Understanding the Structure and Interactions of Polycystin-2 Through Structural and Simulation Studies



Qinrui Wang

Trinity College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

September 2019

This thesis is dedicated to
my beloved family, particularly my parents, 邵军 and 王子斗, who always
encourage and support me to go on every adventure,
especially this one.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Mark Sansom for his exceptional support and guidance throughout my years in Oxford, not only in research, but also in my other endeavours in life. His patience, motivation, and immense knowledge have taught me more than what I expected from a DPhil degree. Thanks also go to my co-supervisor Prof. Liz Carpenter, who gave me the great opportunity to be part of the fast growing field of cryo-EM. Many thanks also to my industrial supervisor Dr Jiye Shi for supporting this project. I was provided with a unique opportunity by all my supervisors to develop multidisciplinary skills, which allow me to grow into a better scientist.

Further thanks go to Dr George Hedger, Dr Robin A. Corey, Dr Prafulla Aryal and all SBCBers past and present for helping me with various parts of the simulation work, for the very insightful and enlightening discussions (science related or not), and for the great company for my time in the lab. I also want to thank Dr Ashley C.W. Pike for sharing his vast knowledge about solving structures with me and for his patience in teaching me how to do cryo-EM data processing. I must also acknowledge Dr Mariana Grieben and Dr Chady Nasrallah for guiding me through the processes of experimental work, Agnese Baronina for making and screening my grids, and the SGC biotech team for helping me with the cell work.

I would also like to thank SABS CDT and Trinity College for their support. I greatly appreciate all my friends at Oxford. They are the ones who made my DPhil life more interesting, and made Oxford feel like home, especially when I had not got the chance to be home all these years. Finally, I have to say that I am extremely fortunate to have an incredible family. Without their love, support and encouragement, I would not be where I am today. 我永远爱你们!

Abstract

Polycystin-2 (PC2) is a member of the transient receptor potential (TRP) superfamily of non-selective cation channels. It has a systemic distribution with relatively high expression in kidney tubules. PC2 shares a common transmembrane fold with other TRP channels, in addition to having a extracellular/luminal domain (TOP domain) unique to TRPP and TRPML channels. Mutations in PC2 are associated with autosomal dominant polycystic kidney disease (ADPKD), which is one of the most prevalent genetic disorders in human. Despite that the genetic link between PC2 and ADPKD is well established, the molecular basis of the disease is still elusive. It has been proposed that dysfunction of ciliary PC2 is the main mechanism of pathology of PC2-associated ADPKD. Given the complex organisation of ciliary membrane, it is important to establish how PC2 interacts with specific lipids in its membrane environment. Lipid binding assay results suggest that detergent-solubilised PC2 can bind a range of phosphatidylinositol phosphates but not other simple anionic (phosphatidylserine, phosphatidic acid) or zwitterionic (phosphatidylethanolamine, phosphatidylcholine) lipids. Combining cryo-electron microscopy (cyro-EM) and multiscale molecular dynamics (MD) simulations, we identified a hydrophobic pocket in PC2 formed between S3, S4 transmembrane helices and S4-S5 linker, which showed a preference for phosphatidylinositol bisphosphate (PIP_2) binding. Simulations of other members of the TRP channel family suggest this lipid-binding site may be shared amongst a number of TRP channels. In addition, our cryo-EM maps revealed a binding site for cholesterol on PC2, which is further characterised with MD simulations. These results help to position PC2 within an emerging model of the complex roles of lipids in the regulation and organisation of ciliary membranes. Moreover, we used MD simulations to probe the ion selectivity of PC2. Our results suggest that PC2 is selective for Na^+ over Ca^{2+}, and not only the selectivity filter residues but also the TOP domain contribute to the selectivity of the channel.

Contents

List of Figures vii

List of Tables x

List of Abbreviations xi

1 Introduction 1
- **1.1 Proteins of TRP Super Family** . . . 2
  - **1.1.1 The TRPC Subfamily** . . . 7
  - **1.1.2 The TRPV Subfamily** . . . 10
  - **1.1.3 The TRPM Subfamily** . . . 11
  - **1.1.4 The TRPA Subfamily** . . . 12
  - **1.1.5 The TRPML Subfamily** . . . 13
  - **1.1.6 The TRPP Subfamily** . . . 14
- **1.2 Proteins of the Polycystin Family** . . . 17
- **1.3 Autosomal Dominant Polycystic Kidney Disease** . . . 19
- **1.4 Protein Lipid Interactions** . . . 22
- **1.5 Overall Aims and Motivation** . . . 26

2 Theory and Methods 27
- **2.1 Cryo-electron Microscopy** . . . 28
  - **2.1.1 Protein Purification** . . . 31
  - **2.1.2 Specimen Preparation** . . . 33
  - **2.1.3 Image Acquisition** . . . 35
  - **2.1.4 Image Processing** . . . 36
  - **2.1.5 Model Building and Validation** . . . 38
- **2.2 Molecular Dynamics Simulations** . . . 39
  - **2.2.1 Force Fields** . . . 40
    - **2.2.1.1 Intramolecular Interactions** . . . 41
    - **2.2.1.2 Non-bonded Interactions** . . . 42
    - **2.2.1.3 Force Field Parameterisation** . . . 44
  - **2.2.2 Simulation Methods** . . . 45

2.2.2.1 Periodic Boundary Conditions 45
2.2.2.2 Integration algorithms 46
2.2.2.3 Thermodynamic Ensembles 47
2.2.3 Coarse-grained Simulations 50
2.2.4 Free Energy Calculations 53
2.2.4.1 Umbrella Sampling 54
2.2.4.2 Potential of Mean Force 55

3 Experimental Studies of Lipid Interactions of Polycystin-2 56
3.1 Introduction 57
3.2 Methods 58
3.2.1 Protein Expression and Purification 58
3.2.2 PIP Strip Assay 60
3.2.3 Label-free Differential Scanning Fluorimetry (nanoDSF) Measurements 60
3.2.4 Cryo-EM Grid Preparation and Data Acquisition . . 62
3.2.5 Image Processing 63
3.2.6 Model Building 63
3.3 Results and Discussion 64
3.3.1 Characterisation of PC2 Lipid Binding Using nanoDSF and Biochemical Assays 64
3.3.2 Structural Characterisation of PC2 in the Presence of PIP Lipids by Cryo-EM 68
3.3.2.1 Data Processing and Model Building for Sample with PI(4,5)P_2 68
3.3.2.2 Unidentified Density in C1 Map of PC2/PI(4,5)P_2 75
3.3.2.3 Data Processing and Model Building for Sample with PI(3,5)P_2 77
3.3.2.4 A Possible Calcium Ion in the Selectivity Filter 80
3.3.2.5 Lipid Interactions Revealed by Cryo-EM . . 82
3.4 Conclusions 85

4 Simulation Studies of Lipid Interactions of Polycystin-2 and Other Channels 88
4.1 Introduction 89
4.2 Methods 90
4.2.1 Cryo-EM Structure-based Molecular Models 90

4.2.2 Coarse-grained Simulation Details 94
4.2.3 Atomistic Simulation Details 96
4.2.4 PMF Calculations 98
4.2.5 Simulation Visualisation and Analysis 99
4.3 Results and Discussion 99
4.3.1 Identification of a Phospholipid Interaction Site . . . 99
4.3.2 Free Energy Landscapes for PC2/Lipid Interactions 104
4.3.3 The PIP_2 Binding Site 113
4.3.4 PIP_2 Interactions in TRP Channels Revealed by CG Simulations 115
4.3.5 Atomistic Study of Phosphoinositide Binding Specificity 117
4.3.6 A Cholesterol Binding Site Revealed by Cryo-EM and Simulations 121
4.4 Conclusions 125

5 Simulation Studies of Ion Interactions of Polycystin-2 127
5.1 Introduction 128
5.2 Methods 130
5.2.1 System setup 130
5.2.2 Coarse-grained Simulation Details 130
5.2.3 Atomistic Simulation Details 132
5.2.4 Simulation Visualisation and Analysis 134
5.3 Results and Discussion 135
5.3.1 Dynamics of the Ion Identified in Cryo-EM Maps . . 135
5.3.2 Ion Permeation and Selectivity of PC2 at the Selectivity Filter and TOP Domain 144
5.3.3 Ion and Water Permeation at Central Cavity and Lower Gate of PC2 152
5.4 Conclusions 155

6 Conclusions and Future Directions 157

References 162

Appendix A 179

Appendix B 184

Appendix C 186

List of Figures

1.1 Number of deposited protein structures 3
1.2 Family tree of mammalian TRP channels 4
1.3 Structure of Kv and TRP channels 5
1.4 TRP channel structures 6
1.5 Schematic representation and cyro-EM structure of PC2 . 17
1.6 Cyro-EM structure of PC1/PC2 complex 19
1.7 Disease progression of ADPKD 20
1.8 Structures showing protein lipid interactions 24

2.1 Schematics of electron scattering 31
2.2 Overview of the steps involved in structural determination by cryo-EM 32
2.3 Schematics of single-particle cryo-EM specimen 35
2.4 Schematic representation of the contributions to a force field 41
2.5 The Lennard-Jones 6-12 potential 43
2.6 Periodic Boundary Conditions 46
2.7 Martini mapping between the chemical structure and the CG particles 52
2.8 ElNeDyn network of PC2 53

3.1 Chemical structures of different PIP_2 used in experiments . 61
3.2 Biological repeats of PC2 binding to PIP strips 66
3.3 Thermal stability of PC2 in the presence of PIP_2 67
3.4 Overlay of PC2 models with claimed detector pixel sizes . . 69
3.5 Cryo-EM data processing for PC2 with $PI(4,5)P_2$ 70
3.6 FSC curves and local resolution for PC2 with $PI(4,5)P_2$ maps 73
3.7 Molecular model of PC2 with $PI(4,5)P_2$ 74
3.8 Discrete density revealed in C1 map 75
3.9 Pixel size estimation for the dataset of PC2 with $PI(3,5)P_2$ 78
3.10 FSC curves and local resolution for PC2 with $PI(3,5)P_2$ maps 79
3.11 Cryo-EM density of putative cation ion in PC2/$PI(4,5)P_2$ and PC2/$PI(3,5)P_2$ datasets 81

3.12 Cryo-EM densities of PC2/PI(4,5)P_2 and PC2/PI(3,5)P_2 maps 82
3.13 Lipid/detergent densities in PC2/PI(4,5)P_2, PC2/PI(3,5)P_2 and PDB ID: 5MKE maps 83
3.14 Lipid density comparison between PC2, TRPV1 and TRPV5 86

4.1 Schematic for CG self-assembly simulation 95
4.2 Possible lipid interaction sites in PC2 identified by simulations 100
4.3 Lateral movement of bound and free lipids 102
4.4 Coarse-grained simulations of PC2 in an *in vivo* mimetic mixed lipid bilayer 103
4.5 Sequence alignment for PC2 from 18 different species . . . 105
4.6 Coarse-grained representations of proteins and lipids used in PMF calculations 108
4.7 Free energy profiles for lipid interactions at the binding site on PC2 and Kir3.2 109
4.8 Convergence of PMF calculations of PIP_2 binding to PC2 . 110
4.9 POPC distribution in the cytoplasmic leaflet of the bilayer around PC2 and Kir3.2 in a PIP_2 PMF simulation 112
4.10 Atomistic details of PIP_2 coordination at the binding site in PC2 114
4.11 Comparison of lipids bound to PC2, TRPV1 and TRPV5 . 116
4.12 Hydrogen-bonding patterns of Kir2.2 with different PIPs . 118
4.13 Binding poses of different PIPs within the binding sites on Kir2.2 120
4.14 Cholesterol interactions with PC2 122
4.15 Atomistic simulations of cholesterol-bound PC2 124

5.1 Sliced view of PC2 129
5.2 Schematic of electric field applied across a bilayer 134
5.3 Cryo-EM density of putative calcium ion in PC2/PI(4,5)P_2 and PC2/PI(3,5)P_2 datasets 136
5.4 Filter Na^+ ion positions in the PC2/PI(4,5)P_2 system . . . 137
5.5 Filter Na^+ ion positions in the PC2/PI(3,5)P_2 system . . . 139
5.6 Filter Ca^{2+} ion positions in the PC2/PI(4,5)P_2 system . . . 141
5.7 Filter Ca^{2+} ion positions in the PC2/PI(3,5)P_2 system . . . 143
5.8 Na^+ ion distribution at TOP domain 145
5.9 Cation within the TOP cavity 147
5.10 Cation within the TOP cavity with imposed electric field . 149

5.11 Cation within the TOP cavity with ion concentration of 350 mM 151
5.12 Water permeation of PC2 153
5.13 Movement of cavity Na^+ ion 154

6.1 Comparison of lipid binding sites of TRP and Kv channels 159

B.1 Cryo-EM data processing for PC2 with $PI(3,5)P_2$ 185

C.1 Probability density of phospholipid in atomistic simulations of different PC2 cryo-EM structures 187
C.2 POPC interaction with PC2 in the atomistic simulations. . 188
C.3 PIP_2 binding in two repeats of CG simulations of PC2 . . . 189
C.4 PIP_2 binding in three repeats of CG simulations of a gain-of-function Phe604Pro mutant of PC2 190
C.5 Cholesterol binding in two repeats of CG simulations of PC2 191
C.6 Two repeats of atomistic simulations of cholesterol-bound PC2 192

List of Tables

1.1 TRP Channels Associated Channelopathies 7
1.2 Nomenclature of Polycystin Proteins 14

2.1 Properties of Electron 29

3.1 Summary of Tested Protein-lipid Ratios 62

4.1 Summary of Equilibrium Simulations 92
4.2 Summary of Non-equilibrium Simulations 93

5.1 Summary of Simulations Performed in Chapter 5 131
5.2 Duration and Force Constant in Simulation Steps 133

A.1 List of Available TRP Channel Structures 180

List of Abbreviations

2D . . . . . Two-dimensional

3D . . . . . Three-dimensional

CG . . . . . Coarse grained

CHOL . . . . . Colesterol

COG . . . . . Centre of geometry

COM . . . . . Centre of mass

Cryo-EM . . . Cryo-electron microscopy

CTF . . . . . Contrast Transfer Function

DDD . . . . . Direct electron detection device

DDM . . . . . N-Dodecyl-β-D-maltoside

FSC . . . . . Fourier shell correlation

GM3 . . . . . Monosialodihexosylganglioside

LJ . . . . . Lennard-Jones

MD . . . . . Molecular dynamics

NMR . . . . . Nuclear magnetic resonance

PC1 . . . . . Polycystin-1

PC2 . . . . . Polycystin-2

PDB . . . . . Protein Data Bank

PI . . . . . Phosphatidylinositol

PI(3,5)P_2 . . . Phosphatidylinositol-3,5-bisphosphate

PI(4,5)P_2 . . . Phosphatidylinositol-4,5-bisphosphate

PIP . . . . . Phosphatidylinositide

PIP_1 . . . . . Phosphatidylinositol monophosphate

PIP_2 . . . . . Phosphatidylinositol bisphosphate

PIP_3 . . . . . Phosphatidylinositol-3,4,5-trisphosphate

PME . . . . . Particle-mesh Ewald

PMF . . . . . Potential of mean force

POPC 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine

POPE 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine

POPS 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylserine

QM . . . . . Quantum mechanical

SEC . . . . . Size-exclusion chromatography

SM . . . . . Sphingomyelin

TEV . . . . . Tobacco etch virus

TRP . . . . . Transient receptor potential

UDM . . . . . N-Undecyl-β-D-Maltopyranoside

VSLD . . . . . Voltage sensor-like domain

WHAM . . . . Weighted histogram analysis method

1

Introduction

Contents

1.1 Proteins of TRP Super Family 2
1.1.1 The TRPC Subfamily 7
1.1.2 The TRPV Subfamily 10
1.1.3 The TRPM Subfamily 11
1.1.4 The TRPA Subfamily 12
1.1.5 The TRPML Subfamily 13
1.1.6 The TRPP Subfamily 14
1.2 Proteins of the Polycystin Family 17
1.3 Autosomal Dominant Polycystic Kidney Disease 19
1.4 Protein Lipid Interactions 22
1.5 Overall Aims and Motivation 26

Cells as the fundamental structural and functional units of living organisms are separated from their surrounding environment by membranes. In eukaryotes, to increase the efficiency of cellular function, organisms have developed different subcellular compartments, which contain specific concentration of the required components for their given functions. The establishment of function specific microenvironment is also achieved by membranes. Biological membranes have a complex composition with different lipids and various sterols distributed across a two-layer structure. Lipidomics analysis has identified over 40,000 biologically relevant lipid species (`http://www.lipidmaps.org/resources/databases/`). The lipid compositions differ in different intracellular organelles. The major lipid species in membranes of eukaryotic cells are glycerophospholipids including phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylglycerol (PG), phosphatidylserine (PS) and phosphatidylinositol (PI). There are also some functionally important minor species such as the phosphoinositides including phosphatidylinositol-(3,5)-bisphosphate ($PI(3,5)P_2$) and phosphatidylinositol-(4,5)-bisphosphate ($PI(4,5)P_2$), which are widely involved in cell signalling and recognition [1]. Many membrane spanning integral proteins including ion channels and transporters act as passageways to transport different substances across membranes in order to maintain the tightly regulated contents within each enclosed region. Ion channels form pores within the lipid bilayers to allow ions to pass through in a controlled manner. The total number of protein structures deposited in the Protein Data Bank (PDB) increases exponentially (Figure 1.1). Amongst these, about 4078 are transmembrane proteins corresponding to approximately 1344 unique proteins from $\sim$ 572 protein families to date (`http://memprotmd.bioch.ox.ac.uk/stats`).

1.1 Proteins of TRP Super Family

The transient receptor potential (TRP) superfamily of non-selective cation channels is a major class of membrane proteins found in all eukaryotes, and is

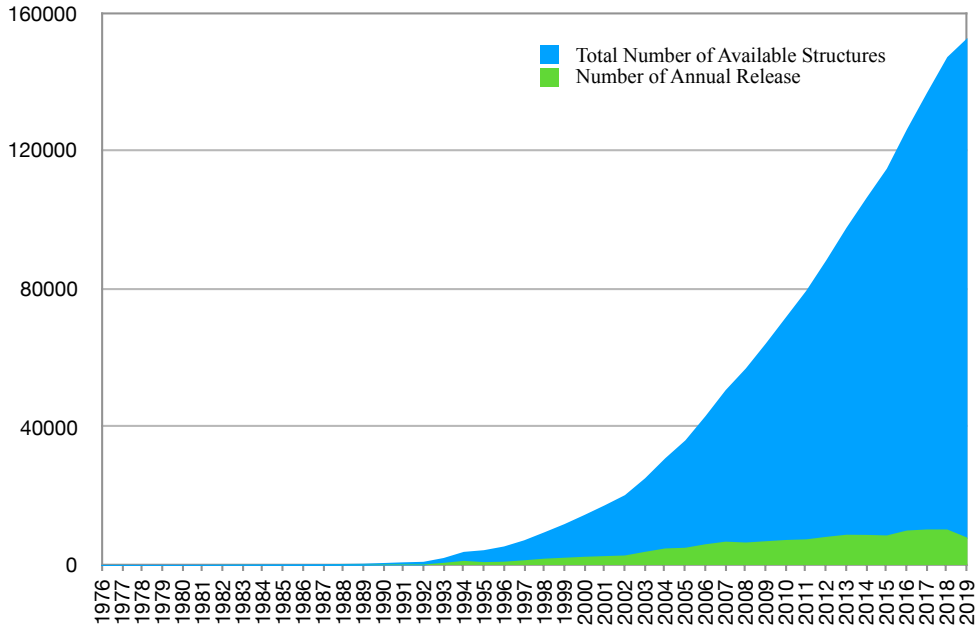


Figure 1.1: Total number of protein structures deposited in the Protein Data Bank (PDB) is shown in blue shading. The number of protein structures released annually is shown in green shading. The data was retrieved from PDB in September 2019.

distantly related to the voltage-gated cation (K^+, Na^+ and Ca^{2+}) channels. Based on sequence homology, mammalian TRP channels may be divided into six subfamilies, the TRPC (Classical or Canonical), TRPV (Vanilloid), TRPM (Melastatin), TRPP (Polycystin), TRPML (Mucolipin) and TRPA (Ankyrin) [2] (Figure 1.2). Genes encoding proteins in the TRPN (no mechanoreceptor potential C; NOMPC) subfamily have only been discovered in *Drosophila melanogaster, Caenorhabditis elegans, Xenopus laevis* and *Danio rerio* so far [3]. Structurally, the TRP channels share strong similarities with the voltage-activated potassium (Kv) channels [4]. Both of these channel families have a tetrameric architecture assembled from identical or similar subunits. Each of the four subunits is composed of six transmembrane helices (S1-S6) with a pore loop region between S5 and S6 (Figure 1.3) [5–16]. Helices S1 to S4 in TRP channels form a voltage sensor-like domain (VSLD), which is packed against the pore domain (S5-Pore-S6) of the adjacent chain. For members

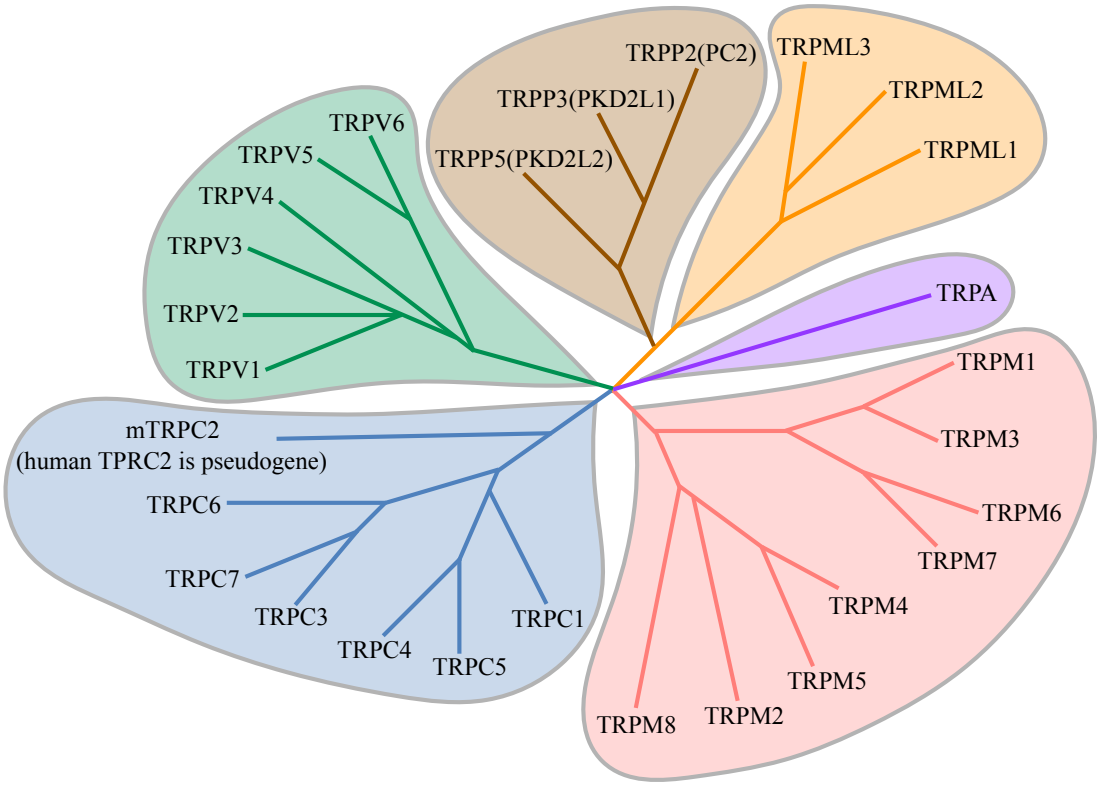


Figure 1.2: Family tree of mammalian TRP channels. Adapted from [26].

of the TRPC, TRPM and TRPN subfamilies, there is a TRP domain consisting of 23 to 25 highly conserved residues on the C-terminal end of the transmembrane domains. The TRPP and TRPML channels are relatively more distant to the other five subfamilies. There is a large insertion between S1 and S2 transmembrane helices forming an extracellular/luminal domain above the transmembrane domain in members of these two subfamilies. A comprehensive list of currently available TRP channel structures is listed in Appendix A, Table A.1. Representative structures from each subfamily are shown in Figure 1.4.

Apart from TRPM4 and TRPM5, which are selective for monovalent cations, all other TRP channels that have been functionally characterised show permeability to Ca^{2+} ions, although the relative selectivity between different cations varies

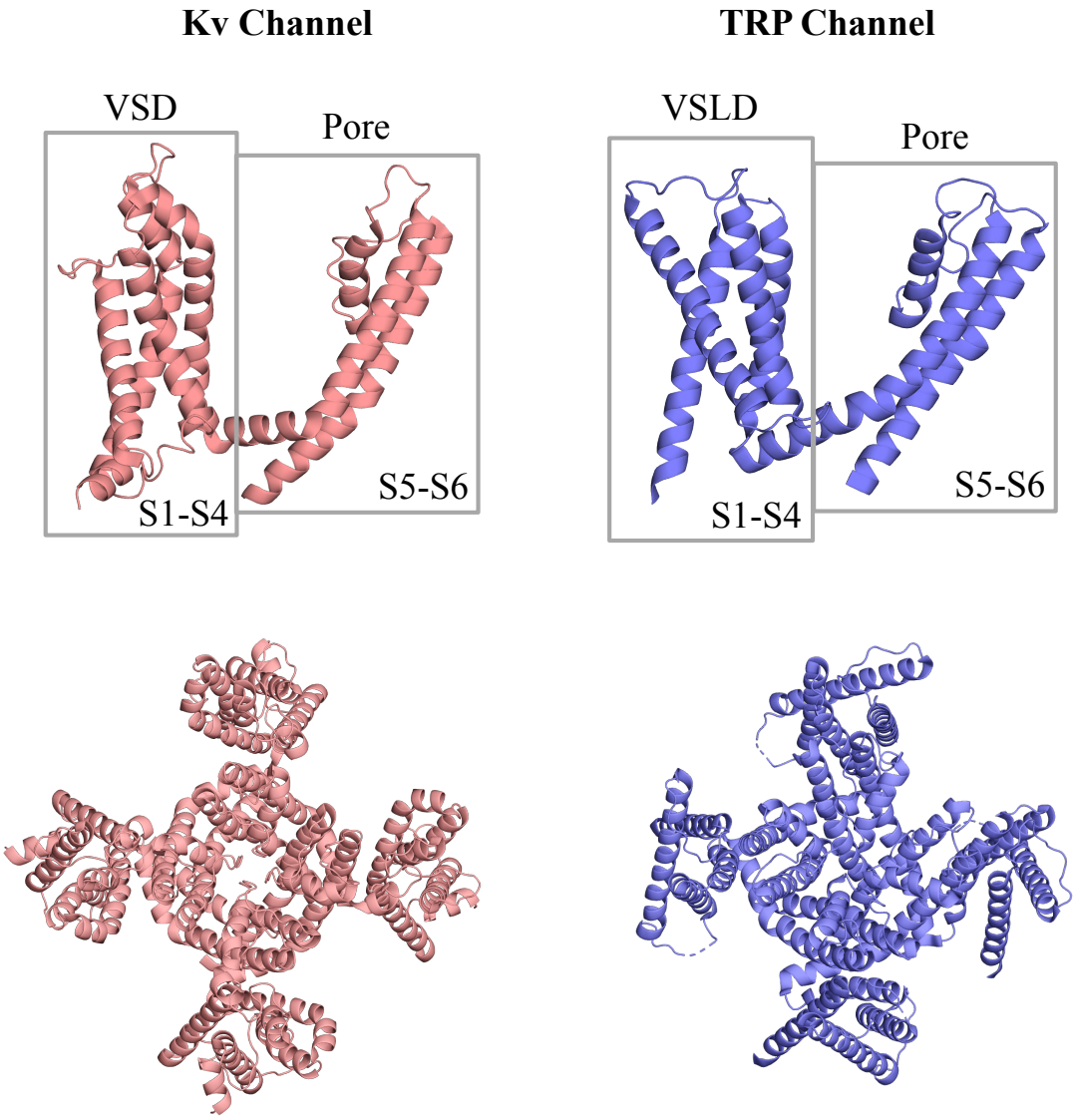


Figure 1.3: Transmembrane region of representative structures of Kv and TRP channels. Kv1.2/Kv2.1 chimera (PDB ID: 2R9R) is shown on the left (pink) and TRPV1 (PDB ID: 3J5P) is on the right (blue). Both channel families form tetrameric structure with each subunit having six transmembrane helices (S1-S4). The voltage sensor domain (VSD) or the voltage sensor-like domain (VSLD) (S1-S4) and the pore domain (S5-S6) are indicated for each channel.

[18]. TRP channels are involved in many aspects of cellular function, including thermosensation, osmotic pressure regulation, mechanosensation, and detection of noxious substances, with wide expression in different tissues including both excitable and nonexcitable cells [19]. TRP channels are activated and inhibited by a

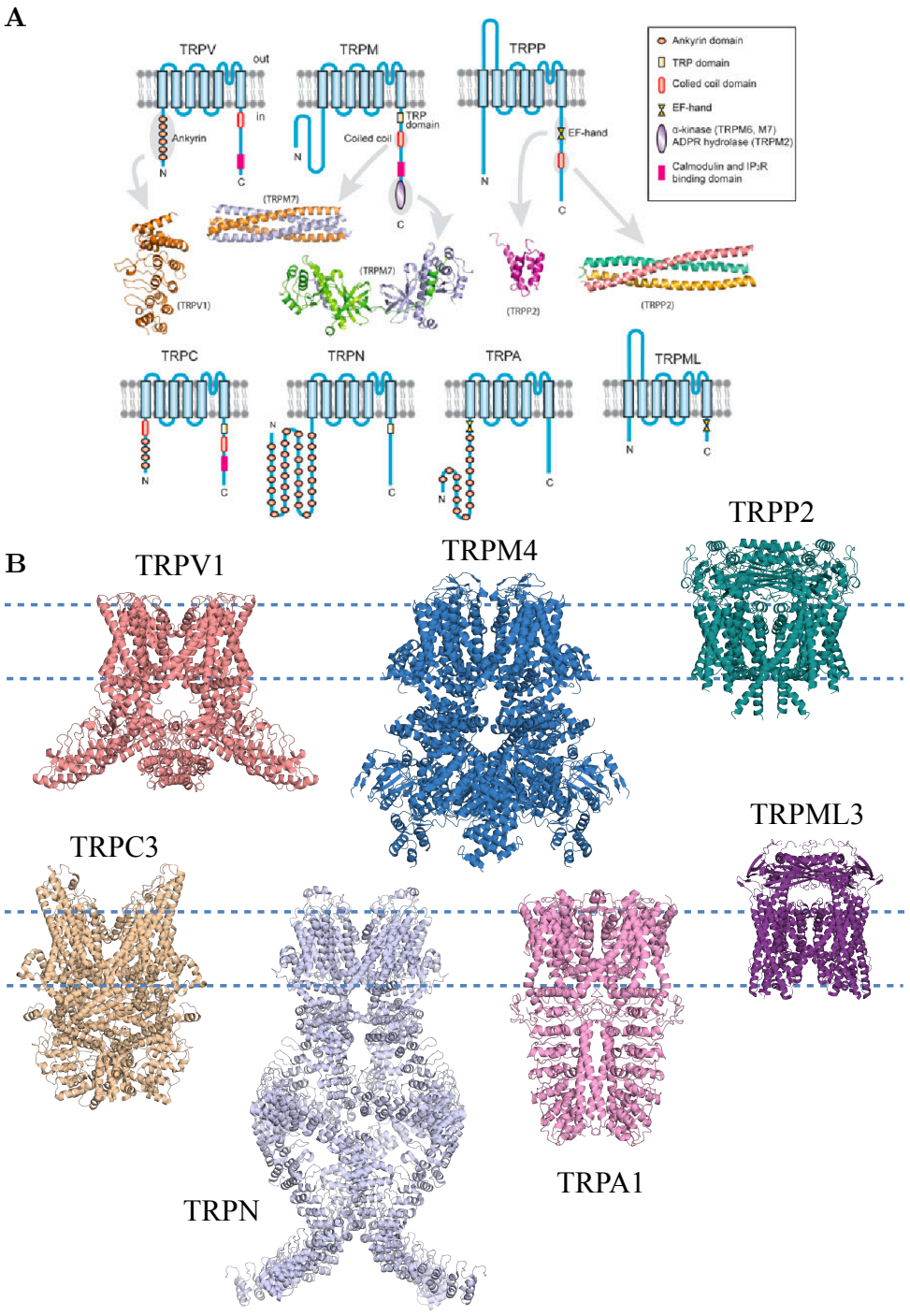


Figure 1.4: TRP channel structures. **A:** Topology of channel subunits for TRP subfamilies. Figure is taken from [17]. **B:** Representative structures from each subfamily of TRP channels: TRPV1 (PDB ID: 3J5P), TRPM4 (PDB ID: 6BCO), TRPP2 (this study), TRPC3 (PDB ID: 6CUD), TRPN (PDB ID: 5VKQ), TRPA1 (PDB ID: 3J9P), TRPML3 (PDB ID: 5W3S). Note that not all domains are resolved in the structures.

range of mechanisms, in response to thermal or chemical stimuli and/or mechanical forces [20], which suggests that their activation/modulation is heavily dependent on the specific cellular context such as lipid composition, phosphorylation and local ligand concentrations. Given the structural characteristics of TRP channels, it is possible for cooperative binding of distinct stimuli further complicating the modulation of the channels [3].

TRP channels have been implicated in a number of different diseases and consequently are of interest as potential drug targets. A summary of channelopathies that have been identified to be linked with TRP channels is shown in Table 1.1 [18, 21].

Table 1.1: TRP Channels Associated Channelopathies

TRP Channel	Channelopathies
TRPC6	Focal and segmental glomerulosclerosis
	Proteinuric renal disease
TRPV4	Spondylometaphyseal dysplasia Kozlowski type
	Scapuloperoneal spinal muscular atrophy
	Charcot–Marie–Tooth disease type 2C
TRPM1	Autosomal recessive congenital stationary night blindness
TRPM2 & TRPM7	Guamanian amyotrophic lateral sclerosis
	Parkinsonism dementia
TRPM4	Autosomal dominant progressive familial heart block type I
TRPM6	Hypomagnesaemia with secondary hypocalcaemia
TRPML1	Mucolipidosis type IV
TRPP2	Autosomal dominant polycystic kidney disease

1.1.1 The TRPC Subfamily

The founding member of the TRP superfamily is the *Drosophila* TRP, which is involved in phototransduction through rhodopsin activated eye-specific beta-type

phospholipase C (PLC). It was discovered in a *Drosophila* mutant that exhibited a transient receptor potential in the photoreceptors in response to light, instead of a plateau-like sustained potential [22], and thus was named *trp*. TRP channels that are the most closely related to *Drosophila* TRP with more than 30% sequence identity are therefore classified into the "Classical/Canonical" (TRPC) subfamily. All seven members (TRPC1−7) in this subfamily share a common TRP box (EWKFAR) in the C-terminal tail and three or four ankyrin repeats on the cytosolic N-terminal. They are divided into four subgroups: TRPC1, TRPC2, TRPC3/6/7 and TRPC4/5 according to sequence homology and functional similarities [23]. It has been shown that both heterotetramers and homotetramers are formed by TRPC channels with different members of the TRP superfamily. For example, TRPC1 has been reported to complex with TRPV1 and TRPC4 to form functional channels [24, 25].

It has been suggested that all mammalian TRPCs with the exception of human TRPC2 are receptor activated. That is, stimulation of G protein–coupled receptors (GPCRs) or receptor tyrosine kinases activate different isoforms of PLCs, which hydrolyse an important signalling lipid phosphatidylinositol 4,5-bisphosphate ($PI(4,5)P_2$) into diacylglycerol (DAG) and inositol-1,4,5-trisphosphate (IP_3). Either the degradation of PIP_2 or the production of DAG or IP_3 can then modulate the TRP channel activity. TRPC proteins have also been widely suggested to be modulated by the intracellular Ca^{2+} stores [26]. However, the dependency of store depletion is highly controversial. For example, human TRPC3 expressed in HEK293 cells has been shown to be activated by PLC-dependent mechanism independent of store depletion, but could be activated by depleted stores of Ca^{2+} when it was expressed in DT40 avian B-lymphocytes [27]. TRPC1 has been reported to be activated by DAG without the presence of Ca^{2+} [28]. Since the mechanistic link between the channels and the filling state of the intracellular stores are largely unclear, it remains questionable whether store dependence is a mechanism of activation or a phenomenon accompanying the activation of channels.

The first identified mammalian TRP channel is TRPC1. It is widely expressed. Despite the early discovery of this channel, the functional studies of TRPC1 remain contentious. Homomeric TRPC1 has been reported to be store-operated, receptor-activated and stretch-activated [28–30]. TRPC1 has also been shown to form functional receptor-regulated heteromeric complexes with TRPC3, TRPC4 and TRPC5 [28, 31]. Human *TRPC2* is considered to be a pseudogene as it encodes a truncated nonfunctional protein. Functional studies haven only been done on mouse and rat TRPC2 showing localisation on neuronal microvilli of vomeronasal organ in rat [32] and in the anterior sperm head in mouse [33].

TRPC4 and TRPC5 are more closely related to TRPC1 comparing to other members of the TRPC subfamily, and share 64% sequence identity. TRPC4 and TRPC5 homotetramers have been shown to be activated in a PLC activity dependent way. However, neither of the two primary products, IP_3 and DAG, is sufficient for the activation of TRPC4 or TRPC5 on their own [34, 35]. For both proteins, there is a PDZ binding domain (VTTRL) at the C-terminal, which is not present in other TRP channels [36]. The PDZ binding motif can interact with the Na^+/H^+ exchanger regulatory factor (NHERF), which is a scaffold protein connecting plasma membrane proteins with the actin cytoskeleton [37]. It is possible that TRPC4 and TRPC5 are part of multiprotein complexes similar to the *Drosophila* transducisome. Heterometric channels composed of TRPC1 and TRPC4 or TRPC1 and TRPC5 have been evidenced *in vivo* functioning as non-selective cation channel in mammalian brain [31].

TRPC3, TRPC6 and TRPC7 share 65% to 78% sequence identity with each other. They exhibit both inward and outward rectifying activity with relatively low selectivity for Ca^{2+} over Na^+. All of them can be directly activated by DGA or $G_{q/11}$-coupled receptors [38]. It has been shown that their activities are sensitive to intracellular Ca^{2+} concentration. TRPC6 and TRPC7 are inversely regulated by Ca^{2+} via the Ca^{2+}/calmodulin interaction [39].

1.1.2 The TRPV Subfamily

The TRPV subfamily of channels is named after the founding member TRPV1 (originally known as vanilloid receptor 1, VR1), which was identified by expression cloning using the inflammatory vanilloid compound capsaicin derived from chillies [40]. Five additional proteins (TRPV2−6) were later classified into this subfamily. There is approximately 25% of homology shared between members of this subfamily and proteins in the TRPC subfamily over the transmembrane helices S5 and S6. All TRPV channels have three to five ankyrin repeats in the N terminal.

TRPV1 is the most well charaterised member of this subfamily. It has been shown that TRPV1 is an outward rectifying cation channel with relative selectivity for Ca^{2+}. Polymodal activation mechanisms have been illustrated for this channel. It is potentiated by capsaicin, resiniferatoxin, photons and heat ($>$ 43 ℃) [41]. TRPV1-knockout mice have been shown to be defective in the transduction of the vanilloid derivatives induced nociceptive, inflammatory and hypothermic effects, demonstrating that TRPV1 plays a vital role in in thermal hyperalgesia [42, 43].

TRPV2 shares 50% sequence identity with TRPV1, but it appears to be insensitive to capsaicin. It also functions as a non-selective cation channel with higher permeability for Ca^{2+} comparing to Na^{+}. Similar to TRPV1, it can be activated by noxious heat with a higher threshold ($>$ 52 ℃) [44]. It has been reported that TRPV2 responds to membrane stretch in vascular smooth muscles, suggesting its role in blood flow regulation [45]. TRPV3 ($>$ 31 ℃) and TRPV4 ($>$ 25 ℃) are activated by non-noxious temperature. TRPV3 is predominately expressed in epithelial cells of skin, tongue and gastrointestinal tract and in the nervous system [46]. TRPV4 has a more systemic expression with presence in liver, kidney, heart and the nervous system [47]. Their roles in thermosensation for both TRPV3 and TRPV4 are controversial with experiments showing different thermo responses in expression systems, native tissues or knockout mice [48–50].

Homomeric and heteromeric channels of TRPV5 and TARV6 are highly expressed in epithelial cells of kidney and intestine [51]. Distinct from other TRP

channels, they are highly Ca^{2+} selective with strong inwardly rectifying currents [52]. It has been suggested that the negatively charged aspartate residues in the selectivity filter contribute to the Ca^{2+} selectivity similar to the voltage-gated Ca^{2+} channels [12, 53]. Their activities are inhibited by high intracellular concentration of Ca^{2+}, which have been proposed to play a role in prevention of Ca^{2+} poisoning of cells [51].

1.1.3 The TRPM Subfamily

There are eight members (TRPM1−8) in the mammalian TRPM (melastatin) subfamily. Unlike TRPC and TRPV channels, TRPM proteins are lack of ankyrin repeats at the N-terminal. Instead, they all have a homology region at the N-terminal unique to the TRPM subfamily, which is functionally important for protein assembly and trafficking [54]. TRPM2, TRPM6 and TRPM7 have exceptional C-terminal regions containing active enzymatic domains. TRPM2 displays adenosine diphosphoribose (ADP-ribose) pyrophosphatase activity [55], and TRPM6 and TRPM7 exhibit protein kinase functions [56, 57].

The first identified TRPM member, TRPM1, was cloned from melanoma and found to be an indicator of melanoma metastasis because of the inverse correlation between TRPM1 expression and melanoma progression [58]. However, the function of TRPM1 is still largely unknown. TRPM2 is a non-selective channel activated by ADP-ribose and nicotinamide adenine dinucleotide, whose binding is mediated by the nudix hydrolase domain. Reactive oxygen species such as H_2O_2 and tumour-necrosis factor-α potentiate TRPM2, suggesting that it functions as intracellular oxidation/reduction sensor [59]. TRPM3 can form homotetramer or heterotetramer with different isoforms of TRPM3 or with TRPM1. The heterologously expressed channels show constitutive activity. However, it is not yet certain that these channels are the same as the ones expressed in native tissues. It has been shown that TRPM3 can be activated by sphingolipids, pregnenolone sulfate, and nifedipine. Its activity is potentiated by cell swelling and extracellular hypotonic solutions [60].

TRPM4 and TRPM5 are distict monovalent-cation selective channels among the TRP superfamily. They are activated by high level of intracellular Ca^{2+} with varied sensitivity depending on membrane potential, temperature, phosphorylation state of the channel and binding of calmodulin or PIP_2 [61]. TRPM5 has been shown to be involved in taste sensations, where reduction in responses to sweet, umami and bitter tastes was seen in double knockout mice, with sour and salty taste unaffected [62]. TRPM6 and TRPM7 have dual functionality exhibiting both ion channel and protein kinase activities. They are permeable to Mg^{2+}, Zn^{2+} and Ca^{2+}, and are inhibited by intracellular free Mg^{2+} in the range of 0.3 to 1.0 mM [63, 64]. Mutations in TRPM7 have been shown to affect cell viability and proliferation due to Mg^{2+} deficiency [64]. Familial hypomagnesemia with secondary hypocalcemia in human is linked to mutations in TRPM6 [65]. These suggest that TRPM6 an TRPM7 are important for Mg^{2+} homeostasis. The significance of the kinase activity is yet clear. It has been reported that the kinase domain is not essential for TRPM7 activation, but its Mg^{2+} sensitivity changes with structural modifications in this region [66].

TRPM8 is a cold ($8 - 28$ ℃) and 'cooling' agents (such as menthol, eucalyptol and icilin) activated nonselective cation channel [67, 68]. Temperature and compounds appear to act on the channel through different mechanisms. Mutagenesis studies have shown that channels with mutations that significantly affect the response to menthol can still be gated by cold [69]. Inhibition of TRPM8 channel activity by depletion of PIP_2 has been reported, and mutations that affect this modulation have been located in the TRP domain [70].

1.1.4 The TRPA Subfamily

TRPA (ankyrin) subfamily has only one member TRPA1, which is distinguished from other members of the superfamily by the absence of TRP motif and the presence of extensive ankyrin repeats ($14 - 17$ repeats) in the N-terminal domain. It is activated by a wide range of structurally diverse exogenous and endogenous

compounds including cysteine and lysine reactive electrophilic molecules like allyl isothiocyanate [71], allicin [72], cinnamaldehyde and acrolein [73], also non-electrophilic compounds such as menthol and cannabinoids [71], and endogenous ligands including bradykinin [74] and oxidants [75]. Whether TRPA1 can be activated by noxious cold is still contentious. Experiments showed that heterologously expressed channels can be activated by noxious cold (< 17 ℃) [74], but controversial results have been reported for proteins in cultured cells [71] and TRPA1-deficient mice [73].

1.1.5 The TRPML Subfamily

There are three (TRPML1−3) relatively small channels in the TRPML (mucolipin) subfamily, each with less than 600 amino acids. All of them contain an extracytosolic/lumenal domain in the S1-S2 linker region. They are mostly located in membranes of intracellular vesicles, mainly late endosomes and lysosomes (LELs), but can have regulated and event dependent trafficking to plasma membrane [76]. A defining feature of TRPML channels is the cytosolic domain consisting of interacting segments of short pre-S1, S2-S3 linker and post-S6 helices [77]. All these members are activated by $PI(3,5)P_2$. Structural studies of TRPML1 and TRPML3 suggest that the $PI(3,5)P_2$ binding site is located at the common cytosolic domain [77–80].

TRPML1 is the first member identified in this subfamily. It has been shown to be permeable to Ca^{2+}, Fe^{2+}, Zn^{2+}, Na^+ and K^+ with strong inwardly rectifying currents (lumen to cytosol). The channel activity is inhibited by $PI(4,5)P_2$ [76]. Given the difference in localisation of $PI(3,5)P_2$ (mostly in intracellular vesicles) and $PI(4,5)P_2$ (mainly in plasma membrane), it is possible that TRPML1 in plasma membrane are predominantly non-active [76]. Mutations in TRPML1 are linked to type IV mucolipidosis, which is an autosomal-recessive lysosomal storage disease characterised by enlargement of LELs and as a result, accumulation of lysosomal storage materials at cellular level leads to severe neurodegeneration [81].

Since there is no known disease implication for TRPML2, it has not been well functionally characterised. Expression of TRPML2 is largely restricted to lymphoid and myeloid tissues, suggesting a role in the immune system [82]. TRPML3 is an inwardly rectifying cation channel with relative selectivity for Ca^{2+} over monovalent cations. It is expressed in intracellular organelles of hair cells and plasma membrane of stereocilia. A gain of function mutation A419P in TRPML3 has been linked to the *varitint-waddler* phenotype in mice, which exhibits pigmentation defect, deafness and circling behaviour [83].

1.1.6 The TRPP Subfamily

The reclassification of members in the TRPP (polycystin) subfamily caused some confusion in the nomenclature of the channels. Proteins in the polycystin-1 family were initially classified in the TRPP subfamily. However, because they have significantly different molecular architecture and functional features to a TRP channel, they were later excluded from the TRPP subfamily. The three real ion channels from the polycystin-2 family remain in the TRPP subfamily. The names of members of both polycystin-1 and polycystin-2 families are listed in Table 1.2. To avoid ambiguity, the protein names will be used in this thesis.

Table 1.2: Nomenclature of Polycystin Proteins

TRPP Channels		
Gene	**Protein**	**TRP nomenclature**
Polycystic kidney disease 2 (PKD2)	Polycystin-2 (PC2)	TRPP2 or TRPP1
PKD2-like 1 (PKD2L1)	Polycystin-2 like 1 (PKD2L1)	TRPP3 or TRPP2
PKD2-like 1 (PKD2L2)	Pplycystin-2 like 2 (PKD2L2)	TRPP5 or TRPP3
Polycystin-1 Family Proteins		
Gene		**Protein**
Polycystic kidney disease 1 (PKD1)		Polycystin-1 (PC1)
PKD1-like 1 (PKD1L1)		Polycystin-1 like 1 (PKD1L1)
PKD1-like 2 (PKD1L2)		Polycystin-1 like 2 (PKD1L2)
PKD1-like 3 (PKD1L3)		Polycystin-1 like 3 (PKD1L3)
PKD and sperm receptor for egg jelly homolog-like (PKDREJ)		polycystin-REJ (PKDREJ)

TRPP channels are Ca^{2+}-permeable cation channels with high sequence homology in the transmembrane region. However, there is a large divergence in their cytosolic domains [84]. The physiological functions of endogenous polycytin-2 like 1 (PKD2L1) and polycytin-2 like 2 (PKD2L2) are still elusive, so it is not clear if the difference in the cytosolic domains has any functional importance. Similar to members of the TRPML subfamily, there is a large insertion between the transmembrane helices S1 and S2 common to all TRPP channels. The TRPP subfamily is one of the most conserved TRP families with presence throughout the animal kingdom with relatively high sequence identity among different species, especially for mammals, highlighting the importance of this class of proteins [84, 85].

The focus of this thesis is polycystin-2 (PC2), which is the founding member of TRPP subfamily encoded by *polycystic kidney disease 2* gene consisting of 968 amino acids. It is widely distributed in many tissues, including kidney, heart, testis, intestine, vascular smooth muscle, pancreas, placenta and bile ducts, with relatively high expression in tubules within kidney cells [86]. The subcellular transport and localisation of PC2 is highly debated. Studies have shown that heterologously expressed PC2 is predominately expressed in intracellular compartments, especially in the endoplasmic reticulum (ER) [87, 88]. It was proposed that there is a ER-retention signal at the Glu787 − Ser820 region in the C-terminal domain preventing PC2 trafficking to the plasma membrane, and it was also shown that PC2 is glycosylated with high-mannose oligosaccharides suggesting that it is a pre-medial-cisternae protein [87]. Contradicting results have reported endogenous expression of PC2 in the Golgi apparatus and the plasma membrane of mouse inner medullar collecting duct (IMCD) cells and Madine–Darby canine kidney (MDCK) cells [89, 90]. Distribution in the basolateral plasma membranes, lamellipodia and mitotic spindles has also been noted [91, 92]. Despite the continued debate over the localisation of PC2, there is a growing consensus that PC2 is present in primary cilia.

Homotetrameric PC2 has the same fold as other TRP channels with the voltage sensor-like domain (VSLD, S1-S4) packed against the pore domain (S5-Pore-S6) of

the adjacent chain. An external (organelle luminal/extracellular) domain (referred to as the TOP domain) is formed from a 218-residue insertion between S1 and S2 transmembrane helices, and interacts with S3 and S4 of the VSLD through a 20-amino-acid extension of the S3-S4 loop (Figure 1.5) [7, 13, 15]. All of the published cryo-electron microscopy (cryo-EM) structures of PC2 are lack of the cytosolic domain. However, both C- and N-terminal regions have been extensively researched due to their implications in channel function, subcellular localisation and interactions with other proteins. It has been shown that the C-terminal domain has four functional regions: a canonical Ca^{2+} binding EF-hand motif ($PC2^{754-781}$), an oligomeric coiled coil domain (CC1: $PC2^{769-796}$/CC2: $PC2^{835-873}$), an ER retention signal ($PC2^{787-820}$) and an acidic amino acid cluster ($PC2^{810-821}$) [87, 93–100]. The EF-hand has been shown to undergo conformational changes upon binding of Ca^{2+}, and thus mediates Ca^{2+} dependent channel gating [98, 100]. The acidic cluster, which acts as a binding site for the phosphofurin acidic cluster sorting protein (PACS)-1 and PACS-2, and the ER retention signal contribute to the localisation of PC2 [87, 94]. There are some controversies concerning the role of the coil-coil domain. It was suggested that both homo- and hetero-oligomerisation of PC2 subunits requires the coil-coil domain [97, 101–103]. There are also reports demonstrating that the absence of the coil-coil domain does not affect the formation of PC2 complexes, and thus suggested that oligomerisation is mediated by the N-terminal domain [104–106]. However, cryo-EM structures of PC2 homotertramer and PC1/PC2 complex showed that neither N- nor C-terminal is essential for oligomerisation [7, 13, 14]. Several phosphorylation sites have been identified. Ser76 at the N-terminal domain has been shown to be phosphorylated by glycogen synthase kinase 3 (GSK3). The phosphorylation state of this GSK3 site regulates the redistribution of PC2. Inhibition of phosphorylation at Ser76 triggers relocation of PC2 from the lateral plasma membrane to the intracellular compartments without disturbance in the cilliary distribution [107].

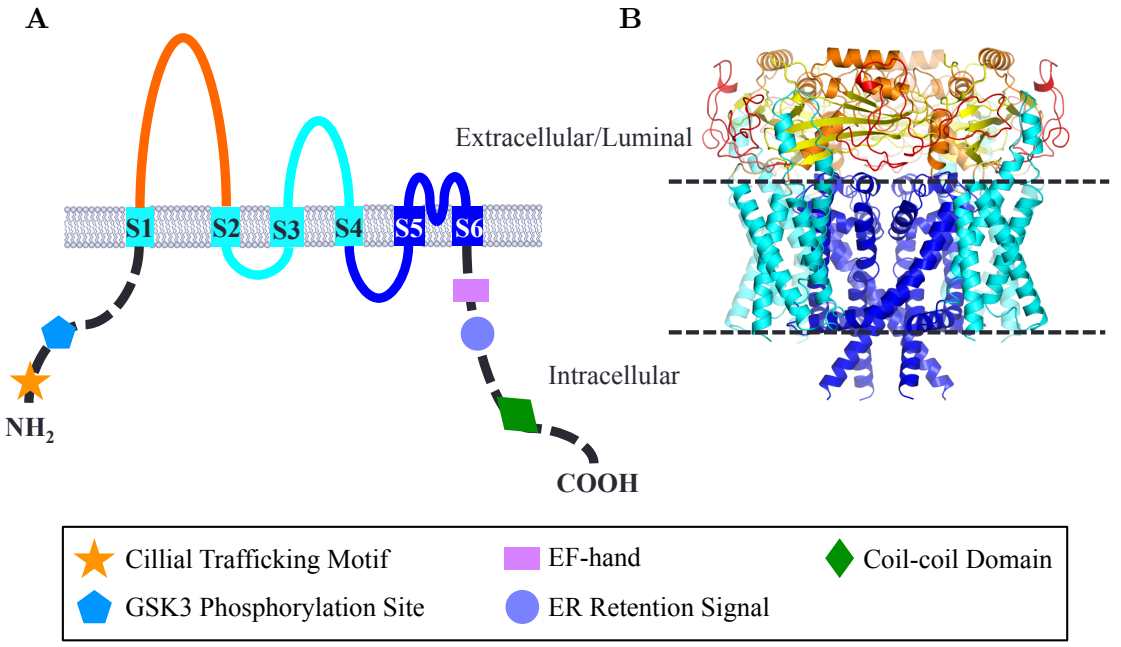


Figure 1.5: Schematic representation (**A**) and cyro-EM structure (**B**) of PC2. The unresolved cytosolic domain is indicated as dashed lines in the schematic representation.

1.2 Proteins of the Polycystin Family

In addition to the members of the TRPP subfamily, the polycystin family also includes five members from the polycystin-1 family: polycystin-1 (PC1), polycystin-1 like 1 (PKD1L1), polycystin-1 like 2 (PKD1L2), polycystin-1 like 3 (PKD1L3) and polycystin-REJ (PKDREJ). They are predicted to have 11 membrane-spanning regions with the C-terminal six transmembrane domains (S6 – S7) sharing high sequence homology with TRPP channels.

PC1 is a large glycoprotein of 4,303 residues with a large extracellular N-teminal domain ($\sim$ 3000 amino acids) and an intracellular C-terminal end ($\sim$ 200 amino acids) further to the transmembrane regions. PC1 has been suggested to function as a cell surface receptor mediating cell-cell and cell-matrix interactions by coupling extracellular chemical or mechanical stimuli to intracellular signalling [108–111]. Studies have shown that all four major classes of G proteins ($G\alpha_{i/o}$, $G\alpha_{q/11}$, $G\alpha_{s}$,

$G\alpha_{12/13}$) can be constitutively activated by heterologously expressed full-length PC1 [112, 113]. Despite the sequence similarity in the last six transmembrane segments with TRP channels, there is no measurable channel activity in PC1 homomers, which may suggest that it does not form ion channel in homomeric state [93, 112].

Germ cell specific expression has been seen for PKDREJ [114]. It is localised on the acrosomal region of the plasma membrane of mature sperm, suggesting that it may be involved in sperm function and fertilisation [114]. *In vivo* experiments have shown that $PKDREJ^{-/-}$ mice have a delayed sperm capacitation comparing to wild type, which leads to a slow development of acrosome reaction [115]. PKD1L1 is predominately expressed in the heart, the mammary gland and the Leydig cells of the testis [116]. PKD1L1 knock-out mice showed significantly lower viability and some with situs inversus phenotype [117, 118], implicating that PKD1L1 plays an important role in developmental function within the embryonic node. PKD1L2 has similar expression pattern as PKD1L1 with yet clear function. It has been reported to bind to different G-protein subunits, which suggests that PKD1L2 may function as a G-protein-coupled receptor (GPCR) [119].

Various heteromeric complexes formed by members of polycystin-1 and -2 families have been reported. The assembly of PC1 and PC2 is thought to be a receptor-channel complex involved in Ca^{2+} dependent signalling pathway. It was proposed that they interact through their C-terminal coil-coil domains [102, 103]. More recent studies show that the extracellular/luminal domain between S1-S2 of PC2 and S6-S7 of PC1 is crucial for the formation of the polycystin complex [120]. A cryo-EM structure of the PC1/PC2 complex provided evidence for the 1:3 PC1:PC2 stoichiometry and showed that they assemble without the C-terminal domains (Figure 1.6) [14]. Co-expression of PKD1L3 and PKD2L1 has been reported in specific subset of taste receptor cells [121–123]. The complex has been proposed to function as a sour taste receptor given its pH-dependent activity in heterologous over-expression systems [121].

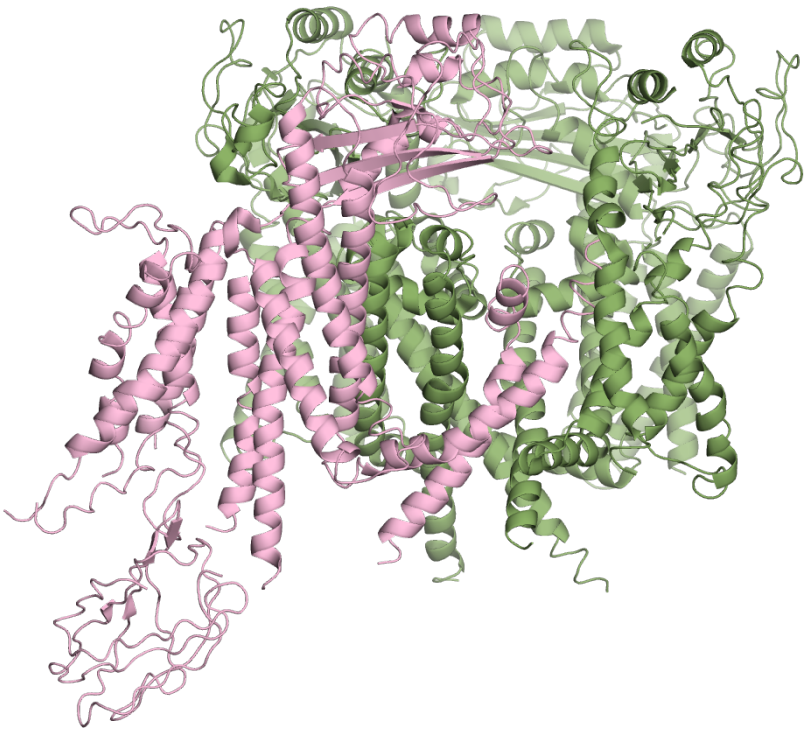


Figure 1.6: Cyro-EM structure of PC1/PC2 complex (PDB ID: 6A70). It has a 1:3 PC1:PC2 stoichiometry. PC2 subunits are shown in green and PC1 subunit is shown in pink.

1.3 Autosomal Dominant Polycystic Kidney Disease

Autosomal dominant polycystic kidney disease (ADPKD) is one of the most prevalent monogenic disorders in human, occurring in approximately one in 800 live births and affecting 4 to 6 million people worldwide [124]. Phenotypically, ADPKD is characterised by the formation and accumulation of fluid-filled renal cysts, which leads to progressive cystic enlargement of both kidneys and ultimately kidney failure (Figure 1.7). Furthermore, cysts are frequently developed in intestines, liver, pancreas and choroid plexus [125]. ADPKD is also associated with increased risk of cardiovascular dysfunctions including aortic aneurysms, hypertension and heart-valve defects [126]. Patients with ADPKD usually present clinical manifestations in their thirties and forties, although early/childhood onsets have also been

reported. The end-stage renal disease generally occurs 10 to 20 years after the initial development of renal deficiency [125].

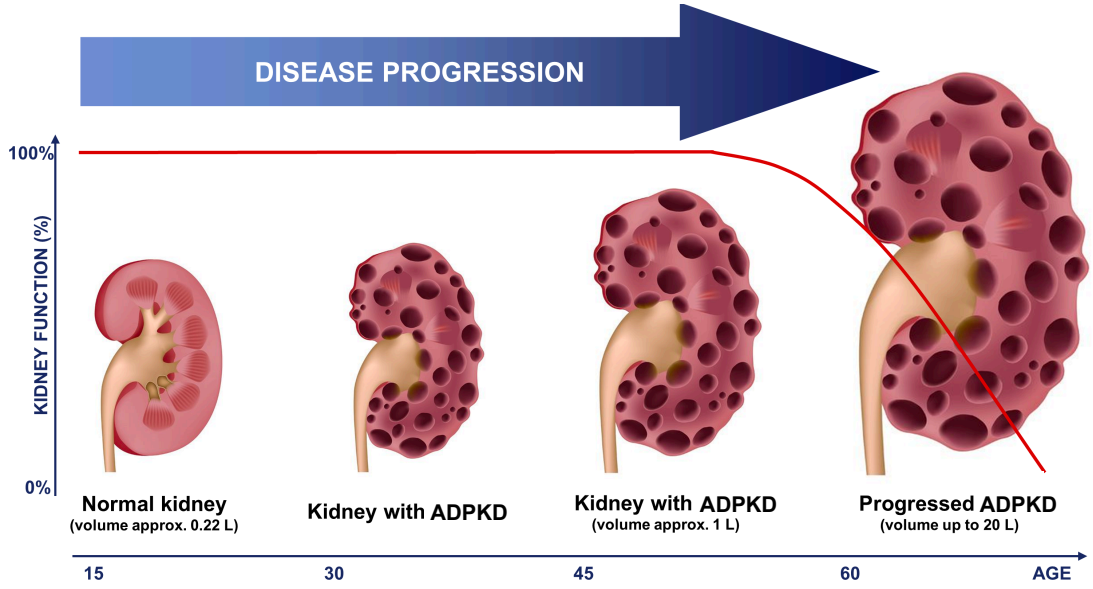


Figure 1.7: Disease progression of ADPKD. Figure is taken from `https://mosaiques-diagnostics.de/mosaiques-diagnostics/Polycystic_kidney_disease`.

Mutations in PC1 and PC2 are responsible for almost all cases of ADPKD. About 85% of cases of ADPKD are due to mutations in PC1, and the remaining $\sim$ 15% are caused by mutations in PC2. In some clinical cases, where typical symptoms of ADPKD are presented, there is no mutations in PC1 or PC2, and thus the existence of a third polycystic kidney disease gene causing rare forms of ADPKD has been suggested. However, it has yet been identified [127]. Although patients with mutations in either genes have similar pathological and physiological features, PC1 mutations lead to earlier onset with a faster rate of progression to renal failure comparing to PC2 mutations. PC1 truncation mutations have the worst renal survival [128]. It has been shown that homozygotes or compound heterozygotes are embryonically lethal [129]. Trans-heterozygous patients with mutations in both PC1 and PC2 usually survive to adulthood but have a more

severe clinical course [130]. Genetic data analysis has shown that there are at least 2323 mutations in PC1 and 278 mutations in PC2 associated with ADPKD [131]. It was initially thought that the onset of the development of renal and hepatic cysts occur after acquiring somatic mutations resulting in a complete loss of both *PKD1* or *PKD2* alleles [132, 133]. However, studies have found that hypomorphic alleles also contribute to cystic phenotype. Therefore, a threshold model was proposed for the development of ADPKD, suggesting that cystogenesis occur if the functions of the polycystin proteins are below a certain dosage [134–136]. The underlying mechanisms by which these mutations lead to ADPKD are still poorly understood. Both loss of function and over-expression of PC1 and PC2 have been demonstrated to cause cyst formation and kidney tubular abnormalities [128, 137, 138].

The primary cilium is a hair-like organelle projecting from the surface of almost all vertebrate cells, and functions as a sensor for extracellular mechanical and chemical stimuli [139–142]. Research across different species have highlighted the possibility of a common mechanism in ADPKD pathology. They showed that most of cyst formation associated proteins are expressed in the primary cilium and the lack of primary cilia in conditionally mutated model organisms leads to the development of cysts, and hence strong evidence have suggested that structural or functional defects in primary cilia play a crucial role in the development of polycystic kidney disease [143–145]. Localisations of PC1 and PC2 in the primary cilia have been reported, and genetic ablation of *PKD1* or *PKD2* caused abnormal elongation of the primary cilia of cystic cells [146].

The PC1 and PC2 receptor-channel complex is proposed to function as a environmental sensor that regulates localised Ca^{2+} influx at the primary cilium. However, the type of stimuli, to which the polycystin complex responds, is still an active area of research. It was thought that the large N-terminal domain of PC1 senses flow-dependent mechanosensory stimuli, and in turn triggers PC2 channel opening to increase ciliary Ca^{2+} concentration [110, 147]. This stimulates downstream signalling pathways to alter gene expressions, and hence regulate cell proliferation and apoptosis, cell differentiation and fluid secretion [147]. The loss of

the flow-mediated regulation leads to cyst formation [110]. Nevertheless, this model have recently been challenged. It was shown that in the presence of fluid flow, there was no cilia-specific Ca^{2+} influx in primary cilia, suggesting that Ca^{2+} signalling is not involve in cilia mechanosensation [148]. Several studies have raised the possibility that the PC1/PC2 complex is ligand-activated, and the ligand-mediated signalling transduction regulates the tubular epithelial cell polarity and size [149, 150]. The relationship between ADPKD and dysfunction of polycystin proteins is not completely understood, and remains to be elucidated.

1.4 Protein Lipid Interactions

Integral membrane proteins inevitably interact with and often regulated by lipids within their local membrane environment. Lipids usually execute their roles in two main ways: firstly acting directly on the proteins through binding, and secondly regulating the lateral pressure through membrane thickness and fluidity. Direct bindings of lipids may exert their effects by directly affecting the activities of membrane proteins, modulating the localisation of membrane proteins or altering membrane-mediated interactions between proteins[152]. The lateral pressure profile underlies membrane properties including surface tension and spontaneous curvature, which can affect the functionality of membrane proteins such as mechanosensitive channels [151].

Both experimental and computational techniques have been developed and employed to characterise protein lipid interactions. Structural identification of lipid binding sites has been largely achieved by high-resolution X-ray crystallography. Although the solubilisation and purification processes usually extract the membrane proteins out of the lipid environment, tightly bound lipids, which are either co-purified with the proteins or added at crystallisation step, have been observed in crystal structures such as the structures of bacteriorhodopsin [153], voltage-dependent K^+ (Kv) channels [154] and the inward rectifier K^+ (Kir) channels [155, 156] (Figure 1.8A). Electron crystallography has been used to solve structures of

reconstituted membrane proteins in lipid bilayers that form 2-dimensional crystals [157–159]. The rapid advances in single particle cryo-EM have greatly broadened the spectrum of membrane protein structures. Atomic details of lipid binding have been seen in cryo-EM structures of membrane proteins such as the TRP channels [6, 53, 160], the nicotinic acetylcholine receptor (nAChR) [161], the $GABA_A$ receptors [162] and the ABC transporter ABCG2 [163] (Figure 1.8B). Native mass-spectroscopy (MS) has been used to quantitatively analyse and identify bound lipids for proteins such as V-type ATPases [164] and GPCRs [165]. It can provide information that is not readily observable from structural studies such as lipid selectivity and thermodynamics of lipid binding. For example, individual lipid binding events of the *E. coli* ammonia channel AmtB were resolved using native MS. The thermodynamic signatures and entropy–enthalpy compensation for binding of different species of lipids or lipids with different acyl chain lengths were characterised to provide insights on lipid binding specificity [166]. Other experimental techniques including nuclear magnetic resonance [167], infrared spectroscopy [168], Förster resonance energy transfer [169] and fluorescence cross-correlation spectroscopy [170] have been utilised to investigate lipid interactions and lipid composition around membrane proteins.

Computationally, molecular simulations are widely employed to characterise the dynamics of protein lipid interactions. The most commonly used models for studying these interactions are atomistic and coarse-grained (CG) models. Both approaches allow membrane proteins to be embedded in a lipid bilayer environment, and therefore, not only a specific lipid binding event can be studied, the impact of the dynamic structure of the membrane can also be accounted for. For example, annulus lipid layer is usually formed around integral membrane proteins. These lipids have relatively low mobility comparing to bulk lipids because of their interactions with the protein. How these lipid shells affect the specific lipid binding to membrane proteins could be addressed by molecular simulations. At atomistic level, the detailed chemical and physical interactions between atoms can be described, and thus the accurate coordination of lipids by proteins can

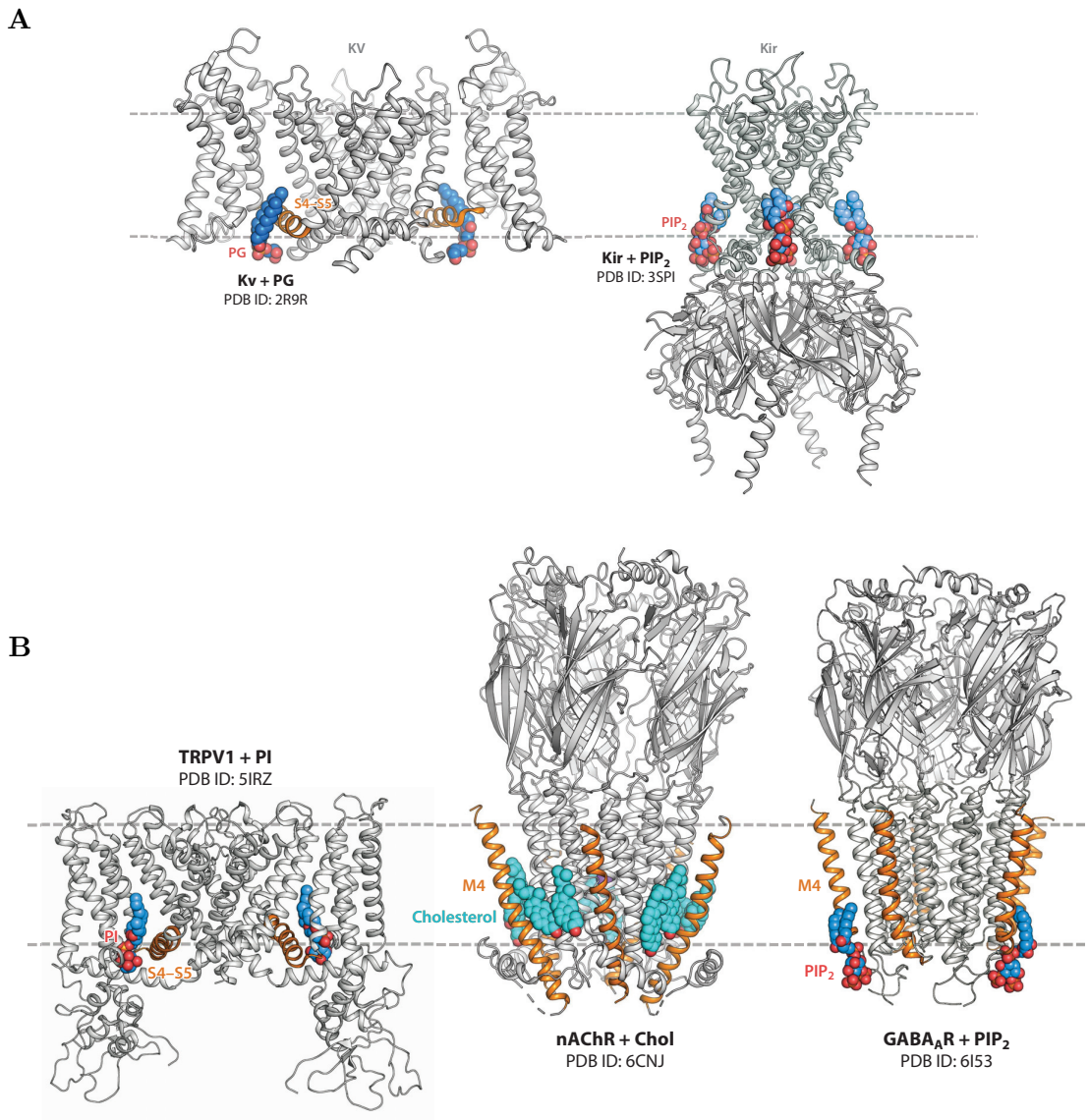


Figure 1.8: Structures showing protein lipid interactions. **A:** Crystal structures of the voltage-dependent K^+ channel, Kv1.2/Kv2.1 chimera with bound PG lipids (PDB ID: 2R9R) and the inward rectifier K^+ channel, Kir2.2 with bound PIP_2 molecule. **B:** Cryo-EM structures of the TRP channel, TRPV1 with bound PI lipid (PDB ID: 5IRZ), the nicotinic acetylcholine receptor (nAChR), $\alpha 4\beta 2$ nicotinic receptor with bound cholesterol molecules (PDB ID: 6CNJ), and the $\alpha 1\beta 3\gamma 2$ $GABA_A$ receptor with bound PIP_2 molecules (PDB ID: 6I53). Proteins are shown in cartoon representation and lipids are shown in spheres. Figures are taken from [171].

be examined, which could provides insights into localised structural changes of proteins upon lipid binding. For example, cholesterol interactions with GPCRs have been extensively characterised atomistically. A long simulation (1.6 μs) study showed that binding of cholesterols perturbs kinks in the transmembrane helices TM1, TM2 and TM7 leading to functionally important structural rearrangement of rhodopsin [172].

However, the size of the system and the timescale accessible by atomistic simulations are rather limited due to the associated computational cost. This has restricted most of atomistic studies of protein lipid interactions to be of a single protein system with limited complexity of lipid mixture. To compensate the limits of atomistic simulations, CG models have been developed to facilitate studies with longer time and more particles at the expense of fine detailed chemical description of molecules. They allow more comprehensive search of lipid binding sites on membrane proteins, studies of lipid-mediated protein protein interactions and also investigations of the effect of membrane structures on integral proteins. The simulation study of the large bacterial mechanosensitive channel (MscL) showing that acute tilting of transmembrane helices caused by lateral membrane tension is essential for channel opening is a good demonstration of the application of CG simulations [173]. A CG simulation of 100 μs was used to identify the binding sites of cardiolipins on the respiratory chain complex III (cytochrome bc_1 dimer) [174]. Large systems like the dengue virion envelope [175] have also been studied at CG level.

Nowadays, multiscale approaches combining atomistic and coarse-grained simulations have become more popular taking advantages of both models. They have been successfully used to predict lipid binding sites and binding poses, even before structural evidence became available. One of the first computational studies of Kir channels applied the multiscale method to investigate the binding of PIP_2 on the bacterial Kir channel KirBac1.1, a Kir3.1-KirBac1.3 chimera channel, and a homology model of the Kir6.2 channel [176]. It successfully located a conserved binding site present in all three Kir channels and identified the residues contributing

to the coordination of PIP_2, which were later confirmed by crystal structures of Kir2.2/PIP_2 complex [155].

1.5 Overall Aims and Motivation

Given the disease implication of PC2, it is of general interest to gain better understanding of the protein. There still is a large amount of unknown/uncertain aspects of PC2 awaiting scientific exploration, such as the mechanisms of lipid modulation, channel gating and ion selectivity. This work aims to combine structural studies and molecular simulations to further explore the structure-function relationship of PC2, and to provide insights into interactions of PC2 with lipids and ions. Cryo-EM was used as the primary structure determination method. Molecular dynamics (MD) simulations were used to probe the protein lipid interactions and ion passage of PC2. This work also aims to provide a demonstration for combining the fast growing cryo-EM and MD simulations as a powerful tool to explore the interactions of membrane proteins.

2

Theory and Methods

Contents

2.1 Cryo-electron Microscopy 28
2.1.1 Protein Purification 31
2.1.2 Specimen Preparation 33
2.1.3 Image Acquisition 35
2.1.4 Image Processing 36
2.1.5 Model Building and Validation 38
2.2 Molecular Dynamics Simulations 39
2.2.1 Force Fields 40
2.2.1.1 Intramolecular Interactions 41
2.2.1.2 Non-bonded Interactions 42
2.2.1.3 Force Field Parameterisation 44
2.2.2 Simulation Methods 45
2.2.2.1 Periodic Boundary Conditions 45
2.2.2.2 Integration algorithms 46
2.2.2.3 Thermodynamic Ensembles 47
2.2.3 Coarse-grained Simulations 50
2.2.4 Free Energy Calculations 53
2.2.4.1 Umbrella Sampling 54
2.2.4.2 Potential of Mean Force 55

The aim of this chapter is to provide an overview of the basic principles and theories of cryo-electron microscopy and molecular simulations that are related to the contents of this thesis. Detailed experimental procedures and simulation settings are described in the Methods section of each subsequent chapter.

2.1 Cryo-electron Microscopy

Cryo-electron Microscopy (cryo-EM) is gradually becoming a mainstream method for studying the structures of biological molecules given the rapid advancement in the imagining techniques to provide high-resolution 3D structural information. In contrast to X-ray crystallography, there is no requirement for cryo-EM samples to be crystallised, which greatly broadens the spectrum of biomolecules that can be studied by cryo-EM. In addition, vitrified cryo-EM specimens are imaged at cryogenic temperatures without application of dyes or fixatives, which decreases the possibility of structure alteration by crystalline packing and allows biological macromolecules to be studied in their native state with more dynamic conformations [177]. Although nuclear magnetic resonance (NMR) spectroscopy also allows detailed study of different conformational states or ligand binding, the size of proteins that can be studied is limited (usually below 50 kDa), because spectral complexity increases with molecular weight. Also, large biomolecules, especially the ones in membranes or detergent, tumble at lower rates and have shorter spin relaxation time, which lead to loss in sensitivity of pulse sequences [178].

Cryo-EM makes it possible to study macromolecules that are difficult to be trapped in a stable crystal lattice, such as biological entities with very dynamic behaviour and large assemblies with complex quaternary structures. Since the scattering ability of biological molecule under investigation has to be sufficiently strong in comparison to its surrounding buffer to provide distinguishable contrast, it imposes a lower limit of mass on biomolecules that can be structurally characterised by cryo-EM. However, the introduction of direct electron detection device (DDD)

cameras has enabled smaller molecules to be studied by cryo-EM. Therefore, unlike X-ray crystallography and NMR, the range of molecular mass of the specimens can span from tens kilo-Daltons to several mega-Daltons for cryo-EM, and the amount of sample required is much smaller [179], which lowers the physical requirement for the type of macromolecules to be studied.

The basic principle of electron microscopy is analogous to light microscopy. Instead of photons, by using electrons as a source of radiation, much smaller components and structures of materials can be investigated, due to the shorter wavelengths of electrons than photons, which allow them to interact with nano-scale matters. Depending on the kinetic energy of the electron, its physical properties can be summarised as shown in Table 2.1:

Table 2.1: Properties of Electron

	Non-relativistic ($E << m_0c^2$)	**Relativistic** ($E \sim m_0c^2$)
Mass	$m = m_0$	$m = m_0(1 - v^2/c^2)^{-1/2}$
Energy	$E = \frac{1}{2}m_0v^2$	$mc^2 = m_0c^2 + E$ $m = m_0(1 + E/(m_0c^2))$
Velocity	$v = (2E/m_0)^{1/2}$	$v = c\left[1 - \frac{1}{(1+E/(m_0c^2))^2}\right]^{1/2}$
Momentum	$p = m_0v = (2m_0E)^{1/2}$	$p = mv = [2m_0E(1 + E/(2m_0c^2))]^{1/2}$ $= \frac{1}{c}(2Em_0c^2 + E^2)^{1/2}$
de Broglie wavelength	$\lambda = \frac{h}{p} = h(2m_0E)^{-1/2}$	$\lambda = \frac{hc}{pc} = hc(E^2 + 2Em_0c^2)^{-1/2}$

1. h: the Planck constant, c: speed of light, E: kinetic energy of motion, m_0: rest mass of electron.

Given these properties, the wavelengths of electrons can be calculated at a given accelerating voltage. The typically used accelerating voltage of an electron microscope is between 80 to 300 kV, which means that the wavelengths of electrons

are approximately ranging between 0.02 to 0.04 Å. These are at least twenty fold smaller than the typical wavelength of X-ray used for crystallography ($\sim$ 1 Å). Therefore, the resolving power of electrons is theoretically much batter than X-ray.

When high-energy electron beam hits a specimen, most of the electrons passes through without interaction, and there is no change in energy and direction. The rest may be scattered in three ways: 1) backwards, which will not be captured by electron detectors; 2) elastically through interactions with Coulomb potential of nucleus; 3) inelastically through interactions with electrons of atoms (Figure 2.1). The specimen-electron interactions cause a phase shift in the scattered electron wave relative to the unscattered electrons. This generates contrast in transmission electron microscope. Elastically scattered electrons do not transferred energy to the sample and contribute to the formation of two-dimensional (2D) images of the specimen. Inelastic scattering results in a loss of energy of the incident electrons, which results in different signals such as X-ray, secondary electrons or plasmons, and thus contribute to the noise [180].

Biological materials that are mainly composed of light elements (C, O, N and H) scatter electrons poorly, which leads to low contrast. In addition, they are vulnerable to radiation damage, which occurs in three stages: the primary radiation damage from the interaction with electrons that results in ionisation of the samples and cleavage of chemical bonds, the secondary damage from further chemical reactions with the secondary electrons and free radicals produced in the previous stage, and the tertiary morphological changes to the materials caused by the hydrogen gas developed within the sample upon interaction with electrons [181]. Therefore, it is difficult to obtaining high-resolution images with a level of electron dose that preserves the structural integrity of the biological sample. One of possible approaches to minimise radiation damage and obtain high-resolution images is to use plunge-frozen specimens and acquire images at either liquid nitrogen or liquid helium temperatures. The remarkable progress in the development of DDD cameras has allowed images to be taken under relatively low electron exposure with reasonably high signal-to-noise ratio, and together with

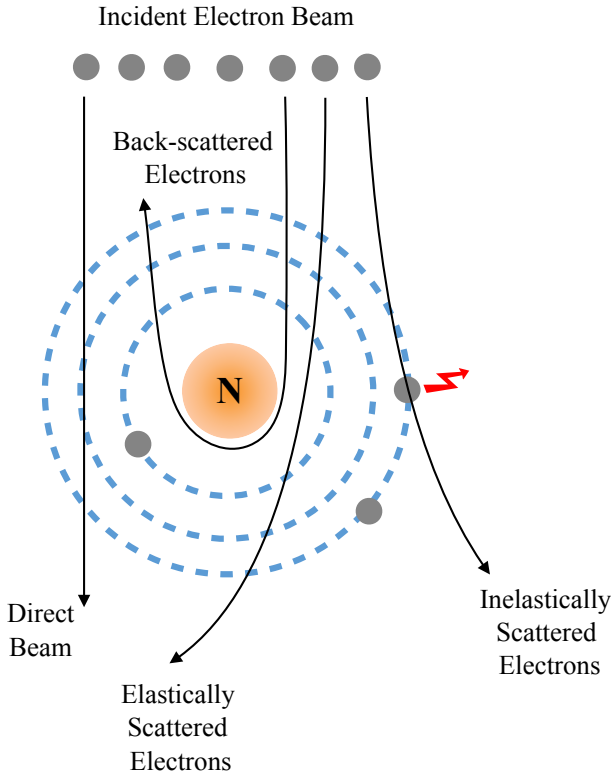


Figure 2.1: Schematics of electron scattering.

rapid improvement in computational algorithms for processing the images such as averaging large number of identical units and correcting specimen movements have enabled reconstruction of high-resolution structures of biological assemblies [182]. The advantages in the physical properties of electrons and the advances in the cryo-EM related technologies have started to shift the structural biology field towards the era of cryo-EM. The goal of this section is to provide an overview of the workflow (Fig. 2.2) for determining a structure using single-particle cryo-EM.

2.1.1 Protein Purification

For single-particle cryo-EM, three-dimensional (3D) reconstruction of the structure relies on computational averaging of a large number of 2D projections of structurally homogeneous particles in different orientation. In order to obtain stable conformationally and compositionally homogeneous particles, biochemical

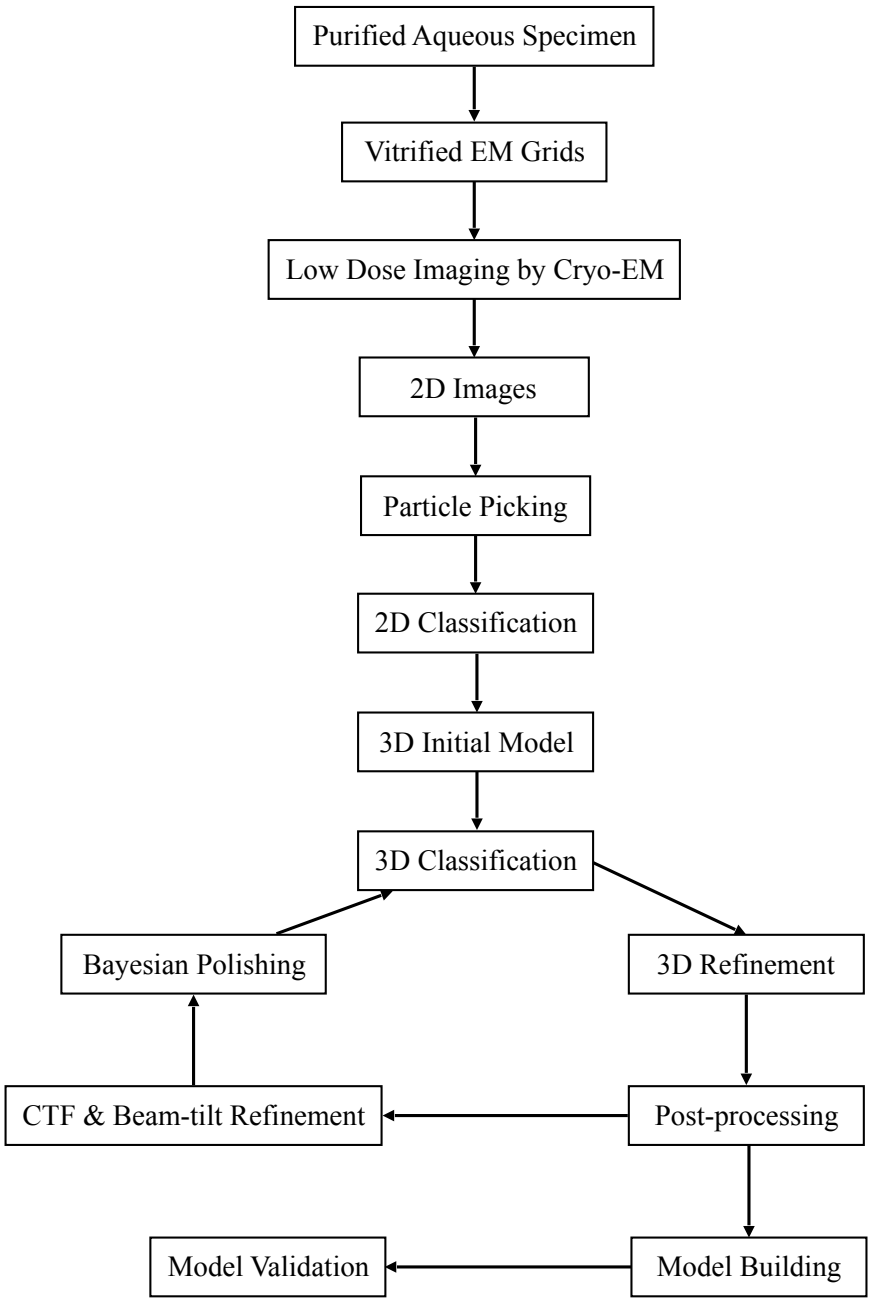


Figure 2.2: Overview of the steps involved in structural determination by cryo-EM.

procedures should be taken to minimise heterogeneity. SDS-PAGE and size exclusion chromatography are primary means to ensure size and to some extent con-

formational homogeneity, but they do not provide sufficient measure of suitability for single-particle cryo-EM. For example, they might not be able to efficiently separate conformationally or compositionally different complexes with very similar size and overall intact shape. Thus negative-stain EM is often used to visually assess the quality of a protein sample. Heavy metal salts such as uranyl and tungstate derivatives are commonly used for staining in EM, because they absorb to organic molecules well and scatter electrons strongly, which allows high phase contrast to be produced. Although the resolution of negative-stain EM is limited to approximately 20 Å, sample quality including the size, oligomeric states, conformational states and the presence of impurities or aggregates can be readily examined [183]. Approaches that have been widely used to minimise structural heterogeneity include screening buffer conditions that stabilise the target protein [184], chemical cross-linking [185] that reduce compositional and conformational heterogeneity and adding molecules that lock the target protein in a certain functional state such as inhibitors, substrates and co-factors.

2.1.2 Specimen Preparation

Electrons can be scattered by both the electrons and the nuclei in any molecules. To avoid scattering by matters in the air, the specimens of interest must be placed in a high vacuum within the electron microscope. The approach adopted by cryo-EM to prepare biological specimens for surviving dehydration caused by the vacuum and radiation damages mentioned above is vitrification. For single-particle cryo-EM, purified macromolecules in aqueous solution are usually put on EM grids with a thin holey film made from carbon, gold or silicon derivates supported by conductive materials such as copper, molybdenum and gold [179]. To increase the hydrophilicity of the grids and thus the adhesion of the sample particles, EM grids are typically pretreated with a glow discharger or plasma cleaner before samples are applied. Excessive solution is removed with filter paper under strictly controlled conditions including the blotting force and time, ambient humidity and temperature. Then the grid is plunge-frozen in a cryogen such as liquid ethane cooled to

liquid nitrogen temperature. These parameters during this sample preparation process can be reasonably managed by commercially available semi-automated plungers. However, optimisation is needed for different samples to produce good ice thickness. The ideal vitrified specimen should have a non-crystalline ice layer with optimal thickness containing well distributed mono-layer particles in a wide range of orientations, which provides complementary information about the 3D construction of the particle of interest (Fig. 2.3). The interaction between the molecules with the air/solution interface might cause the particles to be in one dominant orientation, which would lead to lack of information for 3D reconstruction. This problem could be overcome by altering this surface interaction with thicker ice or different buffer condition (such as adding detergent) that lowers the surface tension [182].

Due to the surface tension of the buffer solution, the ice thickness is not uniform across the hole of the carbon film with the edge being thicker than the centre. This makes relatively large molecules more likely to be situated on the edge than the centre or even to be excluded from the centre. The choice of grids with appropriate hold sizes needs to be made to get good ice thickness with well distributed particles. The concentration of the sample applied onto the grids needs to be tested as well. High density of molecules could decrease the number of images required for structure determination, but particles should be distinguishable from each other. The blotting process removes a large percentage of molecules from the grids, so the concentration of the sample needs to compensate the loss. Different adhesion properties of biomolecules could make some preferentially adsorbed to the carbon film than sitting in the ice layer, so the amount of sample applied to the grids needs to be adjusted accordingly. The ice layer needs to be reasonably transparent to provide clear view of the particle with high contrast compared to the surrounding ice. Ice contamination might occur during the transfer of grids between the plunger and the grid box in liquid nitrogen, which would produce strong contrast projections that would hinder the obtainment of distinguishable molecule images. Therefore, the process needs to be quick to minimise exposure to air.

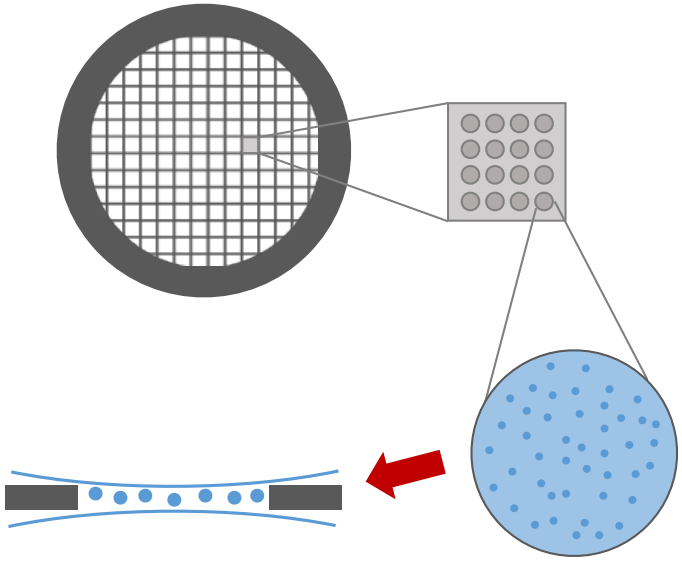


Figure 2.3: Schematics of single-particle cryo-EM specimen.

2.1.3 Image Acquisition

To be able to determine atomistic models of biomolecules, hight-resolution cryo-EM images with high contrast need to be taken on the prepared vitrified specimens. This challenging task requires great care in aligning the electron microscope and choosing the imaging conditions. Because electrons are scattered by not only the biological specimens of interest but also by other molecules in the ice, the contrast of the biomolecules (predominantly light elements) would be very low if images were acquired in focus, and thus cryo-EM images are usually recorded underfocus, which induces phase shifts in the scattered electrons. Contrast is then produced because of the wave interference between the scattered and unscattered electrons. For samples that weakly scatter electron, the densities in the image are linearly proportional to the electrostatic potential of the sample. This relationship between contrast and image in terms of spatial frequency is mathematically described by the Contrast Transfer Function (CTF), which is a quasi-periodic sine function in

reciprocal space whose parameters are dependent on the settings of the microscope. Since CTF is an oscillatory function, it is important to collect images at a range of defocus in order to better sample the reciprocal space so that information lost at zeros can be retrieved [186].

To minimise radiation damage and preserve high-resolution structural information, images are usually taken with low total electron dose, typically below 20 $e^-/Å^2$. Nowadays, DDD cameras have been widely implemented. The significant improvement in the detective quantum efficiency and the underlying technology of the complementary metal-oxide semiconductor have enabled cryo-EM data to be collected as dose-fractionated image stacks rather than a single micrograph [187]. This allows higher total electron dose to be used to boost contrast, and in the mean time the overall signal-to-noise ratios that are affected by the radiation damage can be optimised. The high-frequency noise caused by higher accumulated electron dose from later frames can be discarded or weighted down. Although earlier frames are less affected by radiation damage, they often have fast specimen movements, blurring the high-resolution signals. The fractionated image frames can be aligned to compensate the stage- or beam-induced motion and specimen drift. Therefore, when the image frames are averaged to produce the micrograph for subsequent structural determination, the signal can be improved by optimising the relative weighting for each frame [188].

The pixel size of the camera used for image acquisition depends on the nominal magnification, and it determines the resolution of the final reconstruction, because it is required for defocus determination, CTF correction and the final map interpretation. The pixel size varies between detectors and also depends on the position of the detector in the optical path. This necessitates accurate calibration for each microscope/detector combination as well as each magnification.

2.1.4 Image Processing

A number of image processing software packages have been developed for the analysis of the recorded data. Generally speaking, the image processing involves

motion correction, CTF estimation and correction, particle picking, 2D clustering and averaging, 3D model generation, 3D classification and 3D map refinement. As mentioned earlier, cryo-EM images are usually taken using defocus contrast. The phase-contrast images are modulated by CTF, whose parameters need to be accurately estimated to correct the images for the CTF effects. This is done by fitting the calculated CTF pattern to the amplitude oscillations (Thon rings) seen in the power spectrum of the image. Once the micrographs have been corrected for CTF effects, they can be evaluated based on the visibility of Thon rings to high resolution in all directions [182]. Then the particles can be picked from these good micrographs in a manual, semi-automated or fully automated manner. Caution needs to be taken in this step to avoid orientation bias or model bias in template-based particle picking. After the particles have been identified, they are extracted from the micrographs and collected into a stack. Because the initial coordinates of particles are not precise, the box size used to extract the particles should be at least 30% more than the approximate particle size.

The clustering and analysis of 2D images is very useful for identifying bad or non-particle images, evaluating the angular distribution of the particle views and assessing the quality of the class averages. The algorithm implemented in one of the commonly used image processing package RELION [189] is a maximum-likelihood approach. A set of random images is firstly used as references, and the likelihood of matching to each of the references allowing rotation and translation is calculated for each particle. The 2D Fourier transforms of "similar" particles are averaged and then used as a new set of references. This iterative process is done to produce user-defined number of representative 2D class averages. The generation of initial 3D model can be done with a template or *ab initio*. Templates can be taken from 3D reconstruction of low-resolution negative-stain EM data or EM maps of homologs. Different *ab initio* reconstruction methods have been developed including the random conical tilt approach [190], which requires experimentally determined tilt angles, and stochastic hill climbing algorithm [191], which is a solely computational strategy. The principle of 3D classification is similar to

that of 2D classification. An iterative process of calculating the probability that a certain image arises from each of the 3D references and grouping images is done to identify a subset of images with sufficient homogeneity that gives the best-resolution reconstruction. 3D refinement relies on a 3D projection matching procedure that adjusts the orientation parameters of 2D images to better match the computed 2D projection of the current 3D map. The fitting progress is monitored by the Fourier Shell Correlation (FSC) curve, which provides information on the resolution of the map.

2.1.5 Model Building and Validation

Model building for high-resolution EM maps can be done in the same way as for X-ray crystallographic maps. De novo backbone tracing and subsequent side chain fitting can be performed. EM maps contain both amplitude and phase information. The atomic form factors in cryo-EM maps describing the potential distribution of atoms provide information about the charge states of the specimen, which are not detected in X-ray maps. Also, the phases of the Fourier transform of the specimen that are readily available in cryo-EM experiments need to be additionally acquired in X-ray experiments. Thus, it is possible to construct reliable atomic models even when it is difficult to do so with X-ray crystallographic map at comparable resolution.

The calculation of "gold-standard" FSCs [192] is one standard way of evaluating a cryo-EM map. Reconstructions are obtained from two independent halves of the dataset, and refined separately, so the noise in the half maps used to calculate the FSCs are free from spurious correlations. Comparison between the half FSCs can increase confidence in the overall quality of the EM map. However, it fails to provide information on localised quality of the map, so it has become increasingly important to perform local resolution analysis to assess the quality of different regions of the map. The correlation between the experimental 3D reconstruction and the atomic model can also be determined using the map-to-model FSCs. A small semi-randomised deviation is usually introduced to the model built from the

full cryo-EM map, and then the model is refined against one of the half maps (the working half map). The fit of the model to the remaining half map (the free half map) is calculated using FSC. The level of agreement between the map-to-model FSCs for the half maps and the full map can be used as a measure of model validity [193].

2.2 Molecular Dynamics Simulations

Structural biology has been continuously advancing our understanding of macromolecular structures, which provides basis for studying the mechanisms and regulations of biological processes. However, as static, time- and space- averaged three-dimensional reconstructions of biomolecules, structures fail to provide dynamic information, which are crucial for functionality. Based on the framework made available by structural biology, the exploration of structural dynamics and microscopic interactions can be done with molecular simulations. They bridge the cap between the microscopic properties of molecules such as atomic interactions and velocities and the experimentally measurable macroscopic properties including pressure, diffusion constant and heat capacity. In addition, simulations could provide information of a system of interest under extreme conditions that may be otherwise unachievable in experiments. Molecular dynamics (MD) and Monte Carlo (MC) are two main simulation techniques. Given the stochastic nature of MC simulations, one of the key advantages of MD simulations over MC is that it produces time-evolution trajectories of atomic motions, and thus dynamic properties such as time-dependent interactions and transport efficiency can be addressed [194].

The generation of successive configurations for a system of N atoms in MD simulations are based on Newton's second law:

$$\boldsymbol{F}_i = m_i \boldsymbol{a}_i = \frac{d^2 \boldsymbol{r}_i}{dt^2}, \forall i = 1...N \tag{2.1}$$

where the instantaneous force $\boldsymbol{F}_i$ acting on a particle of mass m_i at position $\boldsymbol{r}_i$ produces an acceleration $\boldsymbol{a}_i$ and thus move the atom in space [195]. The force

for each atom i can be derived from a potential energy function $V(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_N)$, which depends on the complete set of coordinates of all particles in the system:

$$\boldsymbol{F}_i = -\frac{dV}{d\boldsymbol{r}_i} \tag{2.2}$$

This potential energy describes both non-bonded interactions where the electrostatic field from molecule's protons and electrons provides the intermolecular forces and bonded interactions where intramolecular bonding links the motion of associated atoms:

$$V(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_N) = V_{intramolecular}(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_N) + V_{non-bonded}(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_N) \tag{2.3}$$

2.2.1 Force Fields

In order to determine the time evolution of the forces acting on each atom within a system, an accurate description of the potential energy is essential for MD simulations. A force field is a mathematical expression used for the calculation of forces that propagate the system with time. It consists of a differentiable function of the atomic coordinates describing the energy and a set of parameters describing the associated geometric and energetic properties [196]. In force field methods, the electrons are considered to adapt their motions instantly to the changes to nuclear positions, and remain in the ground state. Given the Born–Oppenheimer approximation that allows the motion of atomic nuclei and electrons to be treated separately, the energy of a system is therefore a function of the positions of nuclei only. This assumption renders force field methods incapable of describing properties involving the electronic distribution of a molecule such as electron transfer processes (i.e. chemical bond breaking or formation) and transition states with excited electrons [197].

Force fields as simplified models of the true potential need to be able to reproduce the properties of interest in sufficient details at acceptable cost of efficiency. Parameterisation of force fields is empirical, i.e. they are based on experimental data such as NMR, neutron diffraction and spectroscopy, X-ray crystallographic

data or quantum mechanical calculations [198]. Each force field is generally purposed to reproduce the potential energy surface of a certain type of system. Despite the differences in their oriented systems and parameterisations, they are typically interpreted in a simple functional form:

$$V(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_N) = \sum_{bonds} \frac{k_d}{2}(d - d_0)^2 + \sum_{angles} \frac{k_\theta}{2}(\theta - \theta_0)^2 + \sum_{dihedrals} \frac{k_\phi}{2}(1 + cos(n\phi - \phi_0)) + \sum_{improper} \frac{k_\psi}{2}(\psi - \psi_0)^2 + \sum_{LJ} 4\varepsilon_{ij}[(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^6] + \sum_{elec} \frac{q_i q_j}{\varepsilon_D r_{ij}} \quad (2.4)$$

The contributing terms in Equation 2.4 are schematically represented in Figure 2.4.

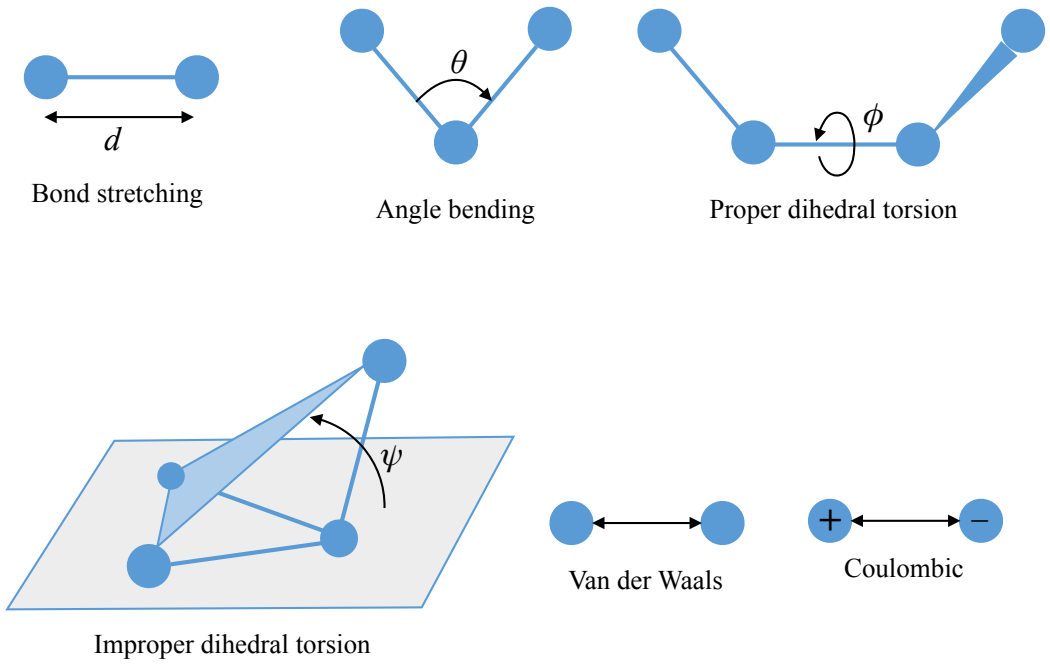


Figure 2.4: Schematic representation of the contributions to a force field.

2.2.1.1 Intramolecular Interactions

The first term in Equation 2.4 controls the bond length between two covalently bonded atoms. This is often modelled by a harmonic potential with a force constant

k_d derived from Hooke's law, which increases the energy as the distance between the bonded atoms (d) deviates from the reference value (d_0). This approximation does not allow bond breakage, which limits the type of processes that force field based simulations can study. The bending of a valence angle is also approximated with a harmonic oscillator (second term in Equation 2.4). As with the bond stretching term, the potential energy change resulted from angle distortion is defined by a spring force constant (k_θ) and a reference angle (θ_0). The proper dihedral as a degree of freedom requires substantially less energy than bond stretching and angle bending, and hence rotation about chemical bonds is one of the major contributors to the structural properties and conformational changes of molecules. The torsional potentials are usually expressed as cosine function such as the third term in Equation 2.4, where ϕ is the torsion angle, ϕ_0 is the phase factor that defines the torsion angle corresponding to energy minima, and n is the multiplicity that determines the number of minima in the energy function between 0 and 2π. The planarity or chirality of certain groups are preserved using an improper torsional term in the potential energy function, since the proper torsional term is not sufficient to describe out-of-plane bending motions. This term penalises deviation from planar structures for groups like amide bond or aromatic rings [195].

2.2.1.2 Non-bonded Interactions

Van der Waals and electrostatic interactions between atoms separated by at least three bonds or in different molecules are described by the non-bonded terms. Van der Waals interactions arises from charge fluctuations of nearby particles. They are highly distance dependent. The dispersion forces from instantaneous dipole result in attraction between a pair of particles (London dispersion force), while they repel each other, when the particles' outer electron clouds come too close (Pauli exclusion principle) [195]. Van der Waals interactions are often modelled with the 12-6 Lennard-Jones (LJ) potential (the fifth term in Equation 2.4). The r_{ij}^{-12} term describes the rapid increase in energy when two particles come closer and the r_{ij}^{-6} term represents the favourable interaction energy at long range. The depth

of the potential well between two particles is ε_{ij} and the distance at which the interaction energy is zero is defined by σ_{ij} (Figure 2.5). Since LJ potential decays relatively quickly with distance and becomes infinitely small at long distance, for computational efficiency, a cut-off distance is usually applied to reduce the number of calculations so that interaction energy between atoms that are apart from each other beyond the cut-off is treated as zero.

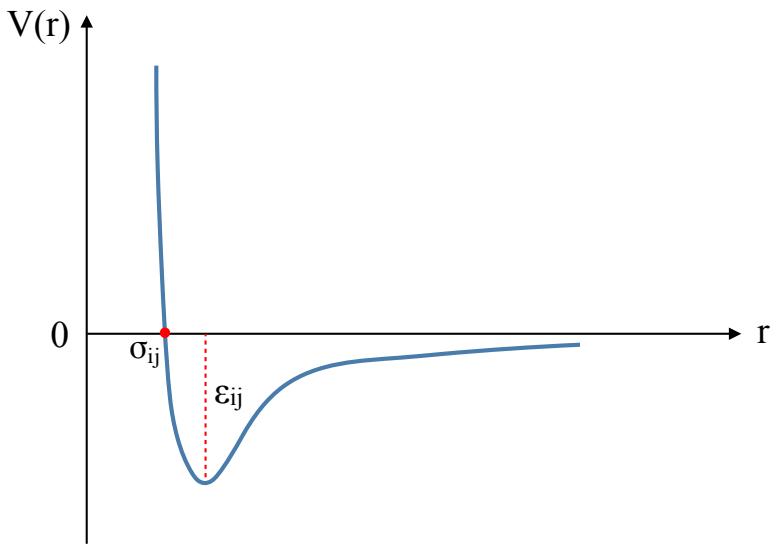


Figure 2.5: The Lennard-Jones 6-12 potential (V) as a function of the particle distance (r).

The electrostatic properties of molecules are usually represented by the last term of Equation 2.4. Instead of calculating the electrostatic potential from the wave function as in high-level quantum-mechanical calculations, MD simulations reduce the complexity by assigning a partial atomic charge to each nucleus. Their contributions to the total electrostatic energy can then be evaluated using Coulomb' s law. The decay of electrostatic potential is less sensitive to distance, and thus the Coulomb interaction is generally used without a cut-off as there would be an abrupt change in the energy at the cut-off [197]. However, calculating all long-range interactions between each pair of atoms is computationally demanding. The most

commonly used method to overcome this is Particle-mesh Ewald (PME), where interactions are separated into short- and long-range potentials. The former is calculated in real space and the latter in reciprocal space using Fourier transforms, which allows substantial speed up.

2.2.1.3 Force Field Parameterisation

Typically, there is a large number of parameters involved in a force field. Parameterisation of a force field is therefore a complicated procedure. The data used to guide the parameterisation process is usually selected based on the intended purpose of a particular force field, which include geometries and the associated conformational energies of a set of key molecules representing certain category of structures of interest. Although experimental data such as vibrational spectra and solvation free energies is commonly used in target datasets, some properties like the torsional energy barriers have been difficult to obtain experimentally [196]. Quantum mechanical (QM) calculations are thus employed to determine these values. QM data has also allowed force fields to be developed to treat classes of molecules that are difficult to handle experimentally.

Due to the cross correlation between all the degrees of freedom and the large number of parameters involved, it is virtually impossible to treat any of the parameters in isolation, especially for the closely coupled torsional and non-bonded contributions. One approach of parameterisation involves iterative refinement of the parameters so that the targeted dataset can be reasonably reproduced by the force field. Another approach makes used of least square method that aims to obtain optimal set of parameters by minimising the sum of squares of the differences between the force-filed based calculations and the observed values [195]. For most of the available force fields to date, the relatively hard degrees of freedom (i.e. bond stretching and angle bending) are treated with high level of similarity. Differences mainly exist in the development of torsional and non-bonded terms. Force fields are often targeted at different subsets of all possible experimental and QM data.

It remains a challenge to develop transferable models, which can describe the majority of chemical space.

2.2.2 Simulation Methods

2.2.2.1 Periodic Boundary Conditions

In most of simulations, the number of particles in a system of interest is relatively small comparing to real bulk. The treatment of boundaries is therefore crucial when evaluating macroscopic properties. If boundaries are treated as rigid walls in a system of N particles, the percentage of particles that are affected by the interactions with the walls and the surface effect is proportional to $N^{-1/3}$. This would render a system of typical MD simulation not appropriate for the study of bulk properties. Using periodic boundary conditions is one way to minimise the boundary effects in a finite system. That is, the main simulation box is replicated infinite times in all dimensions around itself (Figure 2.6) such that once a particle diffuses across the boundary, it is replaced by its periodic image that enters from the opposite side. Only atoms in the main simulation box are considered explicitly, which enable bulk physics to be studied using a simulation with limited number of particles.

Due to the periodicity of the simulated system, care must be taken in choosing the simulation box size. Fluctuations that have longer wavelength than the length of the box can not be achieved, and hence periodic boundary conditions are not appropriate for situations such as phase transition where long wavelength fluctuations are important. In addition, box size should be relatively large comparing with the range of interactions present in the system. Particles in a system should not interact with its own image. MD simulations generally use periodic boundary conditions with the minimum image convention; that is, only one image of each particle (the nearest one) are considered for interactions.

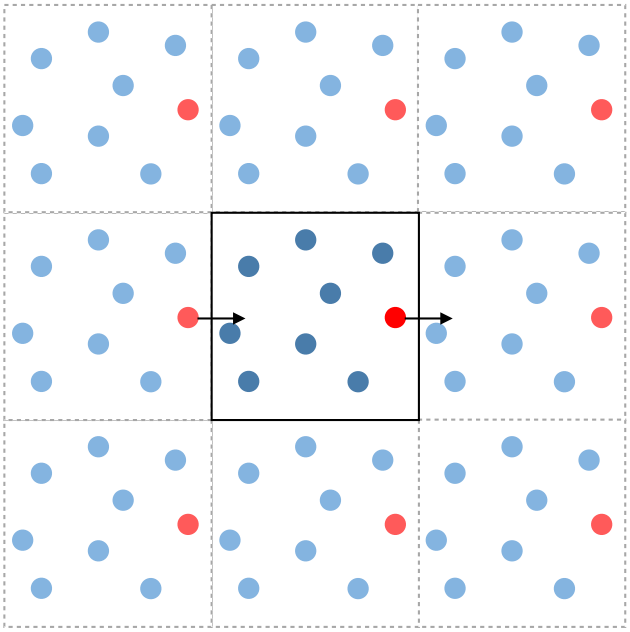


Figure 2.6: Two dimensional representation of periodic boundary conditions. The central simulation box is replicated in all dimensions. All particles are allowed to diffuse across the periodic boundaries. As illustrated by the red dot, as soon as one particle leaves the box, an image particle enters from the opposite side to replace it.

2.2.2.2 Integration algorithms

As shown in Equation 2.1, to update the coordinates of a particle with a given potential energy function, a system of N coupled second order non linear differential equations needs to be solved. However, the complex dependency makes it impossible to be solved exactly for all particles. Therefore, a finite difference method is used to numerically solve the Newton's equations of motion. This involves discretise the trajectory and update the positions of atoms over small time steps, Δt:

$$\boldsymbol{r}_i(t_0) \rightarrow \boldsymbol{r}_i(t_0 + \Delta t) \rightarrow \boldsymbol{r}_i(t_0 + 2\Delta t) \rightarrow \cdots \boldsymbol{r}_i(t_0 + n\Delta t) \tag{2.5}$$

The most widely used integration method in GROMACS (one of the most

commonly used MD simulation package) is the leap-frog algorithm [199], which uses two dynamic properties at two different time points, i.e. the positions $\boldsymbol{r}_i$ at time t and velocities $\boldsymbol{v}_i$ at time $t - \frac{1}{2}\Delta t$. Using the forces $\boldsymbol{F}_i(t)$ determined by the positions at time t, the velocities and positions of particles can be calculated by:

$$\boldsymbol{v}_i(t + \frac{1}{2}\Delta t) = \boldsymbol{v}_i(t - \frac{1}{2}\Delta t) + \frac{\Delta t}{m}\boldsymbol{F}_i(t) \tag{2.6}$$

$$\boldsymbol{r}_i(t + \Delta t) = \boldsymbol{r}_i(t) + \boldsymbol{v}_i(t + \frac{1}{2}\Delta t)\Delta t \tag{2.7}$$

As the propagation of the system relies on velocities from the prior step, a MD simulation needs to be initiated with random starting velocities, which are generally obtained from a Maxwell-Boltzmann distribution at a pre-defined temperature. In addition, an appropriate time step needs to be chosen in a balance of the accuracy of the numerical solution and the computational cost of obtaining certain length of simulation. The size of the time step should be able to capture the fastest intramolecular vibrational motions in a system, for example, the C-H stretches in a biomolecular system. With these high frequency motions, the computational load is relatively large. One way to increase the efficiency of simulation is to apply constraints on the bond lengths so that larger time step could be used. This is typically achieved by using constraint algorithms such LINCS [200] and SHAKE [201]. The LINCS algorithm is more stable and faster than SHAKE, and is used throughout this thesis. It is implemented to reset the covalent bonds to their equilibrium lengths after each integration step.

2.2.2.3 Thermodynamic Ensembles

Numerical integration of Equation 2.1 implies that the number of particles (N), the volume (V) , and the total energy of the system (E) are constant over the course of simulation, which is referred to microcanonical or NVE ensemble. However, biological processes generally occur under constant temperature (T) and/or constant pressure (P). The two most commonly used alternative ensembles are the constant NPT (isotheomal-isobaric) and constant NVT (canonical).

Constant Temperature Dynamics

The temperature of a system is directly associated with the time average of the kinetic energy. For a system without constraints:

$$E_{kinetic} = \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2 = \frac{3}{2} N k_B T \tag{2.8}$$

where m_i and v_i are the mass and velocity of particle i, N is the number of particles in the system, k_B is the Boltzmann constant and T is the instantaneous temperature of the system. If the velocities of all particles are scaled by a factor of λ at time t, the change in temperature would be:

$$\begin{aligned} \Delta T &= \frac{1}{2}\sum_{i=1}^{N} \frac{2}{3}\frac{m_i(\lambda v_i)^2}{N k_B} - \frac{1}{2}\sum_{i=1}^{N} \frac{2}{3}\frac{m_i v_i^2}{N k_B} \\ &= (\lambda^2 - 1)T(t) \end{aligned} \tag{2.9}$$

$$\lambda = \sqrt{T_{new}/T(t)} \tag{2.10}$$

Thus, to achieve a desired temperature of T_{req}, the simplest way is to scale the velocities at each time step t by the factor $\lambda = \sqrt{T_{req}/T(t)}$. However, such scaling can disturb the dynamics of the system undesirably and it does not generate rigorous canonical averages. The more common way of controlling temperature in molecular simulations is the use of thermostat algorithms. One of the widely used thermostats is Berendsen, which couple the system to an external heat bath that supplies or removes thermal energy from the system as appropriate [202]. The velocities of particles are scaled to make the rate of change in temperature proportional to the temperature difference between the system and the bath, so the change in temperature (ΔT) between successive time steps (Δt) and the scaling factor (λ) can be calculated as follow:

$$\frac{dT(t)}{dt} = \frac{1}{\tau_T}(T_{bath} - T(t)) \tag{2.11}$$

$$\Delta T = \frac{\Delta t}{\tau_T}(T_{bath} - T(t)) \tag{2.12}$$

$$\lambda^2 = 1 + \frac{\Delta t}{\tau_T}(\frac{T_{bath}}{T(t)} - 1) \tag{2.13}$$

The strength of temperature coupling between the system and the bath is inversely related to the coupling constant, τ_T. The NVE ensemble is obtained when the coupling is infinitely weak, i.e. $\tau_T \to \infty$. When τ_T is equal to the time step, this method becomes equivalent to the simple velocity scaling method. Similar to velocity scaling, correct NVT ensemble cannot be rigorously sampled with Berendsen thermostat. An extension of this method is the V-rescale thermostat, where an additional stochastic term is included to ensures the generation of a proper canonical ensemble [203].

Constant Pressure Dynamics

Experimental measurements are usually conducted under constant temperature and pressure. To achieve better comparability of the simulation results to experimental data, one may prefer to perform simulations in the NPT ensemble. The constant pressure can be maintained by scaling the volume of the simulation box. Barostat algorithms that control pressure operate in a similar manner as the thermostats. For Berendsen barostat [204], the system of interest is coupled to an external pressure bath to scale the volume at each time step so that the rate of change of pressure is proportional to the difference in pressure between the system and the bath:

$$\frac{dP(t)}{dt} = \frac{1}{\tau_P}(P_{bath} - P(t)) \tag{2.14}$$

where τ_P is the coupling constant, $P(t)$ is the system pressure at time t and P_{bath} is the pressure of the pressure bath. Base on the isothermal compressibility (κ) of the system, the scaling factor for the volume of the simulation box can be calculated as:

$$\lambda = 1 - \kappa\frac{\Delta t}{\tau_P}(P(t) - P_{bath}) \tag{2.15}$$

It has been argued that Berendsen barostat does not generate a rigorous thermodynamic ensemble. However, it is highly efficient in scaling the simulation box. The Parrinello-Rahman barostat [205] is an extended-ensemble pressure coupling, where each unit vector of the simulation cell is treated independently so that

not only the volume but also the shape of the simulation cell can be changed dynamically. This allows control over both stress and pressure, and is capable to sample a more rigorous NPT ensemble. For simulations of systems with lipid bilayer, it is preferred to apply pressure coupling semi-isotropically, i.e. the same scaling factor is used in the x and y direction (the membrane plane), but different in the z direction. This ensures a proper lateral pressure in the bilayer, and thus a more stable bilayer structure.

2.2.3 Coarse-grained Simulations

For biomolecular simulations, the achievable time- and length-scales are still the main limiting factors in studying the properties of complex systems or processes that occur on long time-scales. The special and temporal resolutions of experimental techniques in studying biological processes remain beyond tens of microseconds and nanometers respectively, which are practically inaccessible using standard equilibrium atomistic simulations. Thus, coarse graining that reduces the number of degrees of freedom by approximating a group of atoms with a single particle has been used to bridge the gap between computational and experimental studies. The reduction in the number of particles and elimination of fine-detailed interactions in a coarse-grained (CG) representation allow large-scale systems and long-scale events to be simulated. Depending on the intended molecules, CG force fields with various levels of granularity have been developed. United atom models have been traditionally categorised as "atomistic" because all heavy atoms are explicitly represented. However, these models joins non-polar hydrogens with their covalently bound heavy atoms, such that the highest-frequency interactions and the number of atoms are largely reduced at the expense of some atomic details, so they can already be considered as mild type of CG models [198]. The more commonly used level of coarse graining involves grouping several heavy atoms into a single interaction site. Different approaches have been developed with various CG mapping of a system. They have been successfully applied to study a range of

processes including protein folding [206, 207], DNA denaturation [208] and protein conformational changes [209].

The Martini force field used in this thesis was initially developed for CG lipids [210], and then extended to proteins [211], nucleic acids [212, 213], carbohydrates [214] and organic polymers [215]. This force field has been parameterised to allow a system to be constructed with a mixture of these described building blocks without the need of re-parameterisation. It employs a mapping scheme that on average, one CG interaction site represents four heavy atoms, and all associated hydrogen atoms are neglected. There are four main types of CG particles in Martini force field representing different chemical natures of the underlying atomic structures:

- Polar (P) particles representing neutral hydrophilic groups of atoms;
- Nonpolar (N) particles representing mixed groups which have partial polar and partial apolar natures;
- Apolar (C) particles representing hydrophobic groups of atoms;
- Charged (Q) particles representing ionised groups

For particles of type C and P, five subtypes are further distinguished based on the degree of polarity ranging from low polarity (1) to high polarity (5). For N and Q particles, four subtypes are defined on the basis of hydrogen-bonding capabilities (d = donor, a = acceptor, da = both, 0 = none). This allows the interaction sites to more accurately reflect the chemical natures of the atoms. Higher resolution mapping (S-size particles; 3:1 or 2:1 mapping schemes) has been developed to support the representation of ring-like structures. Examples of Martini mapping between the chemical structure and the CG particles are shown in Figure 2.7. For proteins, all backbone atoms are represented by a single polar particle with polarity assigned depending on the local secondary structure. For water molecules, four atomistic waters are mapped to a single P4 particle in the standard Martini model.

Due to the lack of explicit hydrogen bonds in the CG model, higher-order structures of proteins need to be additionally maintained. The secondary structures

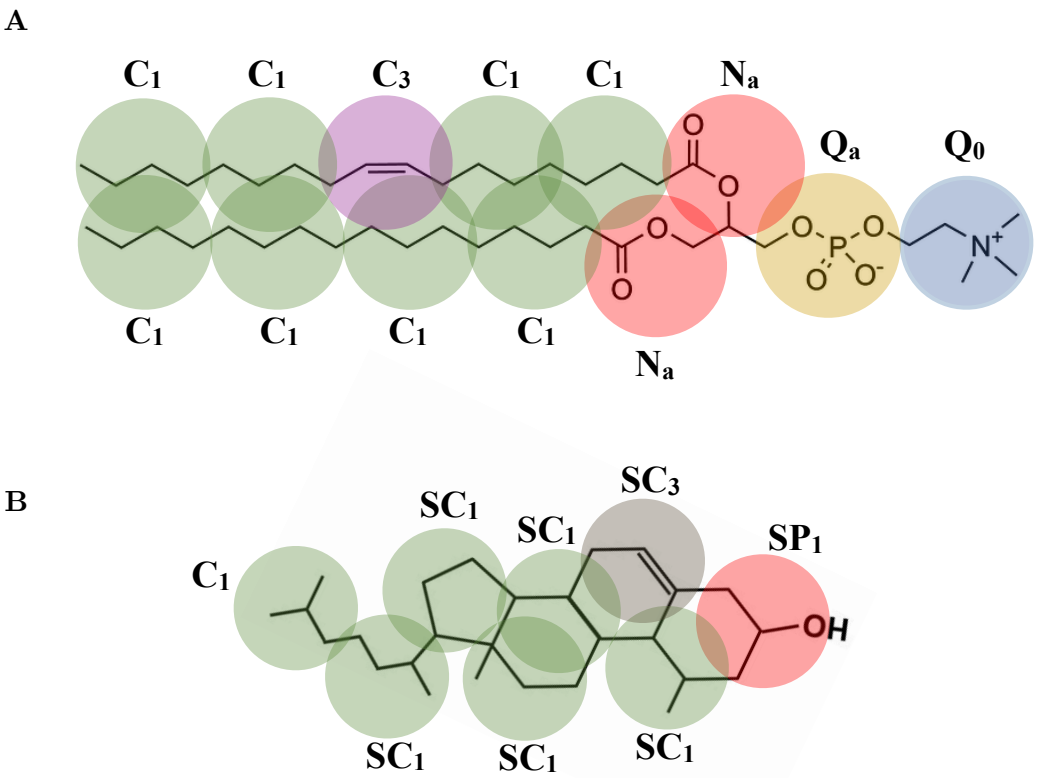


Figure 2.7: Martini mapping of the chemical strusctures of POPC (**A**) and cholesterol (**B**) onto CG particles.

is assigned to constituent particles using DSSP [216] at the beginning of coarse graining. Approximations were made in this step comparing to strict secondary structure classifications. All helical structures including 3_{10}-helix and π-helix are modelled as α-helical representation, and all β-sheets are approximated as extended. Harmonic potentials are applied on the bonds, angles and dihedrals involving only backbone particles to maintain the prescribed secondary structures [211]. The tertiary structure is preserved in Martini topology through two common approaches: the simple elastic network and the ElNeDyn network [217]. An example is shown in Figure 2.8. Both methods apply extra harmonic potentials between all backbone particles within a cut-off distance, which creates a network of long harmonic bonds to conserved the conformation while preserving the local internal dynamics. The simple elastic network is used in combination with the

standard Martini topology. However, quite some changes have been made to the Martini force field in parameterising the ElNeDyn network, and thus it is regarded as different protein force field. For example, the main difference between the two methods is that the backbone particles are placed on the C_α atoms instead of the centre of mass of the backbone atoms in the ElNeDyn network, and therefore, the bonded parameters have been modified to accommodate the change in structural mapping. It is noted that the simplifications made to maintain the overall structural scaffold render the Martini model incapable to fully capture conformational changes in proteins, but it has been proven to be very useful in studying global interactions.

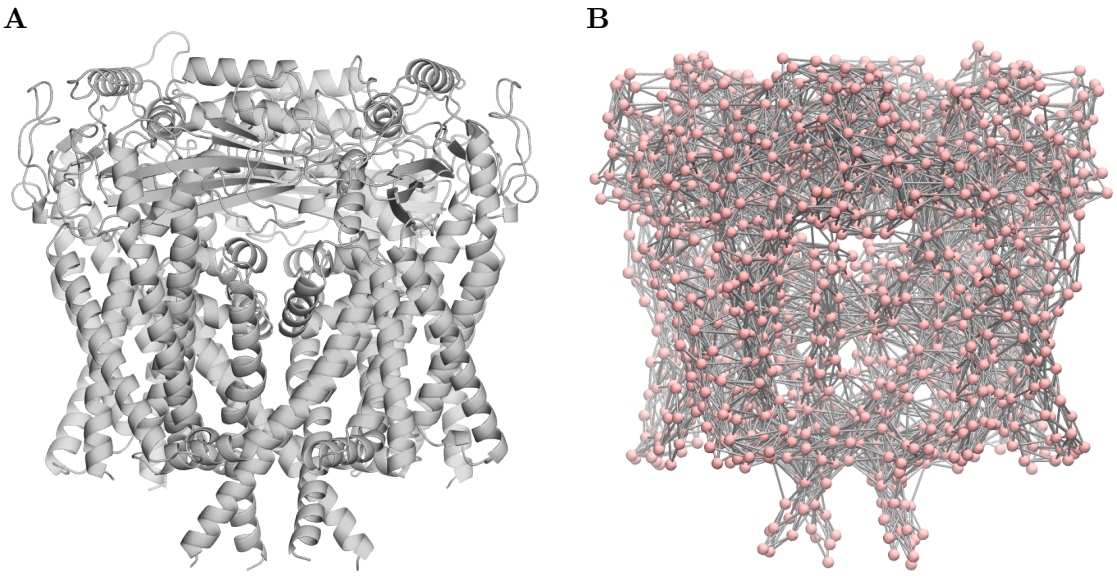


Figure 2.8: **A**: Atomic structure of PC2 with secondary structures shown. **B**: ElNeDyn network representation (gray) of PC2 with CG backbone particles shown as small spheres (pink).

2.2.4 Free Energy Calculations

The probability for a system to reach a particular state is described by the free energy landscape. For biological systems, there usually are many minimum energy configurations separated by energy barriers. In unbiased MD simulations, the finite

length and time scale make it nearly impossible to accurately calculate the absolute free energy due to insufficient sampling. The high barriers between local free energy minima often circumscribe a system into one particular low-energy state and slow down the sampling process in conventional simulations. Thus, enhanced sampling methods have been developed to aid the process by introducing an external bias to drive the system towards certain states of interest.

2.2.4.1 Umbrella Sampling

Umbrella sampling is one of the enhanced sampling techniques that endeavours to overcome the sampling issue by modifying the potential function along reaction coordinates so that the configurations corresponding to "rare" states can be favoured in a controlled manner. Thus, the free energy difference between distinct states can be calculated. In general, multiple windows are simulated with starting configurations set at different reaction coordinate values. For example, if the dissociation of a bound ligand is the chemical process of interest, the distance between the protein and the ligand can be used as reaction coordinate, and simulations can be run with initial structures with different distances. If the potential function at a particular reaction coordinate s is $V(s)$, then the modified potential function $V'(s)$ can be written as:

$$V'(s) = V(s) + U(s) \tag{2.16}$$

where $U(s)$ is a harmonic restraint added on the reaction coordinate s, which often takes the form:

$$U(s) = \frac{k}{2}(s - s_0)^2 \tag{2.17}$$

where k is the spring constant. When the reaction coordinate is far from the equilibrium state s_0, larger bias potential would be applied to push the system towards the reaction coordinate.

2.2.4.2 Potential of Mean Force

Potential of Mean Force (PMF) is the free energy surface along a reaction pathway, which is defined by collective variables (i.e. reaction coordinates) that describe the chemical processes of interest. In the presence of a solvent, the averaged solvent effects and the intrinsic interactions between the particles defining the collective variables are included in the free energy change describe by PMF [195]. To obtain a PMF profile from umbrella sampling, the biased free energy of each window needs to be unbiased and combined to produce a free energy profile along the reaction pathway. This is done using the weighted histogram analysis method (WHAM) [218]. Conceptually, the biased probability distribution of a reaction coordinate $P'(s)$ is known from the simulation of each window, from which the biased free energy of each window can be calculated. The unbiased free energy can then be induced given the umbrella potential applied to that window. Assuming an adequate overlap of the probability distributions of adjacent window, the free energies can be combined to give the free energy profile along the reaction coordinates. Therefore, the choices of window spacing and biasing potential need to be carefully made to ensure sufficient sampling of the reaction pathway, which is central to the calculation of PMF.

3

Experimental Studies of Lipid Interactions of Polycystin-2

Contents

3.1 Introduction 57
3.2 Methods 58
3.2.1 Protein Expression and Purification 58
3.2.2 PIP Strip Assay 60
3.2.3 Label-free Differential Scanning Fluorimetry (nanoDSF) Measurements 60
3.2.4 Cryo-EM Grid Preparation and Data Acquisition 62
3.2.5 Image Processing 63
3.2.6 Model Building 63
3.3 Results and Discussion 64
3.3.1 Characterisation of PC2 Lipid Binding Using nanoDSF and Biochemical Assays 64
3.3.2 Structural Characterisation of PC2 in the Presence of PIP Lipids by Cryo-EM 68
3.3.2.1 Data Processing and Model Building for Sample with $PI(4,5)P_2$ 68
3.3.2.2 Unidentified Density in C1 Map of $PC2/PI(4,5)P_2$ 75
3.3.2.3 Data Processing and Model Building for Sample with $PI(3,5)P_2$ 77
3.3.2.4 A Possible Calcium Ion in the Selectivity Filter 80
3.3.2.5 Lipid Interactions Revealed by Cryo-EM . . . 82
3.4 Conclusions 85

This chapter is partially based upon the manuscript:

"Lipid Interactions of a Ciliary Membrane TRP Channel: Simulation and Structural Studies of Polycystin-2 (PC2)", Qinrui Wang, Robin A. Corey, George Hedger, Prafulla Aryal, Mariana Grieben, Chady Nasrallah, Agnese Baronina, Ashley Pike, Jiye Shi, Elisabeth P. Carpenter, and Mark S.P. Sansom, submitted to Structure.

3.1 Introduction

The TRP superfamily of channels exhibits a wide range of activation mechanisms and cation selectivity, despite the common structural features [20]. The systemic distribution and the large number of biological processes they involved in highlight the implications of TRP channels in different diseases, and thus continuing effort has been devoted in studying the structure and functional dynamics of these channels [20, 21]. Many TRP channels are modulated by membrane lipids, suggesting the possibility of lipid based therapies [219]. One class of the most studied endogenous modulators of TRP channels are the phosphoinositide lipids, especially phosphatidylinositol bisphosphates (PIP_2). PIP_2 regulation has been suggested for almost all subfamilies of TRP channels, despite their possibly diverse activation mechanisms in response to different stimuli [2, 220–222]. Positive regulation by PIP_2 has been indicated in at least five ion channels of the TRPM subfamily [223–227]. For TRPV channels, activation by PIP_2 in excised patches has been reported for TRPV1 [228], TRPV5 [229] and TRPV6 [230]. Indirect inhibition of TRPV1 by PIP_2 in intact cells has also been reported [230]. Given the importance of PIP_2 distribution and dynamics in the membranes of primary cilia [231], we decided to explore possible interactions of this anionic lipid with PC2.

In the past few years, cryo-EM has been used to solve the structures of TRPA [11], TRPC [232–234], TRPM [160, 235, 236], TRPML [77–79], TRPN [9], TRPV

[6, 8, 53, 237, 238] and TRPP [7, 13–15, 239] channels in the TRP family. A number of cryo-EM studies have co-imaged proteins with ligands including lipids, which provide valuable structural information on protein-ligand interactions and the underlying regulatory mechanisms. Recent cryo-EM structures have revealed PIP_2 binding sites on TRPV5 [53], TRPML1 [78] and TRPM8 [160]. A complete list of TRP channel structures with or without lipid/ligand bound is shown in Appendix A, Table A.1. The advances in cryo-EM offer great opportunities to structurally study lipid binding to membrane proteins. For PC2, although previous structures (PDB ID: 5MKE/5MKF) provided some indication of lipid interactions, with the resolution of 4.3Å/4.2Å respectively, it is difficult to confidently interpret the protein-lipid interactions in great detail. In this study, biochemical, biophysical and structural experiments were performed in attempt to extend the resolution of previously published structures of PC2 and to explore the complex roles of lipids in the interaction and regulation of PC2.

3.2 Methods

3.2.1 Protein Expression and Purification

The human *PKD2* gene, which encodes the PC2 (polycystin-2, PC2 or TRPP1) protein was purchased from the Mammalian Gene Collection (MGC, 138466; IMAGE, 8327731, BC112263). A truncated stable construct (hPC2: Pro185 - Asp723) with a C-terminal purification tag containing a tobacco etch virus (TEV) cleavage site, a His_{10} sequence and a FLAG tag was cloned into the expression vector pFB-CT10HF-LIC for insect cell expression and into pHTBV1.1-CT10H-SIII-LIC for mammalian cell expression (available from The Addgene Nonprofit Plasmid Repository). DH10Bac competent cells were used for the production of baculovirus. Recombinant baculoviruses were used to infect *Spodoptera frugiperda* (Sf9) insect cells for virus amplification when cell density reached $\sim$ 2×10^6/ml, grown in suspension in 250 ml of Sf900II serum free, protein-free insect cell medium with L-glutamine (Thermo Fisher Scientific) in 1 L shaker flasks at 27 ℃. 1 L of Sf9 insect

cells in Insect-XPRESS Protein-free Insect Cell Medium with L-glutamine (Lonza) in a 3 L flask was infected with 5 ml of the harvested P2 (second passage) viruses for 65 h at 27 ℃ with 5% CO_2. HEK293S GnTI$^-$ cells at density of $\sim 2 \times 10^6$/ml were infected with 30 ml of P3 (third passage) recombinant baculovirus in Freestyle 293 Expression Medium, and grown for 65 h at 37 ℃ with 5% CO_2. To boost protein expression, sodium butyrate was added to the cell culture to a final concentration of 5 mM at the time of transfection of HEK293S GnTI$^-$ cells. Cells were harvested 65 h post-transduction by centrifugation for 15 min at 1,500 g and 4 ℃.

Extraction buffer containing 50 mM HEPES, pH 7.5, 150 mM NaCl, 20 mM $CaCl_2$, 5% glycerol and Roche protease-inhibitor cocktail was used to re-suspend the cell culture pellets to a volume of 50 ml/L. Cells were lysed on ice with VCX 750 sonicator and 13 mm probe (PRO Scientific Inc.) for 5 min, 3 sec on, 12 sec off and 35% amplitude. 1% (w/v) n-Dodecyl-β-D-maltoside (DDM) was added to the cell lysis and incubated for 1 h at 4 ℃. Cell debris was removed by centrifugation for 1 h at 35,000 g and 4 ℃. To purify by immobilised metal affinity chromatography, the detergent-solubilised protein were batch bound to Co^{2+}-charged Talon resin (Clontech) by gentle rotation at 4 ℃ for 1 h. The resin was washed with 15 column volumes of extraction buffer supplemented with 0.01% DDM and 30 mM imidazole, pH 8.0 and, to exchange the detergent, with another 15 column volumes of the same buffer replacing DDM with 0.035% n-undecyl-β-D-maltopyranoside (UDM). hPC2:185-723 protein was eluted from Talon resin with extraction buffer supplemented with 0.035% UDM and 400 mM imidazole. The eluted protein was divided into two batches. One was further purified via size-exclusion chromatography (SEC) with a Superose 6 increase 10/300GL column (GE Healthcare) pre-equilibrated with SEC buffer (0.035% UDM, 20 mM HEPES, pH 7.5, 200 mM NaCl and 20 mM $CaCl_2$) for PIP strip experiments. The other was buffer exchanged into SEC buffer using a PD-10 column (GE Healthcare). The protein was then treated with 0.4 units of bovine kidney α-L-fucosidase (Sigma) at 18 ℃, overnight. The pH was adjusted to 7.0 before further enzymatic treatment at a ratio of 0.75:0.5:1 (w/w/w) TEV protease, PNGase F, PC2 for another 24 h

at 18 ℃. Reverse His-tag purification was performed to clear out the His_6-tagged TEV protease and uncleaved PC2. The protein was concentrated to 0.5 ml with a 100-kDa-cutoff concentrator (Vivaspin 20, Sartorius), and further purified by SEC as above. The same purification protocol was used for protein expressed in HEK293S GnTI$^-$ cells apart from the removal of glycans.

3.2.2 PIP Strip Assay

PIP strip membranes (Thermo Fisher Scientific P23751) were blocked in 3% (w/v) fatty acid-free BSA (Sigma-Aldrich) in TBST (50 mM Tris-HCl, pH 7.5, 150 mM NaCl and 0.1% (v/v) Tween 20) for 1 h. The membranes were then incubated in the same solution with 0.5 μg/ml of detergent solubilised His_{10}/FLAG-tagged PC2 overnight at 4 ℃ with gentle agitation. The membranes were washed 3 times over 30 min in fatty acid-free BSA-TBST. One more 3 h-incubation with 0.5 μg/ml PC2 and subsequent wash steps were performed at 4 ℃. The membranes were incubated for 1 h with 1:2000 dilution of HRP conjugated anti-FLAG monoclonal antibody (Thermo Fisher Scientific MA1-91878-HRP) at room temperature. Finally, the membranes were washed 6 times over 1 h in TBST, and the protein that was bound to the membrane by virtue of its interaction with phospholipid was detected by enhanced chemiluminescence.

3.2.3 Label-free Differential Scanning Fluorimetry (nanoDSF) Measurements

Low volume differential scanning fluorimetry (nanoDSF) that measures thermal unfolding of proteins by monitoring the intrinsic tryptophan fluorescence at 330 nm and 350 nm was applied to study the possible stabilising effects of PIP_2 on PC2. 18:0-20:4 PI(4,5)P_2, 08:0 PI(4,5)P_2 and 18:1 PI(3,5)P_2 lipid extract (Avanti) (Figure 3.1) dissolved in chloroform was dried under an argon stream. 18:0-20:4 PI(4,5)P_2 ($\sim$ 0.5 mg/ml), 08:0 PI(4,5)P_2 ($\sim$ 0.5 mg/ml) and 18:1 PI(3,5)P_2 ($\sim$ 1 mg/ml) stock was prepared by resolubilising dried lipids in a buffer containing 20 mM HEPES (pH 7.5), and 150 mM NaCl via bath sonication for $\sim$ 1 h.

Before incubating with purified protein, UDM was added to the lipid stocks to a final concentration of 0.035% to prevent protein precipitation after detergent dilution below critical micelle concentration. For $PI(4,5)P_2$, both His_{10}/FLAG-tagged and tag-cleaved PC2 were tested. For $PI(3,5)P_2$, a different biological sample of untagged PC2 were used. Tested protein monomer to lipid molar ratios are listed below (Table 3.1).

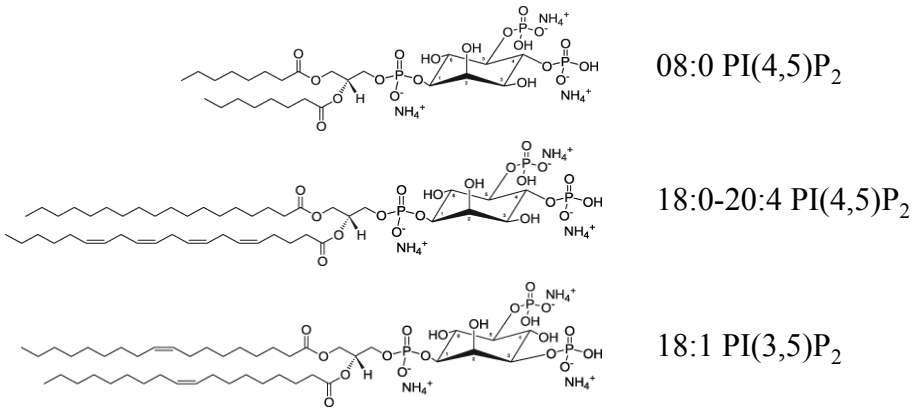


Figure 3.1: Chemical structures of different PIP2 used in this study

Samples were prepared to give a final protein concentration of 0.15 mg/ml. A sample of SEC buffer (see Section 3.2.1) was used as background reference. A volume of lipid buffer (0.035% UDM, 20 mM HEPES (pH 7.5), and 150 mM NaCl) was added to make each sample having the sample amount of lipid buffer as the sample with maximum amount of lipid added. SEC buffer was used to top up each sample to a final volume of 50 μl. After incubation at 4 ℃ for 1 h, samples were loaded into nanoDSF Grade Standard Capillaries (NanoTemper Technologies) in triplicates for $PI(4,5)P_2$ samples and in quadruplicates for $PI(3,5)P_2$ samples. Melting temperature curves were obtained using the Prometheus NT.48 nanoDSF system (NanoTemper Technologies) with a linear thermal ramp (20 ℃ to 80 ℃; 1.0 ℃/min) with an excitation power of 65%. Changes in the emission wavelengths of tryptophan fluorescence at 330 nm, 350 nm and their ratios were used to

determine the transition points of protein unfolding. Data analysis was done using the Prometheus PR. Control software v2.0.3 (NanoTemper Technologies). Melting temperatures (T_m) were measured as the temperature at the first inflection point in the fluorescence ratio curve (350 nm/330 nm), signifying the temperature at which 50% of PC2 is unfolded. Error bars in T_m figures show the standard deviations of the technical repeats in each experiment.

Table 3.1: Summary of Tested Protein-lipid Ratios

Lipid Species	His_{10}/FLAG-tagged	Tested Protein:Lipid Ratios				
		1:0	1:10	1:20	1:30	1:50
18:0-20:4 $PI(4,5)P_2$	✓	✓	✗	✓	✗	✓
	✗	✓	✗	✓	✗	✓
8:0 $PI(4,5)P_2$	✓	✓	✗	✓	✗	✓
	✗	✓	✗	✓	✗	✓
18:1 $PI(3,5)P_2$	✗	✓	✓	✓	✓	✓

3.2.4 Cryo-EM Grid Preparation and Data Acquisition

PIP_2 lipid stocks were prepared as described above (see Section 3.2.3). Purified PC2 was incubated with 18:0-20:4 $PI(4,5)P_2$ at a molar ratio of 1:20 (PC2 tetramer : $PI(4,5)P_2$) and with 18:1 $PI(3,5)P_2$ at a molar ratio of 1:40 (PC2 tetramer : $PI(3,5)P_2$) overnight at 4 ℃. The sample was cleared by centrifugation at 21,000 g for 30 min. For cryo-EM, 3 μl of PC2 with $PI(4,5)P_2$ sample at a protein concentration of $\sim$ 4 mg/ml or PC2 with $PI(3,5)P_2$ sample at a protein concentration of $\sim$ 3.5 mg/ml was applied to glow-discharged Quantifoil 1.2/1.3 holey carbon 300 mesh copper grids. Grids were plunge frozen in liquid ethane using a Vitrobot Mark IV (FEI, Thermo Fisher Scientific) set to 5 ℃, 100% relative humidity, 3.5 s blotting time and -15 blotting force. Data for PC2 with $PI(4,5)P_2$ were collected on a Titan Krios 300-kV transmission electron microscope equipped with a post-column Gatan image filter (GIF; 20 eV slit width) operating in zero-loss mode and a Gatan K2 Summit direct electron detector camera at Central Oxford Structural Molecular Imaging Centre (COSMIC). Movies were

captured for 0.4 s per frame over 8 s. The calibrated pixel size and dose rate were 0.822 Å/pixel and $\sim$ 6.55 electrons/Å^2/s, respectively (total dose 52.4 e$^-$/Å^2). Images were collected in a defocus range between -1.0 and -2.5 μm under focus in 0.25 μm steps. Data for PC2 with PI(3,5)P_2 were collected on a Titan Krios 300-kV transmission electron microscope at the Electron Bio-Imaging Centre (eBic, Diamond Light Source) equipped with a Gatan K2 Summit direct electron detector camera mounted behind a GIF and operated in zero-loss mode (0-20 eV). Movies were captured for 0.2 s per frame over 7 s. The calibrated pixel size and dose rate were 0.816 Å/pixel and $\sim$ 6.0 electrons/Å^2/s, respectively (total dose 42 e$^-$/Å^2). Images were collected in a defocus range between -1.0 and -3.1 μm under focus in 0.3 μm steps.

3.2.5 Image Processing

The beam-induced motion in the movies was corrected and frames were dose-weighted using MotionCor2 [240]. Aligned frames in each movie were averaged to produce a micrograph for further processing. The contrast transfer function (CTF) parameters were estimated using CTFFIND-4.1 [241]. Micrographs with ice contamination or poor CTF cross correlation scores were discarded and the remaining micrographs were processed using RELION 3 [189]. Particle picking was performed using Gautomatch (URL: http://www.mrc-lmb.cam.ac.uk/kzhang/). Local resolutions across the whole map were estimated using RESMAP [242].

3.2.6 Model Building

The previously published model of apo PC2 (PDB ID: 5K47) was used as an initial model and fitted into the cryo-EM maps of lipid-bound PC2 using the *fit-in-map* module from UCSF Chimera [243]. The model was manually adjusted in Coot [244]. A series of B-factor sharpened maps were used to guide model building. Final models of both structures were globally refined and minimised in real space against the RELION-3 C4 post-processed, automatically B-factor sharpened maps with NCS constraints, secondary structure and rotamer restraints imposed and no

Ramachandran restraints applied using the phenix.real_space_refine module in PHENIX [245]. The refinement protocol was validated by taking the final refined models, applying a random shift of up to 0.3 Å to the atomic coordinates and then refining the resultant shifted model against the halfmap1. Model-to-map Fourier shell correlations (FSCs) were calculated using phenix.mtriage against halfmap1 (FSC_{work}), halfmap2 (FSC_{free}) and the full map (FSC_{sum}).

3.3 Results and Discussion

3.3.1 Characterisation of PC2 Lipid Binding Using nanoDSF and Biochemical Assays

To test the ability of PC2 to bind to different lipids *in vitro*, we used a biochemical method, PIP-strips, which are nitrocellulose membranes with lipids including phosphoinositides and other anionic and zwitterionic phospholipids spotted onto their surface [246]. Protein binding to relevant lipids can be detected with antibodies to the tagged protein. This method has been used for example to confirm binding of PIP_2 to the GluA1 ionotropic glutamate receptor [247].

The truncated protein construct of PC2 (hPC2: Pro185 − Asp723) used in this study has been well optimised before [7]. The purification process produced clean protein sample with high degree of monodispersity (Fig. 3.2A & 3.2B). The His_{10}/FLAG-tagged PC2 in UDM micelles was used for the PIP-strip experiments. Five biological repeats were performed using protein expressed in Sf9 insect cells ($N = 4$) or in HEK293S GnTI$^-$ mammalian cells ($N = 1$). The results from this assay (Fig. 3.2) suggest that detergent solubilised truncated PC2 can bind a range of phosphatidylinositol phosphates (PIPs), including all species of PIP_2. In contrast, it does not exhibit interactions with simple anionic (PS, PA, SP1) or zwitterionic (PE, PC) lipids under the experimental condition used. The relatively intensity of HRP signals across different repeats of PIP-strips may suggest that PI(3,5)P_2 has relatively stronger binding to PC2 than the more prevalent PI(4,5)P_2. The same lipid binding patterns shown in the HEK293S GnTI$^-$ cell expressed PC2 may

suggest that lack of complex N-glycans does not affect the ability of PC2 to bind PIPs, especially if the binding site is on the cytosolic half of the transmembrane region given that PIP_2 is only present on the inner leaflet of plasma membranes.

In order to assess whether PIP_2 has any stabilising or destabilising effect upon binding to PC2, the intrinsic tryptophan fluorescence of PC2 protein, as a measure of thermal stability, in the presence of different PIP_2 was compared to PC2 samples without added lipids (apo) using low volume differential scanning fluorimetry (nanoDSF). Since the PIP strip experiments were done with His_{10}/FLAG-tagged PC2, to include the possibility that the protein tags might contribute to the lipid binding, both tagged and untagged proteins were tested with PI(4,5)P_2 with different tail length and saturation (18:0-20:4 and 08:0). 18:1 PI(3,5)P_2 showed relatively consistent high binding intensities between the different repeats of PIP strips, and thus its binding effect on PC2 was also examined. As shown in Figure 3.3, there is no significant shift in melting temperatures (T_m) of PC2 in the presence of either PIP_2 species. Including consideration of the variability in the measurements, the positively charged His_{10}/FLAG tag does not show significant influence on the T_m upon binding of the negatively charged lipids. Although these results might suggest that PIP_2 binding to PC2 would not alter the stability of the truncated protein, and thus probably its conformational state, further experiments need to be done to verify the effect of PIP_2 binding on PC2. Given that the protein construct used for the current experiments lacks the cytoplasmic N- and C-termini, it is possible that the additional coordinations of the PIP_2 head groups from the missing residues would affect the stability of the overall channel.

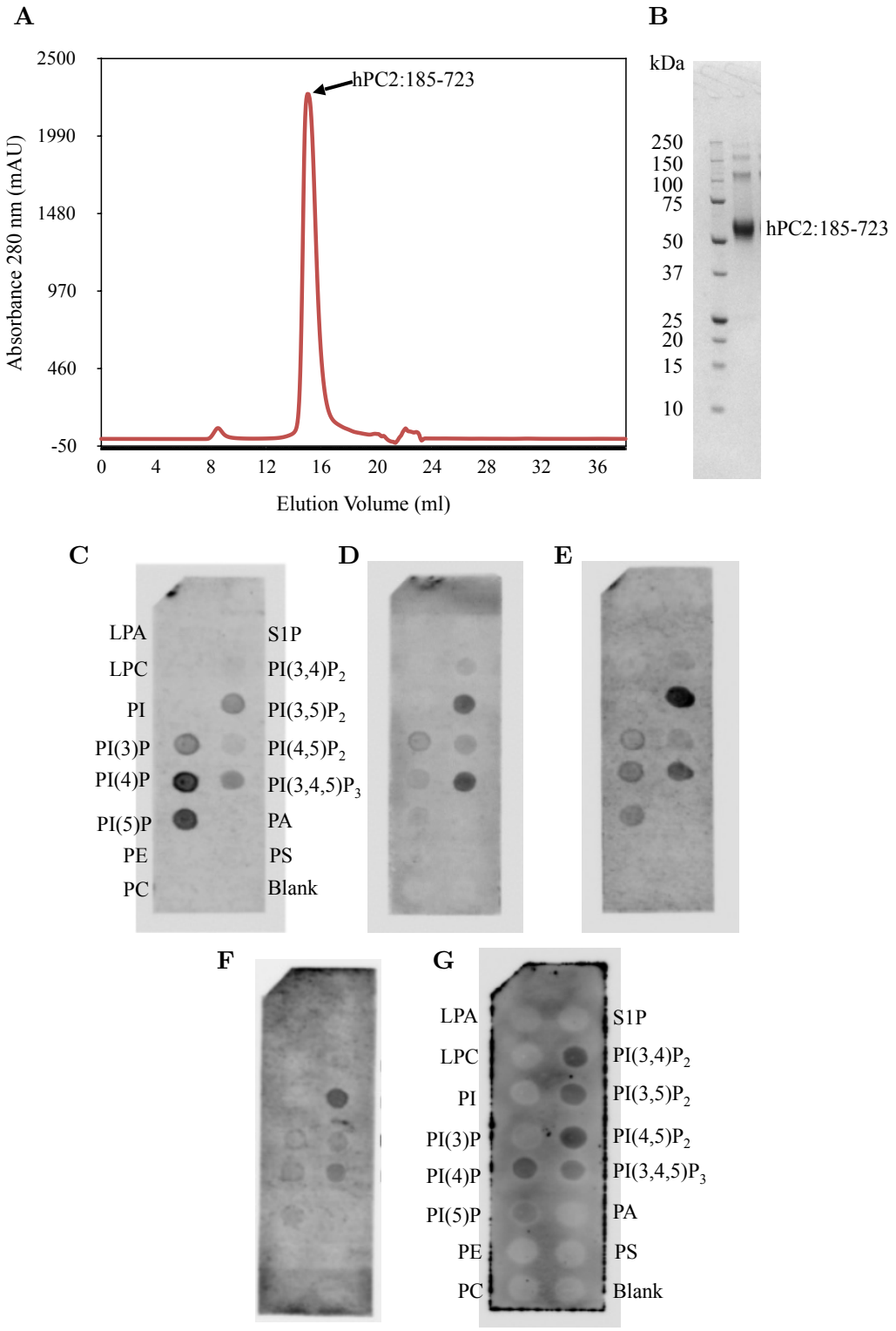


Figure 3.2: Biological repeats of PC2 binding to PIP strips. PC2 protein quality used in this study was analysed by **A:** SEC and **B:** SDS-PAGE. PC2 expressed in different batches of either HEK293S GnTI^{-} (C) or Sf9 (D – G) insect cells were used to repeat the PIP strip experiments. Although binding intensities vary between the different repeats, the bound lipid species are broadly consistent.

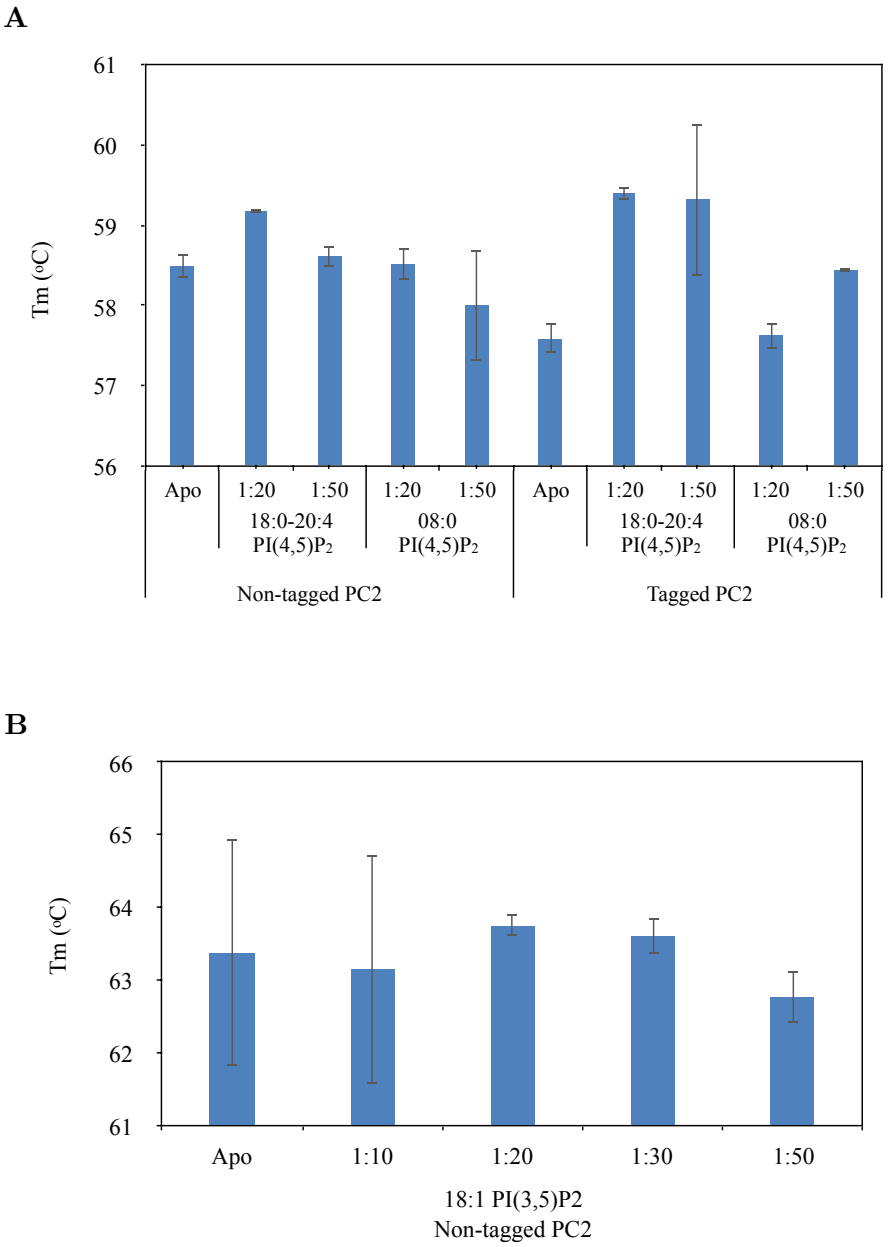


Figure 3.3: Thermal stability of PC2 in the presence of PIP_2. **A:** T_m were measured in absence and the presence of $PI(4,5)P_2$ with different tail length and saturation. Both His_{10}/FLAG-tagged and untagged PC2 were tested. **B:** Effects of different concentrations of $PI(3,5)P_2$ on T_m of untagged PC2 were also investigated. Molar ratios between protein monomer and PIP_2 lipids are indicated.

3.3.2 Structural Characterisation of PC2 in the Presence of PIP Lipids by Cryo-EM

In order to locate the binding site of PIP_2 on PC2, the purified PC2 proteins in detergent were incubated with 18:0-20:4 $PI(4,5)P_2$ at a molar ratio of 1:20 (PC2 tetramer : $PI(4,5)P_2$) or $PI(3,5)P_2$ at a molar ratio of 1:40 (PC2 tetramer : $PI(3,5)P_2$) for analysis by cryo-EM. We used similar conditions to our original 4.2 Å PC2 structure (PDB ID: 5K47), with a truncated construct (residues Pro185 to Asp723), purified in the detergent UDM.

3.3.2.1 Data Processing and Model Building for Sample with $PI(4,5)P_2$

At the time of data collection for both $PI(4,5)P_2$ and $PI(3,5)P_2$ samples, images were set to be record at a nominal magnification of 165,000×, yielding a calibrated pixel size of 0.85 Å/pixel from EPU (automated data acquisition software for single particle workflow). An initial round of processing of both datasets produced a map with a masked resolution of 3.1 Å for PC2 in the presence of $PI(4,5)P_2$, and a map with a masked resolution of 3.5 Å for PC2 in the presence of $PI(3,5)P_2$, as estimated by Fourier shell correlation (FSC). The initial atomic models of both datasets were generated by morphing the previously published structure of PC2 (PDB ID: 5K47) into the cryo-EM maps using PHENIX [245]. We noticed a global inflation of the protein structure for both datasets (Fig. 3.4). In parallel, the Central Oxford Structural Molecular Imaging Centre (COSMIC) recalibrated the electron microscope to show that the actual detector pixel size at the set nominal magnification was 0.822 Å/pixel for the PC2/$PI(4,5)P_2$ dataset, which indicates that the molecular model of the protein based on the pixel size of 0.85 Å/pixel would have $\sim$ 2.8% inflation compared with its actual size. Given that the same type of global change of protein structure was observed for the dataset collected at the Electron Bio-Imaging Centre (eBic, Diamond Light Source), we suspected that the set pixel size was not the actual detector pixel size for PC2/$PI(3,5)P_2$ dataset either. In order to correctly estimate the pixel size for PC2/$PI(3,5)P_2$ dataset, we first reprocessed the PC2/$PI(4,5)P_2$ dataset with the known calibrated pixel size.

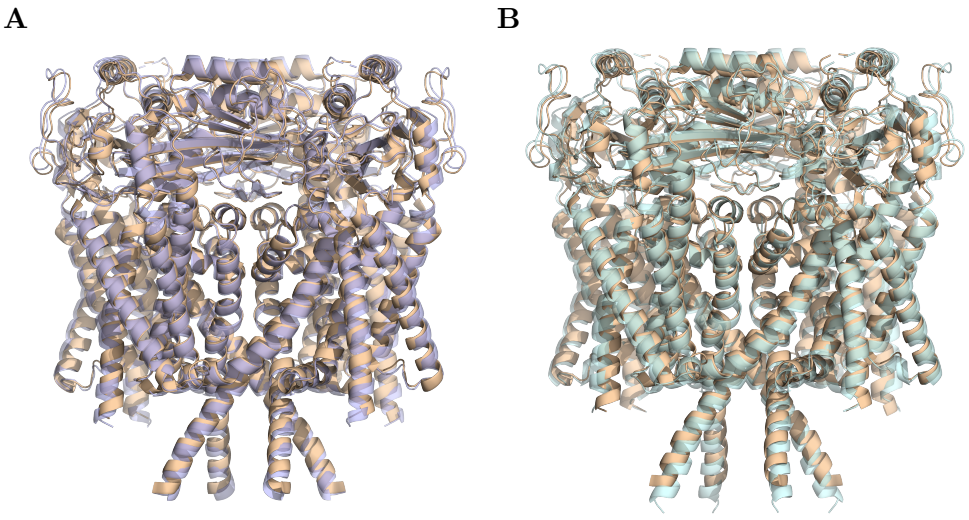


Figure 3.4: Overlay of PC2 models with claimed detector pixel sizes to the structure from PDB ID: 5K47 (light orange). **A:** Model built based on the cryo-EM density of PC2 in the presence of PI(4,5)P_2 processed with pixel size of 0.85 Å/pixel is shown in transparent light purple. **B:** Model built based on the cryo-EM density of PC2 in the presence of PI(3,5)P_2 processed with pixel size of 0.85 Å/pixel is shown in transparent light cyan.

An overview of the data processing procedures for the dataset of PC2 with PI(4,5)P_2 is shown in Fig. 3.5A. Whole-frame movie alignment were done on 1,759 raw micrograph movies with MotionCor2 [240]. A representative micrograph of PC2 particles after initial frame alignment is shown in Figure 3.5B. Estimation and correction of contrast transfer function (CTF) parameters for motion corrected and non-dose-weighted micrographs were carried out with the RELION-3 [189] wrapper of CTFFIND4.1 [241]. The calculated CTF patterns are compared to the Thon rings in the background-subtracted power spectra of corresponding movies. Micrographs were discarded where the fit between experimental and theoretical amplitude spectrums were poor. Particle selection was done in two steps with Gautomatch (URL: http://www.mrc-lmb.cam.ac.uk/kzhang/). An initial set of 77,399 particle images ($\sim$ 207 $\times$ 207 pixels) were obtained from the remaining 1,597 motion corrected micrographs with approximate particle diameter of 170 Å without particle picking templates. The particles were subjected to reference-free

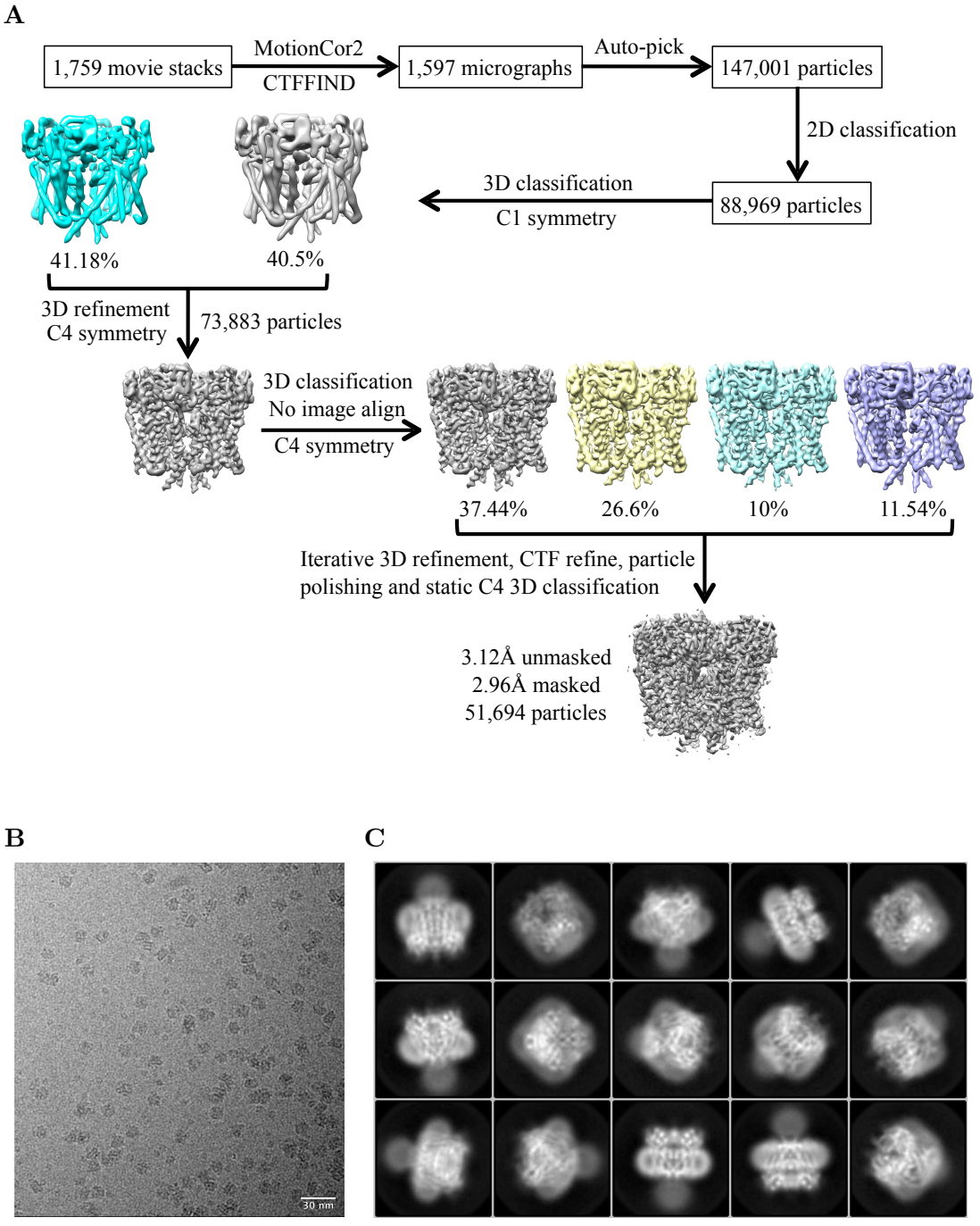


Figure 3.5: **A:** Workflow of data processing for the dataset of PC2 with PI(4,5)P_2 in detergent. **B:** Representative micrograph of PC2 particles recorded on Titan Krios 300-kV transmission electron microscope. **C:** Representative 2D class averages of PC2.

2D classification in RELION-3. Non-particle and low resolution particle images were discarded. Then the class average of 41,840 selected particles were used as templates for reference-based autopicking with a cross-correction cutoff of 0.5 after low-pass filtering to 30 Å. A total of 147,001 particles were automatically picked. For computational efficiency, the particles were down-scaled to half of the original pixel size (Fourier binned 2 × 2, 1.644 Å/pixel). Subsequently, three rounds of iterative 2D classification were performed in RELION-3. Particles in incoherent or poorly resolved classes were removed. A selected set of 2D class averages of detergent solubilised PC2 particles are shown in Figure 3.5C, which consist of a broad distribution of particle orientations with clear features of the extracellular TOP domain and transmembrane helices. The remaining 88,969 particles were extracted at restored pixel size of 0.822 Å.

A low-resolution reference model was generated from the 2D class averages. The initial model was used as template for the first round of 3D classification without symmetry imposed. Particle classes with interpretable features were visually inspected. The two 3D class averages with the highest estimated resolutions in the non-symmetrised 3D classification exhibit consistent C4 symmetry, and there is no visible conformational difference between them. Therefore, a subset of 73,883 particles from these classes was combined and used for subsequent 3D classification with C4 symmetry imposed. First round of 3D 'gold-standard' refinement with the C4-symmetrised particles resulted in an initial reconstruction with a nominal unmasked resolution of 3.5 Å. Subsequent 3D classification without further image alignments was performed on the refined particles, allowing finer angular searches around the consensus refined orientations. The selected particles from static 3D classification were refined and post-processed. To further improve the resolution, the per-particle defocus values and beam-tilt values were estimated using the the refined and post-processed 3D structures as references. The beam-induced motion trajectories of particles were further estimated and corrected with Bayesian polishing implemented in RELION-3. The prior likelihood parameters were trained on a randomly selected subset of motion corrected micrographs containing 10,005

refined particles. All particles were subsequently polished based on the estimated prior. The processes of static 3D classification, 3D refinement, post-processing, CTF and beam-tilt refinement and Bayesian polishing were run iteratively, which resulted a final set of 51,694 particles converged at an unmasked resolution of 3.12 Å. RELION post-processing using unfiltered half maps and a soft-edged mask to exclude the region occupied by the detergent micelle yielded a final B-factor sharpened map ($-$ 84.56 Å^2) with a nominal resolution of 2.96 Å (FSC=0.143) (see Figure 3.6A). Local resolution was calculated and shown on the map in Figure 3.6C. The N- and C-termini and the S3-S4 linker regions are the most flexible with resolution around 4 Å. The core region of the TOP domain and the central pore surrounding helices have the highest resolution ranging from 2.7 Å to 3 Å. The outer helices are relatively less well resolved with resolution around 3.3 Å.

The molecular model of PC2 with PI(4,5)P_2 was built based on our previously published apo structure (PDB ID: 5K47). At 2.96 Å resolution, the backbone of the protein can be clearly traced and most of the side chains can be unambiguously assigned (see Figure 3.7A). The overall conformation of the structures remained the same as the original structure, i.e. a closed conformation. However, there are some inconsistencies. In the original PDB ID: 5K47 structure, Cys331 and Cys344 appeared to be in their reduced form with free thiol groups. In our current dataset, there is clear density showing that a disulphide bond is formed between these cysteines (Figure 3.7C). In addition, at the C-terminal end of the transmembrane helix S3, there is a mismatch for the location of Asn531 between our current structure and PDB ID: 5K47, which leads to subsequent mistrace of residue positions up to Asn543, where the two structures converge again (see Figure 3.7D $-$ 3.7F). As shown in Figure 3.7D, with current resolution, the side chain positions can be confidently identified, and our structure agrees with PDB ID: 5T4D, which was solved at better resolution (3 Å) than our original PDB ID: 5K47 structure. Therefore, the current structure of PC2 protein is most likely to be correct. The final 2.96 Å model was refined against a reconstruction generated from all of the particles. For cross validation of the model, a ”half map” generated

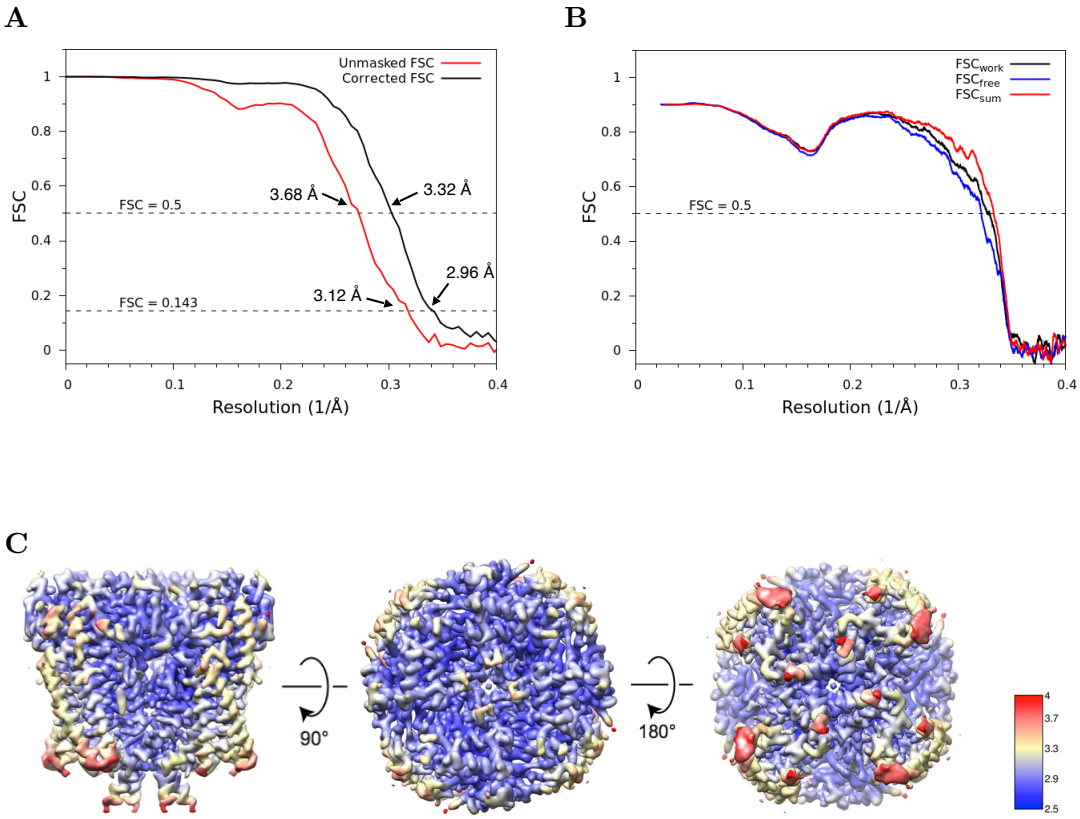


Figure 3.6: **A:** Fourier shell coefficient (FSC) curves of the masked (red) and unmasked (black) maps for PC2 with PI(4,5)P_2 dataset. FSC of 0.143 and 0.5 are indicated with dotted lines. **B:** FSC curves for cross validation comparing the model to half maps 1 (used in test refinement, black) and half map 2 (not used in test refinement, blue), and to summed map (red dotted). **C:** Cryo-EM map of PC2/PI(4,5)P_2 coloured by local resolution estimate (calculated using RESMAP).

from half of the particles was used to refine the semi-randomised (by 0.3 Å) model, and the reconstruction generated from the remaining half of the particles was used to compare with this refined model. The fit of the refined models to the half maps and the full map was assessed using FSC. From Figure 3.6B, it can be seen that the FSC curves have reasonably good agreement indicating our model is not overly fitted to the data.

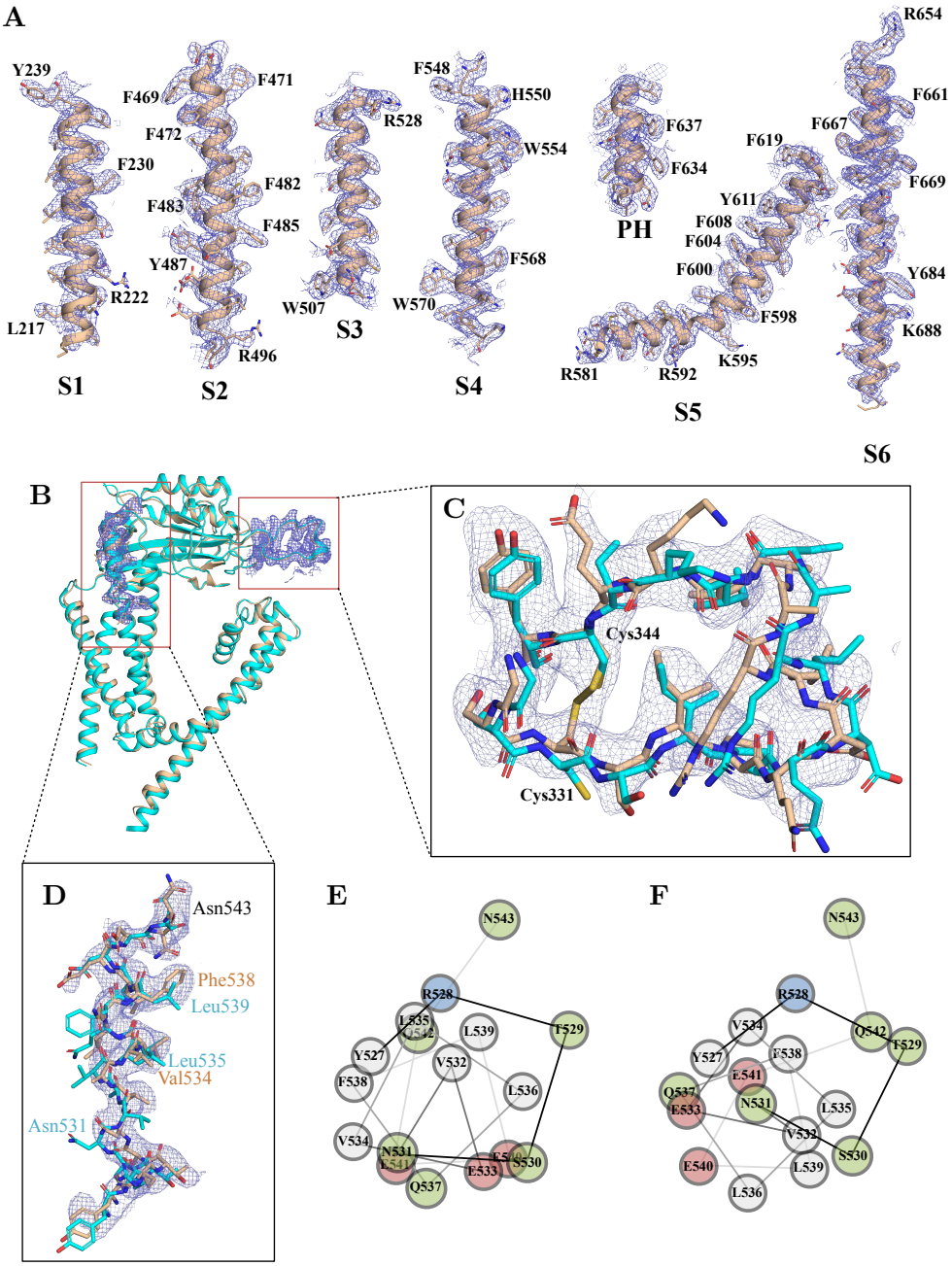


Figure 3.7: **A:** Cryo-EM density maps of current PC2 with PI(4,5)P_2 dataset are shown for the transmembrane helices (S1-S6) and the pore helix (PH). **B:** Comparison between molecular model of PC2 with PI(4,5)P_2 and the structure of PDB ID: 5K47. Overlay of one protein subunit of PDB ID: 5K47 (cyan) and our structure of PC2 with PI(4,5)P_2 (pale orange) are shown in cartoon representation. Densities of our PC2 with PI(4,5)P_2 dataset around the regions of inconsistencies are shown in blue mesh. **C:** Density map of residues from Gly329 to Tyr345 showing the disulphide bonds between Cys331 and Cys344. **D:** Density map of residues from Tyr527 to Asn543 showing the mistrace of residue positions in PDB ID: 5K47 (cyan). **E:** 2D projection of C-alpha atom positions of helix expanding from Tyr527 to Asn543 of PDB ID: 5K47. **F:** 2D projection of C-alpha atom positions of helix expanding from Tyr527 to Asn543 of structure of PC2 with PI(4,5)P_2 in this study.

3.3.2.2 Unidentified Density in C1 Map of PC2/PI(4,5)P_2

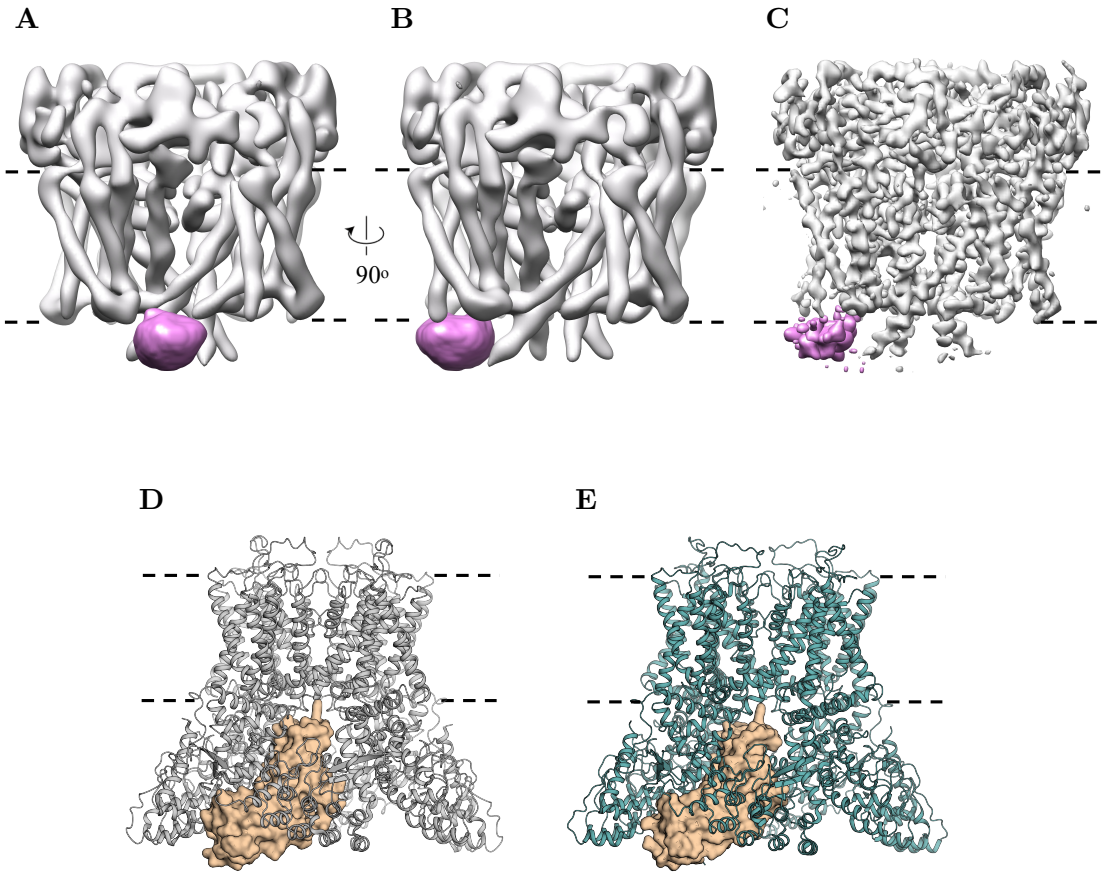


Figure 3.8: A discrete density revealed in a class averaged map with C1 symmetry. **A & B:** Class averaged density map from C1 3D classification. **C:** Post processed C1 map at a resolution of 4.3 Å. **D:** Structure of rabbit TRPV5 in complex with bovine CaM. The protein is shown in grey cartoon, and CaM is in pale orange surface representation. **E:** Structure of the human TRPV6-CaM complex. The protein is shown in light teal cartoon, and CaM is in pale orange surface representation.

One of the non-symmetrised 3D classes with 14,565 particles shows one discrete density just below the transmembrane region of PC2 (Fig. 3.8), which has not been reported before. Given the relative size of the density and that it is located outside of where detergent micelle/membrane bilayer situates, it is likely that this density does not correspond to a lipid head group or a cluster of lipid head groups. The

data was then used for a round of 3D 'gold-standard' refinement followed by post-processing which uses a soft-edged mask that excluded the detergent micelle and performs B-factor sharpening. This produced a 4.3 Å resolution map. However, there is still no clear identifiable features of the extra density (Fig. 3.8C). Although fitting the published PC2 structure (PDB ID: 5K47) in to the protein density showed that the resolved density of the transmembrane helix S6 is shorter than the protein construct used, the discontinuity between the protein and the extra densities suggests that the discrete density is unlikely to be part of PC2 protein density. In our density, there is a slight asymmetry shown in S6 among the tetramers with the S6 density closest to the extra density being longer than the other three. It has been shown that residues Lys590 to Glu600 at the C-terminal region of the related protein PKD2L1 are crucial for the binding of the regulatory calcium binding protein − calmodulin (CaM), and thus for the functionality of PKD2L1 [248]. Sequence and structural alignment of PC2 and PKD2L1 shows that the equivalent region in PC2 is residues Lys710 to Asn720, which are included in our PC2 construct. Thus, there is a possibility that the extra density is part of bound calmodulin that has been co-purified with PC2. Cryo-EM structures of rabbit TRPV5 (PDB ID: 6O20) in complex with bovine CaM and human TRPV6 (PDB ID: 6E2F) in complex with human CaM show that CaM is located directly below the lower gates in both TRPV channels. The C-lobe of CaM interacts with the channels right after S6 helix and before the TRP domain. In comparison, the extra density in our dataset is also at the end of S6 helix but not on the symmetry axis. Since members of the TRPP subfamily do not have the TRP helices, the binding site for CaM observed in TRPV channels may not be preserved in TRPP proteins. Further experiments such as protein complex immunoprecipitation will be needed to test the hypothesis of CaM binding to our truncated PC2 construct. Several frameshift mutations between Lys710 to Asn720 of PC2 (Val713fs2X, Lys714fs1X, Leu717fs19X, Asn720fs16X and Asn720fs4X) have been shown to be definitely pathogenic. This suggests the functional importance of this region, which might be related to the ability of binding CaM.

3.3.2.3 Data Processing and Model Building for Sample with PI(3,5)P_2

Given the observed inflation of the structure, to correctly process the dataset of PC2 with PI(3,5)P_2, we firstly estimated the corresponding pixel size using the PC2/PI(4,5)P_2 structure with the known pixel size of 0.822 Å/pixel. As sanity check, we used the same molecular model as reference for pixel size validation of the cryo-EM map of PC2 with similar resolution (EMD-8354). We sampled all cryo-EM maps at different pixel sizes and fitted our coordinates into each of them. The goodness of fit of the cryo-EM maps to our reference structure was evaluated by correlation coefficient. As shown in Figure 3.9, the pixel sizes corresponding to the highest correlation coefficients for both our PI(4,5)P_2 map and EMD-8354 map reflect the reported pixel sizes very well. Using the same approach, the pixel size of our PI(3,5)P_2 dataset was estimated as the one that gives the best fit to the structure, i.e. 0.816 Å/pixel.

Similar data processing procedures have been carried out for the dataset of PC2 with PI(3,5)P_2 (Appendix B, Figure B.1A). Motion correction, dose weighting and CTF estimation were performed on 3,977 raw micrographs, and 3,353 good micrographs were selected. A set of 323,538 particles were initially picked without reference and classified into 2D classes. Representative well resolved class averages were used as templates for reference-based particle picking after low-pass filtering to 30 Å. A total of 224,402 particles were automatically picked using Gautomatch (URL: http://www.mrc-lmb.cam.ac.uk/kzhang/). Three rounds of iterative 2D classification were performed in RELION-3 [189] to remove bad particles. A low-resolution reference model was generated *ab initio* based on the 141,828 remaining particles. The initial model was used for the first round of 3D classification without symmetry imposed. One further round of 3D classification with C1 symmetry was performed on a set of 97,892 particles from classes with the best estimated resolutions. None of the 3D classes has any extra density as seen in the PC2/PI(4,5)P_2 dataset. A subset of 91,844 particles was used for subsequent 3D classification with

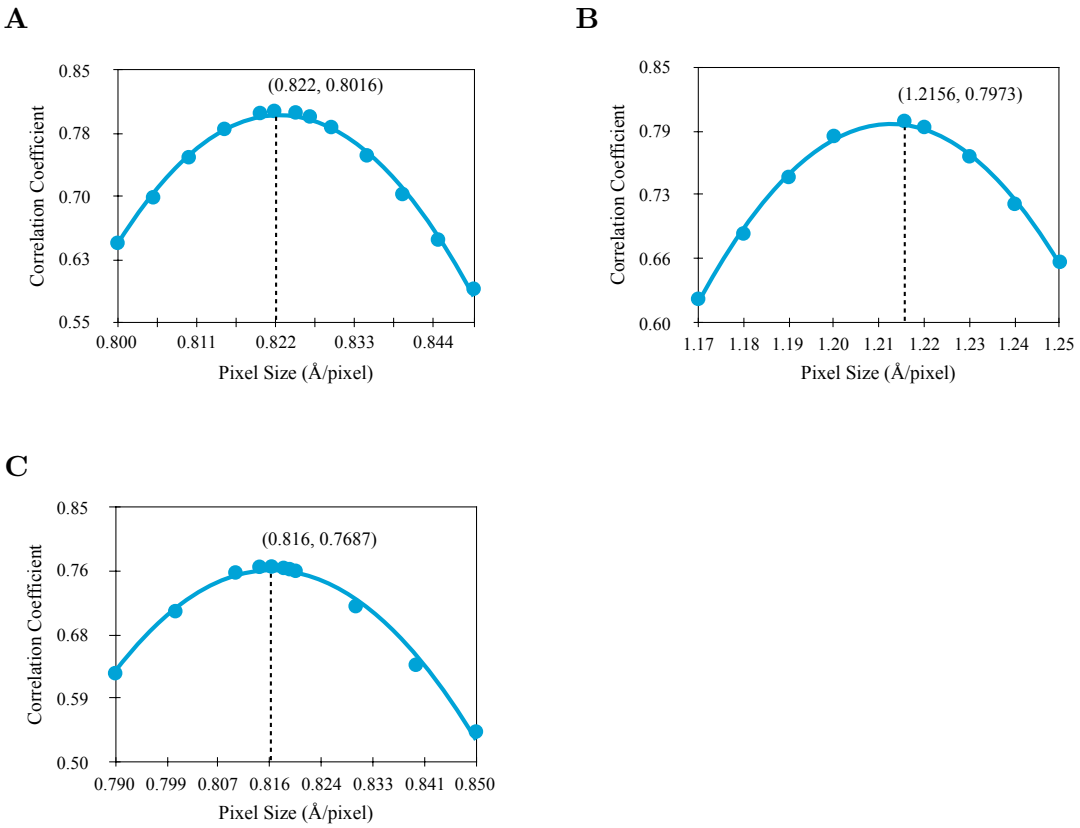


Figure 3.9: Pixel size estimation for the dataset of PC2 with PI(3,5)P_2. Correlation coefficients between our PC2/PI(4,5)P_2 structure and the resampled cryo-EM maps at different pixel sizes from the original **A:** 3 Å PC2/PI(4,5)P_2 map processed at 0.822 Å/pixel, **B:** and 3 Å EMD-8354 map processed at 1.2156 Å/pixel. The reported pixel sizes and corresponding correlation coefficient are labelled with dashed lines. **C:** Estimation of pixel size of the PC2/PI(3,5)P_2 dataset with correlation coefficient between our PC2/PI(4,5)P_2 structure and the PC2/PI(3,5)P_2 map at different pixel sizes. The pixel size (0.816 Å/pixel) that gives the highest correlation coefficient is marked with dashed line.

C4 symmetry imposed. The data was then used for the first round of 3D 'gold-standard' refinement, which resulted in an initial reconstruction at a resolution of 4.1 Å. A final set of 37,297 particles was selected from 3D classification performed without further image alignments. Iterative CTF refinement and Bayesian polishing were performed prior to a final auto-refinement procedure in RELION, which converged at an unmasked resolution of 3.66 Å. Subsequent post-processing, using a soft-edged mask that excluded the detergent micelle, produced a 3.4 Å

resolution map. This reported resolutions were based on a FSC threshold of 0.143 (see Figure 3.10A). Local resolution ranging from 2.9 Å to 4.5 Å are shown on the map in Figure 3.10C. The molecular model of PC2 with PI(3,5)P_2 was built by morphing the PC2/PI(4,5)P_2 coordinates into the 3.4 Å map and local details were adjusted in Coot. The density map of PC2/PI(3,5)P_2 also supports the existence of the disulphide bond. The model was also validated with independent half maps showing good overlap of FSC curves (Figure 3.10B).

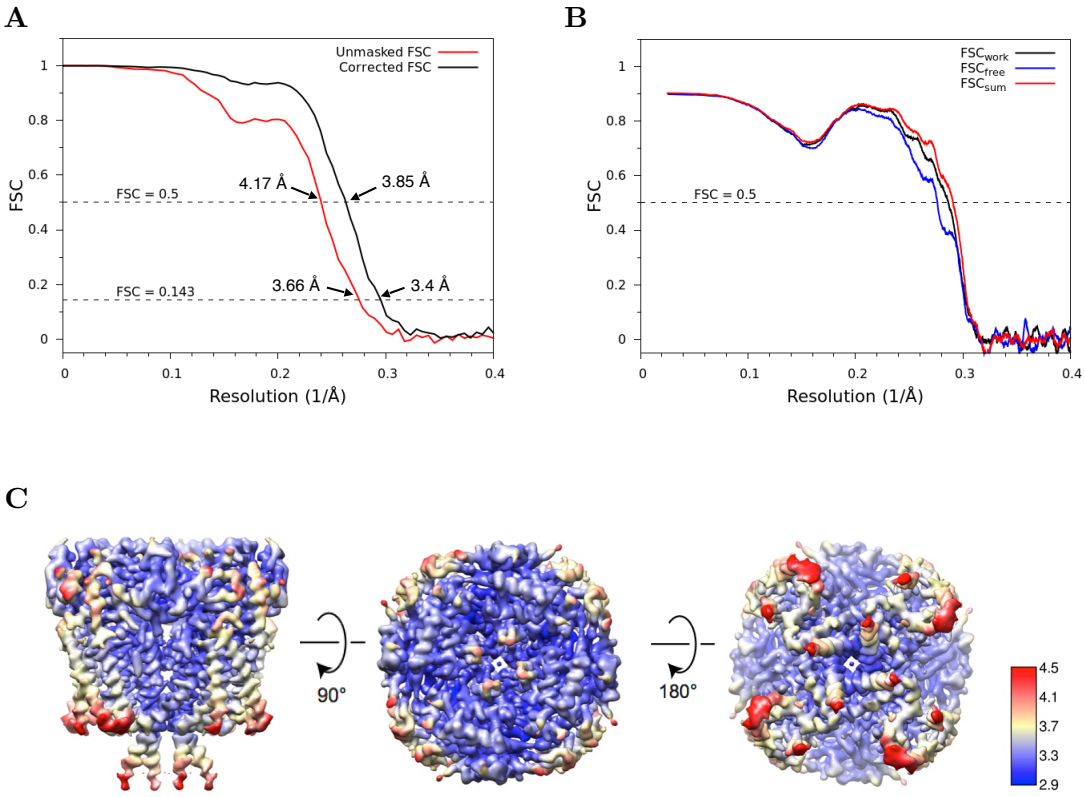


Figure 3.10: **A:** Fourier shell coefficient (FSC) curves of the masked (red) and unmasked (black) maps for PC2 with PI(3,5)P_2 dataset. FSC of 0.143 and 0.5 are indicated with dotted lines. **B:** FSC curves for cross validation comparing the model to half maps 1 (used in test refinement, black) and half map 2 (not used in test refinement, blue), and to summed map (red dotted). **C:** Cryo-EM map of PC2/PI(3,5)P_2 coloured by local resolution estimate (calculated using RESMAP).

3.3.2.4 A Possible Calcium Ion in the Selectivity Filter

Both PC2/PI(4,5)P_2 and PC2/PI(3,5)P_2 structures in the current study appear to be in the closed state. The radii of the selectivity filter region, defined by half of the distance between the carbonyl oxygens of Leu641 from opposite chains, are both 2.6 Å, which are comparable to that of PDB ID: 5T4D (2.5 Å) and 5K47 (2.4 Å). Given that the radii of hydrated physiological cations Na^+, K^+ and Ca^{2+} are more than 3.3 Å, the filter is too narrow for hydrated ions to pass through. However, one spherical density is clearly present below the selectivity filter in each of the PC2/PIP_2 complex cryo-EM maps with the density deeper in the central pore in the PI(3,5)P2 map compared to the PI(4,5)P2 map (see Figure 3.11). These densities are visible at several thresholds, and the peaks are prominent even in C1 maps. In addition, they have similar peak heights as the protein peaks, which indicates that they are significant and could correspond to bound cations. Not surprisingly, it is difficult to distinguish putative ion density from noise and they are not visible at high contour level in the previously published low resolution maps. There is an elongated density present in the equivalent location in the similar-resolution map EMD-8354 (corresponding to PDB ID: 5T4D; 3 Å resolution; [13]). However, this peak has much lower peak height comparing to protein peaks (Figure 3.11B), and therefore likely to be noise. Since the radii of dehydrated Na^+, K^+ and Ca^{2+} ions are 1.2 Å, 1.5 Å and 1.0 Å, taking into account the van der Waals radii of the carbonyl oxygens of Leu641 (1.52 Å), the effective radius of the filter for both of our structures (1.08 Å) could possibly allow a dehydrated Ca^{2+} ion to be coordinated around the filter residues (Leu641 and Gly642). Although the difference in the pore radii of our structures and PDB ID: 5T4D is very small, the available radius of 0.98 Å for PDB ID: 5T4D might contribute to the lack of clear ion density in the map of EMD-8354. The difference in density between our data and EMD-8354 is also likely to originate from the presence of Ca^{2+} ions in our buffer condition, which were not present in that for 5T4D/EMD-8354. Although it is highly likely that the spherical densities present in our current maps correspond to calcium

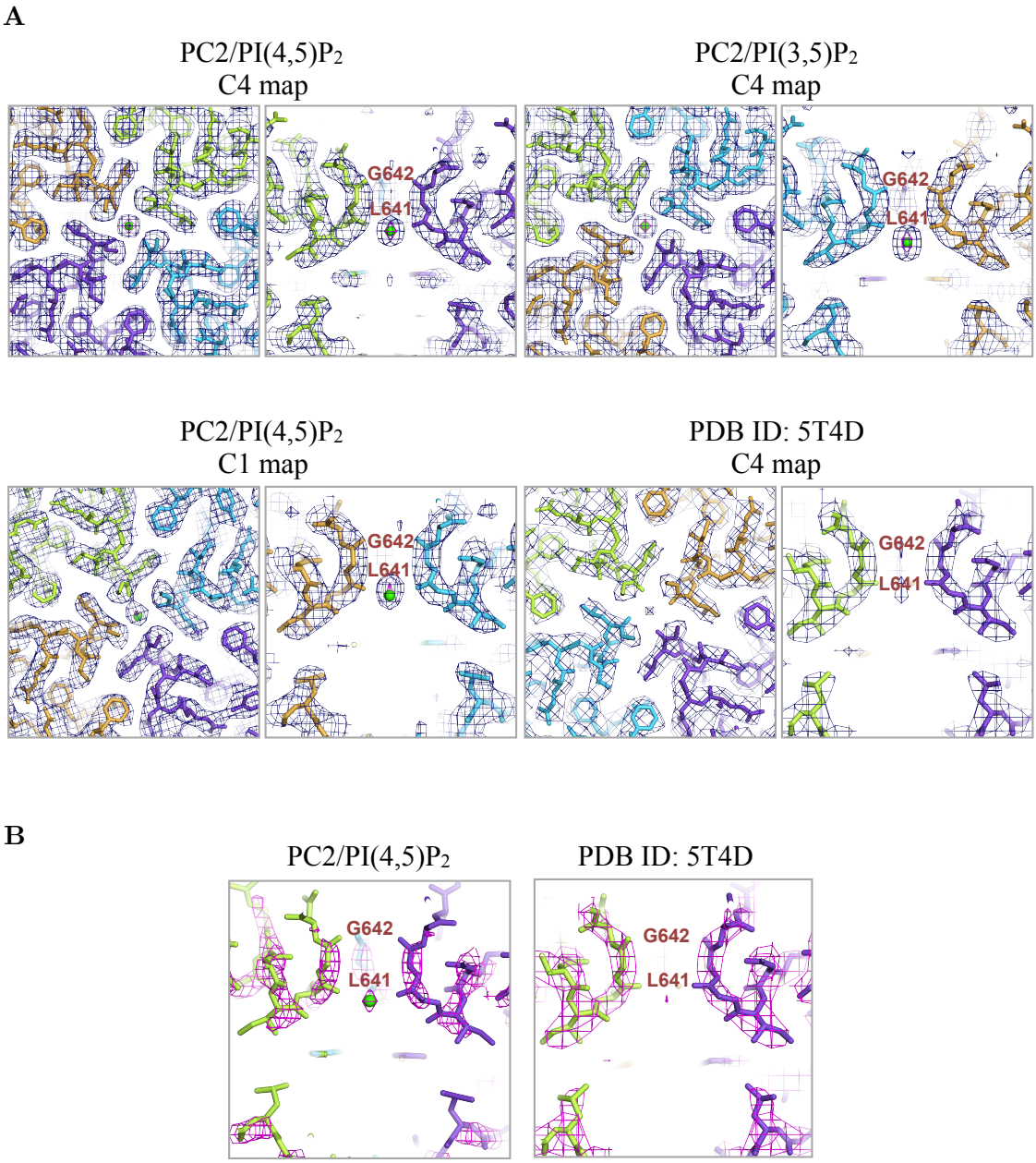


Figure 3.11: Cryo-EM density of putative cation ion. **A:** Density and structure in the region of the bound cation in the ion-conduction pathway from the C4 and C1 maps derived from EMD-8354 (corresponding to PDB ID: 5T4D; 3 Å resolution), the PC2/PI(4,5)P_2 and the PC2/PI(3,5)P_2 datasets. The modelled cation is shown in green. Protein density (dark blue) is shown at 3σ. The putative ion density is contoured at 3σ (dark blue) and at 8σ (magenta). **B:** C4 maps of EMD-8354 and the PC2/PI(4,5)P_2 dataset contoured at 8σ comparing the peak heights between the putative ion and the protein.

ions. Without further confirmation, unknown particles were built into the ion density in both of our structures.

3.3.2.5 Lipid Interactions Revealed by Cryo-EM

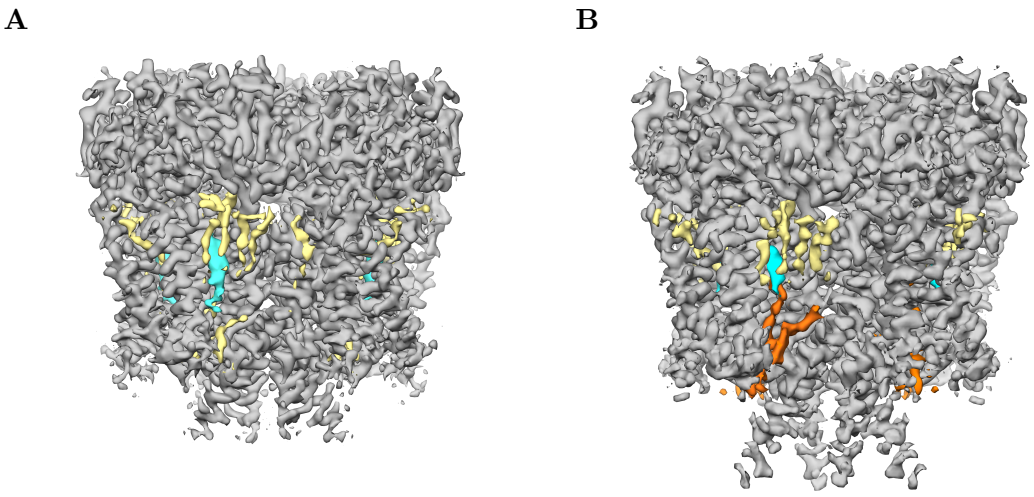


Figure 3.12: A: Cryo-EM density of PC2 with PI(4,5)P_2 dataset. Protein density is shown in grey surface. Densities assigned to UDM detergent molecules are shown in yellow. One clear cholesterol density (cyan) is located in the hydrophobic cleft formed by S3 and S4 helices and S6 of the adjacent subunit. **B:** Cryo-EM density of PC2 with PI(3,5)P_2 dataset. Protein density (grey), UDM detergent densities (yellow) and cholesterol density (cyan) are also shown. One lipid-like density between S3, S4 and S5 helices in the inner-leaflet-facing surface of PC2 is shown in orange. The presumed head group region of the density is not clear enough to unequivocally assign the identity of the lipid, and thus no lipid was built into the density.

In order to locate sites for lipid interaction, we inspected all lipid-like non-protein densities in the transmembrane region for both datasets (Figure 3.12A & 3.12B). One density within the upper segment of S1-S4 helices was found in both cryo-EM maps, which exhibit clear characteristics of UDM detergent molecule. Four other densities in the upper leaflet between S4 of one subunit and S5 and S6 of adjacent subunit were attributed to UDM molecules as well (Figure 3.13A & 3.13B).

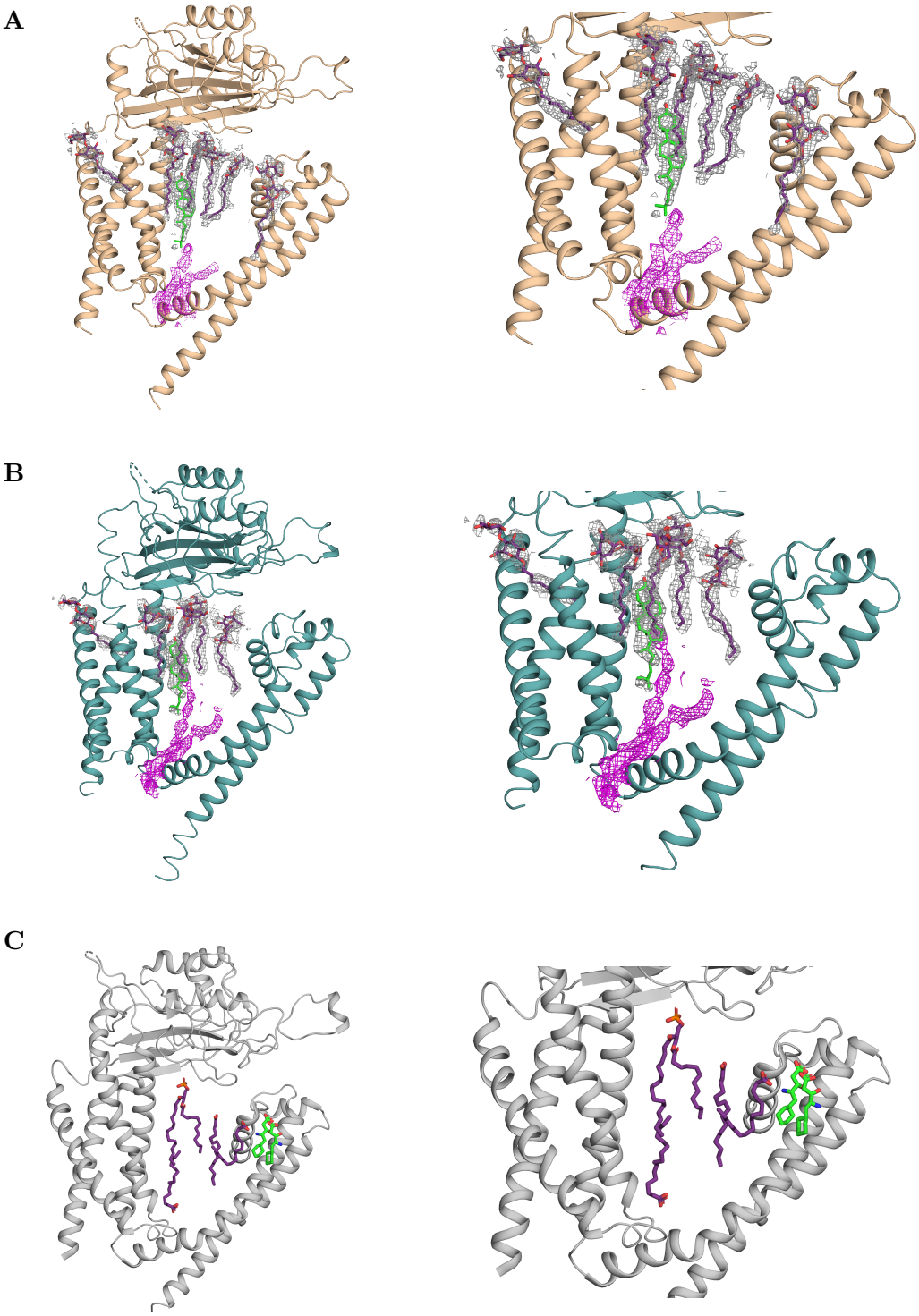


Figure 3.13: **A:** Detergents and lipids built in PC2/PI(4,5)P_2 structure. **B:** Detergents and lipids built in PC2/PI(3,5)P_2 structure. Putative lipid densities in these two datasets are shown in magenta. **C:** Lipids bound in previously published PC2 structure (PDB ID: 5KME). Zoomed-in figures are shown on the right hand side of each panel.

Their positions are close to the lipids/single-chain fatty acids found in previously published structure of PC2 in amphipol (PDB ID: 5KME/5KMF, [15]) (Figure 3.13C). One detergent molecule next to S5 helix in the PC2/PI(4,5)P_2 dataset, which is not well-resolved in the PC2/PI(3,5)P_2 map, is in a similar location to the cholesteryl hemisuccinates (CHS) identified in PDB ID: 5KME/5KMF. One clear density corresponding to one cholesterol molecule is located behind the hydrophobic tails of UDM and coordinated between the S3 and S4 helices and S6 of the adjacent subunit (Figure 3.13A & 3.13B). The cholesterol must have remained bound to PC2 throughout the extraction and purification process, since neither cholesterol nor cholesterol mimetic such as CHS, was used in PC2 purifications. Interestingly, comparable density is also visible in the cryo-EM map for the PDB ID: 5T4D structure. The hydroxyl group of the cholesterol hydrogen bonds to Gln557 and Asn560. The steroid nucleus of cholesterol sits within a shallow hydrophobic pocket formed by a group of isoleucine (Ile561, Ile659), leucine (Leu517, Leu656) and valine (Val564, Val655) residues (Figure 3.14A). This is consistent with previous studies of cholesterol binding in GPCRs, which showed that isoleucine is the main contacting residue within the binding sites ([249, 250]).

As can be seen in Figure 3.12A & 3.12B, there are non-protein densities located in the lower leaflet between S3, S4 and S5 in both structures, suggestive of lipid and/or detergent occupancy. The shape and location resemble the phosphatidylinositol lipid in TRPV1 (PDB ID: 5IRZ) and the PI(4,5)P_2 in TRPV5 (PDB ID: 6DMU) (Figure 3.14). The density observed in the PI(3,5)P_2 structure may correspond to a lipid molecule (Figure 3.12B & 3.14B), although the presumed head group region of the density is not clear enough to unequivocally interpret it as a PIP molecule. With the PI(4,5)P_2 higher resolution structure, the density resembles two detergent molecules (Figure 3.13A), consistent with the observation from the PIP strips that the PI(4,5)P_2 appears to bind less well to PC2 and may not be able to displace high concentrations of detergent. Although the PI(3,5)P_2 density map suggests that a lipid molecule may be bound at this site, there are few

interactions between the region where the head group lies and the potential lipid. Overall the lipid-like density is insufficiently clear to unambiguously assign a given lipid molecule, so no lipid was placed at this site in the final PDB file.

In other TRP channel structures with lipid molecules built into similar locations to the potential PIP site we observed e.g. the studies of TRPV1 [6] and TRPV5 [53], the lipid head groups are coordinated not only by residues on S3, S4 and S5 helices, but also by residues in the cytosolic pre-S1 or post-S6 domains. The lack of cytosolic domains in our construct may contribute to the flexibility of the head group of the bound lipid, helping to explain why the densities observed for these structures are not well resolved in the lipid head group region. Similarly, there are some undefined densities in this region in structures obtained using soy extract polar lipids (which are likely to contain some phosphatidylinositides) [13]. We inspected the cryo-EM density map of the latter structure (EMD-8354), which was determined in nanodiscs [13]. In the region of interest, there is density which could represent a bound phospholipid molecule. Together with our data, this suggests that the hydrophobic pocket formed between S3, S4 and S5 helices of the VSLD is a phospholipid binding site. In addition, with the more soluble and potentially higher-affinity version of PIP_2, there is a small outward movement ($\sim$ 2 Å) of the VSLD at the S2-S3 linker region and an inward movement ($\sim$ 2 Å) towards the central axis of the cytosolic extension of S6 helix in the PI(3,5)P_2 structure compared to the PI(4,5)P_2 structure (Figure 3.14B), which may correlate with the apparently stronger binding of PI(3,5)P_2 seen in the PIP-strip assay.

3.4 Conclusions

Given the importance of lipid interactions to membrane proteins, we investigated lipid binding to the human TRP channel PC2. We solved two structures of PC2 in the presence of PI(4,5)P_2 and PI(3,5)P_2 to 2.96 Å and 3.4 Å resolution respectively. Although the protein samples were prepared in a similar manner as our original study (PDB ID: 5K47), the differences in grid types (copper grids

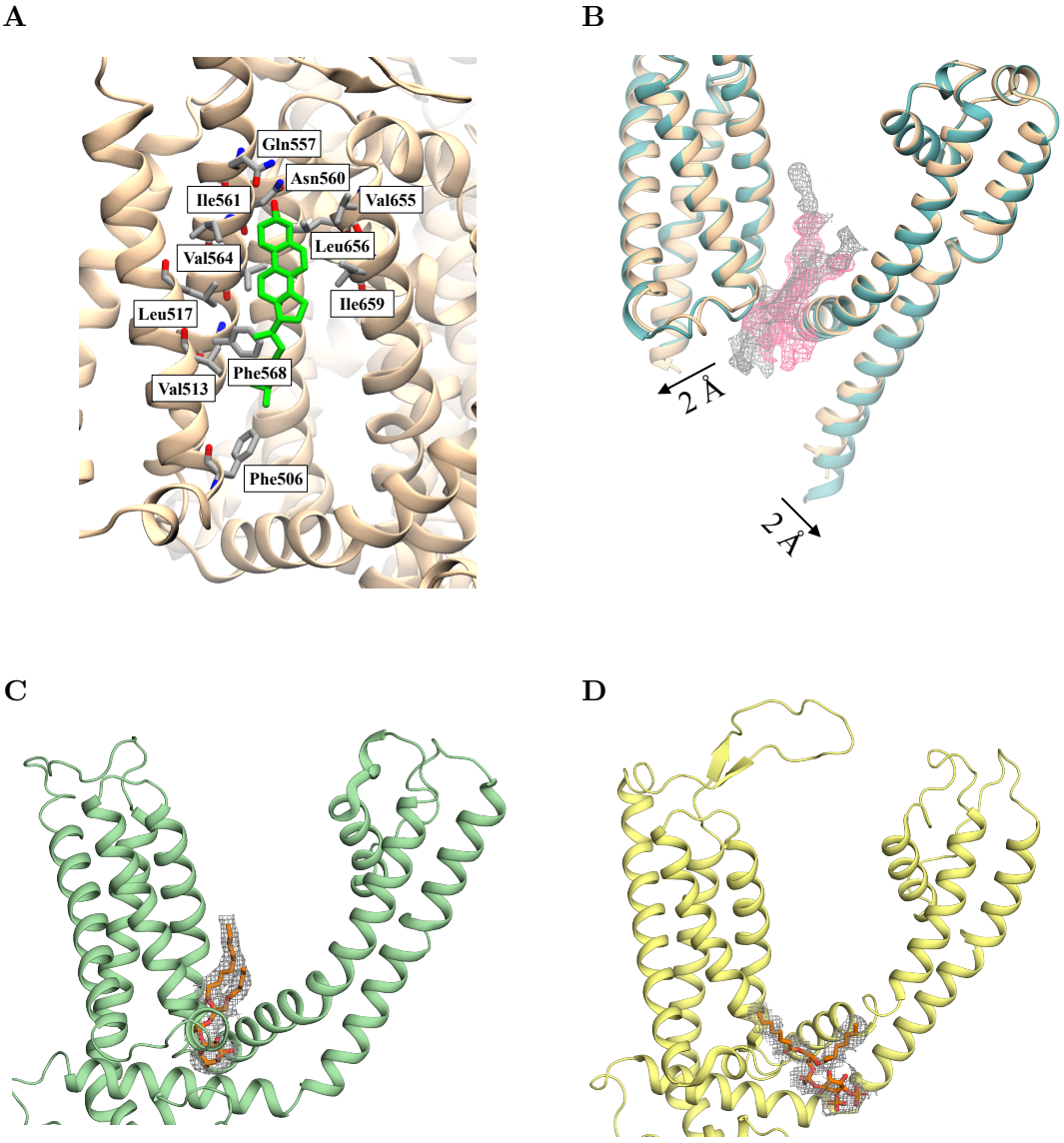


Figure 3.14: Lipid density comparison between PC2, TRPV1 and TRPV5. **A:** Detailed view of the cholesterol binding site revealed by cryo-EM structure of PC2/PI(4,5)P_2 showing key interacting residues. **B:** Expanded view around the potential lipid binding site between S3, S4 and S5 showing density from the 2.96 Å map (pink) and from the 3.4 Å map (grey). PC2/PI(4,5)P_2 structure is shown in light orange cartoon and PC2/PI(3,5)P_2 structure is shown in light teal cartoon. **C:** Phosphatidylinositol lipid in TRPV1 (PDB ID: 5IRZ; EMD-8118) **D:** PI(4,5)P_2 in TRPV5 (PDB ID: 6DMU; EMD-7966).

vs. carbon grids), protein concentrations (3.5−4 mg/ml vs. 4.5 mg/ml), electron microscopes (Titan Krios vs. Tecnai F30 Polara), electron dose rates ($\sim$ 6 electrons/Å/s vs. 9 electrons/Å/s), magnifications (165,000× vs. 37,037×) used and improvement in data processing softwares might collectively contribute to the enhancement of resolutions. Combining with the results from biochemical assays, we identified a cholesterol binding site on the outer-leaflet-facing surface of PC2, and a potential PIP_2 binding pocket between S3, S4 and S5 helices. A distinct PIP_2 site in TRP channels has been proposed, formed between the Pre-S1 and TRP-like domains [251], although the structural basis for this remains uncertain. However, a comparable PIP_2 site was observed in TRPV5 [53], and a PIP_2 site between helices S1, S2 and S3 was seen for TRPML1 [78]. It seems possible that more than one PIP lipid binding site may be present on TRP channels (see also [160]), which could explain some of the apparent diversity of effects (both positive and negative) of PIPs on these channels [221].

Even though the resolution achievable by cryo-EM has improved dramatically, the heterogeneity brought by subtle difference in lipid/detergent densities might be masked by averaging over large amount of particles. Therefore, cryo-EM alone is still difficult to provide definitive answer to lipid binding on membrane proteins. Other methods such as molecular simulations could be valuable complements in studying protein-lipid interactions.

It has been suggested that the basal regions of primary cilia membranes contain PIP_2 whilst the upper regions of the ciliary membrane have an elevated level of PI(4)P [231, 252]. Thus, it is reasonable to propose that PC2 binds to PIP_2 *in vivo*. In this context, it is of interest that OCLR1, a lipid phosphatase that converts PI(4,5)P_2 to PI(4)P, modulates the length of cilia in renal epithelial cells and loss of its function in Lowe syndrome is associated with progressive renal malfunction [253]. In addition, deletion of PC2 in mice leads to the formation of cilia that are more than 3-fold longer than normal cilia [146]. Thus, the interaction between PC2 and PIP lipids may play an important role in modulation of cilia formation.

4

Simulation Studies of Lipid Interactions of Polycystin-2 and Other Channels

Contents

4.1 Introduction 89
4.2 Methods 90
4.2.1 Cryo-EM Structure-based Molecular Models 90
4.2.2 Coarse-grained Simulation Details 94
4.2.3 Atomistic Simulation Details 96
4.2.4 PMF Calculations 98
4.2.5 Simulation Visualisation and Analysis 99
4.3 Results and Discussion 99
4.3.1 Identification of a Phospholipid Interaction Site 99
4.3.2 Free Energy Landscapes for PC2/Lipid Interactions . . 104
4.3.3 The PIP_2 Binding Site 113
4.3.4 PIP_2 Interactions in TRP Channels Revealed by CG Simulations 115
4.3.5 Atomistic Study of Phosphoinositide Binding Specificity 117
4.3.6 A Cholesterol Binding Site Revealed by Cryo-EM and Simulations 121
4.4 Conclusions 125

This chapter is partially based upon the manuscript:

"Lipid Interactions of a Ciliary Membrane TRP Channel: Simulation and Structural Studies of Polycystin-2 (PC2)", Qinrui Wang, Robin A. Corey, George Hedger, Prafulla Aryal, Mariana Grieben, Chady Nasrallah, Agnese Baronina, Ashley Pike, Jiye Shi, Elisabeth P. Carpenter, and Mark S.P. Sansom, submitted to Structure.

4.1 Introduction

Molecular dynamics (MD) simulations provide an important 'discovery tool' for lipid interactions with membrane proteins [254, 255]. They allow integral proteins to be placed in a solvated lipid bilayer, and thus their dynamic behaviour can be characterised within its native environment. This in turn allows the interactions of lipids with membrane proteins to be explored. For example, they have been successfully used to predict lipid binding sites on Kir channels [176, 256], GPCRs [165], epidermal growth factor receptor [257] and aquaporins [258]. With the advances in cryo-EM, the number of structures of membrane proteins has been growing rapidly. However, it is still difficult to unequivocally identify bound lipids and their binding affinity with structural studies alone. The potential of mean force (PMF) obtained in MD simulations provides a way to quantify binding affinities and specificity of interactions between membrane proteins and lipids *in silico* [250, 259, 260], and thus provides an indication of possible mechanisms of activation and allosteric modulation of membrane proteins by lipids. PMF calculations have been successfully used to explore interactions of anionic lipids (e.g. cardiolipin and PIP_2) with a number of transporters (e.g. ANT1 [260]), ion channels (e.g. Kir channels [259]) and receptors (e.g. class A GPCRs [165, 261]). In order to obtain accurate free energy profiles (i.e. PMF), sufficient sampling of the conformational space along a reaction coordinate of interest is necessary. Advanced sampling techniques

such as adaptive force biasing and umbrella sampling and/or reduction in the granularity of the system of interest are usually employed to aid the exploration of the energy surface.

With our cryo-EM studies and the previously published structures of PC2, we have had detailed knowledge about the static atomic structure of PC2 in a closed state. These structures are remarkably similar with an unresolved flexible cytosolic domain, despite the fact that they were solved in a range of environments (detergent, nanodiscs and amphipols) and with both truncated and full length protein. However, none of these structures allows unambiguous interpretation of the bound lipids. Given that the majority of lipid interaction sites are on the outer surface of membrane proteins, where the local resolution is relatively low comparing to the protein core, it is often difficult to determine the nature of the bound lipids with cryo-EM. Molecular simulations could serve as a complemental method in identifying possible bound lipids. In this study, we used multiscale MD simulations and umbrella sampling to further explore and validate the lipid binding sites revealed by cryo-EM (see Chapter 3), and to establish the specificity of these binding sites in the context of biological membranes.

4.2 Methods

4.2.1 Cryo-EM Structure-based Molecular Models

The initial molecular coordinates of human PC2 used for simulations were obtained from cryo-EM structures (our recent PC2/PI(4,5)P_2 structure (see Chapter 3), and also previously published structures PDB ID: 5K47, 5T4D, 5MKF, 6D1W). Multiple molecular models were used to evaluate the convergence of our simulations with related structures. Experimental structures of 5K47 and 5T4D were solved on truncated constructs of PC2. Structure 5MKF was based on full length PC2, and the structure 6D1W with relatively more open lower gate is from a gain-of-function mutant construct (Phe604Pro) of PC2. Comparative studies within the TRP superfamily of ion channels were done on cryo-EM structures

of PKD2L1, TRPA1, TRPC3, TRPM2, TRPM4, TRPML1, TRPML3, TRPV1, TRPV2, TRPV5 and TRPV6 (see Table 4.1). Simulations investigating lipid specificity were also performed on the G protein-gated inward rectifier potassium channels Kir2.2 (PDB ID: 3SPI) and Kir3.2 (PDB ID: 3SYA). Bound N-acetyl-d-glucosamines, lipids and ions were removed from all protein structures. Missing side chains and broken loops were modelled in using MODELLER version 9v10[262]. Models with the lowest value of the MODELLER objective function were visually inspected and the most structurally sound models were chosen for subsequent simulations. All simulations performed in this chapter are summarised in Table 4.1 and Table 4.2 .

Table 4.1: Summary of Equilibrium Simulations

Channel	**PDB ID**	**Bilayer Composition**	**Replicates× Duration (μs)**
Atomistic			
PC2	PC2/PI(4,5)P_2	POPC + CHOL	3×0.25
PC2	PC2/PI(3,5)P_2	POPC + CHOL	3×0.25
PC2	5K47	POPC	3×0.2
PC2	5T4D	POPC	3×0.25
PC2	5MKF	POPC	3×0.25
PC2/F604P	6D1W	POPC	3×0.25
PC2	5K47	POPC + PIP_2	3×0.2 to 0.25
Kir2.2	3SPI	POPC + PI(3,4)P_2 + PI(4,5)P_2[1] + PI(3,4,5)P_2	6×0.25
Coarse Grained			
PC2	5K47	Mixed lipid bilayer[2]	3×5
PC2	6D1W	Mixed lipid bilayer	3×5
PKD2L1	6DU8	Mixed lipid bilayer	3×5
TRPA1	3J9P	Mixed lipid bilayer	2×5
TRPC3	6CUD	Mixed lipid bilayer	3×5
TRPM2	6C07	Mixed lipid bilayer	3×5
TPRM4	6BQV	Mixed lipid bilayer	3×5
TRPML1	5WJ5	Mixed lipid bilayer	3×5
TRPML1	5WJ9	Mixed lipid bilayer	3×5
TRPML3	6AYF	Mixed lipid bilayer	3×5
TRPV1	3J5P	Mixed lipid bilayer	2×10
TRPV2	5AN8	Mixed lipid bilayer	3×5
TRPV5	6B5V	Mixed lipid bilayer	3×5
TRPV6	6B08	Mixed lipid bilayer	3×5

1. One copy of PI(4,5)P_2 is fully ionised with charge of -5, and the other is protonated with charged of -4.
2. The mixed lipid bilayer contained POPC:POPE:SM:GM3:CHOL = 40:10:15:10:25 in the outer leaflet and POPC:POPE:POPS:PIP_2:CHOL = 10:40:15:10:25 in the inner leaflet.
3. POPC: 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine; POPE: 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine, SM: sphingomyelin; GM3: monosialodihexosylganglioside; CHOL: cholesterol; POPS: 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylserine.

Table 4.2: Summary of Non-equilibrium Simulations

Channel	PDB ID	Bilayer Composition (Bulk/Bound)	Duration per Window (μs)	No. of Windows	Window Spacing (nm)	Pulling Group
Umbrella Sampling Simulations Using PLUMED2						
PC2	5K47	POPC/POPC	2	38	0.1	Head group
PC2	5K47	POPC/POPS	2	43	0.05 to 0.1	Head group
PC2	5K47	POPC/PIP_1	3	38	0.1	Head group
PC2	5K47	POPC/PIP_2	3	38	0.1	Head group
PC2	5K47	POPC/PIP_3	3	38	0.1	Head group
PC2	5K47	POPC/POPC	1.5	41	0.1	PO4 bead
PC2	5K47	POPC/POPS	2	38	0.1	PO4 bead
PC2	5K47	POPC/PIP_2	2	38	0.1	RP* beads
Kir3.2	3SYA	POPC/POPC	1.5	37	0.05 to 0.1	PO4 bead
Kir3.2	3SYA	POPC/truncated POPC	0.5	37	0.05 to 0.1	PO4 bead
Kir3.2	3SYA	POPC/PIP_2	1.5	33	0.1	PO4 bead

4.2.2 Coarse-grained Simulation Details

The side chain protonation states of all proteins were modelled using pdb2gmx implemented in GROMACS (www.gromacs.org). Before converting the coordinates into coarse-grained (CG) representations, the structures were energy minimised using the steepest descents algorithm. All models were then converted to CG representations using the martinize.py workflow with dihedral restraints automatically assigned. Equilibrium CGMD simulations of PC2 structures before conversion to atomistic systems were performed using GROMACS version 4.6 [263] with MARTINI version 2.1 force field [211]. All other CG simulations were done using GROMACS version 5.1 [264] with the MARTINI version 2.2 force field [265]. An *ElNeDyn* elastic network [217] was applied with an elastic bond force constant of 500 kJ/mol/nm^2 and an upper elastic bond cut-off of 0.9 nm. The standard MARTINI water particles with van der Waals radii of 0.21nm were used to solvate all systems, which were then neutralised with NaCl at a concentration of 0.15 M. Parameters for lipids and ions were those described in [266]. Phosphatidylinositol phosphate lipids and monosialodihexosylganglioside (GM3) parameters were created in-house as previously described [176, 267].

Each simulation system contains only one copy of the protein. Initially, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipids were randomly placed around the proteins. Correct positioning of the protein in a lipid bilayer was achieved by a 100 ns self-assembly simulation (Figure 4.1) [268]. The CG system for PDB ID: 5K47 was equilibrated for 400 ns after the initial self-assembly. Simulation systems for the other PC2 structures (i.e. PDB ID: 5MKF and 5T4D) were set up by aligning and replacing the protein in the PDB ID: 5K47 POPC-only system, and then equilibrating for 1 μs. Systems used for mixed-lipid simulations were set up by exchanging POPC molecules for other lipids to construct an asymmetric *in vivo* mimetic bilayer with outer (i.e. extracellular or organelle luminal) leaflet containing POPC:POPE:SM:GM3:CHOL = 40:10:15:10:25; and

the inner (i.e. cytoplasmic) leaflet containing POPC:POPE:POPS:PIP_2:CHOL = 10:40:15:10:25 [269, 270].

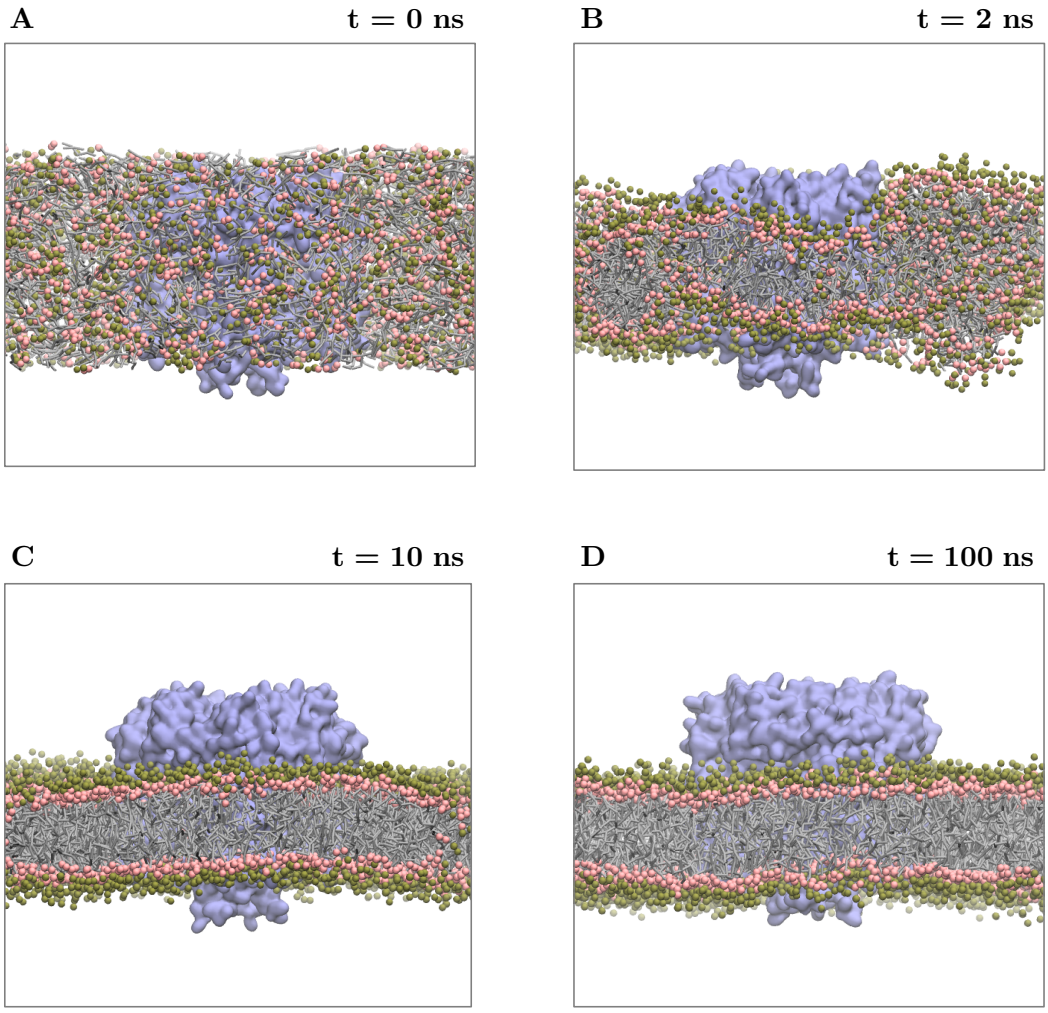


Figure 4.1: Schematic for CG self-assembly simulation. POPC lipids are shown as grey sticks for tails with pink spheres for glycerol particles and tan spheres for phosphate and choline particles. The protein PC2 is shown in iceblue. Water and ions are omitted for clarity.

All CGMD simulations were performed at a temperature of 310 K and a pressure of 1 bar. V-rescale thermostat [203] was used to maintain the temperature with a coupling time constant of $\tau_t = 1$ ps. Protein, lipids, and solvent (water + ions) were coupled separately to the temperature bath. For self-assembly and equilibration

simulations, pressure was controlled by semi-isotropic pressure coupling with the Berendsen barostat [204] with a compressibility of 5×10^{-6}/bar and a time constant of $\tau_p = 4$ ps. Production runs utilised a Parrinello-Rahman barostat [205] with coupling constant of 12 ps and compressibility of 3×10^{-4}/bar for pressure control.

4.2.3 Atomistic Simulation Details

Atomistic simulations of PC2 in a POPC-only bilayer were setup and run using a two-step multiscale procedure [271, 272], starting with a CG simulation to stably insert the protein in a lipid bilayer (see Subsection 4.2.2), followed by conversion to an atomistic representation and subsequent atomistic simulations. A fragment based protocol [272] was used for the CG to atomistic conversion. For the PDB ID: 5K47 system, the final frame of the 400 ns equilibration was converted to atomistic representation. Three repeats of 200 ns atomistic simulations with different initial random seeds were performed after a 6 ns unrestrained equilibration. Simulations were carried out using GROMACS version 4.6 with the CHARMM36 force field [273]. The water model used was TIP3P [274]. Atomistic systems for PDB ID: 5MKF and 5T4D in POPC bilayers were converted from the final frame of the 1 s CG equilibration runs. Three repeats of 250 ns unrestrained atomistic simulations were done by randomising initial velocity of the system for both structures after 1 ns fully restrained (force constant = 1000 kJ/mol/nm^2) equilibration.

Atomistic conversion of the PC2-PIP_2 complex corresponding to the energy minimum in the PMF was done on the final snapshot of the sampling window for the minimum energy configuration. The initial system was energy minimised and equilibrated for 2 ns with position restraints on all atoms of the protein. Then the protein restraints were relaxed to the C_α atoms for 10 ns (force constant = 1000 kJ/mol/nm^2) with harmonic restraints (force constant = 300 kJ/mol/nm^2) applied to the distances between the $1'$, $4'$, and $5'$ phosphates of PIP_2 and the side chains of Arg592, Lys575 and Lys572 respectively. The distance constraint was set to be 6 Å between each pair. A further 30 ns simulation was performed with all protein restraints removed and protein-lipid distance restraints remained.

Unrestrained production simulations of the system was then done in triplets for 200 to 250 ns each using the GROMOS96 53a6 united-atom force field [275] with SPC water model [202] by randomising initial velocity of the system.

An atomistic simulation system of Kir2.2 (PDB ID: 3SPI) with different species of phosphatidylinositol phosphate lipids at each of the crystallographic PIP_2 binding site was set up using the CHARMM-GUI (http://www.charmm-gui.org/).

Systems of PC2 structures with cholesterols were setup using CHARMM-GUI (www.charmm-gui.org). The protein with bound cholesterols were embedded in POPC-only bilayer. Simulations were performed using GROMACS version 5.1 with the CHARMM36m [276] force field and TIP3P water model [274]. Equilibration of the system was done in six steps with gradually decreasing restraining force constants on the protein and cholesterol molecules. Durations of equilibration simulations and force constants applied are: step 1: 25 ps, 2000 kJ/mol/nm^2 on protein sidechain, and 4000 kJ/mol/nm^2 on protein backbone and cholesterols; step 2: 25 ps, 1000 kJ/mol/nm^2 on protein sidechain, and 2000 kJ/mol/nm^2 on protein backbone and cholesterols; step 3: 50 ps, 500 kJ/mol/nm^2 on protein sidechain, and 1000 kJ/mol/nm^2 on protein backbone and cholesterols; step 4: 500 ps, 200 kJ/mol/nm^2 on protein sidechain, and 500 kJ/mol/nm^2 on protein backbone and cholesterols; step 5: 500 ps, 50 kJ/mol/nm^2 on protein sidechain, and 200 kJ/mol/nm^2 on protein backbone and cholesterols; step 6: 1 ns, no restraints on protein sidechain, and 50 kJ/mol/nm^2 on protein backbone and cholesterols. Three repeats of 250 ns unrestrained atomistic simulations were then performed.

For all atomistic MD simulations, long-range electrostatic interactions (< 1 nm) were modelled with the Particle Mesh Ewald (PME) method [277]. Temperature coupling was done with V-rescale thermostat at 310 K. The Parrinello-Rahman barostat [205] with a reference pressure of 1 bar and a compressibility of 4.5×10^{-5}/bar was applied for pressure control. Covalent bonds are constrained to their equilibrium length by the LINCS algorithm [200]. The integration steps of all simulations were set to 2 fs.

4.2.4 PMF Calculations

CG models of PC2 and of Kir3.2 (PDB ID: 3SPI) were embedded in a POPC bilayer using self-assembly simulations, and equilibrated for 1 μs. To reach better computational efficiency, protein truncations were done on the equilibrated CG protein-lipid systems, removing the TOP domain (Ser244 − Leu462) of PC2 with the S3-S4 insertion into the TOP domain intact, and removing the intracellular domain (Ile55 − Arg73, Leu205 − Asn382) of Kir3.2. Further equilibrations were performed with unbiased MD simulations for 100 ns. Position restraints with a force constant of 1000 kJ/mol/nm^2 were used to prevent protein translation during equilibration.

For PC2, the reaction coordinates were defined as the distance between the centre of mass (COM) of Ser505 and Ser591 of PC2 and the COM of the pulling group of the lipids (Table 4.2). Umbrella sampling windows were spaced on the reaction coordinates between −0.7 and 3 nm with a force constant of 500 kJ/mol/nm^2. Position restraints (400 kJ/mol/nm^2 in the xy plane) were applied to the backbone beads of Ser505 and Ser591 in each subunit to prevent rotation and translation of the protein. The reaction coordinates are approximately parallel to the X axis of the simulation box and a weak positional restraint of 100 kJ/mol/nm2 was applied to the corresponding pulling groups of the lipids to limit its movements along the Y axis.

For Kir3.2, the reaction coordinates were defined as the distance between the COM of Lys90, Lys194, Lys199 and Lys200 of Kir3.2 and the COM of the pulling group (Table 4.2). Spacing of window separations were defined between the distance of 0 and 3 nm with a force constant of 500 kJ/mol/nm^2. Lateral movements of the protein within the plane of membrane bilayer were restrained by applying position restraints on the backbone beads of the four lysines with a force constant of 400 kJ/mol/nm^2. The reaction coordinates were defined in the direction of the Y axis. Diffusion along X axis was prevented with a force constant of 100 kJ/mol/nm^2.

The PLUMED2 package [278] was used to define the reaction coordinates and apply biasing potential to pull the lipids. WHAM were used to unbiase all umbrella sampling simulations. For systems of all PIPs and POPS, the dissociation study was done on one of the four bound lipid molecules. For POPC system, the protein was placed in pure POPC bilayers, and one POPC molecule situated in one of the binding sites was chosen to be pulled away from the proteins. To get well converged estimations of the binding energies, the first 800 ns of each window was discarded for all PIP systems, and the first 500 ns for the POPC and POPS systems.

4.2.5 Simulation Visualisation and Analysis

Protein structures were visualised with VMD [279] or PyMOL [280]. Simulation trajectories were processed using GROMACS. Averaged lipid density maps were generated using the VMD Volmap plugin tool with three-dimensional grids every 0.1 nm.

4.3 Results and Discussion

4.3.1 Identification of a Phospholipid Interaction Site

An initial exploration of possible lipid interaction sites on the transmembrane domain of PC2 was made using atomistic molecular dynamics simulations in which the PC2 structure (PDB ID: 5K47, as a representative of several PC2 structures, see following) was embedded in a lipid bilayer made up of a single species of phospholipid (palmitoyl-oleyl-phosphatidylcholine, POPC; Figure 4.2A). This process was repeated for three previously published molecular structures of wild-type full-length or truncated PC2 (PDB ID: 5K47, 5MKF, 5T4D), a gain-of-function mutant (F604P) of PC2 (PDB ID: 6D1W) [239] and for both of our PC2/PI(4,5)P_2 and PC2/PI(3,5)P_2 structures in current study, yielding a total of more than 4 μs of atomistic simulations of PC2 in a POPC bilayer. The simulations were examined in terms of regions of high probability density of occurrence of lipid molecules on the protein surface. In all 18 simulations (i.e. 3 repeats for each of the six structures,

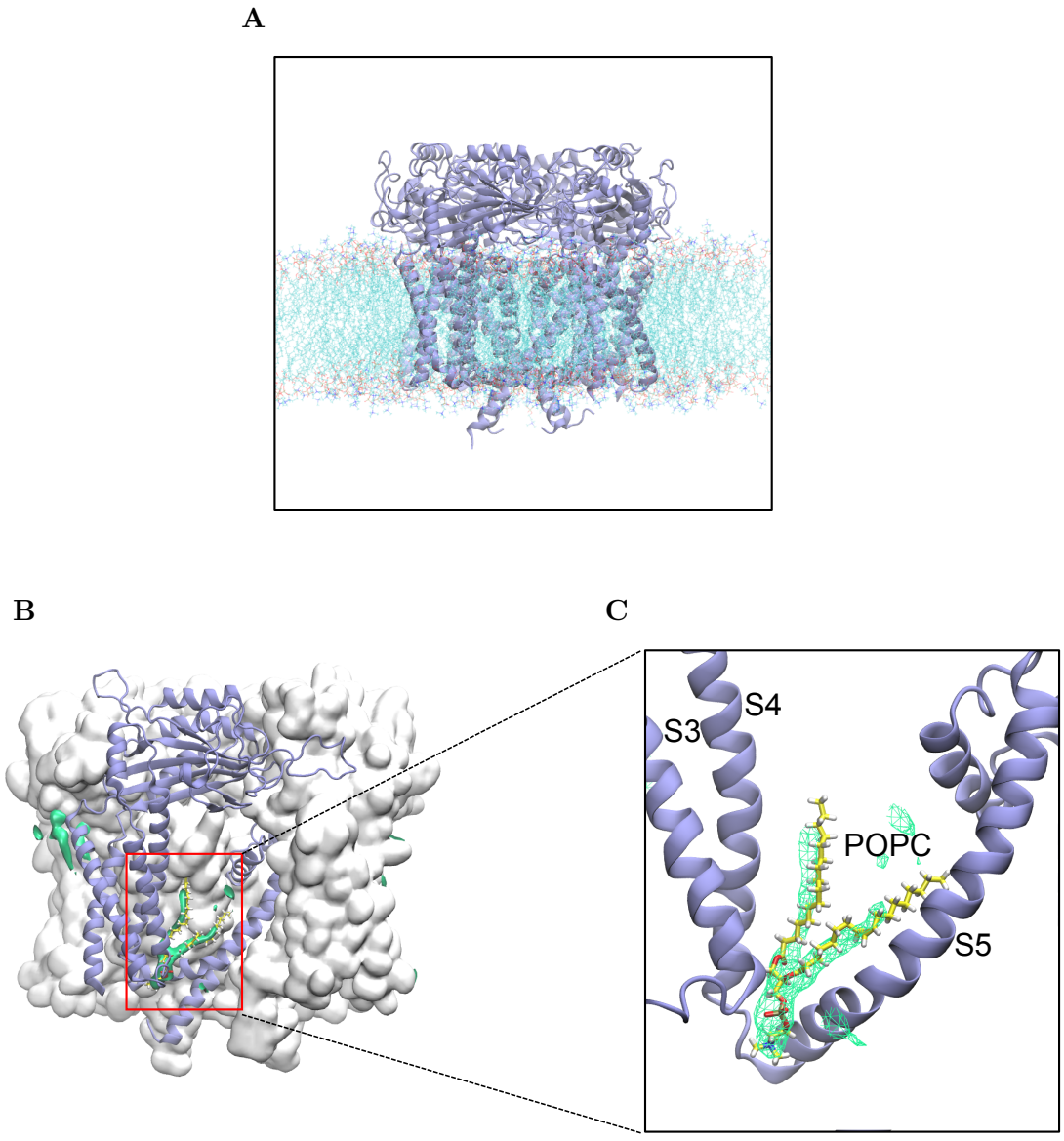


Figure 4.2: A possible phospholipid interaction site identified in atomistic simulations of PC2 in a POPC bilayer. **A:** The PC2 channel (PDB ID: 5K47) embedded in embedded in a phospholipid bilayer, shown using a snapshot from an atomistic simulation of the protein (colour) in a POPC bilayer. The lipid tails are in cyan, phosphates in orange and red, and choline nitrogens in blue. Water molecules are omitted for clarity. **B:** The PC2 protein is shown as a grey surface, viewed perpendicular to the central pore axis, with one subunit depicted as a pale purple cartoon. Green isocontour surfaces represent a high probability of occurrence of phospholipid molecules. **C:** Zoomed in view (red box in B) of the S3/S4/S5 pocket and the high phospholipid occurrence density with a POPC molecule (taken from a simulation snapshot) shown within the density.

PC2/PI(4,5)P_2, PC2/PI(3,5)P_2 structures and PDB ID: 5K47, 5MKF, 5T4D, 6D1W), high lipid occurrence densities (Figure 4.2B) were observed in a pocket exposed to the intracellular leaflet of the lipid bilayer, between transmembrane helices S3, S4 and S5 (Figure 4.2C), corresponding to one POPC lipid molecule bound to each subunit of the PC2 tetramer. These results are illustrated for 5K47 in Figure 4.2B and 4.2C. Similar results for 5MKF, 5T4D and 6D1W are shown in the Appendix C, Figure C.1. This probability density revealed by simulations is in the same location as the putative lipid/detergent density in our cryo-EM studies (see Figure 3.14). The hydrophobic side chains from residues in S3, S4 and S5 create a hydrophobic pocket, within which the acyl tails of the bound lipid molecules reside. The phosphate oxygens of the bound lipid formed hydrogen bonds to the indole ring of Trp507 in S3 and to the hydroxyl group of Ser591 in the S4-S5 linker (Appendix C, Figure C.2).

However, these hydrogen bonds are dynamic and do not persist throughout the simulations, and the lipid head group is not held firmly in that position, which confirms our hypothesis for the reason why there is a lack of clear head group density in both of our cryo-EM datasets. Although the bound lipid remained in the same pocket throughout all simulations, it is possible that this may be a consequence of the short timescale. To investigate this issue, the X-axis coordinates of three randomly chosen upper leaflet lipids and three lower leaflet lipids were compared with all bound lipids in the four subunits of PC2 over time to examine their lateral movement within the bilayer. As illustrated in Figure 4.3, in both atomistic simulation and coarse-grained (CG) simulation before CG to atomistic conversion, the variations in lipid positions for unbound lipids are considerably larger than bound lipids, which validate the hydrophobic pocket as a lipid-binding site. Thus, in a simple model lipid bilayer, we observe a phospholipid binding site on PC2, which is close to the proposed lipid and vanilloid (e.g. capsaicin) binding sites observed in cryo-EM studies of the related TRPV1 and TRPV2 channels [6, 16]. Lipid binding has been observed at a similar location in cryo-EM structures of TRPV6 [281, 282], TRPV5 [53], TRPC3 [232], TRPC4 [234], TRPM4 [283,

284], TRPML3 [77] and TRPM7 [235]. This site has also been suggested to be the binding site for the TRPV5 inhibitor econazole [285], and for the TRPML1 lipid agonist PI(3,5)P_2 [79].

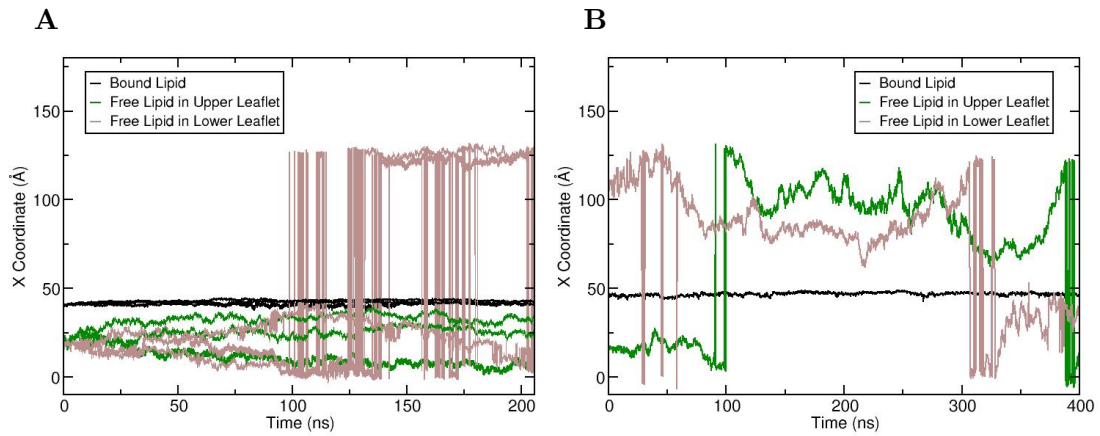


Figure 4.3: Lateral movement of bound and free lipids. **A:** The X-axis coordinates of bound and free lipids in all three repeated 200 ns atomistic simulations showing the lateral movements of these lipids within the bilayer. Only one of the four bound lipid and one free lipid from each upper and lower leaflet were shown here. **B:** The X-axis coordinates of the same lipids in the 400 ns CG simulation of the same system before conversion to atomistic level.

We subsequently used CG simulations of PC2 in a lipid bilayer containing multiple lipid species [270, 286] to explore the possible specificity of the lipid binding site. We embedded PC2 in an *in vivo* mimetic bilayer (Figure 4.4), which contained the anionic lipids phosphatidylserine (POPS) and PIP_2 in the inner leaflet, a glycolipid (GM3) in the outer leaflet, and cholesterol (CHOL) in both leaflets of the bilayer. The lipid composition of the *in vivo* mimetic bilayer membrane provided an approximation to a mammalian plasma membrane [269, 270]. Thus, the outer (i.e. extracellular or organelle luminal) leaflet contained POPC:POPE:SM:GM3:CHOL = 40:10:15:10:25; and the inner (i.e. intracellular) leaflet contained POPC:POPE:POPS:PIP_2:CHOL = 10:40:15:10:25). This provides an overall PIP_2 concentration of 5% which is within the physiological range for a mammalian plasma membrane [1, 269].

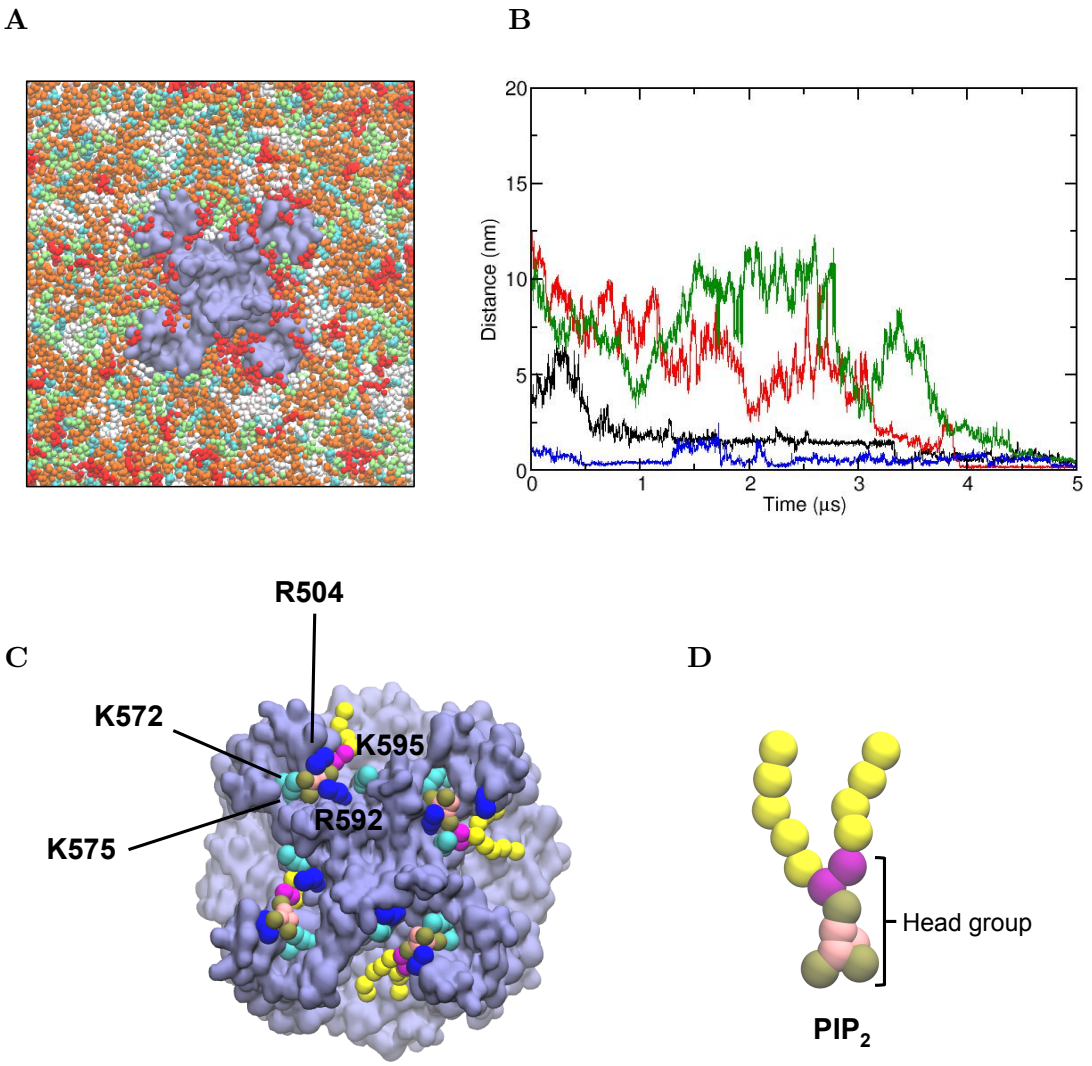


Figure 4.4: Coarse-grained (CG) simulations of PC2 in an *in vivo* mimetic mixed lipid bilayer. **A:** PC2 (pale purple) in a mixed lipid bilayer, viewed from the intracellular face and showing molecules of PIP_2 (red), cholesterol (cyan), PC (white), PE (orange) and PS (lime) in the inner leaflet of the bilayer. **B:** Distance from the binding site as a function of time for four PIP_2 molecules, which bind to PC2 during a simulation in a mixed lipid bilayer. The distance is from the centre of mass of the head group of each PIP_2 molecule to the centre of mass of the Ser505 and Ser591 side chains at the site to which that lipid molecule eventually binds. The four colours correspond to the four different PIP_2 molecules which eventually bind. **C:** Snapshot from the end of the simulation shown in **A** showing PIP_2 molecules (head group phosphate particles in brown, as shown in the CG representation of PIP_2 on the right) at four sites on the channel tetramer. Side chain particles of basic residues at the PIP_2 binding site are shown in blue (arginine) and cyan (lysine). **D:** CG representation of PIP_2 as shown in **C**. The head group region used for the distance calculations in **B** is indicated.

Three independent simulations (each of 5 μs duration) of a single PC2 channel (PDB ID: 5K47 & our PC2/PI(4,5)P_2 structure) inserted in an *in vivo* mimetic lipid bilayer with different random distributions of lipid molecules were performed. Same type of simulations were also performed for the constitutively active F604P PC2 mutant (PDB ID: 6D1W). In each of the simulations, PIP_2 molecules diffused in the bilayer on a timescale of microseconds (Figure 4.4B, Appendix C, Figure C.3 & C.4) resulting in random encounters with the channel molecule followed by binding to the previously identified lipid-binding sites on PC2, as demonstrated by tracking the distance vs. time of the PIP_2 molecules from their eventual binding sites (Figure 4.4B), which is defined by the centre of mass of Ser505 and Ser591. From a final snapshot of one simulation (Figure 4.4C), it can be seen that a PIP_2 molecule has bound to each of the four sites on PC2. The head groups of the PIP_2 molecules each interact with up to five basic residues in the S3/S4/S5 region: Arg504, Lys572, Lys575, Arg592 and Lys595 , which are highly conserved between PC2 in different species (Figure 4.5). This interaction is persistent in both wild type and mutant PC2. Given the presence of other negatively charged lipid species (POPS) in the bilayer and of positively charged patches on the protein surface, the observation that PIP_2 molecules were able to bind to the intracellular binding site suggests that this is specific for PIP_2 and/or related lipids.

4.3.2 Free Energy Landscapes for PC2/Lipid Interactions

Although atomistic simulations allow us to explore detailed protein-lipid interactions, they fail to readily provide information on the specificity or energetics of binding given the high computational cost associated with atomistic MD simulations and slow diffusion of membrane embedded molecules [259]. To explore the energetics and selectivity of the PIP_2/intracellular interaction site and the best binding position of lipids at the binding site, we calculated potentials of mean force (PMFs) based on CG simulations. A PMF provides a one-dimensional free energy landscape for protein-lipid interactions, and allows us to explore the lipid specificity of these interactions.

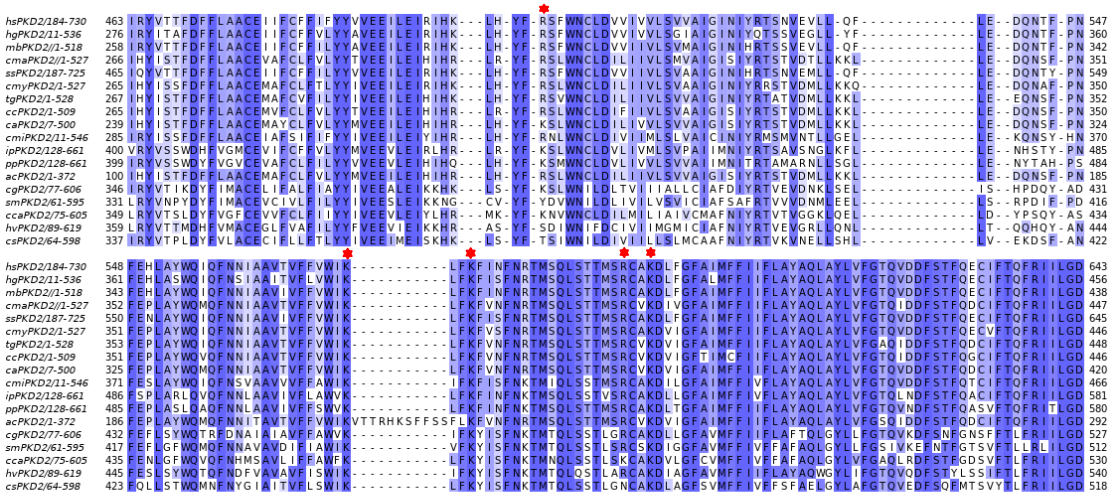


Figure 4.5: Sequence alignment for PC2 from 18 different species. The sequence of human PC2 (M184-Q730) (hsPKD2) is aligned with the corresponding sequences from *H. glaber* (hg), *M. brandtii* (mb), *C. macqueenii* (cma), *S. scrofa* (ss), *C. mydas* (cmy), *T. guttatus* (tg), *C. canorus* (cc), *C. anna* (ca), *C. milii* (cmi), *I. punctatus* (ip), *P. prolifica* (pp), *A. chloris* (ac), *C. gigas* (cg), *S. mimosarum* (sm), *C. candelabrum* (cca), *H. vulgaris* (hv), *C. sinensis* (cs). Residues Arg504, Lys572, Lys575, Arg592 and Lys595, identified to coordinate the PIP_2 headgroup in human PC2, are highlighted by red stars.

In order to meaningfully assess the energy of binding, a reaction coordinate that extends across different energy states needs to be chosen. That is, it should cover a route between the state of interest (e.g. the bound state, with lipid residing in the binding pocket) and a reference state (e.g. the free state, with lipid residing in the bulk lipids). To get the best sampling of the energy landscape along the reaction coordinates, various parameters for umbrella sampling such as restraints, spacing of window separation and the biasing potential need to be determined. Since the potential binding sites of PC2 are relatively unexposed to the bulk lipids, a more straightforward and well characterised model of lipid binding, the G protein-gated inward rectifier potassium channel Kir3.2 (PDB ID: 3SYA), was used as reference. The crystal structure of Kir3.2 has shown a bound PIP_2 molecule at each subunit of the homo-tetrameric ion channel. For computational efficiency, both of the PC2 and Kir3.2 structures were truncated, removing the TOP domain (Ser244 − Leu462) of PC2 and the intracellular domain (Ile55 − Arg73, Leu205 − Asn382)

of Kir3.2, including one lysine residue (Lys64) that is suggested to contribute to the coordination of the bound PIP_2 in the structural study [156] (Figure 4.6). Weak position restraints in the xy plane were applied to the backbone beads of Ser505 and Ser591 for PC2 and of Lys90, Lys194, Lys199, Lys200 for Kir3.2 to prevent translation and rotation of proteins within the plane of membrane bilayer to allow better definition of the reaction coordinates. Simulations performed for the protein-lipid binding free energy estimation is listed in Table 4.2.

To estimate the free energy landscapes for different lipids interacting either with PC2 at the intracellular site defined by S3/S4/S5 or with Kir3.2 at the crystallographic PIP_2 binding site, CG models of a PC2 or a Kir3.2 channel molecule was embedded in a POPC bilayer with one copy of the lipid of interest inserted into each of the four binding sites. The free energy profiles of POPC, palmitoyl-oleoyl phosphatidylserine (POPS), PIP_2 were calculated for both proteins. To test the contribution of the hydrophobic effect of the tail groups to the binding free energies, PMFs were computed for a truncated POPC model, where only one tail bead on each tail group was retained. For PC2, the free energy of binding for phosphatidylinositol monophosphate (PIP_1) and phosphatidylinositol trisphosphate (PIP_3) are also calculated to test if the potential binding site on PC2 has different affinity towards different species of phosphatidylinositol lipids to the extend that CG models could describe.

Initially, a window spacing of 0.1 nm was applied to all systems. In order to derive a continuous energy function, the overlap of neighbouring windows was assessed by histograms of configurations, and additional windows at 0.05 nm separations were added to where overlap was poor. As a first exploration, the group of particles that umbrella biasing potential was applied was the centre of mass the PO4 bead defining the phosphate group of POPC or POPS, or of the RP1, RP2 and RP3 beads defining the inositol ring of PIP_2. Since restraining the lipid positions by applying biasing potential on a small number of particles of the CG lipids limits the area of the free energy surface that lipids could sample, for PC2, to allow the lipid of interest, especially the ones with relatively large head

groups, to have higher degrees of flexibility in exploring the binding conformations, the umbrella biasing potential was extended to the centre of mass of the entire head group of the lipids (POPC/POPS: PO4, GL1, GL2 and NC3/NCO beads; PIP_1/PIP_2/PIP_3: (PO2, PO3)/(PO1, PO2, PO3)/(PO0, PO1, PO2, PO3), RP1, RP2, RP3, GL1 and GL2 beads).

With the lipid bound states as the initial configurations, one of the bound lipid molecules was pulled away from the binding site to assess the free energy of interaction. For PC2, an one-dimensional reaction coordinate was defined as the distance in the plane of the bilayer between the centres of mass (COM) of the lipid particles that umbrella biasing potential were applied and of two serines (Ser505, Ser591) within the lipid binding site of PC2 (Figure 4.6C). Either pulling on the inositol ring or on the entire head group, the free energy profile (i.e. PMF) for PIP_2 shows a clear minimum close to the initial position of the PIP_2 molecule in the binding site, with binding energies converged to approximately the same scale with well depths of ~ -33 kJ/mol and ~ -37 kJ/mol respectively (Figure 4.7A & 4.7C). Comparison of PMF profiles as a function of window simulation duration suggested that convergence had been achieved within 1.5 μs per window and that errors on the well depths of the PMFs are of the order of $\pm$ 3 kJ/mol (Figure 4.8). PIP_1 and PIP_3 exhibit weaker binding to PC2 with free energies of -28 kJ/mol and -20 kJ/mol respectively. In contrast, the anionic lipid POPS showed significantly weaker binding to PC2 with well depths of approximately -7 kJ/mol (PO4 bead) and -9 kJ/mol (head group), indicating a clear selectivity of the binding site for PIP species over POPS. POPC shows even weaker interactions, as would be anticipated given the PMF is evaluated for a single lipid molecule in an environment of a POPC bilayer. The truncated POPC has similar binding energy as the full length POPC, but the position of minimum is slightly different, which suggest that the contribution of hydrophobic interactions from the tail groups is relatively small. However, the presence of the fatty acid chains might influence the binding configurations of lipids, and thus the placement of the head groups, where most binding interactions occur.

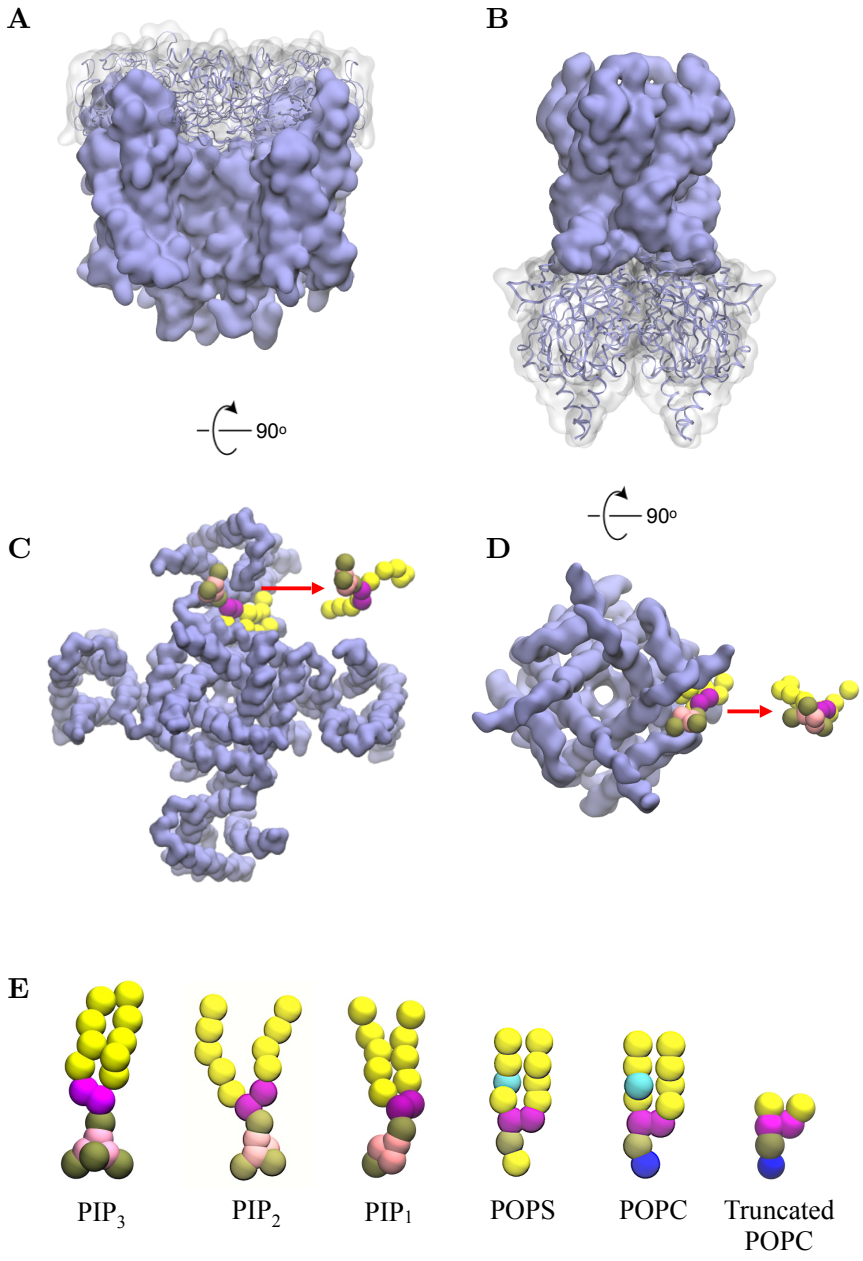


Figure 4.6: Coarse-grained (CG) representations of PC2, Kir3.2 and bound lipids used in PMF calculations. **A:** The PC2 structure used in PMF calculations is shown in ice blue surface. The region that was truncated is shown in transparent surface/tube representation. **B:** The truncated Kir3.2 structure used in PMF calculations is shown in ice blue surface. The transparent surface/tube region was truncated for computational efficiency. **C:** The reaction coordinate for lipid unbinding from PC2 is indicated with the red arrow. The lipid bound state and unbound state are shown. **D:** The reaction coordinate for lipid unbinding from Kir3.2 is indicated with the red arrow, with the lipid bound and unbound states shown. **E:** CG representations of different lipids used in PMF calculations.

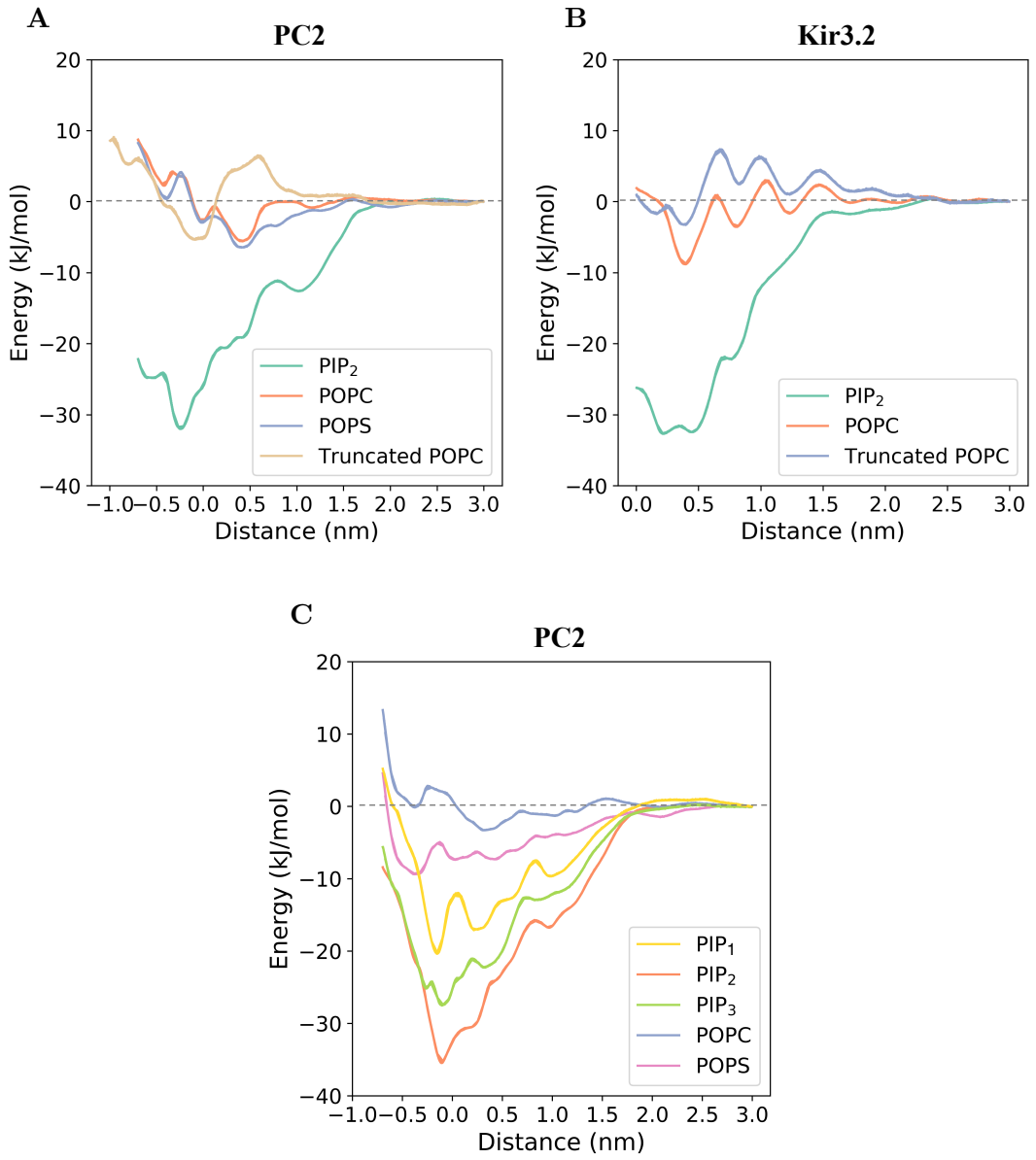


Figure 4.7: Free energy profiles for lipid interactions at the binding site on PC2 and Kir3.2. **A:** Potentials of mean force (PMFs) of the interactions of PIP_2 (green), POPS (blue), POPC (orange) and truncated POPC (brown) with the lipid binding site on PC2. The lipids of interest were pulled on relatively small sets of particles (POPC/POPS: PO4 bead; PIP_2: RP1, RP2 and RP3 beads) **B:** PMFs of the interactions of PIP_2 (green), POPC (orange) and truncated POPC (blue) with the lipid binding site on Kir3.2. The pulling groups of lipids are the same as those for PC2 in **A**. **C:** PMFs of the interactions of PIP_2 (orange), PIP_3 (green), PIP_1 (yellow), POPS (purple) and POPC (blue) with the lipid binding site on PC2. Umbrella biasing potentials were applied to the whole head groups of lipids. The errors for all PMF calculations were estimated with standard deviations calculated from non-overlapping 200 ns time segments. The maximum magnitude of errors is around 3 kJ/mol. An example is shown in Figure 4.8.

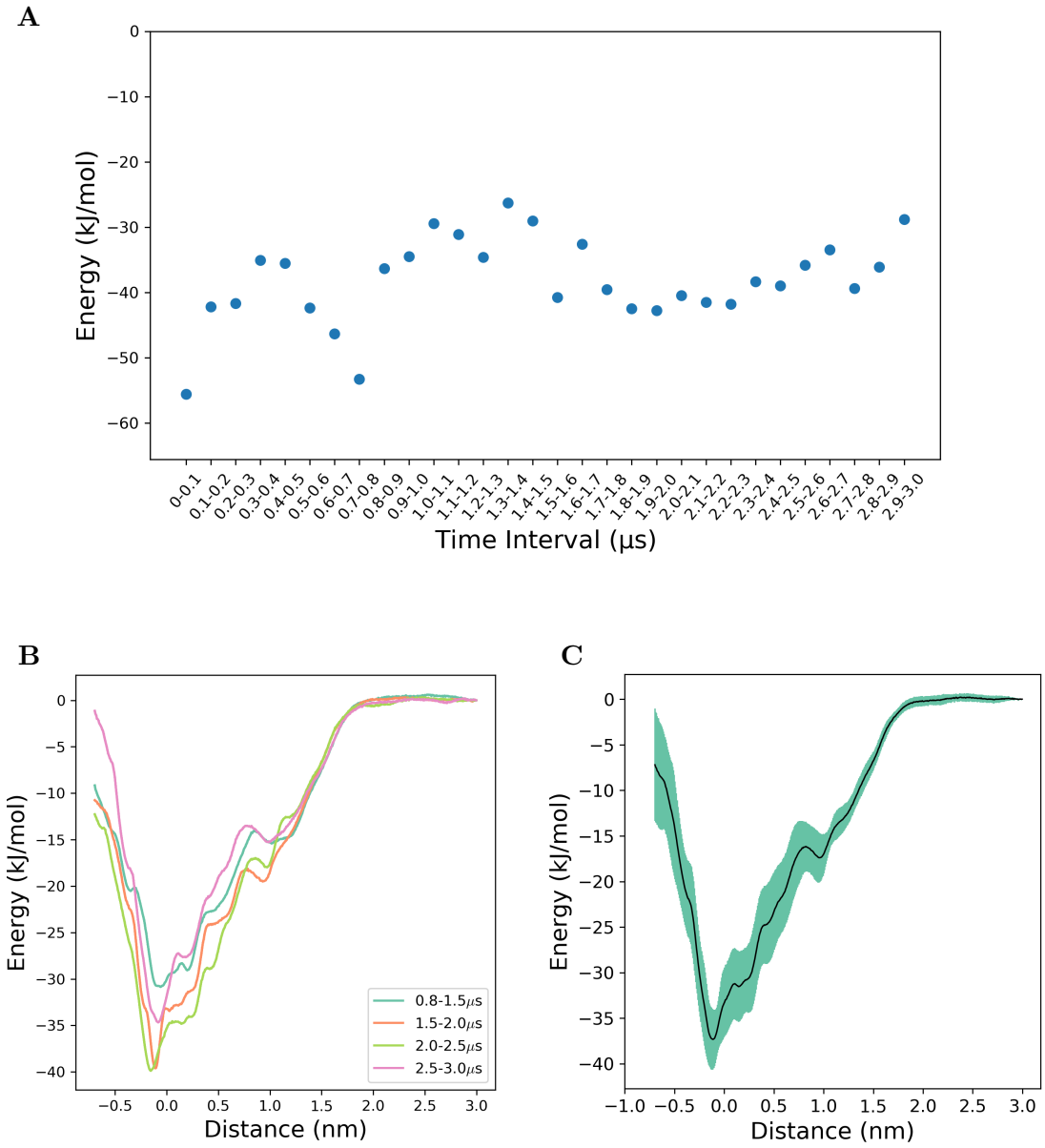


Figure 4.8: Convergence of PMF calculations of PIP_2 binding to PC2. **A:** Convergence of a PMF calculation for PIP_2 binding was assessed by calculating the binding free energy (measured as the energy well depth) derived from non-overlapping 100 ns segments of the simulations. **B:** PMF profiles calculated with non-overlapping 500 to 700 ns time segments are shown. The first 800 ns of each window were excluded from the PMF calculation. **C:** Average PMF (black) and standard deviation (green) calculated from PMFs for non-overlapping 200 ns time segments from 1.6 to 3.0 μs.

The reaction coordinate of Kir3.2 was defined as the distance between the COM of Lys90, Lys194, Lys199 and Lys200 within Kir3.2 and the PO4 bead of POPC or RP* beads of PIP_2. An interaction free energy of approximately -33 kJ/mol was revealed for PIP_2 binding with Kir3.2 (Figure 4.7B), which is comparable to what we observed in PC2. In addition, previous simulation study has reported that the binding energy of PIP_2 in related protein wild-type Kir2.2 is -45 kJ/mol. With R186A mutation in Kir2.2, the number of basic residues around the binding site reduced from six to five, and the binding energy dropped to -32 kJ/mol [259], which is consistent with our calculation for PIP_2 binding with Kir3.2 given that the truncation in our model resulted a number of coordinating basic residues of five at the binding site. The binding energy of POPC in Kir3.2 (~ -9 kJ/mol) is also notably smaller than that of PIP_2, and is in the same energy scale as POPC binding in PC2. The free energy profile for the truncated POPC showed the same pattern as the full length POPC with relatively smaller well depth of ~ -3 kJ/mol, suggesting a degree of hydrophobic contribution from the tail groups with similar binding configuration.

The free energy profiles for both Kir3.2 and PC2 exhibit multiple minima along the reaction coordinates. The formation of concentric lipid shells have been shown in comparable simulations around nanopores [287] and around NanC, a cylindrical bacterial outer membrane protein [288]. The radial distribution and the density plot of POPC around both Kir3.2 and PC2 well demonstrate the existence of ordered lipid shells around the proteins. The minima in PMFs are likely to reflect the annular shells of relatively immobilised lipid molecules around the Kir3.2 and PC2 channels (Figure 4.9), as have been seen for a number of channel and other membrane proteins in simulations [289, 290].

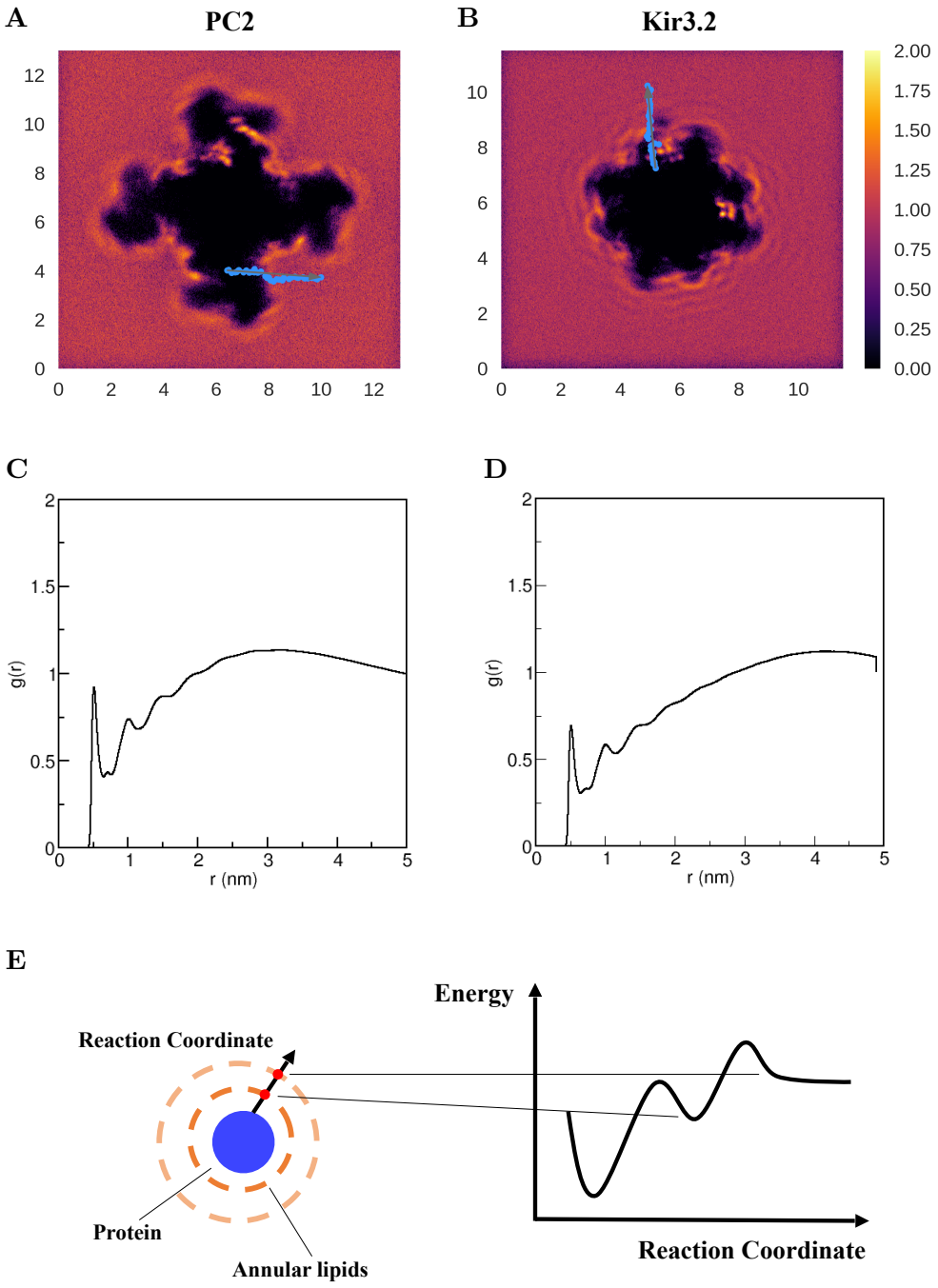


Figure 4.9: POPC distribution in the cytoplasmic leaflet of the bilayer around PC2 and Kir3.2 in a PIP_2 PMF simulation. **A & B:** POPC density around PC2 and around Kir3.2 are plotted with the PIP_2 reaction (i.e. pulling) coordinates indicated as blue dots. The pulling direction is shown with ochre arrow. **C & D:** The radial distribution of POPC around PC2 and around Kir3.2. **E:** Schematic diagram of concentric lipid shells. Blue circle represents a membrane protein, and the rings represent the distributions of lipids around the protein.

4.3.3 The PIP_2 Binding Site

Having established that PC2 binds PIP_2 selectively, we examined the interactions of the protein with the lipid in more detail, based on a CG snapshot structure of the PC2-PIP_2 complex corresponding to the energy minimum in the PMF calculation. As noted above, the head group of PIP_2 interacts with five basic residues: Arg504, Lys572, Lys575, Arg592 and Lys595. In particular, the $1'$ (phosphodiester) phosphate interacts closely with Arg592, and the $4'$ and $5'$ phosphates interact with Lys572 and Lys575 respectively. A binding site for PIP_2 formed by a cluster of basic residues has been seen in other ion channels, e.g. Kir channels [155, 291], and in GPCRs [165, 261].

Given the reduced granularity of the CG MARTINI models, detailed interactions contributing to the binding free energy might have been neglected in the PMF calculations. To evaluate the PIP_2 binding site in finer details, the CG structure corresponding to the final snapshot of the sampling window for the energy minimum configuration was converted into an atomistic representation. A two-stage atomistic simulation was then performed. Firstly, a short (30 ns) simulation was performed in which weak harmonic restraints (force constant = 300 kJ/mol/nm^2) were applied to the distances between the $1'$, $4'$, and $5'$ phosphates of PIP_2 and the side chains of Arg592, Lys575 and Lys572 respectively, in order to relax the atomistic model whilst maintaining the interactions seen in the CG PMF calculations. The distance restraints were then removed and three replicates of an unrestrained simulation (durations 200 to 250 ns) were run to allow the PIP_2 to explore the binding site on PC2. Final snapshots from the restrained and unrestrained simulations are shown (Figure 4.10A). With a distance cut-off of 3 Å and an angle cut-off of 20°, the number of hydrogen bonds formed between PIP_2 and the basic residues around the binding site in PC2 was calculated throughout the unrestrained simulations. In atomistic detail, the side chain of Lys575 impedes the access of the phosphate groups of PIP_2 to Lys572 to the range of hydrogen

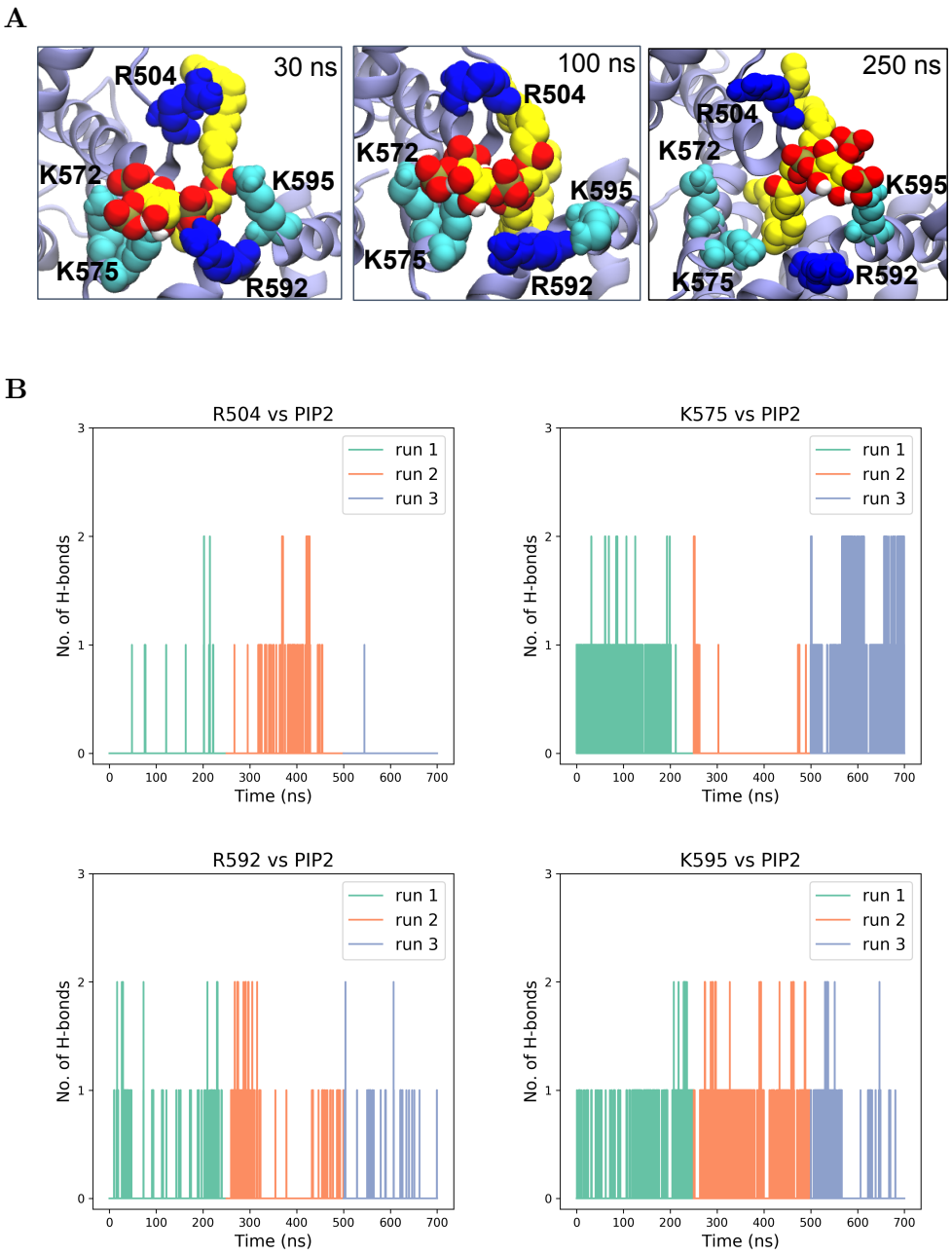


Figure 4.10: Protein-lipid interactions showing the head group of PIP_2 coordinated by five basic residues (Arg504, Lys572, Lys575, Arg592 and Lys595) at the binding site. **A:** Snapshot structure corresponding to the CG energy minimum was converted into an atomistic representation and used as the basis of 30 ns distance restrained simulations (see text for details) followed by 250 ns of unrestrained simulations (with snapshots at 100 and 250 ns shown). **B:** Hydrogen-bonding interactions between key basic side chains and the PIP_2 head group during 3 replicates (run 1: 0−250 ns, run 2: 250−500 ns; run 3: 500−700 ns) of the unrestrained atomistic simulations.

bonding. Under the timescale and cut-off conditions we examined, PIP_2 formed negligible number of hydrogen bonds with Lys572.

The interactions between the lipid head group and the interacting residues are seen to be dynamic and to vary stochastically between replicate simulations. Thus, for residues Arg504, Lys575, Arg592 and Lys595, fluctuating numbers of hydrogen bonds were formed with PIP_2 throughout the simulations (Figure 4.10B). Such fluctuations in PIP_2 binding are not unique to PC2, they have also been reported for PIP_2 molecules bound to Kir channels, e.g. [291].

4.3.4 PIP_2 Interactions in TRP Channels Revealed by CG Simulations

The proposed PIP_2 interaction site on the transmembrane domain of PC2 may be compared with a lipid binding site, which has been seen in the structure of TRPV1, close to the vanilloid ligand-binding site, and has been interpreted as being occupied by a phosphatidylinositol molecule [6]. As an endogenous modulator of TRPV5, binding of $PI(4,5)P_2$ between S3/S4/S5 helices has also been shown in the structure of TRPV5 [53]. Comparison of the three sites (Figure 4.11A) reveals strong similarities, especially the location of the anionic phosphate-containing lipid head group at the N-terminal region of the S3 helix dipole [292], which contributes to the coordination of the negative charges of the head groups. To explore this possible common binding site further, we extended our CG simulations of PC2 in a mixed *in vivo* mimetic lipid bilayer to a number of different TRP channel structures (see Table 4.1). One copy of each TRP channel was embedded in a bilayer with random distribution of lipids. Over the course of each simulation, the duration of PIP_2 contact was calculated as the continuous contact time between the protein and PIP_2 molecules. Contact is defined as within 6 Å of distance. The durations of contact for each residue of the protein were averaged over all repeats of each system. As shown in Figure 4.11B, all simulated TRP channels were structurally aligned, and the average durations of contact were mapped onto the sequences with red as intense contact and white as no contact. The results show

that in some other TRP channels, the residues, which are structurally equivalent to the PIP_2 contacting residues on PC2, also interact with PIP_2 with relatively high intensity, suggesting a degree of conservation of the proposed PIP-channel interaction between different families of TRP channels.

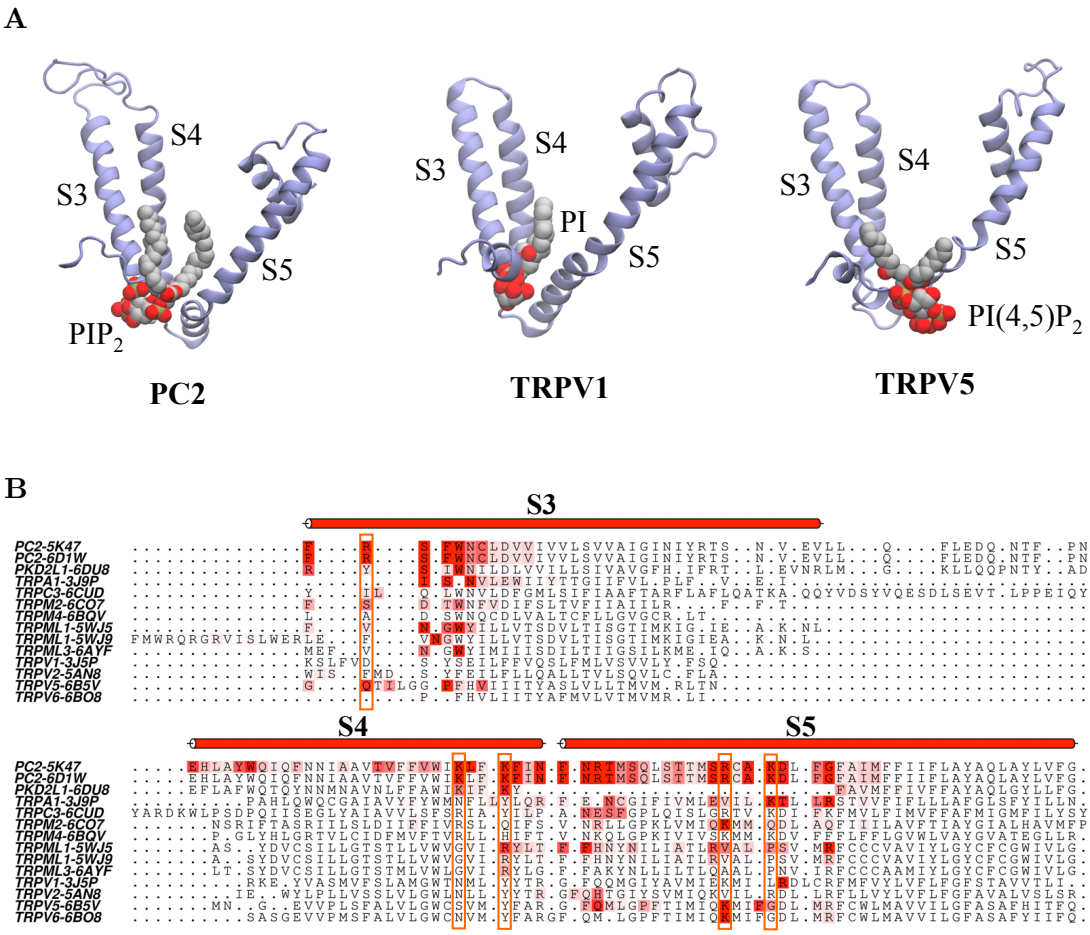


Figure 4.11: Comparison of lipids bound to PC2, TRPV1 and TRPV5. **A:** PIP_2 bound to PC2 (as revealed by the current simulation study); PI bound to TRPV1 (as revealed by cryo-EM, PDB ID: 5IRZ). $PI(4,5)P_2$ bound to TRPV5 (as revealed by cryo-EM, PDB ID: 6DMU). In each case the lipid molecule is located between the S3, S4 and S5 helices of the VSLD. **B:** A structure-based sequence alignment of selected TRP channels coloured on contacts with PIP_2 in the mixed lipid simulations. Residues of the region around the binding pocket between the S3, S4 and S5 helices are coloured (on a white to red scale) based on the mean duration of the interactions of PIP_2 head groups with each residue. The five basic residues of PC2, which form interactions with the head group of PIP_2 at the binding site (i.e. R504, K572, K575, R592 and K595) are boxed.

4.3.5 Atomistic Study of Phosphoinositide Binding Specificity

Given the CG mapping of lipid atoms, the charge distribution of the inositol head group does not allow fine differentiation between different species of PIP_2 at CG level. To explore whether atomistic simulations could provide insights on how lipid-binding proteins discriminate different species of phosphoinositides (PIPs), the well characterised PIP binding ion channel − inward rectifier potassium channel Kir2.2 (PDB ID: 3SPI) was used with CHARMM36 all atom force field to examine possible differences in protein-lipid interactions. The X-ray crystal structure of the chicken Kir2.2 in complex with dioctanoyl (diC_8) derivative of PI(4,5)P_2 showed that the PIP_2 binding site is located at the interface between the transmembrane and cytoplasmic domains. The conserved residues Arg78, Trp79 and Arg80 in the transmembrane domain contribute to the non-specific interaction with $1'$ phosphate, and the cytoplasmic residues Lys183, Arg186, Lys188 and Lys189 facilitate the specific phosphatidylinositol binding [155]. Electrophysiology experiments have shown that PI(3,4)P_2 has almost no activation effect on Kir2.2, and the activation level of PI(3,4,5)P_3 is approximately 40% of that of PI(4,5)P_2 [293], which suggest that Kir2.2 is able to differentially bind to these three species of PIPs. Although within a highly positively charged binding site the PIPs are most likely to be fully ionised, to take into account the effect of protonation state, the binding of four different species of PIPs − $5'$ protonated PI(4,5)P_2, fully deprotonated PI(3,4)P_2, PI(4,5)P_2 and PI(3,4,5)P_3 − was tested (Figure 4.12A).

The topologies of protonated CHARMM PI(3,4)P_2, PI(4,5)P_2 and PI(3,4,5)P_3 lipids were modified to make them fully ionised. The atomistic simulation system of Kir2.2 was set up by aligning one copy of each of the four PIPs to one of the crystallographic PIP_2 positions so that each subunit of Kir2.2 has a different PIP (Figure 4.12A). The total number of hydrogen bonds formed between the lipids and Kir2.2 was counted with an angle cut-off of 20° and a distance cut-off of 3 Å, and averaged over 6 × 250 ns simulations. The averaged continuous time of hydrogen bonds was also calculated. As shown in Figure 4.12B, the difference between the

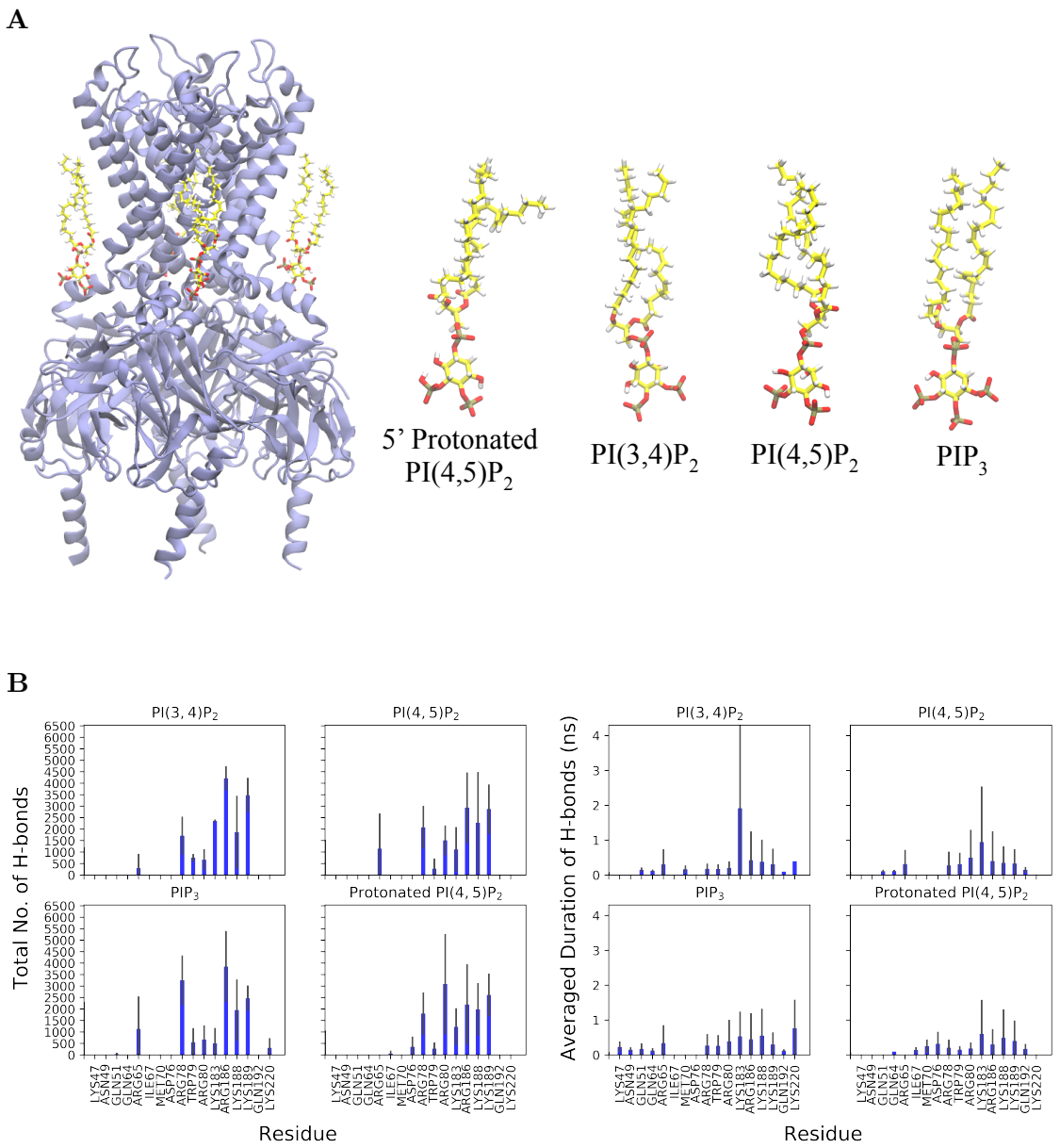


Figure 4.12: Hydrogen-bonding patterns of Kir2.2 with different PIPs. **A:** 5′ protonated $PI(4,5)P_2$, fully deprotonated $PI(3,4)P_2$, $PI(4,5)P_2$ and $PI(3,4,5)P_3$ aligned to the crystallographic PIP_2 locations on Kir2.2 (PDB ID: 3SPI). Kir2.2 is shown in ice-blue cartoon. Atomistic representations of the PIPs are shown in yellow sticks. **B:** Total number (left) and averaged duration (right) of hydrogen bonds between Kir2.2 and each of the four PIPs are calculated with standard deviations across 6 replicates of 250 ns atomistic simulations.

overall hydogen-bonding pattens of the PIPs are minimum. The protein residues involved in the coordination of different PIPs are mostly in common, and the

difference in the persistence of interaction reflected by the averaged duration of hydrogen bonding is within 1 ns across the four different lipids.

Close examination of the binding poses of the PIPs shows that head groups of different lipid tend to adopt slightly different binding angles (Figure 4.13A). To quantify the differences, we calculated the head-tail angle of each PIP, which is defined as the angle between the vector formed from the second carbon of the glycerol moiety (C2) to the phosphorus atom of the 1′ phosphate group (P), and the vector formed from the 1′ carbon (C11) to the 4′ carbon (C14) of the inositol ring. As shown in Figure 4.13B, the head-tail angles of PI(3,4)P_2, PI(3,4,5)P_3 and protonated PI(4,5)P_2 have slightly wider distribution than that of fully deprotonated PI(4,5)P_2. This may suggest that once the fully deprotonated PI(4,5)P_2 adjusts to the binding site on Kir2.2, the binding pose is locked relatively tight by the coordinating residues comparing to the other species of PIPs.

To investigate how the head groups of the PIPs sit relative to the membrane bilayer, we calculated the lipid tilt angle by measuring the angle between the vector $\overrightarrow{C11C14}$ and the negative Z-axis (Figure 4.13C). The head groups positions of PI(4,5)P_2 relative to the bilayer showed the same trend regardless of the protonation state. The head group of PI(3,4)P_2 tend to bend more towards the surface of the bilayer than other species of PIPs. Even though different PIPs showed different binding pose within the binding site on Kir2.2, how these differences translate into PIP specificities is unclear from our simulations. Given the setup of our simulation system, it is possible that the protein local conformation changes as a result of the synergetic effects of binding of the same PIP at each subunit could not be observed. Further studies need to be conducted with single species of PIP present to get better understanding of the difference in their activation mechanisms of Kir2.2. The studies on Kir/PIP binding do not provide a clear indication of the nature of the specificity at the binding site. We therefore decided not to pursue this approach for PC2/PIP interactions, given the uncertainties about experimental specificity of the interactions with PC2 revealed by the PIP strips.

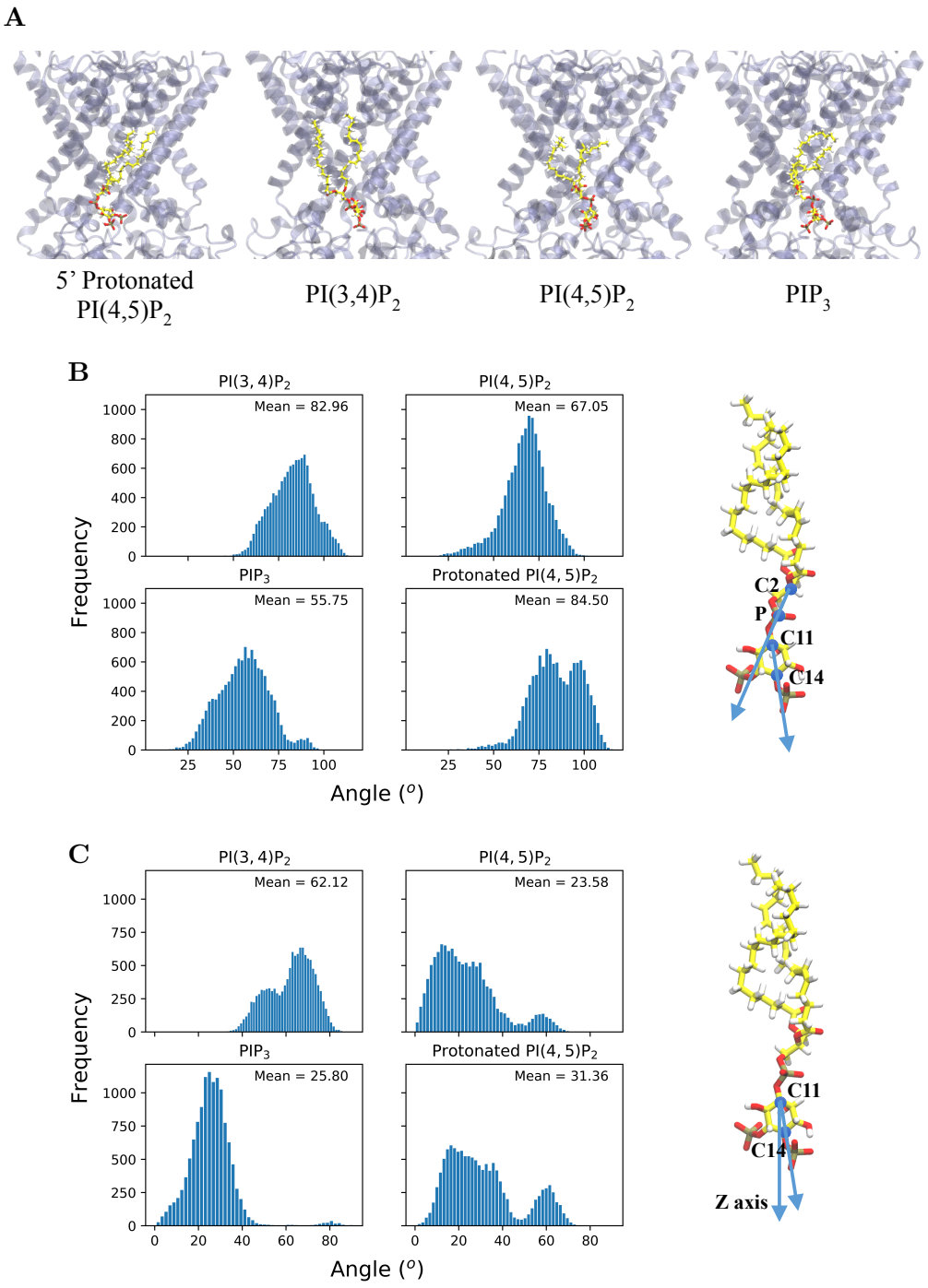


Figure 4.13: Binding poses of different PIPs within the binding sites on Kir2.2. **A:** The position of each PIP from final snapshot of one 250 ns simulation of Kir2.2 with 5′ protonated $PI(4,5)P_2$, fully deprotonated $PI(3,4)P_2$, $PI(4,5)P_2$ and $PI(3,4,5)P_3$. Kir2.2 is shown in transparent ice-blue cartoon. Atomistic representations of the PIPs are shown in yellow sticks. **B:** Histograms of the head-tail angles of the PIPs, which is define by the angle between $\overrightarrow{C2P}$ and $\overrightarrow{C11C14}$. **C:** Histograms of the tilt angles of the PIPs, which is define by the angle between $\overrightarrow{C11C14}$ and the negative Z-axis.

4.3.6 A Cholesterol Binding Site Revealed by Cryo-EM and Simulations

As shown in Chapter 3, our cryo-EM densities revealed a possible site for co-purified cholesterol (Figure 4.14A & 3.13) located between the S3 and S4 helices of one protomer and S6 of the adjacent protomer. Cholesterol has also been shown to play a key role in signalling in ciliary membranes [294] and possibly in their organisation [252]. Cholesterol is known to interact with many ion channels [295, 296] and receptors [297, 298], so we decided to explore its possible interactions with PC2. We analysed our CG-MD simulations of PC2 in an in vivo mimetic (i.e. mixed lipid) bilayer to identify possible interactions of cholesterol with the channel (Figure 4.14B & 4.14C). The results of these simulations revealed a cholesterol interaction site close to the site suggested by the cryo-EM density. This indicates that this site may interact with cholesterol present in the bilayer. We note that recent comparable CG-MD simulations of GPCRs have identified interaction of cholesterol at sites which bind cholesterol in a number of crystal structures [261]. Having identified a cholesterol binding site, we performed CG PMF calculations to estimate the free energy landscape for the cholesterol interaction (Figure 4.14D). This showed a weaker interaction than for PIPs (Figure 4.7), with a free energy well depth comparable to that for binding of cholesterol to other membrane proteins (see e.g. the class F G protein coupled receptor Smoothened [250]).

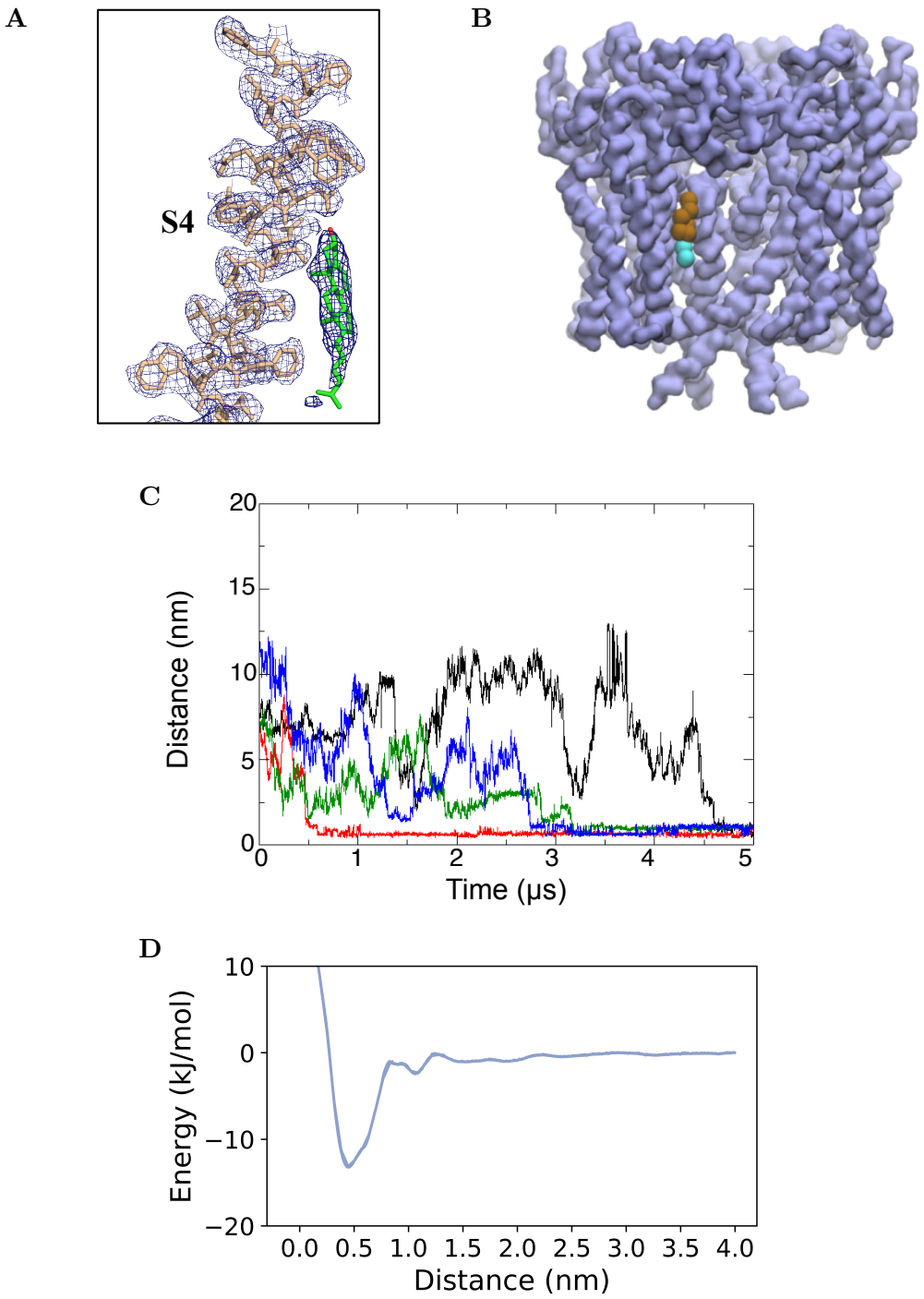


Figure 4.14: Cholesterol interactions with PC2. **A:** Cryo-EM density of the S4 helix with the fitted cholesterol molecule in our 3 Å PC2/PI(4,5)P_2 dataset. **B:** Cholesterol in CG simulations of PC2 in a in vivo mimetic mixed lipid bilayer was observed to bind to the same site as the cholesterol density in our cryo-EM datasets (see Figure 3.13). **C:** Distance from the binding site as a function of time for four cholesterol molecules, which bind to PC2 during a simulation in a mixed lipid bilayer. **D:** Potential of mean force (PMF) for the interaction of cholesterol with the binding site on PC2.

To further test the possible cholesterol binding site, we performed atomistic simulations (3 x 250 ns) starting with a model of cholesterol built into the cryo-EM density. From the simulations it is evident that cholesterol interacts dynamically at this site, which is located at the interface between the VSLD and S6 with the long axis of cholesterol running approximately parallel to helices S3 and S4. Thus, for the four symmetry-related binding sites on the channel, one site exhibits a stably bound cholesterol, one site has a cholesterol which transiently dissociates then rebinds, and the two other sites show intermediate behaviour (Figure 4.15A). Within the timescale of the simulations, none of the cholesterol molecules was completely dissociated from the binding sites. These differences in binding across subunits are probably stochastic. They are consistent with a number of simulation studies of cholesterol on GPCRs, which indicate relatively dynamic, loose binding (Figure 4.15B; see [250] for a more detailed discussion). The hydroxyl group mainly hydrogen bonded to Gln557 and Asn560. The arrangement of the binding pocket and the binding pose of cholesterols are consistent with what we observed in cryo-EM (Figure 4.15C & 3.14A). The hydrocarbon chain of cholesterol showed relatively higher degree of mobility throughout the simulation, which may explain the lack of density around this region in our 3.4 Å resolution cryo-EM map. It is tempting to speculate about the possible biological importance of cholesterol interactions with PC2. For example, cholesterol plays a key role in signalling via the ciliary GPCR Smoothened, and it is possible that different regions of cilia differ in the cholesterol content of their membranes [294]. However, further biochemical studies will be needed to establish a possible role of cholesterol in regulation of PC2 in ciliary membranes.

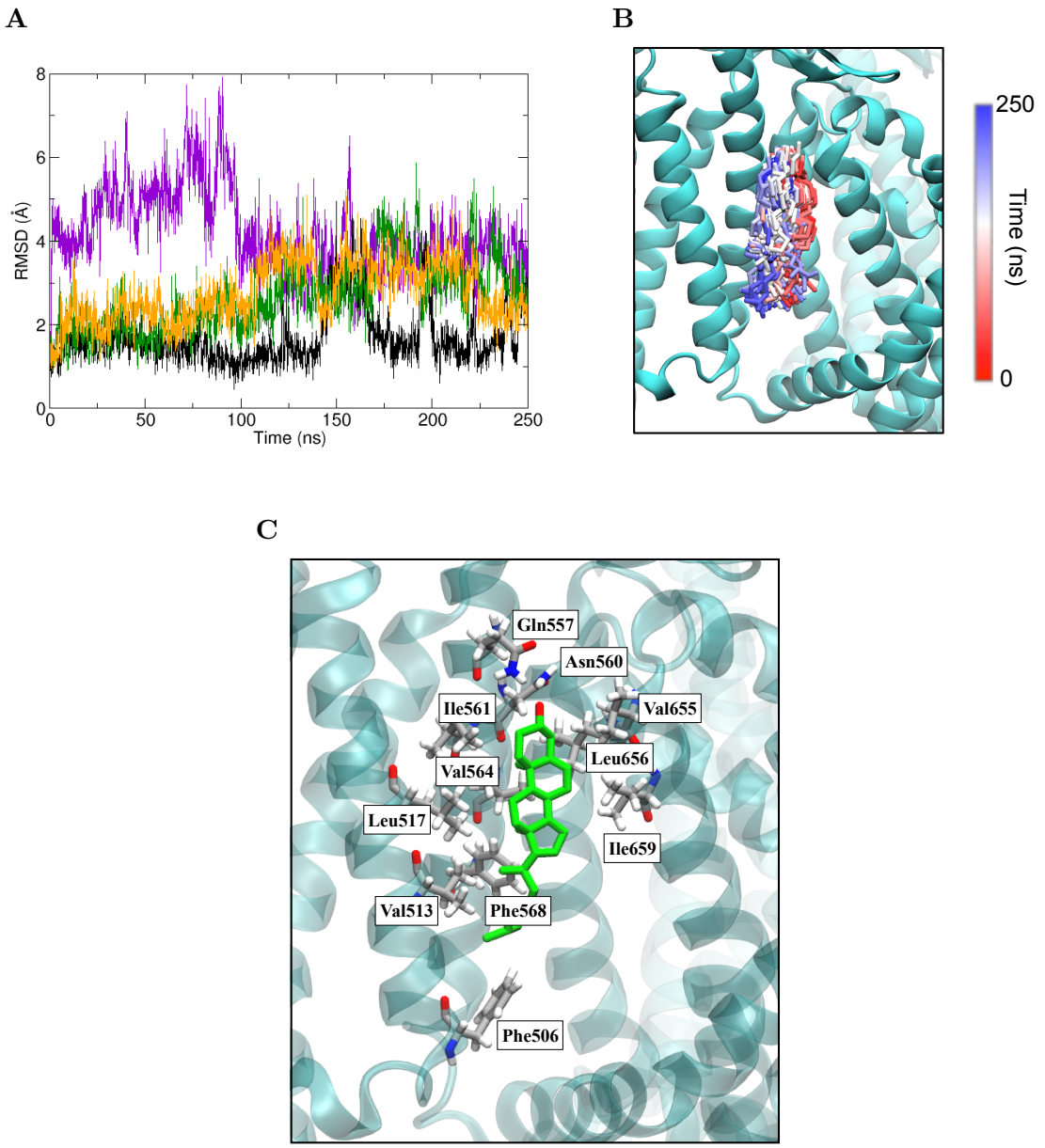


Figure 4.15: Atomistic simulations of cholesterol-bound PC2, with the cholesterol molecule initially built into the cryo-EM density. **A:** RMSD vs. time for the four cholesterol molecules bound to PC2 during a 250 ns atomistic MD simulation. The large fluctuations in RMSD for two molecules (purple and green curves) demonstrate the relatively loose binding of cholesterol at this site. **B:** Snapshot structures of cholesterol (stick representation; taken every 1 ns) at the binding site on PC2. Each of these structures is coloured according to the corresponding simulation time on the RWB colour-scale shown, thus structures at the start of the simulation are coloured red, and those at the end in blue. **C:** Final simulation snapshot (t = 250 ns) showing the arrangement of key binding site residues.

4.4 Conclusions

Using multiscale MD simulations, we identified a lipid binding pocket in the human TRP channel PC2 between the S3 and S4 transmembrane helices and the S4-S5 linker. Free energy calculations suggested that this binding site is selective for PIP_2 and related PIPs. The location of this site is analogous to the shared binding site for vanilloid ligands and phosphatidylinositol lipids in the canonical TRPV1 and TRPV5 channels (Figure 4.11A). This region is distinct from the location between the VSLD and the core pore-forming domain of the PDB ID: 5MKE structure of PC2 [15] where lipid-like density is modelled as phosphatidic acid and as palmitic acids binding within the extracellular leaflet region of the channel protein.

Structural and simulation studies have shown that the movement of the S4-S5 linker is important for gating in voltage-gated cation channels. For example, in the voltage-gated potassium channel Kv1.2, the constriction or dilation of the pore domain is controlled by the voltage sensor domain through the S4-S5 linker [299]. The transmembrane helices of TRP channels and voltage-gated cation channels share high structural similarity, indicating a possible common functional role of the S4-S5 linker. Indeed, it has been shown that TRPV1 [6, 300], TRPV2 [301] and TRPV4 [302] use the same mechanism to gate the channels, and disruption of the interaction configurations at the linker region by ligands/lipids could change the gating probability.

Sequence alignment shows that the basic residues involved in PIP_2 binding (Lys572, Lys575, Arg592 and Lys595) are highly conserved (Figure 4.5), which suggests that these residues may play a crucial role in modulating the channel activity. Disruption of the native intramolecular interactions involving these residues may also alter channel activity. Residue Asp511 is central to the hydrogen-bonding network involving Lys572 and Lys575. The naturally occurring mutation D511V has been well characterised to be channel inactivating [88]. The loss of electrostatic interactions between the Asp and Lys residues could allow Lys572 and Lys575

to have stronger interactions with PIP_2, thus magnifying the regulatory effect of PIP_2 on PC2. The similar binding energy of PIP_2 in PC2, Kir3.2 and Kir2.2 [259] channels is also supportive of the existence of a PIP_2 binding site in PC2.

Our structural and simulation data also reveal a cholesterol-binding site on the outer-leaflet-facing surface of the PC2 molecule. This seems to be close to the site suggested to interact with a phospholipid (modelled as phosphatidic acid; see PDB ID: 5MKF and 5MKE) in the study of Wilkes et al. [15]. Although interactions of TRP channels with cholesterol have not been studied in detail, some physiological data are available [296]. For PC2, mutations of two of the residues forming the cholesterol binding site (Leu517Arg and Leu656Trp) are classified as 'likely pathogenic' and 'likely hypomorphic' respectively (see http://pkdb.mayo.edu). It is tempting to suggest the likely biological importance of cholesterol interactions with PC2. For example, cholesterol plays a key role in signalling via the ciliary GPCR Smoothened, and it is possible that different regions of cilia differ in the cholesterol content of their membranes [294]. However, further biochemical studies will be needed to establish a functional role of cholesterol in regulation of PC2 in ciliary membranes. Given the emerging importance of cholesterol interactions with other ion channels and with receptors [297, 303, 304], it would not be surprising if cholesterol interactions with TRP and related channels were of functional importance.

Our results have further demonstrated the ability of CGMD simulations to reveal the specificity and energetics of the interactions between transmembrane proteins and different lipids. However, more refined difference between closely related lipid species e.g. $PI(3,4)P_2$ and $PI(4,5)P_2$ is beyond the purpose of CG simulations. Using atomistic simulations, we were able to observe some difference in binding poses among different PIPs. Nevertheless, even with atomistic details, the binding specificity can not be differentiated by simply looking at the hydrogen-bonding networks. Other characterisation criteria needs to be evaluated to explain the differences in regulatory effect of different lipids.

5

Simulation Studies of Ion Interactions of Polycystin-2

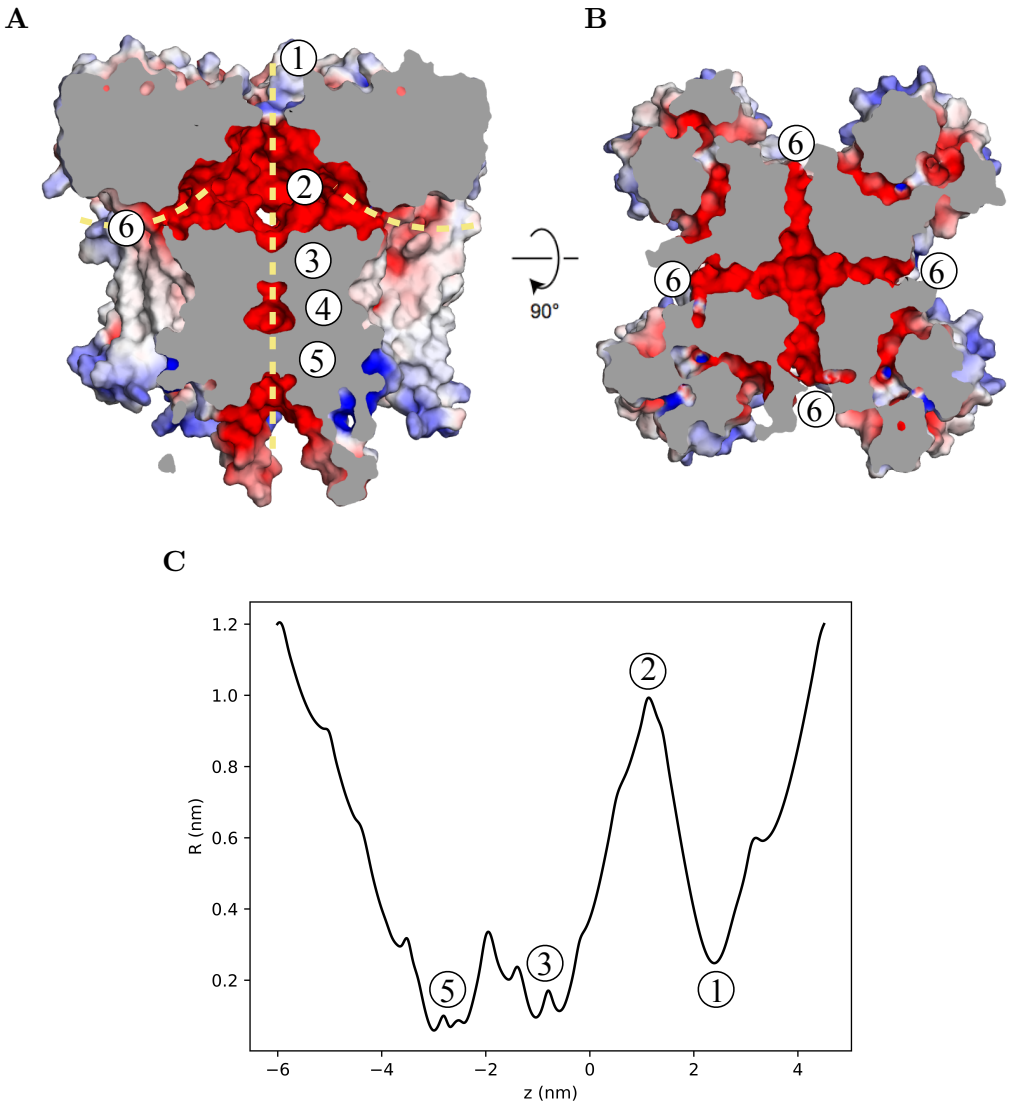
Contents

5.1 Introduction 128
5.2 Methods 130
5.2.1 System setup 130
5.2.2 Coarse-grained Simulation Details 130
5.2.3 Atomistic Simulation Details 132
5.2.4 Simulation Visualisation and Analysis 134
5.3 Results and Discussion 135
5.3.1 Dynamics of the Ion Identified in Cryo-EM Maps 135
5.3.2 Ion Permeation and Selectivity of PC2 at the Selectivity Filter and TOP Domain 144
5.3.3 Ion and Water Permeation at Central Cavity and Lower Gate of PC2 152
5.4 Conclusions 155

5.1 Introduction

The channel function of PC2 has been described in a range of subcellular compartments due to its wide distribution. PC2 has been proposed to contribute to the transduction of extracellular mechanical stimuli caused by bending of the cilia into intracellular Ca^{2+} signals in the primary cilia of kidney epithelium [110]. Direct blocking of PC2 with antibodies in wild-type kidney epithelial cells abolishes Ca^{2+} influx in response to physiological fluid flow. However, more recent studies have suggested that Ca^{2+} signalling may not be involve in ciliary mechanosensation [148]. When fluid flow was applied to primary cilia of various cultured cells including kidney epithelial cells, kidney thick ascending tubules and kinocilia of inner ear hair cells, there were no cilia-specific Ca^{2+} influxes. In addition to its role in primary cilia, it has been proposed that PC2 could function as an intracellular Ca^{2+} release channel in the endoplasmic reticulum (ER) membrane [88]. Nevertheless, studies have reported contradictory findings on the ion selectivity of PC2. Single channel recording of reconstituted PC2 in human syncytiotrophoblast apical membranes showed that it has high selectivity and permeation to Ca^{2+} [305]. Whole cell recording from cells co-expressing polycystin-1 and -2 also suggested higher Ca^{2+} permeability over monovalent ions [93]. However, direct current recording from the primary cilia of renal collecting duct epithelial cells showed that ciliary PC2 is a non-selective Na^+, K^+ channel, with low levels of Ca^{2+} currents, rather than a Ca^{2+} channel [146]. Due to the still unknown mechanism of activation of PC2 and the tendency of ER retention in over-expression systems, electrophysiology studies of PC2 have been proven to be difficult.

As shown in Figure 5.1, there are two vestibules in PC2 that are accessible for water and ions. The upper vestibule is enclosed by the ER luminal or extracellular TOP domain above the selectivity filter, which could be entered through the apical opening at the centre of the TOP domain or the lateral openings between the TOP domain and the transmembrane domain. There is also a central vestibule constrained by the selectivity filter (Leu641 to Gly642) and the lower gate (Leu677



① Apical opening ② TOP cavity ③ Selectivity filter
④ Central cavity ⑤ Lower gate ⑥ Lateral opening

Figure 5.1: Side (**A**) and top (**B**) view of PC2 in sliced surface representation coloured by electrostatic potential. Pore constrictions and vestibules are shown. Possible entry point for water and ion are indicated. **C:** Pore radius profile with corresponding constrictions labelled.

and Asn681). This chapter uses atomistic MD simulations to provide some insights on the ion and water interactions of PC2. Firstly, the selectivity of the filter and

the TOP domain will be explored, and then the water and ion permeation through the channel pore, especially at the lower gate will be discussed.

5.2 Methods

5.2.1 System setup

Initial protein coordinates of PC2 were obtained from our PC2/PI(3,5)P_2 and PC2/PI(4,5)P_2 structures and from its published structures, i.e. PDB ID: 5K47, 5MKE and 6D1W. Bound N-acetyl-d-glucosamines were removed from all protein structures. Bound cholesterols in the systems of our recent PC2/PI(3,5)P_2 and PC2/PI(4,5)P_2 were retained. Other lipids and ions were removed from all systems. Missing atoms were modelled in using MODELLER version 9v10[262]. The MODELLER objective function was used to assess the models, and the ones with the best scores were visually inspected. The most structurally sound models were then chosen for subsequent simulations. Protein ionisation states were determined with pdb2gmx implemented in GROMACS (www.gromacs.org). Simulations additional to the ones listed in Table 4.1 were all performed using GROMACS version 5.1 [264], and are summarised in Table 5.1.

5.2.2 Coarse-grained Simulation Details

Coarse-grained simulations were used to embed the PC2 structures in lipid bilayers. Energy minimisation of the protein structures were done using the steepest descents algorithm implemented in GROMACS before coarse graining with the martinize.py workflow. MARTINI 2.2 force field [265] was used to map protein atoms to MARTINI particles. The *ElNeDyn* elastic network [217] was used for the control of overall protein shape. An elastic bond force constant of 500 kJ/mol/nm^2 and an upper elastic bond cut-off of 0.9 nm were applied with the elastic network. MARTINI 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipids were randomly placed around the coarse-grained (CG) protein structures. The systems were then solvated and neutralised in NaCl solution at a concentration of 0.15 M

Table 5.1: Summary of Simulations Performed in Chapter 5

Channel	PDB ID	Ion Species	Concentration (M)	Force Field	Replicates × Duration (ns)
Equilibrium Simulation					
PC2	PC2/PI(4,5)P_2	NaCl	0.15	CHARMM36	3 × 250
		NaCl/filter Ca^{2+}	0.15	CHARMM36	3 × 250
	PC2/PI(3,5)P_2	NaCl	0.15	CHARMM36	3 × 250
		NaCl/filter Ca^{2+}	0.15	CHARMM36	3 × 250
	5K47	$CaCl_2$	0.15	CHARMM36	3 × 250
Electric Field Simulation					
PC2	PC2/PI(4,5)P_2	NaCl	0.35	CHARMM36	3 × 200
		NaCl	0.35	OPLS	3 × 200
		$CaCl_2$	0.35	CHARMM36	3 × 200
	5K47	NaCl	0.15	CHARMM36	3 × 250
		$CaCl_2$	0.15	CHARMM36	1 × 250
	5MKE	NaCl	0.35	OPLS	3 × 200
	6D1W	NaCl	0.35	OPLS	3 × 200

with the standard MARTINI water particles with Van der Waals radii of 2.1 Å. Parameters for lipids and ions were those described in [266].

An initial 100 ns of CG simulation was performed for each structure to allow the self-assembly of a POPC lipid bilayer [268]. Simulation temperatures were maintained at 323 K using Berendsen thermostat [204] with a coupling time constant of $\tau_t = 1$ ps. Higher than physiological temperature was used to allow faster convergence of the system. Protein, lipids, and solvent (water + ions) were coupled separately to the temperature bath. Pressure was controlled at 1 bar by semi-isotropic pressure coupling with the Berendsen barostat [204] with a compressibility of 5×10^{-6}/bar and a time constant of $\tau_p = 4$ ps. Electrostatics was cut off at 1.1 nm, and Van der Waals potential was switched between 0.9 and 1.1 nm. An integration time step of 20 fs was used. LINCS constraint algorithm was used to control covalent bonds.

5.2.3 Atomistic Simulation Details

Atomistic simulations of PC2 (PDB ID: 5K47, 5MKE and 6D1W) in a pure POPC bilayer were setup and run by converting the end point of the CG self-assembly simulations into atomistic representation using a fragment based protocol [271, 272]. Force fields, ion species and concentrations are summarised in Table 5.1. Simulations performed with CHARMM36 force field [273] were solvated with TIP3P water model [274], and the ones with OPLS force field [306] were solvated with TIP4P/2005 [307].

Systems of our PC2/PI(4,5)P_2 and PC2/PI(3,5)P_2 structures were setup using CHARMM-GUI (www.charmm-gui.org). The protein with bound cholesterols were embedded in a POPC-only bilayer. Simulations were performed using GROMACS version 5.1 with the CHARMM36m [276] force field and TIP3P water model [274]. Energy minimisation and equilibration of the system was done in four steps with gradually decreasing restraining force constants on the protein, cholesterol molecules and the filter ion. Three repeats of a total 175 ns production atomistic simulations were then performed in three steps gradually reducing restraints on the protein and the filter ion. Restraining force constants and duration of each equilibration and production step are listed in Table 5.2.

Table 5.2: Duration and Force Constant in Simulation Steps

Step	Duration	Restraining Object	Force Constant (kJ/mol/nm^2)
M	NA	BB	4000
		SC	2000
		CHOL	4000
		Filter Ion	4000
E1	25 ps	BB	4000
		SC	2000
		CHOL	4000
		Filter Ion	4000
E2	25 ps	BB	2000
		SC	1000
		CHOL	2000
		Filter Ion	2000
E3	2 ns	BB	1000
		SC	1000
		CHOL	0
		Filter Ion	1000
P1	25 ns	BB	1000
		SC	500
		CHOL	0
		Filter Ion	1000
P2	50 ns	BB	500
		SC	0
		CHOL	0
		Filter Ion	500
P3	100 ns	BB	0
		SC	0
		CHOL	0
		Filter Ion	0

1. Abbreviations used for simulation steps are: M: energy minimisation; E(1,2,3): equilibration (1,2,3); P(1,2,3,): production (1,2,3).
2. Abbreviations used for restraining objects are: BB: protein backbone; SC: protein side chains; CHOL: cholesterol; Filter ion: Na^+ or Ca^{2+}.

For all atomistic MD simulations, the long-range electrostatics (< 1 nm) was modelled with the Particle Mesh Ewald (PME) method [277]. Temperature coupling was done with V-rescale thermostat at 310 K. The Parrinello-Rahman barostat [205] with a reference pressure of 1 bar and a compressibility of 4.5×10^{-5}/bar was applied for pressure control. Covalent bonds are constrained to their equilibrium length by the LINCS algorithm [200]. The integration steps of all simulations were set to 2 fs. For computational electrophysiology, the electric field parameters implemented in GROMACS were used to yield a static field with field strength of 50 mV/nm in the negative Z-axis direction, which corresponds to approximately -200 mV voltage difference across bilayer (Figure 5.2).

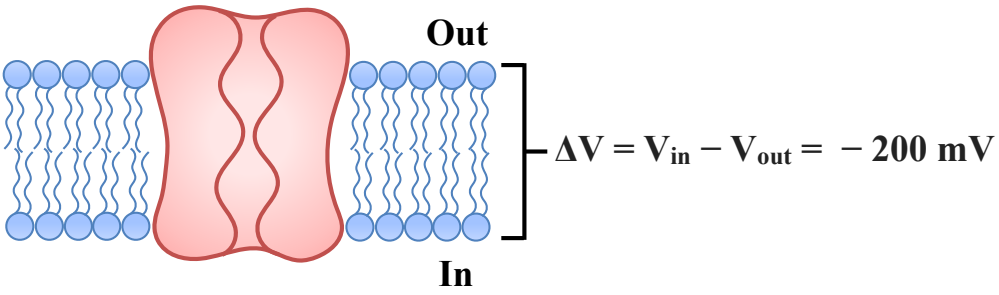


Figure 5.2: Schematic of the electric field applied across the lipid bilayer.

5.2.4 Simulation Visualisation and Analysis

Simulation trajectories were processed using GROMACS. Protein structures and simulation trajectories were visualised with VMD [279] or PyMOL [280]. Time averaged pore radius and free energy of water permeation were calculated with CHAP [308].

5.3 Results and Discussion

5.3.1 Dynamics of the Ion Identified in Cryo-EM Maps

As mentioned in Chapter 3, putative ion densities are clearly present below the selectivity filter in both of our PC2/PI(4,5)P_2 and PC2/PI(3,5)P_2 cryo-EM maps. Without other experimental measurements, the identity of the ion could not be unambiguously determined. Therefore, given that Na^+ and Ca^{2+} are the only cation species present in our cryo-EM buffer conditions, we decided to explore the binding stability of Na^+ vs. Ca^{2+} at the position of the observed electron densities with atomistic simulations. Both of our PC2 structures were placed in POPC bilayers, and in order to remain close to what has been observed in the cryo-EM maps, the bound cholesterols were included. Both systems were solvated in 150 mM NaCl with either one Na^+ or one Ca^{2+} ion positioned at the observed ion density. The systems were energy minimised and then equilibrated with strong position restraints applied on both the protein and the filter ion. After initial equilibration, three replicates were simulated where position restraints on the protein and the filter ion were relaxed in a step-wise manner. Firstly, 25 ns simulation was performed with position restraints applied to the protein backbone and the filter ion with a force constant of 1000 kJ/mol/nm^2. Protein side chains were restrained with slightly weaker force constant (500 kJ/mol/nm^2). Then, for a further 50 ns, the restraining force constant was reduced to 500 kJ/mol/nm^2 for the protein backbone and the filter ion. Protein side chains were fully relaxed. Finally, all restraints were removed and the systems were simulated for 100 ns to allow the filter ion and its surrounding residues to explore possible interactions more fully.

As illustrated by the cryo-EM densities (Figure 5.3), the putative ion in the PC2/PI(4,5)P_2 structure is nicely coordinated by four backbone carbonyls of Leu641. In comparison, the ion position in the PC2/PI(3,5)P_2 dataset is further into the central pore by 0.12 nm. When Na^+ was placed into the density, the ion in the PC2/PI(4,5)P_2 structure stayed dehydrated in the plane of backbone oxygens of Leu641, which is consistent with what we observed in cryo-EM. With

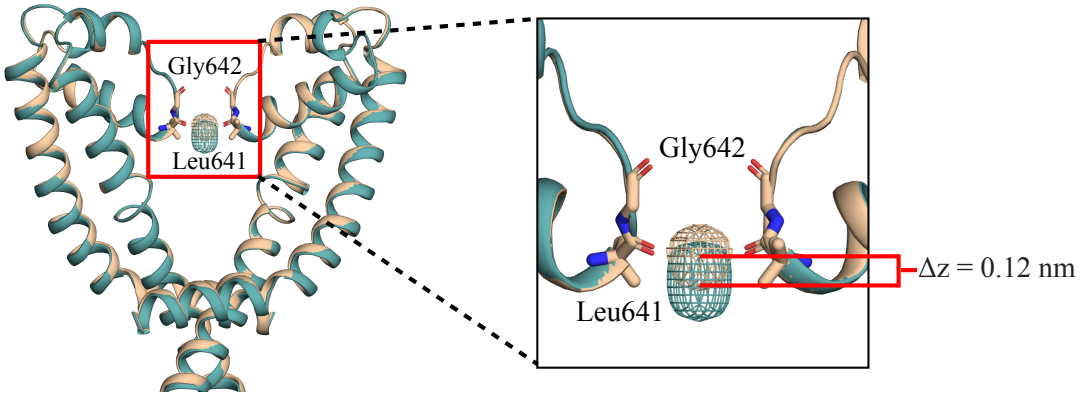


Figure 5.3: Overlay of likely putative ion density in PC2/PI(4,5)P_2 (pale orange) and PC2/PI(3,5)P_2 (light teal) datasets. Zoomed-in view is shown in the right panel with the distance between the centres of the densities indicated.

position restraints, the Na^+ ion stayed relatively stable at the position observed in cryo-EM (Figure 5.4A − 5.4D). When restraints on both the protein and the ion were removed, the filter ion passed through the selectivity filter and entered into the central vestibule in all three replicates of simulations (Figure 5.4D). As shown in the Figure 5.4C, at the end of the simulation, the Na^+ ion sits in the vestibule between the selectivity filter and the hydrophobic Phe669 residues. To assess the dynamics of water and ions through the pore of PC2, a pore region was defined as a cylindric zone centred in the centre of geometry (COG) of the protein with a radius of 15 Å that extends from 20 Å above the COG to 50 Å below the COG. The Z-axis coordinates of the filter ion and water, Na^+ and Cl^- ions within the defined pore region were plotted over the simulated time. Not surprisingly, while the Na^+ ion is coordinated by the backbone oxygens of Leu641, it blocks the entry of water molecules (Figure 5.4E).

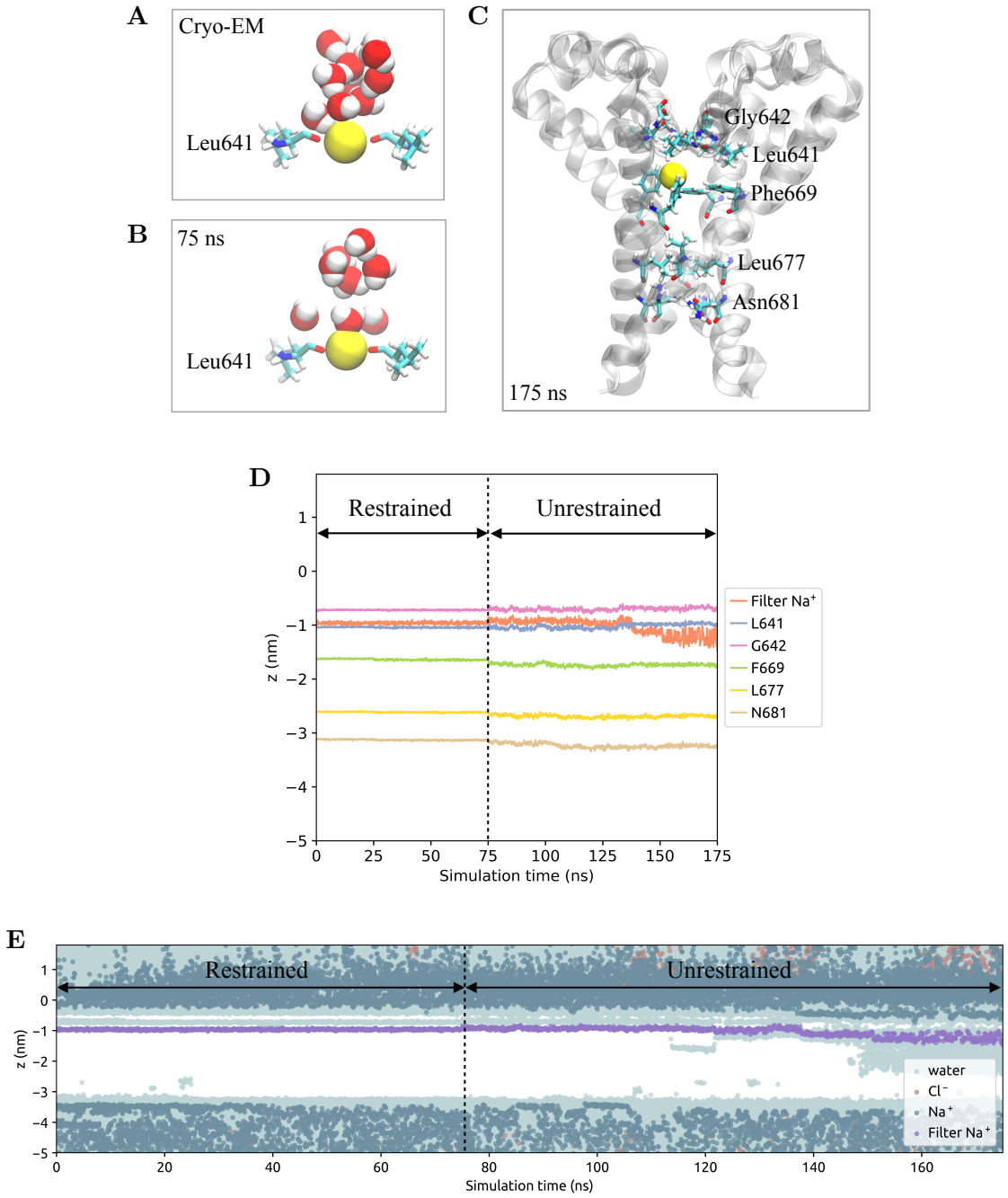


Figure 5.4: A, B: Positions for the filter Na^+ ion in the PC2/PI(4,5)P_2 system corresponding to the cryo-EM position and at the end of position-restrained equilibration simulation respectively. Filter Na^+ ion (yellow) with adjacent water molecules are shown in spheres. Leu641 is shown in sticks. **C:** Final snapshot of the channel pore helices showing the position of the filter ion. **D:** The Z-axis coordinates of the centre of mass of the filter Na^+ ion, the selectivity filter (Leu641, Gly642) and the lower gate (Leu677, Asn681) in the PC2/PI(4,5)P_2 structure were plotted throughout one replica of simulations. **E:** The Z-axis coordinates of the filter ion and the solvent within the pore region throughout one production simulation of the PC2/PI(4,5)P_2 structure.

For the structure of PC2/PI(3,5)P_2, the ion is not stably coordinated by the backbone oxygens of Leu641, and thus at energy minimisation step, it has already adjusted to a position similar to what is seen in the PC2/PI(4,5)P_2 structure (Figure 5.5A & 5.5B). The readjustment leads to larger fluctuation of the ion position comparing to the PC2/PI(4,5)P_2 structure even with strong position restraints applied. As before, under fully relaxed conditions, the filter Na^+ ion was able to move into the central cavity without an external electric potential applied (Figure 5.5D). However, no further Na^+ ion entered. It is possible that in equilibrium simulations, the electrostatic repulsion of the cavity ion is enough to prevent the entry of a second ion. It is noted that once the filter ion enters the central cavity, it fluctuates on either side of Phe669. Together with what has been observed for the filter ion in the PC2/PI(4,5)P_2 structure, where it spent shorter time in the central cavity and stayed solely above Phe669, it is possible that Phe669 imposes a hydrophobic barrier that impede the free movement of ion within the central cavity.

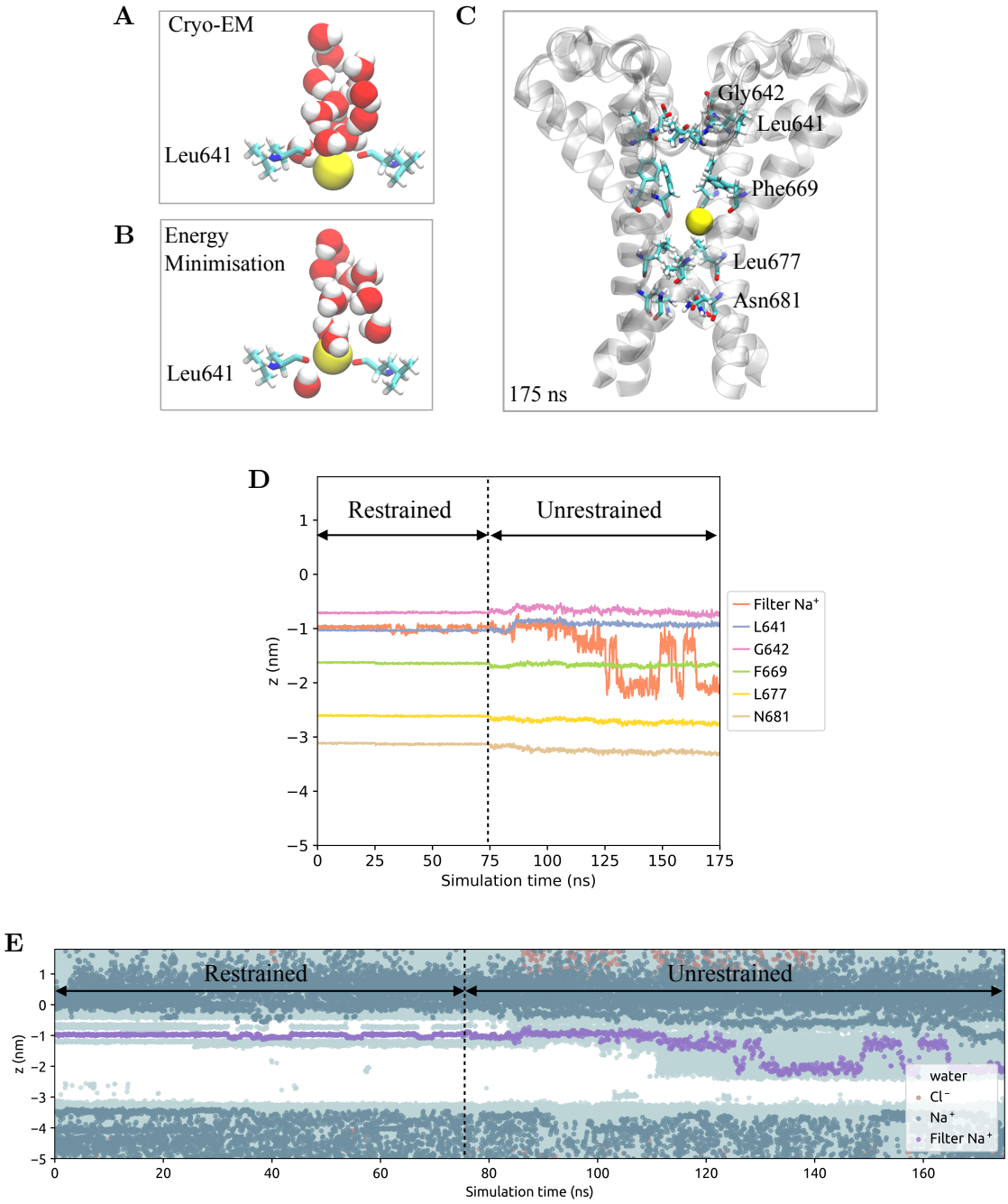


Figure 5.5: **A, B:** Positions for the filter Na^+ ion in the PC2/PI(3,5)P_2 system corresponding to the cryo-EM position and after energy minimisation step respectively. Filter Na^+ ion (yellow) with adjacent water molecules are shown in spheres. Leu641 is shown in sticks. **C:** Final snapshot of the channel pore helices showing the position of the filter ion. **D:** The Z-axis coordinates of the centre of mass of the filter Na^+ ion, the selectivity filter (Leu641, Gly642) and the lower gate (Leu677, Asn681) in the PC2/PI(3,5)P_2 structure were plotted throughout one replica of simulations. **E:** The Z-axis coordinates of the centre of mass ofthe filter ion and the solvent within the pore region throughout one production simulation of the PC2/PI(3,5)P_2 structure.

When Ca^{2+} ion was built into the cryo-EM density, for both PC2/PI(4,5)P_2 and PC2/PI(3,5)P_2 structures, the ion quickly adapted to a new position sitting between Leu641 and Gly642 (Figure 5.6 & 5.7). While it is at the selectivity filter, the ion remains dehydrated in the xy-axis plane. The filter Ca^{2+} ion was ejected outwards relative to the selectivity filter shortly after the systems were completely unrestrained in all simulations (Figure 5.6D & 5.7D). This may reflect the experimentally reported selectivity of PC2 for Na^+ over Ca^{2+} [146]. Given that the hydration energy of Ca^{2+} ion is -1577 kJ/mol comparing to -406 kJ/mol for Na^+ ion, it is possible that the coordination by the selectivity residues is not enough to compensate the loss of energy due to dehydration. Thus, it is more energetically favourable for the filter Ca^{2+} ion to move away from the filter and re-enter the solution environment. Once the Ca^{2+} ion moves to the plane of backbone oxygens of Gly642, it becomes partially hydrated, and water molecules were able to enter the central vestibule through the selectivity filter. However, the binding of Ca^{2+} ion at the selectivity filter blocks the entry of Na^+ ion (Figure 5.6E & 5.7E). When the Ca^{2+} ion moves away from Gly642, Na^+ ion could quickly binds at the selectivity filter and move toward the central pore. Similar to previous filter Na^+ simulations, ion binding at Leu641 prevents entry of water and a maximum of one Na^+ ion is present in the central cavity.

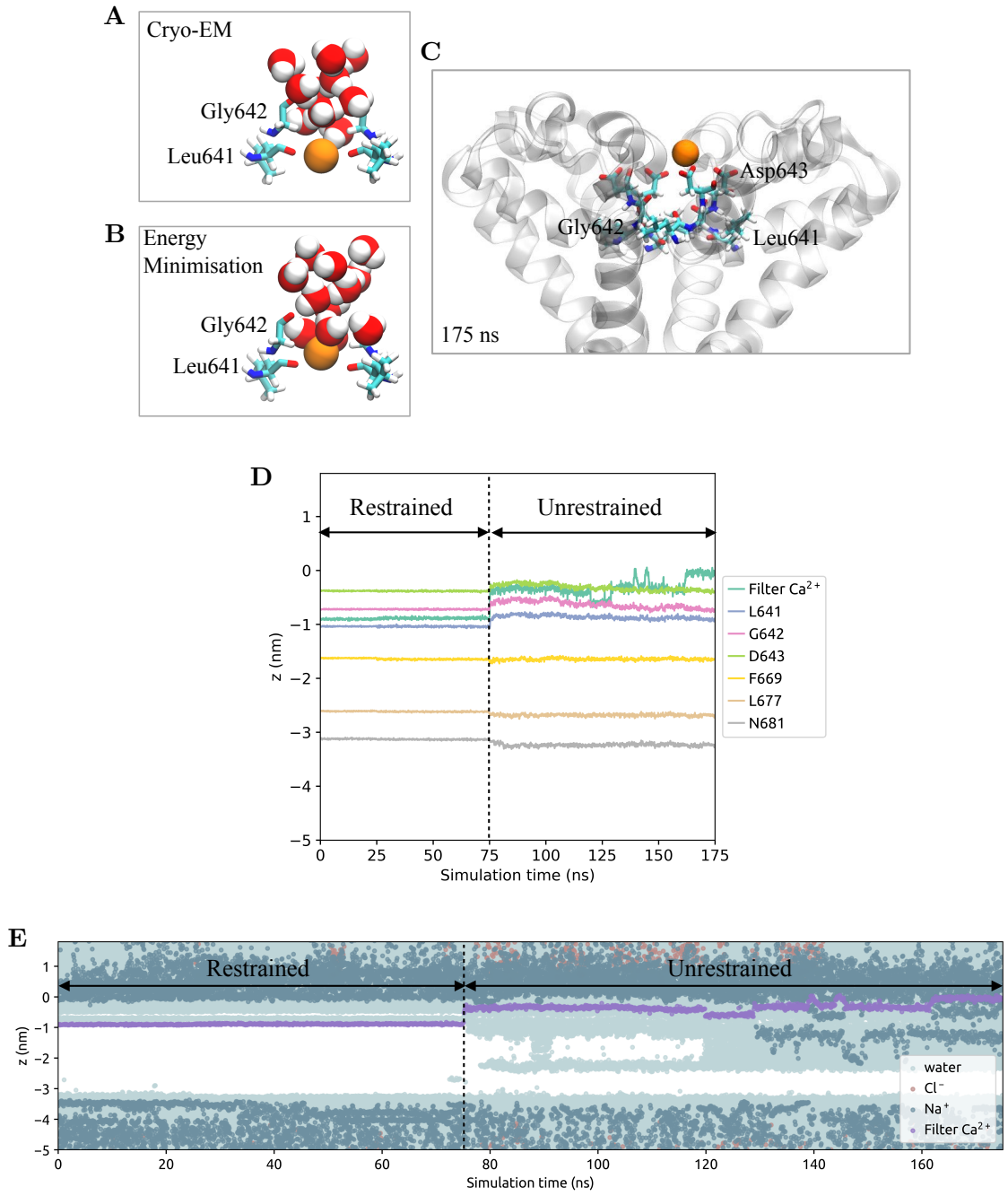


Figure 5.6: **A, B:** Positions for the filter Ca^{2+} ion in the PC2/PI(4,5)P_2 system corresponding to the cryo-EM position and after energy minimisation step respectively. Filter Ca^{2+} ion (orange) with adjacent water molecules are shown in spheres. Residues of the selectivity filter (Leu641, Gly642) are shown in sticks. **C:** Final snapshot of the channel pore helices showing the position of the filter ion. **D:** The Z-axis coordinates of the centre of mass of the filter Ca^{2+} ion, the selectivity filter, Asp643 and the lower gate (Leu677, Asn681) in the PC2/PI(4,5)P_2 structure were plotted throughout one replica of simulations. **E:** The Z-axis coordinates of the filter ion and the solvent within the pore region throughout one production simulation of the PC2/PI(4,5)P_2 structure.

In all simulations, after ejection from the selectivity filter, the Ca^{2+} ion interacts with the acidic residue Asp643, which is located one amino acid above the filter. Because Asp643 sits upper on the pore helix of each subunit, the side chain distance between opposite chains is more than 1 nm throughout the simulations. The Ca^{2+} ion is, therefore, not coordinated by four Asp643 residues, and rather binds to just one of these residues (Figure 5.6C & 5.7C). Electrophysiology experiments on the constitutively open mutant F604P of PC2 showed that when Asp643 is mutated to asparagine, the reversal potential of the single mutant F604P in Na^+ solution shifts from -10.9 mV to 9.7 mV of the double mutant F604P_D643N [309], which suggests that Asp643 is important for ion permeation. It is possible that this negatively charged residue facilitates the concentration of cations near the selectivity filter, and hence aid the ion flow through the pore. However, further investigation in the role of Asp643 in Na^+ vs. Ca^{2+} ion selectivity is still needed. It is interesting to note that in voltage-gated sodium (Na_V) and calcium (Ca_V) channels, there is a highly conserved acidic residue on the P2 helix before the selectivity filter [310–312], which is at a similar location as Asp643 in PC2. This may suggest that Asp643 has similar functional role as the pre-filter acidic residue in Na_V and Ca_V channels.

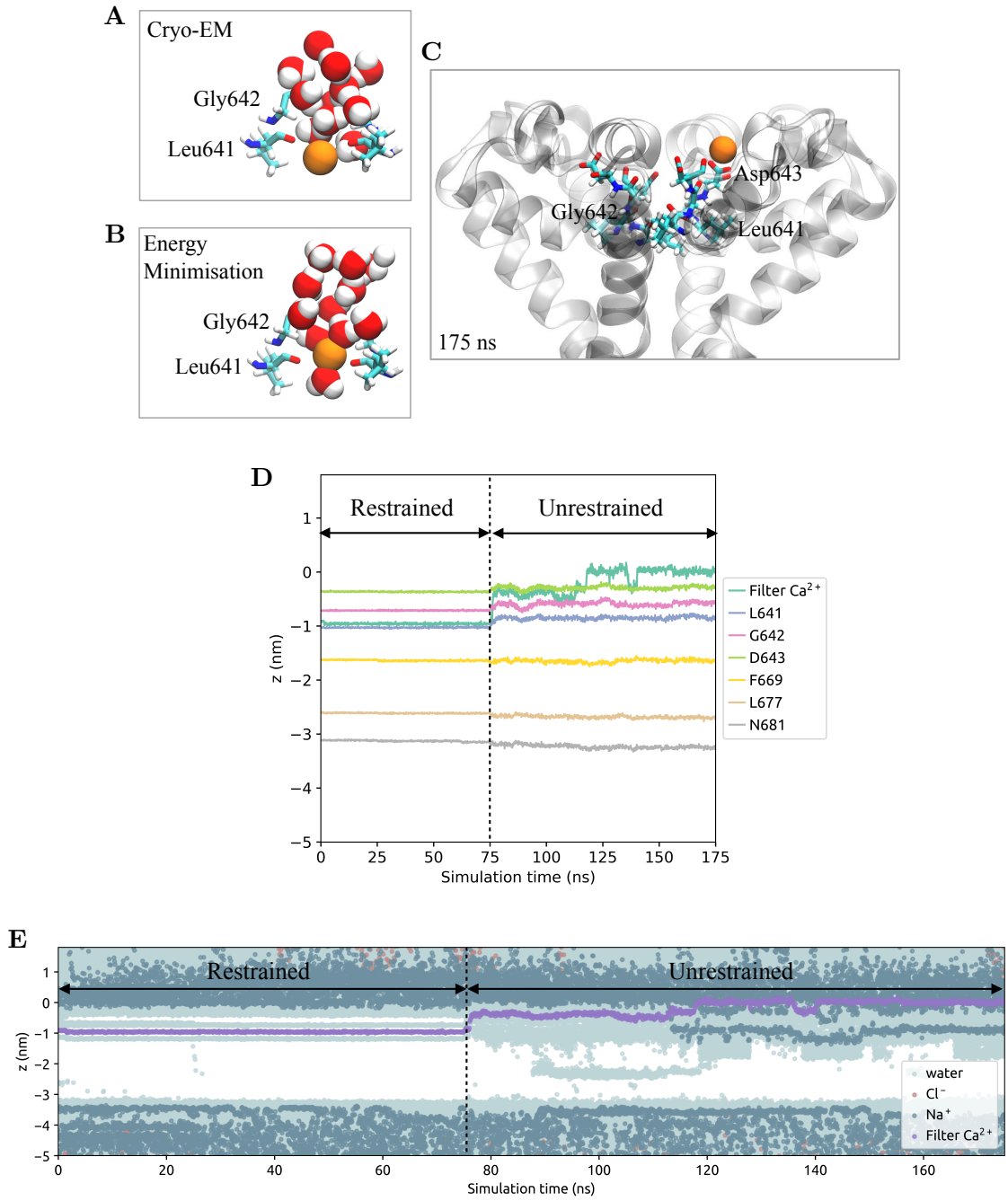


Figure 5.7: **A, B:** Positions for the filter Ca^{2+} ion in the PC2/PI(3,5)P_2 system corresponding to the cryo-EM position and after energy minimisation step respectively. Filter Ca^{2+} ion (orange) with adjacent water molecules are shown in spheres. Residues of the selectivity filter (Leu641, Gly642) are shown in sticks. **C:** Final snapshot of the channel pore helices showing the position of the filter ion. **D:** The Z-axis coordinates of the centre of mass of the filter Ca^{2+} ion, the selectivity filter, Asp643 and the lower gate (Leu677, Asn681) in the PC2/PI(3,5)P_2 structure were plotted throughout one replica of simulations. **E:** The Z-axis coordinates of the filter ion and the solvent within the pore region throughout one production simulation of the PC2/PI(3,5)P_2 structure.

5.3.2 Ion Permeation and Selectivity of PC2 at the Selectivity Filter and TOP Domain

To investigate the passage of ion through the pore of PC2, we first examined all of the equilibrium atomistic simulations of PC2 (see Table 4.1 & 5.1). In order to determine the route of entry for ions into the external vestibule that is enclosed by extracellular/luminal (TOP) domain above the selectivity filter, we traced the movements of 20 distinct Na^+ ions that have been within 3 Å of the selectivity filter in the simulations of PDB ID: 5K47. Although there are fenestrations lateral to the pore axis, they are mostly blocked by the lipid head groups throughout the simulations (Figure 5.8A & 5.8B). It has been seen that almost all of the ions enter the TOP domain through the inter TOP domain groove with only one diffusing through the central apical opening (Figure 5.8C). The side chains of Lys264 and Lys267 extend into the central apical opening (Figure 5.8D). It is possible that the electrostatic repulsion of the positive charges hinders the access of cations. However, water molecules are able to flow through the apical opening freely. In addition, the negatively charged side chains of residues Glu266, Glu277, Asp278 and Asp289 around the inter TOP domain groove could facilitate the assembly of cations near that area (Figure 5.8E), which may help ions passing through the groove.

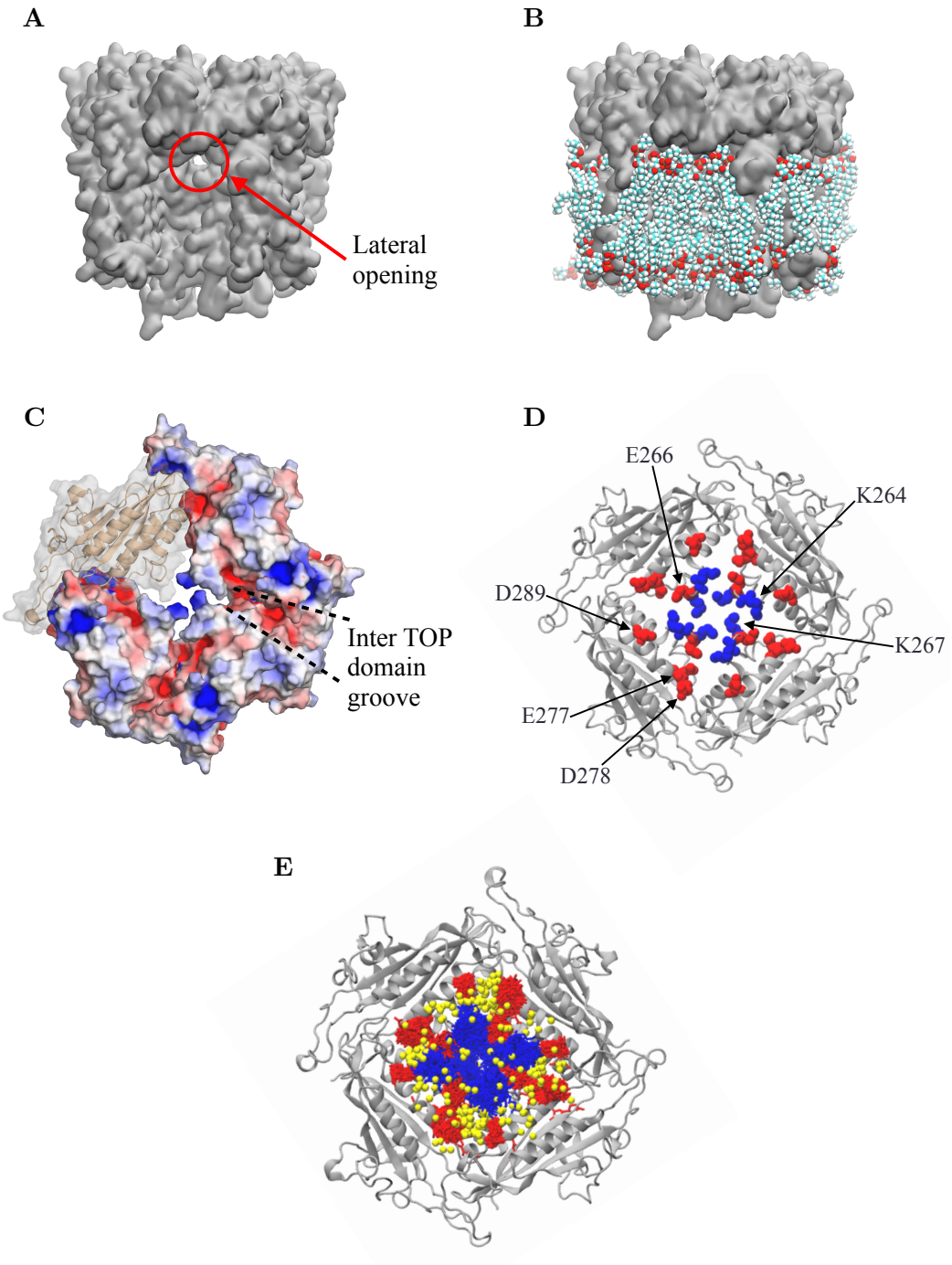


Figure 5.8: **A:** Surface representation of PC2 showing the lateral fenestration on the protein. **B:** A simulation snapshot showing close-by POPC lipids around the protein that block the fenestration. **C:**Top-down view of the TOP domain possible ion pathway labelled. **D:** Side chains of residues Lys264 and Lys267 projecting into the TOP domain apical opening. Residues lining the inter TOP domain grooves are negatively charged. **E:** Side chain positions of residues Lys264, Lys267, Glu266, Glu277, Asp278 and Asp289 during 250 ns atomistic simulation were sampled every nanosecond. Residues are coloured by residue type. Positions of Na^+ ions (small yellow dots) around the TOP domain during 250 ns atomistic simulation were also sampled every nanosecond.

Simulations of all available PC2 structures apart from PDB ID: 5K47 and 5T4D show that one Na^+ ion was able to permeate through the selectivity filter into the central vestibule without an imposed trans-bilayer voltage difference. For simulation of PDB ID: 5K47 and 5T4D, one Na^+ ion was placed inside the central chamber at the beginning of the simulations. Although water molecules are able to move through the selectivity filter, no ion passage was observed, which is consistent with what we observed in Subsection 5.3.1. To exclude the possibility that ion permeation is structure specific, one copy of the simulations was repeated for each of those structures removing the central ion. Both structures show permeation of Na^+ ion at the selectivity filter. Given that protein samples used for cyro-EM structural determination are usually prepared under non-physiological conditions, it is perhaps not surprising that after relaxation within lipid bilayer at physiological temperature, the selectivity filter becomes wide enough for Na^+ ion to pass through, even though the effective radius shown in cryo-EM structures does not permit ion passage.

Liu *et al.* have shown that ciliary PC2 is preferentially selective for monovalent cations over divalent cations [146]. Also, based on the results in Subsection 5.3.1, PC2 is more likely to be Na^+ selective comparing to Ca^{2+}. To investigate the selectivity properties of PC2, we performed equilibrium simulations replacing NaCl with equal molar of $CaCl_2$, even though this is not the physiological concentration of Ca^{2+}. In all simulations, Ca^{2+} ions within the TOP cavity are bound to Asp643, but no Ca^{2+} ion passed through the selectivity filter. This is in good agreement with the results shown in Subsection 5.3.1.

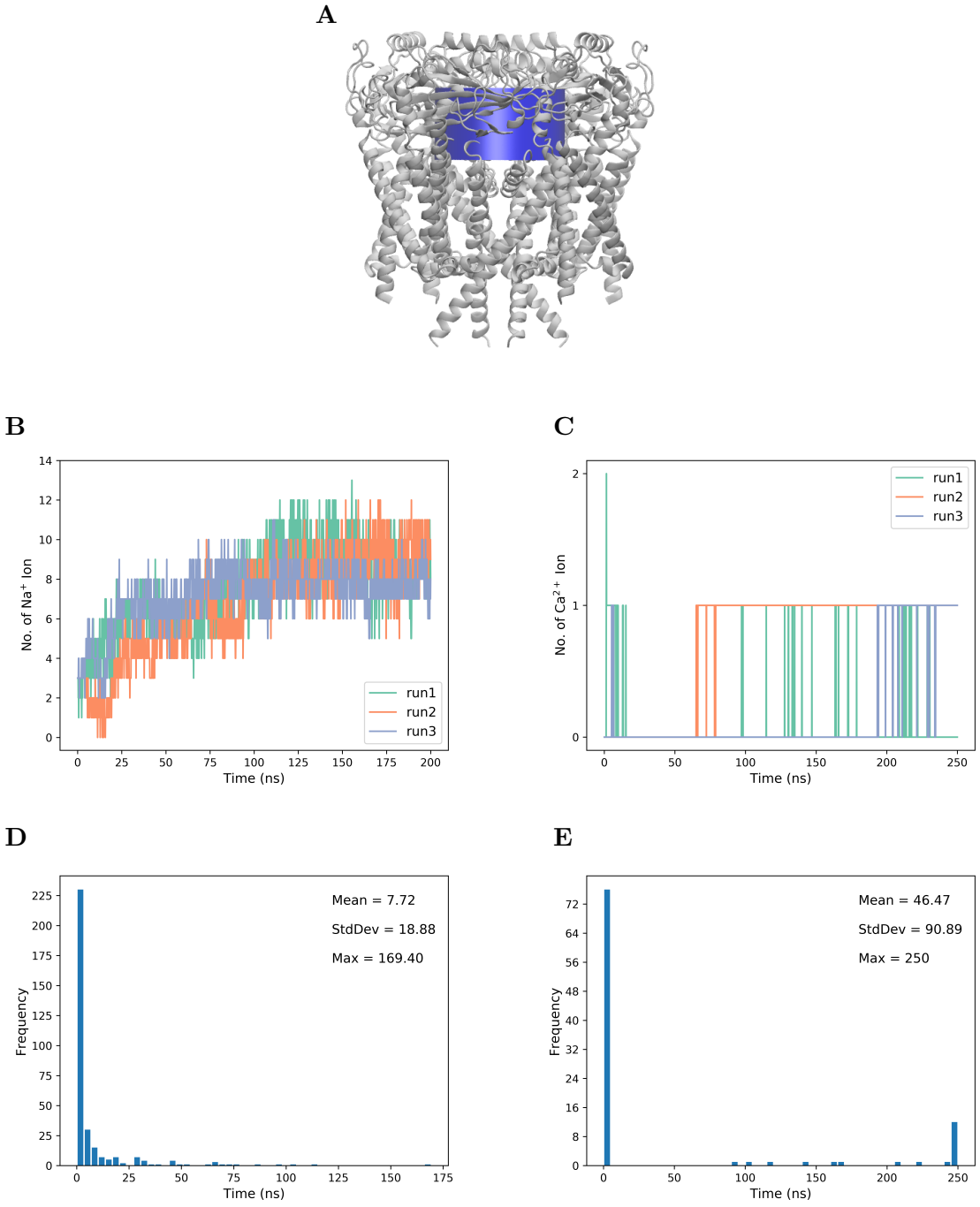


Figure 5.9: **A:** Illustration of the defined TOP domain region. Ions within the blue cylinder were counted as inside of TOP cavity. **B:** Number of Na^+ ion in TOP vestibule in three repeats of equilibrium simulations of PDB ID: 5K47. **C:** Number of Ca^{2+} ion in TOP vestibule in three repeats of equilibrium simulations of PDB ID: 5K47. The distribution of continuous contacting time between the acidic residues lining the inter TOP domain groove and **D:** Na^+ and **E:** Ca^{2+} ion in equilibrium simulations.

It was noticed that Ca^{2+} ions could not enter the TOP domain as easily as Na^{+} ions. To quantify the difference, the TOP cavity was define as a cylindric zone centred in the COG of the protein with a radius of 20 Å that extends from 19 Å above the COG to 1 Å below the COG (Figure 5.9A). Then the number of Na^{+} or Ca^{2+} ions within the TOP vestibule was counted over time (Figure 5.9B & 5.9C). The interior of the TOP domain is relatively electronegative. The number of Na^{+} ions plateaus around eight, which indicates the maximum positive charge that the TOP cavity could contain is around eight given that there is no conductance of ion through the channel and no electric field across membrane. The average Na^{+} ion density within TOP domain is, therefore, ~ 0.32 ion/nm^3, which is higher than that of bulk (~ 0.1 ion/nm^3). The electronegative TOP domain cavity is functioning to concentrate Na^{+}. However, the average Ca^{2+} density within the TOP domain is about half of the bulk density. It can be seen that the entry of Ca^{2+} ions is hindered by the TOP domain. Ca^{2+} ion is depleted from the TOP vestibule in comparison to the surrounding environment. Even though Ca^{2+} is divalent, the maximum charge of Ca^{2+} ions within the TOP cavity is four in the simulations, and thus it is not the saturation of charges that impede the entry of Ca^{2+} ions. Since ions predominantly enter through the TOP domain groove surrounded by negatively charged residues. It is possible that the coordination of Ca^{2+} ion at the grooves are tighter than Na^{+} ion, and hence blocks the pathway for ion to enter the TOP cavity. To evaluate the binding of ions, we counted the continuous binding time of ion to residues Glu266, Glu277, Asp278 and Asp289. Binding is defined as being within 5 Å of these residues. The mean and standard deviation of the binding times were calculated across all replicate simulations. As shown in Figure 5.9D & 5.9E, the average binding time of Ca^{2+} ion is much longer than Na^{+} ion, and once Ca^{2+} ion binds to the TOP domain groove, it tends to stay bound. Without any electrochemical gradient, it lacks forces to push the bound ions into the TOP domain.

To further investigate the ion permeation of PC2, we performed simulations with an electric field of ~ 200 mV applied across the lipid bilayer in the direction of $-$ Z axis (i.e. $\Delta V \simeq -200$ mV; see Figure 5.2). For initial exploration, we started

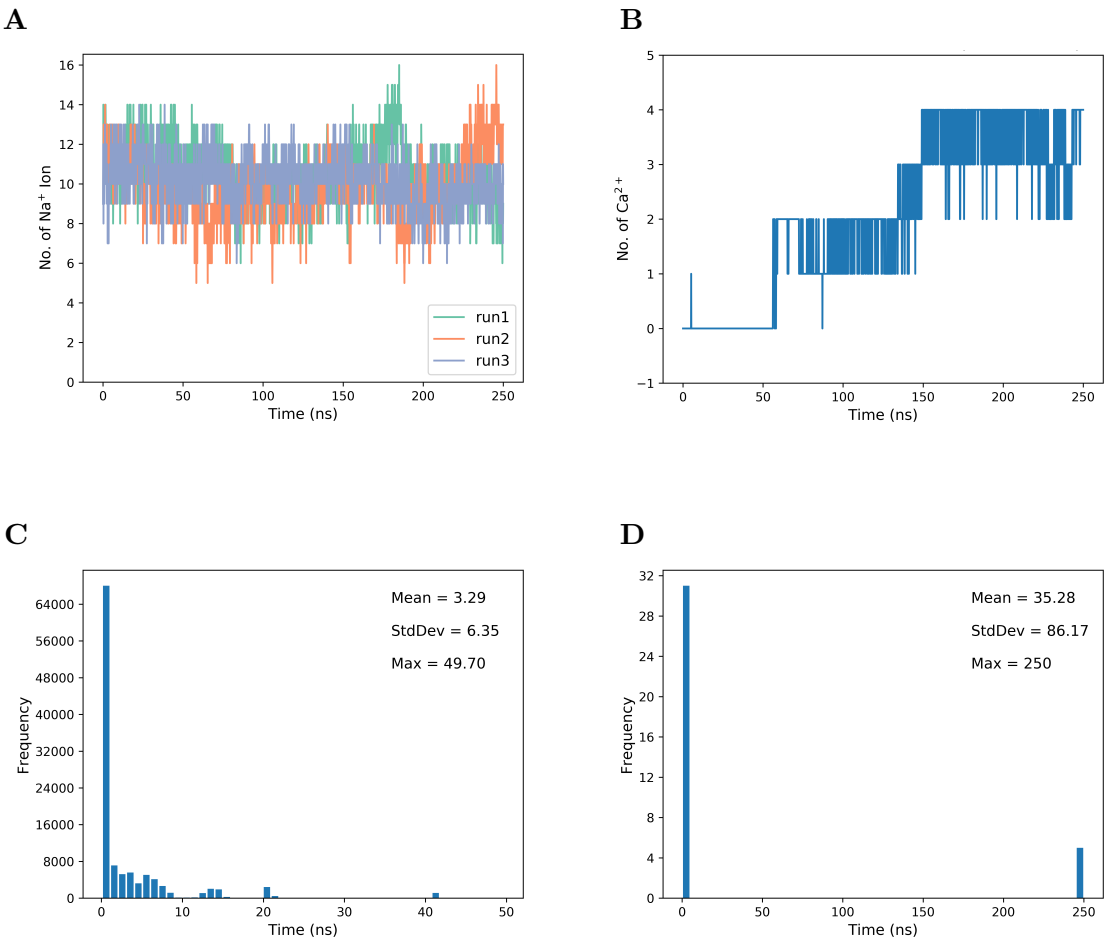


Figure 5.10: **A:** Number of Na^+ ion in TOP vestibule in three repeats of simulations of PDB ID: 5K47 with -200 mV electric field applied across the membrane with 150 mM ion concentration. **C:** Number of Ca^{2+} ion in TOP vestibule in one simulation of PDB ID: 5K47 with -200 mV electric field applied across the membrane with 150 mM ion concentration. The distribution of continuous contacting time between the acidic residues lining the inter TOP domain groove and **D:** Na^+ and **E:** Ca^{2+} ion in computational electrophysiology simulations.

the simulations from the final snapshot of one of the equilibrium simulations of PDB ID: 5K47 and the ion concentration was maintained at 150 mM. The number of Na^+ ion within the TOP vestibule increased slightly to an average around 10 (Figure 5.10A). With applied transmembrane voltage, two Na^+ ions were able to pass through the selectivity filter and enter the central cavity in all simulations, which may reflect the maximum charge capacity of the central cavity at given

electric field. However, there was still no conductance through the lower gate. The binding time of Na^+ at the inter TOP domain groove decreased from an average of 7.7 ns to 3.3 ns. The maximum binding time dropped from 169 ns to 50 ns (Figure 5.10C). This suggests that the electric potential could drive the bound ions off the negatively charge residues lining the groove, and thus facilitate the movement of ion into the TOP domain cavity. For Ca^{2+} ion, the number of entry to the TOP vestibule increase to a maximum of 4 (Figure 5.10B), which holds similar charges as Na^+ ion within the volume. However, there is still no Ca^{2+} passing the selectivity filter with the electric potential applied. The average binding time at the entry grooves decreased from 46 ns to 35 ns (Figure 5.10D). Although the maximum binding time remained at 250 ns with or without external potential, the frequency of fully bound events declined, which may contribute to the increased entry of Ca^{2+} ions.

We then performed computational electrophysiology on our 3 Å PC2/PI(4,5)P_2 structure for 200 ns with increased ion concentration of 350 mM to explore how local concentration would impact ion interactions with PC2. To test the robustness, we ran the simulations of NaCl in both CHARMM36 [273] and OPLS [306] force fields. The results show that with higher bulk ion concentration, the number of Na^+ ion within TOP cavity reached equilibrium at an average of 14 (Figure 5.11A), which is marginally more than what we observed at lower concentration. The binding time at the entry grooves decreased further from an average of 3.3 ns to 0.9 ns (Figure 5.11C). This suggests that with a higher ion concentration, the higher rate of ion collision at the inter TOP domain groove causes higher rate of dissociation for the weakly bound Na^+ ion. However, the number of Ca^{2+} ion within the TOP cavity and the average ion binding time remained in a similar range as in the 150 mM simulations (Figure 5.11B & 5.11D). The strong interaction between Ca^{2+} ion and the negatively charge residues still prevented ion entry into the TOP vestibule.

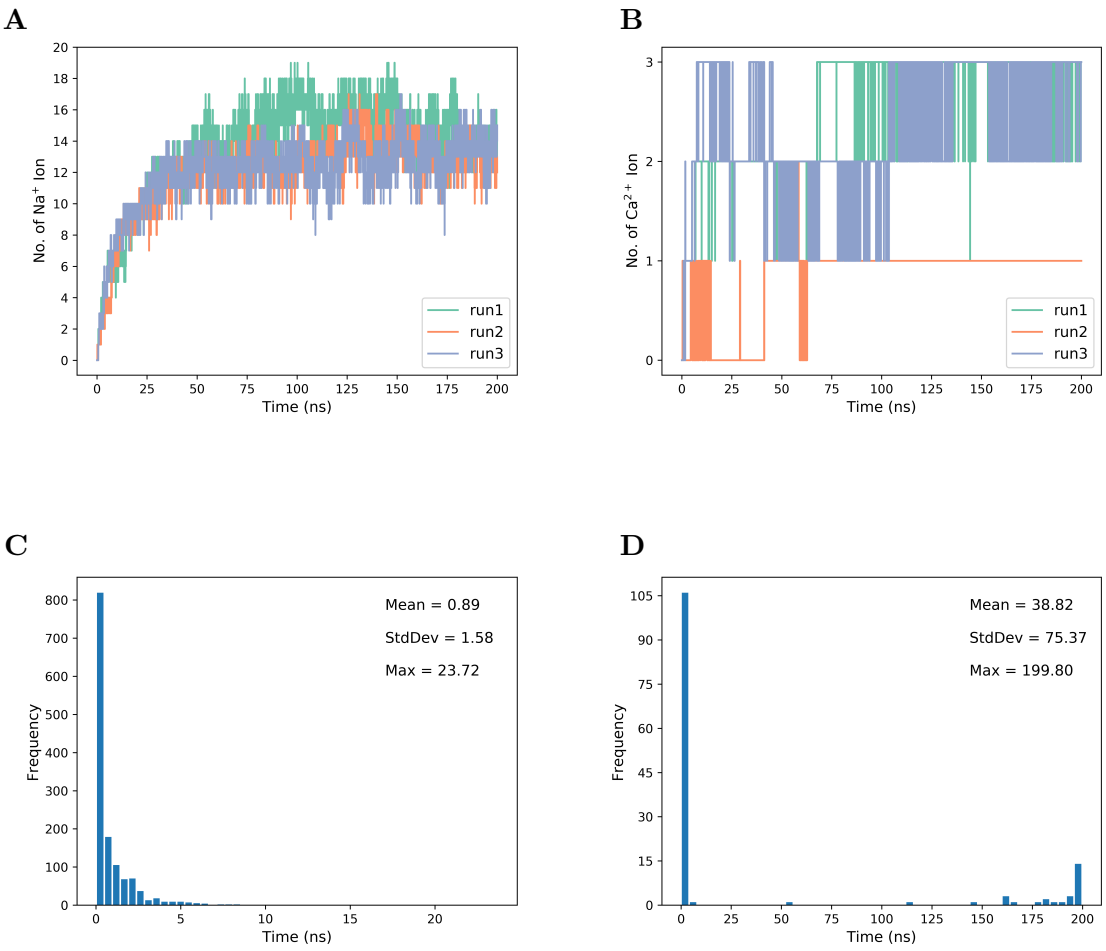


Figure 5.11: **A:** Number of Na^+ ion in TOP vestibule in three repeats of simulations of PC2/PI(4,5)P_2 with -200 mV electric field applied across the membrane with 350 mM ion concentration. **C:** Number of Ca^{2+} ion in TOP vestibule. The distribution of continuous contacting time between the acidic residues lining the inter TOP domain groove and **D:** Na^+ and **E:** Ca^{2+} ion in computational electrophysiology simulations.

5.3.3 Ion and Water Permeation at Central Cavity and Lower Gate of PC2

Visualisation of the ion and water positions over the simulation time confirmed the closed state of the PC2 channel lower gate for all available structures, which has neither ion nor water passing through the gate residue Leu677. The structural study of the constitutively open mutant of PC2 (F604P, PDB ID: 6D1W) showed that the mutation causes the middle of S6 to undergo a transition from a π-helix to a canonical α-helix, this leads to the lower gate residues Leu677 and Asn681 changing from pore-facing to helix-facing, and Phe676 becoming pore-facing instead (Figure 5.12B & 5.12C). This results in a dilation of the pore diameter from 5 Å to 7.6 Å [239]. However, the time averaged radius profile of this structure shows the radius of the constriction at the lower gate region is comparable to other PC2 structures throughout the simulations (Figure 5.12D & 5.12D). During the simulations, the side chain distance between Phe676 on opposite chains fluctuates around an average of 6.5 Å, which is similar to the side chain distance between Leu677 in simulations of wild type PC2. Based on the water density, the free energy of water along the channel pore was calculated using Boltzmann inversion [308]. Both wild type and F604P mutant showed an energetic barrier for water permeation at the lower gate (Figure 5.12F & 5.12G). The functional data of F604P mutant was collected with full length PC2. However, the structure of the mutant was obtained from a truncated construct [239]. It is possible that the structure of the truncated F604P mutant does not fully reflect the functional consequence. Thus, even though the F604P channel is shown to be constitutively open functionally, the lower gate region is still dehydrated throughout the simulations based on the truncated structure.

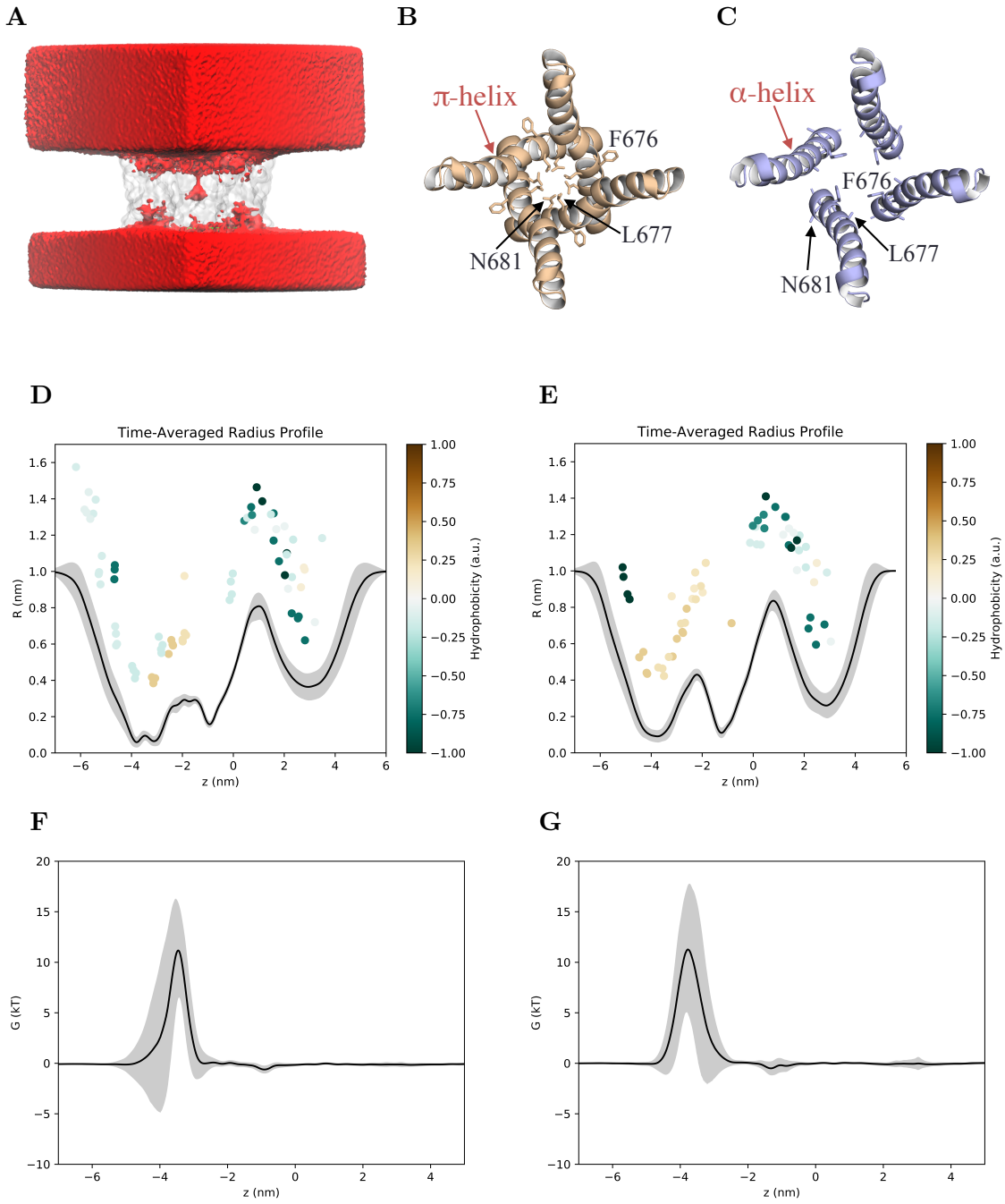


Figure 5.12: Water permeation of PC2. **A:** Probability density of water in simulation of PDB ID: 6D1W showing no water permeation through the lower gate. Time averaged pore radius of PDB ID: 5K47 (**B**) and PDB ID: 6D1W (**C**) with standard deviation shown as grey shading. Hydrophobicity of pore lining residues are indicated along the pore axis. Free energy of water along the channel pore for PDB ID: 5K47 (**D**) and PDB ID: 6D1W (**E**).

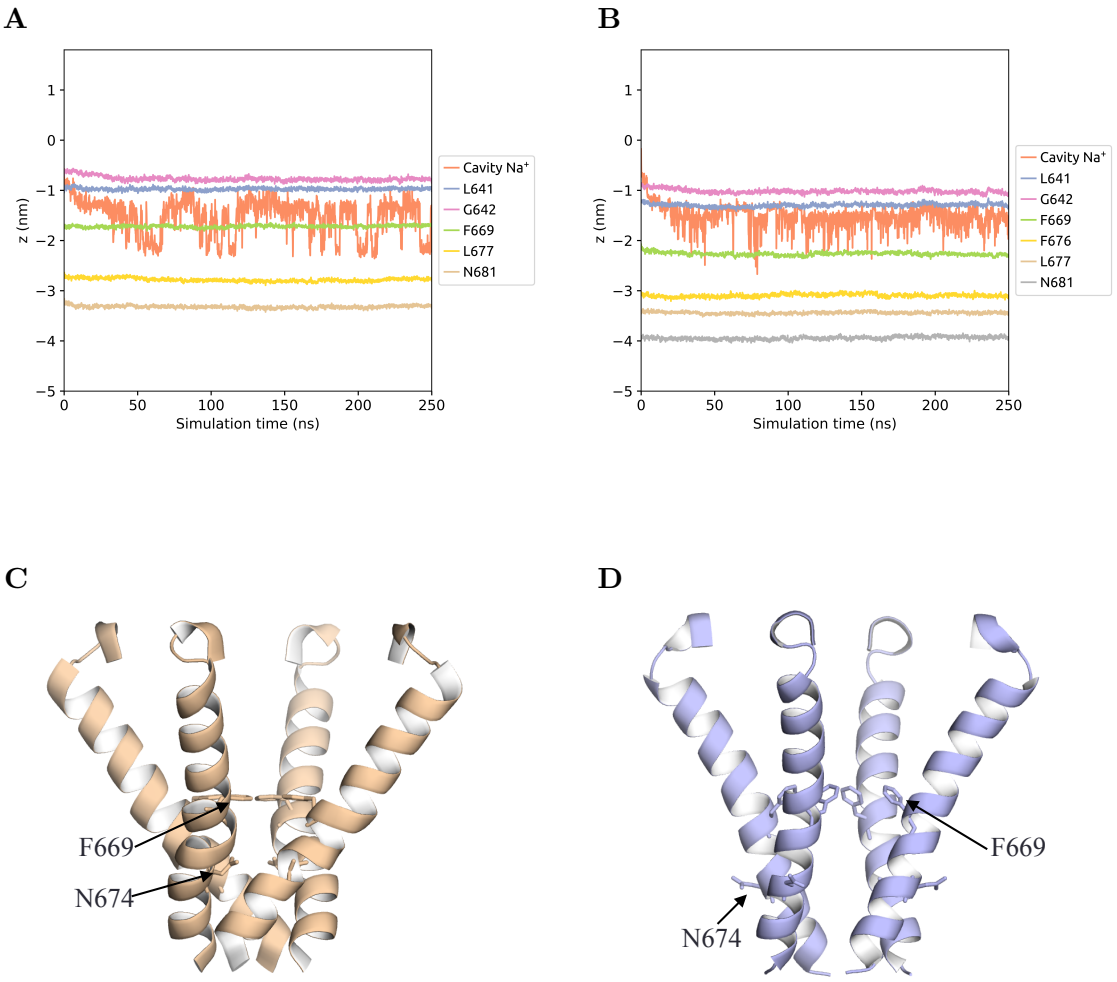


Figure 5.13: The Z-axis coordinates of the cavity Na^+ ion, residues in the selectivity filter, the lower gate and the constriction residue Phe669 shown for wild type PC2 (**A**) and F604P (**B**) over time. Cryo-EM structures of S6 helix in cartoon representation with Phe669 and Asn674 in sticks are shown for our PC2/PI(4,5)P_2 structure (**C**) and PDB ID: 6D1W (**D**).

Analysis of the movements of the cavity ion showed that it is confined within the central vestibule in all simulations (Figure 5.13). However, as observed in Subsection 5.3.1, there is a hydrophobic barrier imposed by Phe669 for the Na^+ ion to move freely within the central cavity. As shown in Figure 5.13A, the Na^+ ion spends most of the time on either side of this hydrophobic residues. It is interesting to note that in the simulations of F604P mutant, the cavity Na^+ ion

largely remained in the upper part of the central cavity above Phe669, close to the carbonyl oxygens of Leu641 (Figure 5.13B). The twisting of S6 in F604P causes the only pore-facing polar residue Asn674 to turn away from the pore (Figure 5.13C & 5.13D), and hence the lower central cavity is more hydrophobic than that of the wild type PC2. Although the entire central cavity is hydrated, the more hydrophobic confined region might not allow the Na^+ ion to be well coordinated.

5.4 Conclusions

Using equilibrium atomistic simulations, we first investigated the identity of ion corresponding to the spherical density observed in our cryo-EM data. The results suggested that the coordination of a dehydrated Ca^{2+} ion at the selectivity filter is not favourable, which might be due to high solvation energy of Ca^{2+}. However, Na^+ ion can be nicely coordinated by the carbonyl oxygens of Leu641 given that the backbone positions of the protein are kept close to the cryo-EM structures. Therefore, under the conditions that cryo-EM data were collected, it is likely that the ion density correspond to a dehydrated Na^+ in contrast to our initial hypothesis.

Examination of all simulations of PC2 suggested that ion preferentially enters the TOP vestibule through the inter TOP domain groove rather than the apical or the lateral openings. Two residues Lys264 and Lys267 on the loop project into the apical opening repels cation from this region. In the mean while, the negatively charge residues Glu266, Glu277, Asp278 and Asp289 along the groove help concentrating cations to the groove and thus facilitate entry of monovalent ions. Simulations of PC2 in $CaCl_2$ showed that the concentration of Ca^{2+} ion within the TOP vestibule is lower than that of bulk even at supraphysiological levels of Ca^{2+} ion. In contrast, the concentration of Na^+ ion within the TOP cavity is three times of bulk. Our simulations suggest that the entry of Ca^{2+} ion into the TOP domain is hindered by the binding of Ca^{2+} at the groove. Previous studies have shown that monovalent current of PC2 can be blocked by extracellular Ca^{2+} [305, 309, 313]. This may be achieved by the behaviour of Ca^{2+} ion observed in our

simulations, i.e. the blockage of the ion entry grooves upon binding of Ca^{2+} ion. For the related TRPML1 and TRPML2 channels, there is also a series of acidic residues in the loop of the extracellular/luminal domain projecting towards the apical opening. It has been shown that Ca^{2+} binds with this loop in both TRPML1 (Asp111, Asp114, and Asp115) and TRPML2 (Asp113, Glu114, Asp115) in a pH dependent manner, which is important for channel regulation [80, 314]. This may suggest that the extracellular/luminal domain unique to the TRPP and TRPML subfamilies might be of function importance. In addition, our simulations showed that Na^{+} ion can pass through the selectivity filter even without transmembrane voltage, but Ca^{2+} binds at Asp643 and does not enter the central vestibule with or without electric field. Overall, the results suggest that both the TOP domain and the filter of PC2 impose a selectivity for Na^{+} over Ca^{2+}. This is consistent with experimental recordings of ciliary currents of PC2 [146].

6

Conclusions and Future Directions

The biological importance of PC2 ion channel has been widely demonstrated in different cell types, especially in the kidney tubular epithelia given its implication in ADPKD. However, there still is a lack of knowledge in the function and modulation mechanisms of PC2. In this thesis, by an approach combining cryo-EM and multiscale MD simulations, the structural dynamics of PC2 was explored with an emphasis on lipid and ion interactions.

We identified a hydrophobic pocket formed between S3, S4 and S4-S5 linker as a lipid binding site, which is at an analogous location to the shared binding site for vanilloid ligands and phosphatidylinositol lipids in TRPV channels. Biochemical assay, cryo-EM data and free energy calculation results provided evidence that this site preferentially binds to PIP_2 and related phosphoinositides. Comparison with other ion channels and their lipid interactions suggest that the PIP_2-binding site on PC2 may be of functional importance. A number of TRP and related channels show anionic lipid interactions in this region, close to the S4-S5 helix, which links the VLSD/VSD to the pore domain [6, 53, 154, 300]. For example, TRPV1 with PI, TPRV5 with PIP_2, and Kv1.2/2.1 with anionic lipid interactions

all occur at similar sites to the PC2-PIP_2 interaction site (Figure 6.1A). PIP_2 regulates a number of TRP channels [2] and also a number of Kv channels [315] via interactions with the S4-S5 linker. This suggests that the PC2 PIP_2-binding site near to the S4-S5 linker may be a regulatory/allosteric site. Interestingly, some pathogenic missense variants (see http://pkdb.mayo.edu/) in PC2 occur in the vicinity of the PIP_2 site, e.g. L517R, D511V and N580K (the latter located at the start of the S4-S5 linker), and might be expected to perturb interactions with PIP_2. It would be interesting to combine simulations and experimental studies to probe the effects of mutations on PIP_2 binding to PC2, but this would first require a reliable physiological assay of channel function.

A cholesterol-binding site on PC2 was also determined in this work. Cholesterol has been shown to be involved in ciliary signalling [294], and the regulatory effect of cholesterol has been reported for various receptors and ion channels including TRP channels [295, 297, 303, 304]. Examination of structures of a number of other TRP channels (e.g. TRPM2, TRPML3) and more distantly related channels (e.g. Kv10.1) reveals multiple cholesterol interaction sites between the VSLD/VSD and the pore domain (Figure 6.1B). In addition, It is interesting to note that pregnenolone sulphate has been shown to enhance the activation of PC2 in the presence of TRPM3 [316]. Given the similarity between the structures of cholesterol and pregnenolone, we speculate that the two molecules could occupy the same binding site, placing the sulphate on the extracellular surface, near the pore, thus activating the channels. Therefore, it would not be surprising if cholesterol interaction with PC2 were functionally important. Again, this could be explored by mutagenesis if functional assay was available.

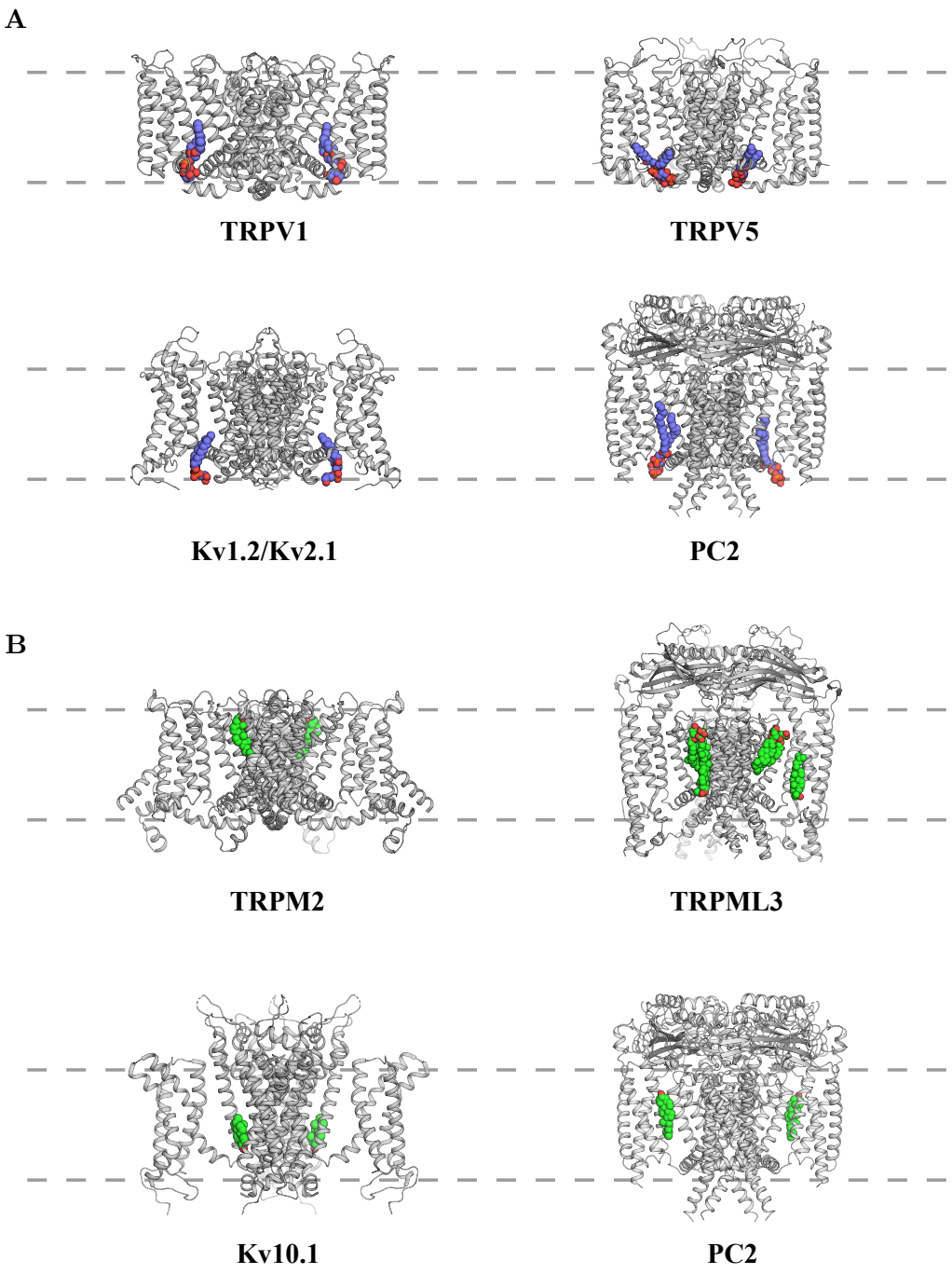


Figure 6.1: Comparison of lipid binding sites of TRP and Kv channels. **A:** Anionic lipid (blue: acyl tails and red: phosphate oxygens) binding sites close to the S4-S5 linker: TRPV1 and PI (PDB ID: 5IRZ), TRPV5 and PIP_2 (PDB ID: 6DMU), Kv1.2/Kv2.1 chimera and PG (PDB ID: 2R9R), PC2 and PIP_2 (this study). **B:** Cholesterol (green: carbons and red: oxygens) binding sites between the VSD/VSLD and the central pore domains of: TRPM2 (PDB ID: 6CO7), TRPML3 (PDB ID: 5W3S), Kv10.1 (PDB ID: 5K7L), and PC2 (this study). In each case the transmembrane region of the channel structure is shown, with the location of the lipid bilayer shown via the grey dashed lines.

Furthermore, in our simulations, it was observed that the electrostatic effect of the TOP domain residues directs the passage of cations through the inter-TOP domain grooves into the external vestibule of PC2. Simulations comparing the transport of Na^+ vs. Ca^{2+} ions showed that the Ca^{2+} ions bind to the groove lining residues more tightly than Na^+ ions, and thus block further ion entry into the external vestibule of PC2. As a result, the concentration of Ca^{2+} ions within the TOP vestibule is depleted comparing to that of Na^+, which indicates that some degree of the ion selectivity happens at the TOP domain before the selectivity filter. The groove lining residues Glu266, Glu277, Asp278 ad Asp289 are all highly conserved, suggesting that the TOP domain may play a crucial role in ion transport and potentially ion selectivity.

The resolution currently achievable in cryo-EM studies of membrane proteins means that it is sometimes difficult to unambiguously identify bound lipids. In particular, certain regions of cryo-EM have lower resolution than other regions of the maps. The regions containing the lipids, usually on the outer surface of the protein, are often of lower resolution, making it more challenging to determine the nature of the bound lipids. Molecular simulations thus provide a valuable tool in assessing possible identities of bound lipids. In our studies, MD simulations are strongly suggestive of PIP_2 and cholesterol binding at the two sites identified on PC2. However, the cryo-EM data only allowed unambiguous confirmation of these predictions in the case of cholesterol, not PIP_2. PIP strip assays were used to identify possible phospholipid species interacting with PC2, but quantification of relative binding affinities for different lipids could not be established. In our PMF calculations, coarse graining of the lipid models does not allow distinction between different PIP species with the same charge. In addition, the PC2 construct used for PIP strips is truncated (as is the construct in the structural studies and the simulations). Therefore, it is uncertain whether full-length PC2 protein would exhibit exactly the same lipid binding behaviour, especially for the PIP binding site which is close to the termini of the truncation.

Having to note the limitations of this study, future work exploring the structural effects of PIP binding on full length PC2 could be proposed in order to better understand the regulation of the channel. If a full length structure were available, then molecular simulations could be used to probe the larger-scale conformational dynamics of the protein in relation to gating. This would probably need a multiscale simulation approach to balance the computational cost and the timescale achievable for exploring protein conformational changes. Although functional studies have proved to be very difficult, mutagenesis studies and development of effective functional assays would be of value to determine the effects of PIP on channel functions. The binding affinity of different PIPs could be further characterised with quantitative binding assays, and free energy calculations with atomistic details might provide a complement in this regard. Additionally, the TOP domain has been shown to be highly N-glycosylated, but the molecular model used for the simulations has the glycans removed. This could introduce structural flexibility in the TOP domain given that de-glycosylation through glucosidase II has been evidenced to decrease biogenesis and stability of proteins [317]. How this potentially extra flexibility affects the function of the channel and how the presence of glycans affects ion interactions with the TOP domain are subjects to further investigation.

Overall, the structural and simulation results presented here suggest that further studies of possible functional effects of PIPs and/or of cholesterol on PC2 channels, and possible functional roles of the TOP domian are merited. From a broader perspective this study demonstrates that cryo-EM and MD together provide a powerful combination for revealing ion and lipid interactions of ion channels, enabling identification of the molecular identity and interactions of lipid-like and ion-like densities observed in structures. This is of particular importance when the resolution of cryo-EM maps is such that possible ion and lipid molecules are difficult to identify unambiguously. Our results help to establish that the interactions of TRP channels with lipids will enable definition of novel druggable sites on this physiologically important class of channel molecules.

References

[1] G. Van Meer, D. R. Voelker, G. W. Feigenson, *Nature Reviews Molecular Cell Biology* **2008**, *9*, 112–124.

[2] T. Rohacs, *Cell Calcium* **2009**, *45*, 554–565.

[3] I. S. Ramsey, M. Delling, D. E. Clapham, *Annual Review of Physiology* **2006**, *68*, 619–647.

[4] J. Kalia, K. J. Swartz, *Scientific Reports* **2013**, DOI 10.1038/srep01523.

[5] E. Cao, M. Liao, Y. Cheng, D. Julius, *Nature* **2013**, *504*, 113–118.

[6] Y. Gao, E. Cao, D. Julius, Y. Cheng, *Nature* **2016**, *534*, 347–351.

[7] M. Grieben, A. C. W. Pike, C. Shintre, E. Venturi, S. El-ajouz, A. Tessitore, L. Shrestha, S. Mukhopadhyay, P. Mahajan, R. Chalk, N. A. Burgess-brown, R. Sitsapesan, J. T. Huiskonen, P Carpenter, S. G. Consortium, O. Particle, I. Centre, *Nature Structural & Molecular Biology* **2016**, *2*, 1–56.

[8] K. W. Huynh, M. R. Cohen, J. Jiang, A. Samanta, D. T. Lodowski, Z. H. Zhou, V. Y. Moiseenkova-Bell, *Nature Communications* **2016**, *7*, 11130.

[9] P. Jin, D. Bulkley, Y. Guo, W. Zhang, Z. Guo, W. Huynh, S. Wu, S. Meltzer, T. Cheng, L. Y. Jan, Y.-N. Jan, Y. Cheng, *Nature* **2017**, *547*, 118–122.

[10] M. Liao, E. Cao, D. Julius, Y. Cheng, *Nature* **2013**, *504*, 107–112.

[11] C. E. Paulsen, J.-P. Armache, Y. Gao, Y. Cheng, D. Julius, *Nature* **2015**, *520*, 511–517.

[12] K. Saotome, A. K. Singh, M. V. Yelshanskaya, A. I. Sobolevsky, *Nature* **2016**, *534*, 506–511.

[13] P. S. S. Shen, X. Yang, P. G. G. DeCaen, X. Liu, D. Bulkley, D. E. E. Clapham, E. Cao, *Cell* **2016**, *167*, 763–773.e11.

[14] Q. Su, F. Hu, X. Ge, J. Lei, S. Yu, T. Wang, Q. Zhou, C. Mei, Y. Shi, *Science* **2018**, *361*, DOI 10.1126/science.aat9819.

[15] M. Wilkes, M. G. Madej, L. Kreuter, D. Rhinow, V. Heinz, S. De Sanctis, S. Ruppel, R. M. Richter, F. Joos, M. Grieben, A. C. W. Pike, J. T. Huiskonen, E. P. Carpenter, W. Kühlbrandt, R. Witzgall, C. Ziegler, *Nature Structural & Molecular Biology* **2017**, *24*, 123–130.

[16] L. Zubcevic, M. A. Herzik, B. C. Chung, Z. Liu, G. C. Lander, S.-Y. Lee, *Nature Structural & Molecular Biology* **2016**, *23*, 180–186.

[17] M. Li, Y. Yu, J. Yang, *Advances in experimental medicine and biology* **2011**, *704*, 1–23.

References

[18] B. Nilius, G. Owsianik, T. Voets, J. A. Peters, *Physiological Reviews* **2007**, *87*, 165–217.

[19] I. Vetter, R. J. Lewis in Advances in Experimental Medicine and Biology, *Vol. 704*, **2011**, pp. 41–85.

[20] K. Venkatachalam, C. Montell, *Annual review of biochemistry* **2007**, *76*, 387–417.

[21] B. Nilius, G. Owsianik, *Pflugers Archiv European Journal of Physiology* **2010**, *460*, 437–450.

[22] D. J. Cosens, A. Manning, *Nature* **1969**, *224*, 285–287.

[23] S. F. Pedersen, G. Owsianik, B. Nilius, *Cell Calcium* **2005**, *38*, 233–252.

[24] A. Dietrich, M. Fahlbusch, T. Gudermann, *Cells* **2014**, *3*, 939–962.

[25] H. L. Ong, L. B. De Souza, I. S. Ambudkar in *Advances in Experimental Medicine and Biology*, *Vol. 898*, Springer New York LLC, **2016**, pp. 87–109.

[26] D. E. Clapham, *Nature* **2003**, *426*, 517–524.

[27] M. Trebak, G. S. J. Bird, R. R. McKay, J. W. Putney, *Journal of Biological Chemistry* **2002**, *277*, 21617–21623.

[28] B. Lintschinger, M. Balzer-Geldsetzer, T. Baskaran, W. F. Graier, C. Romanin, M. X. Zhu, K. Groschner, *Journal of Biological Chemistry* **2000**, *275*, 27799–27805.

[29] P. D. Wes, J. Chevesich, A. Jeromin, C. Montell, C. Rosenberg, G. Stetten, *Proceedings of the National Academy of Sciences of the United States of America* **1995**, *92*, 9652–9656.

[30] R. Maroto, A. Raso, T. G. Wood, A. Kurosky, B. Martinac, O. P. Hamill, *Nature Cell Biology* **2005**, *7*, 179–185.

[31] C. Strübing, G. Krapivinsky, L. Krapivinsky, D. E. Clapham, *Neuron* **2001**, *29*, 645–655.

[32] E. R. Liman, D. P. Corey, C. Dulac, E. R. Liman, D. P. Corey, *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 5791–5796.

[33] M. K. Jungnickel, H. Marrero, L. Birnbaumer, J. R. Lémos, H. M. Florman, *Nature Cell Biology* **2001**, *3*, 499–502.

[34] T. Hofmann, A. G. Obukhov, M. Schaefer, C. Harteneck, T. Gudermann, G. Schultz, *Nature* **1999**, *397*, 259–263.

[35] M. Schaefer, T. D. Plant, A. G. Obukhov, T. Hofmann, T. Gudermann, G. Schultz, *Journal of Biological Chemistry* **2000**, *275*, 17517–17526.

[36] Y. Tang, J. Tang, Z. Chen, C. Trost, V. Flockerzi, M. Li, V. Ramesh, M. X. Zhu, *Journal of Biological Chemistry* **2000**, *275*, 37559–37564.

[37] J. W. Voltz, E. J. Weinman, S. Shenolikar, Expanding the role of NHERF, a PDZ-domain containing protein adapter, to growth regulation, **2001**.

[38] B. Svobodova, K. Groschner, Mechanisms of lipid regulation and lipid gating in TRPC channels, **2016**.

[39] J. Shi, E. Mori, Y. Mori, M. Mori, J. Li, Y. Ito, R. Inoue, *Journal of Physiology* **2004**, *561*, 415–432.

[40] M. J. Caterina, M. A. Schumacher, M. Tominaga, T. A. Rosen, J. D. Levine, D. Julius, *Nature* **1997**, *389*, 816–824.

[41] M. J. Caterina, D. Julius, *Annual Review of Neuroscience* **2001**, *24*, 487–517.

[42] M. J. Caterina, A. Leffler, A. B. Malmberg, W. J. Martin, J. Trafton, K. R. Petersen-Zeitz, M. Koltzenburg, A. I. Basbaum, D. Julius, *Science* **2000**, *288*, 306–313.

[43] J. B. Davis, J. Gray, M. J. Gunthorpe, J. P. Hatcher, P. T. Davey, P. Overend, M. H. Harries, J. Latcham, C. Clapham, K. Atkinson, S. A. Hughes, K. Rance, E. Grau, A. J. Harper, P. L. Pugh, D. C. Rogers, S. Bingham, A. Randall, S. A. Sheardown, *Nature* **2000**, *405*, 183–187.

[44] M. J. Caterina, T. A. Rosen, M. Tominaga, A. J. Brake, D. Julius, *Nature* **1999**, *398*, 436–441.

[45] K. Muraki, Y. Iwata, Y. Katanosaka, T. Ito, S. Ohya, M. Shigekawa, Y. Imaizumi, *Circulation Research* **2003**, *93*, 829–838.

[46] G. D. Smith, M. J. Gunthorpe, R. E. Kelsell, P. D. Hayes, P. Reilly, P. Facer, J. E. Wright, J. C. Jerman, J. P. Walhin, L. Ooi, J. Egerton, K. J. Charles, D. Smart, A. D. Randall, P. Anand, J. B. Davis, *Nature* **2002**, *418*, 186–190.

[47] W. Everaerts, B. Nilius, G. Owsianik, The vanilloid transient receptor potential channel TRPV4: From structure to disease, **2010**.

[48] S. M. Huang, X. Li, Y. Yu, J. Wang, M. J. Caterina, *Molecular Pain* **2011**, *7*, DOI `10.1186/1744-8069-7-37`.

[49] A. D. Güler, H. Lee, T. Iida, I. Shimizu, M. Tominaga, M. Caterina, *Journal of Neuroscience* **2002**, *22*, 6408–6414.

[50] A. Moqrich, S. W. Hwang, T. J. Earley, M. J. Petrus, A. N. Murray, K. S. R. Spencer, M. Andahazy, G. M. Story, A. Patapoutian, *Science* **2005**, *307*, 1468–1472.

[51] J. G. Hoenderop, T. Voets, S. Hoefs, F. Weidema, J. Prenen, B. Nilius, R. J. Bindels, *EMBO Journal* **2003**, *22*, 776–785.

[52] E. den Dekker, J. G. Hoenderop, B. Nilius, R. J. Bindels, The epithelial calcium channels, TRPV5 & TRPV6: From identification towards regulation, **2003**.

[53] T. E. Hughes, R. A. Pumroy, A. T. Yazici, M. A. Kasimova, E. C. Fluck, K. W. Huynh, A. Samanta, S. K. Molugu, Z. H. Zhou, V. Carnevale, T. Rohacs, V. Y. Moiseenkova-Bell, *Nature Communications* **2018**, *9*, DOI `10.1038/s41467-018-06753-6`.

[54] R. Kraft, C. Harteneck, The mammalian melastatin-related transient receptor potential cation channels: An overview, **2005**.

[55] F. J. Kühn, A. Lückhoff, *Journal of Biological Chemistry* **2004**, *279*, 46431–46437.

[56] L. V. Ryazanova, M. V. Dorovkov, A. Ansari, A. G. Ryazanov, *Journal of Biological Chemistry* **2004**, *279*, 3708–3716.

References

[57] K. P. Schlingmann, S. Weber, M. Peters, L. N. Nejsum, H. Vitzthum, K. Klingel, M. Kratz, E. Haddad, E. Ristoff, D. Dinour, M. Syrrou, S. Nielsen, M. Sassen, S. Waldegger, H. W. Seyberth, M. Konrad, *Nature Genetics* **2002**, *31*, 166–170.

[58] L. M. Duncan, J. Deeds, J. Hunter, J. Shao, L. M. Holmgren, E. A. Woolf, R. I. Tepper, A. W. Shyjan, *Cancer Research* **1998**, *58*, 1515–1520.

[59] Y. Hara, M. Wakamori, M. Ishii, E. Maeno, M. Nishida, T. Yoshida, H. Yamada, S. Shimizu, E. Mori, J. Kudoh, N. Shimizu, H. Kurose, Y. Okada, K. Imoto, Y. Mori, *Molecular Cell* **2002**, *9*, 163–173.

[60] K. Held, T. Voets, J. Vriens, *Temperature* **2015**, *2*, 201–213.

[61] A. Zholos, C. Johnson, T. Burdyga, D. Melanaphy, *Advances in experimental medicine and biology* **2011**, *704*, 707–29.

[62] Y. Zhang, M. A. Hoon, J. Chandrashekar, K. L. Mueller, B. Cook, D. Wu, C. S. Zuker, N. J. Ryba, *Cell* **2003**, *112*, 293–301.

[63] T. Voets, B. Nilius, S. Hoefs, A. W. C. M. Van Der Kemp, G. Droogmans, R. J. M. Bindels, J. G. J. Hoenderop, *Journal of Biological Chemistry* **2004**, *279*, 19–25.

[64] M. J. Nadler, M. C. Hermosura, K. Inabe, A. L. Perraud, Q. Zhu, A. J. Stokes, T. Kurosaki, J. P. Kinet, R. Penner, A. M. Scharenberg, A. Fleig, *Nature* **2001**, *411*, 590–595.

[65] R. Y. Walder, D. Landau, P. Meyer, H. Shalev, M. Tsolia, Z. Borochowitz, M. B. Boettger, G. E. Beck, R. K. Englehardt, R. Carmi, V. C. Sheffield, *Nature Genetics* **2002**, *31*, 171–174.

[66] C. Schmitz, A. L. Perraud, C. O. Johnson, K. Inabe, M. K. Smith, R. Penner, T. Kurosaki, A. Fleig, A. M. Scharenberg, *Cell* **2003**, *114*, 191–200.

[67] D. D. McKemy, W. M. Neuhausser, D. Julius, *Nature* **2002**, *416*, 52–58.

[68] A. M. Peier, A. Moqrich, A. C. Hergarden, A. J. Reeve, D. A. Andersson, G. M. Story, T. J. Earley, I. Dragoni, P. McIntyre, S. Bevan, A. Patapoutian, *Cell* **2002**, *108*, 705–715.

[69] M. Bandell, A. E. Dubin, M. J. Petrus, A. Orth, J. Mathur, W. H. Sun, A. Patapoutian, *Nature Neuroscience* **2006**, *9*, 493–500.

[70] T. Rohács, C. M. Lopes, I. Michailidis, D. E. Logothetis, *Nature Neuroscience* **2005**, *8*, 626–634.

[71] S. E. Jordt, D. M. Bautista, H. H. Chuang, D. D. McKemy, P. M. Zygmunt, E. D. Högestätt, I. D. Meng, D. Julius, *Nature* **2004**, *427*, 260–265.

[72] L. J. Macpherson, B. H. Geierstanger, V. Viswanath, M. Bandell, S. R. Eid, S. W. Hwang, A. Patapoutian, *Current Biology* **2005**, *15*, 929–934.

[73] D. M. Bautista, S. E. Jordt, T. Nikai, P. R. Tsuruda, A. J. Read, J. Poblete, E. N. Yamoah, A. I. Basbaum, D. Julius, *Cell* **2006**, *124*, 1269–1282.

[74] M. Bandell, G. M. Story, S. W. Hwang, V. Viswanath, S. R. Eid, M. J. Petrus, T. J. Earley, A. Patapoutian, *Neuron* **2004**, *41*, 849–857.

[75] N. Takahashi, T. Kuwaki, S. Kiyonaka, T. Numata, D. Kozai, Y. Mizuno, S. Yamamoto, S. Naito, E. Knevels, P. Carmeliet, T. Oga, S. Kaneko, S. Suga, T. Nokami, J. I. Yoshida, Y. Mori, *Nature Chemical Biology* **2011**, *7*, 701–711.

[76] K. Venkatachalam, C. O. Wong, M. X. Zhu, The role of TRPMLs in endolysosomal trafficking and function, **2015**.

[77] M. Hirschi, M. A. Herzik, J. Wie, Y. Suo, W. F. Borschel, D. Ren, G. C. Lander, S. Y. Lee, *Nature* **2017**, *550*, 411–414.

[78] M. Fine, P. Schmiege, X. Li, *Nature Communications* **2018**, DOI 10.1038/s41467-018-06493-7.

[79] Q. Chen, J. She, W. Zeng, J. Guo, H. Xu, X. C. Bai, Y. Jiang, *Nature* **2017**, DOI 10.1038/nature24035.

[80] M. Li, W. K. Zhang, N. M. Benvin, X. Zhou, D. Su, H. Li, S. Wang, I. E. Michailidis, L. Tong, X. Li, J. Yang, *Nature Structural and Molecular Biology* **2017**, *24*, 205–213.

[81] W. Wang, X. Zhang, Q. Gao, H. Xu, *Handbook of Experimental Pharmacology* **2014**, *222*, 631–645.

[82] M. P. Cuajungco, J. Silva, A. Habibi, J. A. Valadez, The mucolipin-2 (TRPML2) ion channel: a tissue-specific protein crucial to normal cell function, M.P. Cuajungco, Department of Biological Science, California State University Fullerton, Fullerton, United States, **2016**.

[83] H. J. Kim, S. Yamaguchi, Q. Li, I. So, S. Muallem, *Journal of Biological Chemistry* **2010**, *285*, 16513–16520.

[84] F. Qian, K. Noben-Trauth, *Pflugers Archiv European Journal of Physiology* **2005**, *451*, 277–285.

[85] C. P. Palmer, E. Aydar, M. B. A. Djamgoz, *Biochemical Journal* **2005**, *387*, 211–219.

[86] V. Chauvet, F. Qian, N. Boute, Y. Cai, B. Phakdeekitacharoen, L. F. Onuchic, T. Attié-Bitach, L. Guicharnaud, O. Devuyst, G. G. Germino, M. C. Gubler, *American Journal of Pathology* **2002**, *160*, 973–983.

[87] Y. Cai, Y. Maeda, A. Cedzich, V. E. Torres, G. Wu, T. Hayashi, T. Mochizuki, J. H. Park, R. Witzgall, S. Somlo, *Journal of Biological Chemistry* **1999**, *274*, 28557–28565.

[88] P. Koulen, Y. Cai, L. Geng, Y. Maeda, S. Nishimura, R. Witzgall, B. E. Ehrlich, S. Somlo, *Nature Cell Biology* **2002**, *4*, 191–197.

[89] M. S. Scheffers, *Human Molecular Genetics* **2002**, *11*, 59–67.

[90] Y. Luo, P. M. Vassilev, X. Li, Y. Kawanabe, J. Zhou, *Molecular and Cellular Biology* **2003**, *23*, 2600–2607.

[91] L. Foggensteiner, A. P. Bevan, R. Thomas, N. Coleman, C. Boulter, J. Bradley, O. Ibraghimov-Beskrovnaya, K. Klinger, R. Sandford, *Journal of the American Society of Nephrology* **2000**, *11*, 814–827.

[92] D. R. Rundle, G. Gorbsky, L. Tsiokas, *Journal of Biological Chemistry* **2004**, *279*, 29728–29739.

[93] K Hanaoka, F Qian, a Boletta, a. K. Bhunia, K Piontek, L Tsiokas, V. P. Sukhatme, W. B. Guggino, G. G. Germino, *Nature* **2000**, *408*, 990–994.

[94] M. Köttgen, T. Benzing, T. Simmen, R. Tauber, B. Buchholz, S. Feliciangeli, T. B. Huber, B. Schermer, A. Kramer-Zucker, K. Höpker, K. C. Simmen, C. C. Tschucke, R. Sandford, E. Kim, G. Thomas, G. Walz, *EMBO Journal* **2005**, *24*, 705–716.

[95] A. Ćelić, E. T. Petri, B. Demeler, B. E. Ehrlich, T. J. Boggon, *Journal of Biological Chemistry* **2008**, *283*, 28305–28312.

[96] A. S. Ćelić, E. T. Petri, J. Benbow, M. E. Hodsdon, B. E. Ehrlich, T. J. Boggon, *Journal of Biological Chemistry* **2012**, *287*, 17232–17240.

[97] Y. Yu, M. H. Ulbrich, M.-H. Li, Z. Buraei, X.-Z. Chen, A. C. M. Ong, L. Tong, E. Y. Isacoff, J. Yang, *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 11558–11563.

[98] E. T. Petri, A. Celic, S. D. Kennedy, B. E. Ehrlich, T. J. Boggon, M. E. Hodsdon, *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 9176–81.

[99] A. Giamarchi, S. Feng, L. Rodat-Despoix, Y. Xu, E. Bubenshchikova, L. J. Newby, J. Hao, C. Gaudioso, M. Crest, A. N. Lupas, E. Honoré, M. P. Williamson, T. Obara, A. C. Ong, P. Delmas, *EMBO Journal* **2010**, *29*, 1176–1191.

[100] M. D. Allen, S. Qamar, M. K. Vadivelu, R. N. Sandford, M. Bycroft, *Protein Science* **2014**, *23*, 1301–1308.

[101] J. Zhu, Y. Yu, M. H. Ulbrich, M. H. Li, E. Y. Isacoff, B. Honig, J. Yang, *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 10133–10138.

[102] F. Qian, F. J. Germino, Y. Cai, X. Zhang, S. Somlo, G. G. Germino, *Nature Genetics* **1997**, *16*, 179–183.

[103] L. Tsiokas, E. Kim, T. Arnould, V. P. Sukhatme, G. Walz, *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 6965–6970.

[104] V. Babich, W. Z. Zeng, B. I. Yeh, O. Ibraghimov-Beskrovnaya, Y. Cai, S. Somlo, C. L. Huang, *Journal of Biological Chemistry* **2004**, *279*, 25582–25589.

[105] S. Feng, G. M. Okenka, C. X. Bai, A. J. Streets, L. J. Newby, B. T. DeChant, L. Tsiokas, T. Obara, A. C. Ong, *Journal of Biological Chemistry* **2008**, *283*, 28471–28479.

[106] Y. Cai, S. V. Fedeles, K. Dong, G. Anyatonwu, T. Onoe, M. Mitobe, J. D. Gao, D. Okuhara, X. Tian, A. R. Gallagher, Z. Tang, X. Xie, M. D. Lalioti, A. H. Lee, B. E. Ehrlich, S. Somlo, *Journal of Clinical Investigation* **2014**, *124*, 5129–5144.

[107] A. J. Streets, D. J. Moon, M. E. Kane, T. Obara, A. C. Ong, *Human Molecular Genetics* **2006**, *15*, 1465–1473.

[108] A. Patel, E. Honoré, *Nature Reviews Nephrology* **2010**, *6*, 530–538.

[109] S. Kim, H. Nie, V. Nesin, U. Tran, P. Outeda, C. X. Bai, J. Keeling, D. Maskey, T. Watnick, O. Wessely, L. Tsiokas, *Nature Cell Biology* **2016**, *18*, 752–764.

[110] S. M. Nauli, F. J. Alenghat, Y. Luo, E. Williams, P. Vassilev, X. Li, A. E. H. Elia, W. Lu, E. M. Brown, S. J. Quinn, D. E. Ingber, J. Zhou, *Nature genetics* **2003**, *33*, 129–37.

[111] V. Chauvet, X. Tian, H. Husson, D. H. Grimm, T. Wang, T. Hieseberger, P. Igarashi, A. M. Bennett, O. Ibraghimov-Beskrovnaya, S. Somlo, M. J. Caplan, *Journal of Clinical Investigation* **2004**, *114*, 1433–1443.

[112] P. Delmas, H. Nomura, X. Li, M. Lakkis, Y. Luo, Y. Segal, J. M. Fernández-Fernández, P. Harris, A. M. Frischauf, D. A. Brown, J. Zhou, *Journal of Biological Chemistry* **2002**, *277*, 11276–11283.

[113] S. C. Parnell, B. S. Magenheimer, R. L. Maser, C. A. Rankin, A. Smine, T. Okamoto, J. P. Calvet, *Biochemical and Biophysical Research Communications* **1998**, *251*, 625–631.

[114] Y. Butscheid, V. Chubanov, K. Steger, D. Meyer, A. Dietrich, T. Gudermann, *Molecular Reproduction and Development* **2006**, *73*, 350–360.

[115] K. A. Sutton, M. K. Jungnickel, H. M. Florman, *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 8661–8666.

[116] T. Yuasa, B. Venugopal, S. Weremowicz, C. C. Morton, L. Guo, J. Zhou, *Genomics* **2002**, *79*, 376–386.

[117] P. Vogel, R. Read, G. M. Hansen, L. C. Freay, B. P. Zambrowicz, A. T. Sands, *Veterinary Pathology* **2010**, *47*, 120–131.

[118] C. Karcher, A. Fischer, A. Schweickert, E. Bitzer, S. Horie, R. Witzgall, M. Blum, *Differentiation* **2005**, *73*, 425–432.

[119] T. Yuasa, A. Takakura, B. M. Denker, B. Venugopal, J. Zhou, *Genomics* **2004**, *84*, 126–138.

[120] Z. Salehi-Najafabadi, B. Li, V. Valentino, C. Ng, H. Martin, Y. Yu, Z. Wang, P. Kashyap, Y. Yu, *Journal of Biological Chemistry* **2017**, *292*, 4210–4221.

[121] Y. Ishimaru, H. Inada, M. Kubota, H. Zhuang, M. Tominaga, H. Matsunami, *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 12569–12574.

[122] A. L. Huang, X. Chen, M. A. Hoon, J. Chandrashekar, W. Guo, D. Tränkner, N. J. Ryba, C. S. Zuker, *Nature* **2006**, *442*, 934–938.

[123] N. D. Lopezjimenez, M. M. Cavenagh, E. Sainz, M. A. Cruz-Ithier, J. F. Battey, S. L. Sullivan, *Journal of Neurochemistry* **2006**, *98*, 68–77.

[124] P. D. Wilson, *New England Journal of Medicine* **2004**, *350*, 151–164.

[125] P. Igarashi, S. Somlo, *Journal of the American Society of Nephrology* **2002**, *13*, 2384–2398.

[126] Y. Pirson, *Advances in Chronic Kidney Disease* **2010**, *17*, 173–180.

[127] B. M. Paul, M. B. Consugar, M. Ryan Lee, J. L. Sundsbak, C. M. Heyer, S. Rossetti, V. J. Kubly, K. Hopp, V. E. Torres, E. Coto, M. Clementi, N. Bogdanova, E. De Almeida, D. G. Bichet, P. C. Harris, *Kidney International* **2014**, *85*, 383–392.

[128] A. C. Ong, P. C. Harris, *Kidney International* **2015**, *88*, 699–710.

[129] A. D. Paterson, K. R. Wang, D. Lupea, P. St. George-Hyslop, Y. Pei, *American Journal of Kidney Diseases* **2002**, *40*, 16–20.

[130] G. Wu, *Human Molecular Genetics* **2002**, *11*, 1845–1854.

[131] E. Waldrop, M. A. Al-Obaide, T. L. Vasylyeva, *Frontiers in Genetics* **2019**, *10*, DOI 10.3389/fgene.2019.00044.

[132] T. J. Watnick, V. E. Torres, M. A. Gandolph, F. Qian, L. F. Onuchic, K. W. Klinger, G. Landes, G. G. Germino, *Molecular Cell* **1998**, *2*, 247–251.

[133] Y. Pei, T. Watnick, N. He, K. Wang, Y. Liang, P. Parfrey, G. Germino, P. St. George-Hyslop, *Journal of the American Society of Nephrology* **1999**, *10*, 1524–1529.

[134] S. Rossetti, V. J. Kubly, M. B. Consugar, K. Hopp, S. Roy, S. W. Horsley, D. Chauveau, L. Rees, T. M. Barratt, W. G. Van't Hoff, W. P. Niaudet, V. E. Torres, P. C. Harris, *Kidney International* **2009**, *75*, 848–855.

[135] K. Hopp, C. J. Ward, C. J. Hommerding, S. H. Nasr, H. F. Tuan, V. G. Gainullin, S. Rossetti, V. E. Torres, P. C. Harris, *Journal of Clinical Investigation* **2012**, *122*, 4257–4273.

[136] A. Li, X. Tian, X. Zhang, S. Huang, Y. Ma, D. Wu, G. Moeckel, S. Somlo, G. Wu, *American Journal of Pathology* **2015**, *185*, 2843–2860.

[137] S. Burtey, M. Riera, E. Ribe, P. Pennekamp, E. Passage, R. Rance, B. Dworniczak, M. Fontés, *Nephrology Dialysis Transplantation* **2008**, *23*, 1157–1165.

[138] C. Thivierge, A. Kurbegovic, M. Couillard, R. Guillaume, O. Cote, M. Trudel, *Molecular and Cellular Biology* **2006**, *26*, 1538–1548.

[139] D. N. Wheatley, *Pathobiology* **1995**, *63*, 222–238.

[140] P. M. Jenkins, D. P. Mcewen, J. R. Martens, *Chemical Senses* **2009**, *34*, 451–464.

[141] H. A. Praetorius, K. R. Spring, *Journal of Membrane Biology* **2003**, *191*, 69–76.

[142] H. A. Praetorius, K. R. Spring, *Journal of Membrane Biology* **2001**, *184*, 71–79.

[143] J. M. Gerdes, E. E. Davis, N. Katsanis, The Vertebrate Primary Cilium in Development, Homeostasis, and Disease, **2009**.

[144] F. Hildebrandt, M. Attanasio, E. Otto, *Journal of the American Society of Nephrology* **2009**, *20*, 23–35.

[145] P. Satir, L. B. Pedersen, S. T. Christensen, *Journal of Cell Science* **2010**, *123*, 499–503.

[146] X. Liu, T. Vien, J. Duan, S. H. Sheu, P. G. DeCaen, D. E. Clapham, *eLife* **2018**, DOI 10.7554/eLife.33183.

[147] T. Malekshahabi, N. Khoshdel Rad, A. L. Serra, R. Moghadasali, *Journal of Cellular Physiology* **2019**, *234*, 12451–12470.

[148] M Delling, A. A. Indzhykulian, X Liu, T Xie, D. E. Clapham, T. Ca, C.-s. Ca, E. D. Figs, E. D. Fig, E. D. Fig, E. D. Fig, S. Video, S. Information, S. Video, E. D. Fig, *Nature* **2016**, *531*, 656–660.

[149] K. Piontek, L. F. Menezes, M. A. Garcia-Gonzalez, D. L. Huso, G. G. Germino, *Nature Medicine* **2007**, *13*, 1490–1495.

[150] M. C. Hogan, L. Manganelli, J. R. Woollard, A. I. Masyuk, T. V. Masyuk, R. Tammachote, B. Q. Huang, A. A. Leontovich, T. G. Beito, B. J. Madden, M. C. Charlesworth, V. E. Torres, N. F. Larusso, P. C. Harris, C. J. Ward, *Journal of the American Society of Nephrology* **2009**, *20*, 278–288.

[151] O. H. Samuli Ollila, T. Róg, M. Karttunen, I. Vattulainen, *Journal of Structural Biology* **2007**, *159*, 311–323.

[152] R. Phillips, T. Ursell, P. Wiggins, P. Sens, *Nature* **2009**, *459*, 379–385.

[153] H. Luecke, B. Schobert, H. T. Richter, J. P. Cartailler, J. K. Lanyi, *Journal of Molecular Biology* **1999**, *291*, 899–911.

[154] S. B. Long, X. Tao, E. B. Campbell, R. MacKinnon, *Nature* **2007**, *450*, 376–382.

[155] S. B. Hansen, X. Tao, R. MacKinnon, *Nature* **2011**, *477*, 495–498.

[156] M. R. Whorton, R. MacKinnon, *Cell* **2011**, *147*, 199–208.

[157] G. Wisedchaisri, S. L. Reichow, T. Gonen, *Structure* **2011**, *19*, 1381–1393.

[158] S. Raunser, T. Walz, *Annual Review of Biophysics* **2009**, *38*, 89–105.

[159] R. K. Hite, Z. Li, T. Walz, *EMBO Journal* **2010**, *29*, 1652–1658.

[160] Y. Yin, S. C. Le, A. L. Hsu, M. J. Borgnia, H. Yang, S.-Y. Lee, *Science* **2019**, *363*, pii: eaav9334.

[161] R. M. Walsh, S. H. Roh, A. Gharpure, C. L. Morales-Perez, J. Teng, R. E. Hibbs, *Nature* **2018**, *557*, 261–265.

[162] D. Laverty, R. Desai, T. Uchański, S. Masiulis, W. J. Stec, T. Malinauskas, J. Zivanov, E. Pardon, J. Steyaert, K. W. Miller, A. R. Aricescu, *Nature* **2019**, *565*, 516–520.

[163] S. M. Jackson, I. Manolaridis, J. Kowal, M. Zechner, N. M. Taylor, M. Bause, S. Bauer, R. Bartholomaeus, G. Bernhardt, B. Koenig, A. Buschauer, H. Stahlberg, K. H. Altmann, K. P. Locher, *Nature Structural and Molecular Biology* **2018**, *25*, 333–340.

[164] M. Zhou, N. Morgner, N. P. Barrera, A. Politis, S. C. Isaacson, D. Matak-Vinković, T. Murata, R. A. Bernal, D. Stock, C. V. Robinson, *Science* **2011**, *334*, 380–385.

[165] H. Y. Yen, K. K. Hoi, I. Liko, G. Hedger, M. R. Horrell, W. Song, D. Wu, P. Heine, T. Warne, Y. Lee, B. Carpenter, A. Plückthun, C. G. Tate, M. S. Sansom, C. V. Robinson, *Nature* **2018**, *559*, 424–427.

[166] X. Cong, Y. Liu, W. Liu, X. Liang, D. H. Russell, A. Laganowsky, *Journal of the American Chemical Society* **2016**, *138*, 4346–4349.

[167] D. Huster, *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* **2014**, *1841*, 1146–1160.

[168] G. Güler, R. M. Gärtner, C. Ziegler, W. Mäntele, *Journal of Biological Chemistry* **2016**, *291*, 4295–4307.

References

[169] F. X. Contreras, A. M. Ernst, P. Haberkant, P. Björkholm, E. Lindahl, B. Gönen, C. Tischer, A. Elofsson, G. Von Heijne, C. Thiele, R. Pepperkok, F. Wieland, B. Brügger, *Nature* **2012**, *481*, 525–529.

[170] V. Betaneli, E. P. Petrov, P. Schwille, *Biophysical Journal* **2012**, *102*, 523–531.

[171] A. L. Duncan, W. Song, M. S. Sansom, *Annual Review of Pharmacology and Toxicology* **2020**, *60*, DOI 10.1146/annurev-pharmtox-010919-023411.

[172] G. Khelashvili, A. Grossfield, S. E. Feller, M. C. Pitman, H. Weinstein, *Proteins: Structure Function and Bioinformatics* **2009**, *76*, 403–417.

[173] E. Deplazes, M. Louhivuori, D. Jayatilaka, S. J. Marrink, B. Corry, *PLoS Computational Biology* **2012**, *8*, DOI 10.1371/journal.pcbi.1002683.

[174] C. Arnarez, J. P. Mazat, J. Elezgaray, S. J. Marrink, X. Periole, *Journal of the American Chemical Society* **2013**, *135*, 3112–3120.

[175] T. Reddy, M. S. Sansom, *Structure* **2016**, *24*, 375–382.

[176] P. J. Stansfeld, R. Hopkinson, F. M. Ashcroft, M. S. P. Sansom, *Biochemistry* **2009**, *48*, 10926–10933.

[177] R. M. Glaeser, *Nat Methods* **2016**, *13*, 28–32.

[178] D. P. Frueh, A. C. Goodrich, S. H. Mishra, S. R. Nichols, *Current Opinion in Structural Biology* **2013**, *23*, 734–739.

[179] K. Murata, M. Wolf, *Biochimica et Biophysica Acta - General Subjects* **2018**, *1862*, 324–334.

[180] R. F. Egerton in *Physical Principles of Electron Microscopy*, Springer, Boston, MA, **2005**, Chapter Electron O, pp. 27–55.

[181] L. A. Baker, J. L. Rubinstein, *Methods in Enzymology* **2010**, *481*, 371–388.

[182] Y. Cheng, N. Grigorieff, P. A. Penczek, T. Walz, *Cell* **2015**, *161*, 438–449.

[183] M. Ohi, Y. Li, Y. Cheng, T. Walz, *Biological Procedures Online* **2004**, *6*, 23–34.

[184] U. B. Ericsson, B. M. Hallberg, G. T. DeTitta, N. Dekker, P. Nordlund, *Analytical Biochemistry* **2006**, *357*, 289–298.

[185] B. Kastner, N. Fischer, M. M. Golas, B. Sander, P. Dube, D. Boehringer, K. Hartmuth, J. Deckert, F. Hauer, E. Wolf, H. Uchtenhagen, H. Urlaub, F. Herzog, J. M. Peters, D. Poerschke, R. Lührmann, H. Stark, *Nature Methods* **2008**, *5*, 53–55.

[186] J. Frank, *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*, Oxford University Press, **2010**.

[187] R. S. Ruskin, Z. Yu, N. Grigorieff, *Journal of Structural Biology* **2013**, *184*, 385–393.

[188] S. Wu, J. P. Armache, Y. Cheng, Single-particle cryo-EM data acquisition by using direct electron detection camera, **2016**.

[189] J. Zivanov, T. Nakane, B. O. Forsberg, D. Kimanius, W. J. Hagen, E. Lindahl, S. H. Scheres, *eLife* **2018**, DOI 10.7554/eLife.42166.

[190] M. Radermacher, T. Wagenknecht, A. Verschoor, J. Frank, *Journal of Microscopy* **1987**, *146*, 113–136.

[191] D. Elmlund, H. Elmlund, *Journal of Structural Biology* **2012**, *180*, 420–427.

[192] R. Henderson, A. Sali, M. L. Baker, B. Carragher, B. Devkota, K. H. Downing, E. H. Egelman, Z. Feng, J. Frank, N. Grigorieff, W. Jiang, S. J. Ludtke, O. Medalia, P. A. Penczek, P. B. Rosenthal, M. G. Rossmann, M. F. Schmid, G. F. Schröder, A. C. Steven, D. L. Stokes, J. D. Westbrook, W. Wriggers, H. Yang, J. Young, H. M. Berman, W. Chiu, G. J. Kleywegt, C. L. Lawson, *Structure* **2012**, *20*, 205–214.

[193] D. Lyumkis, *Journal of Biological Chemistry* **2019**, *294*, 5181–5197.

[194] V. E. Lamberti, L. D. Fosdick, E. R. Jessup, C. J. C. Schauble, *Journal of Chemical Education* **2009**, *79*, 601–606.

[195] A. R. Leach, *Molecular modelling : principles and applications*, Second Edi, **2001**.

[196] O. Guvench, A. D. MacKerell in *Methods in Molecular Biology*, Humana Press, **2008**, pp. 63–88.

[197] L. Monticelli, D. P. Tieleman, *Methods in Molecular Biology* **2013**, *924*, 197–213.

[198] M. A. González, *Collection SFN* **2011**, *12*, 169–200.

[199] R. W. Hockney, S. P. Goel, J. W. Eastwood, *Journal of Computational Physics* **1974**, *14*, 148–158.

[200] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *Journal of Computational Chemistry* **1997**, *18*, 1463–1472.

[201] J. P. Ryckaert, G. Ciccotti, H. J. Berendsen, *Journal of Computational Physics* **1977**, *23*, 327–341.

[202] J. Hermans, H. J. C. Berendsen, W. F. Van Gunsteren, J. P. M. Postma, *Biopolymers* **1984**, *23*, 1513–1518.

[203] G. Bussi, D. Donadio, M. Parrinello, *The Journal of Chemical Physics* **2007**, *126*, 014101.

[204] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *The Journal of Chemical Physics* **1984**, *81*, 3684–3690.

[205] M. Parrinello, A. Rahman, *Journal of Applied Physics* **1981**, *52*, 7182–7190.

[206] B. Erman, *Biophysical Journal* **2001**, *81*, 3534–3544.

[207] F. Ding, S. V. Buldyrev, N. V. Dokholyan, *Biophysical Journal* **2005**, *88*, 147–155.

[208] E. J. Sambriski, D. C. Schwartz, J. J. De Pablo, *Biophysical Journal* **2009**, *96*, 1675–1690.

[209] J. W. Chu, G. A. Voth, *Biophysical Journal* **2007**, *93*, 3860–3871.

[210] S. J. Marrink, A. H. De Vries, A. E. Mark, *Journal of Physical Chemistry B* **2004**, *108*, 750–760.

[211] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, S.-J. J. Marrink, *Journal of Chemical Theory and Computation* **2008**, *4*, 819–834.

[212] J. J. Uusitalo, H. I. Ingólfsson, P. Akhshi, D. P. Tieleman, J Siewert, *Journal of Chemical Theory and Computation* **2015**, *11*, 3932–3945.

[213] J. J. Uusitalo, H. I. Ingólfsson, S. J. Marrink, I. Faustino, *Biophysical Journal* **2017**, *113*, 246–256.

[214] C. A. López, A. J. Rzepiela, A. H. de Vries, L. Dijkhuizen, P. H. Hünenberger, S. J. Marrink, *Journal of Chemical Theory and Computation* **2009**, *5*, 3195–3210.

[215] H. Lee, A. H. De Vries, S. J. Marrink, R. W. Pastor, *Journal of Physical Chemistry B* **2009**, *113*, 13186–13194.

[216] W. Kabsch, C. Sander, *Biopolymers* **1983**, *22*, 2577–2637.

[217] X. Periole, M. Cavalli, S.-J. J. Marrink, M. A. Ceruso, *Journal of Chemical Theory and Computation* **2009**, *5*, 2531–2543.

[218] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, *Journal of Computational Chemistry* **1992**, *13*, 1011–1021.

[219] M. G. Ciardo, A. Ferrer-Montiel, *Biochimica et Biophysica Acta - Biomembranes* **2017**, *1859*, 1615–1628.

[220] S. Brauchi, G. Orta, C. Mascayano, M. Salazar, N. Raddatz, H. Urbina, E. Rosenmann, F. Gonzalez-Nilo, R. Latorre, *Proceedings of the National Academy of Sciences* **2007**, *104*, 10246–10251.

[221] T. Rohacs, *Pflugers Archiv European Journal of Physiology* **2007**, *453*, 753–762.

[222] X. Steinberg, C. Lespay-Rebolledo, S. Brauchi, *Frontiers in Physiology* **2014**, *5*, 1–14.

[223] R. L. Daniels, Y. Takashima, D. D. McKemy, *Journal of Biological Chemistry* **2009**, *284*, 1570–1582.

[224] D. Liu, E. R. Liman, *Proceedings of the National Academy of Sciences* **2003**, *100*, 15160–15165.

[225] L. W. Runnels, L. Yue, D. E. Clapham, *Nature Cell Biology* **2002**, 329–336.

[226] B. I. Tóth, M. Konrad, D. Ghosh, F. Mohr, C. R. Halaszovich, M. G. Leitner, J. Vriens, J. Oberwinkler, T. Voets, *The Journal of general physiology* **2015**, *146*, 51–63.

[227] Z. Zhang, H. Okawa, Y. Wang, E. R. Liman, *Journal of Biological Chemistry* **2005**, *280*, 39185–39192.

[228] R. M. Klein, C. A. Ufret-Vincenty, L. Hua, S. E. Gordon, *Journal of Biological Chemistry* **2008**, *283*, 26208–26216.

[229] J. Lee, S.-K. Cha, T.-J. Sun, C.-L. Huang, *The Journal of general physiology* **2005**, *126*, 439–451.

[230] B. Thyagarajan, V. Lukacs, T. Rohacs, *Journal of Biological Chemistry* **2008**, *283*, 14980–14987.

[231] F. Nakatsu, *Developmental Cell* **2015**, *34*, 379–380.

[232] C. Fan, W. Choi, W. Sun, J. Du, W. Lu, *eLife* **2018**, DOI 10.7554/eLife.36852.

[233] Q. Tang, W. Guo, L. Zheng, J. X. Wu, M. Liu, X. Zhou, X. Zhang, L. Chen, *Cell Research* **2018**, *28*, 746–755.

[234] D. Vinayagam, T. Mager, A. Apelbaum, A. Bothe, F. Merino, O. Hofnagel, C. Gatsogiannis, S. Raunser, *eLife* **2018**, DOI 10.7554/eLife.36615.001.

[235] J. Duan, Z. Li, J. Li, R. E. Hulse, A. Santa-Cruz, W. C. Valinsky, S. A. Abiria, G. Krapivinsky, J. Zhang, D. E. Clapham, *Proceedings of the National Academy of Sciences* **2018**, *115*, E8201–E8210.

[236] L. Wang, T. M. Fu, Y. Zhou, S. Xia, A. Greka, H. Wu, *Science* **2018**, *362*, DOI 10.1126/science.aav4809.

[237] A. K. Singh, L. L. McGoldrick, A. I. Sobolevsky, *Nature Structural and Molecular Biology* **2018**, *25*, 805–813.

[238] L. Zubcevic, A. L. Hsu, M. J. Borgnia, S.-Y. Lee, *eLife* **2019**, DOI 10.7554/elife.45779.

[239] W. Zheng, X. Yang, R. Hu, R. Cai, L. Hofmann, Z. Wang, Q. Hu, X. Liu, D. Bulkey, Y. Yu, J. Tang, V. Flockerzi, Y. Cao, E. Cao, X. Z. Chen, *Nature Communications* **2018**, *9*, 2302.

[240] S. Q. Zheng, E. Palovcak, J. P. Armache, K. A. Verba, Y. Cheng, D. A. Agard, *Nature Methods* **2017**, *14*, 331–332.

[241] A. Rohou, N. Grigorieff, *Journal of Structural Biology* **2015**, *192*, 216–221.

[242] A. Kucukelbir, F. J. Sigworth, H. D. Tagare, *Nature Methods* **2014**, *11*, 63–65.

[243] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *Journal of computational chemistry* **2004**, *23*, 1605–1612.

[244] P. Emsley, K. Cowtan, *Acta Crystallographica Section D: Biological Crystallography* **2004**, *60*, 2126–2132.

[245] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *Acta Crystallographica Section D: Biological Crystallography* **2010**, 213–221.

[246] C. M. Shirey, J. L. Scott, R. V. Stahelin, *Analytical Biochemistry* **2017**, *516*, 9–12.

[247] G. Seebohm, E. Wrobel, M. Pusch, M. Dicks, J. Terhag, V. Matschke, I. Rothenberg, O. N. Ursu, F. Hertel, L. Pott, F. Lang, E. Schulze-Bahr, M. Hollmann, R. Stoll, N. Strutz-Seebohm, *Pflugers Archiv European Journal of Physiology* **2014**, *466*, 1885–1897.

[248] E. Y. J. Park, J. Y. Baik, M. Kwak, I. So, *The Korean Journal of Physiology & Pharmacology* **2019**, *23*, 219–227.

[249] G. Gimpl, *Chemistry and Physics of Lipids* **2016**, *199*, 61–73.

[250] G. Hedger, H. Koldsø, M. Chavent, C. Siebold, R. Rohatgi, M. S. Sansom, *Structure* **2019**, *27*, 549–559.

[251] W. Zheng, R. Cai, L. Hofmann, V. Nesin, Q. Hu, W. Long, M. Fatehi, X. Liu, S. Hussein, T. Kong, J. Li, P. E. Light, J. Tang, V. Flockerzi, L. Tsiokas, X. Z. Chen, *Cell Reports* **2018**, *22*, 1560–1573.

[252] G. Garcia, D. R. Raleigh, J. F. Reiter, How the Ciliary Membrane Is Organized Inside-Out to Communicate Outside-In, **2018**.

[253] Y. Rbaibi, S. Cui, D. Mo, M. Carattino, R. Rohatgi, L. M. Satlin, C. M. Szalinski, L. M. Swanhart, H. Fölsch, N. A. Hukriede, O. A. Weisz, *Traffic* **2012**, *13*, 1295–1305.

[254] V. Corradi, E. Mendez-Villuendas, H. I. Ingólfsson, R. X. Gu, I. Siuda, M. N. Melo, A. Moussatova, L. J. Degagné, B. I. Sejdiu, G. Singh, T. A. Wassenaar, K. Delgado Magnero, S. J. Marrink, D. P. Tieleman, *ACS Central Science* **2018**, *4*, 709–717.

[255] G. Hedger, M. S. P. Sansom, *Biochimica et Biophysica Acta - Biomembranes* **2016**, *1858*, 2390–2400.

[256] M. R. Schmidt, P. J. Stansfeld, S. J. Tucker, M. S. P. Sansom, *Biochemistry* **2013**, *52*, 279–281.

[257] G. Hedger, D. Shorthouse, H. Koldsø, M. S. P. Sansom, *Journal of Physical Chemistry B* **2016**, *120*, 8154–8163.

[258] P. J. Stansfeld, E. E. Jefferys, M. S. P. Sansom, *Structure* **2013**, *21*, 810–819.

[259] J. Domański, G. Hedger, R. B. Best, P. J. Stansfeld, M. S. P. Sansom, *The Journal of Physical Chemistry B* **2017**, acs.jpcb.6b08445.

[260] G. Hedger, S. L. Rouse, J. Domański, M. Chavent, H. Koldsø, M. S. P. Sansom, *Biochemistry* **2016**, *55*, 6238–6249.

[261] W. Song, H. Y. Yen, C. V. Robinson, M. S. Sansom, *Structure* **2019**, DOI `10.1016/j.str.2018.10.024`.

[262] B. Webb, A. Sali, *Current protocols in bioinformatics / editoral board Andreas D. Baxevanis ... [et al.]* **2014**, *47*, 5.6.1–5.6.32.

[263] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, *Bioinformatics* **2013**, *29*, 845–854.

[264] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindah, *SoftwareX* **2015**, DOI `10.1016/j.softx.2015.06.001`.

[265] D. H. de Jong, G. Singh, W. F. D. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schäfer, X. Periole, D. P. Tieleman, S. J. Marrink, *Journal of Chemical Theory and Computation* **2013**, *9*, 687–697.

[266] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, A. H. De Vries, *Journal of Physical Chemistry B* **2007**, *111*, 7812–7824.

[267] D. Shorthouse, G. Hedger, H. Koldsø, M. S. Sansom, Molecular simulations of glycolipids: Towards mammalian cell membrane models, **2016**.

[268] K. A. Scott, P. J. Bond, A. Ivetac, A. P. Chetwynd, S. Khalid, M. S. P. Sansom, *Structure* **2008**, *16*, 621–630.

[269] J. L. Sampaio, M. J. Gerl, C. Klose, C. S. Ejsing, H. Beug, K. Simons, A. Shevchenko, *Proceedings of the National Academy of Sciences* **2011**, DOI 10.1073/pnas.1019267108.

[270] H. Koldsø, D. Shorthouse, J. Hélie, M. S. P. Sansom, *PLoS Computational Biology* **2014**, *10*, DOI 10.1371/journal.pcbi.1003911.

[271] P. J. Stansfeld, J. E. Goose, M. Caffrey, E. P. Carpenter, J. L. Parker, S. Newstead, M. S. Sansom, *Structure* **2015**, *23*, 1350–1361.

[272] P. J. Stansfeld, M. S. P. Sansom, *Journal of Chemical Theory and Computation* **2011**, *7*, 1157–1166.

[273] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, A. D. MacKerell, *Journal of Chemical Theory and Computation* **2012**, *8*, 3257–3273.

[274] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *The Journal of Chemical Physics* **1983**, *79*, 926–935.

[275] C. Oostenbrink, A. Villa, A. E. Mark, W. F. Van Gunsteren, *Journal of Computational Chemistry* **2004**, *25*, 1656–1676.

[276] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. De Groot, H. Grubmüller, A. D. MacKerell, *Nature Methods* **2017**, *14*, 71–73.

[277] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *The Journal of Chemical Physics* **1995**, *103*, 8577–8593.

[278] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, *Computer Physics Communications* **2014**, *185*, 604–613.

[279] W. Humphrey, A. Dalke, K. Schulten, *Journal of Molecular Graphics* **1996**, *14*, 33–38.

[280] W. L. DeLano, *Schrödinger LLC www.pymol.org* **2002**, *Version 1.* http://www.pymol.org.

[281] L. L. McGoldrick, A. K. Singh, K. Saotome, M. V. Yelshanskaya, E. C. Twomey, R. A. Grassucci, A. I. Sobolevsky, *Nature* **2018**, DOI 10.1038/nature25182.

[282] A. K. Singh, K. Saotome, L. L. McGoldrick, A. I. Sobolevsky, *Nature Communications* **2018**, DOI 10.1038/s41467-018-04828-y.

[283] H. E. Autzen, A. G. Myasnikov, M. G. Campbell, D. Asarnow, D. Julius, Y. Cheng, *Science* **2018**, DOI 10.1126/science.aar4510.

[284] J. Duan, Z. Li, J. Li, A. Santa-Cruz, S. Sanchez-Martinez, J. Zhang, D. E. Clapham, *Proceedings of the National Academy of Sciences* **2018**, DOI 10.1073/pnas.1722038115.

[285] T. E. Hughes, D. T. Lodowski, K. W. Huynh, A. Yazici, J. Del Rosario, A. Kapoor, S. Basak, A. Samanta, X. Han, S. Chakrapani, Z. H. Zhou, M. Filizola, T. Rohacs, S. Han, V. Y. Moiseenkova-Bell, *Nature Structural and Molecular Biology* **2018**, DOI 10.1038/s41594-017-0009-1.

References

[286] H. I. Ingólfsson, M. N. Melo, F. J. Van Eerden, C. Arnarez, C. A. Lopez, T. A. Wassenaar, X. Periole, A. H. De Vries, D. P. Tieleman, S. J. Marrink, *Journal of the American Chemical Society* **2014**, DOI 10.1021/ja507832e.

[287] R. Garcia-Fandiño, . Piñeiro, J. L. Trick, M. S. Sansom, *ACS Nano* **2016**, *10*, 3693–3701.

[288] T. A. Dunton, J. E. Goose, D. J. Gavaghan, M. S. Sansom, J. M. Osborne, *PLoS Computational Biology* **2014**, DOI 10.1371/journal.pcbi.1003417.

[289] P. S. Niemelä, M. S. Miettinen, L. Monticelli, H. Hammaren, P. Bjelkmar, T. Murtola, E. Lindahl, I. Vattulainen, *Journal of the American Chemical Society* **2010**, *132*, 7574–7575.

[290] J. E. Goose, M. S. Sansom, *PLoS Computational Biology* **2013**, DOI 10.1371/journal.pcbi.1003033.

[291] E. Lacin, P. Aryal, I. W. Glaaser, K. Bodhinathan, E. Tsai, N. Marsh, S. J. Tucker, M. S. Sansom, P. A. Slesinger, *The Journal of General Physiology* **2017**, *149*, 799–811.

[292] W. G. Hol, P. T. Van Duijnen, H. J. Berendsen, *Nature* **1978**, *273*, 443–446.

[293] T. Rohacs, C. M. B. Lopes, T. Jin, P. P. Ramdya, Z. Molnar, D. E. Logothetis, *Proceedings of the National Academy of Sciences* **2003**, *100*, 745–750.

[294] G. Luchetti, R. Sircar, J. H. Kong, S. Nachtergaele, A. Sagner, E. F. Byrne, D. F. Covey, C. Siebold, R. Rohatgi, *eLife* **2016**, DOI 10.7554/eLife.20304.

[295] I. Levitan, D. K. Singh, A. Rosenhouse-Dantsker, Cholesterol binding to ion channels, **2014**.

[296] S. L. Morales-Lázaro, T. Rosenbaum in *Current Topics in Membranes*, **2017**, pp. 139–161.

[297] A. G. Lee, *Biophysical Journal* **2018**, *115*, 522–532.

[298] J. Oates, A. Watts, Uncovering the intimate relationship between lipids, cholesterol and GPCR activation, **2011**.

[299] S. B. Long, E. B. Campbell, R. MacKinnon, *Science* **2005**, *309*, 903–908.

[300] F. Yang, X. Xiao, W. Cheng, W. Yang, P. Yu, Z. Song, V. Yarov-Yarovoy, J. Zheng, *Nature Chemical Biology* **2015**, *11*, 518–524.

[301] F. Yang, S. Vu, V. Yarov-Yarovoy, J. Zheng, *Proceedings of the National Academy of Sciences* **2016**, *113*, E3657–E3666.

[302] J. Teng, S. H. Loukin, A. Anishkin, C. Kung, *Proceedings of the National Academy of Sciences* **2015**, *112*, 3386–3391.

[303] A. N. Bukiya, A. Rosenhouse-Dantsker, *Biochimica et Biophysica Acta - Biomembranes* **2017**, *1859*, 1233–1241.

[304] J. J. De Jesús-Pérez, S. Cruz-Rangel, . E. Espino-Saldaña, A. Martínez-Torres, Z. Qu, H. C. Hartzell, N. E. Corral-Fernandez, P. Pérez-Cornejo, J. Arreola, *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* **2018**, *1863*, 299–312.

[305] S. González-Perrett, K. Kim, C. Ibarra, A. E. Damiano, E. Zotta, M. Batelli, P. C. Harris, I. L. Reisin, M. A. Arnaout, H. F. Cantiello, *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 1182–1187.

[306] M. J. Robertson, J. Tirado-Rives, W. L. Jorgensen, *Journal of Chemical Theory and Computation* **2015**, *11*, 3499–3509.

[307] J. L. Abascal, C. Vega, *Journal of Chemical Physics* **2005**, *123*, DOI 10.1063/1.2121687.

[308] G. Klesse, S. Rao, M. S. Sansom, S. J. Tucker, *Journal of Molecular Biology* **2019**, *431*, 3353–3365.

[309] M. Arif Pavel, C. Lv, C. Ng, L. Yang, P. Kashyap, C. Lam, V. Valentino, H. Y. Fung, T. Campbell, S. G. Møller, D. Zenisek, N. G. Holtzman, Y. Yu, *Proceedings of the National Academy of Sciences* **2016**, *113*, E2363–E2372.

[310] R. F. Stephens, W. Guan, B. S. Zhorov, J. D. Spafford, Selectivity filters and cysteine-rich extracellular loops in voltage-gated sodium, calcium, and NALCN channels, **2015**.

[311] C. E. Naylor, C. Bagnéris, P. G. DeCaen, A. Sula, A. Scaglione, D. E. Clapham, B. Wallace, *The EMBO Journal* **2016**, *35*, 820–830.

[312] H. Shen, Q. Zhou, X. Pan, Z. Li, J. Wu, N. Yan, *Science* **2017**, *355*, DOI 10.1126/science.aal4326.

[313] P. M. Vassilev, M. Kanazirska, C. ping Ye, E. M. Brown, L. Guo, X. Z. Chen, Y. Segal, J. B. Peng, N. Basora, H. Babakhanlou, G. Cruger, M. A. Hediger, J. Zhou, *Biochemical and Biophysical Research Communications* **2001**, *282*, 341–350.

[314] K. K. Viet, A. Wagner, K. Schwickert, N. Hellwig, M. Brennich, N. Bader, T. Schirmeister, N. Morgner, H. Schindelin, U. A. Hellmich, *Structure* **2019**, *27*, 1246–1257.

[315] M. Kruse, G. R. V. Hammond, B. Hille, *Journal of General Physiology* **2012**, *140*, 189–205.

[316] S. J. Kleene, B. J. Siroky, J. A. Landero-Figueroa, B. P. Dixon, N. W. Pachciarz, L. Lu, N. K. Kleene, *PLoS ONE* **2019**, *14*, DOI 10.1371/journal.pone.0214053.

[317] A. Hofherr, C. Wagner, S. Fedeles, S. Somlo, M. Köttgen, *Journal of Biological Chemistry* **2014**, *289*, 14854–14867.

A

This chapter contains supplemental information pertaining to Chapter 1.

Table A.1: List of Available TRP Channel Structures

Protein	PDB	Resolution	Species	Lipid/Ligand
NOMPC	5VKQ	3.55	*D. melanogaster*	PCF
PC2	5K47	4.22	*H. sapiens*	N/A
PC2	5MKE	4.3	*H. sapiens*	PX6, CHS, PLM
PC2	5MKF	4.2	*H. sapiens*	PX6, CHS, PLM
PC2	5T4D	3.0	*H. sapiens*	N/A
PC2	6D1W	3.54	*H. sapiens*	N/A
PC1/PC2	6A70	3.6	*H. sapiens*	N/A
PKD2L1	5Z1W	3.38	*M. musculus*	N/A
PKD2L1	6DU8	3.11	*H. sapiens*	N/A
TRPA1	3J9P	4.24	*H. sapiens*	N/A
TRPC3	5ZBG	4.36	*H. sapiens*	N/A
TRPC3	6CUD	3.3	*H. sapiens*	6OE, FGJ
TRPC4	5Z96	3.28	*M. musculus*	Y01, LPP
TRPC4	6G1K	3.6	*D. rerio*	Y01, 44E
TRPC6	5YX9	3.8	*H. sapiens*	N/A
TRPM2	6CO7	3.07	*N. vectensis*	POV, CLR
TRPM2	6D73	3.8	*D. rerio*	N/A
TRPM2	6MIX	3.6	*H. sapiens*	N/A
TRPM2	6MIZ	6.1	*H. sapiens*	N/A
TRPM2	6MJ2	6.36	*H. sapiens*	N/A
TRPM2	6DRJ	3.3	*D. rerio*	APR-
TRPM2	6DRK	3.8	*D. rerio*	N/A
TRPM4	5WP6	3.8	*H. sapiens*	DVT
TRPM4	6BCJ	3.14	*M. musculus*	N/A
TRPM4	6BCL	3.54	*M. musculus*	N/A
TRPM4	6BCO	2.88	*M. musculus*	ATP
TRPM4	6BCQ	3.25	*M. musculus*	ATP
TRPM4	6BQR	3.2	*H. sapiens*	Y01
TRPM4	6BQV	3.1	*H. sapiens*	Y01
TRPM4	6BWI	3.7	*H. sapiens*	Y01
TRPM7	5ZX5	3.28	*M. musculus*	Y01
TRPM8	6BPQ	4.1	*F. albicollis*	N/A
TRPML1	5WJ5	3.72	*H. sapiens*	N/A
TRPML1	5WJ9	3.49	*H. sapiens*	AQV
TRPML1	5WPQ	3.64	*M. musculus*	N/A
TRPML1	5WPT	3.75	*M. musculus*	N/A
TRPML1	5WPV	3.59	*M. musculus*	N/A
TRPML1	5YDZ	5.8	*M. musculus*	N/A
TRPML1	5YE1	5.8	*M. musculus*	N/A
TRPML1	5YE2	5.8	*M. musculus*	N/A
TRPML1	5YE5	5.8	*M. musculus*	N/A

Continued on next page

Table A.1 – *Continued from previous page*

Protein	**PDB**	**Resolution**	**Species**	**Lipid/Ligand**
TRPML1	6E7P	3.5	*H. sapiens*	HZ7
TRPML1	6E7Y	3.57	*H. sapiens*	PIO
TRPML1	6E7Z	3.73	*H. sapiens*	PIO, AQV
TRPML3	5W3S	2.94	*C. jacchus*	Y01, 3PE
TRPML3	6AYE	4.06	*H. sapiens*	N/A
TRPML3	6AYF	3.62	*H. sapiens*	N/A
TRPML3	6AYG	4.65	*H. sapiens*	N/A
TRPV1	3J5P	3.28	*R. norvegicus*	N/A
TRPV1	3J5Q	3.8	*R. norvegicus*	N/A
TRPV1	3J5R	4.2	*R. norvegicus*	N/A
TRPV1	3J9J	3.275	*R. norvegicus*	N/A
TRPV1	5IRX	2.95	*R. norvegicus*	6OE, 6EU, 6O8, 6O9
TRPV1	5IRZ	3.28	*R. norvegicus*	6OE, 6O8, 6ES
TRPV1	5IS0	3.43	*R. norvegicus*	6ET
TRPV2	5AN8	3.8	*O. cuniculus*	N/A
TRPV2	6BWJ	3.1	*O. cuniculus*	6EU
TRPV2	5HI9	4.4	*R. norvegicus*	N/A
TRPV2	6OO3	2.9	*O. cuniculus*	6EU
TRPV2	6OO4	3.3	*O. cuniculus*	6EU
TRPV2	6OO5	4.2	*O. cuniculus*	6EU
TRPV2	6OO7	3.8	*O. cuniculus*	6EU
TRPV2	6BO4	4.0	*R. norvegicus*	N/A
TRPV2	6BO5	3.6	*R. norvegicus*	N/A
TRPV2	6BWM	3.1	*O. cuniculus*	6EU
TRPV3	6OT2	4.1	*H. sapiens*	N/A
TRPV3	6OT5	3.6	*H. sapiens*	FZ4
TRPV3	6MHO	3.4	*H. sapiens*	N/A
TRPV3	6MHS	3.2	*H. sapiens*	N/A
TRPV3	6MHV	3.5	*H. sapiens*	N/A
TRPV3	6MHW	4.0	*H. sapiens*	N/A
TRPV3	6MHX	4.0	*H. sapiens*	N/A
TRPV3	6DVW	4.3	*M. musculus*	N/A
TRPV3	6DVY	4.0	*M. musculus*	FZ4
TRPV3	6DVZ	4.24	*M. musculus*	FZ4
TRPV4	6BBJ	3.8	*X. tropicalis*	N/A
TRPV5	6B5V	4.8	*O. cuniculus*	ECL
TRPV5	6O1N	2.9	*O. cuniculus*	N/A
TRPV5	6O1P	3.0	*O. cuniculus*	N/A
TRPV5	6O1U	2.8	*O. cuniculus*	N/A
TRPV5	6O20	3.3	*O. cuniculus*	N/A
TRPV5	6DMR	3.9	*O. cuniculus*	N/A
TRPV5	6DMU	4.0	*O. cuniculus*	PIO

Continued on next page

Table A.1 – *Continued from previous page*

Protein	**PDB**	**Resolution**	**Species**	**Lipid/Ligand**
TRPV5	6DMW	4.4	*R. norvegicus*	N/A
TRPV6	5IWK	3.25	*R. norvegicus*	DTB
TRPV6	5IWP	3.65	*R. norvegicus*	DTB
TRPV6	5IWR	3.85	*R. norvegicus*	DTB
TRPV6	5IWT	3.8	*R. norvegicus*	DTB
TRPV6	6BO8	3.6	*H. sapiens*	N/A
TRPV6	6BO9	4.0	*H. sapiens*	N/A
TRPV6	6BOA	4.2	*H. sapiens*	N/A
TRPV6	6BOB	3.9	*H. sapiens*	N/A
TRPV6	6E2F	3.9	*H. sapiens*	N/A
TRPV6	6E2G	3.6	*H. sapiens*	N/A
TRPV6	6D7O	3.45	*R. norvegicus*	DTB, FZ4
TRPV6	6D7P	3.37	*R. norvegicus*	DTB
TRPV6	6D7Q	3.497	*R. norvegicus*	DTB, FZ4
TRPV6	6D7S	4.34	*H. sapiens*	N/A
TRPV6	6D7T	4.44	*H. sapiens*	FZ4
TRPV6	6D7V	4.3	*R. norvegicus*	FYY
TRPV6	6D7X	3.6	*R. norvegicus*	FYY, DTB

Abbreviations of ligand names are listed below.

Abbreviations of Ligand Names

PX6	1,2-DIPALMITOYL-SN-GLYCERO-3-PHOSPHATE
CHS	4-AMINO-5-CYCLOHEXYL-3-HYDROXY-PENTANOIC ACID
PLM	PALMITIC ACID
PCF	1,2-DIACYL-SN-GLYCERO-3-PHOSHOCHOLINE
POV	(2S)-3-(hexadecanoyloxy)-2-[(9Z)-octadec-9-enoyloxy]propyl 2-(trimethylammonio)ethyl phosphate
CLR	CHOLESTEROL
ATP	ADENOSINE-5'-TRIPHOSPHATE
Y01	CHOLESTEROL HEMISUCCINATE
AQV	2-{2-oxo-2-[(4S)-2,2,4-trimethyl-3,4-dihydroquinolin-1(2H)-yl]ethyl}-1H-isoindole-1,3(2H)-dione
3PE	1,2-DIACYL-SN-GLYCERO-3-PHOSPHOETHANOLAMINE
6OE	(2S)-3-{[(S)-(2-aminoethoxy)(hydroxy)phosphoryl]oxy}-2-(hexanoyloxy)propyl hexanoate
6EU	resiniferatoxin
6O8	(4R,7S)-4-hydroxy-N,N,N-trimethyl-4,9-dioxo-7-[(pentanoyloxy)methyl]-3,5,8-trioxa-4lambda∼5∼-phosphatetradecan-1-aminium
6O9	(2S)-2-(acetyloxy)-3-{[(R)-(2-aminoethoxy)(hydroxy)phosphoryl]oxy}propyl pentanoate
6ES	(2S)-1-{[(R)-hydroxy{[(1R,2R,3S,4S,5S,6S)-2,3,4,5,6-pentahydroxycyclohexyl]oxy}phosphoryl]oxy}-3-(pentanoyloxy)propan-2-yl decanoate
6ET	capsazepine
ECL	1-[(2R)-2-[(4-chlorobenzyl)oxy]-2-(2,4-dichlorophenyl)ethyl]-1H-imidazole
DTB	6-(5-METHYL-2-OXO-IMIDAZOLIDIN-4-YL)-HEXANOIC ACID
FZ4	2-aminoethyl diphenylborinate
HZ7	(1R,2S,3S,4R,5S,6R)-5-{[(R)-[(2R)-2,3-bis{[(1S)-1-hydroxyoctyl]oxy}propoxy](hydroxy)phosphoryl]oxy}-2,4,6-trihydroxycyclohexane-1,3-diyl bis[dihydrogen (phosphate)]
PIO	[(2R)-2-octanoyloxy-3-[oxidanyl-[(1R,2R,3S,4R,5R,6S)-2,3,6-tris(oxidanyl)-4,5-diphosphonooxy-cyclohexyl]oxy-phosphoryl]oxy-propyl] octanoate
APR-	ADENOSINE-5-DIPHOSPHORIBOSE
FYY	2-aminoethyl (4-bromophenyl)phenylborinate
FGJ	(2R)-3-hydroxypropane-1,2-diyl dihexanoate
44E	(2R)-3-(phosphonooxy)propane-1,2-diyl dihexanoate
LPP	2-(HEXADECANOYLOXY)-1-[(PHOSPHONOOXY)METHYL]ETHYL HEXADECANOATE
DVT	DECAVANADATE

B

This chapter contains supplemental information pertaining to Chapter 3.

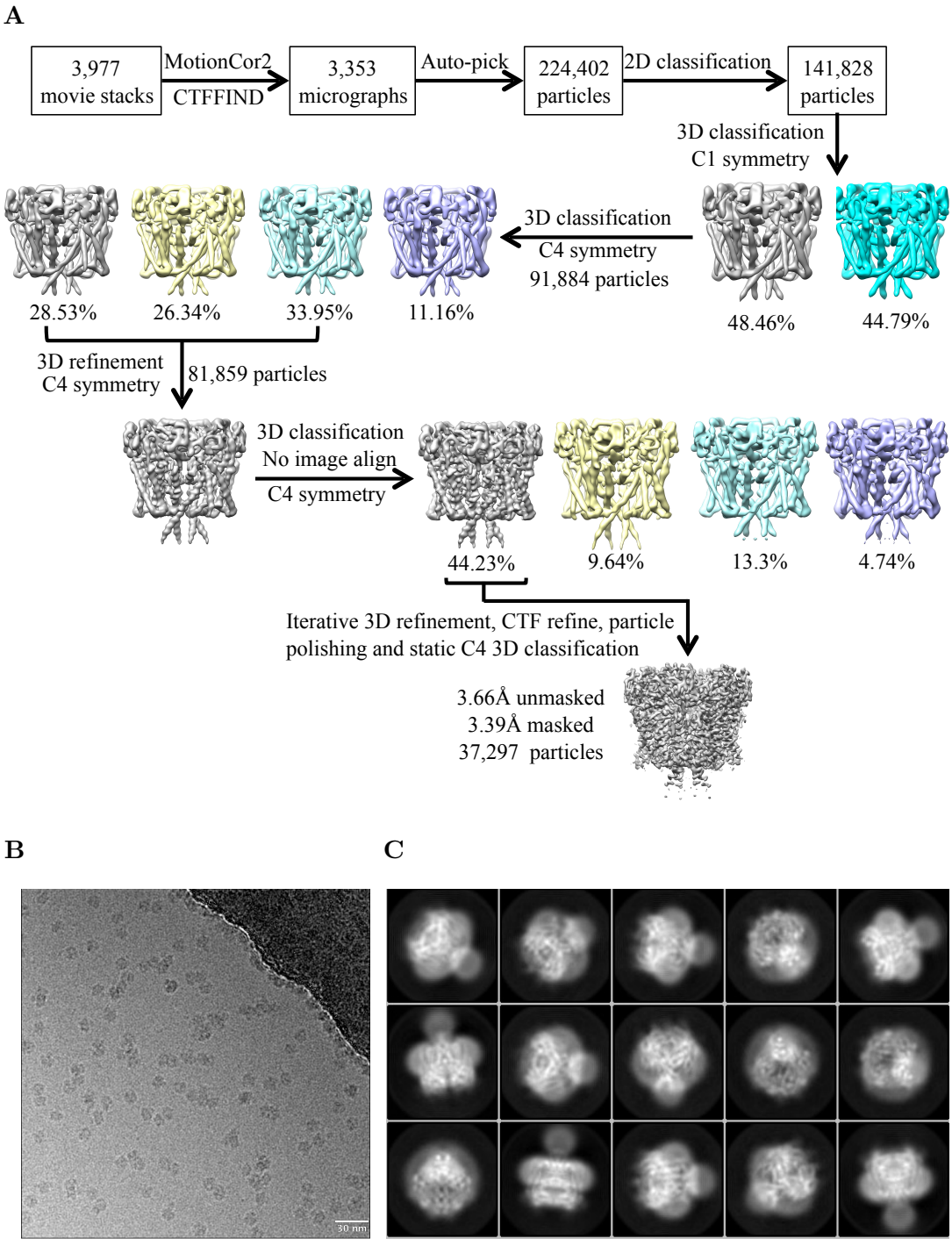


Figure B.1: **A:** Workflow of data processing for the dataset of PC2 with PI(3,5)P_2 in detergent. **B:** Representative micrograph of PC2 particles recorded on Titan Krios 300-kV transmission electron microscope. **C:** Representative 2D class averages of PC2.

C

This chapter contains supplemental information pertaining to Chapter 4.

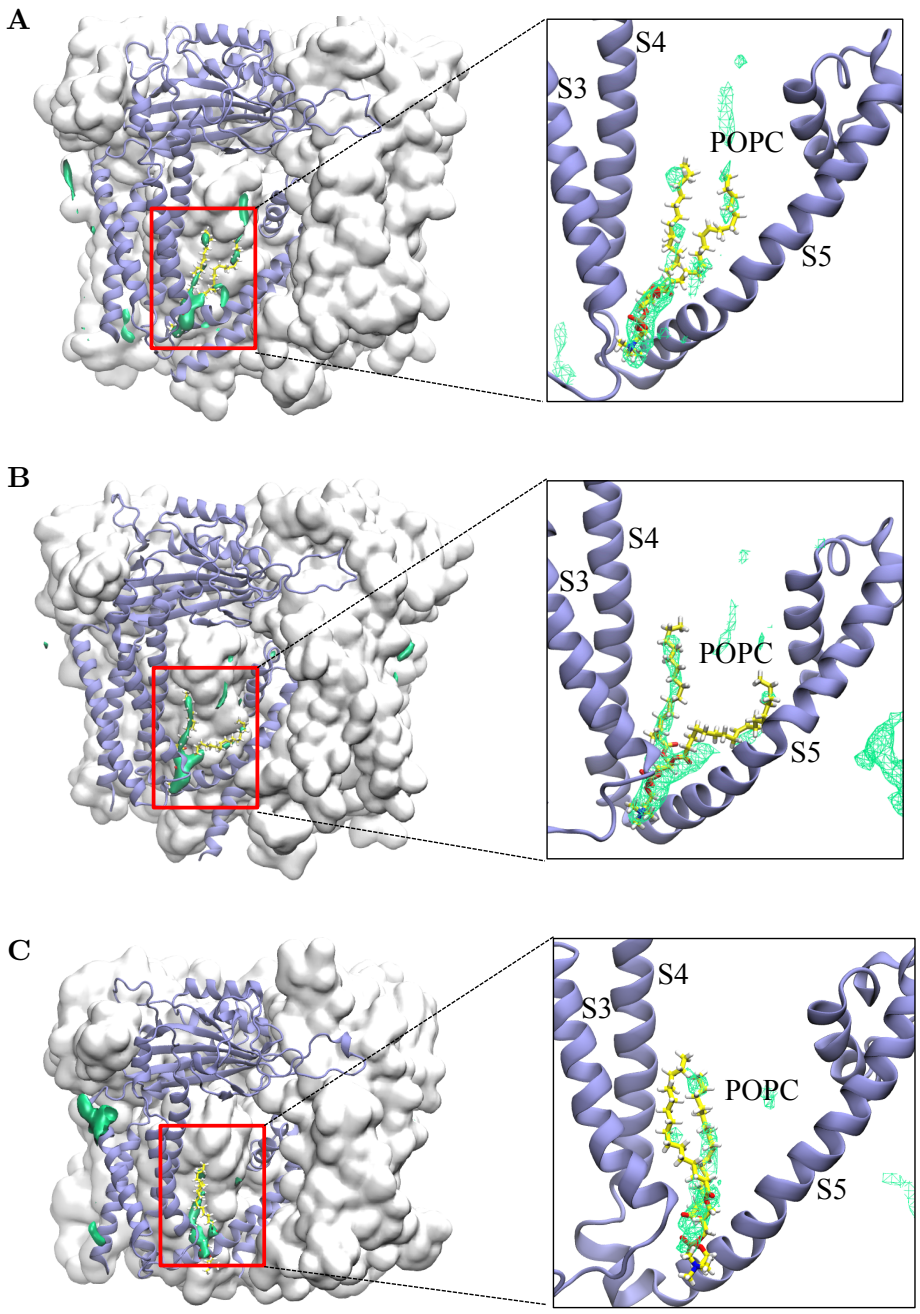


Figure C.1: Probability density of phosphatidylcholine (POPC) in atomistic simulations of different PC2 cryo-EM structures (**A:** 5T4D, **B:** 5MKF, **C:** 6D1W). In each case, in the left panel, the PC2 protein is shown as a grey surface, viewed perpendicular to the central pore axis with one subunit depicted as pale purple cartoon. Green isocontour surfaces represent a high probability of occurrence of phospholipid molecules. In the right panel, the corresponding zoomed in view (red box in left panels) of the S3/S4/S5 pocket and the high phospholipid occurrence density with a POPC molecule taken from a simulation snapshot) shown within the density.

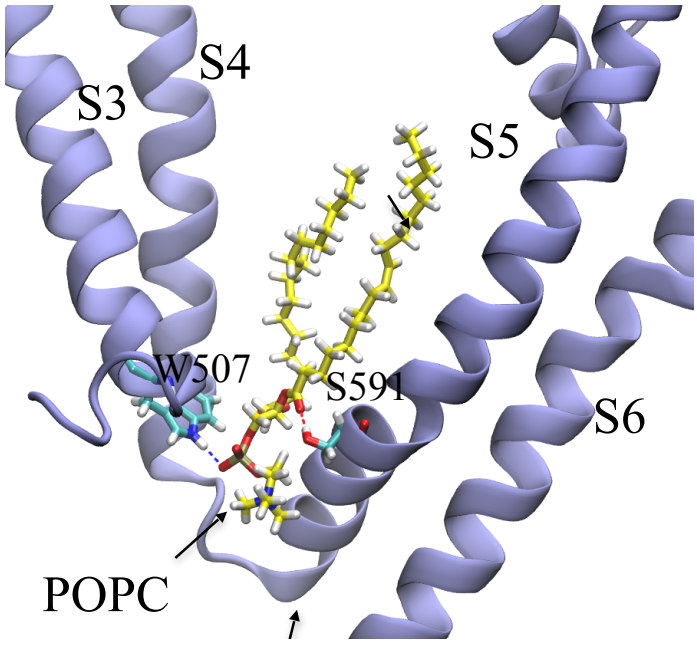


Figure C.2: POPC interaction with PC2 at the possible lipid binding site identified in the atomistic simulations. S3-S6 of PC2 (taken from a simulation snapshot of PDB ID: 5K47) is shown as pale purple cartoon with a bound POPC molecule shown as yellow sticks. Residues Trp507 and Ser591 which formed the most hydrogen bonds between the lipid and protein during the simulations are shown as blue sticks.

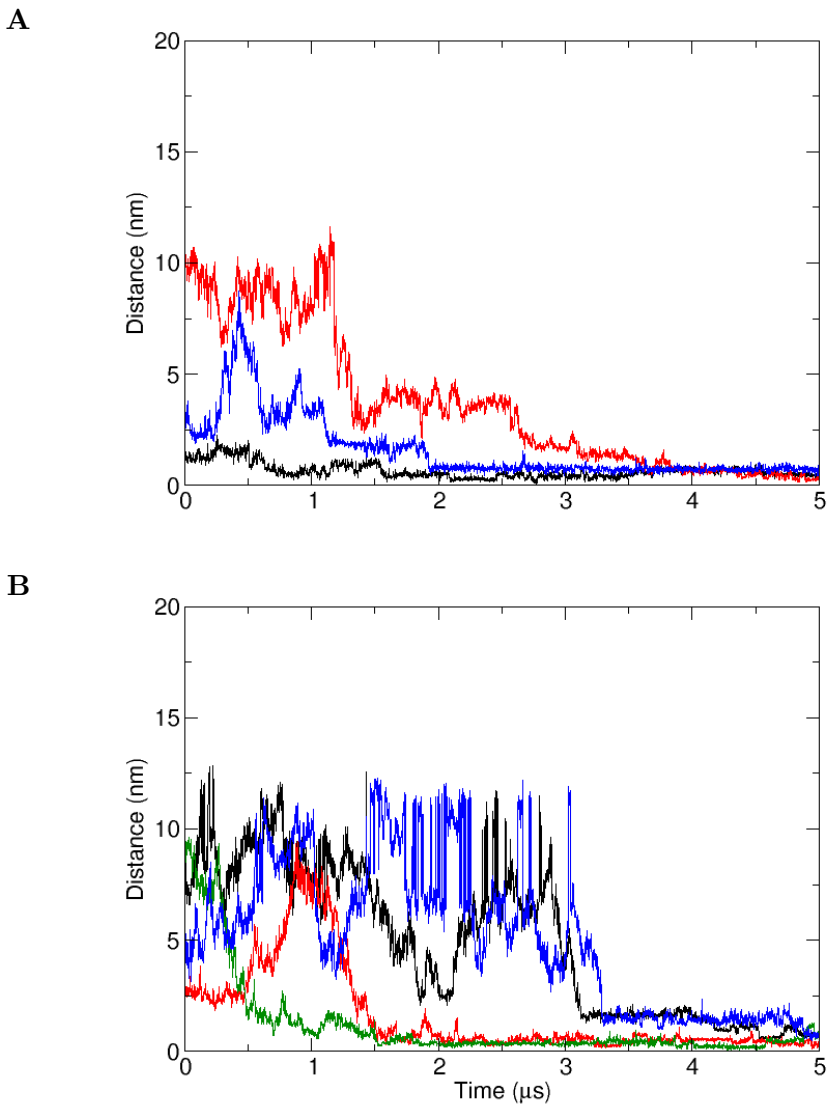


Figure C.3: PIP_2 binding in two repeats of a coarse-grained (CG) simulations of PC2 (structure 5K47) in an *in vivo* mimetic mixed lipid bilayer (also see Figure 4.4B of the main text for the third simulation of this triplicate). The distances from their binding site on PC2 of PIP_2 molecules are shown as functions of time. Distances are measured from the centre of mass of the headgroup of each PIP_2 molecule to the centre of mass of residues S505 and S591 of the site to which that lipid molecule eventually binds. The different coloured lines correspond to the different PIP_2 molecules.

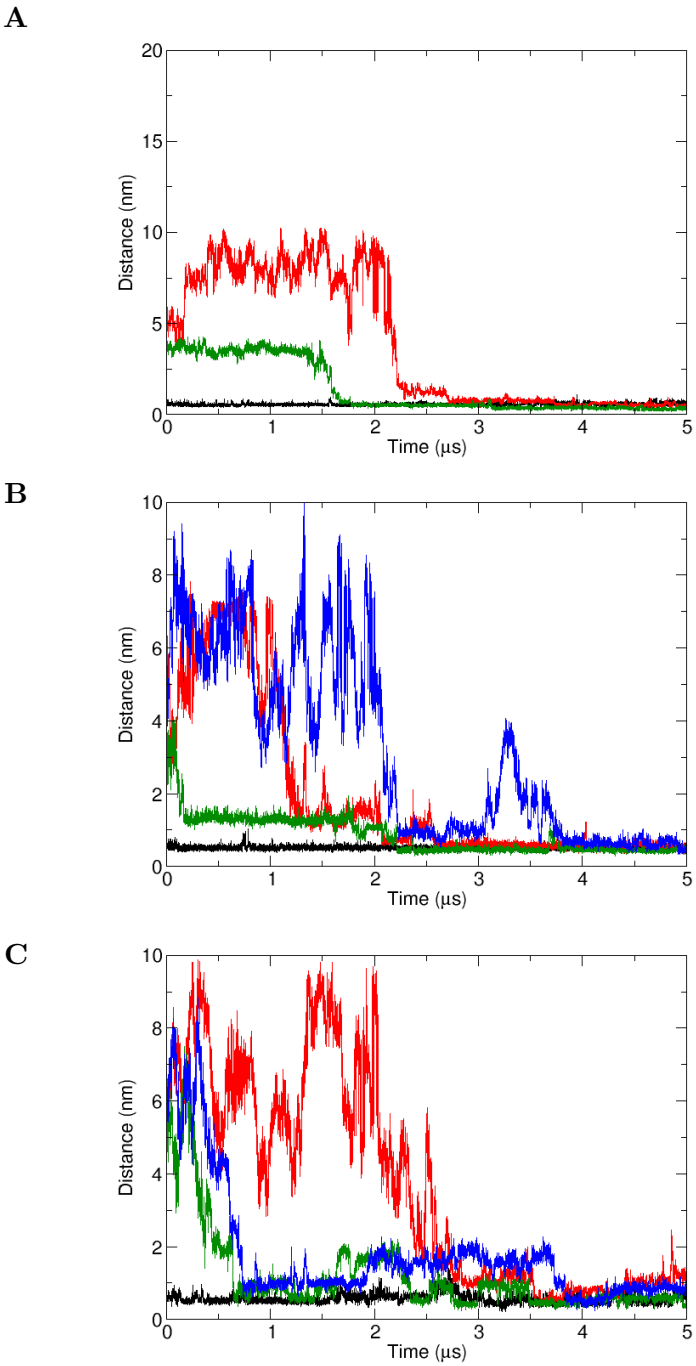


Figure C.4: PIP_2 binding in three repeats of CG simulations of a gain-of-function Phe604Pro mutant of PC2 (PDB ID: 6D1W) in an *in vivo* mimetic mixed lipid bilayer. Other details are as for Figure C.3.

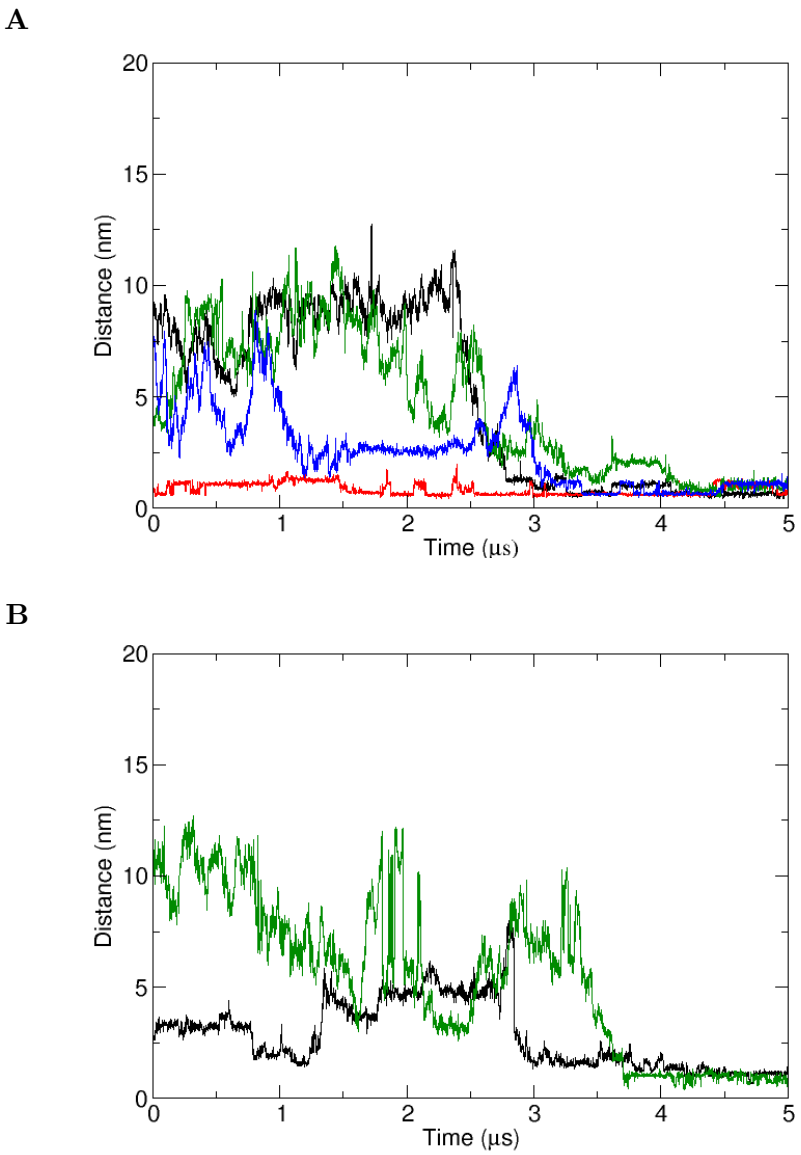


Figure C.5: Cholesterol binding in two repeats of CG simulations of PC2 (PDB ID: 5K47) in an *in vivo* mimetic mixed lipid bilayer. Distances from the binding site as function of time for cholesterol molecules are shown. The distance shown is from the centre of mass of each cholesterol molecule to the centre of mass of Gln557 and Asn560 at the site to which that lipid molecule eventually binds. Different coloured lines correspond to the different cholesterol molecules. This is related to Figure 4.14C.

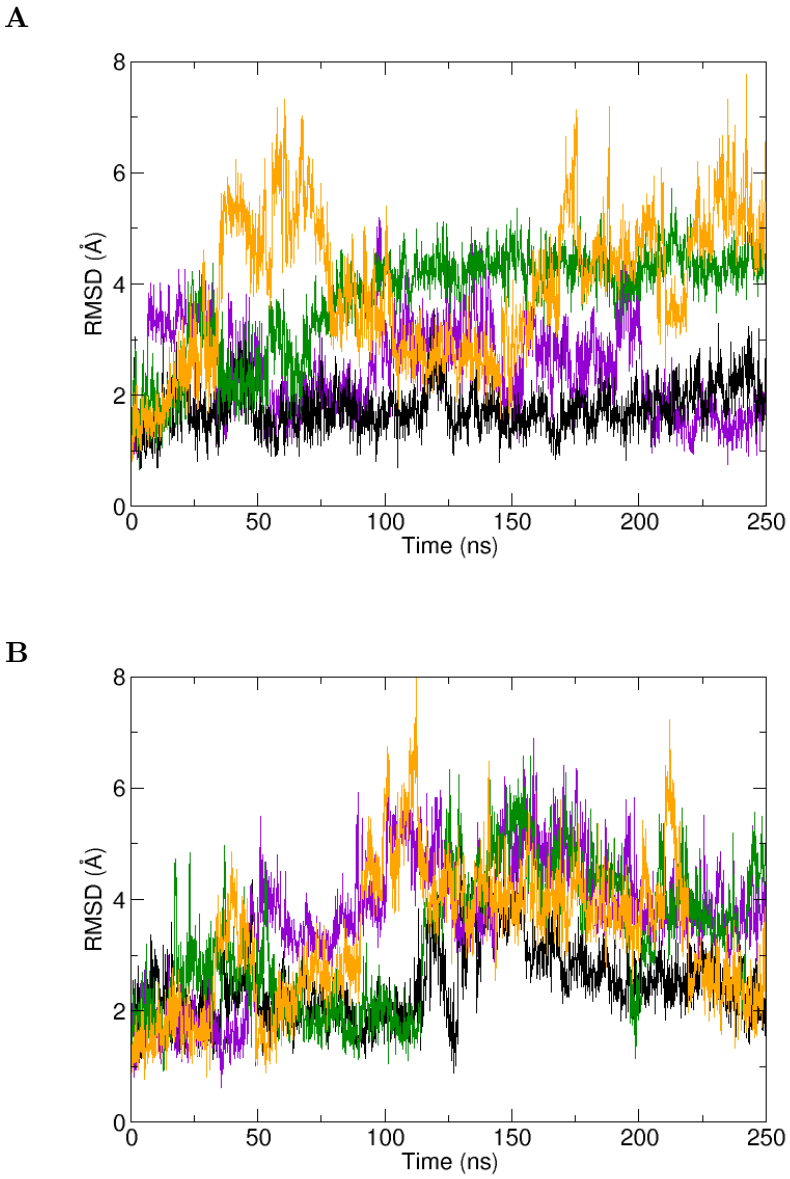


Figure C.6: Two repeats of atomistic simulations of cholesterol-bound PC2, with the cholesterol molecule initially built into the cryo-EM density. RMSD vs. time for the four cholesterol molecules bound to PC2 during 250 ns atomistic MD simulations are shown. Different coloured lines correspond to the different cholesterol molecules. This is related to Figure 4.15A.