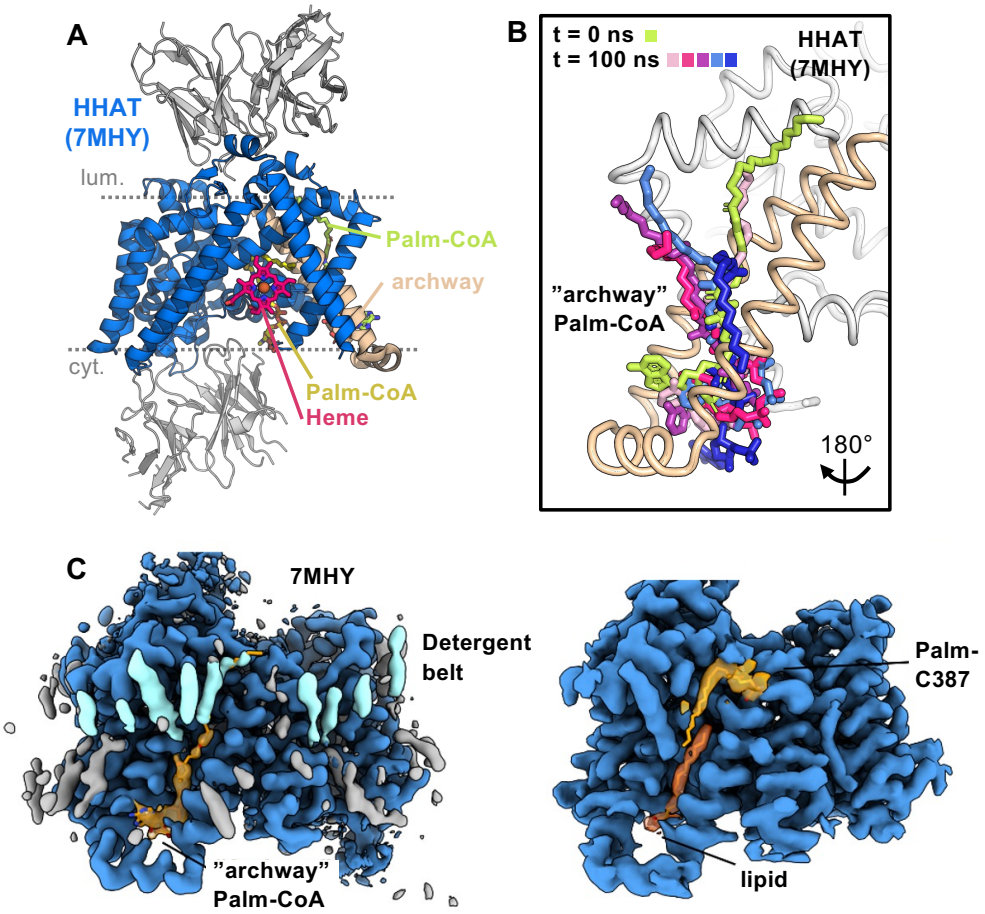


Figure 5.14: Probing the 'archway' Palm-CoA.

A) Cryo-EM structure of HHAT (blue) bound to two antibody fragments (grey), Heme (magenta) and two Palm-CoA molecules within the substrate binding pocket (yellow) and a second 'archway' site (lime) (7MHY)[282]. **B)** Comparison of the 'archway' Palm-CoA at the start (lime) and end (pinks/blues) of 5 x 100 ns atomistic simulations of the HHAT structure in **A** (with the antibody fragments removed). **C)** Comparison of densities surrounding HHAT within the 7MHY structure and the structure described in this thesis. The GDN detergent belt is coloured cyan, the 'archway' lipid is coloured orange and palmitoylated C387 is coloured light orange. **B/C** were adapted from Coupland *et al* [179].

5.3.6 A luminal exit gate for SHH

In CG simulations of HHAT, stable lipid binding was observed at the site of luminal deformation between TM5 and TM9. One tail of the bound lipid would partially insert between TM helices in proximity to H379 (Fig. 5.15) however, owing to the elastic network, expansion of the helices to enable binding of the whole lipid tail within the reaction centre of HHAT was not permissible. We reasoned the lipid tail may be mimicking the behaviour of the palmitoylated-SHH product within the reaction centre.

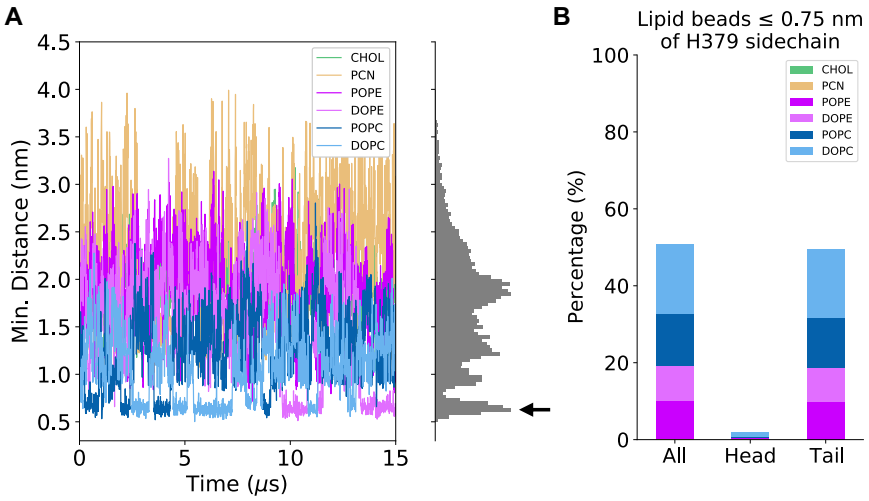


Figure 5.15: Insertion of lipids between HHAT transmembrane helices in CG simulations.

A) Minimum distances of luminal leaflet lipids to H379 sidechain beads across one 15 μs CG trajectory of HHAT. Events corresponding to insertion of lipid between TM5/TM9 (the luminal gate) in proximity to H379 are marked by an arrow. **B)** Relative contribution of lipid head group/tail beads at the luminal gate to HHAT interactions, for those interaction events with a minimum distance < 0.75 nm across 10 x 15 μs simulations.

To examine this further a CG frame with DOPC bound between helices was converted to atomistic resolution and simulated for 5x 200 ns. During simulations the luminal portion of TM5 tilted sideways, enabling full protrusion of one DOPC tail into the HHAT enzymatic core (Fig. 5.16A). Helical expansion only occurred on the luminal portion of TM5 and TM9 (hereafter referred to as the 'luminal gate'). The distance between the cytoplasmic regions of TM5/TM9 was unaffected by DOPC tail entry (Fig. 5.16B). Expansion of the luminal gate was accompanied by inward swivel of the Y377

sidechain to position between TM5 and TM9 (Fig. 5.16A). Luminal gate expansion and Y377 swivel were correlated (Fig. 5.16C-D) suggesting Y377 props open the luminal gate.

We wished to assess whether removal of the lipid was sufficient to permit luminal gate closure of whether additional driving forces were required to reset the gate. To do this, two additional sets of simulations were performed, starting from the end snapshots (t=200 ns) of each replicate. In one set of simulations the DOPC lipid was removed and a further 200 ns of equilibrium MD simulation performed. DOPC removal enabled a slight relaxation of the TM5-TM9 distance however trapping of the Y377 sidechain between helices did not permit full closure of the luminal gate (Fig. 5.16D). In the other set of simulations, the DOPC lipid was pulled into the membrane over a period of 10 ns using steered MD before 40 ns of equilibrium MD. As the lipid moved into the surrounding bilayer, the lipid tail came into contact with Y377 and the biasing force increased. Outward swivel of the Y377 sidechain enabled the lipid to move out of the cavity and relived the build-up of bias force (Fig. 5.16E-F). In cases where the Y377 sidechain remained outwardly swivelled, the luminal gate distance decreased to that prior to opening (Fig. 5.16E).

Using the DOPC lipid tail as palmitoylated-SHH mimetic we therefore propose the following product exit model for future experimental validation (Fig. 5.16G). In the absence of product, the luminal gate formed by TM5-TM9 remains closed and the Y377 sidechain points into the membrane. When the HHAT enzymatic core is occupied by palmitoylated-SHH the luminal gate expands and Y377 swivels inwards to prop open the gate. This also prevents movement of additional lipid tails into the reaction centre from the membrane i.e. ensures only one tail can localise within the pocket. Product movement into the surrounding membrane, forces Y377 outwards, enabling TM5 to tilt inwards and reset the gate. Thus, Y377 forms a unidirectional door (and reset mechanism) to permit closure of the luminal gate. Movement of palmitoylated-SHH into the membrane (rather than luminal solvent) is supported by observations of product inhibition in peptide assays using detergent solubilised HHAT.

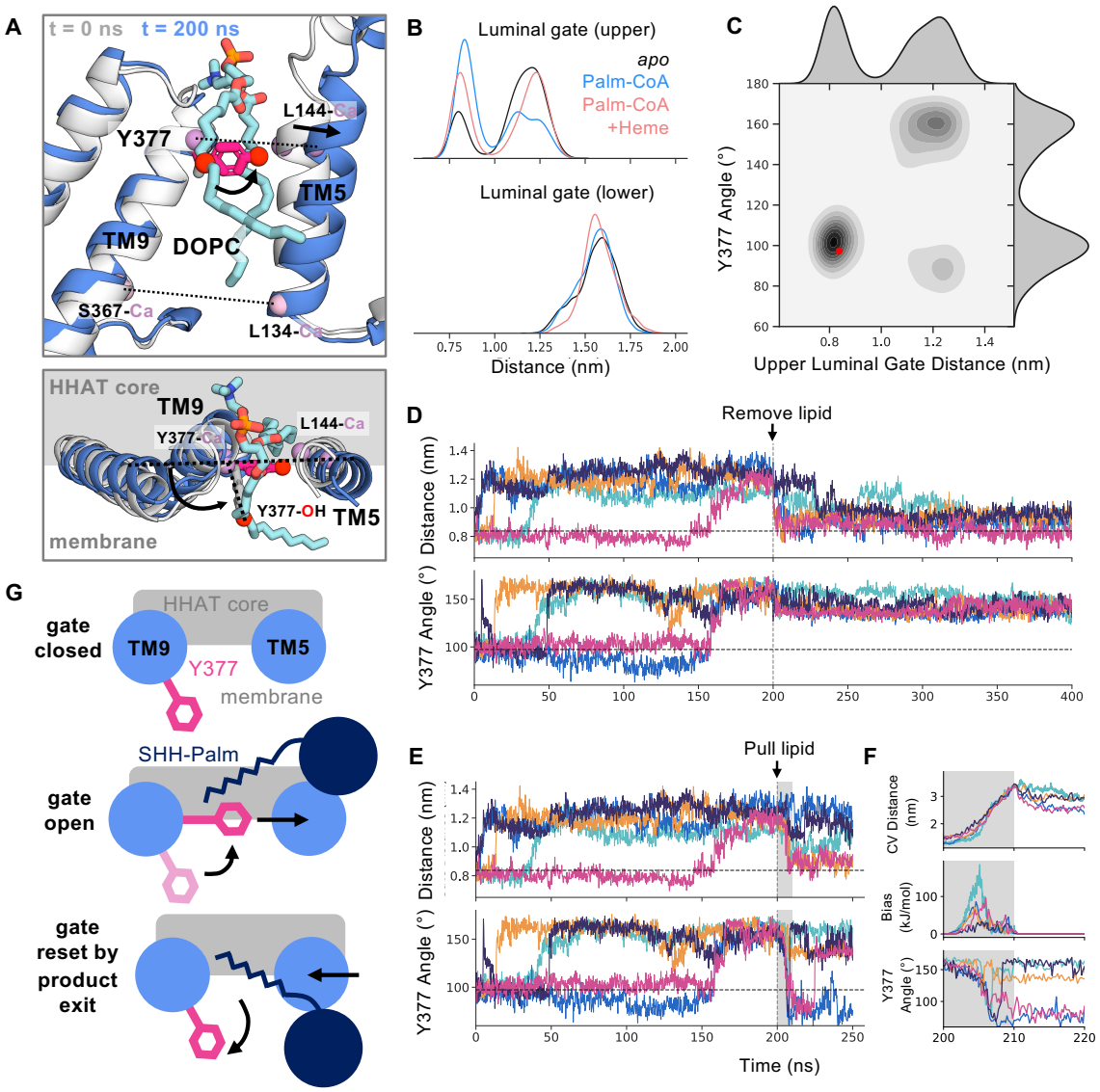


Figure 5.16: A luminal exit gate for palmitoylated-SHH.

A) Comparison of start (grey) and end (blue) snapshots from an atomistic simulation of HHAT initiated with DOPC bound at the luminal gate (cyan sticks), mimicking the palmitoylated-SHH product. Simulations were initiated using a frame backmapped from CG simulations where DOPC stably bound within the luminal gate. Y377 is shown as sticks at the start (grey) and end (magenta) of the simulation. Spheres indicate atoms used for the calculation of luminal gate distances (in e.g. **B**). The Y377 sidechain angle was defined between vectors formed by the hydroxyl-O and Cα atoms of Y377 and the Y377/L144 Cα atoms (top view, lower panel). **B)** Distribution of distances at the upper/lower portions of the luminal gate (as marked in **A**) across 5 x 200 ns atomistic simulations of each HHAT liganded system. 2D correlation (**C**) and correlation with time (**D/E**) of upper luminal gate distances and the Y377 sidechain angle across 5 x 200 ns simulations of *apo* HHAT. The red dot in **C** indicates the distance/angle at t=0 ns. At t=200 ns the DOPC lipid was either removed from the gate (**D**) or pulled laterally into the bilayer (**E**) over a period of 10 ns (grey) (see methods for details). **F)** Zoomed in view of the steered MD simulation in **E** indicating exit of DOPC through the gate as shown by the CV distance (top) as the work bias factor is

increased (middle) and the Y377 sidechain swings outward (bottom) to permit DOPC exit. **G)** Model of the proposed luminal exit gate mechanism for palmitoylated-SHH. In the absence of palmitoylated-SHH product the luminal gate is closed. When palmitoylated-SHH is located within HHAT, the upper region of TM5 tilts outwards allowing the sidechain of Y377 to swing inwards and plug the gate. As palmitoylated-SHH moves laterally into the membrane the Y377 sidechain is pushed outwards, enabling inward movement of TM5 to reset the luminal gate.

5.3.7 HHAT and other MBOATs

This section contains data collected whist supervising Megan Healey (Part II student 2020-2021), or extensions thereof.

We sought to contextualise HHAT dynamics compared to other MBOAT family members. Accordingly, CG and atomistic simulations of 5 other MBOATs (DltB, GOAT, PORCN, ACAT1 and DGAT1) were run, initiated from existing structures (DltB, ACAT1 and DGAT1)[266,267,270] or well-defined AlphaFold2 models (GOAT and PORCN)[127]. Bioinformatics analysis revealed loss of the Heme-coordinating C324 in all other MBOATs. In protein acylases this interface was stabilised by a conserved salt bridge in GOAT and PORCN or by π-π stacking between two phenylalanine residues in DltB. These interactions persisted across atomistic simulations (Fig. 5.17A). In ACAT1/DGAT1 this site instead forms the dimerization interface, suggesting diverse mechanisms for stabilisation the MBOAT fold have evolved across the family. Notably, all these interactions (including Heme binding) are theoretically reversible, which could permit state dependant protein regulation by relative (de)stabilisation of MBOAT tertiary structure.

Additionally, we calculated water movement through MBOAT members at equivalent positions to HHAT gating residues W335/F372 (see Fig. 5.10 for details). All MBOATs were capable of forming direct interactions between equivalent gating residues (Fig. 17B, dashed line at 0.35 nm) however water permeation was only restricted in HHAT, DltB, ACAT1 and to a lesser extent in DGAT1 (Fig. 5.17B). In GOAT/PORCN this region was notably wetted, likely due to replacement of one of the gating residues by a polar residue (S303 in GOAT and Y334 in PORCN). Additionally, in ACAT1/DGAT1 the whole TM helix moved inward to restrict water permeation in contrast to the

individual sidechain motions which occur in HHAT (Fig. 5.10). Thus, water gating in HHAT is most similar to DltB (mediated by sidechain motions) whereas ACAT1 and DGAT1 gate via a combination of sidechain motions and concerted TM helix movement. It remains unclear how (and whether) GOAT and PORCN prevent water movement across the bilayer.

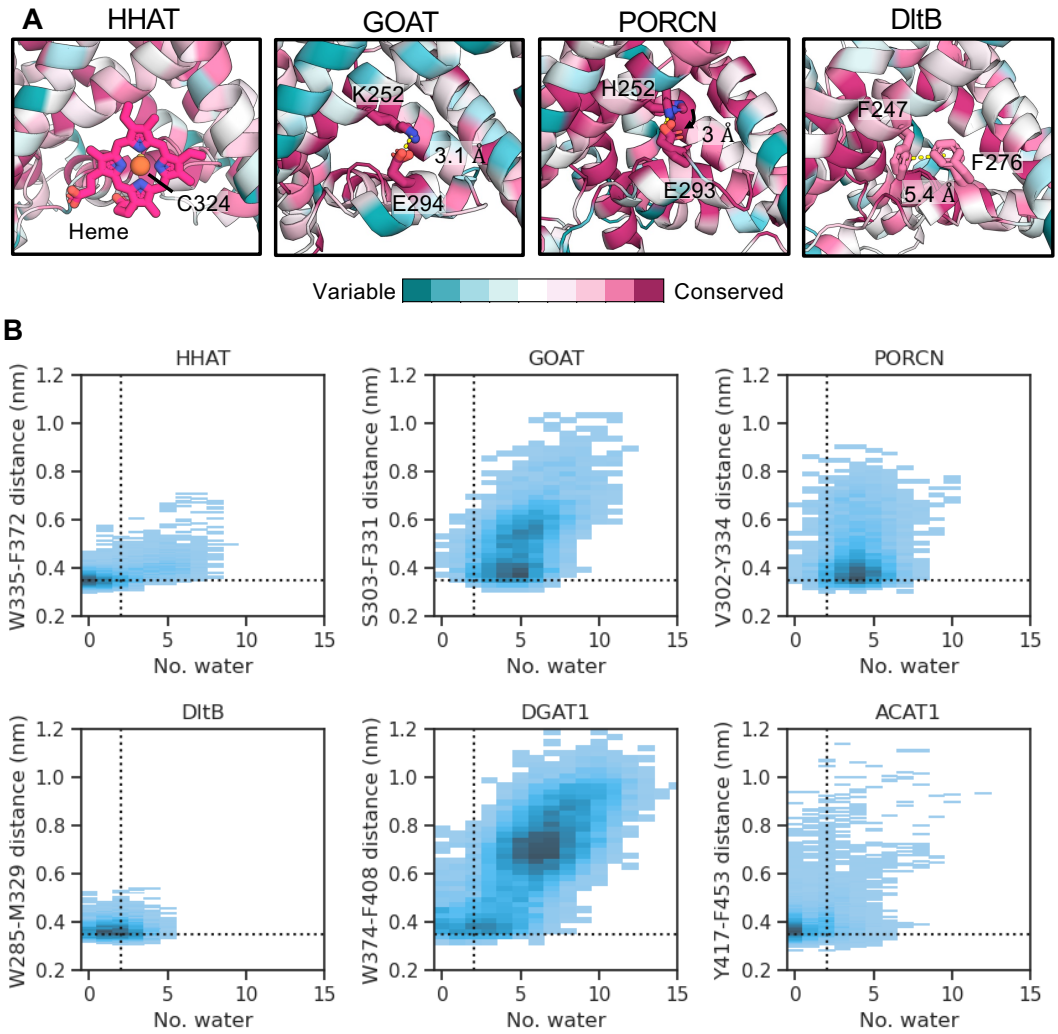


Figure 5.17: Comparison of HHAT with other MBOATs.

A) Sequence conservation of protein MBOAT homologues mapped onto the structures of HHAT (PDB: 7Q1U)[179] and DltB (PDB: 6BUG)[266] or AlphaFold2 models of GOAT and PORCN[127]. Structures are viewed from the equivalent interface to the HHAT Heme site. Sticks forming conserved salt-bridges or π-π stacking interactions are shown at the start and end of 200 ns atomistic simulations. **B)** 2D distribution plots of the minimum distance between gating residues (equivalent positions to W335/F372 in HHAT, see *y* axis for residues) *vs* the number of water within a sphere (radius 4 Å) positioned at the centre of the gating pair Cα atoms (see Fig. 5.10 for details) across 5 x 200 ns simulations of each MBOAT. Dashed lines at *y* = 0.35 nm indicate direct contacts between residues and *x* = 2 waters serve as the cut-off below which water permeation is prevented.

5.4 Conclusion

Simulations reveal a pronounced membrane distortion around HHAT (Fig. 5.5), which may reduce the energetic cost of palmitoyl transfer across the ER membrane and/or enhance the membrane diffusion rate of the protein[289]. Recent cryo-EM structures and simulations of ER transporters suggest protein induced bilayer thinning may be a generalised principle for molecular transport across the ER membrane[65]. Structures of e.g. EMC homologues, YidC and SecYEG, reveal splayed TM helices which form cytosolic funnels towards the bilayer midplane[293–295]. Structures and simulations of the signal peptidase complex also reveal membrane thinning within a 'TM window' which forms a 'molecular ruler' to measure substrate length[296]. Additionally, simulations of the DHHC acyltransferase-20 enzyme show a 1 nm deformation towards the catalytic site[281]. In parallel to the 'electrostatic steering' principle proposed for soluble enzymes, it is possible that membrane embedded enzymes/transporters such as HHAT may 'hydrophilically steer' substrate positioning to deeper within the membrane, thus forming both a physical and energetic funnel for substrate diffusion away from the reaction centre.

The ER membrane is low in cholesterol and glycosphingolipids and thought to have increased levels of tail unsaturation[278]. The resultant membrane is therefore thinned and less rigid in comparison to e.g. the cholesterol/sphingolipid rich plasma membrane[297]. This may render the ER membrane more amenable to deformation around the protein. A similar mechanism was proposed from biochemical assays of the nucleotide sugar transporter Vrg4 which localises in the thinner Golgi membrane[298]. Vrg4 activity was reduced in membranes composed of longer tailed lipids which effectively 'strait-jacket' the protein (i.e. the protein distorts to accommodate the membrane rather than vice versa). Thus, observations of HHAT membrane deformation may be a product of both protein function and localisation/trafficking to enhance enzymatic activity only within the native membrane context.

Several aspects of HHAT function remain uncertain; notably, the mechanism of HHAT catalysis and how palmitoylated-SHH product binds HHAT and is released into the

membrane. Structures of HHAT, solved e.g. in the presence of conformation-specific nanobodies and/or small-molecule mimetics may help to elucidate reaction intermediates. The simulations within this study provide mechanistic clues regarding these remaining queries:

A reversed acid-base pairing for D339/H379 compared to standard pKa values and the observation that D339 faces away from H379 when Palm-CoA substrate suggests D339 may not participate directly in the catalytic mechanism (Fig. 5.11). D339 may play a role in 'priming' H379 for catalysis by sequestering protons away from the imidazole nitrogen to free up the lone pair for nucleophilic attach, or by activation of the Palm-CoA carbonyl (Fig. 5.18). It is also unclear whether HHAT becomes conjugated to substrate as part of the reaction cycle, and whether SHH-C24 palmitoylation proceeds via direct N-palmitoylation or S,N-acyl shift. Mass spectrometry has recently been used to elucidate details of acylated reaction intermediates in the ELOVL fatty-acid elongase[299], an approach which may also be applicable to HHAT.

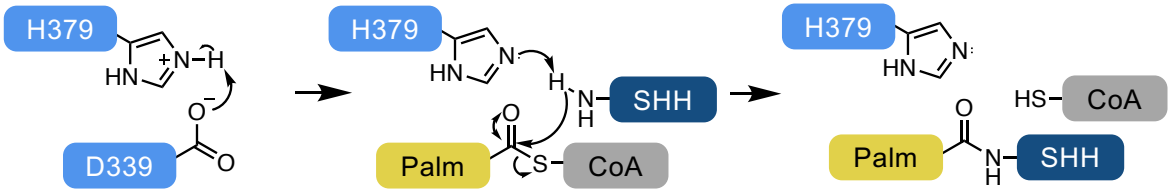


Figure 5.18: A potential catalytic mechanism for HHAT.

Schematic of one potential mechanism for SHH-palmitoylation by HHAT. D339 'primes' H379 for catalysis by making the nitrogen lone pair available for nucleophilic attack. H379 then mediates handover of the palmitoyl chain from Palm-CoA to SHH, producing the palmitoylated-SHH product.

Given observations of product inhibition in detergent solubilised HHAT assays and the experimental difficulties of purifying bi-lipidated SHH, it seems unlikely that SHH is released into the luminal solvent without aid from additional factors. Our simulations suggest an alternative exit route, into the membrane via a luminal gate formed by TM5/TM9 (Fig. 5.16). We observe stable binding of a product-mimetic lipid tail within the HHAT reaction centre, leading to re-arrangement of the luminal portion of TM5 and the Y377 sidechain, thus forming a uni-directional and reversible door for palmitoyl-

SHH exit. Mutational studies locking the TM5/TM9 helices using disulphide or salt bridges are under current investigation.

In addition to luminal *vs* lateral gate product exit mechanisms across the MBOAT family, HHAT shows a surprising degree of structural and dynamical divergence compared to other members. HHAT binds Heme at a conserved position, in place of predicted salt-bridges in GOAT/PORCN, π-π stacking in DltB or the dimeric interface of ACAT1/DGAT1 (Fig. 5.17A). The water gating properties of HHAT are most similar to DltB (Fig. 5.17B) and HHAT can only be structurally aligned to DltB (but not other MBOATs) without manual intervention. This raises the possibility that HHAT evolved from prokaryotic precursors via a distinct lineage to other eukaryotic MBOATs. This may also account for the mechanistic divergence of HHAT as a N- *vs* O- palmitoylase.

6. Anti-Ending

“Nature is chaos and art is pattern.”
Kate DiCamillo

6. Anti-Ending 147
6.1 Context is key 149
6.2 Acknowledge the complexity of the results – an eye to the future 149

This thesis set out to address the role of small-molecule ligands in the Hedgehog signalling pathway. Two ligands, palmitate and cholesterol, emerge as key players in HH pathway regulation and activation, either as free entities or via direct conjugation to SHH. Their diverse roles, explored in Chapters 3-5, are briefly summarised in Fig. 6.1.

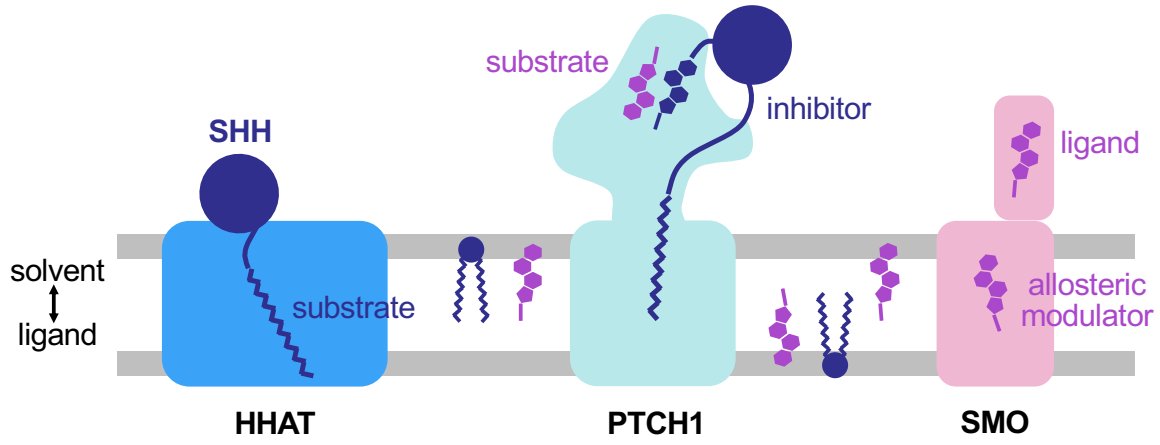


Figure 6.1: Lipids as ligands, modulators and inhibitors in Hedgehog signalling.

Summary schematic of the diverse roles of cholesterol (purple) and palmitate (blue) in HH pathway membrane proteins.

In HHAT palmitate acts as a substrate for transfer between the CoA carrier and SHH[179]. In PTCH1, cholesterol may be the substrate transported through the ECD and both SHH-palmitate and SHH-cholesterol contribute to PTCH1 inhibition via occlusion of the proposed transport tunnels[220]. In SMO, we suggest cholesterol functions as the *in vivo* HH pathway ligand by binding to the CRD and also as an allosteric modulator when bound to the TMD[252]. Furthermore, the lipid bilayer contains heterogenous pools of cholesterol and palmitoyl (as phospholipid tails) which bind externally to HH membrane protein TMDs, thereby functioning as solvent and/or specific ligands[99,120].

In learning to ask better questions, I would like to switch the focus of this chapter from '*What are the roles of cholesterol and palmitate in HH signalling?*' to '*How can cholesterol and palmitate play such diverse roles within the same pathway, protein, or binding site?*'

6.1 Context is key

The first component in answering this question is that lipid binding sites, even for the same lipid species, do not behave the same. Lipid site behaviour is influenced by the tertiary structure of proteins, the arrangement and composition of binding site residues, the site surface exposure to lipids/protein/water and neighbouring ligands[64], resulting in the site kinetics and affinity under a specific set of conditions. When the environment changes, such as e.g. protein conformational transitions, the affinity/kinetics of the site may also change[160]. Transducing these physical (or physically derived) properties to their biochemical outcomes remains challenging but is possible when collaborative approaches are effectively employed. In addition, the development of novel lipid probes[55,56,200], *in vivo* signalling[54,252] and *in situ* imaging approaches[300] greatly facilitate site contextualisation.

6.2 Acknowledge the complexity of the results – an eye to the future

The second consideration is that we must allow ourselves to recognise the complexity of our data. While, reductive and isolative approaches can play key roles in understanding the intricacies of membrane protein function, sometimes a story just isn't very simple. In Chapter 3 we focussed on exploring a transport function for PTCH1. It is also possible that PTCH1 plays multiple transport/flippase functions under different conditions however our current techniques cannot differentiate these. Similarly, dissecting cholesterol signalling and allosteric effects on SMO in Chapter 4 was challenging. MD simulations are often noisy, and it can be difficult to unpick biologically relevant dynamics from background motions, especially when it is unclear biochemically whether multiple ligand binding sites are redundant or function in elegant synergy.

The future of structural biology, I believe, lies in learning to embrace the dynamism of structural data. This may include use of machine learning methods to untangle molecular patterns[301,302], cryo-EM particle heterogeneity[303] or assist protein modelling[127]. It could also include understanding how membrane protein localisation shifts conformational equilibria using e.g., integrative tomographic data and larger

scale modelling[304]. This is particularly relevant to membrane partitioned processes such as those emerging from the cilium.

Additionally, there are several experiments which are required to better understand HH signalling and the feasibility of therapeutic intervention. The mechanism of PTCH1 is still unknown. Within this thesis we provide energetic contextualisation to PTCH1 transport mechanisms coupled to Na^+/K^+ movement. One key experiment to assess whether K^+ is implicated in PTCH1 function would be to measure the membrane potential of the cilium. If similar to the plasma membrane potential, it would not be possible for PTCH1 to be K^+ coupled because the free energy stored across the membrane would approach zero. Recent development of a DNA-based voltmeter which can be targeted to specific membranes could facilitate this[305]. Furthermore, investigations into the feasibility of delivering drugs upstream of SMO are needed. For HHAT, there would need to be a method of intracellular drug delivery and targeting to the ER membrane which incorporates considerations about the relative hydrophilicity of the HHAT Palm-CoA binding site. Evidently, further work is required to 'bridge the gap' between basic research and prevention of severe developmental pathologies.

Finally, I would like to touch on the human aspects of scientific research beyond the immediate focus of study. Research is driven by human conversation and hence subject to underlying societal inequities. There is still a bias in biophysical and computational sciences towards white, male academics, which undoubtedly influences the way science is conducted/funded and who benefits from our research[306]. Future scientific research must centre voices which have been historically or consequentially excluded from scientific spaces via systems of empire, racism, classism, queerphobia, ableism and patriarchy[307–309]. We must actively work to forge democratized science which encompasses human diversity and does not frame itself as politically objective. This may include alternative forms of scientific organisation such as through global collectives[310] which respect indigenous/ancient knowledge systems in parallel to our own. It could also include mentorship and/or financial stewardships schemes specifically for underrepresented people to access University resources. We must consider who benefits from our scientific outputs (especially those leading to e.g. biotechnology and/or drug applications) and lever operations which are financially and morally just compared to e.g. current industrial mechanisms. Ultimately,

our science must be diverse, collaborative, non-hierarchal, open and gregarious to effectively tackle complex planetary problems and mobilise cultural change.

In acknowledging the complexity of our data (thematic, technological and human elements) we learn to sit with uncertainty as a means to effective understanding. At times, structural biology feels like having a photo of a tree in winter and a tree in summer and trying to deduce how leaves grow. MD helps to turn the seasons but perhaps nature is more beautiful because it cannot ever be fully understood.

Biology is weird and magic.

7. References

1. Ingham, P. W., Nakano, Y. & Seger, C. Mechanisms and functions of Hedgehog signalling across the metazoa. *Nat. Rev. Genet.* **12**, 393–406 (2011).
2. Kong, J. H., Siebold, C. & Rohatgi, R. Biochemical mechanisms of vertebrate hedgehog signaling. *Development* **146**, 1–17 (2019).
3. Nusslein-Volhard, C. & Wieschaus, E. Mutations affecting segment number and polarity in Drosphilia. *Nature* **287**, 795–801 (1980).
4. Ingham, P. W. & McMahon, A. P. Hedgehog signaling in animal development: Paradigms and principles. *Genes Dev.* **15**, 3059–3087 (2001).
5. Bumcrot, D. A., Takada, R. & McMahon, A. P. Proteolytic processing yields two secreted forms of sonic hedgehog. *Mol. Cell. Biol.* **15**, 2294–2303 (1995).
6. Porter, J. A. *et al.* The product of hedgehog autoproteolytic cleavage active in local and long-range signalling. *Nature* **374**, 363–366 (1995).
7. Porter, J. A., Young, K. E. & Beachy, P. A. Cholesterol modification of hedgehog signaling proteins in animal development. *Science.* **274**, 255–259 (1996).
8. Pepinsky, R. B. *et al.* Identification of a palmitic acid-modified form of human Sonic hedgehog. *J. Biol. Chem.* **273**, 14037–14045 (1998).
9. Buglino, J. A. & Resh, M. D. Hhat is a palmitoylacyltransferase with specificity for N-palmitoylation of Sonic Hedgehog. *J. Biol. Chem.* **283**, 22076–22088 (2008).
10. Chamoun, Z. *et al.* Skinny Hedgehog, an acyltransferase required for palmitoylation and activity of the Hedgehog signal. *Science.* **293**, 2080–2084 (2001).
11. Burke, R. *et al.* Dispatched, a novel sterol-sensing domain protein dedicated to the release of cholesterol-modified Hedgehog from signaling cells. *Cell* **99**, 803–815 (1999).
12. Tukachinsky, H., Kuzmickas, R. P., Jao, C. Y., Liu, J. & Salic, A. Dispatched and Scube mediate the efficient secretion of the cholesterol-modified Hedgehog ligand. *Cell Rep.* **2**, 308–320 (2012).
13. Ma, Y. *et al.* Hedgehog-mediated patterning of the mammalian embryo requires transporter-like function of dispatched. *Cell* **111**, 63–75 (2002).
14. Briscoe, J. & Thérond, P. P. The mechanisms of Hedgehog signalling and its roles in development and disease. *Nat. Rev. Mol. Cell Biol.* **14**, 418–431 (2013).
15. Rohatgi, R., Milenkovic, L. & Scott, M. P. Patched1 Regulates Hedgehog Signaling at the Primary Cilium. *Science.* **317**, 372–376 (2007).
16. Marigo, V., Davey, R. A., Zuo, Y., Cunningham, J. M. & Tabin, C. J. Biochemical evidence that Patched is the Hedgehog receptor. *Nature* **384**, 176–179 (1996).
17. Stone, D. M. *et al.* The tumour-suppressor gene patched encodes a candidate receptor for Sonic hedgehog. *Nature* **384**, 129–134 (1996).
18. Carpenter, D. *et al.* Characterization of two patched receptors for the vertebrate hedgehog protein family. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 13630–13634 (1998).
19. Wierbowski, B. M. *et al.* Hedgehog Pathway Activation Requires Coreceptor-

Catalyzed, Lipid-Dependent Relay of the Sonic Hedgehog Ligand. *Dev. Cell* **55**, 1–18 (2020).
20. Huang, P. *et al.* Structural basis for catalyzed assembly of the Sonic hedgehog–Patched1 signaling complex. *Dev. Cell* **57**, 670-685.e8 (2022).
21. Kinnebrew, M. *et al.* Patched 1 reduces the accessibility of cholesterol in the outer leaflet of membranes. *Elife* **10**, 1–25 (2021).
22. Luchetti, G. *et al.* Cholesterol activates the G-protein coupled receptor smoothened to promote hedgehog signaling. *Elife* **5**, 1–22 (2016).
23. Happ, J. T. *et al.* A PKA Inhibitor Motif within Smoothened Controls Hedgehog Signal Transduction. *bioRxiv* 1–38 (2021). doi:10.1101/2021.07.05.451193
24. Arveseth, C. D. *et al.* Smoothened Transduces Hedgehog Signals via Activity-Dependent Sequestration of PKA Catalytic Subunits. *PLoS Biol.* **19**, e3001191 (2021).
25. Hui, C.-C., Slusarski, D., Platt, K. A., Holmgren, R. & Joyner, A. L. Expression of Three Mouse Homologs of the Drosphilia Segment Polarity Gene cubitus interruptus, Gli, Gli-2 and Gli-3, in Ectoderm and Mesoderm-derived Tissues Suggests Multiple Roles during Postimplantation Development. *Dev. Biol.* **162**, 402–413 (1994).
26. Wu, F., Zhang, Y., Sun, B., McMahon, A. P. & Wang, Y. Hedgehog Signaling: From Basic Biology to Cancer Therapy. *Cell Chem. Biol.* **24**, 252–280 (2017).
27. Hahn, H. *et al.* Rhabdomyosarcomas and radiation hypersensitivity in a mouse model of Gorlin syndrome. *Nat. Med.* **4**, 619–622 (1998).
28. Johnson, R. L. *et al.* Human homolog of patched, a candidate gene for the basal cell nevus syndrome. *Science.* **272**, 1668–1671 (1996).
29. Hui, C. C. & Angers, S. Gli proteins in development and disease. *Annu. Rev. Cell Dev. Biol.* **27**, 513–537 (2011).
30. Burness, C. B. Sonidegib: First Global Approval. *Drugs* **75**, 1559–1566 (2015).
31. Aditya, S. & Rattan, A. Vismodegib: A smoothened inhibitor for the treatment of advanced basal cell carcinoma. *Indian Dermatol. Online J.* **4**, 365–368 (2013).
32. Atwood, S. X. *et al.* Smoothened Variants Explain the Majority of Drug Resistance in Basal Cell Carcinoma. *Cancer Cell* **27**, 342–353 (2015).
33. Corbit, K. C. *et al.* Vertebrate Smoothened functions at the primary cilium. *Nature* **437**, 1018–1021 (2005).
34. Goetz, S. C. & Anderson, K. V. The primary cilium: A signalling centre during vertebrate development. *Nat. Rev. Genet.* **11**, 331–344 (2010).
35. Plotnikova, O. V., Pugacheva, E. N. & Golemis, E. A. Primary Cilia and the Cell Cycle. *Methods Cell Biol* **94**, 137–160 (2009).
36. Ishikawa, T. Axoneme structure from motile cilia. *Cold Spring Harb. Perspect. Biol.* **9**, 1–16 (2017).
37. Sun, S., Fisher, R. L., Bowser, S. S., Pentecost, B. T. & Sui, H. Three-dimensional architecture of epithelial primary cilia. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 9370–9379 (2019).
38. Hoek, H. van den *et al.* In situ architecture of the ciliary base reveals the stepwise assembly of IFT trains. *Science.* **377**, abm6704 (2022). doi:10.7868/s0869565214210269
39. Kiesel, P. *et al.* The molecular structure of mammalian primary cilia revealed by cryo-electron tomography. *Nat. Struct. Mol. Biol.* **27**, 1115–1124 (2020).
40. Weiss, L. E., Milenkovic, L., Yoon, J., Stearns, T. & Moerner, W. E. Motional dynamics of single Patched1 molecules in cilia are controlled by Hedgehog

and cholesterol. *Proc. Natl. Acad. Sci.* **116**, 5550–5557 (2019).
41. Nachury, M. V. & Mick, D. U. Establishing and regulating the composition of cilia for signal transduction. *Nat. Rev. Mol. Cell Biol.* **20**, 389–405 (2019).
42. Domire, J. S. *et al.* Dopamine receptor 1 localizes to neuronal cilia in a dynamic process that requires the Bardet-Biedl syndrome proteins. *Cell. Mol. Life Sci.* **68**, 2951–2960 (2011).
43. Pigino, G. Intraflagellar transport. *Curr. Biol.* **31**, R530–R536 (2021).
44. Stepanek, L. & Pigino, G. Microtubule doublets are double-track railways for intraflagellar transport trains. *Science.* **352**, 721–724 (2016).
45. Porter, J. A. *et al.* Hedgehog patterning activity: Role of a lipophilic modification mediated by the carboxy-terminal autoprocessing domain. *Cell* **86**, 21–34 (1996).
46. Kowatsch, C., Woolley, R. E., Kinnebrew, M., Rohatgi, R. & Siebold, C. Structures of vertebrate Patched and Smoothened reveal intimate links between cholesterol and Hedgehog signalling. *Curr. Opin. Struct. Biol.* **57**, 204–214 (2019).
47. Nachtergaele, S. *et al.* Oxysterols are allosteric activators of the oncoprotein Smoothened. *Nat. Chem. Biol.* **8**, 211–220 (2012).
48. Byrne, E. F. X. *et al.* Structural basis of Smoothened regulation by its extracellular domains. *Nature* **535**, 517–522 (2016).
49. Huang, P. *et al.* Cellular Cholesterol Directly Activates Smoothened in Hedgehog Signaling. *Cell* **166**, 1176–1187 (2016).
50. Chen, J. K., Taipale, J., Cooper, M. K. & Beachy, P. A. Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothened. *Genes Dev.* **16**, 2743–2748 (2002).
51. Keeler, R. F. & Binns, W. Teratogenic Compounds of Veratrum californicum: Comparison of Cyclopian Effects of Steroidal Alkaloids from the Plant and Structurally Related Compounds from other Sources. *Teratology* **1**, 5–10 (1968).
52. Montesano, R. Inhomogenous Distribution of Filipin-Sterol Complexes in the Ciliary Membrane of Rat Tracheal Epithelium. *Am. J. Anat.* **156**, 139–145 (1979).
53. Taipale, J., Cooper, M. K., Maiti, T. & Beachy, P. A. Patched acts catalytically to suppress the activity of Smoothened. *Nature* **418**, 892–896 (2002).
54. Kinnebrew, M. *et al.* Cholesterol accessibility at the ciliary membrane controls hedgehog signaling. *Elife* **8**, 1–28 (2019).
55. Johnson, K. A. & Radhakrishnan, A. *The use of anthrolysin O and ostreolysin A to study cholesterol in cell membranes*. *Methods in Enzymology* **649**, (Elsevier Inc., 2021).
56. Das, A., Brown, M. S., Anderson, D. D., Goldstein, J. L. & Radhakrishnan, A. Three pools of plasma membrane cholesterol and their relation to cholesterol homeostasis. *Elife* **3**, 1–16 (2014).
57. Zhang, Y. *et al.* Structural Basis for Cholesterol Transport-like Activity of the Hedgehog Receptor Patched. *Cell* **175**, 1–13 (2018).
58. Garcia, G., Raleigh, D. R. & Reiter, J. F. How the Ciliary Membrane Is Organized Inside-Out to Communicate Outside-In. *Curr. Biol.* **28**, R421–R434 (2018).
59. Nechipurenko, I. V. The Enigmatic Role of Lipids in Cilia Signaling. *Front. Cell Dev. Biol.* **8**, 1–10 (2020).
60. Stilling, S., Kalliakoudas, T., Benninghoven-Frey, H., Inoue, T. & Falkenburger,

B. H. PIP2 determines length and stability of primary cilia by balancing membrane turnovers. *Commun. Biol.* **5**, 1–15 (2022).
61. Garcia-Gonzalo, F. R. *et al.* Phosphoinositides Regulate Ciliary Protein Trafficking to Modulate Hedgehog Signaling. *Dev. Cell* **34**, 400–409 (2015).
62. Li, F. *et al.* Highlighting membrane protein structure and function: A celebration of the protein data bank. *J. Biol. Chem.* **296**, 1–50 (2021).
63. Stahelin, R. V., Scott, J. L. & Frick, C. T. Cellular and molecular interactions of phosphoinositides and peripheral proteins. *Chem. Phys. Lipids* **182**, 3–18 (2014).
64. Corradi, V. *et al.* Emerging Diversity In Lipid-Protein Interactions. *Chem. Rev.* **119**, 5775–5848 (2019).
65. Wu, X. & Rapoport, T. A. Translocation of Proteins through a Distorted Lipid Bilayer. *Trends Cell Biol.* **31**, 473–484 (2021).
66. Corradi, V. *et al.* Lipid−Protein Interactions Are Unique Fingerprints for Membrane Proteins. *ACS Cent. Sci* **4**, 709–717 (2018).
67. Habeck, M. *et al.* Stimulation, inhibition, or stabilization of Na,K-ATPase caused by specific lipid interactions at distinct sites. *J. Biol. Chem.* **290**, 4829–4842 (2015).
68. Whorton, M. R. & MacKinnon, R. Crystal structure of the mammalian GIRK2 K+ channel and gating regulation by G proteins, PIP2, and sodium. *Cell* **147**, 199–208 (2011).
69. Albert, A. D., Young, J. E. & Yeagle, P. L. Rhodopsin-cholesterol interactions in bovine rod outer segment disk membranes. *Biochim. Biophys. Acta. Biomembr.* **1285**, 47–55 (1996).
70. Jojoa-Cruz, S. *et al.* Structural insights into the Venus flytrap mechanosensitive ion channel Flycatcher1. *Nat. Commun.* **13**, 1–11 (2022).
71. Ansell, T. B. *et al.* Relative Affinities of Protein-Cholesterol Interactions from Equilibrium Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **17**, 6548–6558 (2021).
72. Stansfeld, P. J., Hopkinson, R., Ashcroft, F. M. & Sansom, M. S. P. PIP2-binding site in Kir channels: Definition by multiscale biomolecular simulations. *Biochemistry* **48**, 10926–10933 (2009).
73. Hansen, S. B., Tao, X. & MacKinnon, R. Structural basis of PIP2 activation of the classical inward rectifier K+ channel Kir2.2. *Nature* **477**, 495–498 (2011).
74. Yen, H. Y. *et al.* PtdIns(4,5)P2 stabilizes active states of GPCRs and enhances selectivity of G-protein coupling. *Nature* **559**, 423–427 (2018).
75. Wilson, K. A., Wang, L., Lin, Y. C. & O'Mara, M. L. Investigating the lipid fingerprint of SLC6 neurotransmitter transporters: a comparison of dDAT, hDAT, hSERT, and GlyT2. *Biochim. Biophys. Acta. Adv.* **1**, 1–14 (2021).
76. Duncan, A. L., Corey, R. A. & Sansom, M. S. P. Defining how multiple lipid species interact with inward rectifier potassium (Kir2) channels. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 7803–7813 (2020).
77. Ingólfsson, H. I. *et al.* Capturing Biologically Complex Tissue-Specific Membranes at Different Levels of Compositional Complexity. *J. Phys. Chem. B* **124**, 7819–7829 (2020).
78. Wilson, K. A., Macdermott-Opeskin, H. I., Riley, E., Lin, Y. & O'Mara, M. L. Understanding the Link between Lipid Diversity and the Biophysical Properties of the Neuronal Plasma Membrane. *Biochemistry* **59**, 3010–3018 (2020).
79. Ingólfsson, H. I. *et al.* Computational Lipidomics of the Neuronal Plasma Membrane. *Biophys. J.* **113**, 2271–2280 (2017).

80. Sharp, L., Salari, R. & Brannigan, G. Boundary lipids of the nicotinic acetylcholine receptor: Spontaneous partitioning via coarse-grained molecular dynamics simulation. *Biochim. Biophys. Acta. Biomembr.* **1861**, 887–896 (2019).
81. Casiraghi, M. *et al.* Functional Modulation of a G Protein-Coupled Receptor Conformational Landscape in a Lipid Bilayer. *J. Am. Chem. Soc.* **138**, 11170–11175 (2016).
82. Corey, R. A. *et al.* Identification and assessment of cardiolipin interactions with E. coli inner membrane proteins. *Sci. Adv.* **7**, eabh2217 (2021).
83. Wang, C. *et al.* Structure of the human smoothened receptor bound to an antitumour agent. *Nature* **497**, 338–343 (2013).
84. Weierstall, U. *et al.* Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nat. Commun.* **5**, 1–6 (2014).
85. Wang, C. *et al.* Structural basis for Smoothened receptor modulation and chemoresistance to anticancer drugs. *Nat. Commun.* **5**, 1–11 (2014).
86. Nachtergaele, S. *et al.* Structure and function of the Smoothened extracellular domain in vertebrate Hedgehog signaling. *Elife* **2**, 1–32 (2013).
87. Bai, Q., Shen, Y., Jin, N., Liu, H. & Yao, X. Molecular modeling study on the dynamical structural features of human smoothened receptor and binding mechanism of antagonist LY2940680 by metadynamics simulation and free energy calculation. *Biochim. Biophys. Acta. Gen. Subj.* **1840**, 2128–2138 (2014).
88. Sinha, N., Chowdhury, S. & Sarkar, R. R. Deciphering structural stability and binding mechanisms of potential antagonists with smoothened protein. *J. Biomol. Struct. Dyn.* **36**, 2917–2937 (2018).
89. Tu, J. *et al.* Molecular modeling study on resistance of WT/D473H SMO to antagonists LDE-225 and LEQ-506. *Pharmacological Research* **129**, 491–499 (2018).
90. Dash, R. C., Maschinot, C. R. & Hadden, M. K. A molecular dynamics approach to identify an oxysterol-based hedgehog pathway inhibitor. *Biochim. Biophys. Acta. Gen. Subj.* **1861**, 168–177 (2017).
91. Nasution, M. A. F., Stephanie, F. & Tambunan, U. S. F. Pharmacophore-based virtual screening and molecular docking simulation of flavonoids as smoothened protein inhibitor of Hedgehog signaling pathways. *AIP Conf. Proc.* **2168**, 1–9 (2019).
92. Che Omar, M. T. Data analysis of molecular dynamics simulation trajectories of β-sitosterol, sonidegib and cholesterol in smoothened protein with the CHARMM36 force field. *Data Br.* **33**, 1–9 (2020).
93. Zhang, X. *et al.* Crystal structure of a multi-domain human smoothened receptor in complex with a super stabilizing ligand. *Nat. Commun.* **8**, 1–10 (2017).
94. Deshpande, I. *et al.* Smoothened stimulation by membrane sterols drives Hedgehog pathway activity. *Nature* **571**, 284–288 (2019).
95. An, X. *et al.* Deciphering the Allosteric Effect of Antagonist Vismodegib on Smoothened Receptor Deactivation Using Metadynamics Simulation. *Front. Chem.* **7**, 1–11 (2019).
96. Agostino, M. & Pohl, S. Ö. G. Activation barriers in Class F G protein-coupled receptors revealed by umbrella sampling simulations. *Org. Biomol. Chem.* **18**, 9816–9825 (2020).
97. Wright, S. C. *et al.* A conserved molecular switch in Class F receptors

regulates receptor activation and pathway selection. *Nat. Commun.* **10**, (2019).
98. Turku, A., Schihada, H., Kozielewicz, P., Bowin, C.-F. & Schulte, G. Residue 6.43 defines receptor function in class F GPCRs. *Nat. Commun.* **12**, 1–14 (2021).
99. Hedger, G. *et al.* Cholesterol Interaction Sites on the Transmembrane Domain of the Hedgehog Signal Transducer and Class F G Protein-Coupled Receptor Smoothened. *Structure* **27**, 549–559 (2019).
100. 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.* **49**, W544–W550 (2021).
101. Sejdiu, B. I. & Tieleman, D. P. Lipid-Protein Interactions are a Unique Property and Defining Feature of G Protein-Coupled Receptors. *Biophys. J.* **118**, 1887–1900 (2020).
102. Kumari, S., Mitra, A. & Bulusu, G. Structural dynamics of Smoothened (SMO) in the ciliary membrane and its interaction with membrane lipids. *Biochim. Biophys. Acta. Biomembr.* **1864**, 1–20 (2022).
103. Yoon, H. J., Jeong, H., Lee, H. H. & Jang, S. Molecular dynamics study with mutation shows that N-terminal domain structural re-orientation in Niemann-Pick type C1 is required for proper alignment of cholesterol transport. *J. Neurochem.* **156**, 967–978 (2020).
104. Martínez-Archundia, M., Hernández Mojica, T. G., Correa-Basurto, J., Montaño, S. & Camacho-Molina, A. Molecular dynamics simulations reveal structural differences among wild-type NPC1 protein and its mutant forms. *J. Biomol. Struct. Dyn.* **38**, 3527–3532 (2020).
105. Elghobashi-Meinhardt, N. Computational Tools Unravel Putative Sterol Binding Sites in the Lysosomal NPC1 Protein. *J. Chem. Inf. Model.* **59**, 2432–2441 (2019).
106. Vanharanta, L. *et al.* High-content imaging and structure-based predictions reveal functional differences between Niemann-Pick C1 variants. *Traffic* **21**, 386–397 (2020).
107. Elghobashi-Meinhardt, N. Cholesterol transport in wild-type NPC1 and P691S: Molecular dynamics simulations reveal changes in dynamical behavior. *Int. J. Mol. Sci.* **21**, 1–14 (2020).
108. Long, T. *et al.* Structural basis for itraconazole-mediated NPC1 inhibition. *Nat. Commun.* **11**, 1–11 (2020).
109. Estiu, G., Khatri, N. & Wiest, O. Computational studies of the cholesterol transport between NPC2 and the N-terminal domain of NPC1 (NPC1(NTD)). *Biochemistry* **52**, 6879–6891 (2013).
110. Hodošček, M. & Elghobashi-Meinhardt, N. Simulations of NPC1(NTD):NPC2 protein complex reveal cholesterol transfer pathways. *Int. J. Mol. Sci.* **19**, 1–14 (2018).
111. Vargiu, A. V. *et al.* Computer simulations of the activity of RND efflux pumps. *Res. Microbiol.* **169**, 384–392 (2018).
112. Seeger, M. A. *et al.* Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. *Science.* **313**, 1295–1298 (2006).
113. Murakami, S., Nakashima, R., Yamashita, E., Matsumoto, T. & Yamaguchi, A. Crystal structures of a multidrug transporter reveal a functionally rotating mechanism. *Nature* **443**, 173–179 (2006).
114. Fischer, N. & Kandt, C. Three ways in, one way out: Water dynamics in the trans-membrane domains of the inner membrane translocase AcrB. *Proteins*

Struct. Funct. Bioinforma. **79**, 2871–2885 (2011).
115. Eicher, T. *et al.* Coupling of remote alternating-access transport mechanisms for protons and substrates in the multidrug efflux pump AcrB. *Elife* **3**, 1–26 (2014).
116. Vargiu, A. V. *et al.* Water-mediated interactions enable smooth substrate transport in a bacterial efflux pump. *Biochim. Biophys. Acta. Gen. Subj.* **1862**, 836–845 (2018).
117. Moseng, M. A. *et al.* Cryo-em structures of cusa reveal a mechanism of metal-ion export. *MBio* **12**, e00452-21 (2021).
118. Su, C. C. *et al.* Structures and transport dynamics of a Campylobacter jejuni multidrug efflux pump. *Nat. Commun.* **8**, 1–11 (2017).
119. Kovachka, S., Malloci, G. & Vargiu, V. Molecular insights into the Patched1 drug efflux inhibitory activity of panicein A hydroquinone : a computational study †. *Phys. Chem. Chem. Phys.* **23**, 8013–8022 (2021).
120. Zhong, C. & Wang, B. Regulation of Cholesterol Binding to the Receptor Patched1 by its interactions With the Ligand Sonic Hedgehog (Shh). *Front. Mol. Biosci.* **9**, 1–11 (2022).
121. Hwang, S., Thangapandian, S. & Lee, K. W. Molecular Dynamics Simulations of Sonic Hedgehog-Receptor and Inhibitor Complexes and Their Applications for Potential Anticancer Agent Discovery. *PLoS One* **8**, e68271 (2013).
122. Gong, X. *et al.* Structural basis for the recognition of Sonic Hedgehog by human Patched1. *Science.* **361**, eaas8935 (2018).
123. Hitzenberger, M., Schuster, D. & Hofer, T. S. The Binding Mode of the Sonic Hedgehog Inhibitor Robotnikinin, a Combined Docking and QM/MM MD Study. *Front. Chem.* **5**, 1–14 (2017).
124. Paris, F. *et al.* A novel variant of DHH in a familial case of 46,XY disorder of sex development: Insights from molecular dynamics simulations. *Clin. Endocrinol. (Oxf).* **87**, 539–544 (2017).
125. Galli, L. M. *et al.* Determination of the membrane topology of PORCN , an O-acyl transferase that modifies Wnt signalling proteins. *Open Biol.* **11**, 200400 (2021).
126. Campaña, M. B. *et al.* The ghrelin O-acyltransferase structure reveals a catalytic channel for transmembrane hormone acylation. *J. Biol. Chem.* **294**, 14166–14174 (2019).
127. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
128. Leach, A. R. *Molecular modelling: principles and applications*. (2001).
129. Monticelli, L. & Salonen, E. *Biomolecular Simulations: Methods and Protocols*. (2013).
130. Corey, R. A., Sansom, M. S. P. & Stansfeld, P. J. Identification and Characterization of Specific Protein–Lipid Interactions Using Molecular Simulation. in *Computational Design of Membrane Proteins, Methods in Molecular Biology* **2315**, Chapter 8 (2021).
131. Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
132. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
133. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing

of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236 (1996).
134. Hansen, H. S. & Hunenberger, P. H. A Reoptimized GROMOS Force Field for Hexopyranose-Based Carbohydrates Accounting for the Relative Free Energies of Ring Conformers, Anomers, Epimers, Hydroxymethyl Rotamers, and Glycosidic Linkage Conformers. *J. Comput. Chem.* **32**, 174–182 (2012).
135. Leonard, A. N., Wang, E., Monje-Galvan, V. & Klauda, J. B. Developing and Testing of Lipid Force Fields with Applications to Modeling Cellular Membranes. *Chem. Rev.* **119**, 6227–6269 (2019).
136. Dickson, C. J. *et al.* Lipid14: The amber lipid force field. *J. Chem. Theory Comput.* **10**, 865–879 (2014).
137. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
138. Kirschner, K. N. *et al.* GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J. Comput. Chem.* **32**, 174–182 (2012).
139. Best, R. B., Buchete, N. V. & Hummer, G. Are current molecular dynamics force fields too helical? *Biophys. J.* **95**, 7–9 (2008).
140. Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ 1 and χ 2 dihedral angles. *J Chem Theory Comput J Chem Theory Comput. Sept.* **11**, 3257–3273 (2012).
141. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
142. MacKerell, A. D., Feig, M. & Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulation. *J. Comput. Chem.* **25**, 1400–1415 (2004).
143. Jo, S., Lim, J. B., Klauda, J. B. & Im, W. CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophys. J.* **97**, 50–58 (2009).
144. Galvelis, R., Doerr, S., Damas, J. M., Harvey, M. J. & De Fabritiis, G. A Scalable Molecular Force Field Parameterization Method Based on Density Functional Theory and Quantum-Level Machine Learning. *J. Chem. Inf. Model.* **59**, 3485–3493 (2019).
145. Skjevik, Å. A., Madej, B. D., Walker, R. C. & Teigen, K. LIPID11: A modular framework for lipid simulations using amber. *J. Phys. Chem. B* **116**, 11124–11136 (2012).
146. 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* **111**, 7812–7824 (2007).
147. Darré, L. *et al.* SIRAH: A structurally unbiased coarse-grained force field for proteins with aqueous solvation and long-range electrostatics. *J. Chem. Theory Comput.* **11**, 723–739 (2015).
148. Limbach, H. J., Arnold, A., Mann, B. A. & Holm, C. ESPResSo-an extensible simulation package for research on soft matter systems. *Comput. Phys. Commun.* **174**, 704–727 (2006).
149. De Jong, D. H. *et al.* Improved parameters for the martini coarse-grained protein force field. *J. Chem. Theory Comput.* **9**, 687–697 (2013).
150. Song, W., Duncan, A. L. & Sansom, M. S. P. Modulation of adenosine A2a

receptor oligomerization by receptor activation and PIP2 interactions. *Structure* **29**, 1312–1325 (2021).
151. Liu, Y., de Vries, A. H., Pezeshkian, W. & Marrink, S. J. Capturing Membrane Phase Separation by Dual Resolution Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **17**, 5876–5884 (2021).
152. Reddy, T. *et al.* Nothing to Sneeze At: A Dynamic and Integrative Computational Model of an Influenza A Virion. *Structure* **23**, 584–597 (2015).
153. Bruininks, B. M. H., Souza, P. C. T., Ingolfsson, H. & Marrink, S. J. A molecular view on the escape of lipoplexed dna from the endosome. *Elife* **9**, 1–16 (2020).
154. Souza, P. C. T. *et al.* Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **18**, 382–388 (2021).
155. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**–**2**, 19–25 (2015).
156. Marrink, S. J. *et al.* Computational Modeling of Realistic Cell Membranes. *Chem. Rev.* **119**, 6184–6226 (2019).
157. Song, W. *et al.* PyLipID: A Python Package for Analysis of Protein-Lipid Interactions from Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **18**, 1188–1201 (2022).
158. Ansell, T. B. *et al.* LipIDens: Simulation assisted interpretation of lipid densities in cryo-EM structures of membrane proteins. *bioRxiv* (2022). doi:10.1101/2022.06.30.49823
159. Hedger, G. *et al.* Lipid-Loving ANTs: Molecular Simulations of Cardiolipin Interactions and the Organization of the Adenine Nucleotide Translocase in Model Mitochondrial Membranes. *Biochemistry* **55**, 6238–6249 (2016).
160. Song, W., Yen, H. Y., Robinson, C. V. & Sansom, M. S. P. State-dependent Lipid Interactions with the A2a Receptor Revealed by MD Simulations Using In Vivo-Mimetic Membranes. *Structure* **27**, 392–403 (2019).
161. Ansell, T. B., Song, W. & Sansom, M. S. P. The Glycosphingolipid GM3 Modulates Conformational Dynamics of the Glucagon Receptor. *Biophys. J.* **119**, 300–313 (2020).
162. Rubin, D. B. The Bayesian Bootstrap. *Ann. Stat.* **9**, 130–134 (1981).
163. McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **109**, 1528–1532 (2015).
164. Zhang, H. *et al.* Structure of the full-length glucagon class B G-protein-coupled receptor. *Nature* **546**, 259–264 (2017).
165. Barbera, N., Ayee, M. A. A., Akpa, B. S. & Levitan, I. Molecular Dynamics Simulations of Kir2.2 Interactions with an Ensemble of Cholesterol Molecules. *Biophys. J.* **115**, 1264–1280 (2018).
166. Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **P10008**, 1–12 (2008).
167. Humphrey, W., Dalke, A. and Schulten, K. VMD -Visual Molecular Dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
168. Shrake, A. & Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* **79**, 361–371 (1973).
169. Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. *Proc. 9th Python Sci. Conf.* **SciPy 2010**, 92–96 (2010).
170. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn.*

Res. **12**, 2825–2830 (2011).
171. Fantini, J. & Barrantes, F. J. How cholesterol interacts with membrane proteins: An exploration of cholesterol-binding sites including CRAC, CARC, and tilted domains. *Front. Physiol.* **4**, 1–9 (2013).
172. Taghon, G. J. *et al.* Predictable cholesterol binding sites in GPCRs lack consensus motifs. *Structure* **29**, 1–8 (2021).
173. Sefah, E. & Mertz, B. Bacterial Analogs to Cholesterol Affect Dimerization of Proteorhodopsin and Modulates Preferred Dimer Interface. *J. Chem. Theory Comput.* **17**, 2502–2512 (2021).
174. Larsen, A. H., Tata, L., John, L. H. & Sansom, M. S. P. Membrane-binding mechanism of the EEA1 FYVE domain revealed by multi-scale molecular dynamics simulations. *PLoS Comput. Biol.* **17**, 1–14 (2021).
175. Qi, C., Minin, G. Di, Vercellino, I., Wutz, A. & Korkhov, V. M. Structural basis of sterol recognition by human hedgehog receptor PTCH1. *Sci. Adv.* **5**, 1–10 (2019).
176. Baker, M. R. *et al.* Cryo-EM structure of type 1 IP3R channel in a lipid bilayer. *Commun. Biol.* **4**, 1–11 (2021).
177. Nadezhdin, K. D. *et al.* Structural mechanism of heat-induced opening of a temperature-sensitive TRP channel. *Nat. Struct. Mol. Biol.* **28**, 564–572 (2021).
178. Zhao, P. *et al.* Structure and activation mechanism of the hexameric plasma membrane H+-ATPase. *Nat. Commun.* **12**, 1–11 (2021).
179. Coupland, C. E. *et al.* Structure, mechanism, and inhibition of Hedgehog acyltransferase. *Mol. Cell* **81**, 1–14 (2021).
180. Pintilie, G. *et al.* Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nat. Methods* **17**, 328–334 (2020).
181. Saotome, K. *et al.* Structures of the otopetrin proton channels Otop1 and Otop3. *Nat. Struct. Mol. Biol.* **26**, 518–525 (2019).
182. Kumar, P., Cymes, G. D. & Grosman, C. Structure and function at the lipid – protein interface of a pentameric ligand-gated ion channel. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 1–9 (2021).
183. Flegler, V. J. *et al.* Mechanosensitive channel gating by delipidation. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 1–8 (2021).
184. Corey, R. A., Stansfeld, P. J. & Sansom, M. S. P. The energetics of protein-lipid interactions as viewed by molecular simulations. *Biochem. Soc. Trans.* **48**, 25–37 (2020).
185. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12562–12566 (2002).
186. Dror, R. O. *et al.* Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci.* **108**, 13118–13123 (2011).
187. Souza, P. C. T. *et al.* Protein – ligand binding with the coarse-grained Martini model. *Nat. Commun.* **11**, 1–11 (2020).
188. Corey, R. A., Vickery, O. N., Sansom, M. S. P. & Stansfeld, P. J. Insights into Membrane Protein–Lipid Interactions from Free Energy Calculations. *J. Chem. Theory Comput.* **15**, 5727–5736 (2019).
189. Mattedi, G., Acosta-Gutiérrez, S., Clark, T. & Gervasio, F. L. A combined activation mechanism for the glucagon receptor. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 15414–15422 (2020).
190. Hub, J. S., De Groot, B. L. & Van Der Spoel, D. g_wham- A free Weighted Histogram Analysis implementation including robust error and autocorrelation estimates. *J. Chem. Theory Comput.* **6**, 3713–3720 (2010).

191. Klimovich, P. V., Shirts, M. R. & Mobley, D. L. Guidelines for the analysis of free energy calculations. *J. Comput. Aided. Mol. Des.* **29**, 397–411 (2015).
192. King, E., Aitchison, E., Li, H. & Luo, R. Recent Developments in Free Energy Calculations for Drug Discovery. *Front. Mol. Biosci.* **8**, 1–24 (2021).
193. Mey, A. S. J. S. *et al.* Best Practices for Alchemical Free Energy Calculations. *Living J Comput Mol Sci.* **2**, 1–48 (2020).
194. Laio, A. & Gervasio, F. L. Metadynamics: A method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports Prog. Phys.* **71**, 126601 (2008).
195. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100**, 1–4 (2008).
196. Lee, J. Y. & Lyman, E. Predictions for cholesterol interaction sites on the A2A adenosine receptor. *J. Am. Chem. Soc.* **134**, 16512–16515 (2012).
197. Genheden, S., Essex, J. W. & Lee, A. G. G protein coupled receptor interactions with cholesterol deep in the membrane. *Biochim. Biophys. Acta. Biomembr.* **1859**, 268–281 (2017).
198. Sharp, L. & Brannigan, G. Spontaneous lipid binding to the nicotinic acetylcholine receptor in a native membrane. *J. Chem. Phys.* **154**, 185102-1-185102–13 (2021).
199. Alessandri, R. *et al.* Martini 3 Coarse-Grained Force Field: Small Molecules. *Adv. Theory Simulations* **5**, 1–19 (2022).
200. Cabanos, C., Wang, M., Han, X. & Hansen, S. B. A Soluble Fluorescent Binding Assay Reveals PIP2 Antagonism of TREK-1 Channels. *Cell Rep.* **20**, 1287–1294 (2017).
201. Ko, D. C., Binkley, J., Sidow, A. & Scott, M. P. The integrity of a cholesterol-binding pocket in Niemann-Pick C2 protein is necessary to control lysosome cholesterol levels. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 2518–2525 (2003).
202. Domicevica, L., Koldsø, H. & Biggin, P. C. Multiscale molecular dynamics simulations of lipid interactions with P-glycoprotein in a complex membrane. *J. Mol. Graph. Model.* **80**, 147–156 (2018).
203. Newport, T. D., Sansom, M. S. P. & Stansfeld, P. J. The MemProtMD database: A resource for membrane-embedded protein structures and their lipid interactions. *Nucleic Acids Res.* **47**, D390–D397 (2019).
204. Xu, P. *et al.* Structural insights into the lipid and ligand regulation of serotonin receptors. *Nature* **592**, 469–473 (2021).
205. Tseng, T. T. *et al.* The RND permease superfamily: An ancient, ubiquitous and diverse family that includes human disease and development proteins. *J. Mol. Microbiol. Biotechnol.* **1**, 107–125 (1999).
206. Qi, X., Schmiege, P., Coutavas, E., Wang, J. & Li, X. Structures of human Patched and its complex with native palmitoylated sonic hedgehog. *Nature* **560**, 128–132 (2018).
207. Rohatgi, R. & Snell, W. J. The ciliary membrane. *Curr. Opin. Cell Biol.* **22**, 541–546 (2010).
208. Radhakrishnan, A., Rohatgi, R. & Siebold, C. Cholesterol access in cellular membranes controls Hedgehog signaling. *Nat. Chem. Biol.* **16**, 1303–1313 (2020).
209. Raleigh, D. R. *et al.* Cilia-Associated Oxysterols Activate Smoothened. *Mol. Cell* **72**, 316–327 (2018).
210. Qi, X. *et al.* Cryo-EM structure of oxysterol-bound human Smoothened coupled

to a heterotrimeric Gi. *Nature* **571**, 279–285 (2019).
211. Qi, X., Schmiege, P., Coutavas, E. & Li, X. Two Patched molecules engage distinct sites on Hedgehog yielding a signaling-competent complex. *Science* **362**, eaas8843 (2018).
212. Qian, H. *et al.* Inhibition of tetrameric Patched1 by Sonic Hedgehog through an asymmetric paradigm. *Nat. Commun.* **10**, 1–9 (2019).
213. Zhang, Y. *et al.* Hedgehog pathway activation through conformational blockade of the Patched sterol conduit. *Proc. Natl. Acad. Sci.* **117**, 28838–28846 (2020).
214. Luo, Y. *et al.* Cryo-EM Study of Patched in Lipid Nanodisc Suggests a Structural Basis for Its Clustering in Caveolae. *Structure* **29**, 1–9 (2021).
215. Yamaguchi, A., Nakashima, R. & Sakurai, K. Structural basis of RND-type multidrug exporters. *Front. Microbiol.* **6**, 1–19 (2015).
216. Petrov, K., Wierbowski, B. M., Liu, J. & Salic, A. Distinct Cation Gradients Power Cholesterol Transport at Different Key Points in the Hedgehog Signaling Pathway. *Dev. Cell* **55**, 1–14 (2020).
217. Wang, Q. *et al.* Dispatched uses Na+ flux to power release of lipid-modified Hedgehog. *Nature* **599**, 320–324 (2021).
218. Myers, B. R., Neahring, L., Zhang, Y., Roberts, K. J. & Beachy, P. A. Rapid, direct activity assays for Smoothened reveal Hedgehog pathway regulation by membrane cholesterol and extracellular sodium. *Proc. Natl. Acad. Sci.* **114**, E11141–E11150 (2017).
219. Tukachinsky, H., Petrov, K., Watanabe, M. & Salic, A. Mechanism of inhibition of the tumor suppressor Patched by Sonic Hedgehog. *Proc. Natl. Acad. Sci.* **113**, E5866–E5875 (2016).
220. Rudolf, A. F. *et al.* The morphogen Sonic hedgehog inhibits its receptor Patched by a pincer grasp mechanism. *Nat. Chem. Biol.* **15**, 975–982 (2019).
221. Šali, A. & Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial restraints. *Journal of Molecular Biology* **234**, 779–815 (1993).
222. 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.* **5**, 2531–2543 (2009).
223. 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. Theory Comput.* **11**, 2144–2155 (2015).
224. Stansfeld, P. J. *et al.* MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes. *Structure* **23**, 1350–1361 (2015).
225. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
226. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
227. Hess Berk, Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
228. Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *J. Chem. Theory Comput.* **7**, 2284–2295 (2011).
229. Lee, J. *et al.* CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER,

OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **12**, 405–413 (2016).
230. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
231. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255–268 (1984).
232. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).
233. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An *N* ·log(*N*) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
234. Gumbart, J., Khalili-Araghi, F., Sotomayor, M. & Roux, B. Constant electric field simulations of the membrane potential illustrated with simple systems. *Biochim. Biophys. Acta. Biomembr.* **1818**, 294–302 (2012).
235. Shirts, M. R. & Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **129**, 1–10 (2008).
236. Wu, X., Yan, R., Cao, P., Qian, H. & Yan, N. Structural advances in sterol-sensing domain-containing proteins. *Trends Biochem. Sci.* **47**, 289–300 (2022).
237. Möckl, L. The Emerging Role of the Mammalian Glycocalyx in Functional Membrane Organization and Immune System Regulation. *Front. Cell Dev. Biol.* **8**, 1–14 (2020).
238. Nikaido, H. RND transporters in the living world. *Res. Microbiol.* **169**, 363–371 (2018).
239. Du, D. *et al.* Multidrug efflux pumps: structure, function and regulation. *Nat. Rev. Microbiol.* **16**, 523–539 (2018).
240. Wang, B., Fallon, J. F. & Beachy, P. A. Hedgehog-Regulated Processing of Gli3 Produces an Anterior/Posterior Repressor Gradient in the Developing Vertebrate Limb. *Cell* **100**, 423–434 (2000).
241. Strigini, M. & Cohen, S. M. A Hedgehog activity gradient contributes to AP axial patterning of the Drosophila wing. *Development* **124**, 4697–4705 (1997).
242. Infante, R. E. *et al.* NPC2 facilitates bidirectional transfer of cholesterol between NPC1 and lipid bilayers, a step in cholesterol egress from lysosomes. *Proc. Natl. Acad. Sci.* **105**, 15287–15292 (2008).
243. Delling, M., Decaen, P. G., Doerner, J. F., Febvay, S. & Clapham, D. E. Primary cilia are specialized calcium signalling organelles. *Nature* **504**, 311–314 (2013).
244. Steck, T. L., Ye, J. & Lange, Y. Probing red cell membrane cholesterol movement with cyclodextrin. *Biophys. J.* **83**, 2118–2125 (2002).
245. Poulsen, H. *et al.* Neurological disease mutations compromise a γ-terminal ion pathway in the Na+/K+-ATPase. *Nature* **467**, 99–102 (2010).
246. Li, J. *et al.* Crystal Structures of Membrane Transporter MmpL3, an Anti-TB Drug Target. *Cell* **176**, 636–648 (2019).
247. Cannac, F. *et al.* Cryo-EM structure of the Hedgehog release protein Dispatched . *Sci. Adv.* eaay7928 (2019).
248. Li, X. *et al.* Structure of human Niemann-Pick C1 protein. *Proc. Natl. Acad. Sci.* **113**, 8212–8217 (2016).
249. Huang, P. *et al.* Structural Basis of Smoothened Activation in Hedgehog Signaling. *Cell* **174**, 312–324 (2018).

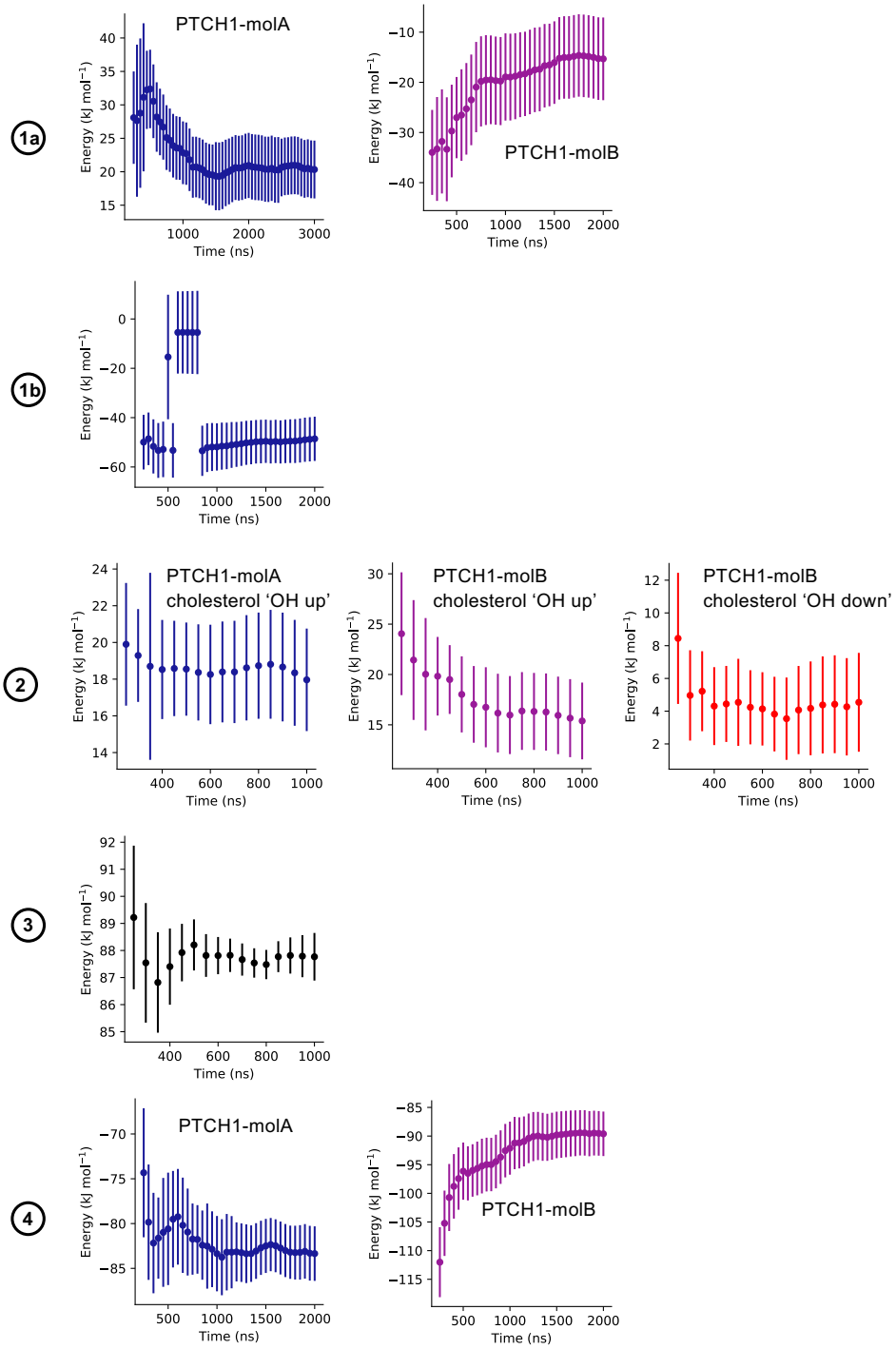
250. Qi, X., Friedberg, L., De Bose-Boyd, R., Long, T. & Li, X. Sterols in an intramolecular channel of Smoothened mediate Hedgehog signaling. *Nat. Chem. Biol.* **16**, 1368–1375 (2020).
251. Ballesteros, J. A. & Weinstein, H. Integrated Methods for the Construction of Three-Dimensional Models and Computational Probing of Structure-Function Relations in G Protein-Coupled Receptors. *Methods Neurosci.* **25**, 366–428 (1995).
252. Kinnebrew, M. *et al.* Patched 1 regulates Smoothened by controlling sterol binding to its extracellular cysteine-rich domain. *Sci. Adv.* **8**, eabm5563 (2022).
253. Xiao, X. *et al.* Cholesterol Modification of Smoothened Is Required for Hedgehog Signaling. *Mol. Cell* **66**, 154–162 (2017).
254. Lee, J. *et al.* CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *J. Chem. Theory Comput.* **15**, 775–786 (2019).
255. Horn, H. W. *et al.* Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **120**, 9665–9678 (2004).
256. Westerlund, A. M., Fleetwood, O., Pérez-Conesa, S. & Delemotte, L. Network analysis reveals how lipids and other cofactors influence membrane protein allostery. *J. Chem. Phys.* **153**, 141103-1-141103–10 (2020).
257. Bonomi, M. *et al.* PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **180**, 1961–1972 (2009).
258. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **185**, 604–613 (2014).
259. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).
260. Latorraca, N. R., Venkatakrishnan, A. J. & Dror, R. O. GPCR Dynamics: Structures in Motion. *Chem. Rev.* **117**, 139–155 (2017).
261. Guixà-González, R. *et al.* Membrane cholesterol access into a G-protein-coupled receptor. *Nat. Commun.* **8**, 1–12 (2017).
262. Wu, X. *et al.* A novel protein encoded by circular SMO RNA is essential for Hedgehog signaling activation and glioblastoma tumorigenicity. *Genome Biol.* **22**, 1–29 (2021).
263. Ermilova, I. & Lyubartsev, A. P. Cholesterol in phospholipid bilayers: positions and orientations inside membranes with different unsaturation degrees. *Soft Matter* **15**, 78–93 (2019).
264. Bennett, W. F. D. & Tieleman, D. P. Molecular simulation of rapid translocation of cholesterol, diacylglycerol, and ceramide in model raft and nonraft membranes. *J. Lipid Res.* **53**, 421–429 (2012).
265. Hofmann, K. A superfamily of membrane-bound O-acyltransferases with implications for Wnt signaling. *Trends Biochem. Sci.* **25**, 111–112 (2000).
266. Ma, D. *et al.* Crystal structure of a membrane-bound O-acyltransferase. *Nature* **562**, 286–290 (2018).
267. Qian, H., Zhao, X., Yan, R., Gao, S. & Sun, X. Structural basis for catalysis and substrate specificity of human ACAT1. *Nature* **581**, 1–30 (2020).
268. Guan, C. *et al.* Structural insights into the inhibition mechanism of human sterol O-acyltransferase 1 by a competitive inhibitor. *Nat. Commun.* **11**, 1–11 (2020).
269. Long, T., Sun, Y., Hassan, A., Qi, X. & Li, X. Structure of nevanimibe-bound

tetrameric human ACAT1. *Nature* **581**, 339–343 (2020).
270. Wang, L. *et al.* Structure and mechanism of human diacylglycerol O-acyltransferase 1. *Nature* **581**, 329–332 (2020).
271. Sui, X. *et al.* Structure and catalytic mechanism of a human triacylglycerol-synthesis enzyme. *Nature* **581**, 323–328 (2020).
272. Herr, P. & Basler, K. Porcupine-mediated lipidation is required for Wnt recognition by Wls. *Dev. Biol.* **361**, 392–402 (2012).
273. Yang, J., Brown, M. S., Liang, G., Grishin, N. V. & Goldstein, J. L. Identification of the Acyltransferase that Octanoylates Ghrelin, an Appetite-Stimulating Peptide Hormone. *Cell* **132**, 387–396 (2008).
274. Resh, M. D. Fatty acylation of proteins: The long and the short of it. *Prog. Lipid Res.* **63**, 120–131 (2016).
275. Konitsiotis, A. D. *et al.* Topological analysis of hedgehog acyltransferase, a multipalmitoylated transmembrane protein. *J. Biol. Chem.* **290**, 3293–3307 (2015).
276. Matevossian, A. & Resh, M. D. Membrane topology of hedgehog acyltransferase. *J. Biol. Chem.* **290**, 2235–2243 (2015).
277. Buglino, J. A. & Resh, M. D. Identification of conserved regions and residues within hedgehog acyltransferase critical for palmitoylation of Sonic hedgehog. *PLoS One* **5**, 17–20 (2010).
278. Jacquemyn, J., Cascalho, A. & Goodchild, R. E. The ins and outs of endoplasmic reticulum-controlled lipid biosynthesis. *EMBO Rep.* **18**, 1905–1921 (2017).
279. Melo, M. N., Ingólfsson, H. I. & Marrink, S. J. Parameters for Martini sterols and hopanoids based on a virtual-site description. *J. Chem. Phys.* **143**, 243152-1-243152–12 (2015).
280. Vickery, O. N. & Stansfeld, P. J. CG2AT2: An Enhanced Fragment-Based Approach for Serial Multi-scale Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **17**, 6472–6482 (2021).
281. Stix, R., Song, J., Banerjee, A. & Faraldo-Gómez, J. D. DHHC20 Palmitoyl-Transferase Reshapes the Membrane to Foster Catalysis. *Biophys. J.* **118**, 980–988 (2020).
282. Jiang, Y., Benz, T. L. & Long, S. B. Substrate and product complexes reveal mechanisms of Hedgehog acylation by HHAT. *Science.* **372**, 1215–1219 (2021).
283. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
284. Notredame, C., Higgins, D. G. & Heringa, J. T-coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).
285. Gabler, F. *et al.* Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Curr. Protoc. Bioinforma.* **72**, 1–30 (2020).
286. Celniker, G. *et al.* ConSurf: Using evolutionary data to raise testable hypotheses about protein function. *Isr. J. Chem.* **53**, 199–206 (2013).
287. Kühl, T. *et al.* Analysis of Fe(III) heme binding to cysteine-containing heme-regulatory motifs in proteins. *ACS Chem. Biol.* **8**, 1785–1793 (2013).
288. Long, J. *et al.* Identification of a family of fatty-acid-speciated sonic hedgehog proteins, whose members display differential biological properties. *Cell Rep.* **10**, 1280–1287 (2015).
289. Kreutzberger, A. J. B., Ji, M., Aaron, J., Mihaljević, L. & Urban, S. Rhomboid distorts lipids to break the viscosity-imposed speed limit of membrane

diffusion. *Science.* **363**, eaao0076 (2019).
290. Chovancova, E. *et al.* CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLoS Comput. Biol.* **8**, 23–30 (2012).
291. Lanyon-Hogg, T. *et al.* Photochemical Probe Identification of a Small-Molecule Inhibitor Binding Site in Hedgehog Acyltransferase (HHAT). *Angew. Chemie - Int. Ed.* **60**, 13542–13547 (2021).
292. Lee, C. *et al.* Crystal structure of the sodium-proton antiporter NhaA dimer and new mechanistic insights. *J. Gen. Physiol.* **144**, 529–544 (2014).
293. Kumazaki, K. *et al.* Structural basis of Sec-independent membrane protein insertion by YidC. *Nature* **509**, 516–519 (2014).
294. Berg, B. Van Den *et al.* X-ray structure of a protein-conducting channel. *Nature* **427**, 36–44 (2004).
295. Pleiner, T. *et al.* Structural basis for membrane insertion by the human ER membrane protein complex. *Science.* **369**, 433–436 (2020).
296. Liaci, A. M. *et al.* Structure of the human signal peptidase complex reveals the determinants for signal peptide cleavage. *Mol. Cell* **81**, 1–15 (2021).
297. Lorent, J. H. *et al.* Plasma membranes are asymmetric in lipid unsaturation, packing and protein shape. *Nat. Chem. Biol.* **16**, 644–652 (2020).
298. Parker, J. L. & Newstead, S. Structural basis of nucleotide sugar transport across the Golgi membrane. *Nature* **551**, 521–524 (2017).
299. Nie, L. *et al.* The structural basis of fatty acid elongation by the ELOVL elongases. *Nat. Struct. Mol. Biol.* **28**, 512–520 (2021).
300. Dahlberg, P. D. & Moerner, W. E. Cryogenic Super-Resolution Fluorescence and Electron Microscopy Correlated at the Nanoscale. *Annu. Rev. Phys. Chem.* **72**, 253–278 (2020).
301. Kaptan, S. & Vattulainen, I. Machine learning in the analysis of biomolecular simulations. *Adv. Phys. X* **7**, 2006080 (2022).
302. Ha, S. & Jeong, H. Unraveling hidden interactions in complex systems with deep learning. *Sci. Rep.* **11**, 1–13 (2021).
303. Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods* **18**, 176–185 (2021).
304. Earnest, T. M. *et al.* Challenges of Integrating Stochastic Dynamics and Cryo-Electron Tomograms in Whole-Cell Simulations. *J. Phys. Chem. B* **121**, 3871–3881 (2017).
305. Saminathan, A. *et al.* A DNA-based voltmeter for organelles. *Nat. Nanotechnol.* **16**, 96–103 (2021).
306. Huang, J., Gates, A. J., Sinatra, R. & Barabási, A. L. Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 4609–4616 (2020).
307. Forrester, N. Diversity in science: next steps for research group leaders. *Nature* **585**, S65–S67 (2020).
308. Cech, E. A. & Waidzunas, T. J. Systemic inequalities for LGBTQ professionals in STEM. *Sci. Adv.* **7**, eabe0933 (2021).
309. Peterson, R. J. We need to address ableism in science. *Mol. Biol. Cell* **32**, 507–510 (2021).
310. Popkin, G. A Fungal Safari: ‘Myconauts’ aim to reveal the often hidden soil fungi that shape ecosystems. *Science.* **377**, 143–147 (2022).

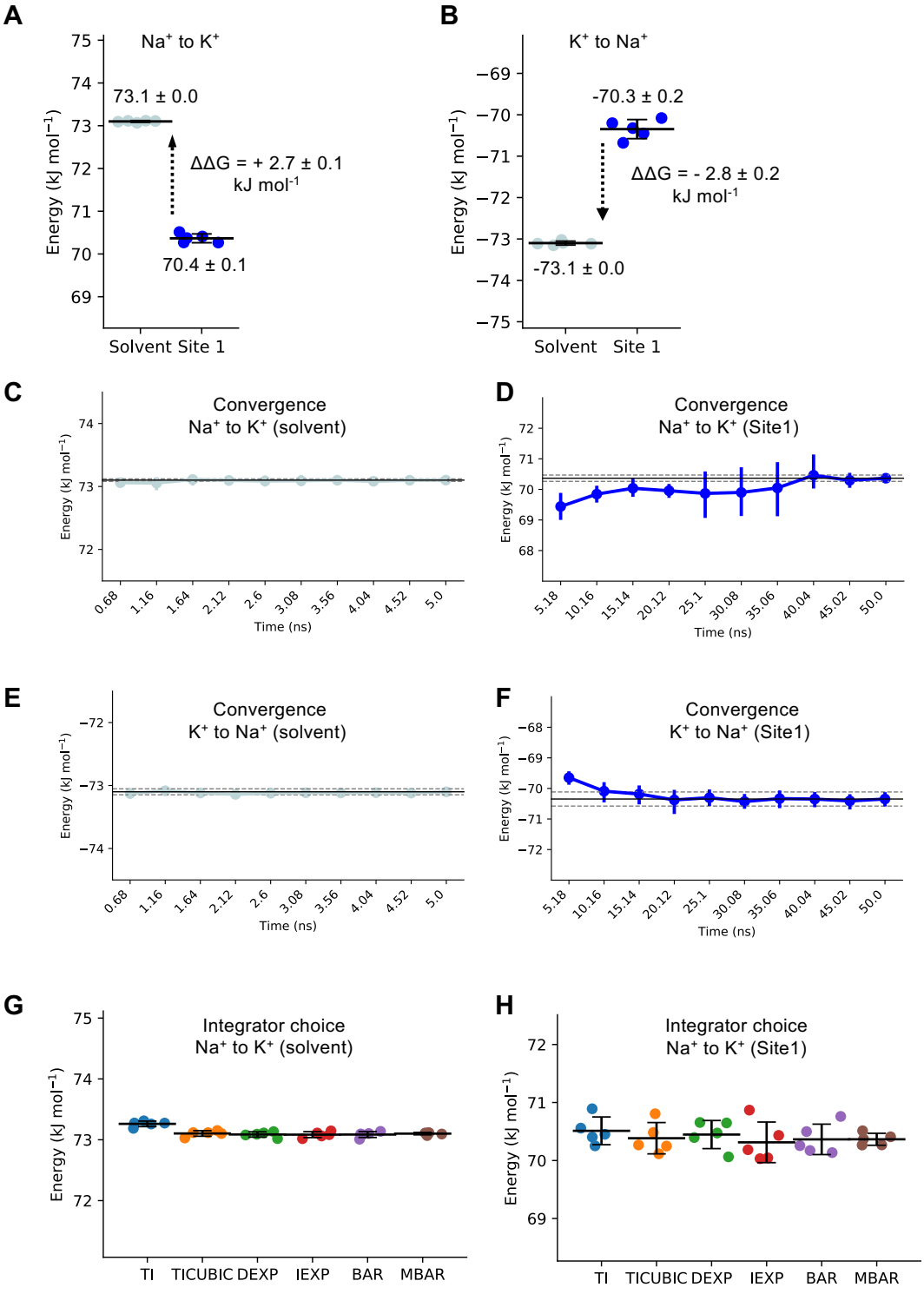
8. Appendices

Appendices for Chapter 3:



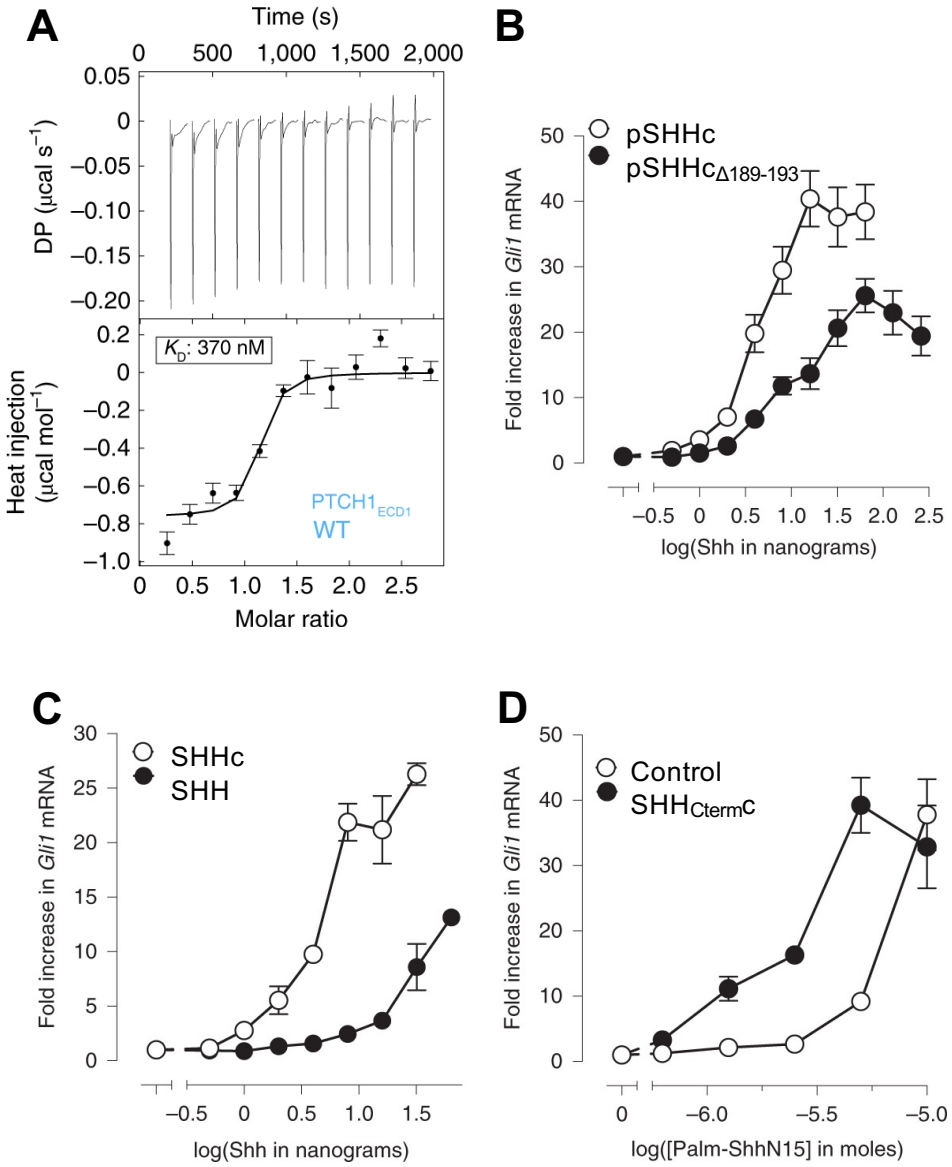
Appendix Figure S3.1: Convergence of PMF calculations.

Free energy from PMF calculations as a function of the fraction of simulation analysed. The first 200 ns of each window was discarded as equilibration time and Bayesian bootstrapping was applied 200 times to estimate the error.



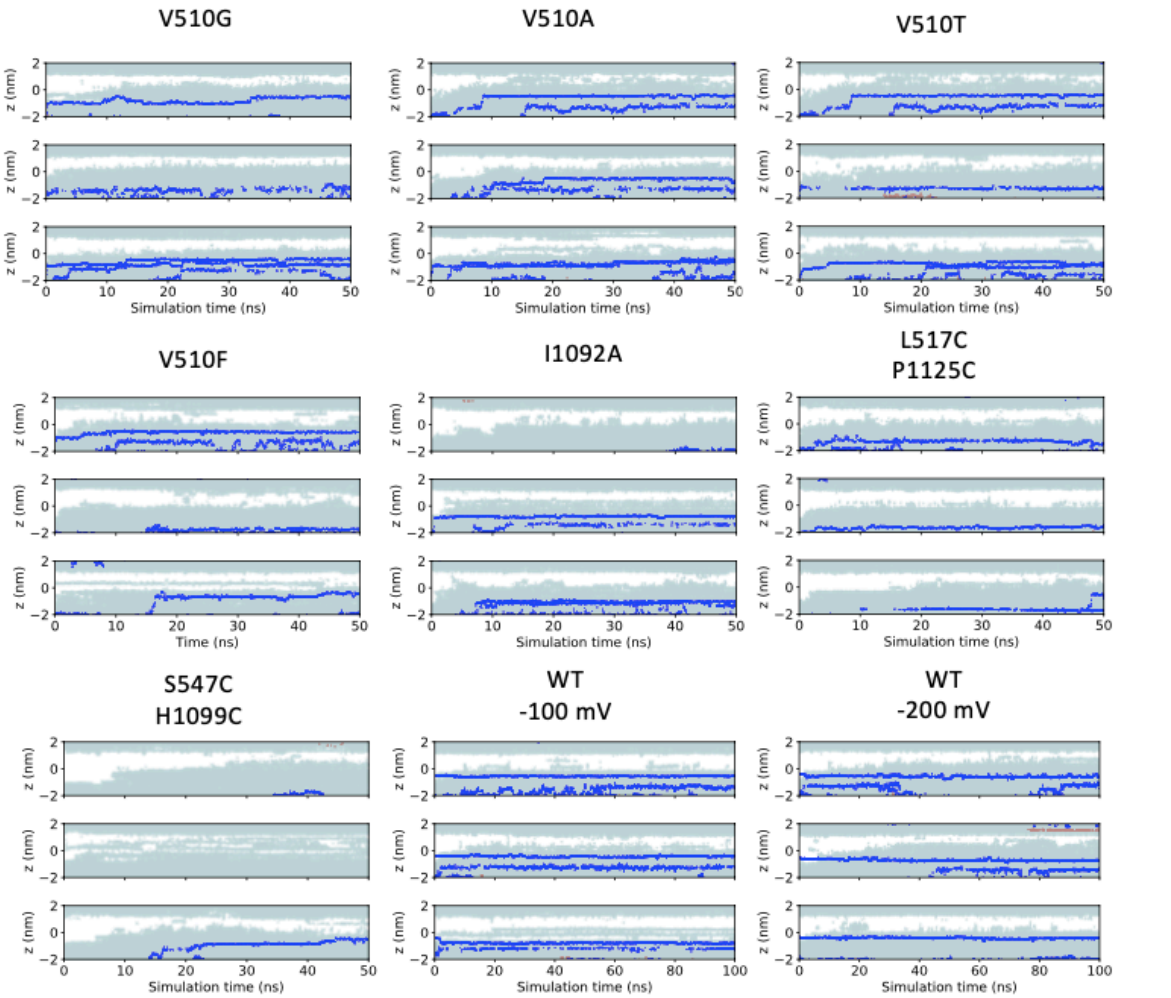
Appendix Figure S3.2: Reverse perturbation and FEP convergence.

FEP values for **A)** Na^+ to K^+ and **B)** K^+ to Na^+ in solvent and when bound to Site 1. **C-F)** Convergence of FEP values as a function of λ window length. **G-H)** FEP values for the Na^+ to K^+ perturbation predicted using different integrators.



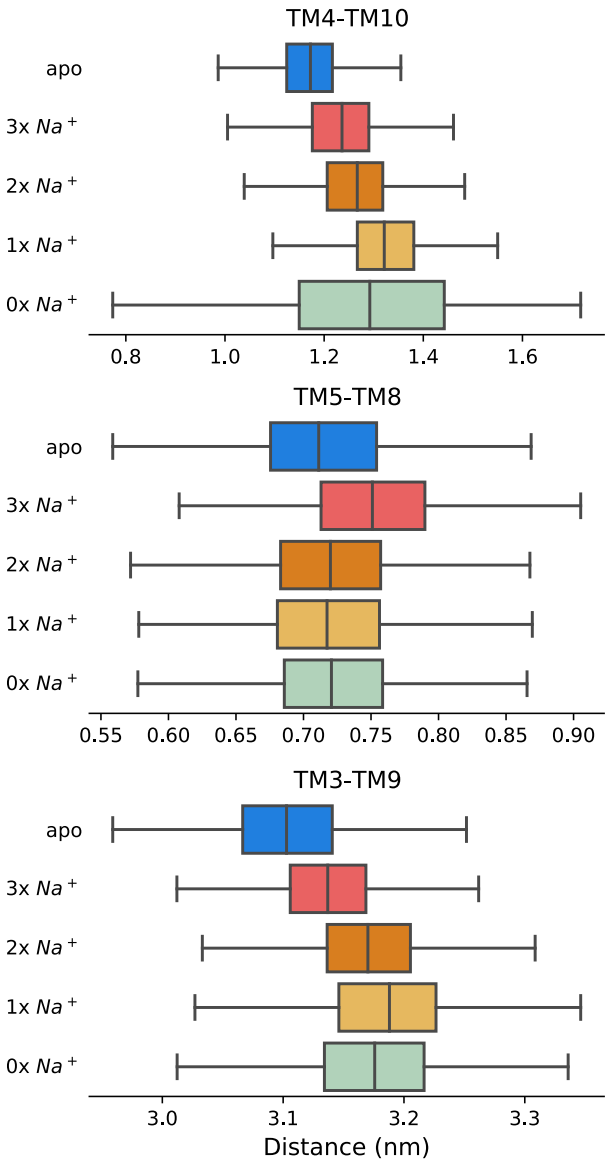
Appendix Figure S3.3: Biochemical support for PTCH1 inhibition by SHH-cholesterol.

A) Isothermal titration calorimetry (ITC) curve for addition of PEGylated-cholesterol to a cyclised PTCH1 ECD1-ECD2 construct, obtained by Dr Amalie Rudolf. B**-D)** Pathway activation (indicated by fold increase in Gli1 mRNA) in mouse cells treated with **B)** dually lipidated SHH (pSHHc) or a shortened C-terminal truncation ($pSHHc_{\Delta 189\text{-}193}$) **C)** cholesterylated SHH (SHHc) and unlipidated SHH and **D)** increasing concentrations of palmitoylated SHH N-terminus in the presence/absence of a cholesterylated-Cterminal SHH peptide. Signalling assays were conducted by members of the Rohatgi lab.



Appendix Figure S3.4: Water and ions within the PTCH1 TMD.

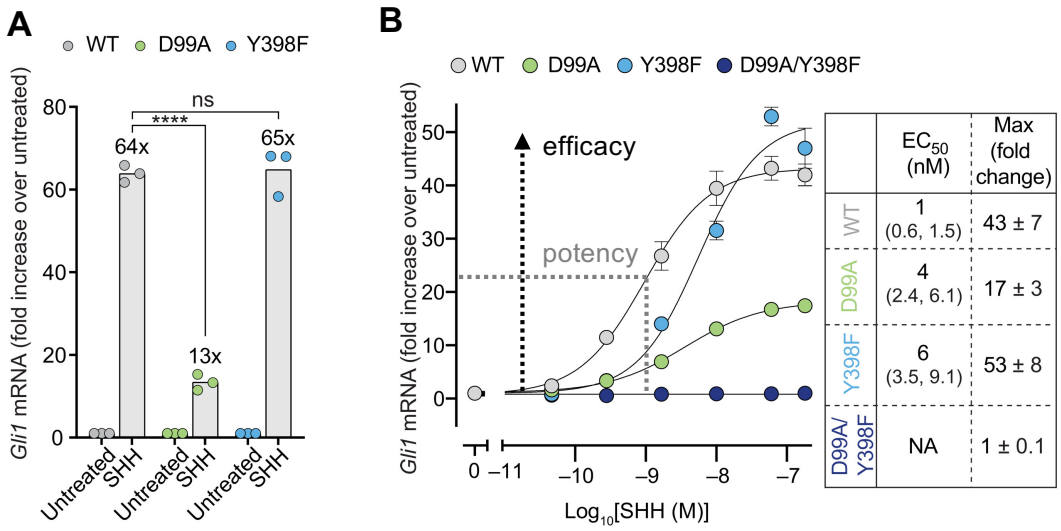
Analysis of water and ions within a cylinder localised to the PTCH1 TMD (see Fig. 3.10) across 3 x 50 ns simulations of WT PTCH1 or *in silico* PTCH1 mutants, or 3 x 100 ns simulations in the presence of -100 mV and -200 mV membrane voltages with Na^+ initially bound at Site 1.



Appendix Figure S3.5: DISP1 extracellular TMD distances.

Minimum distances between the extracellular portions of transmembrane helices across 3 x 100 ns simulations of DISP1 in distinct ion coupled states. Distances were defined between the Cα atoms of A561-T1039 (TM4-TM10), V623-A985 (TM5-TM8) and Y542-V1022 (TM3-TM9).

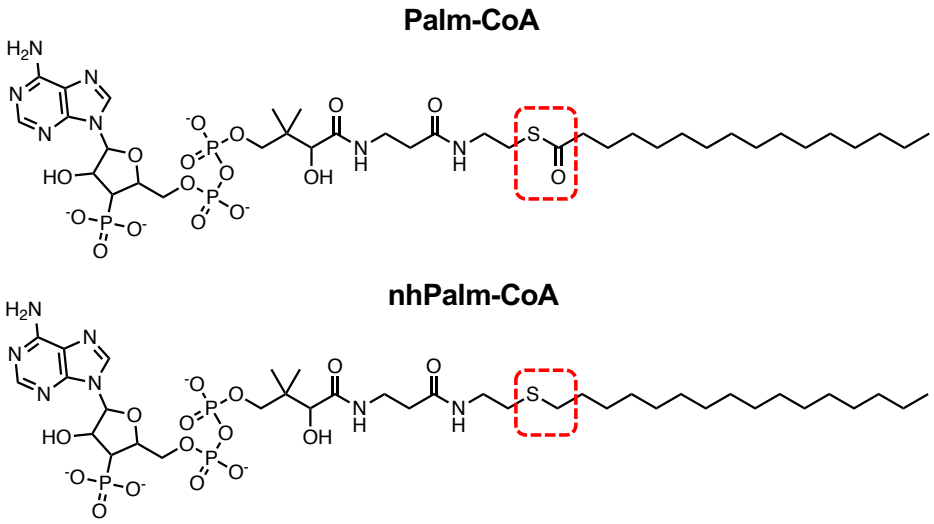
Appendices for Chapter 4:



Appendix Figure S4.1: Biochemical investigation of cholesterol binding to SMO.

A) Fold change in Gli1 mRNA (measured by RT-qPCR) with/without addition of SHH to cells expressing WT mSMO or mSMO with point mutations to the CRD (D99A) and TMD (Y398F) cholesterol binding sites (Fig. 4.1A). **B)** As in A upon addition of increasing concentrations of SHH (dose response curve).
Biochemical assays were performed by members of the Rohatgi lab. Figure adapted from Rudolf *et al* [220].

Appendices for Chapter 5:



Appendix Figure S5.1: Palm-CoA structure.

Chemical structures of palmitoyl-coenzyme A (Palm-CoA) and non-hydrolysable palmitoyl-coenzyme A (nhPalm-CoA).