The Influence of Polarisation on the Behaviour of Anions Within Ion Channels and Nanopores



Linda Xi Phan

Department of Physics

Balliol College, University of Oxford

A thesis submitted in partial fulfilment
of the requirements for the degree of

Doctor of Philosophy

Michaelmas 2023

To my parents
Hai and Ling

Acknowledgements

First and foremost, I would like express my gratitude towards my supervisors Prof. Mark Sansom, Prof. Stephen Tucker and Prof. Jason Crain. Thank you for sharing your knowledge, enthusiasm and invaluable guidance in the pursuit of scientific ideas. I am incredibly privileged and deeply grateful for the opportunity to learn and develop as a scientist alongside you. I would like to thank the Engineering and Physical Sciences Research Council and IBM Research for funding.

SBCB has provided a wonderfully unique environment for research to thrive. Particular thanks go to Dr Shanlin Rao and Dr Gianni Klesse for laying the groundwork to this upcoming work and to Dr Charlotte Lynch for all things watery. Appreciation to Dr Irfan Alibay and Dr Michael Horrell for workstation and cluster maintenance. Thanks also to Dr Laura John, Dr Robin Corey, Dr Astrid Bradner, Patrick Simcock and David Seiferth for sharing molecular dynamics expertise and countless rewarding conversations throughout the DPhil. Special thanks go to Dr Bertie Ansell for their fruitful discussions (especially their opinions on shower gel sandwiches) and being the better half of giggle corner, and to Daniel Crusius for always being willing to procrastinate with me and being hungry when I am. Lastly, thanks to the wider SBCB lab (past and present) for a wholesome and nurturing environment for research. It is an honour to hold the title of *The Last Sansomnite*.

To the rare lifelong friendships that formed during my early teenage years, thank you for sticking with me through each chapter of my story. I am extremely fortunate and thankful to have felt the unconditional support of Helena Bradley, Rosie Morse, Gemma Brandon and Rebecca Velickovic. Your friendships have been a source of joy, balance and great comfort. PortuGal love.

Gratitude is owed to Balliol College for providing me a home (not so far from home) and an outstanding community of people. Within Balliol, I consider myself exceptionally lucky to have an incredible group of friends in Hannah Ogden, Emily Meekel, Ellen Baker, Moa Schafer and Vicky Taylor; the most fiercely strong and kind-hearted group of women who have carried me through my DPhil in more ways than one. Many more people have made my time at Oxford unforgettable. I am beyond thankful for the warmth and unbounded friendships of Finlay Ryburn, Shu Wei, Matthew Jackson and Darius Mcardle - phenomenal human beings. To all the above, thank you for sharing the load. I appreciate each of you for enduring this journey with me - from the belly-aching laughs to the tears and everything in between. 11 stars.

Finally, to my family, in particular my parents, you may never quite understand what it is that I do but I am eternally grateful for your endless love, support and encouragement in my various endeavours.

Abstract

Ion and water interactions within ion channels and nanopores hold significant biological and technological importance. Whilst cation channels have been studied in great depth, general mechanisms underlying anion selectivity and permeation within such pores remain unclear. Electronic polarisation arises from the deformation of electronic shells. In molecular dynamics simulations, polarisation is not included in commonly used additive forcefields which therefore are insufficient for the study of polarisable anions. Previous simulations of halide ions at aqueous/hydrophobic interfaces have demonstrated an anion affinity for the interface which can be markedly influenced by the inclusion of polarisation. This thesis concerns the assessment of anion-hydrophobic interactions of biological significance and the resulting behaviours from modelling electronic polarisation with varying degrees of physical accuracy. To begin these studies, an *in silico* model biomimetic nanopore was designed to facilitate the exploration of Cl^- interactions with hydrophobic surfaces in confinement. Polarisation modelled implicitly through the electronic continuum correction revealed Cl^- adsorption to hydrophobic interfaces was achieved through partial ion dehydration of the first hydration shell. Free energy calculations provided energetic contextualisation of Cl^- interactions relative to Na^+. Expanding upon this, a hydrophobic Cl^--pumping rhodopsin binding site was surveyed using different levels of granularity for capturing polarisation. Simulations employing the mean-field prosECCo forcefield revealed stronger and longer ion binding than traditional additive forcefields suggest and free energy calculations identify a second metastable site where key ion-hydrophobic interactions promote ion rebinding. These simulations rationalise the use of implicit polarisation methods as a proxy for identifying interesting interactions to be probed further with more advanced methods. Mechanistic insight was gained through simulations of the open and partially open states of the human bestrophin 1 Cl^- channel in fully polarisable simulations. The explicit modelling of quadrupolar moments suggest key roles of aromatic residues in anion permeation modulated by the Cl^- hydration response. This behaviour was not captured in non-polarisable simulations. Furthermore, we demonstrate how polarisable simulations can aid the interpretation of ambiguous structural densities from cryo-EM and inform the functional annotation of ion channels. Overall, this thesis demonstrates the importance of including electronic polarisation in realistic and physically accurate models of anion selective membrane channels and transporters.

Publications

The results presented in this thesis have contributed to the following publications:

1. **Phan, L. X.**, Lynch, C. I., Crain, J., Sansom, M. S. P. and Tucker, S. J. Influence of Effective Polarization on Ion and Water Interactions within a Biomimetic Nanopore. *Biophysical Journal*, 121, 11, 2014-2026 (2022)

2. **Phan, L. X.**, Chamorro, V. C., Martinez-Seara, H., Crain, J., Sansom, M. S. P., Tucker, S. J. Influence of Electronic Polarization on the Binding of Anions to a Chloride-Pumping Rhodopsin. *Biophysical Journal*, 122, 8, 1548-1556 (2023)

3. **Phan, L. X.**, Owji, A. P., Yang, T., Crain, J., Sansom, M. S. P., Tucker, S. J. Electronic Polarizability Tunes the Function of the Human Bestrophin 1 Channel. *bioRxiv*, doi.org/10.1101/2023.11.14.567055 (2023)

Contents

List of Figures **x**

List of Tables **xii**

List of Abbreviations **xiii**

1 Introduction **1**
1.1 Ion Channels and Nanopores . . . 1
1.1.1 Biophysical Properties and Physiological Roles . . . 4
1.1.2 From Structure to Function . . . 5
1.1.3 Anion Selective Ion Channels, Transporters and Nanopores . . . 7
1.1.4 Model Nanopores . . . 9
1.2 Properties of Water and Halide Ions . . . 10
1.2.1 Water & Hydrophobic Gating . . . 10
1.2.2 Halide Ions . . . 11
1.3 Molecular Dynamics Simulations for Channels, Transporters and Nanopores 14
1.3.1 Computational Methods for the Study of Channels and Transporters 14
1.3.2 Electronic Polarisation in Biomolecular Simulations . . . 16
1.4 Thesis Outline and Objectives . . . 18

2 Theory and Methods **20**
2.1 Classical Molecular Dynamics . . . 21
2.1.1 Pairwise Additive Forcefields . . . 21
2.1.2 Molecular Dynamics Algorithms, Settings and Parameters . . . 31
2.2 Polarisation Effects in Classical Molecular Dynamics . . . 39
2.2.1 Polarisable Atomic Multipoles . . . 41
2.2.2 Charge Scaling and Mean-field Methods . . . 46
2.3 Non-equilibrium Methods . . . 49
2.3.1 Potential of Mean Force Calculations . . . 49

3 A Model Biomimetic Nanopore **51**
3.1 Introduction . . . 52
3.2 Objectives . . . 54
3.3 Aqueous/Decane Interface System . . . 54
3.3.1 Methods . . . 55
3.3.2 Results & Discussion . . . 57
3.4 Model Biomimetic Nanopore . . . 63
3.4.1 Model Nanopore Design . . . 63
3.4.2 Methods . . . 65
3.4.3 Results and Discussion . . . 68
3.5 Summary & Conclusions . . . 85

4 Anion-Hydrophobic Interactions in a Protein **86**
4.1 Introduction . . . 87
4.2 Objectives . . . 89
4.3 Methods . . . 90
4.3.1 Protein Truncation and System Preparation . . . 90
4.3.2 Equilibrium Molecular Dynamics Simulations . . . 92
4.3.3 AMOEBA Forcefield Simulations . . . 93
4.3.4 Umbrella Sampling . . . 94
4.4 Results and Discussion . . . 96
4.4.1 Influence of Effective Polarisation on the Energetics of Cl^- Ion Binding 96
4.4.2 Influence of Polarisation on Ion Binding Dynamics . . . 101
4.4.3 Cl^--Hydrophobic Interactions within the Binding Site . . . 103
4.4.4 Mechanisms of Cl^- Binding . . . 106
4.5 Summary & Conclusions . . . 109

5 Simulation Studies of New Channels **111**
5.1 Introduction . . . 112
5.2 Objectives . . . 116
5.3 Methods . . . 117
5.3.1 Structural Model and System Preparation . . . 117
5.3.2 Non-polarisable MD Simulations . . . 119
5.3.3 Constant Electric Field Simulations . . . 120
5.3.4 AMOEBA Simulations . . . 120
5.3.5 Binding Site Identification and Analysis . . . 121
5.4 Results and Discussion . . . 122
5.4.1 Model System Validation . . . 122
5.4.2 Interpretation of Cryo-EM Densities From Fully Polarisable MD Simulations . . . 125
5.4.3 Functional Advantages of Pore Asymmetry . . . 131
5.4.4 The Role of Water in Cl^- Permeation . . . 135
5.4.5 Pathway for Conformational Change Between Open and Partially Open States . . . 138
5.4.6 The Partially Open Neck state . . . 139
5.4.7 Analysis of the Partially Open Neck State . . . 140
5.5 Summary & Conclusions . . . 148

6 Conclusions and Future Directions **150**

Bibliography **156**

List of Figures

1.1 Molecular Transport Through Biological Membranes 2
1.2 Ion Channels in the PDB 6
1.3 Examples Anion Channels Structures 8
1.4 Definition of an Interface 13
1.5 Timescale Ranges of Ion Channels 15

2.1 Bonded Interactions 23
2.2 Non-bonded Interactions 25
2.3 Rigid Fixed-Charge Water Models 29
2.4 Comparison of Fixed Charge and Polarisable Models 40
2.5 Polarisable Multipole Water Model 43
2.6 Schematic diagram of Forcefield Types Based on the Modelling of Electrostatics 47

3.1 Schematic of Aqueous/Decane Model Interface System 55
3.2 Aqueous/Hydrophobic Interfacial System Tests 59
3.3 Aqueous/Hydrophobic Interfacial System 61
3.4 AMOEBA Simulation of Hydrophobic Interface 62
3.5 Schematic of the Biomimetic Model Nanopore Design 64
3.6 Symmetrised Number Density Profiles as a Function of Salt Concentration . 69
3.7 Symmetrised Number Density Profiles as a Function of Pore Radius 72
3.8 Potentials of Mean Force Ion Profiles 74
3.9 Convergence of PMF Profiles 75
3.10 Variation in Ion First Hydration Shell 76
3.11 Cl^- Hydration Structure Analysis 77
3.12 Cl-O RDFs 78
3.13 Na^+ Hydration Structure 80
3.14 ECCR Number Density 82
3.15 ECCR PMF Comparison and Convergence 84

4.1 Structure of the Cl^--Pumping Rhodopsin 87
4.2 Definition of Systems Prepared 91
4.3 Cl^--O RDFs 96
4.4 Potential of Mean Force Profiles 97
4.5 Potential of Mean Force Convergence Tests for Cl^- Profiles 98
4.6 Potential of Mean Force Profiles of Cl^- vs Br^- 100
4.7 Ion-distance Plots 101
4.8 Ion-distance plots of Cl^--H-A44 102
4.9 Ion-distance Plots as a Function of Forcefield 104
4.10 Snapshot of Cl^- in Concave Site Conformation 106
4.11 Large Fragment Binding Site Dynamics 107
4.12 Ion-distance Plots from Full Protein Simulations 108

5.1 Example of an Anion-Aromatic Interaction 112
5.2 Example of Aromatic Sidechains in an Anion Channel 113
5.3 Open and Partially Open States of the Human Bestrophin 1 Channel 115
5.4 Structural Model Preparation 118
5.5 Trajectories of Water and Ions Within Simulations of Whole Protein-Embedded Systems 123
5.6 Protein Fragment System Validation 125
5.7 The Cl^- Binding Site 127
5.8 Alternative Binding Sites Identified with PyLipid 129
5.9 RMSD of F80 Sidechains in the Open State 132
5.10 Capturing Phenylalanine Sidechain Reorientation 133
5.11 F80 Sidechain Flipping Motion 135
5.12 Cl^- Partial Dehydration is Compensated by Anion-π Interactions 136
5.13 RMSD of F80 Sidechains from the Open State 137
5.14 Partially Open State Sidechain Dynamics of F80 139
5.15 Cryo-EM Map and Model of the Partially Open State at F84 140
5.16 Uniform External Electric Field Applied to the Partially Open State 142
5.17 The Partially Open State 143
5.18 Coordination Number of Cl^- in the Partially Open State 145
5.19 Partially Open State Sidechain Dynamics of F84 147

6.1 Summary of Cl^--Hydrophobic Interactions in this Thesis 151

List of Tables

2.1 Water Model Properties 30
2.2 AMOEBA03 Water Model Properties 44

3.1 ECC Test Combinations 56

4.1 Table of Forcefield Performance Comparison 92
4.2 Average Hydration Number of Cl^- 104

5.1 Detected AMOEBA Simulation Binding Sites 129

List of Abbreviations

AA . . . . . All-atom

AMOEBA . . . Atomic Multipole Optimised Energetics for Biomolecular Applications

AMBER . . . . Assisted Model Building with Energy Refinement

Br^- . . . . . Bromide

c36 . . . . . CHARMM36

c36m . . . . . CHARMM36 modified

CG . . . . . Coarse-grained

CHAP . . . . . CHannel Annotation Package

CHARMM . . . Chemistry at HARvard Macromolecular Mechanics

Cl^- . . . . . Chloride

CLC . . . . . Chloride Channel

CNT . . . . . Carbon nanotube

COG . . . . . Center of geometry

COM . . . . . Center of mass

cyro-EM . . . . Cryogenic electron microscopy

CV . . . . . Collective variable

DFT . . . . . Density Functional Theory

ECC . . . . . Electronic Continuum Correction

ECCR . . . . . Electronic Continuum Correction with Rescaling

F^- . . . . . Fluoride

FEP . . . . . Free energy perturbation

FF . . . . . Forcefield

GDS . . . . . Gibbs dividing surface

GROMOS . . . GROningen MOlecular Simulation

I^- . . . . . Iodide

LJ . . . . . Lennard-Jones

MD . . . . . Molecular dynamics

NMR . . . . . Nuclear magnetic resonance

NP . . . . . Non-polarisable

OPLS/AA . . . . Optimised Potentials for Liquid Simulations All-Atom

PBC . . . . . Periodic boundary conditions

PDB . . . . . Protein Data Bank

pLGICs . . . . . Pentameric Ligand-Gated Ion Channels

PME . . . . . Particle Mesh Ewald

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

prosECCo . . . Polarisation Reintroduced by Optimal Scaling of Electronic Continuum Correction Origin

QM . . . . . Quantum mechanical

QM/MM . . . . Quantum-mechanics/molecular-mechanics

RDF . . . . . Radial distribution function

RMSD . . . . . Root Mean Square Deviation

r-RESPA . . . . Reversible reference system propagation algorithm

SF . . . . . Scaling factor

SFG . . . . . Sum-frequency generation

SHG . . . . . Second harmonic generation

SPC . . . . . Simple point charge

SPC/E . . . . . Extended simple point charge

SOR . . . . . Successive over-relaxation

TIP . . . . . Transferable intermolecular potential

TM . . . . . Transmembrane

TMD . . . . . Transmembrane domain

vdW . . . . . van der Waals

V-rescale . . . . Velocity-rescale

WHAM . . . . . Weighted Histogram Analysis Method

w.r.t . . . . . with respect to

"The richness I achieve comes from nature, the source of my inspiration."

— Claude Monet

1 Introduction

Contents

1.1 Ion Channels and Nanopores 1
1.1.1 Biophysical Properties and Physiological Roles 4
1.1.2 From Structure to Function 5
1.1.3 Anion Selective Ion Channels, Transporters and Nanopores 7
1.1.4 Model Nanopores 9
1.2 Properties of Water and Halide Ions 10
1.2.1 Water & Hydrophobic Gating 10
1.2.2 Halide Ions 11
1.3 Molecular Dynamics Simulations for Channels, Transporters and Nanopores . 14
1.3.1 Computational Methods for the Study of Channels and Transporters 14
1.3.2 Electronic Polarisation in Biomolecular Simulations 16
1.4 Thesis Outline and Objectives 18

1.1 Ion Channels and Nanopores

Biological cell membranes are typically composed of amphipathic lipid molecules that self-assemble into two layers to form lipid bilayers. Membranes serve as a physical barrier between the intracellular and extracellular environments of the cell or to separate regions within the cell into distinct aqueous compartments. Additionally, they are considered to be the boundary separating two salt reservoirs at different concentrations, giving rise to an electric potential difference across the cell membrane [1]. The central region within these bilayers is hydrophobic and is relatively impermeable to ions and other polar molecules. Therefore, to enable the transport of ions into and out of the cell or between

different intracellular compartments, transmembrane (TM) proteins known as ion channels and transporters are required (see Figure 1.1) [2].

Transporters expend energy for ion translocation. Primary active transporters such as ion pumps utilise energy, derived from ATP hydrolysis, to drive ions uphill, against their electrochemical potential [3]. Secondary active transporters use energy stored within their electrochemical gradient which can be released as ions move down their gradients. Examples of these include cotransporters; proteins that carry one ion against and another along its electrochemical gradient simultaneously. Symporters are proteins that transport two ions of opposite charge in different directions whilst antiporters carry two ions of the same charge and move them in opposite directions. This thesis focuses primarily on ion channels and structurally similar nanopores.

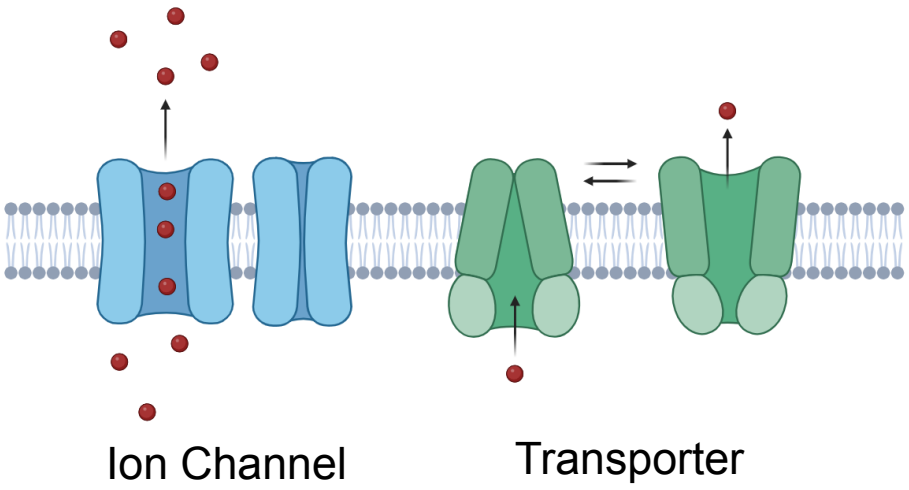

Figure 1.1 | Molecular Transport Through Biological Membranes Biological membrane proteins facilitate the transport of solutes (represented by red spheres) across membranes. Ion channels enable the diffusion of selected ionic species down their electrochemical gradients. Transporters utilise energy to transport ions against their electrochemical gradient. Typically, this energy comes from coupling to other ionic gradients or directly from ATP hydrolysis. Figure created with *BioRender.com*.

Ion channels, the main class of transmembrane proteins investigated in this thesis, usually form water-filled pores within cell membranes to facilitate and regulate the passage of specific ions across the membrane. They are often composed of multiple different subunits or domains that typically surround a central ion-conducting pore [4]. Unlike active transporters, they enable the rapid and passive diffusion of ions thermodynamically driven by the existing electrochemical gradient without requiring an external source of energy [3]. Ion channels achieve higher ion transport rates of $\sim 1 \times 10^9$ s^{-1}, compared to that of transporters which only reach rates of $\sim 1 \times 10^3$ s^{-1} [3, 5]. These high conductance rates can be estimated by considering the treatment of membrane embedded ion channel systems as electrical circuits; the membranes can be understood to act as parallel-plate capacitors that are capable of storing charge through creating a potential difference. In 1889, Nernst published the famous equation relating voltage to the ratio of ion concentrations which proved to be instrumental to the understanding of equilibrium potentials. For a channel in an open, conductive state the electrochemical potential in equilibrium can be estimated for a specific ion species X, from the Nernst equation,

$$\Delta E_X = E_1 - E_2 = \frac{RT}{zF} \ln \frac{[X]_{out}}{[X]_{in}} \tag{1.1}$$

where E_X is the electrochemical potential for the ion, $[X_{out}]$ and $[X_{in}]$ represent the concentrations of the ion species outside and inside the cell respectively, R is the Boltzmann gas constant, T is the temperature, F represents Faraday's constant and z denotes the charge of the ion species [5–7].

High rates of electrodiffusion through ion channels are exploited by physiological processes that rely on rapid electrical signalling. Ion channels play fundamental roles in enabling living organisms to perform various essential tasks from the propagation of nerve impulses for muscle contraction to overall cellular homeostasis. These processes

rely on the correct functioning of channels or otherwise channel malfunction and dysregulation results in numerous disease states known as *channelopathies* [8]. These include sleep apnea [9], cystic fibrosis [10], cardiac arrhythmias including long QT syndrome [11] and epilepsy [12]. Consequently, ion channels make suitable therapeutic targets; they constitute the second largest family of drug targets after G-coupled protein receptors (GPCRs) and are responsible for 18% of existing FDA-approved small molecule drugs [13, 14].

1.1.1 Biophysical Properties and Physiological Roles

Channel functionality is determined by a combination of factors. Ion channels can *activate* by undergoing conformational rearrangements to transition from closed (impermeable) to open (permeable) states to allow ions to pass through the channel pore. These transitions may correspond to changes in the physical dimensions of the pore and may be energetically driven [15]. The process of channel activation can be triggered through a range of stimuli such as the binding of a ligand to an allosteric site, changes in membrane potential or mechanical tension depending on their physiological functions. The regulation of the opening and closing of ion channels in response to specific internal or external stimuli is known as *gating* [2, 5].

Once activated, open channel pores have the ability to discriminate with high fidelity, allowing the passage of a restricted class of ions to permeate whilst simultaneously excluding others. This property is known as *selectivity*. To achieve selective permeability of a given species, specific binding sites within the pore lining of the ion conduction pathway allow a protein to sense ions, permitting only the desired ones to pass [16]. Most channels can discriminate between anions and cations however some have evolved to differentiate between ions of the same charge and of similar size e.g. Na^+ and K^+ ions. For example, K^+ channels contain a narrow constriction known as the *selectivity filter*

composed of highly conserved amino acids. Here, backbone carbonyls form a number of sites that constitute the selectivity filter in which a passing ion can interact with the protein directly [6, 17, 18]. Remarkably, permeation rates of K^+ channels are $\sim$ 100 times faster than Na^+ ions in spite of the smaller size of Na^+ suggesting that selectivity for K^+ is not achieved solely through steric occlusion [2, 19]. More recently, crystal structures of a double barrelled fluoride channel (Fluc) exhibited exceedingly high specificity ($> 10^4$) for F^- over Cl^-, which was proposed to arise from anion coordination with quadrupolar edges of conserved phenylalanine rings [20].

The selectivity and gating mechanisms of cation channels have been vastly more studied relative to anion channels, which is typically due to the weaker selectivity of anion channels (see Section 1.1.3) [21]. Whilst progress has been made on gaining a firm understanding of the general principles of selectivity and gating, work still remains to be done to elucidate the molecular details, in particular within anion selective channels and transporters.

1.1.2 From Structure to Function

The field of structural biology focuses on elucidating the relationship between three-dimensional protein structures, their assembly, and underlying molecular mechanisms of protein function. The development of powerful structural techniques such as X-ray crystallography, nuclear magnetic resonance (NMR) and cryogenic electron microscopy (cryo-EM) have revolutionised the field [22]. To date, the total number of unique ion channel structures deposited in the Protein Data Bank (PDB) [23] has experienced an unprecedented growth in the number of ion channel structures being resolved (Figure 1.2). The first high-resolution structures of ion channels were resolved by X-ray crystallography, including the ground-breaking study of the K^+ channel, KcsA in 1998 [24]. However, the protein crystallisation process remains the major bottleneck of structural

determination by this technique due to the detergent required to extract proteins from their host cell membranes [22, 25].

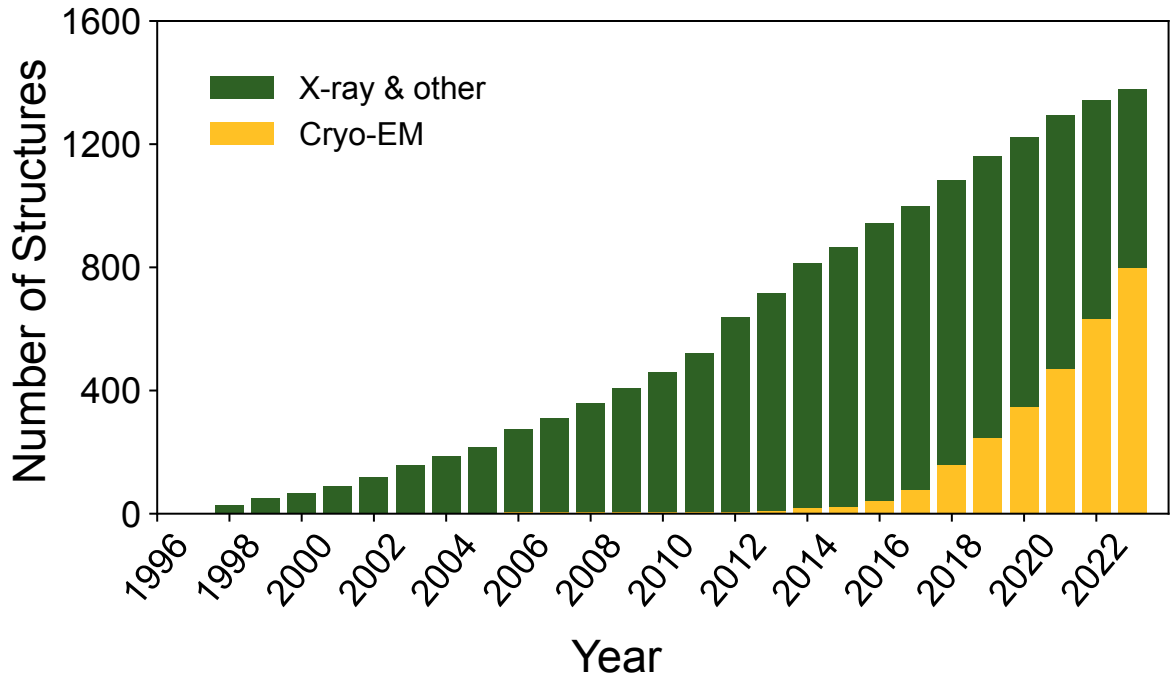

Figure 1.2 | Ion Channels in the PDB The cumulative number of ion channel structures deposited in the Protein Data Bank (PDB) (`https://www.rcsb.org/`) retrieved on the 1st of October 2023. The majority of these structures were solved using X-ray diffraction (green) however there has been a rapid growth in the use of cryo-EM (yellow) in more recent years.

In the last decade, single-particle cryo-EM has provided significant breakthroughs in reaching near atomic resolution as crystallisation is not required. This has uncovered a wealth of structures with $\sim$ 37% of structures in the PDB being determined using cryo-EM to date (Figure 1.2). Recent technological improvements in sample treatment, grid preparation, microscopic hardware and image processing have made accessing higher resolution structures better than 0.3 nm achievable [26, 27]. However, several caveats remain including radiation damage resulting in lower signal-to-noise ratios, protein denaturation and beam-induced sample movement [28].

The resulting experimentally determined structures represent a fraction of the billions of known protein sequences. The emergence of ground-breaking computational methods are essential to addressing this gap. Computational structure prediction methods such as AlphaFold, developed by DeepMind, utilises deep learning to predict the three-dimensional structures of proteins to a high degree of accuracy based solely from their amino acid sequence [29]. Such methods provide valuable insights into protein folding patterns and contribute towards understanding protein function however they currently may not be able to capture transient states of a protein [30].

In conjunction with structural studies, electrophysiology experiments, namely the patch clamp method developed by Sakmann and Neher [31, 32], are able to monitor the electrical activity of a single channel [7]. Together, these experimental techniques are capable of capturing a range of conformational states; accessing not only the open and closed states but also transient intermediate states. Static structures may imply intrinsic properties about a channel's functional state however the structure itself is not able to directly explain mechanisms of channel gating, ion selectivity or modulation mechanisms. Leaning on the availability of high-resolution structures, computational approaches such as molecular dynamics (MD) and free energy calculations, the main techniques used in this thesis, offer powerful avenues to observe and quantify molecular mechanisms and make testable predictions for channel function and annotation [3, 33].

1.1.3 Anion Selective Ion Channels, Transporters and Nanopores

The underlying selectivity mechanisms within anion channels remain poorly understood relative to their cation counterparts. The anion of most biological significance is Cl^-. Unlike cation channels that have high affinity and selectivity, Cl^--selective channels are comparably less selective and are permeable to other anions for example, the cystic fibrosis transmembrane regulator (CFTR) channel is not only permeable to Cl^- but

permits the passage of HCO^{3-} ions too [34]. The chloride channel (CLC) family, involved in muscle excitability and transmembrane electrical potentials, select for Cl^- over other anions in the sequence $Cl^- > Br^- > NO_3^- > I^-$. CLC channels provide an opportunity to explore whether anion channels use analogous structural features for ion selectivity compared with cation channels [21, 35, 36].

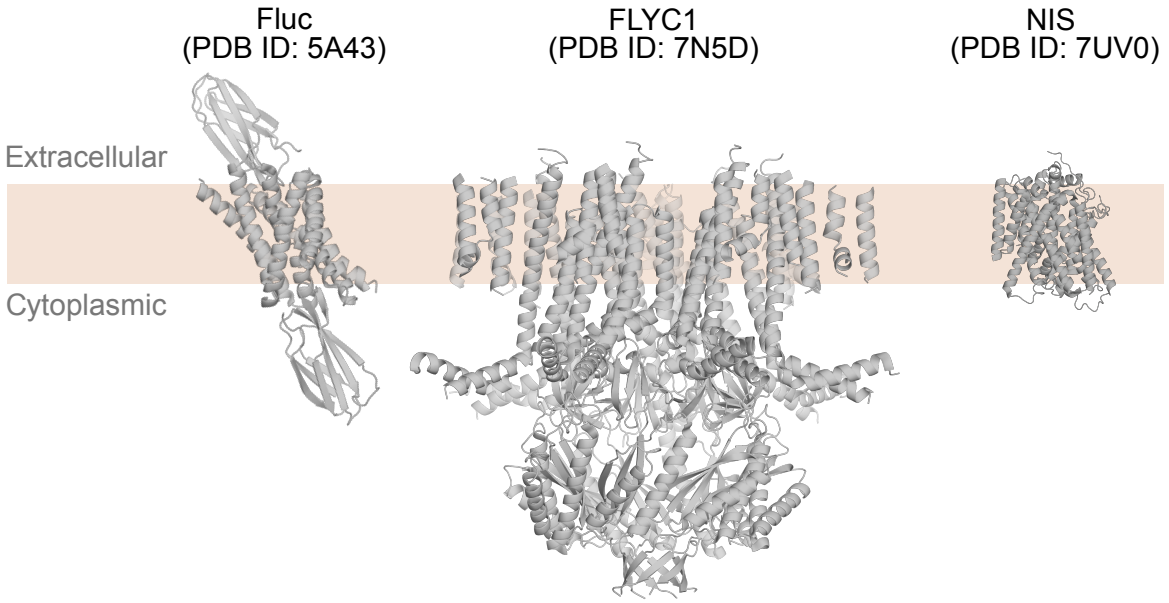

Figure 1.3 | Example Anion Channel Structures Examples of anion selective channels and transporters (grey ribbons, visualised in the plane of the membrane using PyMOL (`https://pymol.org/2/`)). The F^- selective Fluc channel, Cl^- selective Flycatcher 1 (FLYC1) channel and the I^- selective Na^+/I^- symporter (NIS) with Protein Data Bank (PDB) IDs 5A43, 7N5D and 7UV0 are shown respectively. Structures are shown approximately with respect to a membrane (orange) and oriented with the intracellular/cytoplasmic side at the bottom.

With the emergence of new structures for anion selective channels, there is a growing interest in elucidating the key aspects influencing anion selectivity and permeation mechanisms which can be learned through examining structures. Some notable anion-selective membrane protein structures solved within the past decade include the Na^+/I^- symporter (NIS) [37] which transports I^- but not Cl^- even though the extracellular Cl^- concentration is over 10^5 times that of I^- (Figure 1.3). The Fluc F^-

channels have exceedingly high specificity (> 10^4) for F^- over Cl^- which is proposed to arise from very narrow pores and unusual anion coordination through exploiting anion-quadrupolar interactions of conserved phenylalanines (Figure 1.3) [38–40]. The light-driven Cl^--pumping rhodopsin contains two defined Cl^- binding sites, of which the cytosolic site is predominately composed of hydrophobic residues [41]. Other Cl^- channels with high Cl^- affinity include the transmembrane protein 16A (TMEM16A) channel [42, 43], the mechanosensitive Flycatcher 1 (FLYC1) channel [44] and the family of bestrophin channels [45–47]. Further to these are some pentameric ligand-gated ion channels (pLGICs), namely glycine receptors (GlyRs)[48, 49] and $GABA_A$ receptors [50, 51], where multiple structures are known. Due to the structural diversity across anion selective channels and nanopores, it is difficult to draw patterns that universally explain how proteins discriminate between ions of the same charge and similar size or how they are modulated.

1.1.4 Model Nanopores

Membrane embedded nanopores are not limited to biological function. Nanopores in technology span a wide range of applications including water desalination [52], DNA sequencing and biosensing, all of which exploit their ability to conduct and often differentiate between charged species [16, 40, 53]. Technological and synthetic nanopore designs have drawn inspiration from aspects of biological ion channels. Conversely, biological nanopores contain thousands of atoms where molecular interactions are complex in nature; therefore model biomimetic nanopores offer opportunities to reduce a system down to a minimal working model to examine fundamental biophysical mechanisms governing ion selectivity, gating and transport [54]. This can further inform the design of novel nanoscale devices that are tuned for a specific function.

Examples of existing artificial biomimetic nanopores include porous organic cages that act as synthetic water channels [55], replicating the high water permeability of aquaporins and providing alternative channel morphologies for water desalination applications. Self-assembling cyclic peptide nanotubes have provided insight into the influence of internal channel geometry and the chemical properties of pore-linings on ion and water transport, suggesting possible criteria for selectivity of cations over anions [56, 57].

Carbon nanotubes (CNTs) are a well-studied choice of model nanopore. CNTs are particularly attractive as structural templates because they can imitate many key characteristics of biological pores including high transport efficiency, tunable pore sizes, surface functionalisation, and well-defined hydrophobic interiors. The relevant properties of CNT nanopores have been extensively studied, both experimentally and with simulations (see Section 3.1 for further details) [52, 58–61].

1.2 Properties of Water and Halide Ions

1.2.1 Water & Hydrophobic Gating

Functional properties of ion channels and nanopores are dependent on the behaviour of water at the nanoscale. Water under nanoconfinement exhibits anomalous physico-chemical properties compared to bulk liquid. These properties are sensitive to the chemical surface composition; in particular at hydrophobic surfaces, water experiences a repelling effect [62, 63]. The shape, pore dimensions as well as the nature of the pore-lining atoms are major determinants of channel function. Within sufficiently narrow nanopore containing a hydrophobic inner surface, water may undergo stochastic transitions between liquid and vapour states. This is referred to as a *hydrophobic gate* or sometimes known as a *vapour lock*. Although a pore may be wide enough to allow for a

specific ion to permeate, in a hydrophobic gate the pore may spontaneously dewet, presenting an energetic barrier to ion permeation without steric occlusion [64–66]. Simulations of open state hydrophobic pores have revealed pores with internal radius $R \geq 0.7$ nm are fully hydrated, whereas narrow pores with $R \leq 0.4$ nm are considered completely dewetted and non-conductive. Between these radii water exhibits liquid-vapour transitions [65]. Hydrophobic gates can be opened in many ways by e.g. increasing the pore radius, changing the surface hydrophobicity through allosteric transitions or applying an external electric field to induce pore wetting [67].

Measurements of the physical dimensions of ion channels can be routinely performed using the HOLE [68] and the Channel Annotation Package (CHAP) [69] programmes. Based on pore radius and local hydrophobicity, CHAP is able to rapidly identify hydrophobic gates in the permeation pathway and inform functional annotation of novel ion channel structures [69, 70].

1.2.2 Halide Ions

The halide ions: F^-, Cl^-, Br^- and I^- derived from halogens, play vital roles in biological processes. In a biological context, halide ions form components of buffers to maintain the ionic strength of solution which is especially relevant for Cl^- ions. In 1888, Franz Hofmeister formulated the Hofmeister series: $F^- > Cl^- > Br^- > I^-$. The series is based on the solubility of proteins and it has since been invoked in a plethora of effects beyond this original concept [40, 71].

Halide ions can be found to abide by the inverse Hofmeister series at a variety of hydrophobic interfaces ranging from simple planar interfaces to more complex protein surfaces. Extensive studies have been carried out on electrolyte solutions at planar

hydrophobic interfaces such as the aqueous/air [72, 73] or aqueous/oil interface, whereby intermolecular interactions in this region characterise the phenomenon of adsorption. Adsorption results in a difference in the concentration of an ionic species in the interfacial region relative to bulk solution. The force that drives this surface ion adsorption arises from the cavitational energy whereby the presence of ions disturbs the hydrogen bonding network of water with an associated energetic cost [73, 74]. This interfacial region of varying concentration defines the *Gibbs surface excess* (Figure 1.4 B), Γ_i, which can be related to the surface tension γ and the chemical potential μ_i via the Gibbs adsorption isotherm for two adjoining phases [75, 76],

$$-d\gamma = \Gamma_\alpha d\mu_\alpha + \Gamma_\beta d\mu_\beta \tag{1.2}$$

The Gibbs surface excess, Γ_i is given by

$$\Gamma_i = n_i^\sigma / A_s \tag{1.3}$$

where n_i^σ is the surface excess amount of component i and A_s is the area of the interface. The theoretical picture of an interface consists of an idealised system where two substances meet, known as the *Gibbs dividing surface* (GDS) (Figure 1.4 A). At this location, the surface excess equals zero ($\Gamma_\alpha = 0$) and the Gibbs adsorption isotherm is reduced to $-d\gamma = \Gamma_\beta d\mu_\beta$. However, recently MD simulations implementing explicitly polarisable forcefields [77, 78] or effectively modelling polarisation [79] have revealed a propensity for halide ions to associate at an aqueous/hydrophobic interface following the inverse Hofmeister series, i.e. F^- < Cl^- < Br^- < I^-. This order is directly correlated to the polarisability of the anions such that the smaller "hard" F^- ions retain their hydration shells and therefore are excluded from the interface. On the contrary, the larger "soft" polarisable halides, which have more flexible hydration shells, exhibit an increasing affinity towards the interface with an increase in ion size [80, 81]. The anisotropy of such

interfaces induces a dipole in polarisable ions which are otherwise not observed in bulk solution. Interactions between the induced dipole and surrounding water molecules compensate for the partial loss of ion hydration as it adsorbs to the interface where the ion is stabilised (Figure 1.4 C) [82, 83].

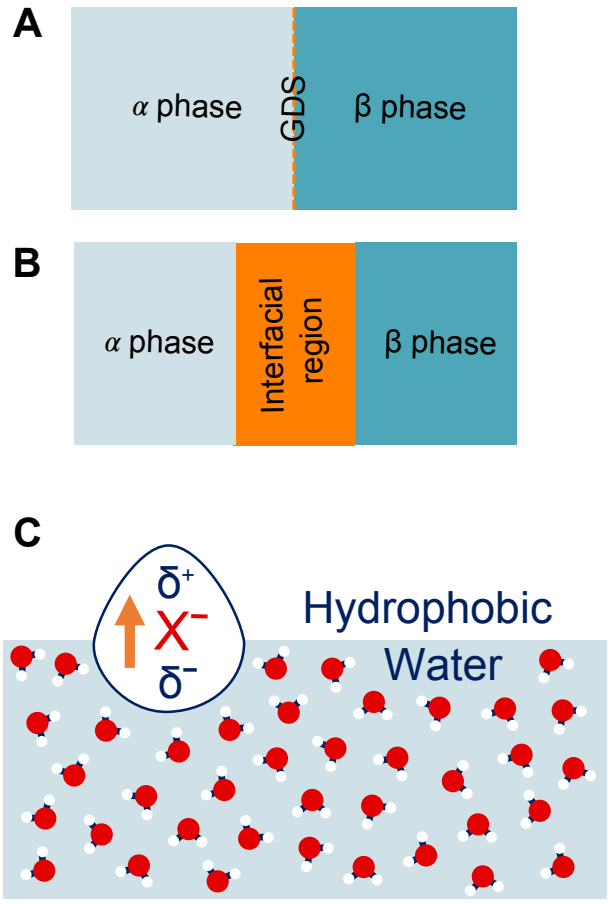

Figure 1.4 | Definition of an Interface A The theoretical picture of an interface: the orange dashed line represents the Gibbs Dividing Surface (GDS) which corresponds to the idealised geometrical surface where two bulk phases, α and β meet. **B** In a realistic system, there is an interfacial region where the concentration of a solute differs from bulk known as the surface excess. **C** Schematic diagram of an induced dipole in a polarisable halide ion, X^-, adsorbed to a hydrophobic/water interface.

Direct experimental observations of ion behaviours in interfacial regions remains a challenge. Nonetheless, interfacial ion properties can be inferred from spectroscopic experiments that are capable of probing the surface compositions of electrolyte solutions which have shown to be supportive of ion specific effects in accordance with the

Hofmeister series [84–87]. Simulation studies of ion solvation within small aqueous droplets and clusters using polarisable models (see Section 2.2) have demonstrated that halide movement towards the surface was found to be enthalpically driven by the organisation of water around partially dehydrated ions which allow for more favourable water-water interactions compared to bulk [83, 88].

1.3 Molecular Dynamics Simulations for Channels, Transporters and Nanopores

1.3.1 Computational Methods for the Study of Channels and Transporters

For several decades, computational modelling has been a useful tool for the study of ion channels and transporters to complement and overcome the shortcomings of experimental techniques, offering opportunities to bridge the gap between structural and functional data [89]. Continuum approaches where protein, water and membrane are modelled as continuous dielectric media and ions are modelled by a continuous density distribution, have been used to explore solute transport on longer timescales and can provide estimates for relevant observables characterising transport. However, these approaches neglect the granular nature concerning ion channels [90].

More recently, atomistic MD simulations offer particle-based approaches that capture time-dependent atomic motions and fluctuations, providing all the biochemical information necessary to study the molecular bases of channel gating, ion selectivity, hydrophobic control, etc. [3, 33]. *In silico* mutagenesis studies have enabled the influence of (point) mutations associated with known channelopathies to be explored over a range of ion channels [91]. *In silico* electrophysiology can be achieved through either the application of an external electric field or creating a charge imbalance over a double-membrane configuration. Modelled in this way, ion channel permeation can be

estimated and directly related to patch clamp electrophysiology experiments [33, 92]. The scope of MD simulations can be greatly expanded by enhanced sampling techniques such as free energy perturbations (FEP), potential of mean force (PMF) calculations and metadynamics approaches to name a few, which allow a system to cross large energetic barriers to sample regions of conformational space that would otherwise not be accessible within a given timescale [90][3]. Collective variable approaches help reduce the complexity of a system e.g. by applying geometric criteria, to enable enhanced sampling methods which are useful tools to probe ion permeation and channel inactivation [93]. Alchemical methods such as FEPs are valuable in understanding the basis for ion selectivity, however due to thermodynamic selectivity, FEPs may be insufficient to explain selectivity in all ion channels [17, 94].

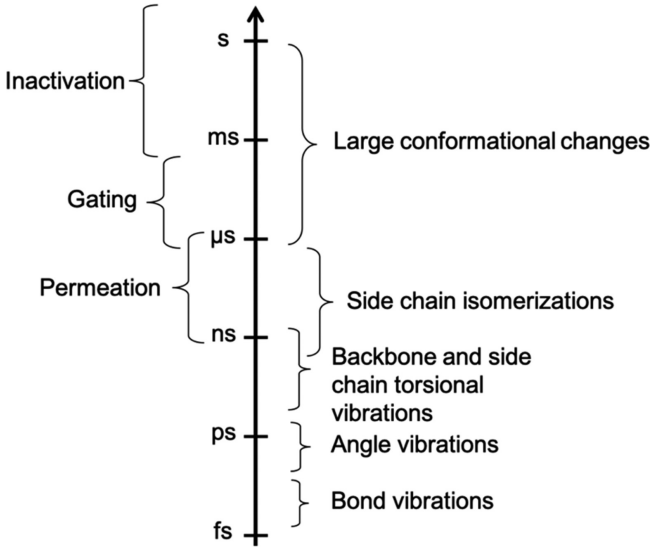

Figure 1.5 | Timescale Ranges of Ion Channels Timescale ranges associated with ion channel activity (left) relating to structural fluctuations (right). Reprinted (adapted) with permission from [33]. Copyright 2023 American Chemical Society.

There are many challenges that still remain in MD simulations for channels and transporters. The time scales for channel gating involving large conformational changes may span a wider range of timescales that are currently not feasible to run on most standard computer clusters (Figure 1.5) [33]. This has implications for e.g. observing

multi-ion translocation events which have demonstrated to be important in studies of the selectivity filter of KcsA K^+ channel [95, 96]. Additionally, the ability to move between different time and length scales e.g. from the molecular level to the macroscopic scale will become necessary and progress has already been made in this area [91]. It is important to keep in mind that results can be sensitive to forcefield inaccuracies or limitations of the forcefield [90]. In particular, this thesis is concerned with the modelling of electronic polarisation and its effects of ion selectivity and permeation mechanisms.

1.3.2 Electronic Polarisation in Biomolecular Simulations

Electronic polarisation refers to the redistribution of charges within an atom or molecule as a result of an external electric field. Polarisation can be modelled in MD simulations implicitly and/or explicitly (see Section 2.2 for further details), the latter being significantly more computationally intensive than the former. Therefore, the choice to explicitly model polarisation in a simulation is system dependent and limited by the availability of resources capable of performing such calculations. Nonetheless, advances in computing hardware and improved simulation methods have recently made polarisable forcefields more accessible such that they are being increasingly applied to biomolecular systems [97].

To date, explicitly modelling polarisation has provided new insight into a number of ion channel systems. Studies of ion transport across the short hydrophobic channel, gramicidin A [98, 99] have revealed lowered energetic barriers to permeation for Na^+ and K^+ and have further reproduced experimental single channel K^+ conductance rates using polarisable CHARMM Drude forcefield simulations (see Section 2.2).

Simplified models of full protein-embedded systems offer opportunities to explore the interactions of isolated regions of interest in more detail and making them more computationally feasible to model with polarisability. For example, interactions of permeating water and Cl^- ions have been explored within reduced models of the 5-HT3 serotonin receptor channel that include only the transmembrane pore-lining helices of the parent protein [100]. Inclusion of induced polarisation resulted in a partial loss of first hydration shell of Cl^- to interact with the hydrophobic pore-lining while it permeates the pore.

Methods such as the Electronic Continuum Correction (ECC) [101] (see Section 2.2.2.1) and Non-bonded FIX (NBFIX) [102] (see Section 2.2.2.3) are approaches that include polarisation effects in a mean-field way with no additional computational overhead relative to additive forcefields. The permeation mechanism of K^+ through the selectivity filter of the K^+ channel, KcsA, has been explored using explicit polarisation (AMOEBA) and compared with a scaled charge method (ECC) to reveal similar ion occupancy and deviating only in relative stabilities between configurations [103]. The NBFIX method has also shown potential for ion channel applications. Simulations of the CLC-ec1 channel implementing NBFIX showed no Cl^- dissociation from a binding site over $\sim 4\ \mu s$ whereas with an additive forcefield, the ion was seen to dissociate without reoccupying the site again [104].

Applications of polarisable forcefields have provided many new insights and generally agree better with experimental findings due to improved physical descriptions. They are steadily becoming more popular and more accessible as systems require accurate treatment of electrostatics to yield realistic dynamics [97].

1.4 Thesis Outline and Objectives

An increasing number of anion selective membrane transport protein structures are beginning to emerge. Current MD simulation studies largely employ non-polarisable forcefields that do not explicitly capture the effects of induced polarisation; a factor that could prove to be essential for accurate characterisation of anion interactions. This thesis primarily seeks to explore the influence of modelling polarisation implicitly or explicitly for more physically realistic anion interactions and to observe the resulting anion behaviour within ion channels and biologically relevant nanopores. The presentation of results from these studies is structured as follows.

Chapter 2 provides an overview of the underlying theory of the computational methods employed. The focus lies on atomistic representations in MD simulations and a review of next-generation forcefields that either explicitly or implicitly model induced polarisation. An introduction to potential of mean force calculations is also detailed.

A recurring theme in this thesis is the question of how does modelling polarisation affect the behaviour of water and ions within nanoconfined hydrophobic environments? In Chapter 3, the ground work is laid by first evaluating the applicability of the ECC method to a simple aqueous/hydrophobic interface. Once validated, the application of ECC to an *in silico* model biomimetic nanopore, designed to mimic the general charge distributions of pLGICs, was used to probe ion and water interactions within a confined hydrophobic environment and the associated energetic costs of permeation were considered.

In Chapter 4, we move on to study a more biologically relevant protein system. Simulations of a truncated Cl^--pumping rhodopsin [41], including only the hydrophobic Cl^- binding site of interest and some surrounding residues. A systematic comparison of

the ion and water dynamics modelled using non-polarisable, scaled-charge and full polarisable forcefields were explored. We evaluate the energetics and binding dynamics of Cl^- ions in the binding site and identify a second metastable binding site governed by interactions between Cl^-, water and some nearby hydrophobic residues. Critically, we demonstrate that only by including polarisation in the molecular model are these complex anion-hydrophobic interactions captured and binding is observed.

Chapter 5 details the exploration of in-pore interactions of Cl^- and water within the hydrophobic neck of the human bestrophin 1 (hBest1) channel [105]. Implementation of the fully polarisable forcefield, AMOEBA provided insight into the potential roles that edge-wise anion-π interactions mediated by phenylalanine residues play in Cl^- selectivity and permeation. It is observed that Cl^- partially dehydrates to interact directly with the hydrophobic contact through "hydrophobic solvation". In addition, the functional advantages of pore asymmetry in the neck were explored. Furthermore, this study demonstrates how polarisable MD simulations can aid the assignment of ambiguous cryo-EM densities which have otherwise been labelled as putative waters. A discussion of limitations that current non-polarisable forcefields pose with emphasis placed on the importance of physically accurate models in quantitative investigations involving weakly polarisable anions completes this chapter.

Finally, this thesis concludes with a summary of key findings and a discussion of future research directions in Chapter 6.

"Those who do not want to imitate anything, produce nothing."

— Salvador Dali

2 Theory and Methods

Contents

2.1 Classical Molecular Dynamics 21
2.1.1 Pairwise Additive Forcefields 21
2.1.1.1 Bonded Interactions 22
2.1.1.2 Non-Bonded Interactions 24
2.1.1.3 Forcefields 26
2.1.1.4 Water Models 28
2.1.1.5 Ion Parameters 30
2.1.2 Molecular Dynamics Algorithms, Settings and Parameters 31
2.1.2.1 Energy Minimisation 31
2.1.2.2 Integration Algorithms 31
2.1.2.3 Pressure and Temperature Control 33
2.1.2.4 The Particle Mesh Ewald Method 38
2.1.2.5 Periodic Boundaries Conditions and Long Range Forces 38
2.2 Polarisation Effects in Classical Molecular Dynamics 39
2.2.1 Polarisable Atomic Multipoles 41
2.2.1.1 The AMOEBA Forcefield 41
2.2.1.2 MD Algorithms for Polarisable Atomic Multipoles 45
2.2.2 Charge Scaling and Mean-field Methods 46
2.2.2.1 Electronic Continuum Correction (ECC) 46
2.2.2.2 The prosECCo Forcefield 48
2.2.2.3 NBFIX 49
2.3 Non-equilibrium Methods 49
2.3.1 Potential of Mean Force Calculations 49

2.1 Classical Molecular Dynamics

Molecular Dynamics (MD) is a computational technique that enables us to simulate the time-dependent behaviour of molecular systems. The physical movements of interacting of atoms and molecules are realised through iterating Newton's equations of motion for all particles in the system.

MD simulations aim to simplify the complex quantum mechanical (QM) nature of atomic interactions. QM calculations are associated with an increased computational expense which restricts the practical applications to short timescales and small systems or specific regions of larger systems using a hybrid quantum-mechanics/molecular-mechanics (QM/MM) methods [106].

In classical MD, the Born-Oppenheimer approximation makes it possible for the wave functions of atomic nuclei and electrons in a molecule to be treated separately on the basis that nuclei are much heavier than electrons and therefore move on comparatively slower timescales [107]. This fundamental assumption enables the simulation of larger systems and/or ability to achieve longer and biologically relevant timescales by approximating the potential energy of the system at each timestep to be a function of nuclear coordinates and neglecting to explicitly model electrons.

2.1.1 Pairwise Additive Forcefields

In molecular mechanics, the underlying physical model is called a *forcefield* (FF). This refers to the collection of mathematical equations that describe the inter- and intra-molecular interactions for a system containing N atoms with Cartesian coordinates $\mathbf{r}^N$ associated with their nuclei positions,

$$\mathbf{r}^N = \{x^1, y^1, z^1, ..., x^N, y^N, z^N\} \tag{2.1}$$

The potential energy function $U(\mathbf{r}^N)$ underlying pairwise additive forcefields are generally calculated as the summation of bonded (intramolecular) and non-bonded (intermolecular) terms,

$$U(\mathbf{r}^N) = U_{bonded} + U_{non-bonded} \tag{2.2}$$

Pairwise additive forcefields (i.e. non-polarisable) refer to the energy due to the interaction between two non-bonded atoms being independent of the position of all other atoms. They are not exclusively pair-interactions, as we will see below in Section 2.1.1.1, they can be 3-body interactions e.g. bond angle bending or even 4-body e.g. dihedral angle interactions. On the contrary, polarisable forcefields discussed in Section 2.2.1, facilitate many-body interactions which give rise to a more diverse range of physical effects however at a significantly greater computational expense [108].

2.1.1.1 Bonded Interactions

The first term of the potential energy function (Equation 2.2) represents the energy contributions associated with interactions between atoms that are covalently linked. These intramolecular interactions are deconstructed into terms for bond stretching, bond angle bending, torsional rotations and improper dihedral angles which are summed over pairs of bonded atoms to constitute U_{bonded},

$$\begin{aligned} U_{bonded} = &\frac{1}{2}\sum_{bonds} k_b(b-b_0)^2 + \frac{1}{2}\sum_{angles} k_\theta(\theta-\theta_0)^2 \\ &+ \frac{1}{2}\sum_{dihedrals} k_\omega(1+cos(n\omega-\phi)) + \frac{1}{2}\sum_{improper} k_\psi(\psi-\psi_0)^2) \end{aligned} \tag{2.3}$$

The term for bond stretching is modelled as a harmonic potential and therefore the associated energy varies with the square of the displacement of the bond length b from its equilibrium position b_0 with force constant k_b, in accordance with Hooke's Law.

Similarly, a harmonic potential is used to model deviations in the bond angle between a triplet of atoms (θ) from their equilibrium positions (θ_0) where the corresponding force constant k_θ is generally less than k_b.

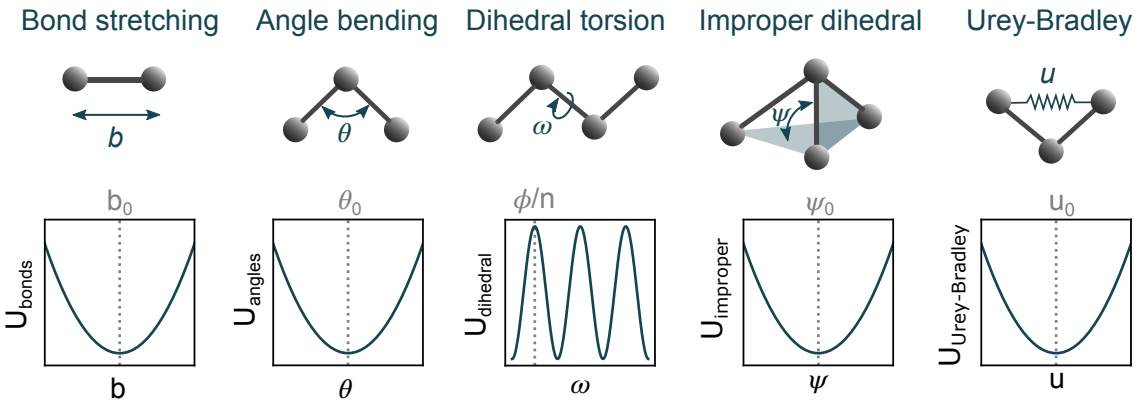

Figure 2.1 | Bonded Interactions The fundamental types of bonded interactions typically considered in molecular mechanics force fields. Schematic depictions of atoms and bonds involved are illustrated with their corresponding potential energy functional forms that together constitute the bonded contributions to the overall potential energy function.

The dihedral potential describes the torsional movement about a dihedral angle, ω i.e. the angle between two planes which pass through the same bond. The multiplicity n signifies the number of maxima or minima in the energy function and ϕ is the associated phase factor which determines their positions. The torsional force constant k_ω indicates the barrier height to rotation described by the cosine function. The improper dihedral characterises out of plane deviation (ψ) and is modelled with a harmonic potential. It is often used to preserve planarity for groups with flat geometries (e.g. aromatic rings) or to prevent significant changes to tetrahedral geometries from their reference planes (ψ_0). Together, these terms describe the bonded interactions in the potential energy function, U_{bonded} and are illustrated in Figure 2.1.

Whilst these terms describing bond lengths, angles and torsion energies provide a convenient approximation to QM behaviours, they do not account for the

interdependence of bonds, angles and torsions on each other. Therefore, although pairwise additive forcefields are capable of reproducing structural properties, they may be unable to reproduce vibrational frequencies which might be dependent on the coupling of multiple terms [107].

The Urey-Bradley potential, for example, is a harmonic potential that describes changes in energy with angle bending as a function of the distance u between two non-covalently bonded atoms that are separated by the two bonds forming the angle (i.e. 1-3 pair of atoms) with force constant k_u (Figure 2.1). The Urey-Bradley functional form is as follows,

$$U_{Urey-Bradley} = \frac{1}{2} \sum_{Urey-Bradley} k_u (u - u_0)^2 \tag{2.4}$$

Multi-term coupling is more commonly used in more advanced forcefields which will be discussed further in Section 2.2.

2.1.1.2 Non-Bonded Interactions

In pairwise additive forcefields, the intermolecular interactions can generally be decomposed into the sum of non-bonded electrostatic and van der Waals (vdW) interactions which are described by the Coulomb potential (Equation 2.8) and Lennard-Jones 12-6 (LJ) potential respectively (see Equation 2.5). Non-bonded interactions are defined as those between all pairs of atoms i and j which are separated by three or more intramolecular bonds or are from different molecules. The vdW term is commonly modelled through the LJ potential which takes the form,

$$U_{vdW} = \sum_{i=1}^{N-1} \sum_{j=i+j}^{N} \left\{ 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \right\} \tag{2.5}$$

where r_{ij} is the inter-nuclear distance between two atoms (i, j). The function contains two terms; a repulsion term describing two atoms in close proximity (typically $\sim$ < 0.3 nm), due to the Pauli exclusion principle given by $(\frac{\sigma_{ij}}{r_{ij}})^{12}$. The attractive term

originates from dispersion forces between instantaneous dipoles, which arise from fluctuations in electronic charge distributions and is denoted by $(\frac{\sigma_{ij}}{r_{ij}})^6$. σ_{ij} gives the distance between two atoms for which the potential is equal to zero and ϵ_{ij} is the depth of the potential well defining the interaction strength (Figure 2.2 A).

The values for σ_{ij} and ϵ_{ij} are often calculated using the Lorentz-Berthelot combination rules,

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \tag{2.6}$$

and

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \tag{2.7}$$

where σ_{ii}, σ_{jj} and ϵ_{ii}, ϵ_{jj} reflect the collision diameters and well-depths of individual atom types i and j given by the forcefield.

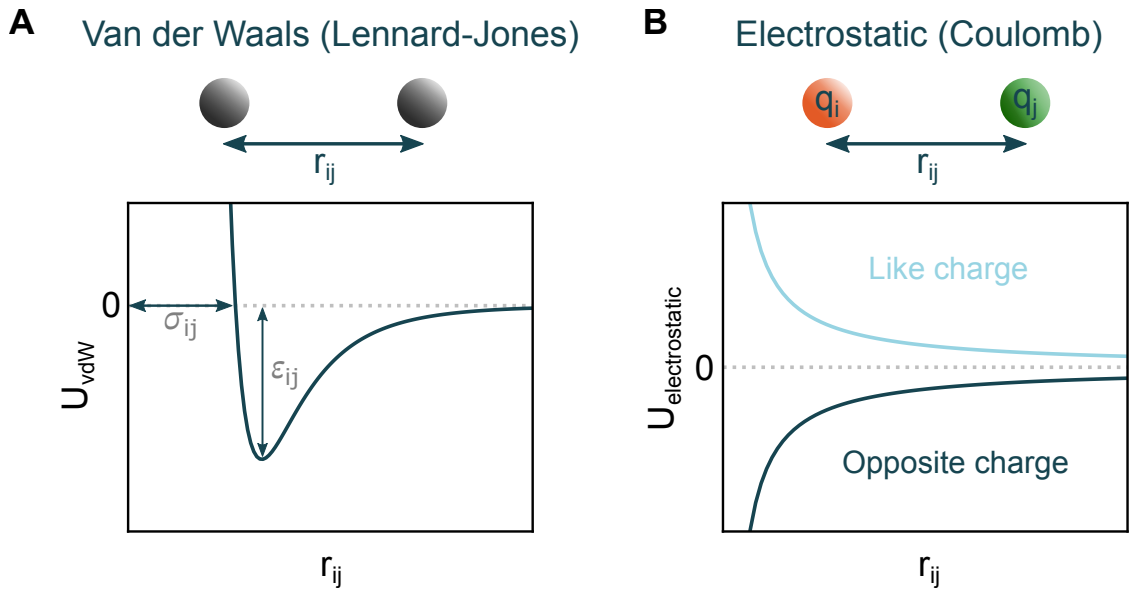

Figure 2.2 | Non-bonded Interactions The fundamental types of non-bonded interactions describing **A** van der Waals and **B** electrostatic interactions between atoms i and j in pairwise additive forcefields. Their corresponding potential energy functional forms are modelled through the Lennard-Jones and Coulomb potentials respectively.

Electrostatic interactions in pairwise additive forcefields are modelled by the Coulomb potential which takes the form,

$$U_{electrostatic} = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \tag{2.8}$$

where ϵ_0 is the vacuum permittivity, r_{ij} is the distance between atoms i and j with partial charges q_i and q_j (Figure 2.2 B). Electrostatic interactions are related inversely proportional to r_{ij}, so are appropriate for short-range interactions, however different treatment for long-range electrostatics is required which is where current challenges lie.

2.1.1.3 Forcefields

As discussed above, a forcefield consists of a set of equations and parameters describing the potential energy of a system. Each forcefield contains a list of atom types with associated parameters detailing bonds, angles, dihedrals, improper dihedrals and non-bonded interactions. In this thesis, we mainly use full atomistic representations which model each individual atom, including hydrogen atoms. These are encapsulated in forcefields such as the Chemistry at Harvard Macromolecular Mechanics (CHARMM) [109] and Optimised Potentials for Liquid Simulations All-Atom (OPLS/AA) [110] forcefields and are the additive forcefields implemented in this work. Other widely used additive forcefields include the Assisted Model Building with Energy Refinement (AMBER) [111] and Groningen Molecular Simulation (GROMOS) [112].

These forcefields are able to accurately reproduce many protein characteristics to some extent. However, due to simplifications in modelling the electrostatics, namely the lack of electronic polarisability, there are limitations on the achievable accuracy of some properties. For example, the relative orientation of aromatic moieties of e.g. benzene rings that have significant quadrupolar charge distributions [113] are affected by electrostatic simplifications. Coarse-grained (CG) models such as the Martini forcefield

[114] group together typically four heavy atoms to form “beads”, resulting in longer achievable timescales and smoothed energy landscapes. Additionally, much larger systems can be modelled with CG than with atomistic forcefields including virions [115], an entire cell [116] and complex membranes [117, 118].

The parameterisation of atomistic forcefields contain several steps in the workflow. Forcefields are derived by first defining a functional form (see Sections 2.1.1.1 and 2.1.1.2) which expresses the energy of the system as an analytical and differentiable function of particle coordinates [119, 120]. This is followed by selecting reference data to parameterise against which might be obtained from *ab initio* QM and density functional theory (DFT) calculations or be experimentally determined through nuclear magnetic resonance (NMR) spectroscopy, infrared spectroscopy, X-ray crystallography, etc. [121–123].

An objective function is then constructed to measure the levels of disagreement between the reference data and corresponding simulation results. Optimisation algorithms are applied to minimise the objective function through iterative procedures. Some of the more popular optimisation methods include simulated annealing, first-order iterative methods and second-order Newton-Raphson approaches which take a random guided search approach [124]. Alternatively, force-matching methods fit the potential energy surface derived from QM calculations [125]. To validate the refined set of parameters, properties of the system are computed that were not included in the parameterisation process and the reproducibility of properties is evaluated.

No universal forcefield exists to encompass modelling all molecular interactions accurately as each is parameterised with an emphasis on accurate optimisation of targeted properties. These might include system-specific molecular geometries, free

energies of hydration, liquid densities etc. Therefore, appropriate forcefield selection and usage for a given system need to be considered in order to achieve physically accurate dynamics [126].

2.1.1.4 Water Models

The choice of water model is an important consideration. The treatment of water in MD simulations can be broadly categorised into implicit solvent, coarse-grained and explicit atomistic models. Water models are often parameterised to describe physical properties analysed by experimental techniques including X-ray scattering and neutron diffraction which measure the structure of liquid water [127–129]. Generally, the radial pair distribution functions g_{OO}, g_{OH} and g_{HH}, together with the self-diffusion coefficient (2.3×10^{-9} m^{2}s^{-1}) are used to characterise the structure and dynamics of water at different temperatures [130]. The ability of water models to accurately reproduce experimental properties of bulk water such as the dielectric constant, surface tension and bulk density, provide an indication of the quality and transferability of the model [63].

The simplest and most widely used group of water models are the rigid non-polarisable models. These models represent atoms and properties of water as points that serve as LJ interaction sites. Typical rigid non-polarisable water models contain between three to seven interaction sites. There are two notable families of water models; the transferable intermolecular potential (TIP) and simple point charge (SPC) models which are based on the geometries of molecules within water vapour and ice respectively.

Within the TIP family, TIP3P is the 3-point model which consists of three interaction sites arranged in a chevron shape with a negative charge on the oxygen and a positive charge on each of the hydrogens (Figure 2.3 A). Moreover, TIP4P is the 4-point model in the TIP family and it includes a fourth massless interaction site (called the M-site) near the

oxygen atom located at the of the bisector of the H–O–H angle (Figure 2.3 B). This additional site was introduced to displace the charge away from the oxygen in an attempt to imitate the large quadrupole moment of water molecules. Reparameterisations of existing models are a common approach to try to achieve greater accuracy of specific properties. For example, Vega and Abascal adjusted the TIP4P model to improve several properties including enthalpies of vaporisation and the density of liquid water at ambient conditions to form the improved TIP4P/2005 model [131].

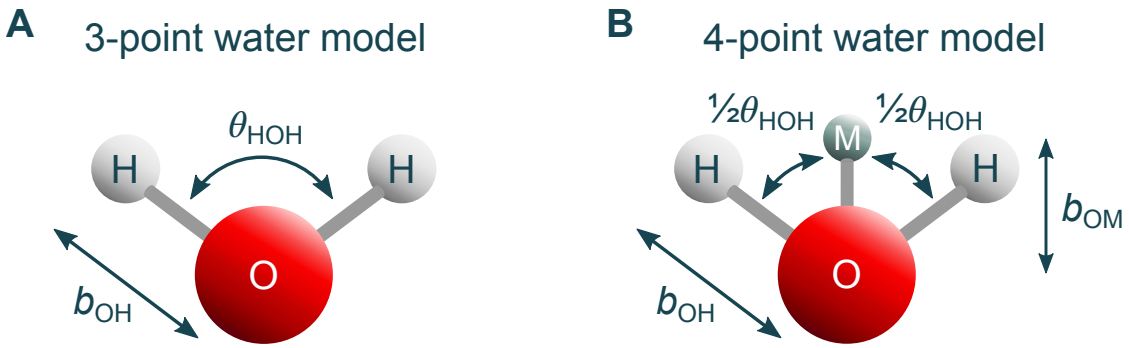

Figure 2.3 | Rigid Fixed-Charge Water Models Geometry of 3-point (**A**) and 4-point (**B**) rigid water models. 3-point models typically consist of three Coulombic interactions sites corresponding to the nuclei of the oxygen (red) and hydrogen (white) atoms. 4-point water models contain an additional, massless virtual site (grey) that is located at the H–O–H bisector and carries a negative charge.

The parallel 3-point model in the SPC family used in this thesis is the SPC/E water model ("extended simple point charge"). This is also a model improved upon the SPC water model which has been modified to include a correction to the self-energy interaction of the dipole moments leading to an improvement in the oxygen–oxygen radial distribution function (RDF), density and the diffusion constant at 300 K and 1 bar [132]. The choice of water model is key. Although models are transferable, it is important to note that they are often parameterised in conjunction with a given forcefield, e.g. the AMBER, CHARMM and OPLS forcefields are based on the TIP family and the SPC family are based on the GROMOS forcefields [107]. Consequently, the most suitable model is often the one

originally used in parameterisation of that forcefield. Many simulations in this thesis were carried out using the CHARMM36m forcefield combined with the mTIP3P water model which contains additional LJ sites on the hydrogen atoms [133]. A summary of the geometries and interaction parameters for water models used in this thesis can be found in Table 2.1.

		Water model				
Parameter	Units	TIP3P [134]	mTIP3P [135]	SPC/E [132]	TIP4P [136]	TIP4P/2005 [131]
LJ sites	–	3	3	3	4	4
b_{OH}	nm	0.09572	0.09572	0.1	0.09572	0.09572
b_{OM}	nm	–	–	–	0.0150	0.01594
θ_{HOH}	°	104.52	104.52	109.47	104.52	104.52
q_O	e	-0.834	-0.834	-0.848	–	–
q_H	e	0.417	0.417	0.4238	0.520	0.5564
q_M	e	–	–	–	-1.04	-1.1128
σ_O	nm	0.315066	0.315066	0.3166	0.3154	0.31589
σ_H	nm	–	0.0200	–	–	–
ϵ_O	kJ/mol	0.63627	0.63627	0.650	0.6480	0.7749
ϵ_H	kJ/mol	–	0.1925	–	–	–

Table 2.1 | Water Model Properties Geometry, electrostatic and vdW parameters for rigid, fixed-point water models employed in this thesis.

2.1.1.5 Ion Parameters

Ion interactions have fundamental roles in biomolecular simulations. A benchmark for accurate parameterisation of ions is the ability to reproduce accurate ion hydration structures which provide an indication of ion behaviours in solution. As with forcefield parameterisation, LJ parameters are fitted to reproduce structural and thermodynamic data from e.g. spectroscopic studies as well as *ab initio* calculations, but atomic charges in additive forcefields are assigned a priori as the formal charges of ions as if they were in the gas phase [137–140]. A degree of disagreement between calculated and measured values of hydration free energies, $\Delta G_{hydration}$ and/or water RDF and coordination

numbers for which ion parameters are targeted to reproduce give rise to different ion parameter sets. These are also water model and forcefield dependent [141]. Discrepancies in parameters may lead to simulated unphysical ion pairing, artificial ion clustering effects at high concentrations or the inability to reproduce specific ion effects [142]. Some of these properties arise from omission of electronic polarisation effects of the solvent [139] where methods to combat this in additive forcefields are discussed further in Section 2.2.2.1. Ion parameterisation remains an area of constant improvement.

2.1.2 Molecular Dynamics Algorithms, Settings and Parameters

2.1.2.1 Energy Minimisation

Prior to performing MD simulations, energy minimisation is executed to ensure any steric clashes or other instabilities are reduced for a more stable initial configuration. In this thesis, all energy minimisations are performed using the method of steepest descent which seeks the minimum potential energy through gradient optimisation. For each minimisation step, the force on each particle is calculated and then displaced according to this force such that if the potential in the new location is higher, the particle will be returned to its previous position and other directions are searched in subsequent steps. This is repeated for a number of steps until the force reaches a minimum or sufficient tolerance has been achieved [143].

2.1.2.2 Integration Algorithms

Having defined the potential energy function $U(\mathbf{r}_i)$ for a molecular system of interacting atoms, the net force $\mathbf{F}_i$ experienced by each atom i based on its position r_i is the negative partial derivative of $U(\mathbf{r}_i)$,

$$\mathbf{F}_i = -\frac{\partial U(\mathbf{r}_i)}{\partial \mathbf{r}_i} \tag{2.9}$$

Newton's equation of motion relates the instantaneous acceleration (the second derivative of the position w.r.t time) to the effecting force $\mathbf{F_i}$ and inversely proportional to the mass m_i by,

$$\frac{\partial^2 \mathbf{r_i}}{\partial t^2} = \frac{\mathbf{F_i}}{m_i} \tag{2.10}$$

thus, forming the basis of time-dependent MD simulations.

As the force on each atom is dependent on its position relative to all other interacting atoms, we cannot solve these equations analytically; instead, numerical approaches such as the finite difference method are used to obtain solutions. The integration is divided temporally into a large number of small finite steps separated by a fixed time interval δt. The initial conditions of a system provide the starting particle positions and velocities for each atom are assigned at random based on the Maxwell-Boltzmann distribution [143].

Per timestep, the accelerations of atoms are calculated from the forces; these are then combined with the positions and velocities of the current timestep (t) to compute those in the subsequent timestep $(t + \delta t)$. The cycle is iterated to progress the simulation.

By default, the GROMACS [144] simulation software uses the leap-frog algorithm to integrate equations of motion. To update both positions and velocities, the leap-frog algorithm uses positions $\mathbf{r_i}$ at time t and velocities $\mathbf{v_i}$ at half time intervals $(t - \frac{\delta t}{2})$. The forces $\mathbf{F}(t)$ on atoms, and hence the accelerations, are determined by the positions at time t. Together, they follow the relations below,

$$\mathbf{v_i}(t + \frac{\delta t}{2}) = \mathbf{v_i}(t - \frac{\delta t}{2}) + \delta t \mathbf{a_i}(t) \tag{2.11}$$

$$\mathbf{r_i}(t+\delta t) = \mathbf{r_i}(t) + \delta t \mathbf{v_i}(t+\frac{\delta t}{2}) \tag{2.12}$$

Other commonly used algorithms include the Verlet method which calculates both the velocities and positions at time t based on the second order Taylor expansion [145, 146].

It is important to select an appropriate timestep that is small enough to capture molecular motions of interest. For atomistic simulations, these might be the bond oscillations of a C–H bond at $\sim$ 10 fs, so an integration timestep on the order of femtoseconds (fs) would be suitable. However, if bond oscillations are not of interest, it is common practice to constrain bond lengths in order to employ larger timesteps. A variety of algorithms have been developed to apply such holonomic constraints to particles with defined coordinates. They are commonly applied through the SHAKE algorithm [147] which is an iterative method that incorporates the constraints into the equations of motion through solving a set of Lagrange multipliers. For the special case of rigid water molecules, the SETTLE method [148] solves the Lagrangian constraint equations analytically. A faster method is the non-iterative LINCS algorithm [149] that enforces constraints through a two-step procedure by first updating bond lengths and then correcting bond rotations.

2.1.2.3 Pressure and Temperature Control

Numerical integration of the equations of motion produces the microcanonical or NVE ensemble whereby the number of particles N, volume of the system V and total energy E of the system are conserved. However, biomolecular systems typically occur under the isothermal-isobaric or NPT ensemble where the pressure P and temperature T of the system are kept constant; to do so, simulations need to employ barostat and thermostat

algorithms respectively to maintain these conditions. There are three main thermostats and two types of barostats used in investigations in this thesis.

2.1.2.3.1 Berendsen Thermostat

In statistical mechanics, the temperature T of an unconstrained system can be estimated from the average kinetic energy of a system containing a sufficiently large number of particles by the relation,

$$K_{NVT} = \sum_{i=1}^{N} \frac{1}{2} m_i v_i^2 = \frac{3}{2} N k_B T \tag{2.13}$$

where k_B is the Boltzmann constant, m_i are the masses and v_i are the velocities of the ith particle.

One possible way to maintain a constant temperature is to rescale particle velocities. Weak-coupling methods such as the Berendsen thermostat couples a system to an external heat bath with a fixed temperature T_0 and then proportionally rescales the particle velocities such that particles which deviate from the specified temperature T_0 are corrected according to:

$$\frac{dT(t)}{dt} = \frac{T_0 - T(t)}{\tau_T} \tag{2.14}$$

where τ_T is the temperature coupling time constant which determines how tightly coupled the bath and system are. This gives an exponential decay of the system towards the desired temperature [143]. The Berendsen thermostat suppresses fluctuations in kinetic energy which means it does not generate the proper canonical ensemble however this error is much smaller for systems with a large number of particles, therefore the ensemble averages are not affected [145].

2.1.2.3.2 Velocity Rescaling Temperature Coupling

The velocity-rescale (V-rescale) method is very similar to the Berendsen thermostat however it contains an additional stochastic term that ensures the correct kinetic energy distribution according to:

$$dK = (K_0 - K)\frac{dt}{\tau_T} + 2\sqrt{\frac{KK_0}{N_f}}\frac{dW}{\sqrt{\tau_T}} \tag{2.15}$$

where N_f is the number of degrees of freedom of the system and dW is a Wiener process. V-rescale produces the correct canonical ensemble whilst sharing the same advantages as the Berendsen thermostat of first-order temperature deviations without oscillations [145, 150].

2.1.2.3.3 Andersen Thermostat

By the Andersen approach, a system is thermally coupled to a fictitious heat bath of a desired temperature. The Andersen thermostat maintains temperature by randomising the velocities of a subset of particles through extracting their velocities from a Maxwell Boltzmann distribution corresponding to the target temperature [151]. The randomisation of particles that suffer collisions is generated by calculating the time intervals between successive collisions of a particle with distribution governed by a Poisson process,

$$P(t) = ve^{-vt} \tag{2.16}$$

where v represents the frequency of stochastic collisions determining the strength of coupling to the heat bath. $P(t)dt$ is the probability that the next collision is between the interval [t, $t + \Delta t$] [152]. The algorithm is effective in avoiding ergodicity issues as energy is unable to flow between energetically decoupled components. However, the algorithm can decrease the kinetics of the system therefore making the Andersen thermostat unsuitable for calculating kinetics or transport properties [145].

2.1.2.3.4 Nosé-Hoover Thermostat

The Nosé-hoover thermostat [153, 154] is an extended-ensemble approach where the Hamiltonian contains an additional term corresponding to a thermal reservoir and friction. The frictional force is proportional to the product of a particle's velocity and a frictional parameter ξ which has its own momentum p_ξ and equation of motion.

$$\frac{d^2\mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i} - \frac{p_\xi}{Q}\frac{d\mathbf{r}_i}{dt} \tag{2.17}$$

The Hamiltonian of the extended system becomes

$$H_{Nosé} = \sum_{i=1}^{N} \frac{\mathbf{p_i}^2}{2m_i} + U(\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_N}) + \frac{p_s^2}{2Q} + gkT \ln s \tag{2.18}$$

where g is equivalent to the number of degrees of freedom, U is the potential energy and s is a time-scale variable with associated momentum p_s and Q describes the coupling strength [154, 155]. The Hamiltonian generates the Boltzmann probability distribution independent of the choice for $H_{Nosé}$ and Q.

2.1.2.3.5 Parrinello-Rahman Barostat

Parrinello and Rahman extended the Andersen method to allow for the system cell to change its volume. Volume is inversely related to pressure by Boyle's law for an ideal gas therefore constant pressure can be maintained through allowing the system cell shape to change [156]. The system is coupled to a pressure bath with reference pressure P_0 by

$$\frac{dP(t)}{dt} = \frac{P_0 - P(t)}{\tau_p} \tag{2.19}$$

where τ_p is the pressure coupling constant [157]. For systems involving a membrane, semi-isotropic coupling is employed instead and pressure changes in the x-y plane are treated independently of those in z.

2.1.2.3.6 Monte-Carlo Barostat

The Monte-Carlo barostat [158, 159] uses the target pressure and system energy to probabilistically choose increases or decreases in system cell size [160]. It is a process of a simple volume-rescaling via trial moves where a random change of volume is proposed and is either accepted or rejected [161]. A trial volume change, ΔV is chosen from a uniform distribution from the interval $[-\Delta V_{max}, \Delta V_{max}]$ such that,

$$V' = V + \Delta V \tag{2.20}$$

where V' is the new volume, V is the old volume and ΔV is the trial volume change. V_{max} is typically chosen such that the MC acceptance ratio of 40-50% is attained. The lengths, l_i of the simulation box and particle positions, r_i are then scaled according to

$$l_i' = l_i^3 \sqrt{\frac{V'}{V}} \tag{2.21}$$

and

$$r_i' = (r_i - c_i) \sqrt[3]{\frac{V'}{V}} + c_i \tag{2.22}$$

respectively where $i = 1, 2, 3$ and c_i denotes the coordinates of the center of mass (COM) of the periodic box. The scaling is done isotropically but anisotropic scaling is also possible [159].

$$\Delta W = U_{pot}' - U_{pot} + P_0 \Delta V - N k T_0 \ln \frac{V'}{V} \tag{2.23}$$

where P_0 and T_0 are the desired pressure and temperature and N is the number of molecules in the system.

The volume move is then accepted with a probability given by

$$P(\Delta V) = \begin{cases} 1 & \text{if } \Delta W \leq 0 \\ exp(-\frac{\Delta W}{k_B T}) & \text{if } \Delta W > 0 \end{cases} \tag{2.24}$$

A successful move is completed by updating forces on particles to yield a consistent set of positions and accelerations to form the new configuration in the isothermal-isobaric (NPT) ensemble [158].

2.1.2.4 The Particle Mesh Ewald Method

The particle mesh Ewald (PME) algorithm [162] is an efficient method for the treatment of electrostatics in periodic systems. Electrostatic interactions are split into short and long-range potentials. The short range evaluates particle-particle interactions in real space. Long-range contributions are converted to reciprocal space and expressed as a discrete convolution on a grid, thus evaluation by a Fourier transform enables rapid convergence. The Ewald sum is the solution to Poisson's equation for the electrostatic potential with periodic boundary conditions in reciprocal space. Forces are calculated by interpolating the derivative of the potential back to the particle positions. A cutoff distance is applied for the short-range terms. The PME algorithm reduces the time complexity of calculating electrostatic interactions from an order N^2 computational problem to order $Nlog(N)$ [163, 164].

2.1.2.5 Periodic Boundaries Conditions and Long Range Forces

Systems have a finite number of particles due to limited computational resources. A finite simulation cell introduces unphysical behaviour and artefacts for particles that are forbidden from crossing cell boundaries or are in proximity of a vacuum. Periodic boundary conditions (PBC) present a better approach to approximate the behaviour of particles representative of the bulk phase. Imposition of PBC on a simulation cell has the effect of surrounding a cell by an infinite number of identical periodic images in all

directions (x, y and z) such that if a particle leaves the cell, it will re-enter through the opposite boundary. Particle interactions are calculated across boundaries, therefore it is important that the simulation cell is chosen such that proteins do not "feel" themselves across boundaries. The minimum-image convention is applied along with cutoff schemes to approximate short-range forces such that each particle may only interact with image particles residing within less than half the dimensions of the unit cell. Long-range electrostatic interactions are treated by the PME method (Section 2.1.2.4). This method comes with some drawbacks, including undesirable long-range ordering and the absence of wavelengths greater than the length of the unit cell [143].

2.2 Polarisation Effects in Classical Molecular Dynamics

Pairwise additive potentials are the main type of forcefields implemented for biomolecular MD simulations. They have been instrumental in understanding structures and biomolecular systems over the past few decades yet despite continuous forcefield improvements, they are in part limited in physical accuracy due to their inability to model induced polarisation and many-body effects [165].

In pairwise additive simulations, each atom is assigned a fixed partial charge before the simulation begins and polarisation is ignored between atoms within the same molecule. This is an oversimplification because in reality, the charge density surrounding each atom is constantly shifting in response to their environment, so the partial charges of atoms would be better represented with dynamic and responsive models [166, 167].

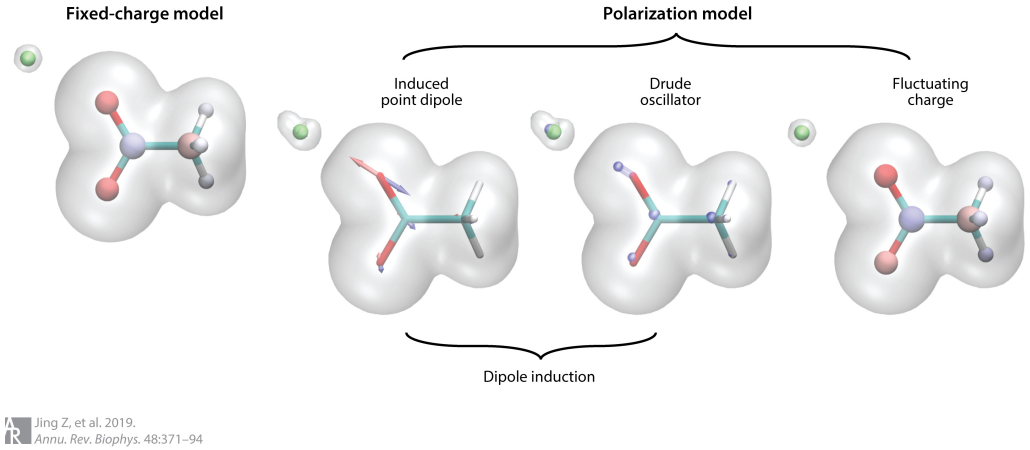

Figure 2.4 | Comparison of Fixed Charge and Polarisable Models Schematic representations of the electron cloud surrounding fixed-charge models compared with polarisable models. The induced point dipole e.g. AMOEBA forcefield and the Drude oscillator e.g. CHARMM Drude models can represent the deformation of electronic shells whereas the fluctuating charge model can represent the redistribution of charge. Figure reprinted from Jing *et al.* [97]

There are three main methods for modelling polarisation in classical forcefields: fluctuating charge, Drude oscillator and polarisable atomic multipole models [113] (Figure 2.4). Fluctuating charge models are based on the principle that partial charges on atoms are allowed to vary over time. Drude oscillator forcefields model electronic degrees of freedom by means of additional massless Drude particles that spatially displace charge relative to each atom's centroid. Finally, polarisable atomic multipole models calculate inducible point dipoles and higher-order static multipoles [166]. There are short-comings of the fluctuating charge approach as it does not consider the out-of-plane polarisation or polarisation of monatomic ions [97]. Here, we use the polarisable atomic multipole models which are described in more depth in Section 2.2.1.

These forcefields provide new and improved levels of predictive power over non-polarizable forcefields; however, they incur additional computational costs. Advances in computing hardware and polarisation algorithms, in particular GPU computing, have significantly reduced the computational overhead of polarisable forcefields relative to additive forcefields, making them more widely accessible [97].

2.2.1 Polarisable Atomic Multipoles

The atomic multipole approach is to include higher order multipoles in the calculation of electrostatic interactions.

2.2.1.1 The AMOEBA Forcefield

A popular forcefield based on the polarisable atomic multipole method is the Atomic Multipole Optimised Energetics for Biomolecular Applications (AMOEBA) forcefield. It was initially developed for water [167] but has since been extended to cover ions [168], organic molecules [169], lipids [170] and proteins [171]. The AMOEBA forcefield uses an induced dipole to model polarisation and includes higher order permanent multipole moments up to the quadrupole [97, 166, 167]. The electrostatic energy includes contributions from both permanent and induced multipoles but only the dipole moment is treated as inducible (the monopole and quadrupole moments are fixed) [172].

The general functional form of the AMOEBA forcefield is given by,

$$U = U_{bond} + U_{angle} + U_{b\theta} + U_{oop} + U_{torsion} + U_{vdW} + U_{ele}^{perm} + U_{ele}^{ind} \tag{2.25}$$

where the first five terms define the short-range bonded interactions: bond stretching, angle bending, bond-angle cross term, out-of-plane bending and torsion rotation. The last three terms define the non-bonded vdW and electrostatic contributions: vdW, permanent electrostatic and induced electrostatic contributions. The potentials consist of anharmonic bond and angle terms as well as Urey-Bradley terms with associated spring constants (see Figure 2.1 and Equation 2.4) to describe the intermolecular geometry and vibrations. A buffered 14-7 potential for vdW captures the repulsion and dispersion interactions [167, 169].

The electrostatic potential is given by multipole moments up to and including the quadrupole. These are modelled as,

$$U_{ele} = M_i^{\mathsf{T}} T_{ij} M_j \tag{2.26}$$

where M_i is the multipole moment vector with respect to i and T_{ij} is the multipole-multipole interaction tensor with elements obtained according to

$$T = \tfrac{1}{R} \quad T_\alpha = \nabla_\alpha T \quad T_{\alpha\beta} = \nabla_\alpha T_\beta \quad T_{\alpha\beta\gamma} = \nabla_\alpha T_{\beta\gamma} \quad ... \quad (\alpha, \beta, \gamma = 1, 2, 3) \tag{2.27}$$

with $4\pi\epsilon_0$ omitted for clarity [167].

The permanent multipole vector is expressed as,

$$\mathbf{M}_i = \begin{pmatrix} q_i & \mu_i^x & \mu_i^y & \mu_i^z & Q_i^{xx} & Q_i^{xy} & Q_i^{xz} & ... & Q_i^{zz} \end{pmatrix}^{\mathsf{T}} \tag{2.28}$$

where q, μ and Q are the monopole, dipole and quadrupole moments respectively. The multipole expansion is truncated at the quadrupoles and is defined in local coordinate frames with atomic sites at the origins.

The induced dipole at each atomic site according to the electric field arising from both the permanent and induced multipole moments at that site is computed by

$$\mu_{i,\alpha}^{ind} = \alpha_i E_{i,\alpha} = \alpha_i (\sum_j T_\alpha^{ij} M_j^{perm} + \sum_{k \neq i} T_{\alpha\beta}^{ik} \mu_{k,\beta}^{ind}) \tag{2.29}$$

where α_i is the atomic polarisability and $E_{i,\alpha}$ represents the electric fields generated by both the permanent (first term) and induced (second term) multipoles. M_j^{perm} contains the permanent multipole moments and μ_k^{ind} contains an induced component which adapts to the local electric field caused by other atoms in the environment. T_α^{ij} is the multipole interaction tensor between sites i and j with components introduced in

Equation 2.27 [167, 173, 174]. If two inducible dipoles come into very short ranges of each other, the dipolar interaction between them will mutually enhance their dipoles to an infinite size. This is known as the "polarisation catastrophe". To avoid this, Thole's damped model is implemented [175] where ρ is used to replace the point dipole moment and is defined as

$$\rho = \frac{3a}{4\pi} exp(-au^3) \tag{2.30}$$

where $u = R_{ij}/(\alpha_i\alpha_j)^{1/6}$ is the effective distance as a function of atomic polarisablities of sites i and j. The a factor is a dimensionless width parameter of smeared charge distributions and controls the damping strength [167, 176].

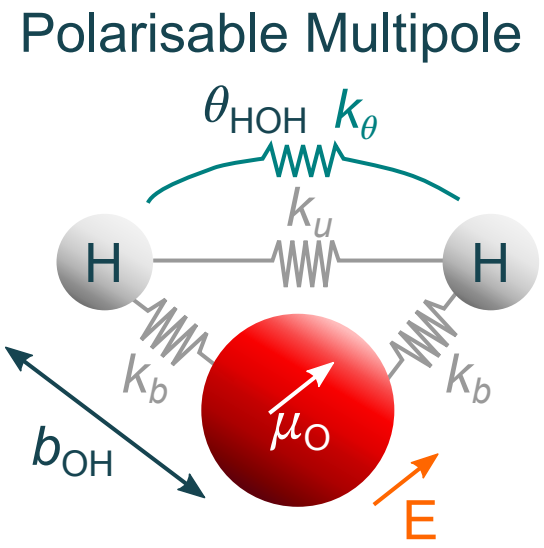

Figure 2.5 | Polarisable Multipole Water Model Geometry of the AMOEBA03 water model. Three interactions sites are located at the positions of the hydrogen (white spheres) and oxygen (red sphere) atoms. The geometry of the model is flexible and defined by O–H bond lengths b_{OH} and the H–O–H bond angle which are subject to anharmonic potentials with force constants k_b and k_θ. The Urey-Bradley potential is applied with force constant k_u to restrain the H–H distance. μ_O represents an inducible dipole on the oxygen atom and E represents an external electric field.

Several explicit water models have been parameterised for the AMOEBA forcefield. The AMOEBA03 model (Figure 2.5) [167] was the first model which served as the foundation for the parameterisation of the protein forcefield. Subsequently, the iAMOEBA model (inexpensive AMOEBA) [177] followed by a reparameterisation using ForceBalance [177] yielded the AMOEBA14 water model [63, 174]. Only the AMOEBA03 water model is implemented in the work presented here and its parameters are presented in Table 2.2.

	Parameter	Units	AMOEBA03 [167, 174]
O–H bond	b_{OH}	nm	0.09572
H–H Urey-Bradley	u_{HH}	nm	0.15326
	θ_{HOH}	°	108.50
Force constants	k_b	MJ/mol/nm^2	232.99
	k_u	kJ/mol/nm^2	-3.18
	k_θ	kJ/mol/rad^2	203.80
O multipoles	q_O	e	-0.51966
	μ_z^O	e bohr	0.14279
	Q_{xx}^O	e bohr^2	0.37928
	Q_{yy}^O	e bohr^2	-0.41809
	Q_{zz}^O	e bohr^2	0.03881
H multipoles	q_H	e	0.25983
	μ_x^H	e bohr	-0.03859
	μ_z^H	e bohr	-0.05818
	Q_{xx}^H	e bohr^2	-0.03673
	Q_{yy}^O	e bohr^2	-0.10739
	Q_{xz}^O	e bohr^2	-0.00203
	Q_{zz}^O	e bohr^2	0.1441
vdW	ϵ_O	kJ/mol	0.4600
	ϵ_H	kJ/mol	0.0565
	R_O	nm	0.1703
	R_H	nm	0.1328

Table 2.2 | AMOEBA03 Water Model Properties Geometry, electrostatic and vdW parameters for the polarisable AMOEBA03 water model. The Bohr radius $a_0 = 5.29$ pm. See Figure 2.5 for the definition of these parameters.

An alternative approach to modelling polarisability (however not explored in this thesis) is with Drude particles as in the CHARMM Drude forcefield. This type of model is known as a *Charge-on-Spring* model as it includes a massless Drude particle that enables the displacement of charge from the atomic center of particles [178]. The Drude oscillator model is a direct extension of additive models and retains many of the pairwise features [179].

2.2.1.2 MD Algorithms for Polarisable Atomic Multipoles

The treatment of induced dipole moments requires careful consideration as their functional forms are not of an additive manner, therefore additional algorithms are required. The solution to Equation 2.29 can be rewritten as the self-consistent equation

$$\mu_{i,\alpha}^{ind}(n+1) = \mu_{i,\alpha}^{ind}(n) + \alpha_i \sum_{\{k\}} T_{\alpha\beta}^{ik} \mu_{k,\beta}^{ind}(n) \quad (n = 0, 1, 2, ...) \tag{2.31}$$

where $\mu_{i,\alpha}^{ind}(0) = \alpha_i \sum T_{\alpha\beta}^{ik} M_j$ is the "direct" induced dipole on site i. Equation 2.31 can be solved iteratively through successive over-relaxation (SOR) to achieve faster convergence [167]. Alternatively, treatment of induced dipoles can be through the "direct" approximation which is a non-iterative approach that essentially neglects to calculate the coupling between induced dipoles. Instead, the extrapolated perturbation theory which finds $\mu_i^{ind}(\infty)$ through extrapolation of the first few iterations of Equation 2.31 is used to reintroduce this mutual coupling back into the direct algorithm. This method is computationally more efficient compared to iterating the self consistent equation [180].

Additionally, in AMOEBA certain interactions may evolve with time more rapidly than others. The fast varying short-range interactions need to be evaluated more frequently than the long-range interactions. Therefore, multi-time-stepping is essential for more efficient treatment of force calculations [181]. The reversible reference system propagation algorithm (r-RESPA) method is a multiple timestep integrator [182, 183].

Within this method, forces are grouped according to how rapidly they vary over time. Each group is then evaluated at its own timestep whilst maintaining accuracy and numerical stability which accelerates the computation of AMOEBA simulations [143].

2.2.2 Charge Scaling and Mean-field Methods

2.2.2.1 Electronic Continuum Correction (ECC)

Explicit treatment of electrostatics is computationally expensive. For this reason, it has been desirable to modify existing parameters in additive force fields to improve their physical accuracy and to provide a more efficient alternative to full polarisable forcefields (Figure 2.6).

Electronic polarisation can be understood as the deformation of electronic shells which is an additive quantity in the linear approximation. The total electronic polarisation can be partitioned into two parts: the nuclear or "inertial" polarisation which depends on the position of the nuclei and the electronic or "inertialess" polarisation which corresponds to the pure electronic response arising from the external field. In simple terms, nuclear polarisation dominates the low-frequency response while the high-frequency response is due to electronic polarisation. In non-polarisable MD models, the pure electronic component of polarisation is not included in the electronic description, meanwhile the remaining nuclear component is effectively captured in the definitions of atomic partial charges and other non-electrostatic parameters [184]. In practice, this means that ions and other highly charged components in additive forcefields have been observed to dramatically overbind to proteins. The overbinding originates from neglecting to model the pure electronic polarisation from which effects such as charge screening arise [185].

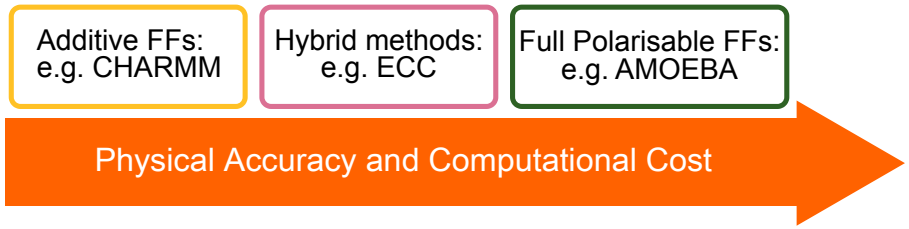

Figure 2.6 | Schematic diagram of Forcefield Types Based on the Modelling of Electrostatics Schematic diagram representing the range of forcefields based on the modelling of electrostatics. Starting with additive forcefields modelled with fixed-point charges, for example CHARMM and ranging to full polarisable forcefields with explicit modelling electrostatics. In between are hybrid methods that employ e.g. charge scaling via the ECC method.

The Electronic Continuum Correction (ECC) was first proposed to implicitly introduce electronic polarisation into additive models [186]. The idea was to embed the entire system into a dielectric continuum with a dielectric constant corresponding to the electronic component of the polarisation response i.e. the high frequency dielectric constant [187]. Through manipulation of Coulomb's law, one can show that the scaling of charges can be described by the relationship,

$$q_i^{eff} = \frac{q_i}{\sqrt{\epsilon_{el}}} \tag{2.32}$$

where ϵ_{el} is the high frequency dielectric constant. For water, $\epsilon_{el} \sim 1.78$ and the charge scaling factor is then $\frac{1}{\sqrt{\epsilon_{el}}} \sim 0.75$ which is a suitable approximation for aqueous environments including those of biological relevance [101]. This method is equivalent to placing charged moieties into a medium rather than in vacuo by replacing the permittivity of free space with the permittivity of the medium in Coulomb's law [187]. Accounting for the polarisation of the surrounding medium is important because roughly half of the solvation free energy of ions is due to electronic polarisation of the medium and electronic screening reduces the interaction between two charges by half compared to in vacuum [188].

Despite its success in improving electrostatic descriptions, ECC has its limitations. The applicability of the ECC approach is restricted to systems with uniform high frequency dielectric constant i.e. electronically homogeneous. Therefore, ECC is not applicable for systems that are abruptly discontinuous in high frequency dielectric constant for example the air/water interface [79, 101].

Furthermore, a general consensus has not been reached over the compatibility of non-polarizable water models and respective scaling factors within the ECC framework. It is also dependent on the chosen forcefield. There are several scaled-charge forcefields with a range of scaling factors in the interval 0.85-0.94 typically applied to monatomic ions [101, 139, 189]. Later, the ECC framework was extended to further refine the capabilities of charge-scaling in the Electronic Continuum Correction with Rescaling (ECCR) approach which applies an additional 5-10 % decrease in ion radius which enables us to recover the correct hydration structure [101, 139].

2.2.2.2 The prosECCo Forcefield

Recently, the Polarisation Reintroduced by Optimal Scaling of Electronic Continuum Correction Origin (prosECCo) forcefield was developed to encapsulate charge scaling for easy application to biomolecular systems. prosECCo is a derivative of the additive forcefield CHARMM36 (c36) and incorporates the ECCR framework through rescaling the charge on highly charged groups by a factor of 0.75 and is further accompanied by adjustments to Lennard-Jones (L-J) parameters by decreasing the atomic radii by 5-10% in order to recover the correct hydration structure [101]. This has been systematically applied for a wide range of lipids, proteins, sugars and ions [190] `https://gitlab.com/sparkly/prosecco/prosECCo75`. We apply the prosECCo forcefield to a biological protein system in Chapter 4.

2.2.2.3 NBFIX

An alternative method with similar objectives to the ECC method is the Non-bonded FIX (NBFIX) approach. In NBFIX, pair-specific LJ parameters are defined for specific interacting pairs of ions or charged group in a system in order to fine-tune properties according to experimental osmotic pressures [102, 104]. This approach presents limitations including the need to parameterise each charge-pair independently and experimental data (e.g. osmotic pressures [104]) is required as a parameterisation target not to mention it can be tedious [101]. NBFIX is not used in the work presented in this thesis however it is a conceptually comparative method to ECC.

2.3 Non-equilibrium Methods

2.3.1 Potential of Mean Force Calculations

The potential of mean force (PMF), $\mathcal{W}$, describes the change in free energy along a given reaction coordinate (ξ), otherwise known as a *collective variable* (*CV*). The *CV* can be obtained from e.g. steered MD simulations where a force constant is applied to a selection of atoms [191]. In this thesis, we have a particular interest in calculating the 1D PMF of an ion moving through a pore or away from a binding site. For these cases, the *CV* is defined by taking the position of a single ion along an axis (typically the z-axis). The PMF is related to the probability $P(\xi)$ of the system at a given *CV* by the Boltzmann equation,

$$\mathcal{W}(\xi) = -k_B T \ln P(\xi) + \mathcal{W}_0 \tag{2.33}$$

where $\mathcal{W}_0$ is an arbitrary constant, often chosen with respect to some suitable reference state which may correspond to a free energy of zero [143, 192]. In principle, $P(\xi)$ can be estimated directly from an MD trajectory through Boltzmann averaging however this approach is confined to the conformational space that can be sampled within a feasible

time frame and is often subject to large energy barriers along ξ. Therefore, to overcome these barriers, more efficient enhanced sampling approaches have been developed such as umbrella sampling [193]. In umbrella sampling, a harmonic biasing potential $w_i(\xi)$ is introduced to ensure the entire domain of ξ is sampled. In practice, the forcefield potential is amended by

$$w_i(\xi) = \frac{1}{2}k_\xi(\xi - \xi_i)^2 \tag{2.34}$$

with force constant k_ξ which restrains a selection of atoms to the COM of a "window" i along the *CV* during umbrella sampling. The choice of window separation should be spaced with sufficient overlap such that the PMF is not affected by gaps along the profile. Probability distributions $P_i(\xi)$ can be estimated from these biased simulations though they need to be combined with the overall distribution such that the bias due to $w_i(\xi)$ is corrected. Unbiasing can be achieved through the weighted histogram analysis method (WHAM) [194] which is based on iteratively solving a set of coupled non-linear equations,

$$P(\xi) = \frac{\sum_i^{N_w} n_i P_i(\xi)}{\sum_j^{N_w} n_j exp(-\beta(w_j(\xi) - F_j))} \tag{2.35}$$

$$F_j = -\beta \ln\left(\int d\xi \frac{P(\xi)}{exp(\beta w_j(\xi))}\right) \tag{2.36}$$

where N_w is the number of biased window simulations, n_i is the number of independent data points used to construct $P_i(\xi)$, $\beta^{-1} = k_B T$ is the thermodynamic temperature factor and F_j is the free energy associated with introducing the window potential [192]. WHAM can estimate the statistical uncertainty associated with the PMF through a resampling technique called *bootstrapping* [195]. Popular implementations of WHAM and those used in this thesis include the Grossfield lab implementation (http://membrane.urmc.rochester.edu/wordpress/?page_id=126") and the g_wham tool in GROMACS [195].

"Computers are useless. They can only give you answers."

— Pablo Picasso

3

A Model Biomimetic Nanopore

Contents

3.1 Introduction 52
3.2 Objectives 54
3.3 Aqueous/Decane Interface System 54
3.3.1 Methods 55
3.3.1.1 Equilibrium MD Simulations 55
3.3.1.2 AMOEBA Forcefield Simulations 57
3.3.2 Results & Discussion 57
3.4 Model Biomimetic Nanopore 63
3.4.1 Model Nanopore Design 63
3.4.2 Methods 65
3.4.2.1 Equilibrium Molecular Dynamics 65
3.4.2.2 Umbrella Sampling 66
3.4.2.3 Analysis 67
3.4.3 Results and Discussion 68
3.4.3.1 Influence of Effective Polarisation within the Model Nanopore . 68
3.4.3.2 Cl^- Accumulation at the Hydrophobic Pore Wall 71
3.4.3.3 Energetics of Ion Permeation 73
3.4.3.4 Ion Hydration Within the Pore 76
3.4.3.5 Robustness and Sensitivity of ECC 81
3.5 Summary & Conclusions 85

This chapter is based upon the paper:

"*Influence of Effective Polarization on Ion and Water Interactions within a Biomimetic Nanopore.*" **Phan, L. X.**, Lynch, C. I., Crain, J., Sansom, M. S. P. and Tucker, S. J. *Biophysical Journal*, 121, 11, 2014-2026 (2022)

3.1 Introduction

Ion and water permeation through sub-nanometer pores are influenced not only by the pore radius, but also by the local hydrophobicity of the pore lining. The behaviour of halide ions at hydrophobic interfaces has been studied extensively using both simulation and experiments (reviewed in Section 1.2.2). Despite existing studies, contradictory trends in the energetic contributions of anions and water at such interfaces persist and a universal consensus on their influences has not yet been reached [80, 196]. There remains a sustained interest in anions at hydrophobic interfaces due to a number of recent structural studies which have indicated that hydrophobic contacts within ion channels and nanopores may act as favourable interaction sites for anions, in particular within Cl^--selective channels [100, 105, 197].

Model systems are a useful tool to reduce the dimensionality of a problem and isolate components of interest to investigate in more depth. To begin exploring the mechanisms of anion interactions with hydrophobic contacts, it is first of interest to examine the behaviour of anions at the most simple model surface: a planar aqueous/hydrophobic interface, before proceeding to study anions in more complex environments resembling biological ion channels i.e. a model biomimetic nanopore. In doing so, it is important to consider how ions and water are modelled in MD simulations. Thus far, MD simulations largely employ non-polarisable (NP) forcefields which neglect to provide a physically accurate description for polarisation; namely insufficient screening and inaccuracies in describing short-range ion-water and ion-ion interactions may result in artificially strong interactions for mainly ion-water interactions [198, 199]. As discussed in Section 2.2.2.1, the ECC method attempts to reintroduce electronic polarisation through a mean-field approach by rescaling integer charges on monatomic ions in electrolyte solutions. Vazdar *et al.* [79] have demonstrated the effective utility of the ECC method in simulations of an

aqueous/oil interface, yielding increased halide ion affinities to the surface. This reproduced the effects observed by explicitly polarisable simulations of similar hydrophobic interfaces [77, 200–202] that were otherwise not observed in corresponding NP simulations, thus emphasising the importance of modelling polarisation in ion forcefields. In principle, the ECC method is a good approximation that is valid for aqueous environments however it is partially dependent on water model compatibility as some water models have been shown to perform better than others however no general consensus on the most appropriate water model has been reached at this time. Similarly the scaling factor is not definitive [101].

One question to contextualise these observations thus far is: do halide ions exhibit similar affinities to hydrophobic interfaces within confined environments e.g. within ion channels? One approach is to use model nanopores to mimic key features of biological channels of known structure and function. Carbon nanotubes (CNTs) (see Section 1.1.4) provide a simple biomimetic platform to explore hydrodynamic flows and their molecular characteristics. Much like biological channels, CNTs have demonstrated rapid water transport properties as a result of weak attraction to the hydrophobic pore walls and strong water-water interactions [203]. Depending on CNT geometry and chirality, water flows within such confined spaces do not necessarily exhibit bulk water transport properties but rather can undergo wetting/dewetting transitions [63, 204]. In addition, CNTs have been instrumental to identifying the molecular-level basis for the partitioning and transport of ions within nanopores. Dye-quenching assays combined with first principle MD simulations [60] have revealed a strong differential anion selectivity upon entry to narrow 0.8 nm diameter CNTs determined by the ability for the anion to partially dehydrate in order to make direct contact with the pore surface. This phenomenon has been observed in several other simulation studies [59, 205–207].

3.2 Objectives

In this chapter, we first begin with a short preliminary study to assess the suitability and robustness of the ECC method for the systems we are interested in. This is tested in simulations of an aqueous/decane interface with a combination of variables for which results are qualitatively benchmarked against the work presented by Vazdar *et al.* [79]. Once validated, we apply the ECC-rescaled ionic charges to examine the interactions of Cl^-, Na^+ and water within a model biomimetic nanopore formulated from a CNT designed to imitate pLGICs, in particular the strongly hydrophobic region of the pore. We investigate the role of (implicit) polarisability on the behaviour and energetics of Cl^- ions within the hydrophobic region of the pore.

To summarise, this chapter explores three key questions:

1. Is the ECC method an effective approach for including the effects of induced polarisability?
2. How do Cl^- ions behave at planar hydrophobic interfaces?
3. How does Cl^- behave when placed in hydrophobic nanoconfinement i.e. the dimensions of ion channels?

3.3 Aqueous/Decane Interface System

We begin by laying the groundwork with simulations of a test interfacial system to explore the effect of implicit polarisation through the ECC method on the behaviour of ions and water at a hydrophobic interface.

3.3.1 Methods

Equilibrium simulations of the model hydrophobic interface systems were built by first solvating a $3 \times 3 \times 4$ nm^3 simulation box with water and 1 M of NaCl or NaBr followed by two simulation boxes of $3 \times 3 \times 7$ nm^3 of decane molecules. These three boxes were concatenated along the z-axis to form the total unit cell containing two hydrophobic-aqueous interfaces (Figure 3.1).

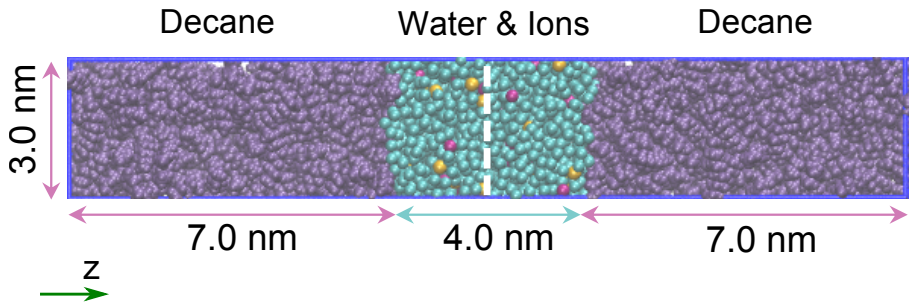

Figure 3.1 | Schematic of Aqueous/Decane Model Interface System The model hydrophobic interface consists of decane (purple), water (cyan), Na^+ (magenta) and Cl^- (yellow). The white dashed line represents the axis of symmetry used for symmetrising plots in the analysis.

3.3.1.1 Equilibrium MD Simulations

All simulations of interfacial systems were carried out using GROMACS version 2018 [144] using the OPLS all-atom forcefield [110]. Energy minimisation was performed using the steepest descent algorithm to correct any steric clashes followed by equilibration in the NVT ensemble for 2.5 ns and a further 2.5 ns in the NPT ensemble. The temperature was controlled by a V-rescale thermostat [150] and was maintained at 300 K with a coupling constant of 0.1 ps and the Parrinello-Rahman barostat [157] was used to maintain isotropic pressure at a compressibility of 4.5×10^{-5} bar^{-1}, approximately that of water. Periodic boundary conditions were applied in all directions. The electrostatics of the system were treated using the PME method [163] and the long range electrostatic interactions were cut off at either 1.0 nm or 1.2 nm using the nearest neighbour count taken by the Verlet scheme. The LINCS algorithm [149] was applied to constrain all bonds

and the equations of motion were integrated using the leap-frog algorithm [144] with a timestep of 2 fs. MD production runs consisted of 50 ns of simulation time and analysis was performed on the final 10 ns of the simulation and averaged over three independent repeats.

Due to the ambiguity surrounding the usage of the ECC method, we performed a set of simulations to test the robustness of the approach. Different combinations of vdW cutoffs, water models and ECC scaling factors were explored. The set of parameter combinations simulated in this thesis are detailed in Table 3.1. After establishing an appropriate set of parameters, they were then applied to explore anion affinity to the surface with 0.25 M incremental changes up to 1 M NaCl.

Water Model	Anion	ECC Scaling Factor	vdW Cutoff (nm)	Concentration (M)	Figure 3.2
SPC/E	Cl^-	0.75	1.20	1.0	C, D, G
	Cl^-	0.75	1.00	0.5	B
	Cl^-	0.70	1.00	1.0	E
	Cl^-	0.75	1.00	1.0	E
	Cl^-	0.80	1.00	1.0	E
	Cl^-	1.00	1.00	1.0	A
	Br^-	0.75	1.00	1.0	H
TIP3P	Cl^-	0.75	1.00	1.0	F
	Cl^-	1.00	0.70	1.0	F
	Cl^-	1.00	0.65	1.0	F
TIP4P	Cl^-	0.75	1.20	1.0	A
TIP4P/2005	Cl^-	0.75	1.00	1.0	C,D

Table 3.1 | ECC Test Combinations Independent hydrophobic slab systems set up with a combination of varying water model, anion in the system, electrostatic cutoff range, ECC scaling factor and ion concentration. The corresponding normalised number density profiles are illustrated in the specified plots in Figure 3.2.

3.3.1.2 AMOEBA Forcefield Simulations

The AMOEBA interfacial system simulation was carried out in OpenMM 7.4.2 (www.openmm.org) and all components were modelled with the AMOEBA13 [171] parameters set and the AMOEBA03 water model [167] in a 0.5 M NaCl solution. Starting configurations for these simulations were taken from the final frame of the non-polarisable forcefield simulations and therefore have been sufficiently equilibrated. Electrostatic multipole moments were treated using the PME method using a real-space cutoff of 0.8 nm, an Ewald error tolerance of 5×10^{-4} and interpolation was performed through fifth order B-splines. VdW forces were calculated with a cutoff distance of 1.2 nm and long-range dispersion corrections were applied to account for interactions beyond this cutoff [208]. Induced dipoles were treated using the extrapolated perturbation theory method [180]. This protocol closely follows that described previously in (https://github.com/Inniag/openmm-scripts-amoeba)[100]. Production runs lasted for 50 ns and time integration was achieved using the r-RESPA method with an outer timestep of Δt =2 fs and an inner timestep of δt = 0.25 fs. The temperature was maintained at 310 K using the Andersen thermostat [152] and the pressure was maintained at 1 bar using the isotropic Monte Carlo barostat [158]. Three independent repeats were performed.

3.3.2 Results & Discussion

Simulations of the model hydrophobic interface have revealed ion surface adsorptions in agreement with prior similar simulations implementing non-polarisable (NP) forcefields [202]. Figure 3.2 A illustrates the density profiles of the NP system using the SPC/E water model. Both the Cl^- and Na^+ ions reside equidistant from the Gibbs dividing surface (GDS), where they remain more fully solvated. However, when polarisation is modelled implicitly using the ECC method (Figure 3.2 B), Cl^- localises in close proximity to the interface (indicated by the graph peak of the blue line at $\sim$ 1.8 nm) with the same region

found to be completely devoid of Na^+ ions altogether. As Cl^- ions are attracted to the surface, a negative surface layer builds, and Na^+ ions, that are repelled from the surface, are now attracted to the negative layer of anions, rendering an effective electrical double layer structure. This leads to increased surface tension corresponding to a negative surface excess (by Equation 1.2) [209]. Anion surface activity depends on the detailed balance between electrostatic forces, repelling surface charges and the polarisability acting to attract the anion to the surface [82]. Away from the interface, all components of the system produce bulk-like solution behaviours.

As discussed above, the use of an appropriate water model within the ECC framework is still open for discussion. Here, we have explored the transferablity of ECC to the choice of water model by simulating the model interface using a range of widely used NP water models: SPC/E, TIP3P, TIP4P and TIP4P/2005. Figure 3.2 C demonstrates that all water models produce a small Cl^- surface adsorption effect with ECC applied indicated by the shift towards the GDS relative to the density profile yielded from the NP SPC/E simulations. Only the SPC/E water model with ECC is able to qualitatively reproduce the experimental surface adsorption effect illustrated by not only the shift towards the interface but with an additional Cl^- accumulation peak [82, 84, 210]. However, the corresponding Na^+ profiles with ECC are repelled from the interface and the implementation of different water models does not affect the surface properties (Figure 3.2 D).

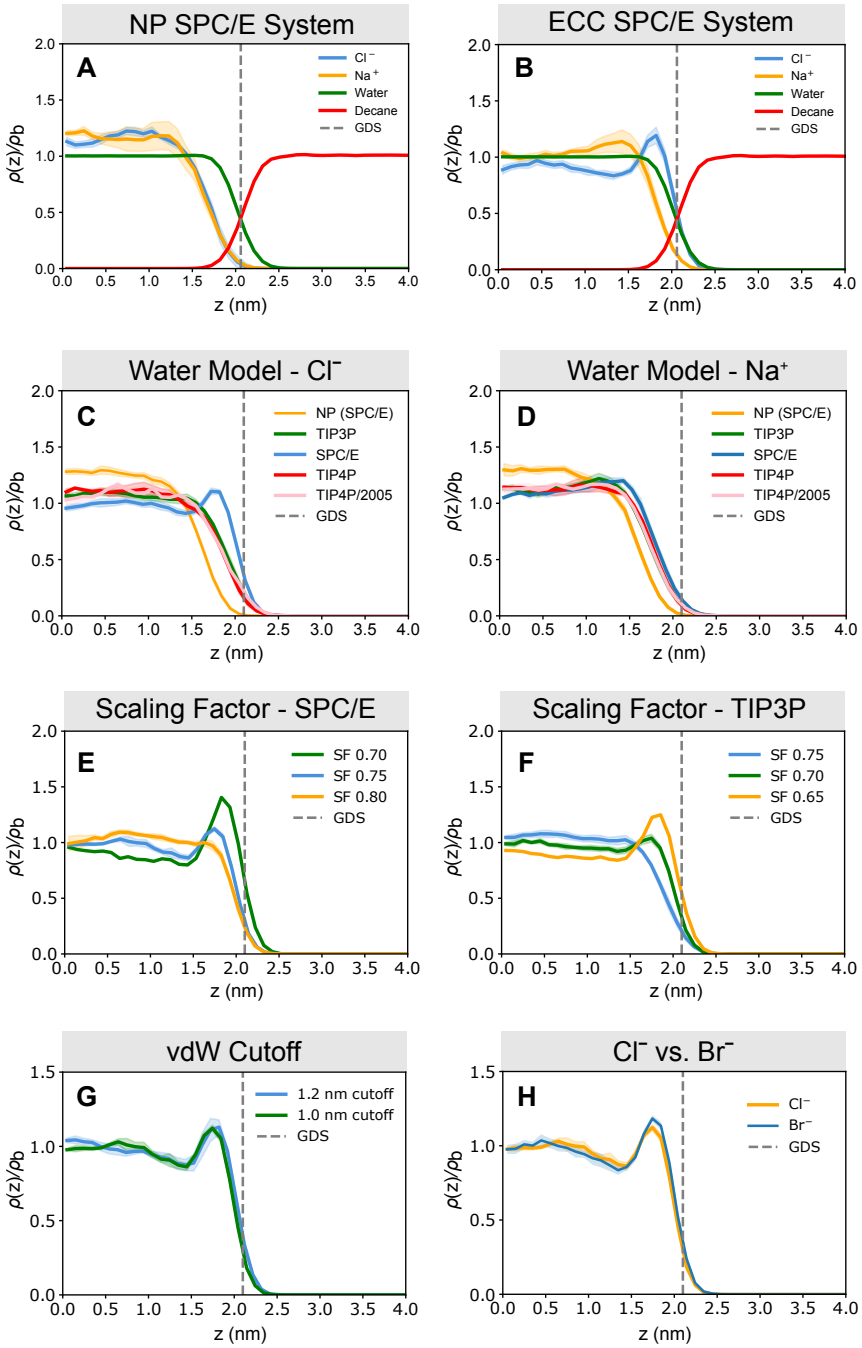

Figure 3.2 | Aqueous/Hydrophobic Interfacial System Tests Comparison plots testing different combinations of variables. The plots detail symmetrised number density profiles $\rho(z)/\rho_b$ calculated by normalising by the bulk density ρ_b. The variable z is the distance from the center of the simulation box and the vertical dashed line corresponds to the aqueous/decane interface or GDS. **A** Non-polarisable (NP) interfacial system with SPC/E water model. **B** ECC interfacial system with SPC/E water model. **C** Cl^- density profile of systems with varying water model. **D** Na^+ density profile of systems with varying water model. **E** Cl^- density profile with varying ECC scaling factor (SF) using the SPC/E water model. **F** Cl^- density profile with varying ECC scaling factor (SF) using the TIP3P water mode. **G** Cl^- density profile with varying vdW cutoff values. **H** Density profile of Cl^- vs Br^-.

The ECC scaling factor (SF) appears to be water model dependent (Figure 3.2 E & F) [211]. The generally accepted SF $\simeq$ 0.75 (calculated from $1/\sqrt{\epsilon_{el}}$) yields an adsorption peak for Cl^- using SPC/E water however, equivalent simulations implementing the TIP3P water model did not produce an adsorption peak with the same SF. Instead, a smaller SF of 0.65 was required to achieve the same effect. Water models are uniquely parameterised, therefore the dielectric constant is subject to variations depending on the model and hence the SF might also differ accordingly. Small differences in the long-range vdW cutoffs for electrostatic interactions were shown to have negligible effects on the surface adsorption of Cl^- (Figure 3.2 G). Finally, the more polarisable halide ion, Br^- was tested in an equivalent model system. Br^- was expected to display stronger surface adsorption relative to Cl^- given its position in the inverse Hofmeister series however when applying ECC to the ions, no significant difference was observed between the density distributions (Figure 3.2 H).

Based on these preliminary studies on the robustness of the ECC method, we have chosen to use the SPC/E water model as it is the most widely used model in conjunction with ECC in existing literature [101, 139, 212]. In addition, we have established a vdW cutoff of 1.2 nm and a SF of 0.75 to be used for further exploration with the ECC method in both the model slab and nanopore.

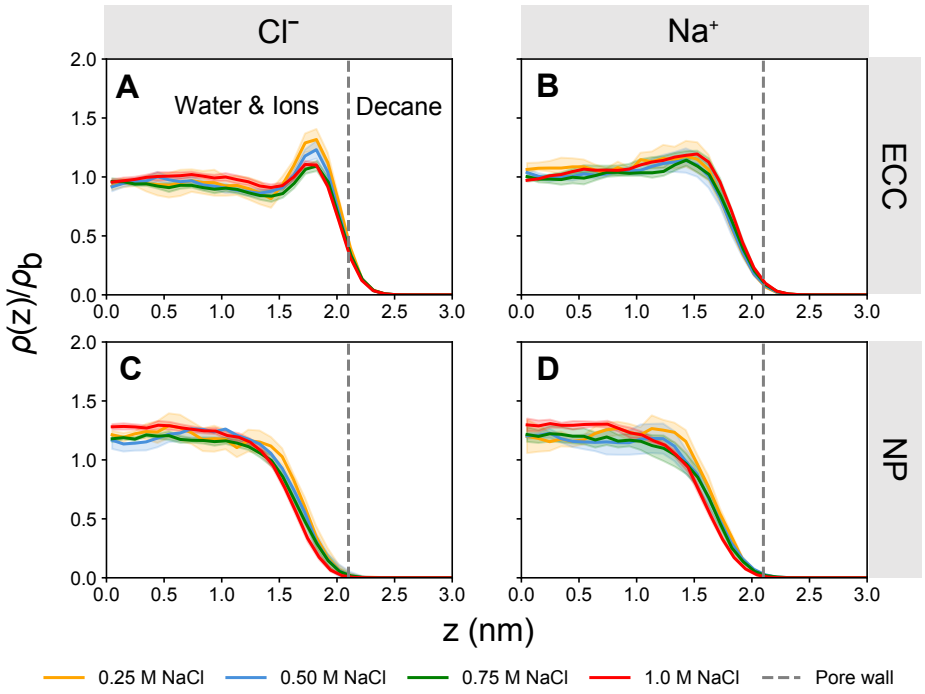

Figure 3.3 | Aqueous/Hydrophobic Interfacial System Symmetrised number density profiles are shown for Cl^- (**A** & **C**) and Na^+ (**B** & **D**) at the aqueous/decane interface with ECC-rescaled charges and a non-polarisable forcefield at various salt concentrations. $\rho(z)/\rho_b$ represents the symmetrised number density, $\rho(z)$, which has been normalised by bulk density, ρ_b. The variable z is the distance from the center of the simulation box and the vertical dashed line corresponds to the aqueous/decane interface. Errors bars reflect the standard errors treated across 3 independent repeats.

The relationship between salt concentration and surface adsorption was examined. In all cases, this revealed a consistent Cl^- adsorption peak towards the interface and corresponding depletion from the Na^+ ion with increasing concentration (Figure 3.3). Again, the surface adsorption effects were otherwise not observed when polarisation was not modelled and both ions reside equidistant from the GDS as previously seen.

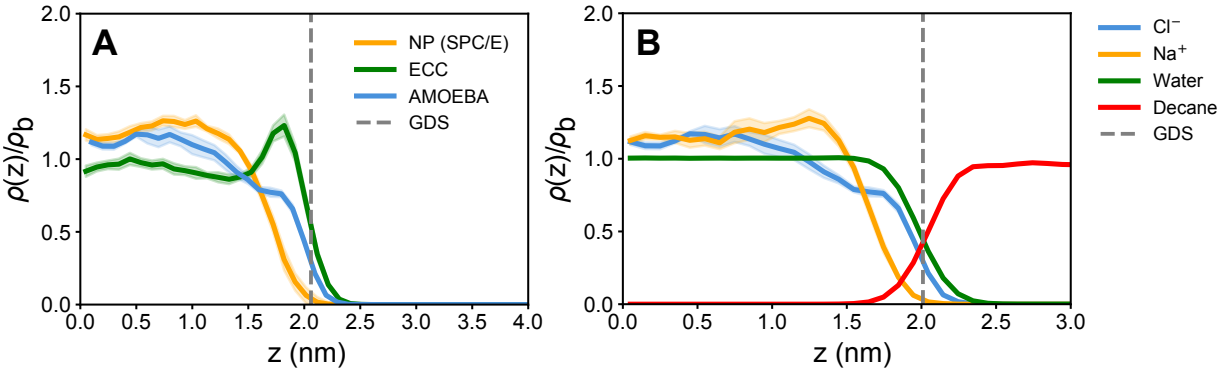

Figure 3.4 | AMOEBA Simulation of Hydrophobic Interface A Symmetrised number density profiles ($(\rho(z)/\rho_b)$) of Cl^- comparing profiles between NP, ECC and AMOEBA simulations. The plots have been aligned relative to the GDS. **B** Symmetrised number density profiles ($(\rho(z)/\rho_b)$) for the AMOEBA forcefield simulation of the aqueous/hydrophobic interface. The variable z is the distance from the center of the simulation box and the vertical dashed line denotes the Gibbs Dividing surface.

As a point for comparison, simulations implementing the fully polarisable AMOEBA forcefield were performed on the same model hydrophobic interface at concentration of 0.5 M NaCl (Figure 3.4 B). The density profiles for each component of the system were in broad agreement with the ECC simulations, namely, the formation of a double electric layer (Figure 3.4 B). However, the degree of surface adsorption of Cl^- in the AMOEBA simulations is less than the ECC simulations indicated by the more modest surface accumulation towards the GDS (Figure 3.4 A) nonetheless a surface effect is present. In conclusion, the ECC method is effective at capturing surface adsorption effects that are observed in spectroscopic experiments [84, 213]. ECC is able to reproduce similar effects of more physically realistic simulations incorporating explicit polarisation however, the extent to which these surface effects occur remains unclear. Next, we move on to explore the effects of ECC on ions and water within a model biomimetic nanopore system.

3.4 Model Biomimetic Nanopore

The test interfacial system showed that electronic polarisation effects were accounted for in an effective way using ECC through analysis of ion adsorption to the GDS. Having obtained an appropriate set of simulation parameters, we proceed by applying the ECC method to explore the behaviour of ions and water within a confined environment through assembly of a model biomimetic nanopore.

3.4.1 Model Nanopore Design

In designing a model nanopore, we sought to explore the interactions of Cl^- with hydrophobic interfaces in a simplified system representative of a biological ion channel. Based on concepts from previous studies of pLGICs [100, 214], we set out to construct a model nanopore that could mimic the general charge distributions of different regions inside such channel pores. pLGICs are a family of ion channels that mediate fast neurotransmission [215]. Within their pores, a highly conserved hydrophobic pore-lining region is associated with channel gating [214]. There are also narrow regions near the entrances of the pore formed by polar and charged residues [216, 217]. We therefore mimicked these two regions in a model nanopore built from a CNT.

The polar regions at either mouth of the nanopore were constructed by applying a harmonic restraining potential between a set of water molecules and the interior pore walls near the openings. We chose to build polar regions in this way to avoid bias toward any particular chemistry of the CNT which may influence ion selectivity [218] and aimed to minimise the introduction of multiple different molecular species. Meanwhile, the central internal hydrophobic cavity was left exposed to resemble the hydrophobic gating region and is the region of interest. This central cavity was considered representative of a (nearly) pure hydrophobic environment as modelling the CNT with higher order

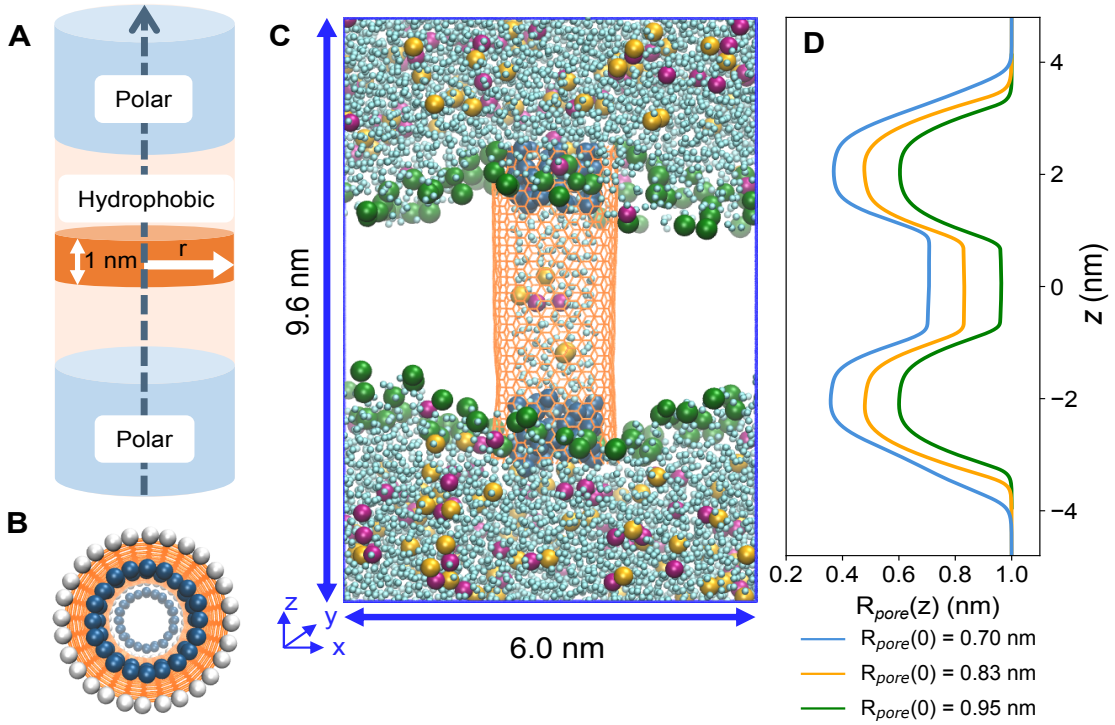

Figure 3.5 | Schematic of the Biomimetic Model Nanopore Design The model nanopore has a radius r, such that $0 < r < R_{pore}$. **B** Top-down view of the pore, showing water molecules (blue) restrained to the CNT pore walls (orange) to create the polar regions at each mouth of the pore. The nanopore ends are capped with hydrogen atoms (white). **C** Snapshot of the simulation setup. The biomimetic nanopore is embedded in a POPC bilayer (lipid headgroup phosphates in green) and solvated with water (light blue), Cl^- (yellow) and Na^+ (pink). **D** Pore radius profile showing the maximum value of R_{pore} as a function of axial position z (approximately aligned with the simulation snapshot in **C**).

electronic moments was expected to have negligible effects since charge distributions become delocalised in the aromatic ring network. Liang *et al.* demonstrated with quantum calculations that within the internal part (region away from the edges) of graphene, charge distribution was relatively uniform and close to zero [219]. The resultant model nanopore has three defined regions of alternating polarity and hydrophobicity along its pore axis (Figure 3.5 A & B). This pore was then embedded into a phospholipid bilayer to span the thickness of the membrane ($\sim$ 4.7 nm) and form a stable membrane-embedded nanopore (Figure 3.5 C).

Pristine armchair (14,14) CNTs were generated and capped with hydrogen atoms to form the hydrophobic pore using the molecular builder, Avogadro [220], and VMD [221]. The length of the CNT was $\sim$ 4.7 nm (and therefore capable of spanning the thickness of the membrane) and the internal diameter was $\sim$ 1.4 nm. A smaller armchair (10,10) CNT was built as a template for insertion and restraining of water molecules in selected positions to create the polar regions of the pore. A harmonic restraining potential was applied between the oxygen of the water molecules and the carbon atoms of the CNT pore wall interiors with a force constant of 1200 kJ mol^{-1}nm^{-2} and maximum distance of 0.143 nm ($\sim$ C-O bond length). For simulations investigating the effects of pore radius, (14,14), (16,16) and (18,18) armchair CNTs were used with internal radii of 0.70 nm, 0.83 nm and 0.95 nm respectively. (10,10), (12, 12) and (14,14) armchair CNTs were used as templates for water molecule insertion to create polar regions for the (14,14), (16,16) and (18,18) CNTs respectively. Pore radius profiles (Figure 3.5 D) of the resultant model nanopores were calculated using the Channel Annotation Package (CHAP)[69].

3.4.2 Methods

3.4.2.1 Equilibrium Molecular Dynamics

We performed 50 ns atomistic MD simulations of the model nanopore embedded in a POPC (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine) bilayer. The nanopore was inserted into the POPC bilayer by the InflateGRO protocol whereby the nanopore was placed in the membrane and equilibrated [222]. The system was then solvated with a 0.50 M NaCl solution. All systems were first equilibrated for 10 ns and this period was not included in analysis. The simulations were carried out using GROMACS version 2020 [144] with the OPLS/AA forcefield [110] and the SPC/E water model [223]. Typically OPLS forcefields are used in conjunction with the TIP family of water models however as a result of transferability of models, SPC models also yield similar results [224]. Although no general consensus has been reached with regards to the most appropriate water model to

use in conjunction with ECC [185], some water models perform better than others as shown previously with the SPC/E water model [225].

The integration timestep was 2 fs. All bonds were constrained using the LINCS algorithm [149]. A Verlet cutoff scheme was applied and long-range electrostatics were treated by the particle mesh Ewald (PME) method [164] with a short-range cutoff of 1 nm and a Fourier spacing of 0.16 nm. Three-dimensional periodic boundary conditions were applied. Simulations were performed in the isothermal-isobaric ensemble. The temperature was maintained at 300 K with a coupling constant of $\tau_t = 0.5$ ps with a Nosé-Hoover thermostat [153, 154]. Pressure was maintained semi-isotropically using the Parrinello-Rahman barostat [157] at 1 bar with a coupling constant of $\tau_t = 2.0$ and compressibility of 4.5×10^{-5} bar^{-1}. For simulations exploring the effects of ion concentration, NaCl concentrations from 0.25 M to 1.0 M were considered in 0.25 M increments. Electronic polarisation effects were introduced to the system in a mean-field approach by applying the ECC method [101, 184]. This was realised by rescaling all ionic charges by a scaling factor of 0.75. Three independent repeats were carried out for each individual parameter combination. Data was analysed using GROMACS and custom written code using MDAnalysis [226–228].

3.4.2.2 Umbrella Sampling

Umbrella sampling [192] was performed to obtain one-dimensional potential of mean force profiles for Na^+ and Cl^- ions moving through the model nanopore using both the non-polarisable forcefield and ECC method. Simulation details follow a similar protocol to those detailed above. However, to prevent the nanopore from tilting in the bilayer, the carbon atoms of the CNT were placed under a harmonic restraint with a force constant of 1000 kJ mol^{-1} nm^{-2}. Equilibration simulations were performed for 15 ns and the starting

configurations for the umbrella windows were produced from the final state of these simulations.

The reaction coordinate was defined as the z-axis which corresponds approximately to the pore axis and direction normal to the lipid membrane. A target ion was relocated to subsequent positions along the z-axis followed by 10 steps of energy minimisation to remove any steric clashes between the target ion and surrounding water molecules. A harmonic biasing potential was applied to the z-coordinate of the target ion with a force constant of 1000 kJ mol^{-1} nm^{-2}. Umbrella windows covered the entire length of the model nanopore and up to 1 nm outside the pore. This setup corresponds to 69 windows along the z-axis with a distance of 0.1 nm between successive windows. Each umbrella window was simulated for 10 ns. PMF profiles were obtained through unbiasing with the Weighted Histogram Analysis Method (WHAM) using the Grossfield lab implementation in version 2.0.9 (http://membrane.urmc.rochester.edu/wordpress/?page_id=126"). The first 5 ns of each simulation were discarded as equilibration, meaning that the final PMF profile was calculated from the final 5 ns of the simulation time. Each resulting PMF was zeroed with respect to the environment outside the nanopore. Convergence was assessed by comparing the cumulative free energy profiles computed from 1 ns fractions of simulation time (see Figure 3.9). The PMF errors were estimated through 100 trials of Monte Carlo bootstrapping error analysis.

3.4.2.3 Analysis

A 1 nm thick segment of the central hydrophobic region of the nanopore was sampled. Number density profiles were calculated by reading the simulation frame by frame and then histogramming specified atoms onto a 3D grid of bin spacing 0.01 nm. The grid was then averaged over the z-axis to produce a 2D plot and symmetrised due to the symmetrical nature of the system. Densities were then normalised by bulk (ρ_b) which was

equivalent to dividing by the total number of given ion species in solution per volume of simulation box.

Structural properties of ionic hydration shells were characterised by calculating the radial distribution function (RDF) between two particles a and b,

$$g_{ab}(r) = (N_a N_b)^{-1} \sum_{i=1}^{N_a} \sum_{j=1}^{N_b} \langle \delta(|\mathbf{r_i} - \mathbf{r_j}| - r) \rangle \tag{3.1}$$

In this thesis, we mainly compute the distribution of water oxygen (O) atoms around a specified ion i.e. where $a \in \{\text{Na}^+, \text{Cl}^-\}$ and b represents the water oxygen (O) atoms. From the RDF, the average ion first hydration shell number was determined by calculating the radial cumulative distribution [229] given by,

$$N_{ab}(r) = \frac{4\pi N}{V} \int_0^r g_{ab}(r) r^2 dr \tag{3.2}$$

where N_{ab} represents the number of particles b (i.e. O), surrounding particle a (i.e. the ion) and V is the simulation box volume. r denotes the limiting radius of the first hydration shell which is determined by the first local minima of the RDF after its first peak. Both the number density calculations and RDFs were determined using MDAnalysis [226, 227].

3.4.3 Results and Discussion

3.4.3.1 Influence of Effective Polarisation within the Model Nanopore

The model biomimetic nanopore system described in Section 3.4.1 was simulated using the OPLS/AA additive forcefield and separately with ECC-rescaled charges applied. Based on preliminary studies from Section 3.3.2, the SPC/E water model combined with an ECC scaling factor of 0.75 and a vdW cutoff of 1.2 nm were chosen. Simulations with varying ionic concentrations in 0.25 M NaCl increments from 0.25 M up to 1.0 M were performed. Number density profiles of ions and water were derived radially to the pore axis from a central hydrophobic region of the pore (see Figure 3.5 A) sampled from a 1 nm thick slice

along the z-axis i.e. from z = -0.5 to +0.5 nm. The final 10 ns of each 50 ns simulation repeat was used for analysis.

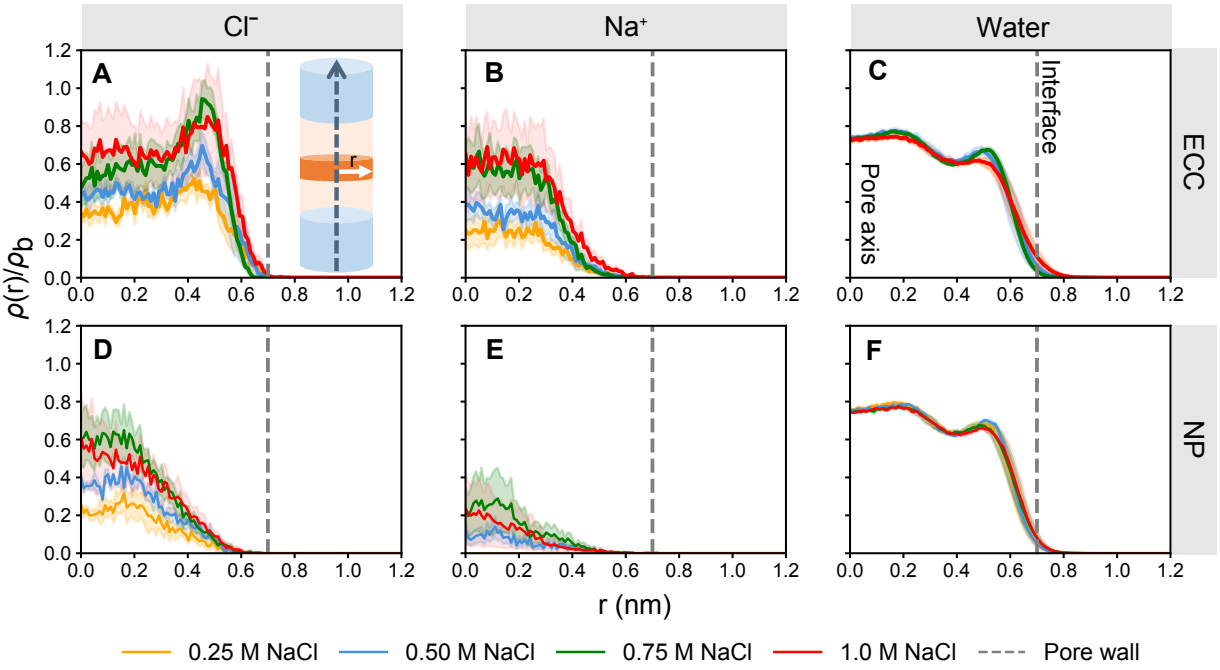

Figure 3.6 | Symmetrised Number Density Profiles as a Function of Salt Concentration Symmetrised number-density profiles of Cl^- (**A** and **D**), Na^+ (**B** and **E**), and water (**C** and **F**), with ECC ionic charges and a non-polarisable forcefield for all other atoms, at various salt concentrations. $\rho(r)/\rho_b$ represents the symmetrised number-density, $\rho(r)/\rho_b$, sampled from the hydrophobic region of the nanopore (orange section of schematic (Figure 3.5 A)) and normalised by bulk density, ρ_b. The variable r is the radius of the nanopore, which extends from 0 (pore axis) to R_{pore} (the interface where the salt solution meets the wall of the nanopore). The gray vertical dashed line represents R_{pore} and the error bars represent the standard error.

Radial density profiles of ions when employing ECC-rescaled charges exhibited increased propensity Cl^- association with the hydrophobic pore-lining, as indicated by a density peak displaced towards the water/CNT interface (Figure 3.6 A, at $r \sim 0.5$ nm). By comparison, the Na^+ ions are largely excluded from the immediate vicinity of this hydrophobic surface and are localised closer to the pore axis where they remain more fully solvated (Figure 3.6 B). Importantly, these ion distributions, which matched those seen in simulations of ions in water nanodroplets [88, 230] and at hydrophobic interfaces

(see Section 3.3.2), were only observed when implicitly including polarisation through the ECC method. This is in marked contrast to simulations with the NP forcefield in which both anions and cations were equidistant from the hydrophobic pore wall, preferring to reside close to the pore axis where they are more fully solvated (Figure 3.6 D & E). Our results can be explained by the breaking of solvation shell symmetry. This can be realised when considering polarisable anions near a hydrophobic surface such as graphene. Anions and cations in the interfacial layer experience an asymmetric and unequal water polarisation response which is non-reciprocal on exchange of the sign of the ion [231] and has also been seen in studies with the Drude polarisable forcefield [232] where the loss of solvation free energy was sufficiently compensated by strong anion polarisation, resulting in qualitative enrichment at the interface. The Na^+ depletion from the interfacial layer is due to a weaker orientational preference and smaller ion-induced dipole hence the formation of a double electric layer where Na^+ is attracted to the surface adsorbed Cl^- [79]. Furthermore, these results reveal similar surface effects to those observed at the aqueous/decane interfacial system (see Section 3.3.2) and are in good agreement with earlier studies of aqueous/air interfaces [202, 233, 234] as well as first principles MD and polarisable simulations of NaCl inside CNTs [235–238].

There is some degree of structuring of the ions and water inside these pores. Notably, the water density distributions (Figure 3.6 C & F) form a layered structure inside the nanopore with the outermost layer forming an ordered concentric ring and inner layers demonstrating more bulk-like profiles. In simulations with ECC, it is this outermost layer that is shared with Cl^- ions. Comparable structured behaviour of water in concentric shells inside pristine CNT porins has previously been reported [239–241].

Given the pore is open, and ions and water are freely allowed to permeate, the area under the normalised number density curve indicates the total number of particles in the

sampling region. At any given salt concentration, there are significantly fewer Na^+ ions present in the hydrophobic core of the nanopore relative to the number of Cl^- (Figure 3.6 A, B & D, E). This observation is more pronounced when applying the NP forcefield, with virtually no Na^+ ions present in the hydrophobic core at lower concentrations, suggesting there is competition between Cl^- and Na^+ to occupy more bulk-like regions close to the pore axis. With the improved electronic description using ECC, an unequal permeability ratio between ions persists, whilst interfacial properties are captured. This could be understood by considering how the interfacial adsorption of Cl^- reduces the competition between ions in the center of the pore by vacating the more bulk-like regions, therefore more Na^+ ions (that favour being fully solvated) permeate the sample region. Another aspect to consider is the cotransport of different ion species under nanoconfinement. It is thought that ion pairs of unusually long lifetimes can form within these dimensions [242]. In contrast, DFT calculations have suggested two contradictory effects observed in narrow CNTs: interactions with the image charge favours ion insertion however this weakens the binding between ions [206].

3.4.3.2 Cl^- Accumulation at the Hydrophobic Pore Wall

We next examined the behaviour of ions and water inside model nanopores of varying radii to explore the influence of different degrees of confinement (Figure 3.7). Model nanopores with radii (R_{pore}) for the hydrophobic region of approximately 0.70 nm, 0.83 nm and 0.95 nm, were built. These pore dimensions are comparable to the hydrophobic gate of a simplified open-state β-barrel structure [239]. All systems were simulated with 0.50 M NaCl solution. Analysis protocols were the same as those detailed for Section 3.4.3.1 however these data are now additionally normalised by the surface area of the pore.

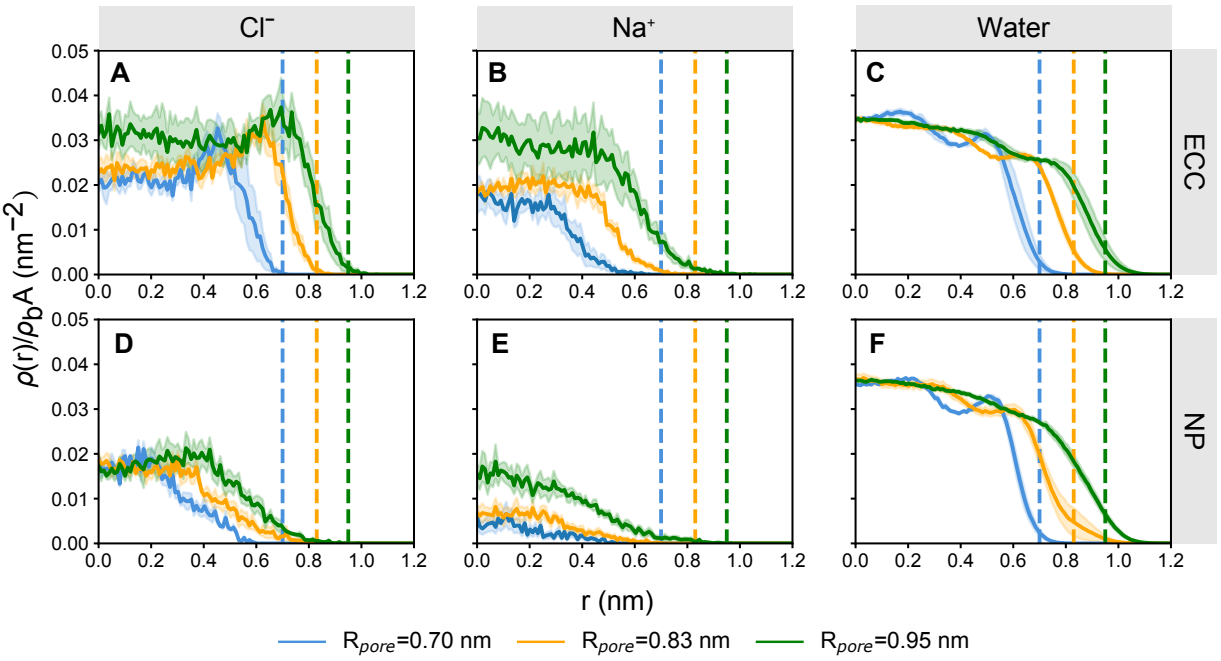

Figure 3.7 | Symmetrised Number Density Profiles as a Function of Pore Radius Symmetrised number-density profiles of Cl^- (**A** and **D**), Na^+ (**B** and **E**), and water (**C** and **F**), with ECC-rescaled ionic charges and a standard non-polarisable forcefield for all other atoms, in nanopores of different radii. $\rho(r)/\rho_b$ represents the symmetrised number-density, $\rho(r)$, sampled from the hydrophobic region of the nanopore and normalised by bulk density, ρ_b, and internal surface area, A, of the nanopore. The vertical dashed lines indicate the interface where the aqueous solution meets the wall of the nanopore (at radius R_{pore}) and the errors bars represent the standard error.

In ECC simulations, the extent of Cl^- accumulation at the hydrophobic pore wall is approximately consistent across nanopores of all sizes indicated by the relative height of the interfacial peaks (Figure 3.7 A). This suggests that at a concentration of 0.5 M NaCl, the surface propensity of Cl^- has reached saturation at the interface and any additional ions entering the hydrophobic region contribute to the density of inner, more bulk-like regions. In contrast, no such interfacial ion effects are observed in the NP simulations (Figure 3.7 D). Instead, an increase in pore radius simply yields a gradual increase in Cl^- density in the central pore region whilst the interfacial layer remains devoid of ions.

Na^+ is excluded from the hydrophobic interface and shows a preference for the middle of the pore within all nanopore sizes in both the ECC and NP simulations (Figure 3.7 B & E).

Given that the smaller Na^+ ions favour being fully solvated, the increase in pore radius (and hence the increase in volume of the bulk-like water regime) enables more Na^+ to retain their hydration shells leading to an increase in number density. These effects are again comparable to those seen at aqueous/air and aqueous/decane interfaces [76, 79].

The water number density profiles remain similar for both NP and ECC simulations. In the smallest nanopore, the water profiles (Figure 3.7 C & F) suggest that water is more densely packed towards the axis of the nanopore, whereas in larger nanopores the density of water becomes progressively more bulk-like with the introduction of more annular rings as the pore radius increases. Similar water structure and packing inside CNTs has been observed in MD simulations and experimentally using Raman spectroscopy methods [240, 243].

Overall, these results suggest the preferential accumulation of Cl^- ions (alongside exclusion of Na^+ ions) close to the hydrophobic pore walls are effects due to surface electrostatic interactions involving the ion rather than as a consequence of confinement *per se*. The inclusion of polarisation effects through ECC appears to not only influence the structure of water and ions inside the nanopore, but also substantially increase the Cl^- density within the interfacial layer - an effect that is not observed with NP forcefields.

3.4.3.3 Energetics of Ion Permeation

To explore the energetic influence of the preferential Cl^- adsorption to the pore walls on ion permeation, we estimated free energy profiles along the pore axis for the R_{pore} = 0.70 nm nanopore (Figure 3.8) and compared the effects of different ion models. To this end, umbrella sampling simulations were performed for both the ECC and NP forcefields to obtain symmetrised one-dimensional potentials of mean force (PMF) for

both a Na^+ and Cl^-. Convergence analysis indicated the resulting PMF profiles for Cl^- and Na^+ had converged (Figure 3.9) (i.e., taking 1 kJ/mol change between each fraction of time as an indication of convergence) for both ion parameter sets.

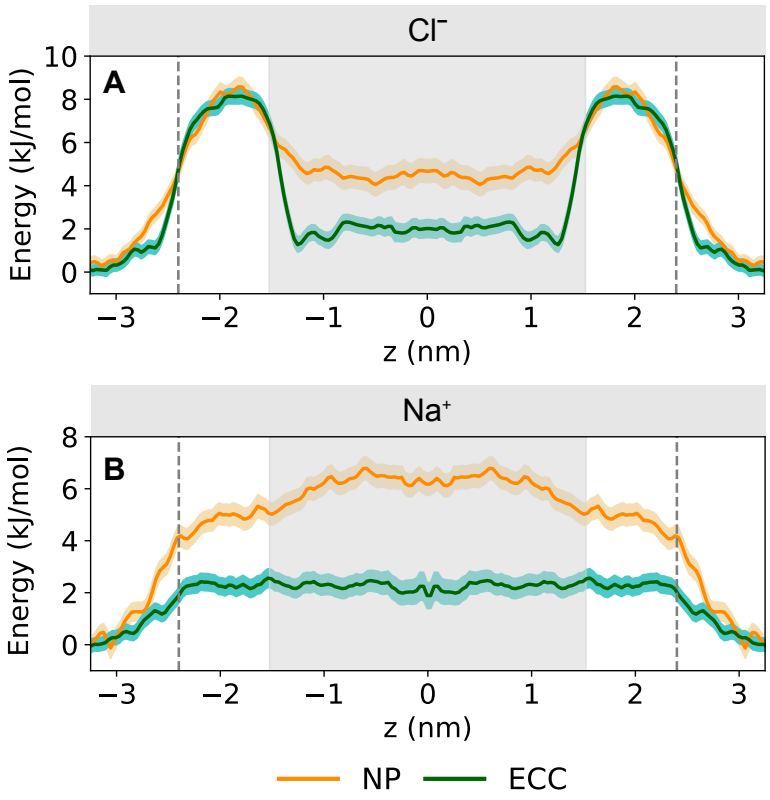

Figure 3.8 | Potentials of Mean Force Ion Profiles Single-ion PMF profiles for **A** Cl^- and **B** Na^+ permeating the model nanopore with ECC-rescaled ionic charges (green) and standard non-polarisable forcefields (yellow). The distance between the ion and the model nanopore center of mass is denoted by z, where $z = 0$ represents the center of the pore. The solid lines indicate the free energy profile calculated from the final 5 ns of each umbrella window. Confidence bands were calculated by taking the standard error over independent 1 ns sampling blocks over the time period sampled. The dashed vertical lines denote the extent of the nanopore and the grey shaded region (-1.5 < z < 1.5) represents the hydrophobic section.

For Na^+, both the NP and ECC simulations produced PMF profiles with a broad energetic barrier and a maximum in the hydrophobic center of the nanopore (i.e. at $z = 0$). This barrier height was $\sim 6.2 \pm 0.4$ kJ/mol for the NP PMF and was reduced $\sim$ 3-fold relative to the corresponding ECC PMF barrier of $\sim 2.0 \pm 0.3$ kJ/mol. Interestingly, PMF calculations for Na^+ along the length of a hydrophobic channel formed by gramicidin A have yielded a

$\sim$ 3-fold reduction in the central barrier height when comparing CHARMM27 with the AMOEBA polarisable forcefield [99] and a $\sim$ 4-fold reduction comparing CHARMM27 with the CHARMM DRUDE polarisable forcefield [98].

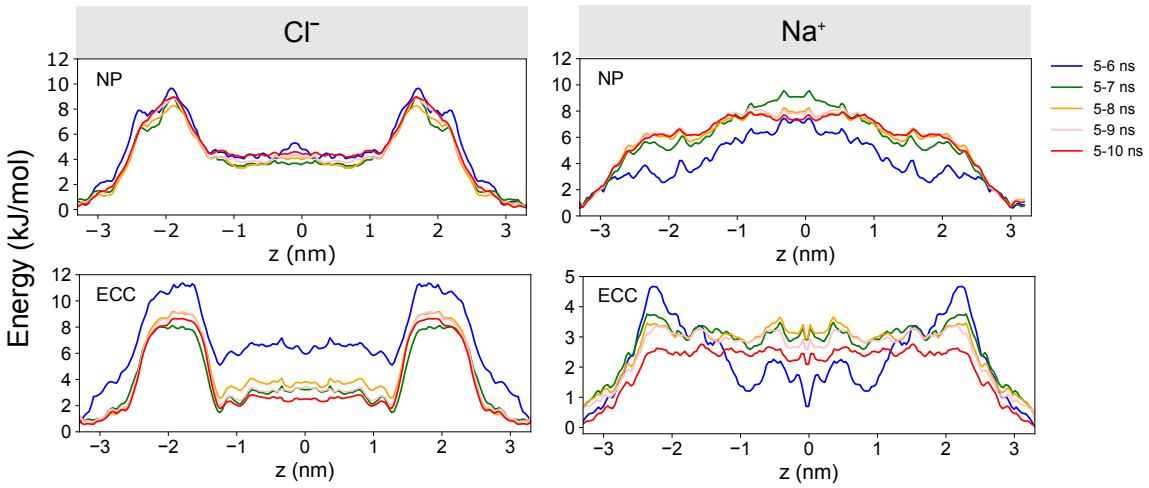

Figure 3.9 | Convergence of PMF Profiles Single-ion PMFs convergence analysis for Cl^- and Na^+ using a non-polarisable (NP) forcefield and ECC method. Convergence analysis was performed by calculating 1 ns cumulative sampling blocks over the sampling time (last 5 ns of simulation).

The shape of the Cl^- PMFs are more complex, but overall are preserved between the NP and ECC simulations. In both cases there is an energy barrier of $\sim$ 8.3 $\pm$ 0.3 kJ/mol at z = 2 nm corresponding to the narrow polar regions at the entrance to the pore (radius $\sim$ 0.4 nm; Figure 3.5 D). In this region, ions likely experience steric effects from the restrained water molecules. In addition, this energetic penalty in the polar region may result from the requirement for Cl^- to partially strip its hydration shell, releasing $\sim$ 2 water molecules that are incompletely compensated by less favourable interactions with the restrained water molecules form the pore lining (Figure 3.10). The height of this barrier is similar for both the NP and ECC PMFs.

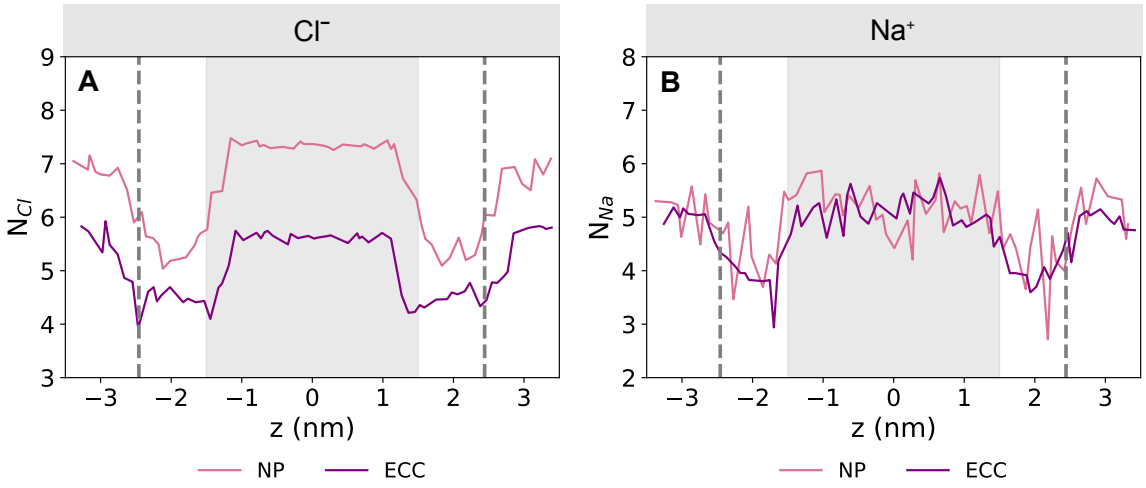

Figure 3.10 | Variation in Ion First Hydration Shell Variation in the first hydration shell of **A** Cl^- and **B** Na^+ as a function of z, the distance between the ion and the model nanopore center of mass. N_{Cl} and N_{Na} denote the coordination numbers of water-oxygen atoms in the first hydration shell of Cl^- and Na^+ respectively, with ECC-rescaled ions (purple) and a non-polarisable (NP) forcefield (pink). The dashed vertical lines represents the extent of the nanopore and the grey shaded region (-1.5 < z < 1.5) represents the hydrophobic section.

For both the NP and ECC Cl^- PMFs there is a broad energetic well centred around z = 0. However, for the NP simulations, this well is $\sim$ +4.6 $\pm$ 0.5 kJ/mol relative to the solution outside the pore, whereas for the ECC simulations the difference (bulk to hydrophobic pore region) is $\sim$ +1.0 $\pm$ 0.4 kJ/mol at z = 1 nm, rising to +2.0 $\pm$ 0.4 kJ/mol in the center. Thus, at z = 1 nm (just inside the central hydrophobic region) Cl^- is stabilised nearly 4-fold in the ECC simulations relative to in the NP simulations. This correlates well with our simulations of the model hydrophobic interface (Section 3.3.2) that suggest Cl^- is preferentially stabilised at a water/hydrophobic interface when ECC or a fully polarisable model are employed in simulations and corresponds to the number density profiles above (Figure 3.6 A & 3.7 A).

3.4.3.4 Ion Hydration Within the Pore

To examine the molecular origins from which different energetic profiles arise, we considered the changes in ion hydration at different radial locations inside the central

hydrophobic region. This region was divided into four sections radially from the axis of the pore up to the pore wall in increments of 0.175 nm (Figure 3.11 A). The following analysis was performed on the final 5 ns of each equilibrium simulation. For ions present in each region, radial distribution functions (RDFs) were computed between the ion and oxygen atoms of the water molecules, i.e. $g_{Ion-O}(r)$, for both the NP and ECC simulations. Ion-water coordination numbers were obtained by evaluating the cumulative ion-oxygen RDFs up to the first minimum, corresponding to the number of oxygens in the first hydration shell. All RDFs were calculated on systems solvated with 0.50 M NaCl solution and with nanopore radius of 0.70 nm.

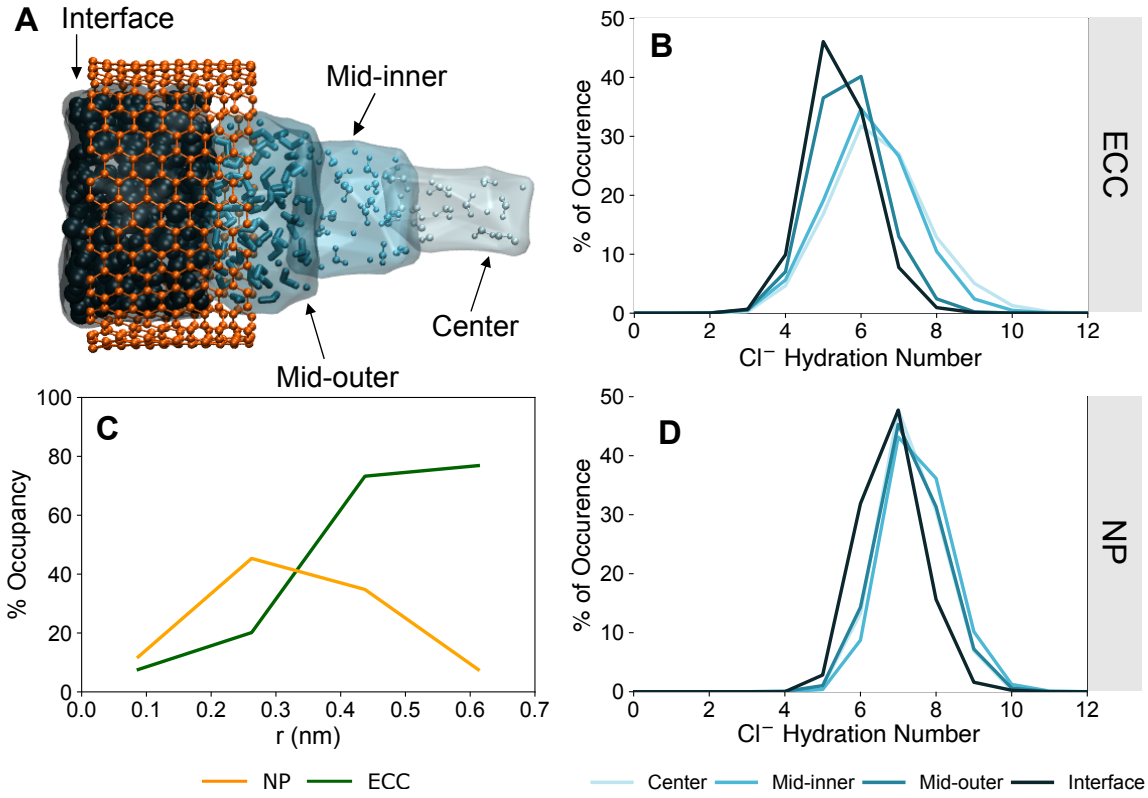

Figure 3.11 | Cl^- Hydration Structure Analysis Cl^- hydration structure inside radial sections of the hydrophobic region of the pore. **A** Schematic of the hydrophobic region of the pore divided into four 0.175 nm radial sections coloured in decreasing shades of blue. The proportion of Cl^- with various hydration numbers in defined radial regions with **B** the ECC method and **D** the non-polarisable forcefields. **C** shows the percentage occupancy of each radial section by Cl^-. With ECC (green line), Cl^- spends a significantly greater proportion of time within the interfacial layer, whereas with the non-polarisable forcefields (orange line), Cl^- tends to occupy regions away from the pore wall.

The Cl^--O RDFs display a change in radius of the first hydration shell both between ion models and ion location. For ions outside the pore, the RDFs achieve the positions of first maxima for ECC and NP models at 0.33 nm (Figure 3.12 B & C) and 0.32 nm (Figure 3.12 B & D) respectively, whilst sharing relatively similar positions for the first minima (at $\sim$ 0.39 nm). Density functional theory calculations report an RDF first peak distance between 0.31 and 0.32 nm which also agrees with experimental data [244, 245]. Similar differences have been noted in simulations comparing Cl^- in bulk solution using NP and Drude polarisable forcefields [242]. RDFs with the ECC method show a reduction in peak intensity for the first peak which is associated with a difference in coordination number between forcefields. Outside the nanopore (i.e. in bulk solution), the coordination number of Cl^- is 5.9 for ECC compared to 7.1 for the NP forcefield. The ECC value falls within range of values 5.1 - 6.3 predicted by *ab initio* calculations [229, 246–248] and 6.1 reported by first principles MD simulations [236].

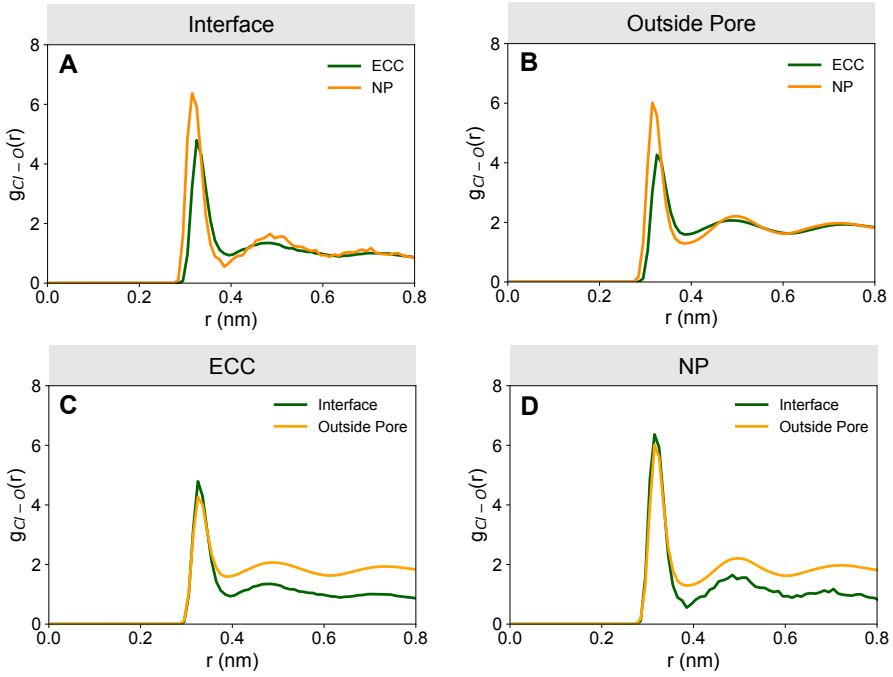

Figure 3.12 | Cl-O RDFs Radial distribution function (RDF), $g_{Cl-O}(r)$, of water oxygen atoms around Cl^- at the interface **A** and outside the pore **B** with ECC and standard non-polarisable (NP) forcefield. **C** & **D** show the same RDFs as a comparison between the interface and outside the pore for a given forcefields parameter set.

Earlier neutron scattering data report hydration numbers of 7.0 $\pm$ 0.4 [245] whilst other experiments yielded values of 5.5 $\pm$ 0.4 [249]. Therefore, the coordination numbers for bulk electrolyte predicted in this study from the ECC simulations are consistent with previous theoretical and experimental data. (We note that our normalised RDFs for outside the pore do not reach a value of 1 at large distances. This is due to the simulation box dimensions and the method of calculation which involves water-oxygen atoms from the whole system including those near the bilayer-pore complex. Therefore, this is a non-homogeneous system).

The Cl^- first hydration number distributions indicated a significant shift towards lower values for ions at the interface of the pore (i.e., within 0.175 nm of the CNT pore wall), attaining an average value of $\sim$ 5 with a significant fraction of ions with hydration numbers of $\leq$ 4.5 (Figure 3.11 B). In comparison, towards the (radial) center of the nanopore, where water density is more bulk-like, the hydration number is on average $\sim$ 6.1. Between the interface and radial center of the pore, the average hydration number ranges within these limits (Figure 3.11 B). The shape of the distributions become progressively broader with increasing distance away from the interface which indicates that the structure of the hydration shell becomes more flexible approaching more bulk-like water behaviour. In comparison, the Cl^- hydration shell is less flexible in simulations with the NP forcefield. This is reflected in the hydration number distributions which remain tighter for NP (Figure 3.11 D) compared to ECC and are centred on an average coordination number of $\sim$ 7.0 in all radial sections inside the nanopore, suggesting the hydration shell remains predominantly intact.

Furthermore, it is interesting to consider the proportion of time Cl^- spends in various regions or, in other words, the percentage occupancy of each radial section (Figure 3.11 C). With ECC, Cl^- spends a significantly greater proportion of time in the

interfacial layer whereas with the NP forcefield, Cl^- is more inclined to occupy regions away from the pore wall. These findings align with the number density distributions (Figure 3.6 A & Figure 3.7 A).

In contrast to Cl^-, the RDFs for Na^+ are in good agreement inside compared to outside (i.e. bulk electrolyte) the pore for the first hydration shell with the ECC forcefield (Figure 3.13 B), suggesting that Na^+ retains their solvation shell under hydrophobic confinement within these pores. The coordination number of Na^+ at the interface is $\sim$ 4.2, compared to outside the pore (i.e. in the bulk electrolyte) where Na^+ has a coordination number of $\sim$ 4.8 when employing ECC Figure 3.13 A).

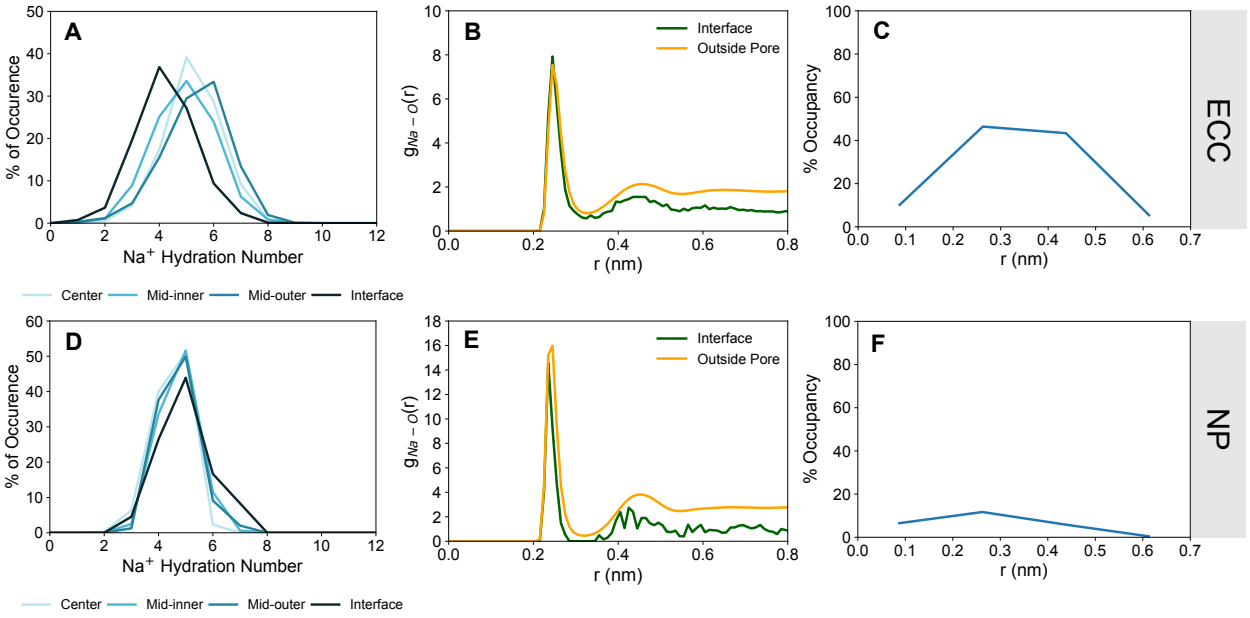

Figure 3.13 | Na^+ Hydration Structure Na^+ hydration structure inside radial sections of the hydrophobic region of the pore with the ECC method and non-polarisable (NP) forcefields. **A** & **D** show the proportion of Na^+ with various hydration numbers in defined radial regions for ECC and NP respectively. **B** & **E** show the radial distribution function of water oxygen atoms around Na^+ in the interfacial layer inside the pore (green) and outside the pore in bulk solution (yellow) for ECC and NP respectively. **C** & **F** show the percentage occupancy of each radial section by Na^+. With both ECC and NP forcefields, Na^+ tends to occupy regions away from the hydrophobic pore interface.

To compare, a coordination number of ~ 5.0 is seen for Na^+ regardless of location using the NP forcefield (Figure 3.13 D). Coordination numbers within the range 4.9 - 6.1 have previously been predicted for Na^+ in bulk water using *ab initio* methods [229, 250, 251]. Older X-ray, neutron diffraction and Raman spectroscopy experiments predict coordination numbers between 4 - 8 for Na^+ in aqueous solution [252]. The proportion of frames analysed with Na^+ present indicated that they were more likely to occupy regions away from the interface (Figure 3.13 C & F). These results reinforce the findings from number density calculations.

Overall, using the ECC parameters, Cl^- has a more flexible hydration shell and is more inclined to partially desolvate to favourably interact with the hydrophobic pore wall. Conversely, the Cl^- hydration shell remains predominantly intact using the NP forcefield and occupies regions away from the interface. Na^+ prefers to remain more solvated by moving through the nanopore closer to the pore axis where the structure of water is more bulk-like relative to the pore wall.

3.4.3.5 Robustness and Sensitivity of ECC

There are some variations in the published ECC model [78, 79, 139, 187, 212]. Specifically, the more recent ECCR model [139] includes some additional small changes to ionic vdW parameters. We therefore compared ion density profiles as a function of radial position for our nanopore models using the ECC model in which only ionic charges have been modified and using the ECCR model in which vdW radii were also adjusted. From Figure 3.14, the results show some sensitivity to the ion model employed however for all three different pore radii and ion parameters. The fundamental observation that Cl^- accumulates at the hydrophobic pore wall (Figure 3.14 A, D, G) and an associated Na^+ depletion (Figure 3.14 B, E, H) is consistent when simulated using the ECC or ECCR whereas this accumulation is not observed with the NP forcefield. The height of the Cl^-

peak at the interface differs between ECC and ECCR which suggests the extent of Cl^- accumulation at the interface is more for ECC than ECCR. Nonetheless, in all three pore models, for both the ECC and ECCR treatments, the difference in ion distributions at the pore wall (calculated as $\Delta r(Cl^- - Na^+)$, i.e. the difference in distance from the nanopore wall at which the ionic concentration rises to 50%; see Figure 3.14) is $\sim$ 0.2 nm. We are therefore confident that our observations are robust to reasonable variations in the implementation of the ECC model.

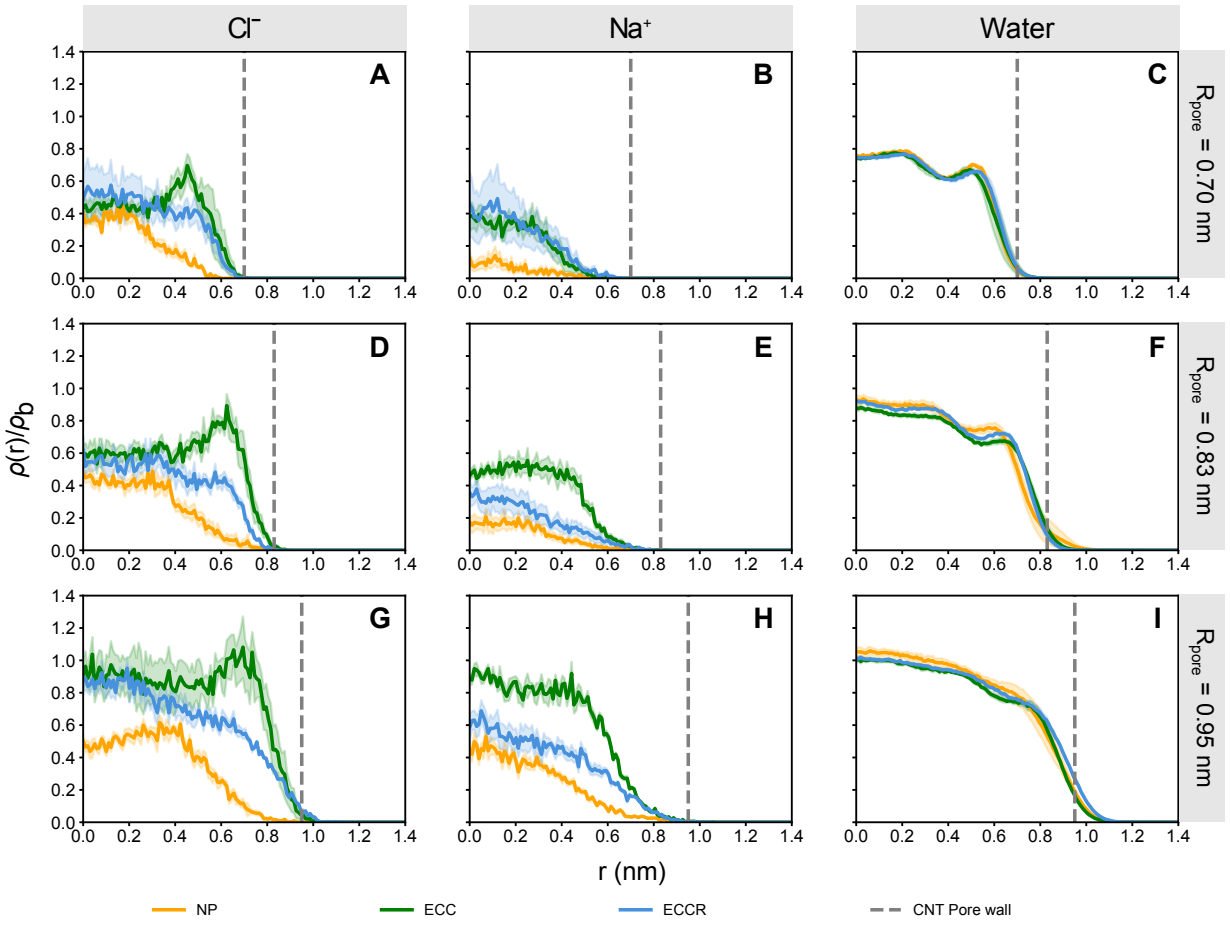

Figure 3.14 | ECCR Number Density Profile Comparison Symmetrised number density ($\rho(r)/\rho_b$) profiles of Cl^-, Na^+ and water with a non-polarisable forcefield (NP), ECC-rescaled ionic charges (ECC) and ECCR (ECC with an additional small van der Waals radii rescaling). The variable r is the radius of the nanopore which extends from 0 (pore axis) to R_{pore} (the interface where the salt solution meets the wall of the nanopore). Plots compare density distributions of nanopores with radii R_{pore} = 0.70 nm (**A - C**), R_{pore} = 0.83 nm (**D - F**) and R_{pore} = 0.90 nm (**G - I**), where the grey vertical dashed lines represent R_{pore} for respective plots.

Free energy profiles were calculated using the ECCR model as another test of robustness following the same protocol as for the ECC and NP PMFs above. As seen in Figure 3.15, the shape of the Cl^- PMFs are preserved between ECCR, ECC and NP simulations, exhibiting similar notable features corresponding to the polar regions at the pore entrances and a central barrier minimum in the hydrophobic core (z = 0 nm). For Cl^- in the polar regions, the energy barrier for ECCR is $\sim 1.0 \pm 0.3$ kJ/mol less compared to ECC and NP (Figure 3.15 A). This can be attributed to the additional vdW radius rescaling applied in ECCR and hence permeating ions likely experience fewer steric clashes with a physically smaller ion. The free energy barrier for Cl^- using the ECCR parameters in the center of the nanopore (z = 0 nm) is $\sim 4.1 \pm 0.4$ kJ/mol which is slightly lower than the NP energy at this location. This is perhaps somewhat unexpected given the number density profiles for Cl^- employing ECCR and ECC share similarities associated with the enhanced localisation of anions to the hydrophobic pore interface (Figure 3.14). For Na^+, an energetic maximum of $\sim 3.5 \pm 0.3$ kJ/mol is reached at $z \sim 1$ nm (Figure 3.15 B) where the ion begins to move into the hydrophobic core of the nanopore and away from the polar regions. The free energy at z = 0 nm is maintained at $\sim 2.5 \pm 0.3$ kJ/mol relative to outside the nanopore. Contrary to the pattern seen with Cl^-, the free energy associated with ECCR for Na^+ is closer to the PMF profile for ECC in the hydrophobic region which corresponds to the comparable number density plots in Figure 3.14. Nonetheless, for both ionic species the free energy of ions in the center of the pore approximately follows the order NP > ECCR > ECC. Thus, this additional work highlights the sensitivity of nanoscale effects to ion parameters in MD simulations.

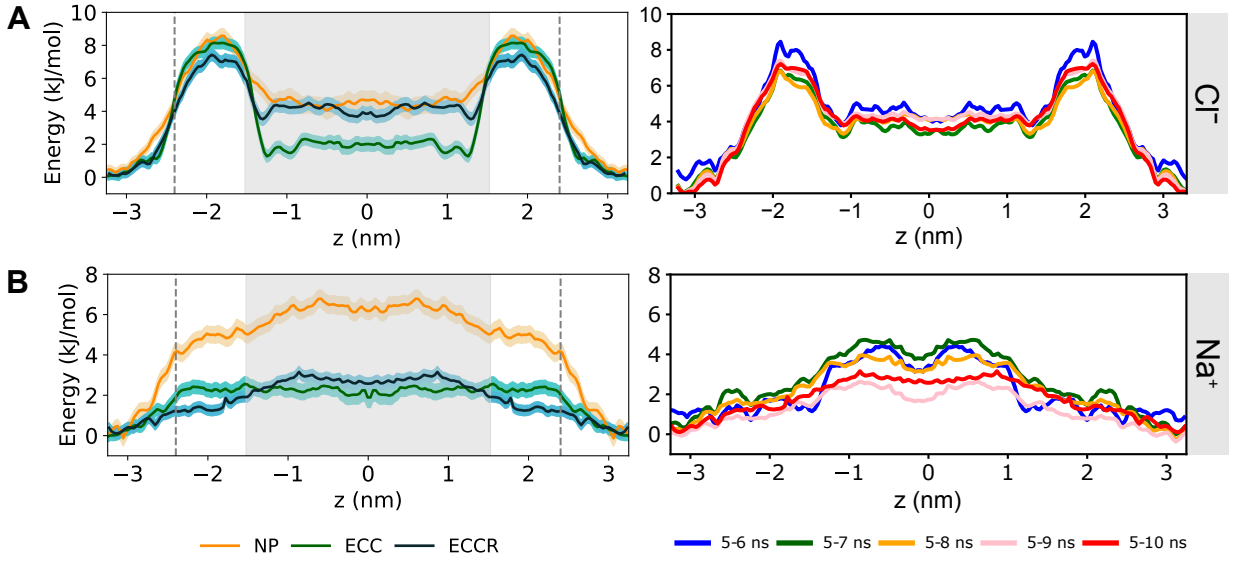

Figure 3.15 | ECCR PMF Comparison and Convergence Single-ion PMF profiles for **A** Cl^- and **B** Na^+ permeating the model nanopore with ECC-rescaled ionic charges (green), standard non-polarisable (NP) forcefields (yellow) and ECCR (ECC with an additional small van der Waals radii rescaling) (dark blue). The distance between the ion and the model nanopore center of mass is denoted by z where $z = 0$ represents the center of the pore. The solid lines indicate the free energy profile calculated from the final 5 ns of each umbrella window. Confidence bands were calculated by taking the standard error over independent 1 ns sampling blocks over the time period sampled. The dashed vertical lines denote the extent of the nanopore and the grey shaded region (-1.5 < z < 1.5) represents the hydrophobic section. The corresponding PMF convergence analysis for **A** Cl^- and **B** Na^+ using the ECCR ion parameters are also presented. Convergence analysis was performed by calculating 1 ns cumulative sampling blocks over the sampling time (last 5 ns of simulation).

Together these results suggest that our observations of local Cl^- accumulation at the hydrophobic pore wall are robust to variations in the ECC model, and that this model is capable of mimicking more computationally demanding polarisable simulations. Additionally, it is helpful to consider how well these simulations mimic the experimental realism of anions close to a graphene-like hydrophobic surface. A recent study [253] compared anion adsorption to graphene/water interfaces as measured by surface-sensitive spectroscopy and NP MD simulations. This demonstrated how the experimental free energy of anion adsorption to such a interface could be reproduced by scaling the anion (I^-) charge by ~ 0.8, as is the case in the ECC model which supports the application of ECC to our CNT-derived model nanopore.

3.5 Summary & Conclusions

Preliminary simulations of a model hydrophobic interface have laid the groundwork for both the usage of the ECC method and establishing the behaviour of Cl^- ions at a planar hydrophobic interface. This chapter moves on to focus on Cl^- behaviour in a hydrophobically confined environment within a model nanopore designed to mimic aspects of pLGICs. Number density profiles showed that the larger, more polarisable Cl^- ions preferentially reside in the outermost interfacial layer of the hydrophobic region within the pore (and at the interface), whereas the smaller "hard" Na^+ ions are repelled from the interface and occupy more bulk-like regions as they prefer to remain more fully solvated. These observations were consistent across a range of electrolyte concentrations and pore radii. Through analysis of ion hydration structure within the hydrophobic core, ion solvation as a function of radial position from the pore axis was investigated. This revealed how the mechanism of Cl^- adsorption to the hydrophobic pore wall could be associated with the partial loss of the anion's first hydration shell which was shown to be energetically favourable. Crucially, these ion and water effects were only observed when effective polarisation was modelled in the system and otherwise Cl^- and Na^+ exhibited similar behaviours.

These results contribute to the longstanding debate over forcefield accuracy and whether more explicit treatment of electrostatics is necessary to capture more physically accurate anion interactions that occur at hydrophobic interfaces [97, 254]. With the evolution of better and more powerful computational resources, extensive simulations examining interactions between Cl^- and hydrophobic interfaces using explicitly polarisable forcefields and even quantum mechanical methods may provide new mechanistic insights into anion selectivity.

"Art should comfort the disturbed and disturb the comfortable."

— Banksy

4

Anion-Hydrophobic Interactions in a Protein

Contents

4.1 Introduction 87
4.2 Objectives 89
4.3 Methods 90
4.3.1 Protein Truncation and System Preparation 90
4.3.2 Equilibrium Molecular Dynamics Simulations 92
4.3.3 AMOEBA Forcefield Simulations 93
4.3.4 Umbrella Sampling 94
4.3.4.1 Analysis 95
4.4 Results and Discussion 96
4.4.1 Influence of Effective Polarisation on the Energetics of Cl^- Ion Binding . . 96
4.4.2 Influence of Polarisation on Ion Binding Dynamics 101
4.4.3 Cl^--Hydrophobic Interactions within the Binding Site 103
4.4.4 Mechanisms of Cl^- Binding 106
4.5 Summary & Conclusions 109

This chapter is based on the publication:

"*Influence of Electronic Polarization on the Binding of Anions to a Chloride-Pumping Rhodopsin.*" **Phan, L. X.**, Chamorro, V. C., Martinez-Seara, H., Crain, J., Sansom, M. S. P. and Tucker, S. J. *Biophysical Journal*, 122, 8, 1548-1558 (2023)

4.1 Introduction

Recently, an increasing number of high-resolution structures of anion-selective ion channels and transporters have been resolved. Amongst these are members of a class of membrane transport proteins related to bacteriorhodopsins. Rhodopsins are formed from seven transmembrane domains and contain a retinal bound to a lysine residue via a Schiff base linkage that acts as the light-absorbing chromophore [255, 256].

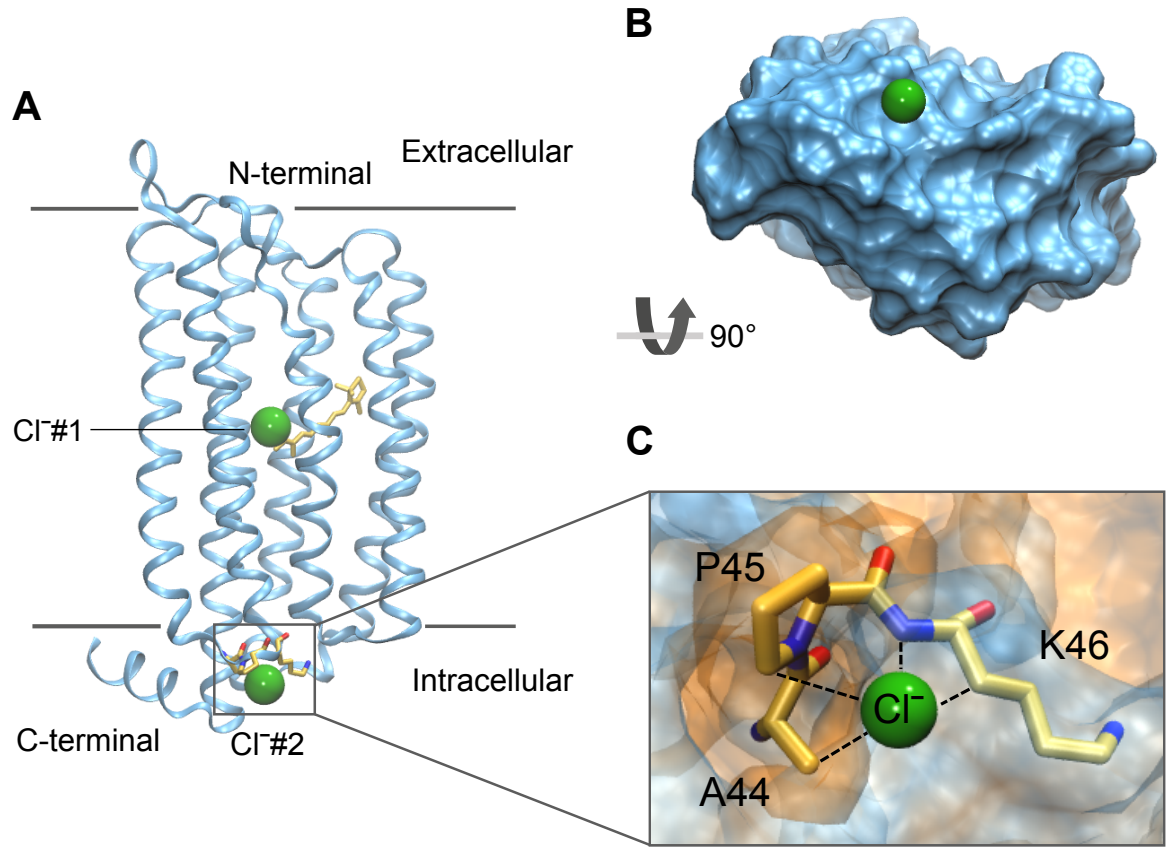

Figure 4.1 | Structure of the Cl^--Pumping Rhodopsin A Crystal structure of Cl^--pumping rhodopsin (ClR) (PDB ID: 5G28). The first Cl^- binding site (Cl^-#1) is located in the active center of the protein and is in contact with the all-*trans* retinal illustrated by the yellow stick representation in the center of the protein. The second site (Cl^-#2) is located on the intracellular side (grey box) and is the focus of this study. The grey lines depict the extent of the lipid bilayer and Cl^- ions are represented by the green spheres. **B** A surface representation of the protein rotated by 90 °to illustrate the bottom face where the Cl^-#2 site is located. The Cl^- ion is situated outside the protein. **C** A close-up illustration of the Cl^-#2 binding site. Cl^-#2 is comprised of the hydrophobic residues A44 and P45 (orange surface) and polar residue K46 (blue surface). The black dashed lines indicate the known contacts with Cl^- in the site.

In this chapter, we have chosen to explore the microbial Cl^--pumping rhodopsin (ClR) (PDB ID: 5G28, 0.16 nm resolution) [41]. ClR functions as a light-driven inward Cl^- pump. Upon photoactivation, retinal isomerisation induces a conformational change (from the all-*trans* to 13-*cis* conformation) to facilitate Cl^- transport through the central pore of the protein [256, 257]. ClR contains two Cl^- binding sites: the first (Cl^-#1) is located near the retinal in the vicinity of the Schiff base at the active center, whilst the second (Cl^-#2) is located on a cytoplasmic loop (Figure 4.1 A) where its role in the Cl^- transfer pathway is Cl^- release into the cytosol. This cytosolic binding site is of particular interest because it is primarily composed of hydrophobic residues yet is known to bind Cl^-. Here, Cl^- is coordinated by an aliphatic hydrogen of A44 (H-A44), a hydrogen from the aromatic ring of P45 (H-P45), a hydrogen from the sidechain of K46 (H-K46) and is considered to be hydrogen-bonded to the backbone nitrogen of K46 (N-K46) (Figure 4.1 C). Interestingly, similar pumping activities were observed for both Cl^- and Br^- within this structure from diffraction studies. Anomalous scattering peaks in Fourier maps of Br^- showed two strong halide ion binding sites in the ClR structure, providing confidence in the classification of these sites as halide ion binding sites [41]. Despite the availability of such structures and the breadth of experimental studies, molecular mechanisms of ion transport are still not well understood and are under-explored [257] in part due to the computational challenges faced in the molecular modelling of polarisable components.

As discussed in Chapter 3, Cl^- ions can form favourable interactions with hydrophobic planar interfaces and model nanopores by means of partially shedding the first hydration shell of the ion to make direct contact. Fundamentally, these behaviours arise from induced polarisation that are not fully captured within non-polarisable forcefields. Ongoing developments in hybrid methods (namely charge-scaling approaches) that possess the computational efficiency of additive forcefields, contain more physically accurate parameters which better represent electronic responses. Here, we implement

the prosECCo forcefield [190] (see Section 2.2.2.2) that encapsulates the charge-scaling of the ECC method in combination with appropriate changes to vdW radii. An alternative method is the NBFIX (Section 2.2.2.3) approach that alters the LJ parameters between charged atoms [258] and has shown to influence the binding dynamics of a Cl^--specific transporter CLC-ec1 consistent with experiment [104]. Therefore, these optimised models can be shown to demonstrate more realistic dynamics.

4.2 Objectives

Having established the applicability of the ECC approach and examined the behaviour of Cl^- ions within a hydrophobic model nanopore in the preceding chapter, this chapter now turns to focus on Cl^- interactions within a more complex, biological hydrophobic binding site (Cl^-#2) from ClR. Here, we examine the free energy landscapes of Cl^- moving out of the Cl^-#2 site into bulk solution. Associated Cl^- binding dynamics were explored within a smaller protein fragment that mimics this site using the full polarisable AMOEBA forcefield, the mean-field prosECCo forcefield and the c36 additive forcefield. Finally, we investigated the role of hydrophobic residues of an adjacent loop on the mechanism of Cl^- rebinding.

This chapter aims to address the following questions:

1. How does electronic polarisation influence the behaviour of Cl^- in a biological binding site?

2. Are Cl^--hydrophobic interactions important to the molecular mechanisms of ion transport?

4.3 Methods

4.3.1 Protein Truncation and System Preparation

One of the main aims of this chapter was to explore the influence of modelling different "coarseness" levels of polarisation on the behaviour of Cl^- ions in a biological setting. Therefore, it was of interest to perform simulations using a more fine-grained full polarizable forcefield (AMOEBA). However, there are considerable methodological and computational demands of implementing such higher level forcefields which pose barriers to running sufficiently long simulations of these larger systems. As a possible solution, we have designed reduced systems containing only the binding site of interest and a few surrounding residues that neither interact with, nor obstruct the Cl^- pathway.

To this end, the original protein structure was truncated to form two protein fragment systems both containing residues 44 to 46 which represent the hydrophobic binding site of interest (Cl^-#2) (Figure 4.2 B). The first fragment system, referred to as the "small fragment", was composed of residues 42 to 49 from the full protein which represented the binding site and a few extra surrounding residues (Figure 4.2 C). The second fragment system consisted of the small fragment with additional residues 252 to 256 from a nearby loop (Figure 4.2 D). We refer to this model as the "larger fragment". These protein fragments were solvated in $\sim$ 0.15 M NaCl and position restraints were applied to the C_α atoms with a force constant of 1000 kJ/mol/nm^2 to ensure the structures did not deviate significantly from their conformation in the parent protein structure. Simplified models are able to retain some structural complexity of the complete protein whilst being sufficiently simple to enable in-depth study of isolated regions of interest [100, 259].

Nonetheless, it should be emphasised that these models can only capture certain aspects of the full protein. These fragment systems substantially reduced the system size, making both systems more computationally feasible to run on the available computational resources (Table 4.1).

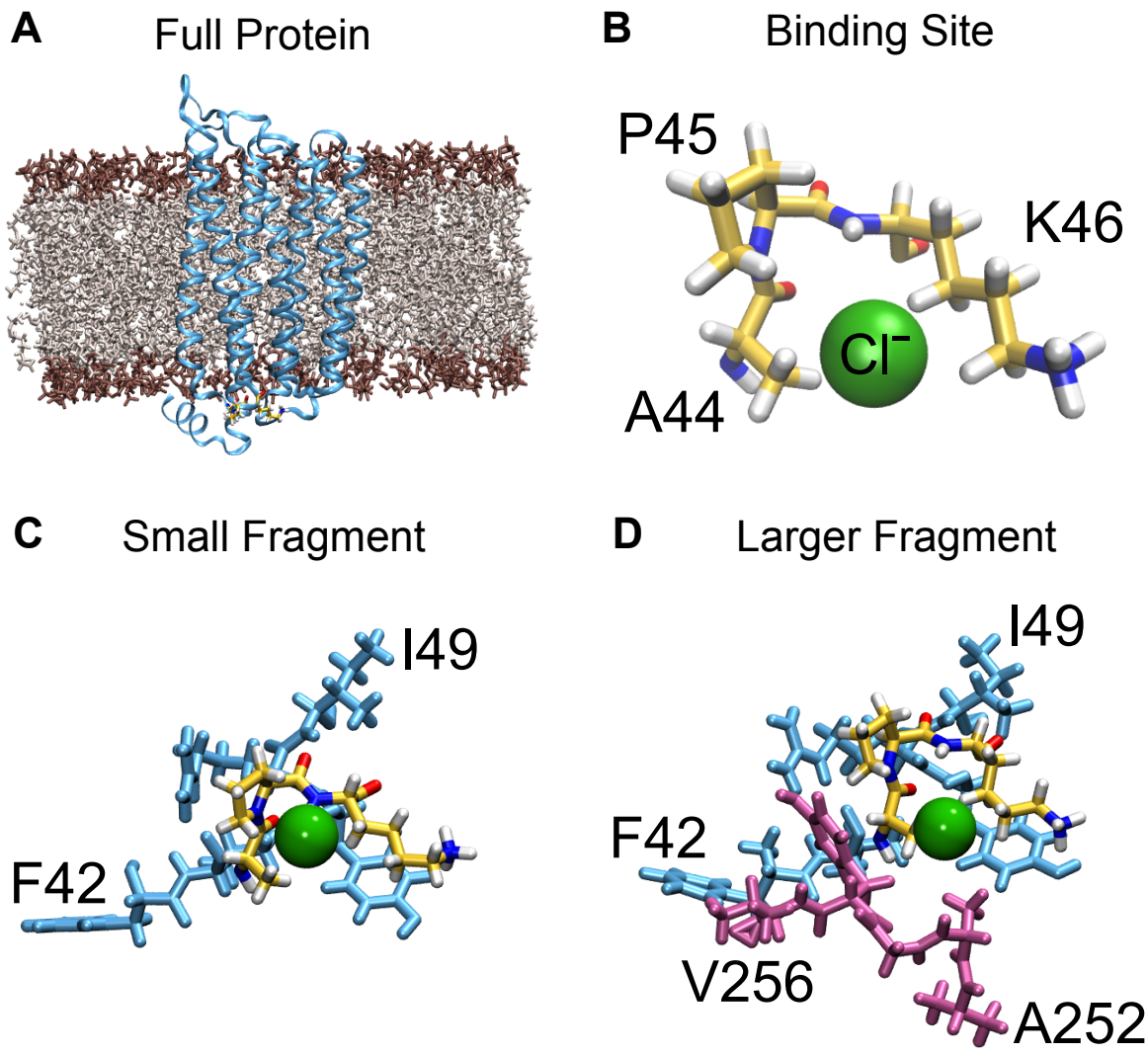

Figure 4.2 | Definition of Systems Prepared A Full ClR protein (blue) (PDB ID: 5G28) embedded in a POPC lipid bilayer (lipids head groups in darker brown, lipid tails in light brown) and binding site in licorice representation. **B** Binding site (Cl^-#2) composed of residues A44, P45 and K46. The green sphere denotes the Cl^- ion initially bound to the structure. **C** Small protein fragment consisting of residues 42 to 49 from the original protein (blue) and inclusive of the binding site (yellow). **D** Larger protein fragment consisting of residues 42 to 49 (small fragment, blue) and residues 252 to 256 from a nearby loop (pink). Water and ions are omitted for clarity.

System	Size	Forcefield	ns/day
Full protein embedded in POPC bilayer	$\sim$ 82300	c36	110
		prosECCo	107
		AMOEBA	N/A
Fragments in solution (small and large)	$\sim$ 12100	c36	690
		prosECCo	697
		AMOEBA	28

Table 4.1 | Table of Forcefield Performance Comparison Computational performance comparison of different system sizes per forcefield. All systems were benchmarked using 1 GPU (NVIDIA Quadro RTX8000), 1 node with 6 cores (CPUs: Intel Xeon Platinum 8268). c36 and prosECCo have very similar computational costs by running on GROMACS 2020. AMOEBA was run with the OpenMM 7.4.2 implementation which enables GPU acceleration and is considerably more computationally expensive relative to c36 and prosECCo

For simulations of the full protein using c36 and prosECCo (Figure 4.2 A), we started from the experimentally determined protein structure (PDB ID: 5G28) [41]. All ligands including the retinal, were removed. We believe that it was not necessary to parameterise and simulate the retinal for the purposes of this study as we were only interested in the molecular interactions occurring at the cytosolic binding site far from the active center. The whole protein was embedded in a POPC bilayer and prepared using the CHARMM-GUI protocol [109, 260, 261]. Unlike the fragment simulations, we chose not to apply position restraints to the full protein simulations to allow for unrestricted movement.

4.3.2 Equilibrium Molecular Dynamics Simulations

Prior to setting up simulations, the protonation states of ionisable residues were predicted using the PROPKA3 tool [262, 263], revealing that no residues were required to be modelled in a non-standard form. This was particularly important to check as one of the constituent residues of the binding site is a lysine (K46).

For comparison of polarisation modelling methods, non-polarisable atomistic simulations were performed using CHARMM36 (c36) with associated lipid parameters where relevant [264]. Simulations with the mean-field approach to including polarisation were achieved using the prosECCo forcefield (https://gitlab.com/sparkly/prosecco/prosECCo75) [190]. Finally, explicit polarisation was modelled in simulations using the AMOEBA forcefield (see Section 4.3.3).

The c36 and prosECCo simulations were assembled following the same protocol. These systems were equilibrated through a staged protocol whereby structural restraints on the protein backbone and Cl^- were gradually weakened over a 15 ns equilibration period. Subsequently, all production runs of 55 ns were performed with an integration timestep of 2 fs. All systems were solvated with $\sim$ 0.15 M NaCl using the SPC/E water model [223]. The temperature was maintained at 303.15 K with a coupling constant of 1.0 ps using the Nosé-Hoover thermostat. Pressure was maintained at 1 bar using the Parrinello-Rahman barostat with a coupling constant of 5.0 ps. Semi-isotropic pressure coupling was used for systems containing a bilayer and isotropic pressure coupling for systems of proteins in solvent. The Verlet cutoff scheme [150] was applied and electrostatics were treated with the particle mesh Ewald method [164]. The LINCS algorithm was used to constrain H-containing bonds only [149]. Three independent repeats were carried out for each system. These simulations were performed in GROMACS [144] version 2020 (www.gromacs.org) and analysed using MDAnalysis [226, 227].

4.3.3 AMOEBA Forcefield Simulations

Polarisable atomic multipole simulations were carried out in OpenMM 7.4.2 (www.openm.org). All components were modelled with the AMOEBA polarisable forcefield using the AMOEBA13 protein parameter set [171] and the AMOEBA03 water model [167]. Starting configurations for these simulations were taken from the final

frames after 15 ns of equilibration using c36 as described above. Simulations were then set up using a similar method to that described previously in Chapter 3 (https://github.com/Inniag/openmm-scripts-amoeba) [100]. Simulations using the AMOEBA forcefield were performed only for the small and large fragments in solvent therefore all (C_α) atoms were placed under a harmonic restraint with force constant 1000 kJ/mol/nm^2 to prevent the reduced structures deviating from the experimentally determined structure. Production runs lasted 55 ns and time integration was achieved using the r-RESPA method [182] with an outer timestep of 2 fs and an inner timestep 0.25 fs. The temperature was maintained at 303.15 K using the Andersen thermostat and the pressure was maintained at 1 bar using the isotropic Monte Carlo barostat. Three independent repeats were performed for each system.

4.3.4 Umbrella Sampling

To assess the energetics of a Cl^- ion moving from the crystallographic binding site into bulk solution, one-dimensional potential of mean force (PMF) profiles were obtained. These simulations were carried out on the whole protein embedded in a POPC bilayer and solvated with $\sim$ 0.15 M NaCl solution using the c36 and prosECCo forcefields in combination with the SPC/E water model. Systems were simulated with and without the use of backbone restraints. The collective variable (*CV*) was defined as the distance between the ion and the center of geometry (COG) of the protein binding site (residues 44 to 46, Cl^-#2) in the negative direction parallel to the z-axis of the simulation box. Starting configurations for the umbrella windows were obtained from equilibrated simulations. The target ion was relocated to subsequent positions parallel to the z-axis and the lateral positions of the ion were set to that of the binding site COG. During equilibration and sampling, a harmonic biasing potential of 1000 kJ/mol/nm^2 was applied to restrain the collective variable. Umbrella windows covered the distance from the ion binding site into the bulk water regime. This corresponded to 50 windows with a 0.05 nm distance interval

between two successive windows. Following ion relocation, 10 steps of energy minimisation were performed to remove any steric clashes between the target ion and surrounding atoms. A further 1.0 ns of isothermal-isobaric equilibration was simulated per window. Each umbrella window was simulated for 1 ns with c36 and 2.5 ns with prosECCo. Simulation details were similar to those detailed above for MD simulations. Unbiasing was performed through WHAM: the Weighted Histogram Analysis Method using the Grossfield lab implementation in version 2.0.9 (http://membrane.urmc.rochester.edu/wordpress/?page_id=126"). PMF profiles were shifted so that the ion in bulk solution corresponds to 0 kJ/mol. The PMF errors were estimated through 100 trials of Monte Carlo bootstrapping error analysis. PMF convergence was assessed by calculating 200 ps and 500 ps sampling intervals over the 1.0 ns and 2.5 ns simulations for c36 and prosECCo respectively (see Figure 4.5).

4.3.4.1 Analysis

To determine whether an ion was classified as "bound" to a residue/contact or not, ion-residue interactions were first filtered by defining an initial cutoff. An ion was considered within binding range if it resided within a spherical radius of 0.5 nm of a given residue or set of residues. This initial distance was appropriately chosen based on halide interactions within other protein binding sites [40]. Following this, distances between specified residue contact atoms and ions selected from the initial filtering were analysed using MDAnalysis [226–228]. These ion-contact distances were plotted. Ions were considered bound to the specified contacts if interactions occurred at a distance of approximately the first hydration shell of the ion, a distance realised through calculating the ion-oxygen RDF ($\sim$ 0.39 nm for all forcefields).

Simulation trajectories and binding events were visualised and inspected using VMD [221]. The analysis of hydration shell and hydration number followed the same protocol as that detailed in Section 3.4.2.3 of Chapter 3.

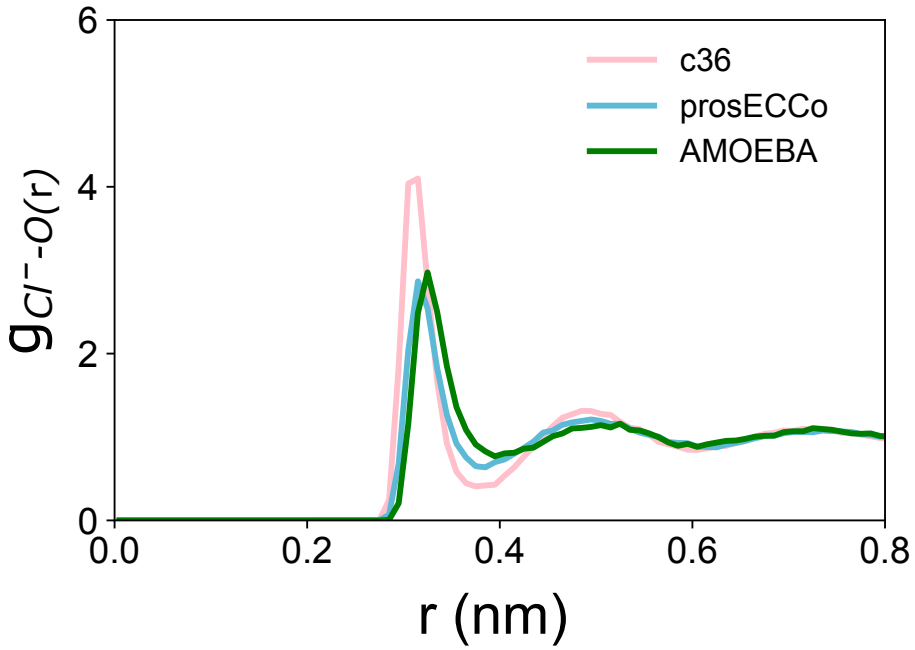

Figure 4.3 | Cl^--O RDFs Radial distribution functions (RDFs) $g_{Cl^- - O}(r)$, of water oxygen atoms around a Cl^- ion in bulk solution with the c36, prosECCo and AMOEBA forcefields.

4.4 Results and Discussion

4.4.1 Influence of Effective Polarisation on the Energetics of Cl^- Ion Binding

We firstly examined the energetics of Cl^- and Na^+ in the binding site and the influence of including effective polarisation by means of the prosECCo forcefield in comparison with the additive c36 forcefield. Additionally, we compared how the energetics vary when backbone restraints were applied to prevent the protein from deviating from the experimental structure. Examination of the umbrella histograms revealed good overlap between windows and comparison of cumulative PMF profiles indicated that the resulting PMF profiles had converged (Figure 4.5).

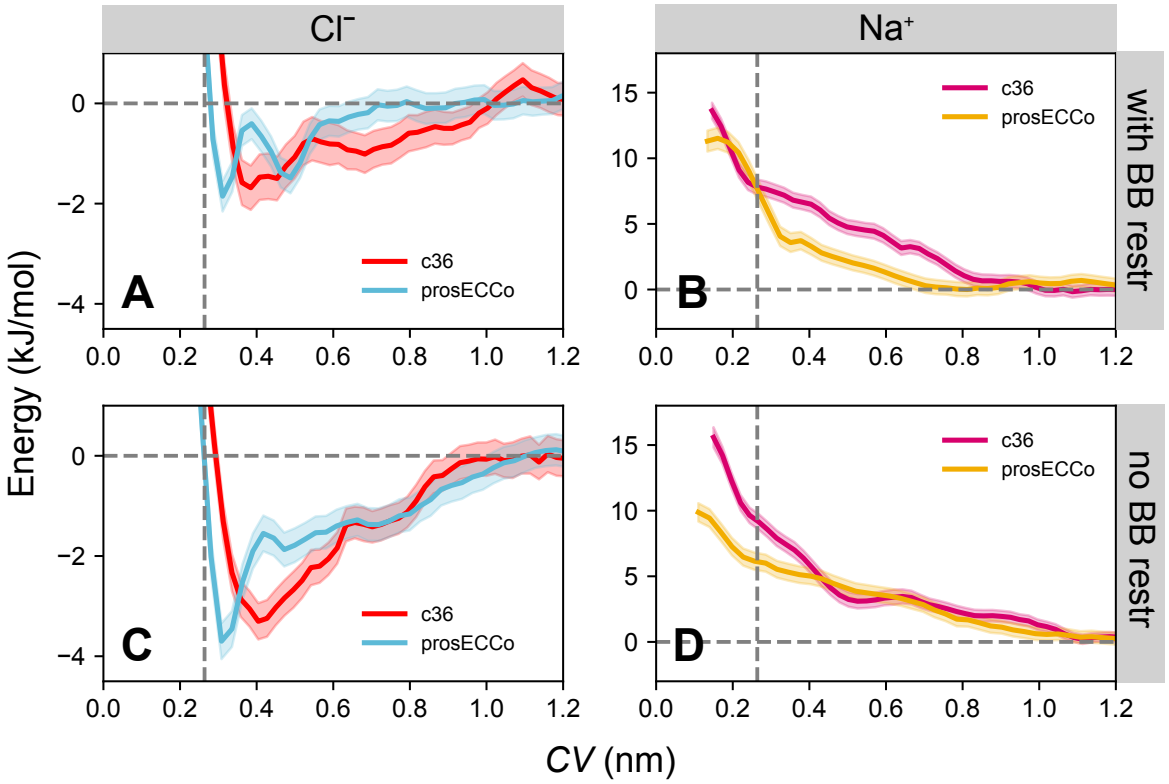

Figure 4.4 | Potential of Mean Force Profiles Potential of mean force profiles of a Cl^- (blue and red) and Na^+ (pink and yellow) ions moving away from the binding site into bulk solution using the c36 and prosECCo forcefields. The collective variable (*CV*) is defined as a straight line parallel to the z-axis starting from the center of geometry of the binding site and extending into bulk solution. Free energy profiles of simulations with (**A** & **B**) and without (**C** & **D**) the application of protein backbone restraints were calculated. Confidence bands were obtained by calculating the standard error over 200 and 500 ps sampling blocks for c36 and prosECCo respectively. The horizontal grey dashed line represents the free energy of a given ion in bulk solution. The vertical grey dashed line indicates the location of the Cl^- ion bound in the experimentally determined crystal structure in terms of *CV*.

The PMF profiles with backbone restraints using the prosECCo forcefield reveals a free energy minimum of $\sim$ -1.9 $\pm$ 0.3 kJ/mol at $CV \sim 0.30$ nm, a distance of ~ 0.06 nm further from the binding location of the Cl^- ion in the crystal structure (Figure 4.4 A). With c36, there is an energy well of $\sim$ -1.7 $\pm$ 0.4 kJ/mol situated at $CV \sim 0.40$ nm which corresponds to a distance of $\sim$ 0.13 nm relative to the experimentally determined ion location. Interestingly, the prosECCo free energy profile exhibits a distinctive second local free energy minimum at $CV \sim 0.50$ nm with a free energy $\sim$ -1.5 $\pm$ 0.3 kJ/mol. This second local minimum is less pronounced when backbone restraints are not applied

(Figure 4.4 C) yet is within error of the PMF profile with backbone restraints which was not observed with c36. On closer inspection of the Cl^- ion at this position, this second energetic well (Figure 4.4 A) appears to be associated with the displacement of the Cl^- ion from the binding site by a water molecule. The ion then proceeds to form contacts with the methyl group of A253 and hydrogen atoms from the aromatic ring of Y255 (see Section 4.4.4).

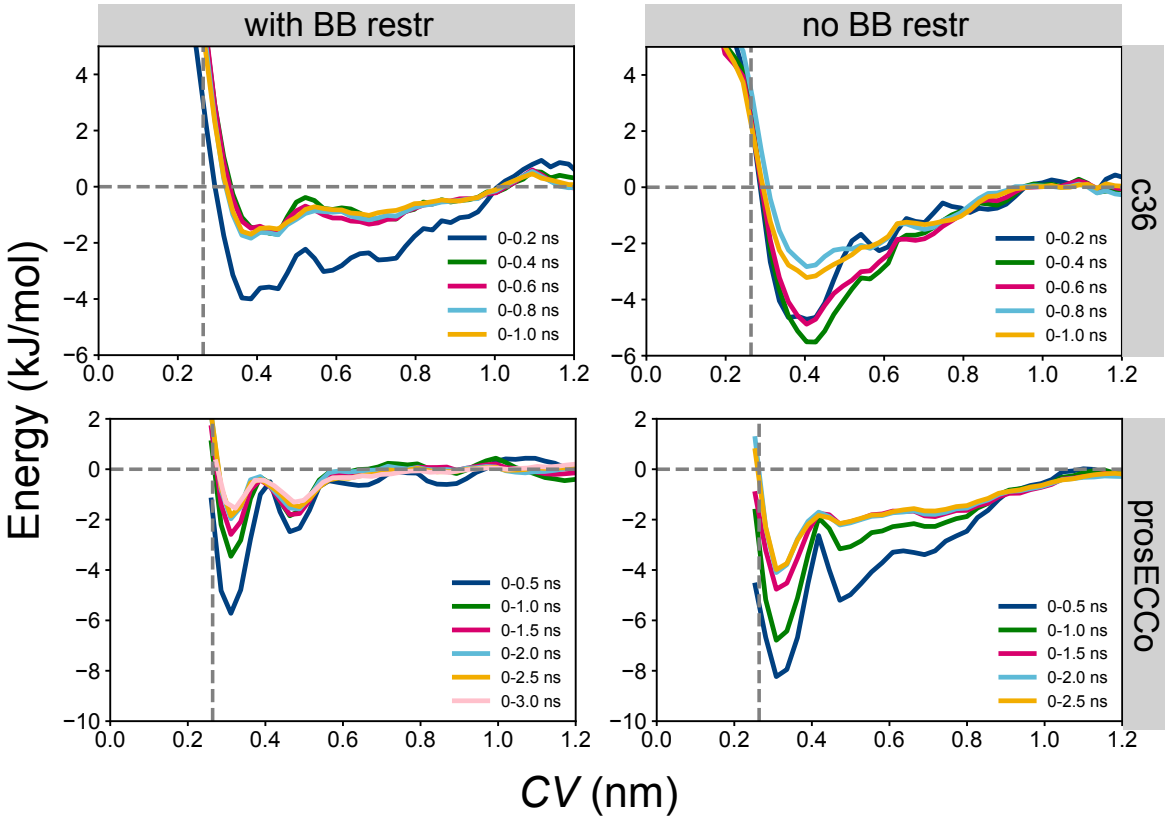

Figure 4.5 | Potential of Mean Force Convergence Tests for Cl^- Profiles Cl^- PMF convergence analysis using the CHARMM36 (c36) and prosECCo forcefields with and without the use of protein backbone restraints. Convergence analysis was performed by calculating 0.2 ns and 0.5 ns sampling blocks over the sampling time for 1.0 ns and 2.5 ns for c36 and prosECCo respectively. The vertical grey dashed line represents the position of the crystal structure binding site in terms of *CV*.

When backbone restraints are relaxed, enabling protein flexibility and unrestrained dynamics, different free energy landscapes are observed. For the prosECCo forcefield, a free energy minimum of $\sim$ -3.7 $\pm$ 0.4 kJ/mol relative to bulk solution at $CV \sim 0.30$ nm can be observed. This corresponds to a distance of $\sim$ 0.06 nm relative to the crystal structure

(Figure 4.4 C) consistent with the location of the first minimum of the PMF profile implementing backbone restraints which suggests that the conformation of the binding site does not deviate much from the crystal structure when the ion is bound. With the c36 forcefield, a profile shift is exhibited with an energy well of $\sim$ -3.3 $\pm$ 0.4 kJ/mol relative to bulk solution at a *CV* $\sim$ 0.40 nm (Figure 4.4 C), $\sim$ 0.16 nm from the crystal structure. Here, the location of the free energy minimum corresponds to a dissociated Cl^- ion forming favourable interactions with a nearby lipid head group, hence the crystallographic binding site goes undetected without the inclusion of polarisation.

In both cases (with/without backbone restraints), the interactions with Cl^- occur more closely to the crystal structure ion binding site when employing prosECCo than c36 which could be associated with partial ion dehydration in the site (discussed in Section 4.4.3, Table 4.2). We find approximately two-fold stronger binding when no backbone restraints are applied (Figure 4.4 B), an effect that is consistent across both forcefields.

By comparison, the energetics of a cation within the site were explored. The smaller, positively charged Na^+ ion PMF profiles do not present a free energy minimum which indicates that Na^+ does not bind to the site (Figure 4.4 B & D). This observation holds for both forcefields regardless of the implementation of backbone restraints, confirming that indeed this site is an anion binding site.

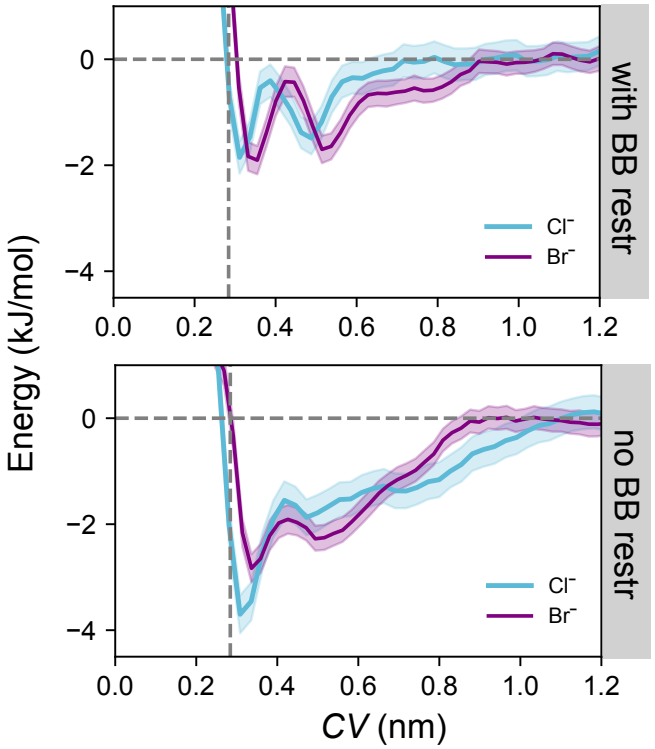

Figure 4.6 | Potential of Mean Force Profiles of Cl^- vs Br^- Potential of mean force profiles for a Cl^- and Br^- ion moving away from the binding site into bulk solution employing the prosECCo forcefield. Simulations with protein backbone restraints applied (top) and without (bottom) were explored. The horizontal grey dashed line represents the free energy of the ion in bulk solution and the vertical grey dashed line is representative of the location of the corresponding ion in the resolved crystal structure.

To locate the ion-binding sites, an anomalous signal showed ClR exhibits similar ion-pumping activity with Cl^- and Br^- [41]. Therefore, we also performed analogous PMF calculations for a Br^- ion moving out of the binding site into bulk solution using the prosECCo forcefield (Figure 4.6). Comparing these PMF profiles, the larger, more polarisable Br^- ion binds in similar locations as the Cl^- and with similar relative free energies with backbone restraints applied, thus supporting experimental observations. A second minimum is consistently observed at $CV \sim 0.50$ nm in the corresponding Br^- profile with backbone restraints. However, unlike the Cl^- profile without backbone restraints, there is a shallow second minimum in the Br^- profile. Upon visualisation, this minimum is associated with interactions with A253 from the adjacent loop. This suggests more residues may be involved in the binding of a halide ion to this site and is explored further in Section 4.4.4.

4.4.2 Influence of Polarisation on Ion Binding Dynamics

We next examined the dynamics of ion binding using the c36, prosECCo and AMOEBA forcefields applied to simulations of the small protein fragment in solution. Within the binding site, the key site contacts thought to stabilise Cl^- binding are illustrated in Figure 4.7 A [41]. Analysis of the distances between Cl^- in simulations and these key contacts were calculated. We considered Cl^- to be "bound" if it resided within the distance defined by the first hydration shell, $\sim$ 0.39 nm for all forcefields (see Section 4.3.4.1). In this section we focused only on distances between Cl^- and an aliphatic hydrogen atom from A44 (H-A44) (Figure 4.7 A).

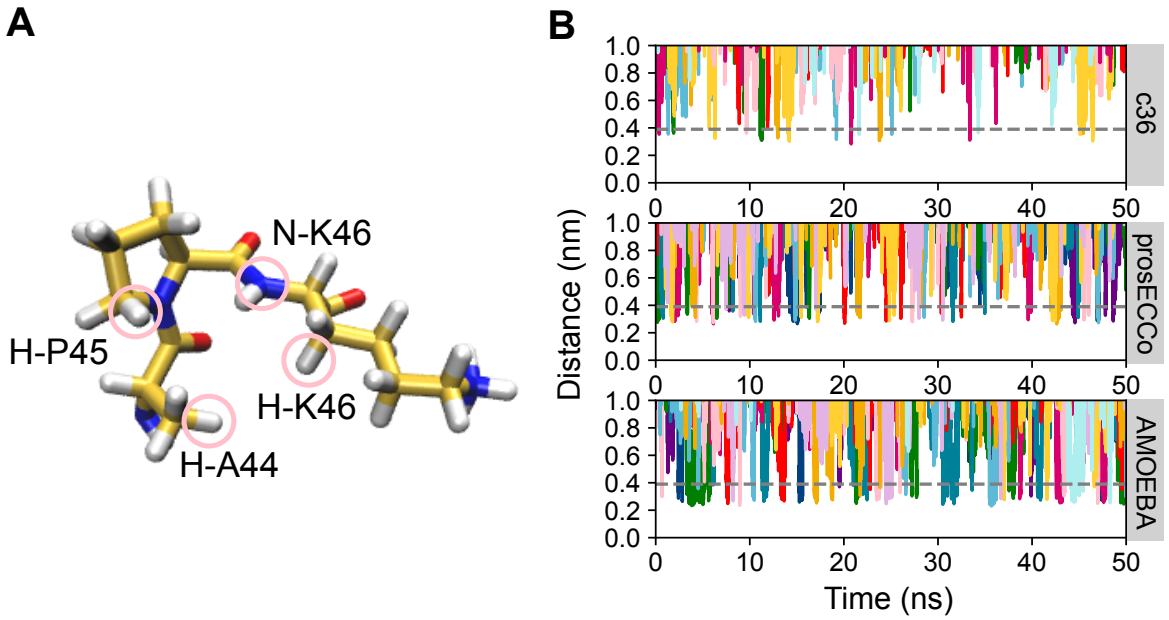

Figure 4.7 | Ion-distance Plots A Schematic diagram of hydrophobic binding site (Cl^-#2). The atoms circled in pink correspond to the atoms used for Cl^- ion-distance calculations. **B** Ion-distance plots from Cl^- ions to an aliphatic hydrogen atom of A44 over 55 ns within simulations of the smaller protein fragment. Each colour represents an individual ion trajectory, plotting only the ions that initially come into proximity of 0.50 nm to the binding site. The grey dashed line indicates the binding distance taken as the Cl^--oxygen atom of water, i.e., the distance of the first hydration shell (0.39 nm).

In simulations with the c36 forcefield (Figure 4.7 B), Cl^- ions occasionally approached the binding site to interact momentarily before dissociating back into solution. These interactions were brief with an average duration of < 0.5 ns and a minimum interaction

distance of $\sim$ 0.30 nm, just within the first hydration shell distance (Figure 4.7, grey line). In comparison, considerably more Cl^- ions come into proximity of the binding site in simulations using prosECCo forcefield yet the dynamics remain relatively transient with Cl^- spending $\sim$ 0.5 ns interacting with H-A44 in the binding site at a minimum distance of $\sim$ 0.29 nm. With the AMOEBA forcefield, a similar number of ions approach the binding site as with prosECCo however, the approach distance is even closer at $\sim$ 0.25 nm and the duration of occupancy is $\sim$ 2 ns.

Together, this data suggests that polarisability leads to stronger interactions between Cl^- ions and A44 in comparison to standard fixed charge models. The short interaction times observed here are expected because the functional role of this binding site is to aid ion release into the cytosol. Simulations and analysis of the complete membrane-embedded protein systems with the c36 and prosECCo forcefields qualitatively support these conclusions (Figure 4.8). Estimated transport rates in other rhodopsins span two orders of magnitude due to experimental challenges in measuring slow transport processes [265].

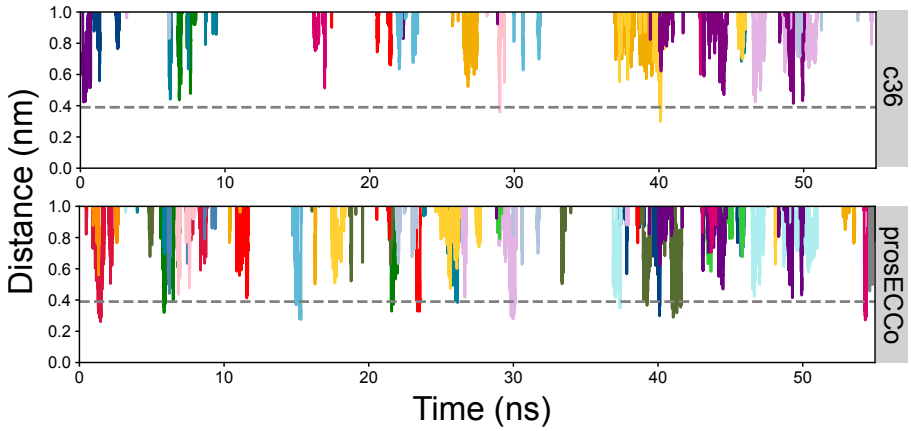

Figure 4.8 | Ion-distance plots of Cl^--H-A44 Ion-distance plots between Cl^- ions to an aliphatic hydrogen atom of A44 within simulations of the full protein embedded in a lipid bilayer with the c36 (top) and prosECCo (bottom) forcefields. Each colour represents an individual ion trajectory, plotting only the ions that initially come within a distance of 0.50 nm to the binding site.

4.4.3 Cl^--Hydrophobic Interactions within the Binding Site

Further analysis of the small fragment simulations were carried out to examine the contribution of other key coordinating atoms within the site to Cl^- binding. We computed the distances between the ion and hydrogen atoms from A44 (H-A44), P45 (H-P45) and K46 (H-K46) as well as the backbone amide nitrogen from K46 (N-K46) (Figure 4.7 A). This analysis was prepared for each forcefield and enabled us to gain a more comprehensive understanding of the molecular mechanisms behind Cl^- binding to this hydrophobic site.

Overall, a notable trend was found across all forcefields: the Cl^- comes into closest proximity with the hydrophobic contacts of the binding site, namely H-A44 and H-P45 (Figure 4.9 A-C) which are approximately equidistant. This suggests that it is more favourable for Cl^- to associate with these hydrophobic contacts [82]. Cl^- localisation was achieved through the partial loss of the ion's hydration shell; an effect that was previously seen in Chapter 3 for Cl^- ions within the model hydrophobic nanopore. As a consequence of modelling polarisation and the resulting induced dipoles (in AMOEBA), the balance between ion-water and ion-protein interactions therefore shifts towards favouring the interactions with the hydrophobic protein contacts over water. This can be realised by calculating the average hydration numbers around Cl^- ions within ~ 0.39 nm of the selected binding site atoms for each given forcefield (Table 4.2).

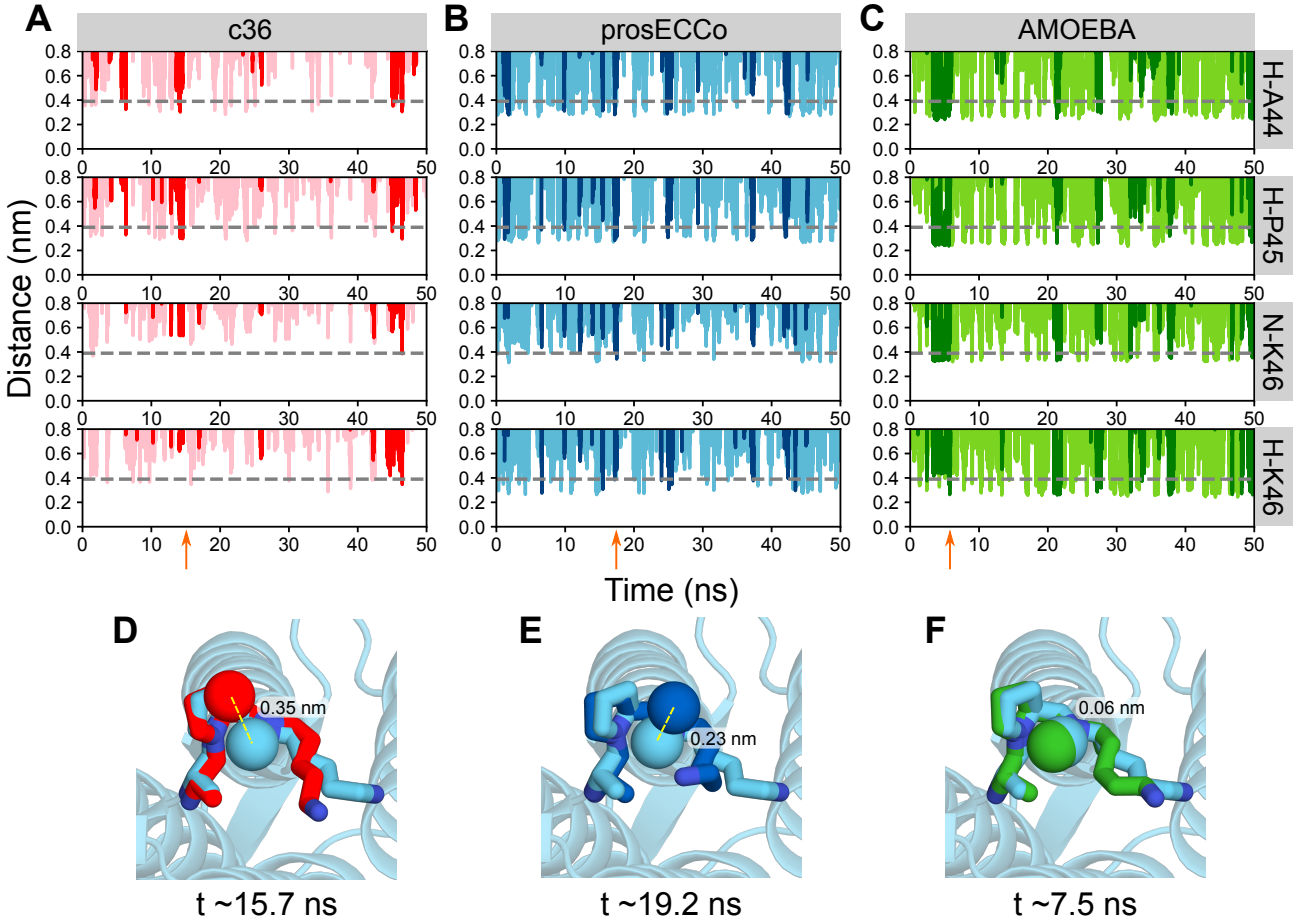

Figure 4.9 | Ion-distance Plots as a Function of Forcefield Examples of single Cl^- ion-distance plots (darker colours). Other Cl^- ion-distance plots are also shown as a function of time (lighter colours). These are illustrated for each forcefield: the **A** c36, **B** prosECCo and **C** AMOEBA forcefields. The ion distances are measured between a Cl^- ion to an aliphatic hydrogen atom of A44 (H-A44), a hydrogen of the aromatic ring of P45 (H-P45), the backbone amide nitrogen of K46 (N-K46), and a hydrogen from the sidechain of K46 (H-K46) (see Figure 4.7 A). Grey dashed lines indicate the binding distance taken as the Cl^--oxygen atom of water, i.e., the distance of the first hydration shell 0.39 nm. Snapshots of the crystal structure Cl^- ion (cyan sphere) in the Cl^-#2 binding site (licorice representation, cyan) are shown as a reference to the binding site conformation and Cl^- localisation in simulations with **D** c36 (red), **E** prosECCo (dark blue) and **F** AMOEBA (green) forcefields. The orange arrows indicate the time points where the representative snapshots were taken.

Average Hydration Number					
Forcefield	**H-A44**	**H-P45**	**N-K46**	**H-K46**	**Bulk**
c36	6.9	6.7	6.3	6.7	7.4
prosECCo	5.0	4.9	3.9	4.3	6.4
AMOEBA	5.2	5.0	4.3	5.5	7.3

Table 4.2 | Average Hydration Number of Cl^- The average hydration number of Cl^- ions in proximity to specified site contacts per forcefield.

For c36, the average hydration number remains similar for each hydrophobic contact. A slight reduction is observed compared to Cl^- in bulk solution. This retention of hydration shell correlates with the ion-distance plots (Figure 4.9) where Cl^- ions do not come into as close proximity with the binding site contacts compared with prosECCo and/or AMOEBA. With the prosECCo forcefield, significant Cl^- ion dehydration occurs in the binding site relative to bulk solution. Based on Chapter 3, this effect is anticipated and is again reflected in the Cl^- approach distance illustrated in Figure 4.9. Moreover, this dehydration is energetically advantageous as it promotes ion binding (Figure 4.4). The largest dehydration is observed in simulations with the AMOEBA forcefield where Cl^- loses ~ 2 water molecules relative to bulk in vicinity of the hydrophobic contacts. The direct ion-residue interactions enable Cl^- to exist in a more stably bound state reflected by the longer binding durations (Figure 4.9) and increased ion occupancy (Figure 4.7 B). This is also consistent with the interactions and behaviours associated with anion hydration that have been more thoroughly explored studies by Rempe *et al.* [20, 248].

Furthermore, the ion-distance plots (Figure 4.9 A-C) suggest that N-K46 interacts with Cl^- less closely than H-A44 and H-P45 within the site. Visually, this is can be understood by considering the distance between Cl^- in the system relative to the Cl^- modelled in the crystal structure site. Figure 4.9 F exemplifies how a Cl^- in AMOEBA simulations can localise in the binding site at positions closely corresponding to that of the experimentally determined structure and with a similar binding site conformation. Whereas for c36 (Figure 4.9 D) and prosECCo (Figure 4.9 E) simulations, the shape of the binding site is more variable which may influence Cl^- interactions. There is a tendency for the K46 tail to enclose the anion such that N-K46 is positioned deeper inside the concave cavity formed by the binding site as a whole (see Figure 4.10).

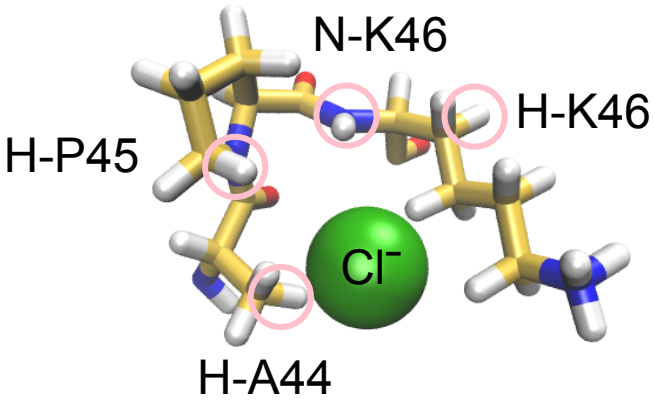

Figure 4.10 | Snapshot of Cl^- in Concave Site Conformation Snapshot of a Cl^- ion in the binding site. This is an example of where the K46 tail encloses the ion to form a deeper concave shape that facilitates ion binding.

4.4.4 Mechanisms of Cl^- Binding

Based on findings from the PMF profiles discussed above, we further explored the mechanisms that give rise to the metastable free energy minimum observed in simulations employing the prosECCo forcefield and backbone restraints (second minimum at $\sim$ 0.50 nm, Figure 4.4 A). On closer observation of the umbrella sampling windows, the residues A253 and Y255 in a nearby loop (residues 252-256) appeared to coordinate Cl^- at the location of the second free energy minimum. Notably, these residues are hydrophobic. To explore the influence of these hydrophobic loop contacts, we simulated the larger protein fragment systems in solution with the c36, prosECCo and AMOEBA forcefields. Again, the distance between Cl^- and the coordinating binding site atoms as before, but now also calculating the distance to a hydrogen from the methyl group of A253 (H-A253) and a hydrogen of CD1 belonging to the aromatic ring of Y255 (H-Y255) (Figure 4.11 A).

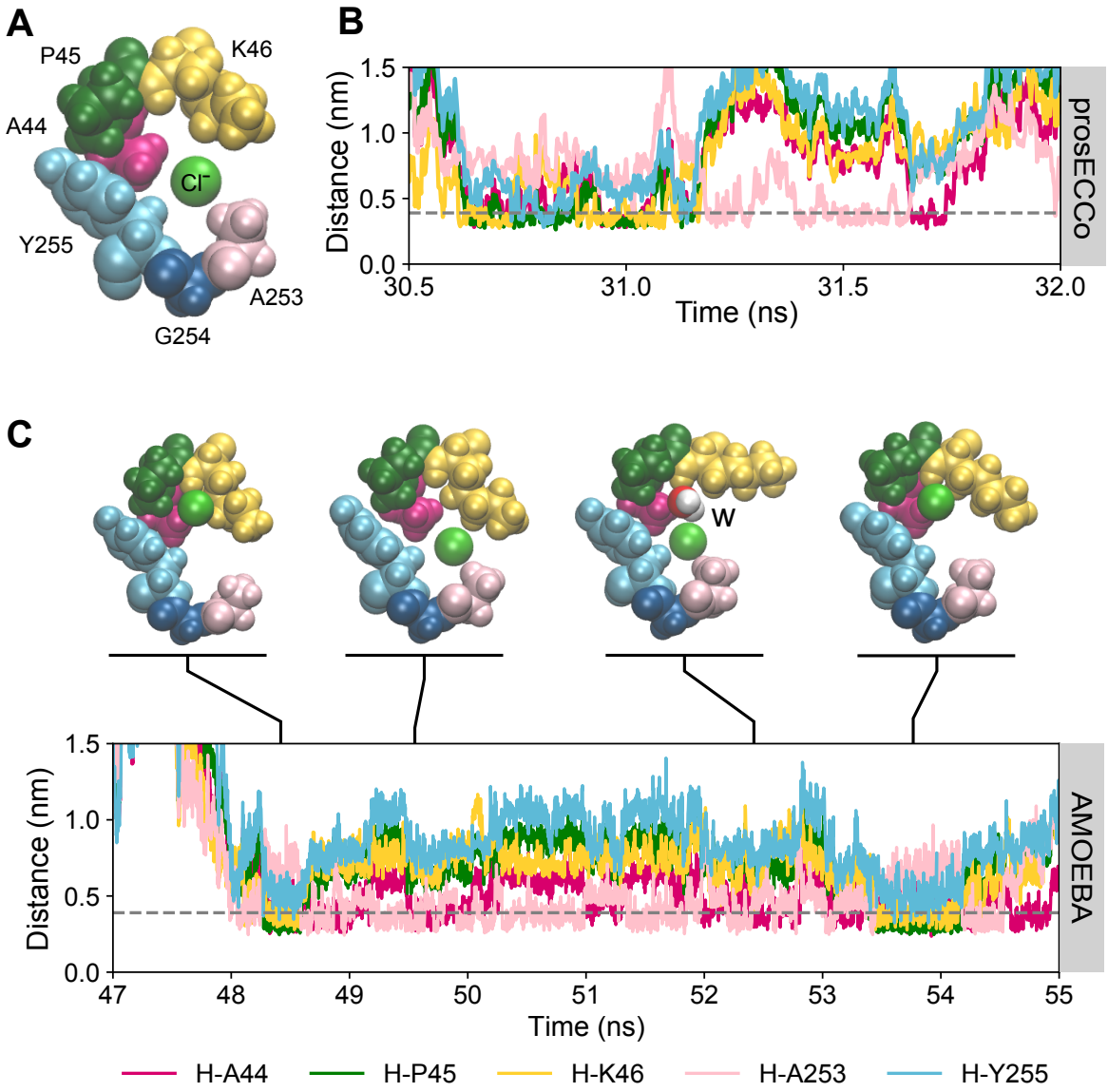

Figure 4.11 | Large Fragment Binding Site Dynamics A Larger protein fragment containing the Cl^-#2 binding site and residues from a nearby loop. The binding site is composed of A44 in magenta, P45 in dark green, and K46 in yellow. The additional residues of interest from the nearby loop are illustrated by A253 in pink, G254 in navy, and Y255 in light blue. **B** Example of an ion-distance plot as a function of time for a Cl^- ion in the binding site simulated using the prosECCo forcefield. **C** An ion-distance plot as a function of time for a Cl^- in the binding site simulated using the AMOEBA forcefield. Snapshots of a Cl^- ion in the binding site illustrate the ion dissociation and rebinding mechanism. A water molecule, marked by 'w', can be observed to displace the Cl^- ion from Cl^-#2 while the ion favourably interacts with nearby hydrophobic residues A253 and Y255.

No prolonged interactions between Cl^- ions and the larger fragment complex were identified with c36 simulations. However, with the prosECCo forcefield (Figure 4.11 B) we observed instances where Cl^- was initially bound to its binding site, then dissociated to interact with the nearby loop (e.g., A253) and later proceeded to rebind to its

site once more. Due to the intriguing nature of this observation, we repeated the simulations with a fully polarisable force field.

With AMOEBA, the ion also dissociated from the binding site in exchange with a water molecule (e.g., Figure 4.11 C, snapshot marked "w"). From here, the ion could then form favourable interactions with A253 and/or Y255 to remain in close proximity to the binding site, thus providing the opportunity for the ion to rebind or otherwise dissociate into bulk solution. An example of such a rebinding event is illustrated in Figure 4.11 C. These conclusions are supported by simulations of the full protein embedded in a lipid bilayer with the c36 and prosECCo forcefields where the rebinding mechanism was not observed with c36 but did occur with prosECCo (Figure 4.12). This mechanism is therefore only observed when polarisability is considered. It is interesting to note that the ion binding/rebinding occurs on a longer timescale with AMOEBA in comparison to c36 and prosECCo suggesting that the rate of transport of Cl^- may be lower than that predicted by non-polarisable forcefields.

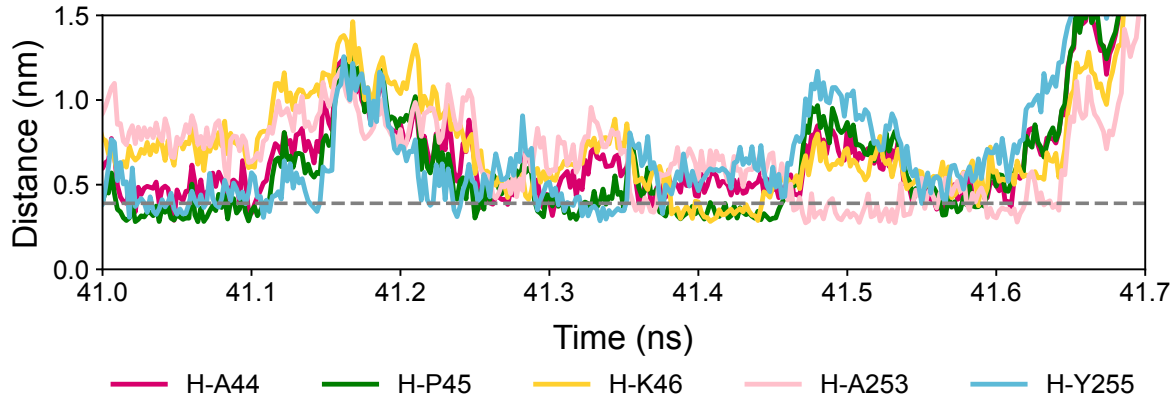

Figure 4.12 | Ion-distance Plots from Full Protein Simulations Ion-distance plot as a function of time for a Cl^- ion in the binding site and residues from a nearby loop. The ion distances were measured from simulations of the whole protein embedded in a lipid bilayer using the prosECCo forcefield. The Cl^- is bound to the site (contacts H-A44, H-P45 & H-K46) then dissociates (at $\sim$ 41.5 ns) and forms interactions with residues from the nearby loop (H-A253 & H-Y255). Cl^- rebinding can be observed (at $\sim$ 41.6 ns) and the ion is then released into bulk solution (at $\sim$ 41.7 ns).

Improving common fixed charged methods to better capture the effects of polarisation (e.g. prosECCo) preserves computational efficiency. Table 4.1 compares the computational performance of each forcefield per system size. As demonstrated here, prosECCo can act as a proxy for identifying potentially interesting anion interactions at no additional computational cost to a standard non-polarisable forcefield. These interactions can then be examined further by performing simulations using explicitly polarisable models such as AMOEBA [171], CHARMM Drude [266], etc. To some degree, prosECCo can replicate the qualitative behaviour produced by more advanced forcefields, however, to gain a higher level of physical accuracy, extensive simulations employing fully polarisable forcefields may be required or even *ab initio* MD calculations.

4.5 Summary & Conclusions

The theme of this chapter has been to assess the impact of modelling polarisation on the interactions of Cl^- ions within a biologically relevant binding site. The Cl^--pumping rhodopsin contains a defined hydrophobic binding site providing a starting point for detailed examinations of anion permeation and selectivity. Simulations of small and larger sections of truncated protein containing the binding site revealed how the inclusion of electronic polarisation leads to stronger anion binding events and longer binding durations. Additionally, these effects result from energetically favourable interactions with a site-adjacent loop concurrent with partial ion dehydration to directly interact with the hydrophobic contacts of both the binding site and the loop. Crucially, these binding mechanisms were only observed when polarisation (effects) were incorporated into the simulations. These results are consistent with the conclusions drawn in Chapter 3.

Future studies might compare two different mean-field approaches such as NBFIX and prosECCo against simulations employing full polarisable forcefields for a given biological system. As demonstrated by Orabi *et al.* [104], NBFIX has been effective at producing more physically realistic dynamics in simulations of CLC-ec1 Cl^-/H^+ antiporter. It would be interesting to compare Cl^- interactions in prosECCo and AMOEBA simulations of this system.

"I put my heart and soul into the work and have lost my mind in the process."

— Vincent Van Gogh

5 Simulation Studies of New Channels

Contents

5.1 Introduction 112
5.2 Objectives 116
5.3 Methods 117
5.3.1 Structural Model and System Preparation 117
5.3.2 Non-polarisable MD Simulations 119
5.3.2.1 Protein Fragment Simulations 119
5.3.2.2 Full Protein Membrane-Embedded Systems 119
5.3.3 Constant Electric Field Simulations 120
5.3.4 AMOEBA Simulations 120
5.3.5 Binding Site Identification and Analysis 121
5.4 Results and Discussion 122
5.4.1 Model System Validation 122
5.4.2 Interpretation of Cryo-EM Densities From Fully Polarisable MD Simulations 125
5.4.3 Functional Advantages of Pore Asymmetry 131
5.4.4 The Role of Water in Cl^- Permeation 135
5.4.5 Pathway for Conformational Change Between Open and Partially Open States 138
5.4.6 The Partially Open Neck state 139
5.4.7 Analysis of the Partially Open Neck State 140
5.5 Summary & Conclusions 148

This chapter is based upon the manuscript:

"*Electronic polarizablity Tunes the Function of the Human Bestrophin 1 Cl^- Channel*"

Phan, L. X., Owji, A. P., Yang T., Crain, J., Sansom, M. S. P. and Tucker, S. J. *bioRxiv*, doi.org/10.1101/2023.11.14.567055 (2023)

5.1 Introduction

Chapters 3 and 4 have laid the foundations for investigating anion-hydrophobic interactions within model and biological environments. The modelling of polarisation for such interactions has shown to be essential in producing physically realistic dynamics and providing new mechanistic insights. The current chapter builds upon the work from previous chapters by now probing in-pore interactions of Cl^- ions within a biological ion channel and in particular exploring the roles of anion-π (a.k.a. anion-quadrupole) interactions and their contributions to protein function.

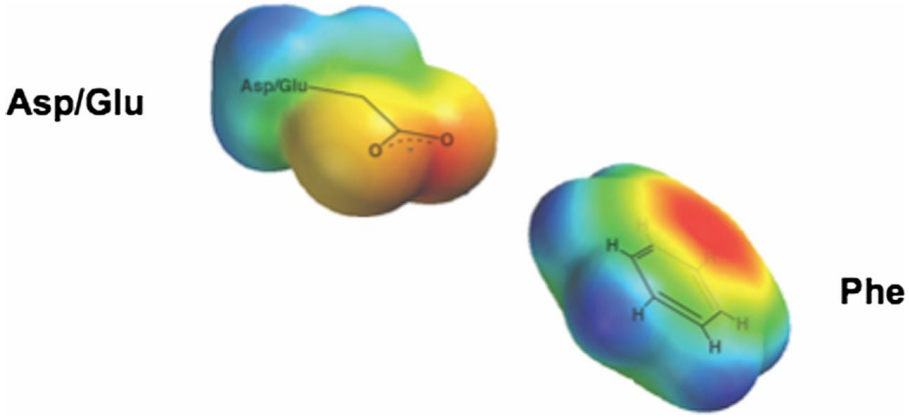

Figure 5.1 | Example of an Anion-Aromatic Interaction Schematic diagram of an anion-π interaction. The negatively charged sidechain of aspartic acid (Asp) or glutamic acid (Glu) is situated next to the aromatic phenylalanine (Phe) sidechain with a positive electrostatic potential around the ring edge. Copyright (2013) Nation Academy of Sciences [267].

Anion interactions with aromatic edges are a well-established phenomenon across multiple fields and have gained considerable interest [268–271]. Anion-π interactions are non-covalent interactions between anions and electron deficient aromatic systems. Electrons in the π bonds of an aromatic ring form a quadrupolar moment which creates a partial negative charge on both faces of the aromatic system and a corresponding partial positive charge around the edge of the aromatic ring (Figure 5.1) [268, 272, 273]. More recently, anion-π interactions have been recognised to play important roles in buried

regions of proteins [271, 274] and within the ion conduction pathway of some anion channels whereby aromatic residues such as phenylalanine form part of the pore-lining. Known structures of anion channels that exploit these types of interactions include the mechanosensitive Cl^- channel, Flycatcher1 (FLYC1) where the narrowest constriction ($\sim$ 0.28 nm radius) is created by a ring of phenylalanine sidechains [44] (Figure 5.2), the slow anion channel (SLAC1) which is occluded by a highly conserved phenylalanine residue responsible for channel gating (< 0.2 nm) [275] and the mechanosensitive channel of small conductance from *Escherichia coli* (EcMscS) in which phenylalanines appear to constitute the hydrophobic gate ($\sim$ 0.4 nm) [276]. Although such anion-aromatic interactions are readily observed, the degree to which dilation is required for anion passage in the presence of pore-lining phenylalanine residues is variable and only partially understood. Therefore, there is a need for further biophysical characterisation to better understand the functional roles of such pore-lining aromatics and their contributions towards anion selectivity and permeation.

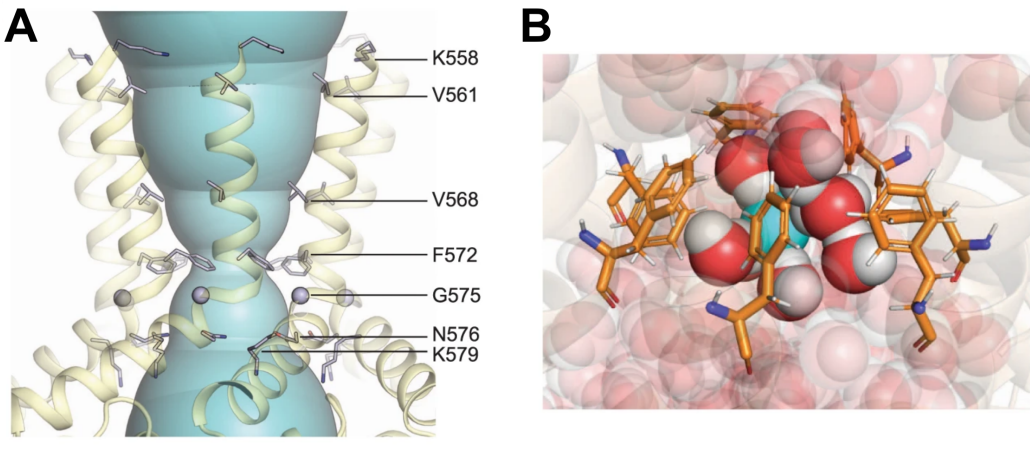

Figure 5.2 | Example of Aromatic Sidechains in an Anion Channel Structure of the Cl^--selective channel - Flycatcher1 (FLYC1). **A** Depiction of the pore-lining helices from the TM domain of FLYC1. The tightest constriction is formed from a set of aromatic phenylalanine residues, F572, shown in licorice representation. **B** A representative snapshot of a Cl^- passing through the F572 ring whereby Cl^- is required to partially dehydrate to permeate this region. Figure reprinted and adapted from *Jojoa-Cruz et al.* [44].

Cryo-electron microscopy (cryo-EM) is a powerful structural technique that has revolutionised the field of structural biology. Recent technological improvements in sample treatment, grid preparation, microscope hardware, and image processing have made it achievable to obtain high-resolution maps better than 0.3 nm [26]. Despite these advancements, there are several caveats, such as radiation damage, resulting in lower signal-to-noise ratios, protein denaturation and beam-induced sample movement, which all present limitations to achieving even higher resolutions [28]. The resolution of single-particle cryo-EM is also often insufficient to provide detailed information on small molecules or bound ligands, such as water or ions, which require a resolution of at least 0.25 nm [277], and so it can be difficult to differentiate and interpret these small densities with confidence when using this technique. Molecular dynamics (MD) simulations can extend the capabilities of cryo-EM by capturing conformational variability and short-lived states which can then be related to evidence from cryo-EM [44]. Furthermore, MD simulations can assist in the assignment and interpretation of cryo-EM data to access improved atomic resolution structures [278].

As emphasised in previous chapters, the majority of MD simulations employ pairwise additive forcefields that model electrostatic interactions as Coulombic forces between fixed point partial charges [178]. Additive forcefields do not capture the effects of induced polarisation that arise from redistributions of charge density in response to local field gradients. This becomes problematic for modelling polarisable components; in particular, aromatic rings with delocalised systems of electrons that do not contain dipole moments, only quadrupole moments. The AMOEBA forcefield models atomic monopoles through to quadrupole moments within a classical MD framework and is therefore capable of capturing effects resulting from fluctuations in charge density [176] (refer to Section 2.2).

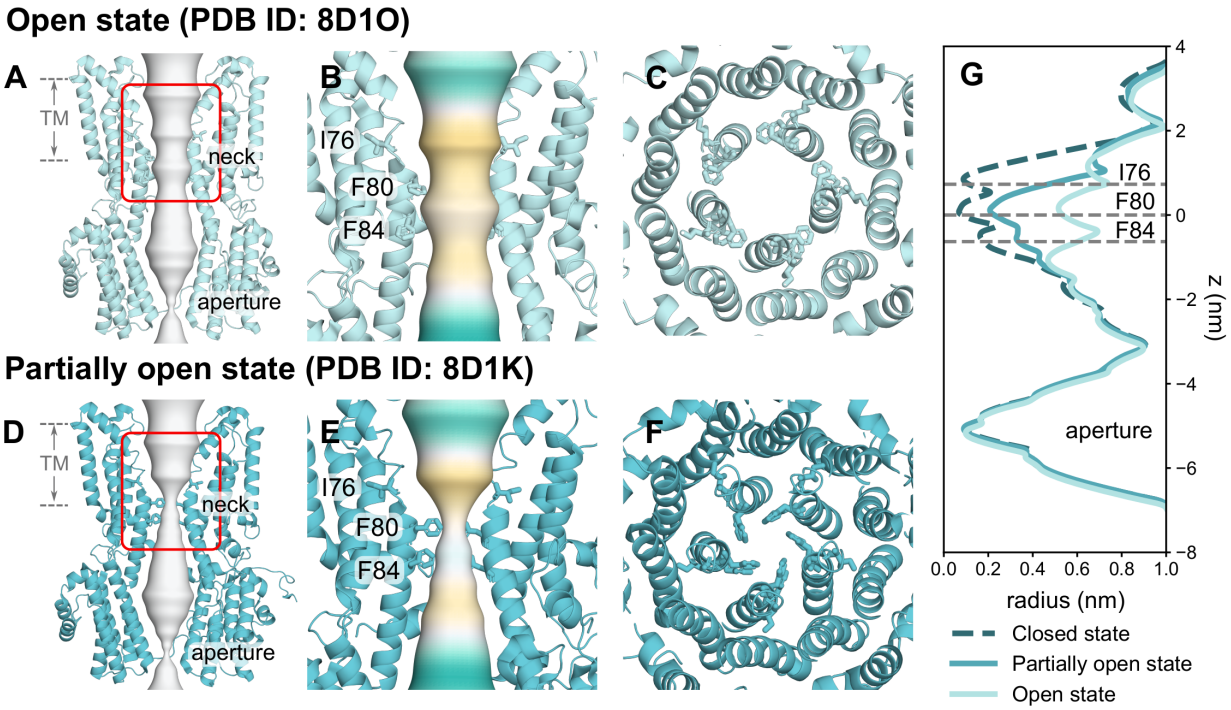

Figure 5.3 | Open and Partially Open States of the Human Bestrophin 1 Channel A Open state hBest1 (PDB ID: 8D1O) (cyan) and **D** partially open state hBest1 (PDB ID: 8D1K) (teal) structures visualised with ion permeation pathway in white. The TM neck region of the open state **B** and partially open state **E** structures, with ion permeation pathway coloured by hydrophobicity with pale brown corresponding to maximum hydrophobicity and green corresponding to maximum hydrophilicity. Top-down view of **C** the open state neck and **F** partially open state neck. **G** Pore radius profiles of the open, partially open and closed (PDB ID: 8D1I) states.

Bestrophin channels are a family of Ca^{2+}-activated Cl^- channels (CaCCs) [46, 279]. Four paralogs (Best1-4) have been identified in eukaryotes and are responsible for a diverse range of functions in the kidneys, airways, central nervous system etc. The best-known physiological role of Best1 is in the eye. Best1 is predominantly expressed in the retinal pigment epithelium where disease-causing mutations cause retinal degenerative disorders called bestrophinopathies [45, 280]. To date, bestrophins all exhibit the same overall pentameric architecture of four TM helices per protomer, forming a central ion conducting pathway [281]. The structures of bestrophin channels contain two key constrictions within the ion conduction pathway (Figure 5.3 G). A permeating ion from the extracellular side will first encounter the *neck* region, which forms a gate composed of three hydrophobic pore-facing residues: isoleucine 76 (I76), phenylalanine (F80) and

phenylalanine 84 (F84) that are highly conserved across homologs (Figure 5.3 B & E). After passing this gate, the ion will traverse the interior of the cytosolic vestibule, followed by a second, shorter constriction at the cytosolic exit, called the *aperture*, which displays significant divergence across paralogs [46]. Together, the neck and aperture are thought to collectively comprise part of the gating mechanism for bestrophin channels, although their relative contribution to gating and/or Cl^- selectivity remain unclear [45, 46, 105, 282].

Here, we examine and compare the human Bestrophin 1 channel (hBest1) in the fully open state (PDB ID: 8D1O, 0.24 nm resolution) (Figure 5.3 A-C, G) with the partially open neck state (PDB ID: 8D1K, 0.23 nm resolution) (Figure 5.3 D-G) [105]. The conformational differences in the pore-facing residues of the hydrophobic neck (Figure 5.3 B & E) in the partially open state relative to the open state allow us to explore the conformational pathway to channel opening/closing as well as probe the mechanisms of ion permeation and selectivity through the neck.

5.2 Objectives

In this chapter, we explore the effects of electronic polarisation on hBest1 function by employing a fully polarisable forcefield. We present an example of how such forcefields provide detailed novel insights into channel function and the behaviour of Cl^- ions in the neck of hBest1. The roles of aromaticity and pore asymmetry in ion permeation were investigated along with the influence of including polarisation in the model. In particular, the relationship between rotameric transitions in the phenylalanine sidechains and Cl^- permeation were examined in more depth. Crucially, we demonstrate how the inclusion of realistic polarisation can be used to interpret ambiguous ion-like densities observed in experimental high-resolution cryo-EM structures, thus enabling more accurate functional annotation.

Simulation studies of the human bestrophin 1 channel in this chapter aim to explore the following questions:

1. How does electronic polarisation influence the behaviour of Cl^- within an ion channel?

2. What roles do anion-π interactions play in Cl^- permeation?

3. Can MD simulations implementing explicitly polarisable forcefields help elucidate the identities of ambiguous cryo-EM densities?

5.3 Methods

5.3.1 Structural Model and System Preparation

Cryo-EM structures of human Bestrophin 1 (hBest1) in the Ca^{2+}-bound open state (PDB ID: 8D1O, 0.24 nm resolution) and the hBest1 Ca^{2+}-bound partially open neck state (PDB ID: 8D1K, 0.23 nm resolution) [105] were obtained from the Protein Data Bank (PDB). As in Section 4.3.1, due to the methodological and high computational demands for implementing explicitly polarisable forcefields, we have reduced the protein structures to include the pore-lining segments spanning most of the TMD (Figure 5.4 A & B). Furthermore, the aperture presents a second tight constriction in the intracellular region where ion dehydration would be required to pass. The conformation of the aperture remains closed in all known structures of bestrophin to date and the open state aperture is yet to be discovered [105, 282]. Electrophysiological studies have concluded that the aperture does not function as an in/activation gate but rather as a size-selective filter responsible for the permeability of the lyotropic series [283, 284]. Thus, the truncation of the proteins to remove the cytosolic region, enabling the passive diffusion of water and ions in the neck to be studied, is justified. We perform preliminary simulations of the full protein to demonstrate the implications of modelling the aperture

for studying the neck region in Section 5.4.1. Therefore, we reduce the protein to fragments composed of residues 56 to 99, which encompass the neck region of interest (residues I76, F80 and F84) (Figure 5.4 C). Protein fragment systems were solvated in a 0.5 M NaCl solution and were prepared using the CHARMM-GUI protocol [261].

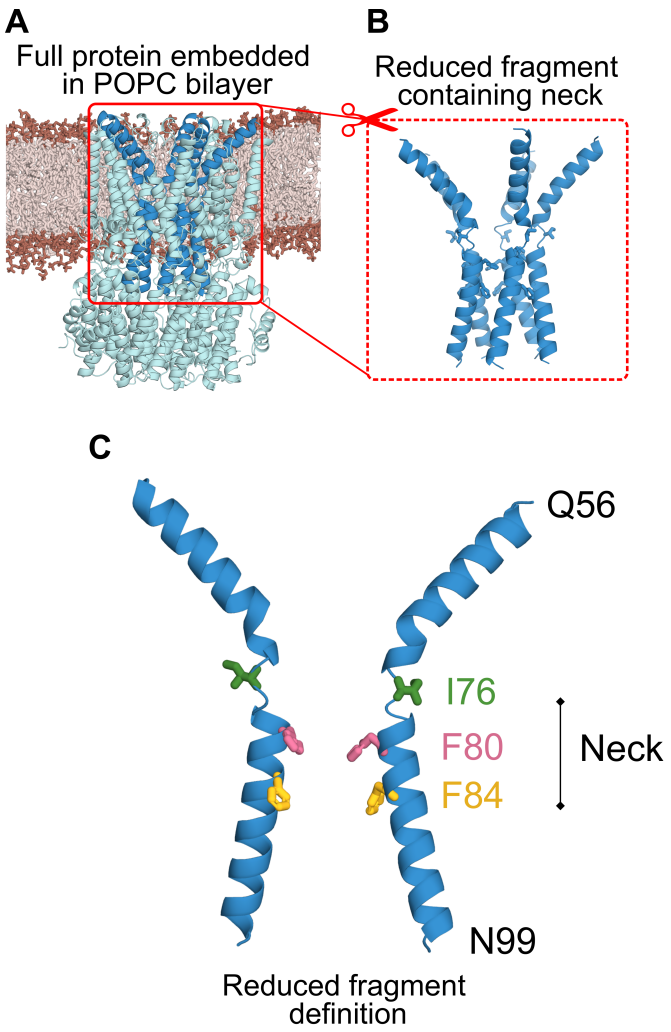

Figure 5.4 | Structural Model Preparation A The open state full protein (cyan) (PDB ID: 8D1O) embedded in a POPC bilayer (lipid heads in brown, lipids tails in tan). The full protein membrane-embedded system in solution contains $\sim$ 200500 atoms. **B** The reduced protein fragment was truncated from approximately the TM portion of the protein and was used in simulations due to the high computational demands of the AMOEBA forcefield. **C** The reduced fragment protein (mid-blue) was obtained by truncating the full protein at Q56 and N99 such that the remaining fragment system contains the conserved hydrophobic neck region of interest (I76, F80, F84). The reduced protein fragment in solution consists of $\sim$ 57000 atoms. Water and ions are omitted from the figure for clarity.

5.3.2 Non-polarisable MD Simulations

5.3.2.1 Protein Fragment Simulations

MD simulations employing a non-polarisable forcefield were performed in GROMACS [144] version 2021 using the CHARMM36m (c36m) forcefield in conjunction with the mTIP3P water model [135]. Protein fragment systems were solvated in 0.5 M NaCl solution to achieve better sampling despite being much higher than physiological conditions. The system was first subjected to energy minimisation followed by a 2 ns NVT equilibration period. Simulations in the NPT ensemble were conducted for 100 ns whereby the first 20 ns of each simulation were discarded as an equilibration. Therefore, the final 80 ns of the simulation were used for analysis.

Simulations were carried out using the leap-frog integrator [144] with a timestep of 2 fs. The temperature was maintained at 310 K using the Nosé-Hoover thermostat [155] with a time coupling constant of 1.0 ps. The pressure was maintained at 1 bar using the Parrinello-Rahman barostat [157] with a time coupling of 5.0 ps. Short-range electrostatics were treated with the Verlet cutoff scheme with a cutoff at 1.2 nm and long-range electrostatics were treated with the particle mesh Ewald (PME) algorithm [163]. The LINCS algorithm [149] was used to constrain H-containing bonds. Backbone atoms were placed under harmonic restraints with a force constant of 1000 kJ/mol/nm^2 to prevent the structures from deviating too much from the experimental coordinates. Three independent repeats were carried out for each system.

5.3.2.2 Full Protein Membrane-Embedded Systems

The full protein systems were prepared using a multiscale procedure. The protein was first coarse-grained (CG) and then embedded into a POPC bilayer and solvated with water and $\sim$ 0.5 M NaCl using Martini version 2.2 [114] and GROMACS 2021 package [144].

This system is subject to 100 ns of equilibration in CG before being converted to atomistic representation with the CG2AT protocol [285]. An equilibration period of 20 ns followed by a production run of 100 ns using the c36m forcefield [135] and mTIP3P water model was performed. The temperature was maintained at 310 K with coupling constant 1.0 ps by the Nosé-Hoover thermostat [155]. Pressure was maintained at 1 bar with coupling constant 5.0 ps by the Parrinello-Rahman barostat [157]. Short-range electrostatics were treated with the Verlet cutoff scheme at a 1.2 nm cutoff and long-range electrostatics were treated with PME method [163]. C_α atoms were placed under harmonic restraints with a force constant of 1000 kJ/mol/nm^2 to prevent the structure from deviating too much from the experimental structure.

5.3.3 Constant Electric Field Simulations

MD simulations employing a constant external electric field were carried out for the partially open membrane-embedded whole protein system. System set up and equilibration followed the same protocol as in Section 5.3.2. An electric field, E = 0.00699 V/nm was applied in the z-direction, orthogonal to the plane of the membrane ($\Delta V = E \cdot L_z$, V = +100 mV, L_z = 14.3 nm is the length of the PDB simulation box) (see Figure 5.16). This gives rise to a force $q_i \cdot E$ that applies to all q_i charges in the system and is directed perpendicular to the membrane plane [286]. No electric field was applied during (NVT) equilibration of these systems, only during the production runs.

5.3.4 AMOEBA Simulations

Fully polarisable forcefield simulations were carried out in OpenMM 7.4.2 (www.openmm.org) and all components in the system were modelled with the AMOEBA polarisable forcefield using the AMOEBA18 parameter set which includes the protein

parameters from the AMOEBA13 forcefield [171, 287]. Starting configurations for these simulations were obtained from the end of the NVT ensemble equilibration period using c36m as described above. Simulations were then setup following a similar procedure to a method described previously in Chapter 4 using (https://github.com/Inniag/openmm-scripts-amoeba) [100].

AMOEBA forcefield simulations proceeded by performing 1000 steps of energy minimisation to resolve any divergent energies due to the induced dipoles. The production run was simulated for 60 ns, with the first 10 ns of the simulation discarded for equilibration; therefore, the final 50 ns were used for analysis. Time integration was performed using the r-RESPA multiple timestep integration algorithm [182] with an inner timestep of 0.25 fs and an outer timestep of 2 fs. The temperature was maintained at 310 K using the Andersen thermostat [151] and pressure was maintained at 1 bar using the isotropic Monte Carlo barostat [158, 159]. Electrostatic multipole interactions were evaluated by the PME method [162] with a real-space cutoff of 0.8 nm and tolerance of (5×10^{-4}) and a fifth order B-spline interpolation. VdW interactions were calculated explicitly up to a distance of 1.2 nm and interactions beyond this cutoff treated with an analytical long-range dispersion correction. All C_α atoms were placed under a harmonic restraint with force constant 1000 kJ/mol/nm^2 to prevent the protein from deviating from the experimental structure.

5.3.5 Binding Site Identification and Analysis

The interactions of Cl^- within each protein structure, binding site detection and quantification of dynamics were calculated using PylipID (https://github.com/wlsong/PyLipID) [288]. Binding sites were defined by Cl^- contact with three or more residues and by a dual-cutoff scheme for which if Cl^- were considered

bound, they resided between 0.33 nm and 0.48 nm based on data from the first hydration shell of Cl^- using the AMOEBA forcefield. The first hydration shell of Cl^- was determined by calculating the $Cl^- - O$ RDFs, as detailed in Section 4.3.4.1. As bestrophins are 5-fold symmetric structures, the Cl^- interactions were averaged over the 5 repeating units.

The Root Mean Square Deviation (RMSD) [289] as a function of time t was calculated to provide a measure of similarity between two superimposed sets of atomic coordinates over a trajectory. This was computed by,

$$RMSD(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(r_i(t) - r_i^{ref})^2} \tag{5.1}$$

where averaging is performed over the N number of coordinates, $r_i(t)$ denotes the coordinates at time t and r_i^{ref} represents the reference coordinates.

Alignment and visualisation of structures were achieved using PyMOL (https://pymol.org/2/). Trajectory analysis was performed using MDAnalysis [226, 227] and GROMACS analysis tools [144]. The pore radius profiles were obtained using CHAP (www.channotation.org) [69].

5.4 Results and Discussion

5.4.1 Model System Validation

Exploratory equilibrium simulations of whole protein membrane-embedded systems were performed for both the open (Figure 5.5 A) and partially open state (Figure 5.5 B) structures using the c36m forcefield. Initial assessment of ion and water trajectories revealed a consistently dewetted region at the aperture in both structures. In the open state (Figure 5.5 A), the neck is hydrated and ions appear to traverse this region, with more Cl^- than Na^+ entering the neck. However, the intracellular vestibule i.e. the region

between the neck and aperture, sees the formation of a double electric layer possibly due to the accumulation of charge as a result of restricted access to/from the intracellular domain.

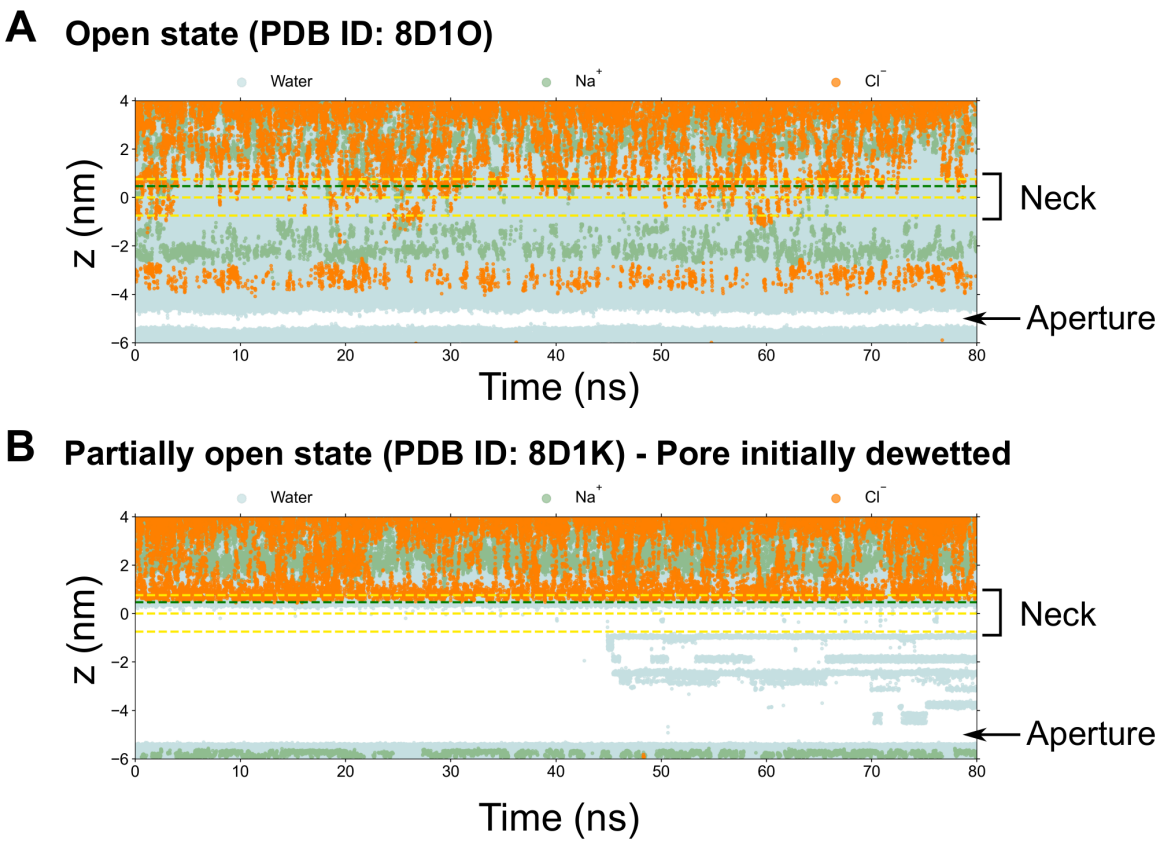

Figure 5.5 | Trajectories of Water and Ions Within Simulations of Whole Protein-Embedded Systems The z-coordinates of water (cyan), Na^+ (green) and Cl^- (orange) plotted as a function of time in simulations of **A** the open state (PDB ID: 8D1O) and **B** the partially open state (PDB ID: 8D1K) full protein structures embedded in a POPC bilayer using the c36m forcefield. In the partially open state, the pore was initially dewetted. The yellow dashed lines represent I76, F80 and F84 in the neck and the green dashed line represents the location of X739 from the experimental structure.

In the partially open state (Figure 5.5 B), the pore is initially dehydrated. Over the course of the simulation, the region between the hydrophobic neck and aperture remains dewetted until $\sim$ 45 ns where a few water molecules are able to permeate the neck to occupy the vestibule region where they are trapped due to the tight hydrophobic aperture. The region between the neck and aperture remain impermeable to ions. Truncating the structure removes the aperture and permits the passive diffusion of particles into the vicinity of the neck for the intentions of studying ion and water interactions with the neck. Non-protein densities have been captured bound to the partially open state structure determined by cryo-EM so we are confident that the vestibule region is solvent accessible (see Figure 5.15) [105].

In Chapter 4 and previous studies, reduced systems of ion channels focusing on isolated sections have shown that they can provide an accurate representation of the dynamics within the original protein [69, 259]. To validate the protein fragment, we performed simulations of the full open state (PDB ID: 8D1O) protein embedded in a lipid bilayer compared with the corresponding protein fragment in solution. The average water number density was analysed. The resulting profiles are similar, are within error of each other, and the average water density remains close to bulk outside the neck region (Figure 5.6). This gives us confidence that the protein fragments in our reduced systems offer a representative model of the fully open and hydrated pore from the parent protein. However, it is important to note that these reduced systems are limited and can only capture aspects of the full protein.

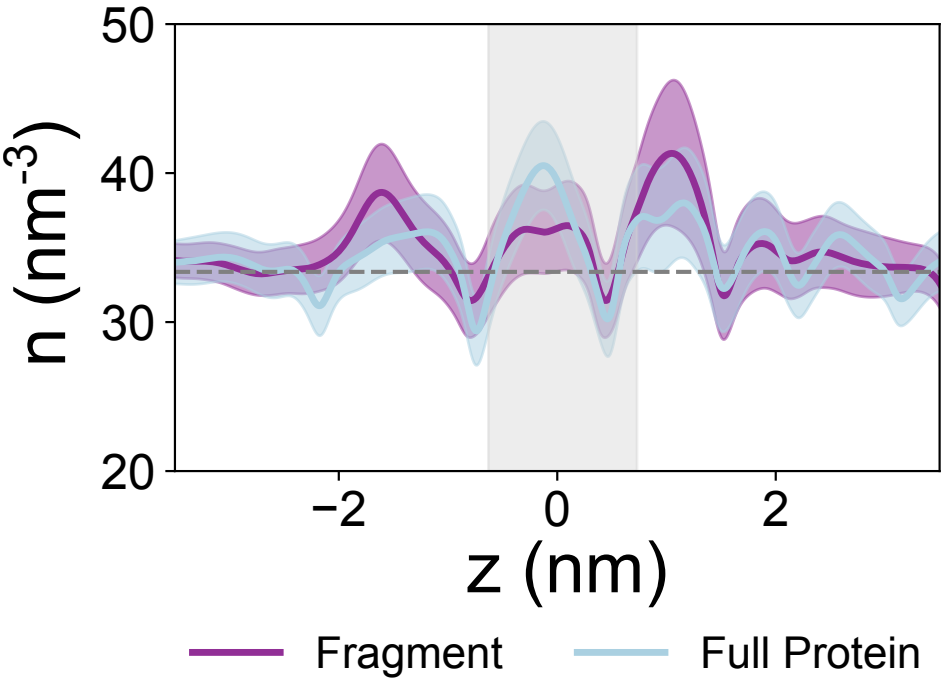

Figure 5.6 | Protein Fragment System Validation Time-averaged water density profiles within the pore for the full protein embedded system (light blue) compared with the protein fragment in solution (purple) using the c36m forcefield. The shaded region represents the neck (I76, F80 and F84), and the dashed grey line corresponds to the density of bulk water (33.37 nm^{-3}). Confidence bands represent the standard deviation over the simulation.

5.4.2 Interpretation of Cryo-EM Densities From Fully Polarisable MD Simulations

In the experimental structure of the open state hBest1, a number of non-protein, ion-like densities have been observed, which have been modelled as water molecules in the published structure. A set of these water molecules (corresponding to HOH739 and HOH539 in the PDB) are located within the neck region and appear to bind to the helix dipole [1] at a distance of 0.32 nm to the backbone NH of F80 of each chain (Figure 5.7 A). Similar non-protein densities are also observed consistently in proximity to the helix dipole in the open state structure of human bestrophin 2 (hBest2) (PDB ID: 8D1N).

[1]The Helix dipole moment refers to a macroscopic cumulative effect originating from contributions of each residue backbone unit such that there is an overall macrodipole from the C-terminus to the N-terminus [290, 291].

Due to experimental limitations, there is ambiguity in the molecular identity of these densities. Assessment of the local chemistry and comparisons with known Br^- binding sites from anomalous scattering studies of a Best1 homolog [47] suggest the location of these densities are comparable to the putative water molecules in the hBest1 structure. Therefore, it has been suggested that this site could possibly serve as a Cl^- binding site. Here, we have used the AMOEBA forcefield to explore the possibility that these densities could instead be representative of Cl^-. We refer to the density X739 as the ion-like density previously labelled as a water molecule (HOH739) in the PDB.

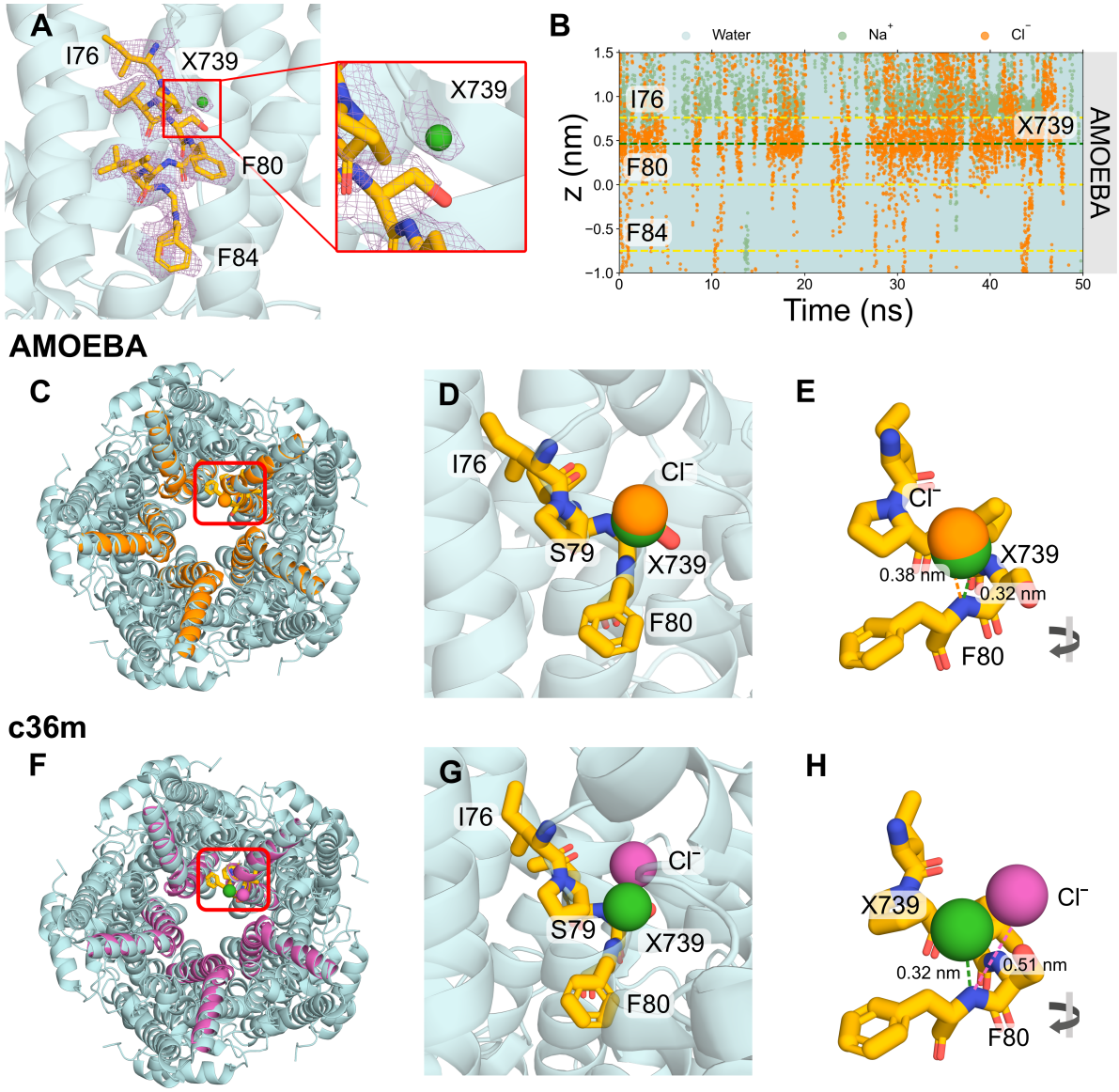

Figure 5.7 | The Cl^- Binding Site A Close-up representation of neck residues (I76, F80 and F84) from a single chain highlighted in licorice representation of the open state structure (PDB ID: 8D1O). The cryo-EM densities are shown as a mesh representation in purple. A panel is zoomed in on the non-protein ion-like density in proximity to F80 labelled as X739, which corresponds to the water molecule HOH739 in the PDB. **B** The trajectories of water (cyan), Na^+ (green) and Cl^- (orange) ions in z-coordinates as a function of time within the neck region of the pore. The plot indicates that Cl^- tends to cluster in a region between I76 and F80. The green dashed line represents the location of X739 from the experimental structure. AMOEBA forcefield simulations: Top-down view of the protein fragment in **C** AMOEBA (orange) or **F** c36m (magenta) aligned to the full protein structure (cyan). The location of X739 is indicated by the green sphere. The top detected binding pose identified with PyLipID [288] shows **D** significant overlap between the AMOEBA Cl^- (orange sphere) and X739 indicating this site likely functions as a Cl^- binding site. On the contrary, **G** the c36m Cl^- (magenta sphere) shares no overlap with X739. **E, H** A rotated view showing the distance of the Cl^- and X739 to the backbone N of F80 in the experimental structure.

Analysis of the z-positions of ions and water as a function of time revealed that the open state neck is wetted and permeable to ions. There is a clear accumulation of Cl^- in the region between I76 and F80 of the neck (indicated by the dense orange region in Figure 5.7 B). An electric double layer also forms and Na^+ ions are largely excluded from regions below I76 (Figure 5.7 B). To pinpoint precisely where Cl^- is clustering in three dimensions, we have used the computational tool, PyLipID, which is capable of analysing protein-ligand interactions [288]. Taking the top-ranked binding pose for the detected binding site, we focus on interactions of a single chain as the structure of hBest1 is 5-fold symmetric, Cl^- interactions have been averaged over the 5 repeating units. The simulated fragment structures have been aligned with the original PDB structure (chain E and HOH739) for comparison (Figure 5.7 C).

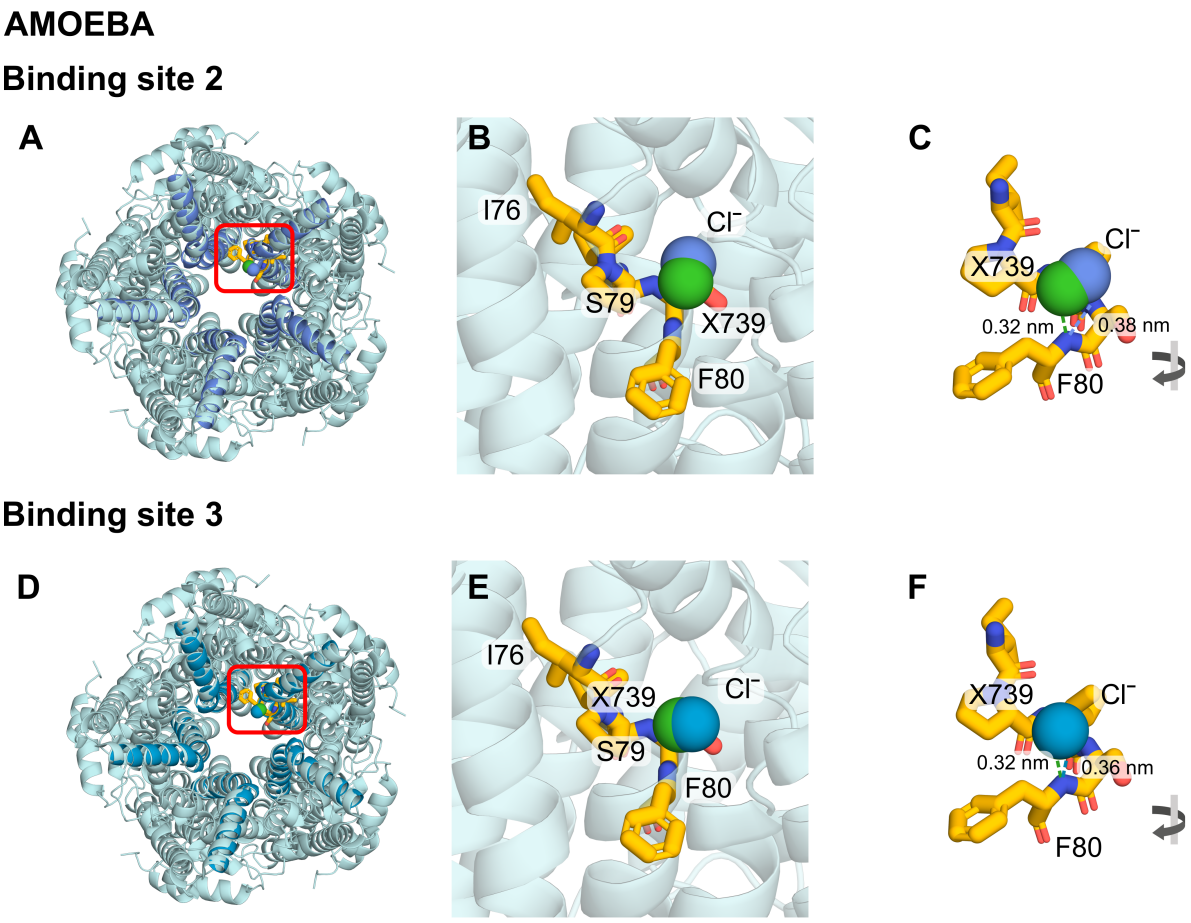

Figure 5.8 | Alternative Binding Sites Identified with PyLipid Alternative Cl^- binding sites of the open state structure (PDB ID: 8D1O) simulated using the AMOEBA forcefield. The location of the structural density X739 is represented by the green sphere. **A-C** The Cl^- ion (indigo sphere) is located at a distance of 0.38 nm to the backbone N of F80 in Binding Site 2 whereas in Binding Site 3, the Cl^- (blue sphere) is located at a distance is 0.36 nm (**D-E**).

Binding Site	Site Residence Time (ns)	k_{off}	Duration (ns)	Occupancy (%)	R-squared	Distance to N of F80 (nm)	Distance to X739 in PDB	Figure
3	0.59	1.7	0.75	14	0.9987	0.38	0.12	5.7 C-E
12	0.43	2.3	0.63	28	1	0.36	0.07	5.8 A-C
2	0.34	3.0	0.63	18	0.9999	0.38	0.06	5.8 D-F

Table 5.1 | Detected AMOEBA Simulation Binding Sites Table of binding site statistics from PyLipid [288] where binding site 3 is the site described in Figure 5.7 C-E and binding sites 2 and 12 correspond to Figure 5.8 A-C and Figure 5.8 D-F respectively.

The results of the AMOEBA simulations correlate very well with the cryo-EM densities. Here, we first show the binding site that is most representative of the density associated with HOH739 in the original PDB structure (Figure 5.7 C-E). A significant overlap can be seen between the Cl^- and the water molecule modelled in the PDB structure (Figure 5.7 D). The distance of the Cl^- to the backbone nitrogen of F80 is 0.38 nm compared with 0.32 nm in the experimental structure (Figure 5.7 E) and it is equidistant from the pore axis. A number of alternative binding sites identified in the neck region are within 0.15 nm of the ion-like density (Figure 5.8) which represent other possible ion binding conformations in the immediate vicinity of the experimental site (Table 5.1, binding sites 12 and 2) when considering metrics such as percentage occupancy and geometrical distance to reference atom N of F80. The site displayed in Figure 5.7 was chosen as the most representative according to binding duration and residence time. Studies of CLC channels have revealed that pore-lining backbone amides can influence ion selectivity and permeation [21]. Therefore, it is feasible that the ion-like density in this experimental structure is not water but a Cl^-. This difference in ion location between the simulation and structural densities could be a result of the imposition of 5-fold symmetry during structural processing, which could marginally shift the cryo-EM density in comparison to allowed asymmetric interaction modes. Conversely, from a simulation perspective, the localisation of Cl^- can be forcefield dependent and conditional on the realism of the model.

To demonstrate the impact of modelling polarisability, equivalent simulations were also performed with the non-polarisable forcefield, CHARMM36m (c36m) (Figure 5.7 F). In contrast to the AMOEBA simulation, these simulations suggest that Cl^- tends to localise at a distance of 0.51 nm from the backbone nitrogen of F80 (Figure 5.7 H) and prefers to interact with the backbone NH, hydroxyl group of S79, and the backbone NH of I78 (Figure 5.7 G). The c36m simulations do not detect the same Cl^- binding geometries as

the ones yielded by the AMOEBA simulations and experimental density. Furthermore, no alternative binding conformations within the neck region are detected for c36m simulations.

5.4.3 Functional Advantages of Pore Asymmetry

It is common practice in cryo-EM to apply rotational symmetry in the map construction and refinement process. Regions of structure that present heterogeneous features are averaged over several images. Averaging leads to the loss of information of individual differences between protomers. It is possible to extract heterogeneous regions through methods such as the application of 3D masks during classification and refinement [26, 292]. However, applying a small mask to putative ions or water molecules is difficult due to the size of the features and may be unsuccessful in providing any additional structural information.

MD simulations enable examination of the functional roles that asymmetry may play in the neck which would otherwise not be easily deduced from experimental structures. In the open state structure, the sidechains of F80 and F84 are pore-facing whilst the side chain of I76 turns away from the ion-conducting pore. The pore radius in the neck region is sufficiently large to accommodate a fully hydrated Cl^- (radius $\sim$ 0.4 nm [246]) and from Figure 5.7 B, the pore is wetted and permeable to ions in the neck. Therefore, it is of interest to study how these ions are coordinated, their permeation mechanisms and whether pore asymmetry presents any functional advantages in this neck region.

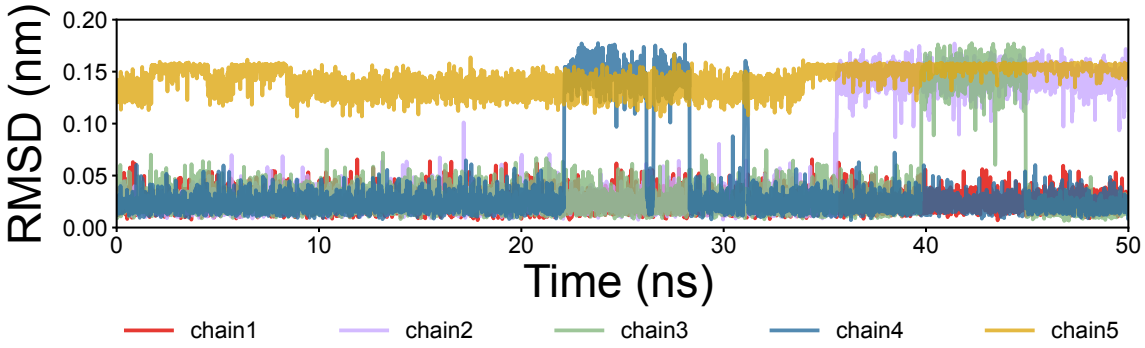

Figure 5.9 | RMSD of F80 Sidechains in the Open State RMSD of F80 sidechain per chain of the open state (PDB ID: 8D1O) protein fragment over the AMOEBA simulation.

Throughout the AMOEBA forcefield simulation, asymmetric states of the pore at F80 can be observed. The sidechains of F80 exhibit two distinct conformations illustrated in the RMSD of F80 per chain of the protein (Figure 5.9) with dynamical pathways between them. The RMSD of $\sim$ 0.14 nm seen in the trajectory of chain 5 corresponds to a 180° rotation of the aromatic ring after equilibration; nonetheless it represents an equivalent conformational change compared to the other chains. To dissect the dynamics of the sidechains further, we characterised the motions into three components: firstly, the torsion angle, (χ_2) (CA-CB-CG-CD1) which captures the orientation of the aromatic ring (Figure 5.8 A). Secondly, the angle (θ) between vectors defined by F80 CA and F84 CA of the same chain and F80 CA with the carbon atom, CZ from the aromatic ring, defines the vertical up/down motion of the F80 sidechain in the z-direction (Figure 5.8 B). Finally, xy defines the average distance in the $x-y$ plane between atom CZ of F80 from one chain (N) and the CA atom from F80 in the neighbouring chain (N+1). This is taken as a moving average over 50 windows (equivalent to a time of 0.5 ns). xy characterises the F80 sidechain conformation as facing inward towards the pore axis or outwards as found in the experimental structure (Figure 5.8 C).

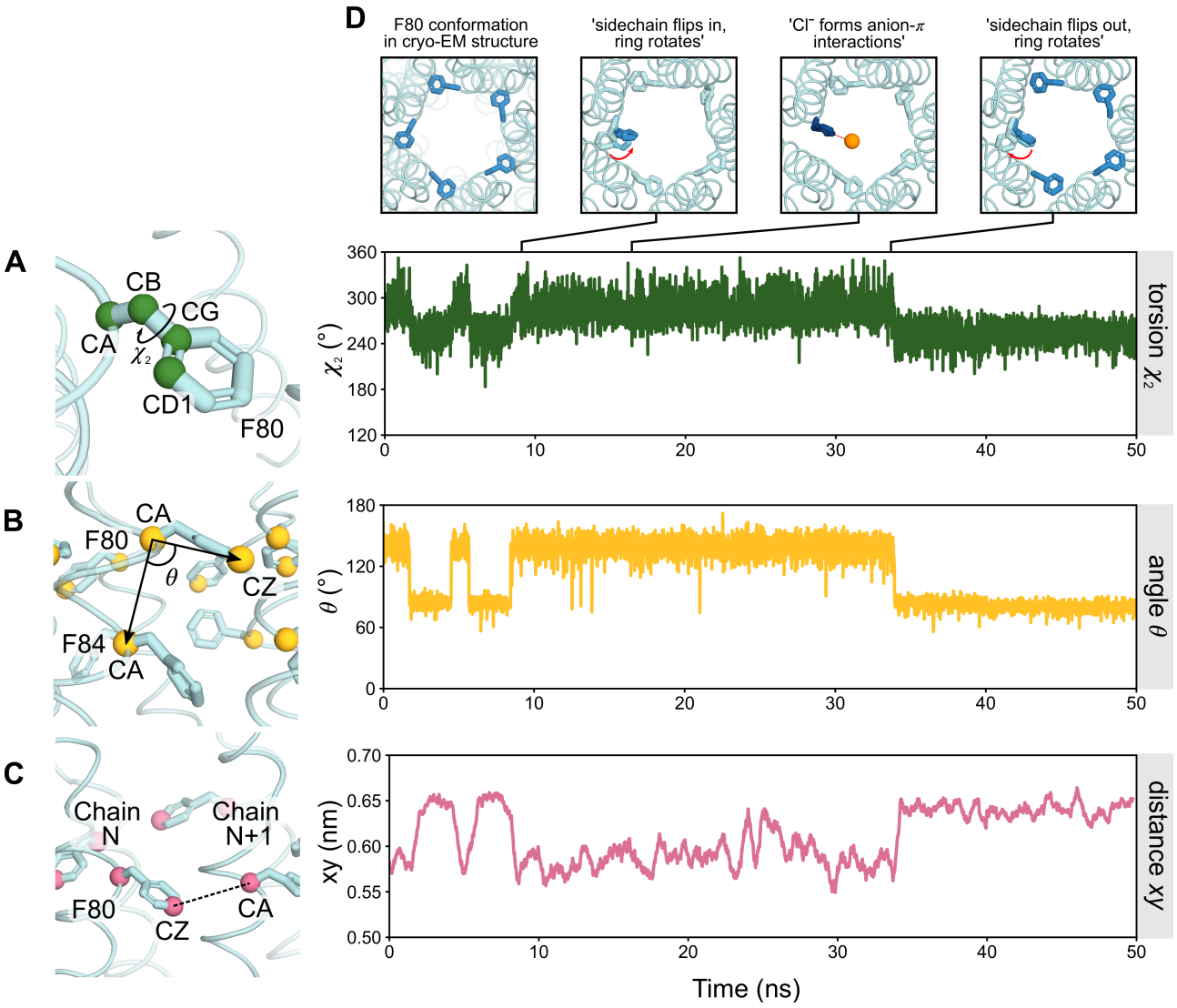

Figure 5.10 | Capturing Phenylalanine Sidechain Reorientation A Torsion angle χ_2 is defined as the angle CA-CB-CG-CD1. The torsion angle exhibits a discrete two state behaviour corresponding to the ring in a flat conformation where $\chi_2 \sim 255$ ° or vertical $\chi_2 \sim 285$ °. **B** Angle θ is defined as the angle between vectors given by F80 CA and F84 CA of the same chain, and F80 CA with F80 CZ from the aromatic ring. θ gives an indication of the vertical movement of the sidechain in the z-direction. **C** Distance xy defines the distance between atom CZ from one chain and the CA atom from F80 in the adjacent chain. This characterises the sidechain conformation as facing either inward towards the pore axis or outwards as it were in the experimental structure. The graph depicts the moving average over the last 50 frames of the simulation. **D** Snapshots exemplifying the flipping motion of a single F80 sidechain (chain 5). The sidechain simultaneously rotates from a flat aromatic ring position (χ_2), moves vertically in the positive z-direction (θ), and flips inwards (xy). Once in the inwards flipped conformation, the aromatic ring may form edgewise anion-π interactions with Cl^-. The flipping motion of the F80 sidechain is reversible.

In the fully open state, the sidechains of F80 are oriented in an outwards conformation relative to the pore axis. Here, we focus on the dynamics of a single chain (chain 5). At time $t \sim 10$ ns, the sidechain simultaneously moves in a vertical upwards direction (θ) whilst flipping inwards towards the pore (xy) and rotating the aromatic ring from a flat orientation to a vertical upright position (χ_2) (Figure 5.8 D). For the duration of the vertical ring-oriented conformation, a nearby Cl^- may then form edgewise anion-π interactions with the ring (discussed in Section 5.4.4). We observe that these Cl^--π interactions are transient. After ~ 25 ns, the sidechain spontaneously reorients back to the outwards-flipped conformation. Although the presence of a Cl^- does not necessarily induce the inward/outward flipping of F80 (Figure 5.11), the flipped sidechain may be functionally advantageous for ion permeation. However, there is no clear correlation between the motion of these sidechains and interactions with Cl^- in the immediate vicinity; at times, flipping occurs concurrently in the presence of a Cl^-, and at other times, it occurs spontaneously. This dynamic behaviour is reversible and is observed across multiple chains (Figure 5.9). Additionally, it is suggestive of more conformational diversity than is captured in the symmetrised cryo-EM structure. The breakdown of 5-fold symmetry has also been observed in GlyRs [293] and 7-fold symmetry in FLYC1 [44].

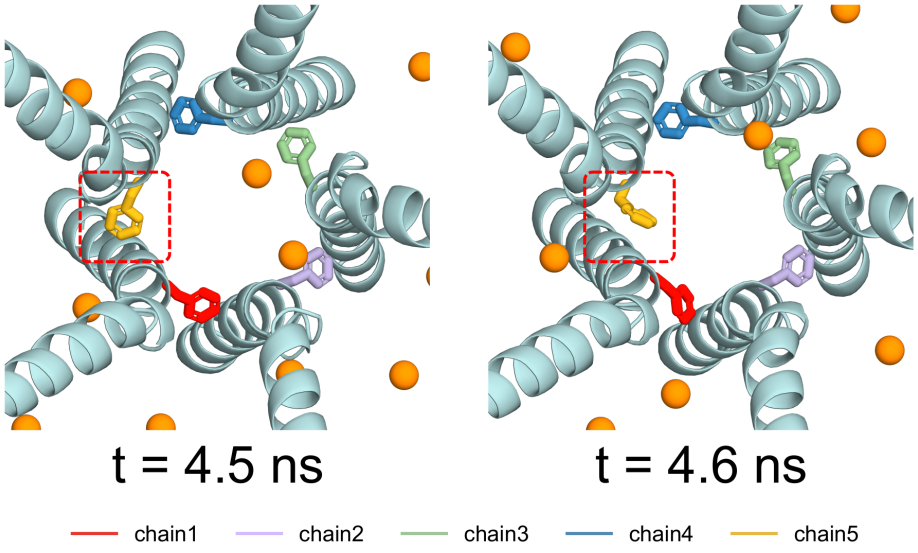

Figure 5.11 | F80 Sidechain Flipping Motion Snapshots demonstrating the spontaneous reversible inwards flipping of an F80 sidechain in the AMOEBA simulation. This motion is not necessarily induced by a passing Cl^- ion (orange spheres).

5.4.4 The Role of Water in Cl^- Permeation

Despite the dimensions of the open neck being wide enough to accommodate a hydrated Cl^-, we observe that permeation through the neck can be associated with the stripping of $\sim$ 1-3 water molecules from the first hydration shell of the ion (Figure 5.12 A). The dynamic inward ring flipping behaviour of F80 can aid the permeation of ions; the loss of a single water molecule from the first hydration shell of Cl^- is compensated by an interaction with the benzene ring from a single F80 sidechain and, similarly, the loss of more than one water molecule correlates to an interaction with two inwards-flipped sidechains from adjacent protomers [20, 294] (Figure 5.12 B). A maximum of two inwards-flipped sidechains can be observed in a single instance and always occur between two adjacent protomers.

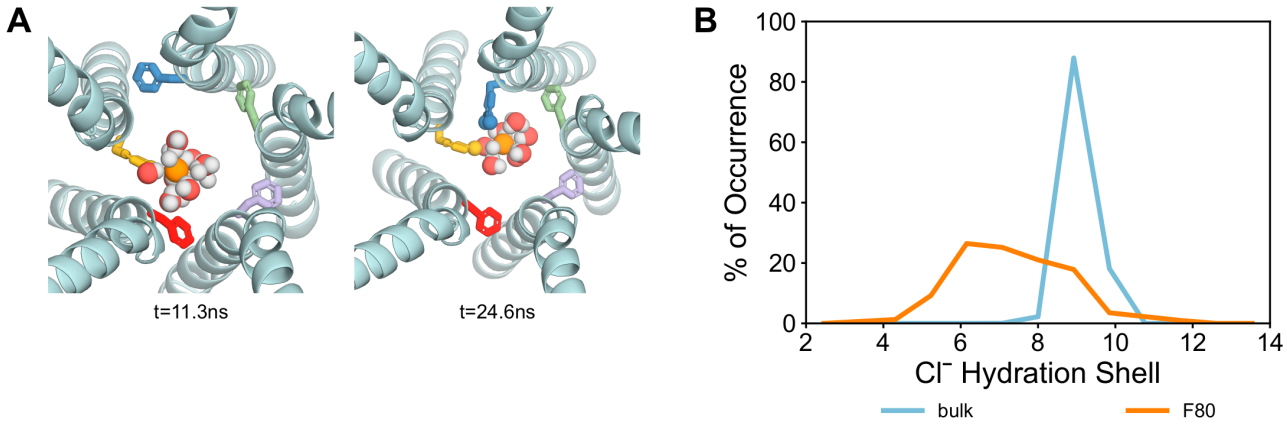

Figure 5.12 | Cl^- Partial Dehydration is Compensated by Anion-π Interactions A The partial loss of the first hydration shell is compensated by a direct interaction with an inward flipped aromatic ring from F80. Two adjacently flipped aromatic rings can facilitate the loss of multiple water molecules and be compensated through hydrophobic solvation. **B** Coordination number of Cl^- in bulk (blue) compared with the coordination of Cl^- at the z-position of F80 (orange) within the pore. Cl^- loses $\sim$ 1-3 water molecules in its first hydration shell at F80 relative to bulk.

The removal of water molecules allows for direct contact between Cl^- and the hydrophobic aromatic rings mediated through anion-π interactions. The benzene ring of the phenylalanine has a large negative quadrupole moment that creates a partial negative charge on both faces of the π-system and subsequently a partial positive charge around the edge of the ring [270, 272, 273, 295]. As a result, the partial loss of solvation of Cl^- can be compensated by favourable interactions with these aromatic rings which could be regarded as "hydrophobic solvation". It has been hypothesised [45, 47, 105, 282] that such a mechanism may mediate the stabilisation of a dehydrated Cl^- at F80 and F84 through anion-π interactions based on the initial X-ray structure of chicken Best1 (cBest1), although this structure was later determined to be in the closed conformation. Mutagenesis studies of the cBest1 neck residues to alanine later showed no effect on ion selectivity which suggested that the role of anion-π interactions may reduce the energy barriers for Cl^- and other anions to permeate the neck region but not be the determining factor for charge selectivity [284, 296]. Similarly, the energetically favourable partial

dehydration of Cl^- ions within simulations of the biomimetic nanopore in Chapter 3 and within the hydrophobic protein binding site in Chapter 4 reflect hydrophobic solvation mechanisms.

These anion-hydrophobic interactions were observed only when electronic polarisation was included in the molecular model and were not observed in equivalent simulations of the open state hBest1 when using fixed-charge (c36m) descriptions. We attribute this to the fact that anion-π interactions arise mainly from two factors: electrostatic interactions of the quadrupole moment and ion-induced polarisation [295]; the latter of which is not captured in c36m [297]. This is clearly reflected in the single chain RMSDs for c36m (Figure 5.13).

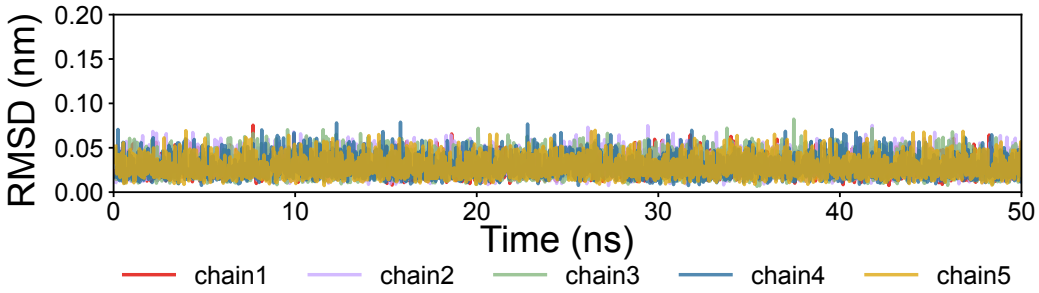

Figure 5.13 | RMSD of F80 Sidechains from the Open State RMSD of F80 sidechains in the open state protein fragment (PDB ID: 8D1O) using the c36m forcefield. The sidechains remain stable in a single conformation throughout the simulation.

5.4.5 Pathway for Conformational Change Between Open and Partially Open States

Simulations of the open state hint at a possible conformational transition pathway between the open and partially open states. The conformation of the inwards-flipped F80 sidechains in the open state resemble the configuration of F80 in the partially open state. In the partially open state, the F80 sidechains remain stable in an inward facing arrangement throughout the simulation with $\chi_2 \sim 275°$ (Figure 5.14 A), a torsion angle between that of the open state fluctuations (flat $\sim 255°$, vertical $\sim 285°$) (Figure 5.10 A). In the partially open state, the angle $\theta \sim 140°$ (Figure 5.14 B) is approximately the same as the position of F80 in the inwards-flipped state of the open neck (Figure 5.10 B at e.g. t $\sim$ 20 ns) which suggests the sidechain moves via an elevated vertical position towards a more closed state. The xy distances in Figure 5.14 C remain stable in their conformation. These fluctuations seen in F80 sidechains in the open state mimic the structure of the partially open state, suggesting a potential mechanism for transitioning of the neck residues from an open, conductive state to a partially open state and vice versa.

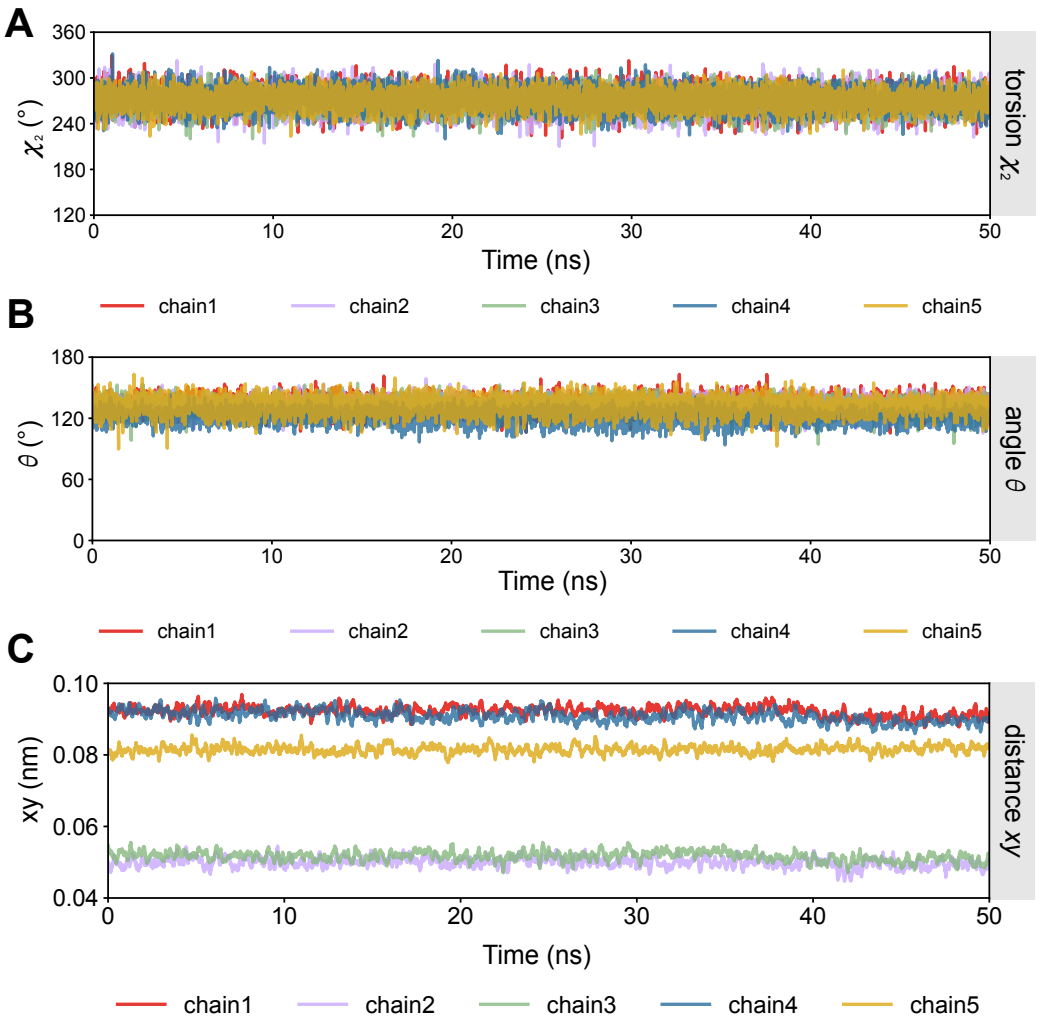

Figure 5.14 | Partially Open State Sidechain Dynamics of F80 Per chain **A** torsion angle χ_2, **B** angle θ and **C** distance xy of F80 in the AMOEBA simulation of the partially open state (PDB ID: 8D1K). All sidechains remain stable in this conformation.

5.4.6 The Partially Open Neck state

The partially open structure of hBest1 presents a more constricted neck region (Figure 5.3 E). Critical conformational changes occur in residues F282, F283, and F276 concomitant with changes in neck residues I76, F80 and F84. I76 remains facing away from the pore axis as in the fully open state, while residues F80, F84 and the hydrophobic gating apparatus adopt a closed-like conformation (Figure 5.3 G)[105]. In the experimental structure, an ion-like density is consistently captured at the edge of F84,

even when processed with C1 refinement (Figure 5.15 C adapted from [105]) i.e. with no symmetry applied. This density could potentially represent a chemical constituent of the buffer system, in which Cl^- is the predominant ion, therefore it may represent a dehydrated Cl^- ion. Furthermore, asymmetry in the neck can be observed as there is continuous movement between the closed and intermediate states in this region. We proceeded to investigate the identity of this experimentally observed density.

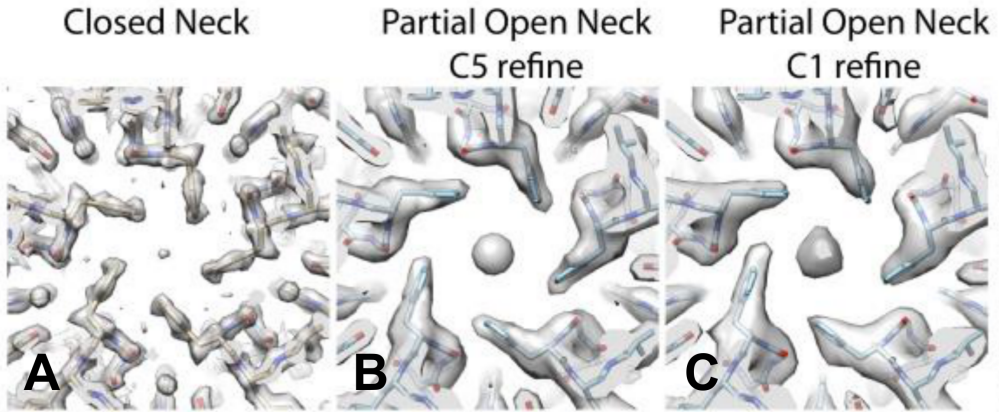

Figure 5.15 | Cryo-EM Map and Model of the Partially Open State at F84 Cryo-EM map and model of the **A** closed neck (PDB ID: 8D1I), **B** partially open neck (PDB ID: 8D1K) with C5 refinement applied and **C** the partially open neck with no symmetry applied. This depicts the presence of an ion-like density bound to the edge of F84. Figure adapted and reprinted from Owji *et al.* [105] (Figure S1 D) is licensed under a creative commons licence (CC by 4.0).

5.4.7 Analysis of the Partially Open Neck State

An initial assessment of this conformation was derived using a simple heuristic method developed by Rao *et al.* [70] which predicts the likelihood of channel dewetting or the existence of a hydrophobic gate based on the local pore radius and corresponding sidechain hydrophobicity. A heuristic score cut-off value of $\Sigma d > 0.55$ describes a non-conducting state. When this heuristic method was applied to the open, partially open and closed (PDB ID: 8D1I) structures and the aperture region was discarded from the calculation (i.e. removal of the cytoplasmic region as we were primarily interested in

the neck region), heuristic scores of 0, 2.71 and 4.57 were obtained respectively. Interestingly, this simulation-free heuristic model annotated this conformational state to be functionally closed i.e. dewetted and non-conducting, which may be expected given the minimum radius was $\sim$ 0.22 nm and the neck is very hydrophobic.

Equilibrium simulations of the embedded full protein in the partially open state with the additive c36m forcefield revealed a persistently dewetted neck region that is impermeable to ions (Figure 5.16 B). Consistent with the heuristic prediction, the hydrophobicity and inwards orientation of the aromatic sidechains of F80 and F84 obstruct the ion-conducting pathway to some degree. The tightest constriction of $\sim$ 0.22 nm radius is at F80 (see Figure 5.3 G) which presents a significant energetic barrier to permeation without complete steric occlusion. However at F84, the pore radius is $\sim$ 0.33 nm which is sufficiently large to accommodate for a partially dehydrated Cl^- [245], so contrary to observations from these simulations, we might anticipate to occasionally see Cl^- approach this region.

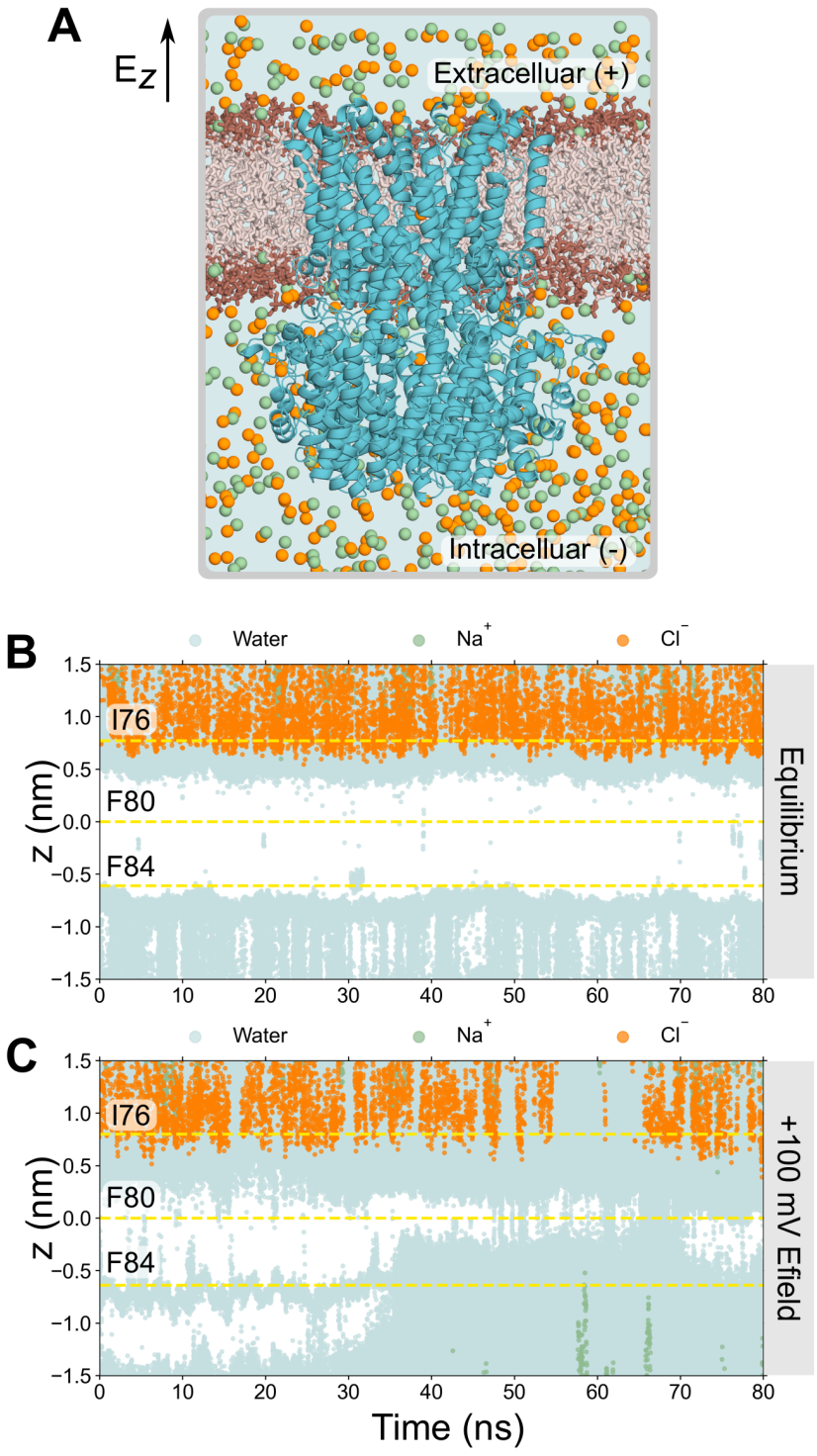

Figure 5.16 | Uniform External Electric Field Applied to the Partially Open State A Schematic of full protein embedded system. The partially open state protein (teal) is embedded into a lipid bilayer (brown) and solvated with water (light blue), Na^+ (green spheres) and Cl^- (orange spheres). Water and ion z-coordinate positions as a function of time within the neck region of the the full protein in **B** equilibrium and **C** external potential simulations.

External potential simulations were performed on the partially open state full protein to increase the possibility of observing pore wetting (and/or ion permeation). The neck remained largely dewetted up until time $t \sim 35$ ns and thereafter transient hydration of regions deeper inside the pore can be observed (Figure 5.16 C). Full water translocation events do occur (e.g. $t \sim 65 - 73$ ns), however these are not sustained and the neck continues to be dewetted, particularly between residues F80 and F84. Continuous water occupation just below F84 can be consistently observed over the whole simulation period which appears to correspond to interactions with the backbone of F84.

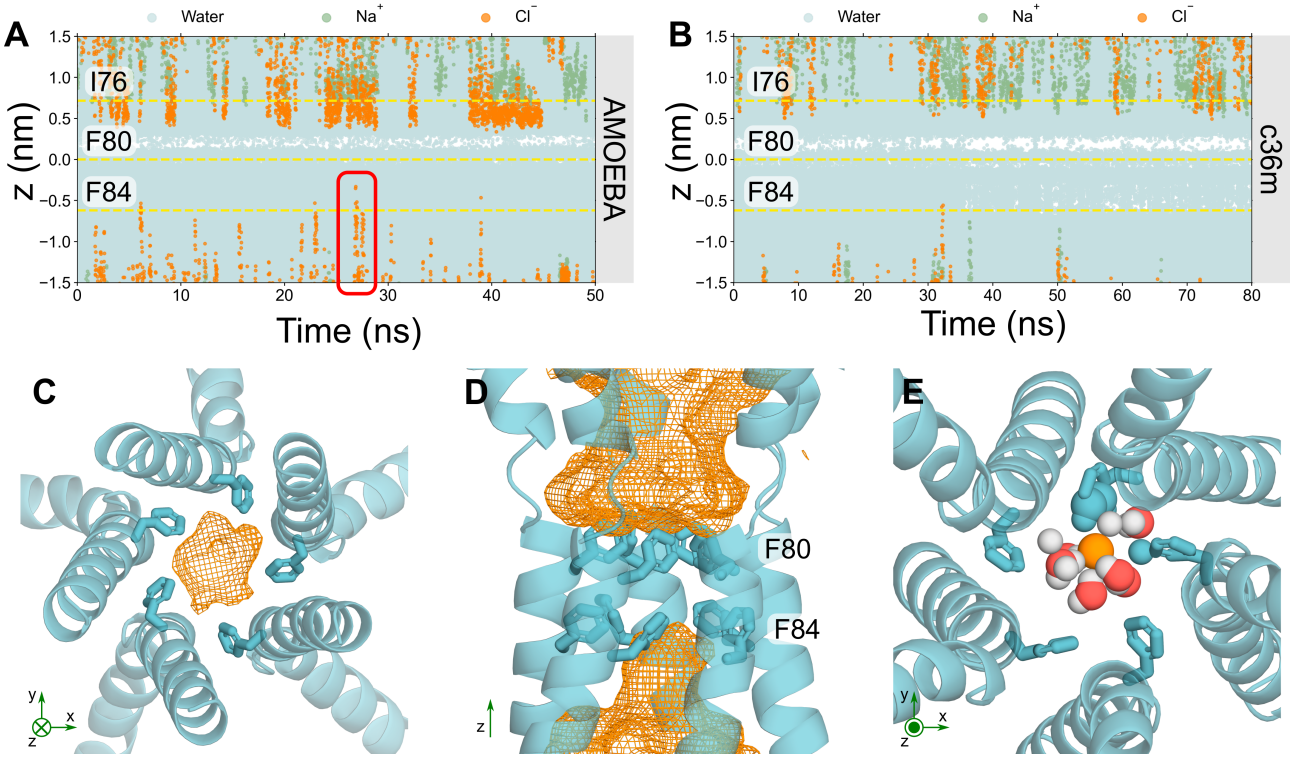

Figure 5.17 | The Partially Open State Trajectories of water (cyan), Na^+ (green) and Cl^- (orange) ions in z-coordinates as a function of time within the neck region of the partially open state pore (PDB ID: 8D1K) using **A** the AMOEBA forcefield or **B** c36m forcefield. The red box in **A** highlights a Cl^- that can occupy locations above F84 in the neck whereas nearly no ions can be seen in **B** with c36m. **C** Bottom-up view of F84 of the partially open state neck. The volume that Cl^- may occupy over the simulation is represented by the orange mesh. **D** Side-view of the neck region. F84 of the neck may accommodate for Cl^-; however, the neck is not permeable to Cl^-. **E** Top-down view of F84. Cl^- can be seen to interact with the aromatic ring through edgewise anion-π interactions. The coordinating atoms within the first solvation shell of the Cl^- (orange sphere) are represented as spheres.

Following this, we now explore the influence of polarisation on the behaviour of ions and water within the partially open state using the AMOEBA forcefield. Trajectories of ions and water in the partially open state protein fragment (Figure 5.17 A) see traces of Cl^- occupying locations just above F84 in the z-direction. This suggests that F84 is capable of accommodating Cl^-; however, the pore is functionally closed and no ions are conducted across the neck, primarily due to the tight hydrophobic constriction formed by F80. Given pore hydration acts as a precursor to ion permeation, a small, transiently dewetted region can be found within the neck (represented by the white band in Figure 5.17 A), corresponding to residues S79 and F80, which reiterates that the partially open state is functionally closed. This is more clearly depicted in Figure 5.17 C & D where the orange mesh depicts the volume of space that Cl^- occupies throughout the simulation and therefore the region between F80 and F84 is devoid of Cl^- (Figure 5.17 D). Mutagenesis studies have demonstrated that the conserved "IFF" motif of the neck acts as an effective gate, rather than a selectivity filter, because ionic selectivity remained constant in the mutant channel [214, 284] which is in agreement with our results.

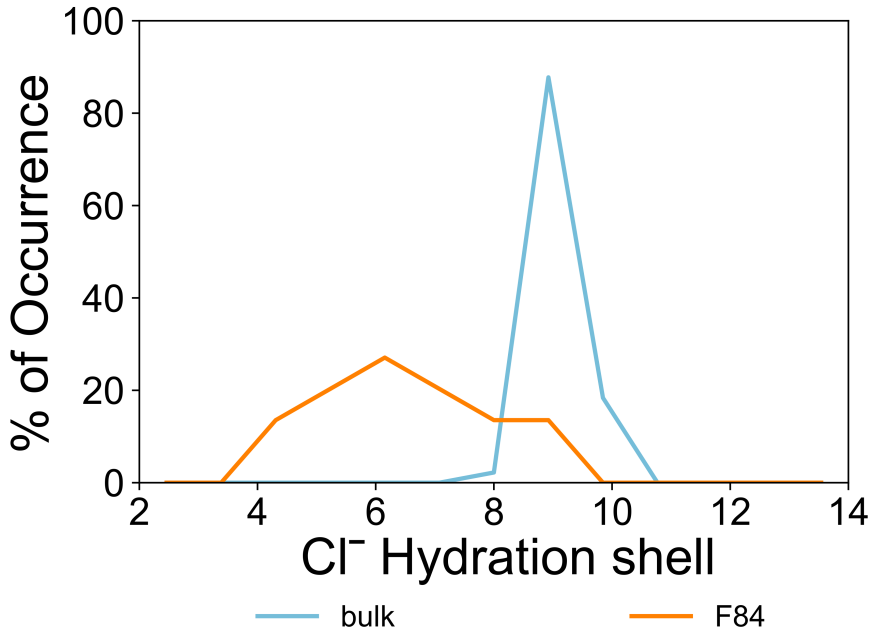

Figure 5.18 | Coordination Number of Cl^- in the Partially Open State Coordination number of Cl^- ions in bulk (blue) compared with Cl^- in the partially open state structure (PDB ID: 8D1K) at z-position corresponding to F84 (orange) within the pore in the AMOEBA simulation. Cl^- loses $\sim$ 1-3 water molecules in its first hydration shell at F84 relative to bulk. This is comparable to the dehydration that occurs at F80 in the fully open state (PDB ID: 8D1O)

F84 accommodates for a Cl^- through a similar mechanism to that seen in the previous Section 5.4.4: the Cl^- partially dehydrates (Figure 5.18) to directly interact with up to two adjacent pore-facing phenylalanines simultaneously, mediated by their edge-wise interaction with the partially dehydrated anion (Figure 5.17 E). The sidechains of F84 remain stable in their inward-facing conformation throughout the simulations as determined by the low RMSD (Figure 5.19 A). This two-state behaviour seen across the RMSD and torsion angle (χ_2) largely corresponds to a 180° reorientation of the aromatic ring (Figure 5.19 A & B). Chain 4 experiences some deviating motion which correlates with an outwards flipping of the sidechain for a duration of $\sim$ 7 ns (Figure 5.19 C & D) but otherwise the conformation of F84 remains relatively stable. Again, this hints at a possible transition pathway between the partially open and fully open neck states.

Together, these observations support functional annotation of the channel as partially open. It is indeed "on-the-way" to opening/closing with the neck containing a consistently dewetted region and presenting a barrier to ion permeation, yet the neck (F84) can accommodate Cl^-. This barrier to ion permeation may not necessarily be due to steric occlusion, but rather to the inability of Cl^- to shed more of its first hydration shell [298]. The partially open neck is wide enough to accommodate a dehydrated Cl^- (Figure 5.17 A & B) but it is energetically unfavourable to lose more than $\sim$ 2-3 water molecules in the neck region which would be required for permeation at the tightest constriction and therefore ions do not pass through this region.

For comparison, in the absence of polarisation, c36m simulations suggest the neck is functionally closed and there is a more prominent dewetted region at S79 and F80. F84 does not accommodate for Cl^- but water can occupy this region (Figure 5.17 B).

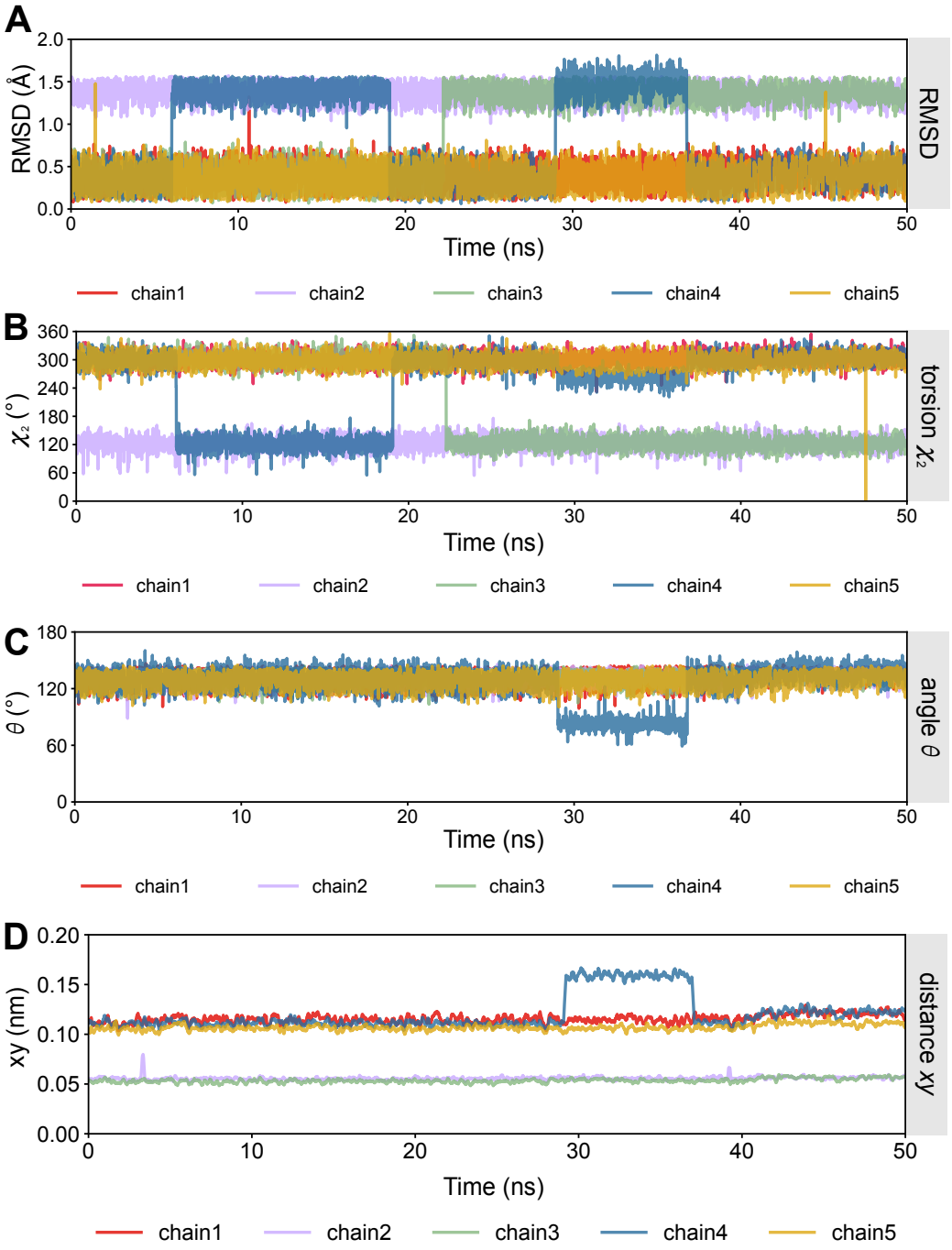

Figure 5.19 | Partially Open State Sidechain Dynamics of F84 Per chain **A** RMSD, **B** torsion angle χ_2, **C** angle θ and **D** distance xy of F84 in the partially open state (PDB ID: 8D1K) AMOEBA simulation. The sidechains are largely stable. The two-state behaviour seen in the RMSD and torsion angle correspond to a 180° rotation of the aromatic ring; however, due to the symmetric properties of the aromatic ring, this does not represent an alternative ring conformation.

5.5 Summary & Conclusions

In this chapter, we have performed MD simulations of the neck region of hBest1 channels in the open and partially open neck states using a fully polarisable forcefield (AMOEBA). Our results demonstrate that electronic polarisation plays a key role in Cl^- accumulation at the helix dipole (the backbone nitrogen of F80) in the open state with significant overlap with an ion-like density previously assigned to a water molecule in the cryo-EM structure. Together, our data suggest that this location functions as a Cl^- binding site that contributes to anion selectivity in the open neck conformation of bestrophins. In contrast, fixed charge (c36m) simulations predict a distinctly different location for Cl^- accumulation that is not supported by experimental data, thus underscoring the importance of forcefield choice in simulation-based functional annotation of polarisable systems. Cl^- permeation through the open neck region may be facilitated by edgewise anion-π interactions with the aromatic sidechains of F80 where, again, the influence of electronic polarisation appears to exert decisive influence over the functional state of the channel. These sidechains dynamically and reversibly flip inwards towards the pore axis to interact directly with a partially dehydrated Cl^-. Simulations of the partially open neck state also support this permeation mechanism as neck dilation at F84 is able to accommodate Cl^-; however, the pore remains functionally closed and is therefore "partially" open. Conversely, c36m simulations predict this partially open state structure to be closed. This chapter therefore provides new insights into bestrophin channel function although further studies are required to separate the individual contributions of structural elements to Cl^- permeation and selectivity within the ion pathway.

Overall, our results demonstrate the importance of modelling polarisation in situations where polarisable moieties play an essential role in protein function. Appropriate treatment of electrostatics reveals more physically accurate behaviour and provides new

mechanistic insights into Cl^- permeation. MD simulations employing explicit polarisation can therefore complement cryo-EM and other structural techniques in the identification and assignment of ambiguous non-protein densities such as anions.

Significant advancements towards elucidating anion selectivity mechanisms have already informed the design of small synthetic anion carriers such as macrocyclic anionophores called *biotin[6]uril*. These synthetic carriers can purposefully exploit hydrophobic CH–Cl^- interactions to selectively transport Cl^- ions across membranes [172, 299]. Similarly, novel artificial channel molecules can be tuned for Cl^- selectivity and to mimic the shape and function of CLC channels [300]. These structurally simple designs exploit anion-π interactions within the binding sites to facilitate Cl^- dehydration. The role of anion-π interactions offer exciting new applications however their role in biological anion transport is still in its infancy.

Since submission, an error was found in the data. The AMOEBA forcefield simulations have been rerun and analysis of data was able to recover the correct Cl^--O RDF and hydration shell in line with previous studies in Chapter 4 and lie within the experimental range of Cl^- hydration numbers. Consequently, this has in fact lead to even better agreement for Cl^- localisation with the experimental density when PyLipid was run to locate the binding site. Please refer to the paper for the most updated findings.

"There are always flowers for those who want to see them."

— Henri Matisse

6

Conclusions and Future Directions

This thesis has sought to investigate the influence of electronic polarisation on anion-hydrophobic interactions within nanopores and ion channels. Starting with simple model systems in Chapter 3 and moving progressively towards more realistic biological systems in Chapters 4 and 5, we have studied the effect of polarisation on interactions of Cl^- and water. The overarching results of this thesis are summarised in Figure 6.1.

In Chapter 3, the groundwork was laid with preliminary simulations of aqueous/hydrophobic (decane) interfaces to assess the applicability and robustness of the ECC method [101, 184] as a means for improved descriptions of induced polarisation effects within additive forcefields. Model biomimetic nanopores were designed as templates for investigating anion-hydrophobic interactions within a confined environment. Analysis of in-pore Cl^- interactions revealed surface adsorption to the hydrophobic pore wall concomitant with partial Cl^- dehydration to directly interact with the hydrophobic interface (Figure 6.1 B). Modelling induced polarisation is required to reproduce these experimentally-supported behaviours which are otherwise not captured in standard non-polarisable forcefields.

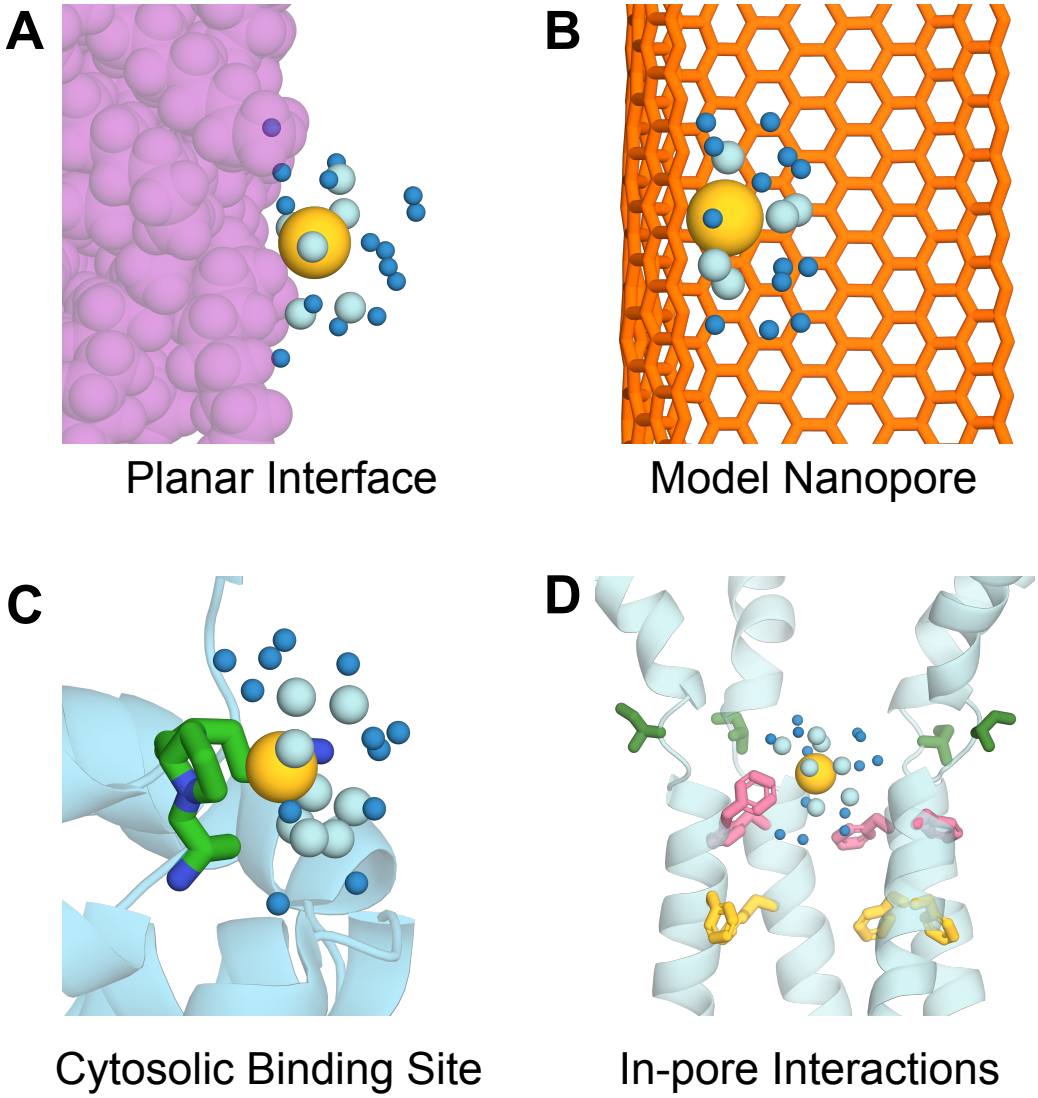

Figure 6.1 | Summary of Cl^--Hydrophobic Interactions in this Thesis The overarching theme of this thesis demonstrates the partial dehydration of Cl^- ions (yellow spheres) to directly interact with hydrophobic surfaces when polarisation is modelled in the system. Chapter 3 explored Cl^- interactions at a **A** planar aqueous (omitted)/decane (purple) interface and within a **B** model biomimetic nanopore (orange). Chapter 4 described Cl^- interactions with a **C** hydrophobic binding site (green, side view) from ClR. Finally, Chapter 5 described how the partial Cl^- dehydration is compensated by **D** anion-π interactions from aromatic residues (F80 in pink) within the hydrophobic neck of hBest1 (chain removed for clarity). The first hydration shell of Cl^- is represented by the cyan spheres and subsequently the smaller, darker blue spheres denote the second ion hydration shell.

Whilst the ECC method has demonstrated to be an effective compromise to explicit modelling of electrostatics, it is limited to systems of homogeneous dielectric mediums and the assumption that all particles have equal polarisability [184, 301]. With the development of more powerful computational resources, next generation forcefield simulations, longer timescales and *ab initio* calculations are becoming more accessible,

enabling forcefield parameterisation to reach quantum-level accuracy. Importantly, parameterisation of physically accurate water models capable of predicting structural, electronic and dynamic properties that can reproduce a range of experimental benchmarks across the phase diagram continues to sustain interest [302–304]. Machine learning potentials are rapidly emerging to optimise model development and used in combination with existing techniques have shown promise in bridging the accuracy gap between empirical forcefields and *ab initio* methods [305, 306].

Building upon Chapter 3, Chapter 4 revealed analogous anion-hydrophobic interactions within a defined hydrophobic binding site from the Cl^- rhodopsin in simulations employing the prosECCo [190] and AMOEBA forcefields but not c36. Mechanistically, Cl^- partially dehydrates to directly associate with the hydrophobic contacts of the site (Figure 6.1 C) to achieve stronger and longer binding events when (effective) polarisation is modelled or otherwise Cl^- does not bind. Associated energetic profiles reflected these findings and further revealed a second metastable free energy minimum when backbone restraints were applied to maintain the crystal structure in prosECCo simulations. A closer look suggested Cl^- formed favourable interactions with hydrophobic residues from a nearby cytosolic loop to facilitate rebinding events.

Chapter 4 exemplified the use of the computationally efficient prosECCo forcefield to act as a proxy for detecting potentially interesting interactions for further exploration using more advanced molecular models and methods. Methodological questions are raised regarding the appropriate level of granularity or molecular detail of models required to capture the correct physics for the study of polarisable moieties and their complex interactions. This chapter examined the cytosolic binding site in isolation and interactions of freely diffusing particles in bulk solution. For a more realistic perspective, further research could dissect Cl^- interactions along the entire pore-conducting pathway

of ClR from Cl^- uptake at the extracellular side through to Cl^- release at the cytosolic binding site [256, 307].

Chapter 5 proceeds to investigate in-pore interactions of Cl^- within the neck region of the open and partially open states of the human bestrophin 1 channel. We have demonstrated how MD simulations employing more physically realistic forcefields can be used to aid assignment and interpretation of ambiguous densities from cryo-EM data to access improved atomic resolution structures. Incorporating polarisation revealed permeation through the neck was achieved through hydrophobic solvation of Cl^-, concurrent with ion dehydration, and is mediated by edgewise anion-π interactions with the aromatic sidechains of pore-facing phenylalanine residues (F80 and F84) (Figure 6.1 D). Such interactions occurring in the open neck state originate from asymmetric modes of the pore driven by spontaneous sidechain reorientation of F80. A similar molecular mechanism is observed for Cl^- at the neck dilation of F84 in the partially open state which are supported by cryo-EM densities at similar locations, supporting the functional annotation of the channel as partially open. Ultimately, including polarisation in the model tunes the state of hBest1; neglecting to results in failure to capture the correct physics from which anion-π interactions arise.

Chapter 5 contributes towards the long-standing discussion surrounding factors governing anion selectivity. In particular, selectivity between anions amongst the halide series could be probed further by investigating new anion selective structures, for example, the Na^+/I^- symporter (PDB ID: 7UV0) [37] is selective for I^- over Cl^- despite the concentration of Cl^- in the extracellular environment being 10^5 times of I^-. Additionally, I^- is much larger and more polarisable (~ 0.006 nm^3) compared with Cl^- (~ 0.003 nm^3) [308]. Furthermore, pore asymmetry and asymmetric movements are being increasingly observed in ion channels [44, 293, 309]. However, existing tools and methods for

heterogeneous reconstruction present a bottleneck in the accessible resolution of structures resulting from simplified assumptions based on observed heterogeneity. Development of advanced algorithms for heterogeneous reconstruction seeking to model the continuous nature of molecules using deep neural networks have begun to make headway [310–312].

Finally, obtaining a more complete understanding of the molecular mechanisms underlying anion selectivity can inform the design of synthetic anion-selective structures. This has implications not only for biomedical and therapeutic applications [172, 300, 313] but also holds potential for impact in technological applications e.g. water desalination technologies [52, 61, 298].

Bibliography

1. Phillips, R., Kondev, J., Theriot, J., Garcia, H. G. & Orme, N. *Physical Biology of the Cell* 2nd ed. (Garland Science, 2013).
2. Ashcroft, F. *Ion Channels and Disease* (Academic Press, 2000).
3. Roux, B. Ion channels and ion selectivity. *Essays in Biochemistry* **61,** 201–209 (2017).
4. Di Resta, C. & Becchetti, A. *Integrins and Ion Channels.* 9–22 (Springer, 2010).
5. Hille, B. *Ionic Channels of Excitable Membranes* 3rd ed. (Sinauer Associates Inc, 2001).
6. Corry, B. Understanding ion channel selectivity and gating and their role in cellular signalling. *Molecular BioSystems* **2,** 527–535 (2006).
7. Zheng, J. & Trudeau, M. C. *Handbook of Ion Channels* 1st ed. (CRC Press, 2016).
8. Dutzler, R., Campbell, E. B. & Mackinnon, R. Gating the selectivity filter in ClC chloride channels. *Science* **300,** 108–112 (2003).
9. Sörmann, J. *et al.* Gain-of-function mutations in KCNK3 cause a developmental disorder with sleep apnea. *Nature Genetics* **54,** 1534–1543 (2022).
10. Linsdell, P. Anion conductance selectivity mechanism of the CFTR chloride channel. *Biochimica et Biophysica Acta - Biomembranes* **1858,** 740–747 (2016).
11. Roden, D. M. & George, A. L. The cardiac ion channels: Relevance to management of arrhythmias. *Annual Review of Medicine* **47,** 135–183 (1996).
12. Catterall, W. A. Sodium channels, inherited epilepsy, and antiepileptic drugs. *Annual Review of Pharmacology and Toxicology* **54,** 317–338 (2014).
13. Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nature Reviews Drug Discovery* **16,** 19–34 (2016).
14. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nature Reviews Drug Discovery* **5,** 993–996 (2006).
15. Lynch, C. I., Klesse, G., Rao, S., Tucker, S. J. & Sansom, M. S. P. Water nanoconfined in a hydrophobic pore: Molecular dynamics simulations of Transmembrane Protein 175 and the influence of water models. *ACS Nano* **15,** 19098–19108 (2021).
16. Gouaux, E. & Mackinnon, R. Principles of selective ion transport in channels and pumps. *Science* **310,** 1461–1465 (2005).
17. Roux, B. *et al.* Ion selectivity in channels and transporters. *Journal of General Physiology* **137,** 415–426 (2011).
18. Kopec, W., Rothberg, B. S. & de Groot, B. L. Molecular mechanism of a potassium channel gating through activation gate-selectivity filter coupling. *Nature Communications* **10,** 1–15 (2019).

19. Dudev, T. & Lim, C. Factors governing the Na+ vs K+ selectivity in sodium ion channels. *Journal of the American Chemical Society* **132,** 2321–2332 (2010).

20. Gomez, D. T., Pratt, L. R., Asthagiri, D. N. & Rempe, S. B. Hydrated anions: from clusters to bulk solution with quasi-chemical theory. *Accounts of Chemical Research* **55,** 2201–2212 (2022).

21. Leisle, L. *et al.* Backbone amides are determinants of Cl selectivity in CLC ion channels. *Nature Communications* **13,** 1–11 (2022).

22. Cheng, Y. Membrane protein structural biology in the era of single particle cryo-EM. *Current Opinion in Structural Biology* **52,** 58–63 (2018).

23. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Research* **28,** 235–242 (2000).

24. Doyle, D. A. *et al.* The structure of the potassium channel: Molecular basis of K+ conduction and selectivity. *Science* **280,** 69–77 (1998).

25. Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. Overcoming the challenges of membrane protein crystallography. *Current Opinion in Structural Biology* **18,** 581–586 (2008).

26. Renaud, J. P. *et al.* Cryo-EM in drug discovery: Achievements, limitations and prospects. *Nature Reviews Drug Discovery* **17,** 471–492 (2018).

27. Earl, L. A., Falconieri, V., Milne, J. L. & Subramaniam, S. Cryo-EM: Beyond the microscope. *Current Opinion in Structural Biology* **46,** 71–78 (2017).

28. D'Imprima, E. & Kühlbrandt, W. Current limitations to high-resolution structure determination by single-particle cryoEM. *Quarterly Reviews of Biophysics* **54,** 1–15 (2021).

29. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589 (2021).

30. Bertoline, L. M., Lima, A. N., Krieger, J. E. & Teixeira, S. K. Before and after AlphaFold2: An overview of protein structure prediction. *Frontiers in Bioinformatics* **3,** 1–8 (2023).

31. Neher, E. & Sakmann, B. Single-channel currents recorded from membrane of denervated frog muscle fibres. *Nature* **260,** 799–802 (1976).

32. Sakmann, B. & Neher, E. Patch clamp techniques for studying ionic channels in excitable membranes. *Annual Review of Physiology* **46,** 455–72 (1984).

33. Flood, E., Lev, B., Vorobyov, I. & Allen, T. W. Atomistic simulations of membrane ion channel conduction, gating, and modulation. *Chemical Reviews* **119,** 7737–7832 (2019).

34. Massey, M. K., Reiterman, M. J., Mourad, J. & Luckie, D. B. Is CFTR an exchanger?: Regulation of HCO3- Transport and extracellular pH by CFTR. *Biochemistry and Biophysics Reports* **25,** 1–8 (2021).

35. Park, E., Campbell, E. B. & MacKinnon, R. Structure of a CLC chloride ion channel by cryo-electron microscopy. *Nature* **541,** 500–505 (2017).

36. Yeagle, P. L. *The Membranes of Cells* 3rd ed. (Academic Press, 2016).

37. Ravera, S. *et al.* Structural insights into the mechanism of the sodium/iodide symporter. *Nature* **612,** 795–801 (2022).

38. Stockbridge, R. B. *et al.* Crystal structures of a double-barrelled fluoride ion channel. *Nature* **525,** 548–551 (2015).

39. Yue, Z., Wang, Z. & Voth, G. A. Ion permeation, selectivity, and electronic polarization in fluoride channels. *Biophysical Journal* **121,** 1–12 (2022).

40. Skitchenko, R. K., Usoltsev, D., Uspenskaya, M., Kajava, A. V. & Guskov, A. Census of halide-binding sites in protein structures. *Bioinformatics* **36,** 3064–3071 (2020).

41. Kim, K. *et al.* Crystal structure and functional characterization of a light-driven chloride pump having an NTQ motif. *Nature Communications* **7,** 1–10 (2016).

42. Kunzelmann, K. TMEM16, LRRC8A, bestrophin: Chloride channels controlled by Ca2+ and cell volume. *Trends in Biochemical Sciences* **40,** 535–543 (2015).

43. Le, S. C., Liang, P., Lowry, A. J. & Yang, H. Gating and regulatory mechanisms of TMEM16 ion channels and scramblases. *Frontiers in Physiology* **12,** 1–14 (2021).

44. Jojoa-Cruz, S. *et al.* Structural insights into the venus flytrap mechanosensitive ion channel Flycatcher1. *Nature Communications* **13,** 1–11 (2022).

45. Owji, A. P. *et al.* Structural and functional characterization of the bestrophin-2 anion channel. *Nature Structural and Molecular Biology* **27,** 382–391 (2020).

46. Owji, A. P., Kittredge, A., Zhang, Y. & Yang, T. Structure and function of the Bestrophin family of calcium-activated chloride channels. *Channels* **15,** 604–623 (2021).

47. Dickson, V. K., Pedi, L. & Long, S. B. Structure and insights into the function of a Ca2+-activated Cl- channel. *Nature* **516,** 13–218 (2014).

48. Zhu, H. & Gouaux, E. Architecture and assembly mechanism of native glycine receptors. *Nature* **599,** 513–517 (2021).

49. Zhu, H. Structure and mechanism of glycine receptor elucidated by cryo-electron microscopy. *Frontiers in Pharmacology* **13,** 1–7 (2022).

50. Laverty, D. *et al.* Cryo-EM structure of the human $\alpha 1\beta 3\gamma 2$ GABAA receptor in a lipid bilayer. *Nature* **565,** 516–520 (2019).

51. Zhu, S. *et al.* Structure of a human synaptic GABAA receptor. *Nature* **559,** 67–88 (2018).

52. Corry, B. Designing carbon nanotube membranes for efficient water desalination. *Journal of Physical Chemistry B* **112,** 1427–1434 (2008).

53. Zhu, Z., Wang, D., Tian, Y. & Jiang, L. Ion/molecule transportation in nanopores and nanochannels: from critical principles to diverse functions. *Journal of the American Chemical Society* **141,** 8658–8669 (2019).

54. Kuyucak, S. & Bastug, T. Physics of ion channels. *Journal of Biological Physics* **29,** 429–446 (2003).

55. Yuan, Y. D. *et al.* Porous organic cages as synthetic water channels. *Nature Communications* **11,** 1–10 (2020).

56. Calvelo, M., Lynch, C. I., Granja, J. R., Sansom, M. S. & Garcia-Fandiño, R. Effect of water models on transmembrane self-assembled cyclic peptide nanotubes. *ACS Nano* **15,** 7053–7064 (2021).

57. Ghadiri, M. R., Granja, J. R. & Buehler, L. K. Artificial transmembrane ion channels from self-assembling peptide nanotubes. *Nature* **369,** 301–304 (1994).

58. Amiri, H., Shepard, K. L., Nuckolls, C. & Hernández Sánchez, R. Single-walled carbon nanotubes: Mimics of biological ion channels. *Nano Letters* **17,** 1204–1211 (2017).

59. Buchsbaum, S. F. *et al.* Fast permeation of small ions in carbon nanotubes. *Advanced Science* **8,** 1–7 (2021).

60. Li, Z. *et al.* Strong differential monovalent anion selectivity in narrow diameter carbon nanotube porins. *ACS Nano* **14,** 6269–6275 (2020).

61. Qin, L., Huang, H. & Zhou, J. Modulation of ion transport through nanopores in water desalination: a molecular dynamics study. *Molecular Simulation* **49,** 1742–1757 (2023).

62. Yazdani, M., Jia, Z. & Chen, J. Hydrophobic dewetting in gating and regulation of transmembrane protein ion channels. *Journal of Chemical Physics* **153,** 1–9 (2020).

63. Lynch, C. I., Rao, S. & Sansom, M. S. Water in nanopores and biological channels: A molecular simulation perspective. *Chemical Reviews* **120,** 10298–10335 (2020).

64. Beckstein, O., Biggin, P. C. & Sansom, M. S. A hydrophobic gating mechanism for nanopores. *Journal of Physical Chemistry B* **105,** 12902–12905 (2001).

65. Beckstein, O. & Sansom, M. S. P. Liquid-vapor oscillations of water in hydrophobic nanopores. *Proceedings of the National Academy of Sciences of the United States of America* **100,** 7063–7068 (2003).

66. Beckstein, O. *et al.* Ion channel gating: Insights via molecular simulations. *FEBS Letters* **555,** 85–90 (2003).

67. Beckstein, O. & Sansom, M. S. The influence of geometry, surface character, and flexibility on the permeation of ions and water through biological pores. *Physical Biology* **1,** 42–52 (2004).

68. Smart, O. S., Neduvelil, J. G., Wang, X., Wallace, B. A. & Sansomt, M. S. P. HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *Journal of Molecular Graphics* **14,** 354–360 (1996).

69. Klesse, G., Rao, S., Sansom, M. S. P. & Tucker, S. J. CHAP: A versatile tool for the structural and functional annotation of ion channel pores. *Journal of Molecular Biology* **431,** 3353–3365 (2019).

70. Rao, S., Klesse, G., Stansfeld, P. J., Tucker, S. J. & Sansom, M. S. P. A heuristic derived from analysis of the ion channel structural proteome permits the rapid identification of hydrophobic gates. *Proceedings of the National Academy of Sciences of the United States of America* **116,** 13989–13995 (2019).

71. Okur, H. I. *et al.* Beyond the Hofmeister series: ion-specific effects on proteins and their biological functions. *Journal of Physical Chemistry B* **121,** 1997–2014 (2017).

72. Peng, M. & Nguyen, A. V. Adsorption of ionic surfactants at the air-water interface: The gap between theory and experiment. *Advances in Colloid and Interface Science* **275,** 1–20 (2020).

73. Schwierz, N., Horinek, D., Sivan, U. & Netz, R. R. Reversed Hofmeister series—The rule rather than the exception. *Current Opinion in Colloid and Interface Science* **23,** 10–18 (2016).

74. Levin, Y. Polarizable ions at interfaces. *Physical Review Letters* **102,** 1–4 (2009).

75. Mitropoulos, A. C. What is a surface excess? *Journal of Engineering Science and Technology Review* **1,** 1–3 (2008).

76. Jungwirth, P. & Tobias, D. J. Specific ion effects at the air/water interface. *Chemical Reviews* **106,** 1259–1281 (2006).

77. Tůma, L., Jeníček, D. & Jungwirth, P. Propensity of heavier halides for the water/vapor interface revisited using the Amoeba force field. *Chemical Physics Letters* **411,** 70–74 (2005).

78. Jungwirth, P. & Tobias, D. J. Molecular structure of salt solutions: A new view of the interface with implications for heterogeneous atmospheric chemistry. *Journal of Physical Chemistry B* **105,** 10468–10472 (2001).

79. Vazdar, M., Pluhařová, E., Mason, P. E., Vácha, R. & Jungwirth, P. Ions at hydrophobic aqueous interfaces: Molecular dynamics with effective polarization. *Journal of Physical Chemistry Letters* **3,** 2087–2091 (2012).

80. Tobias, D. J., Stern, A. C., Baer, M. D., Levin, Y. & Mundy, C. J. Simulation and theory of ions at atmospherically relevant aqueous liquid-air interfaces. *Annual Review of Physical Chemistry* **64,** 339–359 (2013).

81. Kang, B., Tang, H., Zhao, Z. & Song, S. Hofmeister series: Insights of ion specificity from amphiphilic assembly and interface property. *ACS Omega* **5,** 6229–6239 (2020).

82. Petersen, P. B. & Saykally, R. J. On the nature of ions at the liquid water surface. *Annual Review of Physical Chemistry* **57,** 333–364 (2006).

83. Caleman, C., Hubb, J. S., Van Maaren, P. J. & Van Der Spoel, D. Atomistic simulation of ion solvation in water explains surface preference of halides. *Proceedings of the National Academy of Sciences of the United States of America* **108,** 6838–6842 (2011).

84. Liu, D., Ma, G., Levering, L. M. & Allen, H. C. Vibrational spectroscopy of aqueous sodium halide solutions and air-liquid interfaces: Observation of increased interfacial depth. *Journal of Physical Chemistry B* **108,** 2252–2260 (2004).

85. Garrett, B. C. Ions at the Air/Water Interface. *Science* **303,** 1146–1147 (2004).

86. Tong, Y., Zhang, I. Y. & Campen, R. K. Experimentally quantifying anion polarizability at the air/water interface. *Nature Communications* **9,** 1–8 (2018).

87. Piatkowski, L., Zhang, Z., Backus, E. H., Bakker, H. J. & Bonn, M. Extreme surface propensity of halide ions in water. *Nature Communications* **5,** 1–7 (2014).

88. Thaunay, F., Ohanessian, G. & Clavaguéra, C. Dynamics of ions in a water drop using the AMOEBA polarizable force field. *Chemical Physics Letters* **671,** 131–137 (2017).

89. DeMarco, K. R., Bekker, S. & Vorobyov, I. Challenges and advances in atomistic simulations of potassium and sodium ion channel gating and permeation. *Journal of Physiology* **597,** 679–698 (2019).

90. Guardiani, C. *et al.* Computational methods and theory for ion channel research. *Advances in Physics: X* **7,** 1–74 (2022).

91. Musgaard, M., Paramo, T., Domicevica, L., Andersen, O. J. & Biggin, P. C. Insights into channel dysfunction from modelling and molecular dynamics simulations. *Neuropharmacology* **132,** 20–30 (2018).

92. Kutzner, C., Grubmüller, H., De Groot, B. L. & Zachariae, U. Computational electrophysiology: The molecular dynamics of ion channel permeation and selectivity in atomistic detail. *Biophysical Journal* **101,** 809–817 (2011).

93. Oakes, V., Furini, S. & Domene, C. Insights into the mechanisms of K+ permeation in K+ channels from computer simulations. *Journal of Chemical Theory and Computation* **16,** 794–799 (2020).

94. Åqvist, J. & Luzhkov, V. Ion permeation mechanism of the potassium channel. *Nature* **404,** 881–884 (2000).

95. Köpfer, D. A. *et al.* Ion permeation in K+ channels occurs by direct coulomb knock-on. *Science* **346,** 352–355 (2014).

96. Kopec, W. *et al.* Direct knock-on of desolvated ions governs strict ion selectivity in K+ channels. *Nature Chemistry* **10,** 813–820 (2018).

97. Jing, Z. *et al.* Polarizable force fields for biomolecular simulations: Recent advances and applications. *Annual Review of Biophysics* **48,** 371–394 (2019).

98. Ngo, V. *et al.* Polarization effects in water-mediated selective cation transport across a narrow transmembrane channel. *Journal of Chemical Theory and Computation* **17,** 1726–1741 (2021).

99. Peng, X. *et al.* Accurate evaluation of ion conductivity of the gramicidin A channel using a polarizable force field without any corrections. *Journal of Chemical Theory and Computation* **12,** 2973–2982 (2016).

100. Klesse, G., Rao, S., Tucker, S. J. & Sansom, M. S. Induced polarization in molecular dynamics simulations of the 5-HT3 receptor channel. *Journal of the American Chemical Society* **142,** 9415–9427 (2020).

101. Duboue-Dijon, E., Javanainen, M., Delcroix, P., Jungwirth, P. & Martinez-Seara, H. A practical guide to biologically relevant molecular simulations with charge scaling for electronic polarization. *Journal of Chemical Physics* **153,** 1–15 (2020).

102. Yoo, J. & Aksimentiev, A. New tricks for old dogs: Improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions. *Physical Chemistry Chemical Physics* **20,** 8432–8449 (2018).

103. Jing, Z. *et al.* Thermodynamics of ion binding and occupancy in potassium channels. *Chemical Science* **12,** 8920–8930 (2021).

104. Orabi, E. A., Öztürk, T. N., Bernhardt, N. & Faraldo-Gómez, J. D. Corrections in the CHARMM36 parametrization of chloride interactions with proteins, lipids, and alkali cations, and extension to other halide anions. *Journal of Chemical Theory and Computation* **17,** 6240–6261 (2021).

105. Owji, A. P. *et al.* Structures and gating mechanisms of human bestrophin anion channels. *Nature Communications* **13,** 1–11 (2022).

106. Csizi, K. S. & Reiher, M. Universal QM/MM approaches for general nanoscale applications. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **13,** 1–29 (2023).

107. Monticelli, L. & Salonen, E. *Biomolecular simulations: Methods and protocols* (Springer, 2013).

108. Lemkul, J. A. Pairwise-additive and polarizable atomistic force fields for molecular dynamics simulations of proteins. *Progress in Molecular Biology and Translational Science* **170,** 1–71 (2020).

109. Brooks, B. R. *et al.* CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **30,** 1545–1614 (2009).

110. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society* **118,** 11225–11236 (1996).

111. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general Amber force field. *Journal of Computational Chemistry* **25,** 1157–1174 (2004).

112. Christen, M. *et al.* The GROMOS software for biomolecular simulation: GROMOS05. *Journal of Computational Chemistry* **26,** 1719–1751 (2005).

113. Ponder, J. W. & Case, D. A. Force fields for protein simulations. *Advances in Protein Chemistry* **66,** 27–85 (2003).

114. De Jong, D. H. *et al.* Improved parameters for the martini coarse-grained protein force field. *Journal of Chemical Theory and Computation* **9,** 687–697 (2013).

115. Yu, A. *et al.* A multiscale coarse-grained model of the SARS-CoV-2 virion. *Biophysical Journal* **120,** 1097–1104 (2021).

116. Stevens, J. A. *et al.* Molecular dynamics simulation of an entire cell. *Frontiers in Chemistry* **11,** 1–9 (2023).

117. Im, W. & Khalid, S. Molecular simulations of gram-negative bacterial membranes come of age. *Annual Review of Physical Chemistry Annu. Rev. Phys. Chem. 2020* **71,** 171–188 (2020).

118. Khalid, S., Piggot, T. J. & Samsudin, F. Atomistic and Coarse Grain Simulations of the Cell Envelope of Gram-Negative Bacteria: What Have We Learned? *Accounts of Chemical Research* **52,** 180–188 (2019).

119. Vanommeslaeghe, K. & Mackerell, A. D. CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochimica et Biophysica Acta - General Subjects* **1850,** 861–871 (2015).

120. Wang, L. P., Martinez, T. J. & Pande, V. S. Building force fields: An automatic, systematic, and reproducible approach. *Journal of Physical Chemistry Letters* **5,** 1885–1891 (2014).

121. Van der Spoel, D. Systematic design of biomolecular force fields. *Current Opinion in Structural Biology* **67,** 18–24 (2021).

122. Betz, R. M. & Walker, R. C. Paramfit: Automated optimization of force field parameters for molecular dynamics simulations. *Journal of Computational Chemistry* **36,** 79–87 (2015).

123. Mayne, C. G., Saam, J., Schulten, K., Tajkhorshid, E. & Gumbart, J. C. Rapid parameterization of small molecules using the force field toolkit. *Journal of Computational Chemistry* **34,** 2757–2770 (2013).

124. Adcock, S. A. & McCammon, J. A. Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chemical Reviews* **106,** 1589–1615 (2006).

125. Waldher, B., Kuta, J., Chen, S., Henson, N. & Clark, A. E. ForceFit: A code to fit classical force fields to quantum mechanical potential energy surfaces. *Journal of Computational Chemistry* **31,** 2307–2316 (2010).

126. Cole, D. J., Vilseck, J. Z., Tirado-Rives, J., Payne, M. C. & Jorgensen, W. L. Biomolecular force field parameterization via atoms-in-molecule electron density partitioning. *Journal of Chemical Theory and Computation* **12,** 2312–2323 (2016).

127. Amann-Winkel, K. *et al.* X-ray and neutron scattering of water. *Chemical Reviews* **116,** 7570–7589 (2016).

128. Iwashita, T. *et al.* Seeing real-space dynamics of liquid water through inelastic X-ray scattering. *Science Advances* **3,** 1–6 (2017).

129. Soper, A. K. The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa. *Chemical Physics* **258,** 121–137 (2000).

130. Mark, P. & Nilsson, L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *Journal of Physical Chemistry A* **105,** 9954–9960 (2001).

131. Abascal, J. L. & Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *Journal of Chemical Physics* **123** (2005).

132. Berendsen, H. J., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *Journal of Physical Chemistry* **91,** 6269–6271 (1987).

133. Sajadi, F. & Rowley, C. N. Simulations of lipid bilayers using the CHARMM36 force field with the TIP3P-FB and TIP4P-FB water models. *PeerJ* **2018,** 1–26 (2018).

134. Vega, C., Abascal, J. L. F., Conde, M. M. & Aragones, J. L. What ice can teach us about water interactions: a critical comparison of the performance of different water models. *Faraday Discussions* **141,** 251–276 (2009).

135. Huang, J. *et al.* CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods* **14,** 71–73 (2016).

136. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79,** 926–935 (1983).

137. Robertson, M. J., Tirado-Rives, J. & Jorgensen, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *Journal of Chemical Theory and Computation* **11,** 3499–3509 (2015).

138. Jensen, K. P. & Jorgensen, W. L. Halide, ammonium, and alkali metal ion parameters for modeling aqueous solutions. *Journal of Chemical Theory and Computation* **2,** 1499–1509 (2006).

139. Kohagen, M., Mason, P. E. & Jungwirth, P. Accounting for Electronic Polarization Effects in Aqueous Sodium Chloride via Molecular Dynamics Aided by Neutron Scattering. *Journal of Physical Chemistry B* **120,** 1454–1460 (2016).

140. Bankura, A. *et al.* A systematic study of chloride ion solvation in water using van der Waals inclusive hybrid density functional theory. *Molecular Physics* **113,** 2842–2854 (2015).

141. Joung, I. S. & Cheatham, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *Journal of Physical Chemistry B* **112,** 9020–9041 (2008).

142. Kohagen, M., Mason, P. E. & Jungwirth, P. Accurate description of calcium solvation in concentrated aqueous solutions. *Journal of Physical Chemistry B* **118,** 7902–7909 (2014).

143. Leach, A. R. *Molecular Modelling: Principles and Applications* Second (Pearson Education Limited, 2001).

144. Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2,** 19–25 (2015).

145. Lindahl, E., Abraham, M. J., Hess, B. & van der Spoel, D. *GROMACS 2020 Manual (Version 2020)* 2020.

146. Jefferies, D. & Khalid, S. Atomistic and coarse-grained simulations of membrane proteins: A practical guide. *Methods* **185,** 15–27 (2021).

147. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *Journal of Computational Physics* **23,** 321–341 (1977).

148. Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry* **13,** 952–962 (1992).

149. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J Comput Chem* **18,** 1463–1472 (1997).

150. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *Journal of Chemical Physics* **126,** 1–7 (2007).

151. Basconi, J. E. & Shirts, M. R. Effects of temperature control algorithms on transport properties and kinetics in molecular dynamics simulations. *Journal of Chemical Theory and Computation* **9,** 2887–2899 (2013).

152. Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics* **72,** 2384–2393 (1980).

153. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics* **81,** 511–519 (1984).

154. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* **31,** 1695–1697 (1985).

155. Evans, D. J. & Holian, B. L. The Nose-Hoover thermostat. *The Journal of Chemical Physics* **83,** 4069–4074 (1985).

156. Nosé, S. & Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Molecular Physics* **50,** 1055–1076 (1983).

157. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52,** 7182–7190 (1981).

158. Chow, K.-H. & Ferguson, D. M. Isothermal-isobaric molecular dynamics simulations with Monte Carlo volume sampling. *Computer Physics Communications* **91,** 283–289 (1995).

159. Åqvist, J., Wennerström, P., Nervall, M., Bjelic, S. & Brandsdal, B. O. Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chemical Physics Letters* **384,** 288–294 (2004).

160. Harger, M. & Ren, P. Virial-based Berendsen barostat on GPUs using AMOEBA in Tinker-OpenMM. *Results in Chemistry* **1,** 1–5 (2019).

161. Faller, R. & De Pablo, J. J. Constant pressure hybrid molecular dynamics-Monte Carlo simulations. *Journal of Chemical Physics* **116,** 55–59 (2002).

162. Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik* **369,** 253–287 (1921).

163. Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103,** 8577–8593 (1995).

164. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98,** 10089–10092 (1993).

165. Nerenberg, P. S. & Head-Gordon, T. New developments in force fields for biomolecular simulations. *Current Opinion in Structural Biology* **49,** 129–138 (2018).

166. Cieplak, P., Dupradeau, F. Y., Duan, Y. & Wang, J. Polarization effects in molecular mechanical force fields. *Journal of Physics Condensed Matter* **21,** 1–21 (2009).

167. Ren, P. & Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *Journal of Physical Chemistry B* **107,** 5933–5947 (2003).

168. Grossfield, A., Ren, P. & Ponder, J. W. Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. *Journal of the American Chemical Society* **125,** 15671–15682 (2003).

169. Ren, P., Wu, C. & Ponder, J. W. Polarizable atomic multipole-based molecular mechanics for organic molecules. *Journal of Chemical Theory and Computation* **7,** 3143–3161 (2011).

170. Chu, H. *et al.* Polarizable atomic multipole-based force field for DOPC and POPE membrane lipids. *Molecular Physics* **116,** 1037–1050 (2018).

171. Shi, Y. *et al.* Polarizable atomic multipole-based AMOEBA force field for proteins. *Journal of Chemical Theory and Computation* **9,** 4046–4063 (2013).

172. Seiferth, D., Tucker, S. J. & Biggin, P. C. Limitations of non-polarizable force fields in describing anion binding poses in non-polar synthetic hosts. *Physical Chemistry Chemical Physics* **25,** 17596–17608 (2023).

173. Liu, C., Qi, R., Wang, Q., Piquemal, J. P. & Ren, P. Capturing Many-Body Interactions with Classical Dipole Induction Models. *Journal of Chemical Theory and Computation* **13,** 2751–2761 (2017).

174. Laury, M. L., Wang, L. P., Pande, V. S., Head-Gordon, T. & Ponder, J. W. Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *Journal of Physical Chemistry B* **119,** 9423–9437 (2015).

175. Thole, B. T. Molecular Polarizabilities Calculated with a Modified Dipole Interaction. *Chemical Physics* **59,** 341–350 (1981).

176. Ponder, J. W. *et al.* Current status of the AMOEBA polarizable force field. *Journal of Physical Chemistry B* **114,** 2549–2564 (2010).

177. Wang, L. P., Chen, J. & Van Voorhis, T. Systematic parametrization of polarizable force fields from quantum chemistry data. *Journal of Chemical Theory and Computation* **9,** 452–460 (2013).

178. Lopes, P. E., Roux, B. & MacKerell, A. D. Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: Theory and applications. *Theoretical Chemistry Accounts* **124,** 11–28 (2009).

179. Lemkul, J. A., Huang, J., Roux, B. & Mackerell, A. D. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chemical Reviews* **116,** 4983–5013 (2016).

180. Simmonett, A. C., Pickard, F. C., Shao, Y., Cheatham, T. E. & Brooks, B. R. Efficient treatment of induced dipoles. *Journal of Chemical Physics* **143,** 1–12 (2015).

181. Lagardère, L., Aviat, F. & Piquemal, J. P. Pushing the limits of multiple-time-step strategies for polarizable point dipole molecular dynamics. *Journal of Physical Chemistry Letters* **10,** 2593–2599 (2019).

182. Tuckermar, M., Berne, B. J. & Martyna, G. J. Reversible multiple time scale molecular dynamics. *Journal of Chemical Physics* **97,** 1990–2001 (1992).

183. Masella, M. The multiple time step r-RESPA procedure and polarizable potentials based on induced dipole moments. *Molecular Physics* **104,** 415–428 (2006).

184. Leontyev, I. V. & Stuchebrukhov, A. A. Electronic continuum model for molecular dynamics simulations. *Journal of Chemical Physics* **130,** 1–8 (2009).

185. Duboué-Dijon, E. *et al.* Binding of divalent cations to insulin: capillary electrophoresis and molecular simulations. *Journal of Physical Chemistry B* **122,** 5640–5648 (2018).

186. Leontyev, I. V., Vener, M. V., Rostov, I. V., Basilevsky, M. V. & Newton, M. D. Continuum level treatment of electronic polarization in the framework of molecular simulations of solvation effects. *Journal of Chemical Physics* **119,** 8024–8037 (2003).

187. Kirby, B. J. & Jungwirth, P. Charge scaling manifesto: A way of reconciling the inherently macroscopic and microscopic natures of molecular simulations. *Journal of Physical Chemistry Letters* **10,** 7531–7536 (2019).

188. Leontyev, I. V. & Stuchebrukhov, A. A. Electronic continuum model for molecular dynamics simulations of biological molecules. *Journal of Chemical Theory and Computation* **6,** 1498–1508 (2010).

189. Blazquez, S., Conde, M. M., Abascal, J. L. & Vega, C. The Madrid-2019 force field for electrolytes in water using TIP4P/2005 and scaled charges: Extension to the ions F-, Br-, I-, Rb+, and Cs+. *Journal of Chemical Physics* **156,** 1–19 (2022).

190. Nencini, R. *et al.* Prosecco: polarization reintroduced by optimal scaling of electronic continuum correction origin in MD simulations. *Biophysical Journal* **121,** 157a (2022).

191. Corey, R. A., Vickery, O. N., Sansom, M. S. & Stansfeld, P. J. Insights into membrane protein-lipid interactions from free energy calculations. *Journal of Chemical Theory and Computation* **15,** 5727–5736 (2019).

192. Roux, B. t. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications* **91,** 275–282 (1995).

193. Yang, Y. I., Shao, Q., Zhang, J., Yang, L. & Gao, Y. Q. Enhanced sampling in molecular dynamics. *Journal of Chemical Physics* **151,** 1–9 (2019).

194. Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A. & Rosenbergl, J. M. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The Method. *Journal of Computational Chemistry* **13,** 1011–1021 (1992).

195. Hub, J. S., De Groot, B. L. & Van Der Spoel, D. G-whams-a free weighted histogram analysis implementation including robust error and autocorrelation estimates. *Journal of Chemical Theory and Computation* **6,** 3713–3720 (2010).

196. Levin, Y. & Dos Santos, A. P. Ions at hydrophobic interfaces. *Journal of Physics: Condensed Matter* **26,** 1–11 (2014).

197. Ruan, Z., Orozco, I. J., Du, J. & Lü, W. Structures of human pannexin 1 reveal ion pathways and mechanism of gating. *Nature* **584,** 646–651 (2020).

198. Bruce, E. E. & Van Der Vegt, N. F. Does an electronic continuum correction improve effective short-range ion-ion interactions in aqueous solution? *Journal of Chemical Physics* **148,** 1–12 (2018).

199. Bedrov, D. *et al.* Molecular dynamics simulations of ionic liquids and electrolytes using polarizable force fields. *Chemical Reviews* **119,** 7940–7995 (2019).

200. Dos Santos, A. P. & Levin, Y. Ions at the water-oil interface: Interfacial tension of electrolyte solutions. *Langmuir* **28,** 1304–1308 (2012).

201. Jungwirth, P. & Tobias, D. J. Surface effects on aqueous ionic solvation: A molecular dynamics simulation study of NaCl at the air/water interface from infinite dilution to saturation. *Journal of Physical Chemistry B* **104,** 7702–7706 (2000).

202. D'Auria, R. & Tobias, D. J. Relation between surface tension and ion adsorption at the air-water interface: A molecular dynamics simulation study. *Journal of Physical Chemistry A* **113,** 7286–7293 (2009).

203. Rasaiah, J. C., Garde, S. & Hummer, G. Water in nonpolar confinement: From nanotubes to proteins and beyond. *Annual Review of Physical Chemistry* **59,** 713–740 (2008).

204. Liu, D., Li, J., Wu, J. & Lu, D. Ab initio molecular dynamics simulation of water transport through short carbon nanotubes. *ACS Omega* **7,** 40466–40479 (2022).

205. Frolov, A. I., Fedorov, M. V. & Rozhin, A. G. Ion interactions with the carbon nanotube surface in aqueous solutions: Understanding the molecular mechanisms. *ChemPhysChem* **11,** 2612–2616 (2010).

206. Mohammadzadeh, L., Quaino, P. & Schmickler, W. Interactions of anions and cations in carbon nanotubes. *Faraday Discussions* **193,** 415–426 (2016).

207. Beckstein, O., Tai, K. & Sansom, M. S. Not ions alone: Barriers to ion permeation in nanopores and channels. *Journal of the American Chemical Society* **126,** 14694–14695 (2004).

208. Shirts, M. R., Mobley, D. L., Chodera, J. D. & Pande, V. S. Accurate and efficient corrections for missing dispersion interactions in molecular simulations. *Journal of Physical Chemistry B* **111,** 13052–13063 (2007).

209. Hantal, G., Horváth, R. A., Kolafa, J., Sega, M. & Jedlovszky, P. Surface affinity of alkali and halide ions in their aqueous solution: insight from intrinsic density analysis. *Journal of Physical Chemistry B* **124,** 9884–9897 (2020).

210. Petersen, P. B. & Saykally, R. J. Probing the interfacial structure of aqueous electrolytes with femtosecond second harmonic generation spectroscopy. *Journal of Physical Chemistry B* **110,** 14060–14073 (2006).

211. Le Breton, G. & Joly, L. Molecular modeling of aqueous electrolytes at interfaces: Effects of long-range dispersion forces and of ionic charge rescaling. *Journal of Chemical Physics* **152,** 1–7 (2020).

212. Pluhařová, E., Fischer, H. E., Mason, P. E. & Jungwirth, P. Hydration of the chloride ion in concentrated aqueous solutions using neutron scattering and molecular dynamics. *Molecular Physics* **112,** 1230–1240 (2014).

213. Padmanabhan, V. *et al.* Specific ion adsorption and short-range interactions at the air aqueous solution interface. *Physical Review Letters* **99,** 1–4 (2007).

214. Rao, S. *et al.* Water and hydrophobic gates in ion channels and nanopores. *Faraday Discussions* **209,** 231–247 (2018).

215. Dämgen, M. A. & Biggin, P. C. A refined open state of the glycine receptor obtained via molecular dynamics simulations. *Structure* **28,** 130–139 (2020).

216. Aryal, P., Sansom, M. S. P. & Tucker, S. J. Hydrophobic gating in ion channels. *Journal of Molecular Biology* **427,** 121–130 (2015).

217. Zhu, F. & Hummer, G. Gating transition of pentameric ligand-gated ion channels. *Biophysical Journal* **97,** 2456–2463 (2009).

218. Leung, K., Rempe, S. B. & Lorenz, C. D. Salt permeation and exclusion in hydroxylated and functionalized silica pores. *Physical Review Letters* **96,** 1–4 (2006).

219. Liang, Z. F. *et al.* Charge distribution in graphene from quantum calculation. *Chinese Physics B* **27,** 1–7 (2018).

220. Hanwell, M. D. *et al.* Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics* **4,** 1–17 (2012).

221. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* **14,** 33–38 (1996).

222. Kandt, C., Ash, W. L. & Peter Tieleman, D. Setting up and running molecular dynamics simulations of membrane proteins. *Methods* **41,** 475–488 (2007).

223. Fuentes-Azcatl, R., Mendoza, N. & Alejandre, J. Improved SPC force field of water based on the dielectric constant: SPC/ϵ. *Physica A: Statistical Mechanics and its Applications* **420,** 116–123 (2015).

224. Jorgensen, W. L. & Tirado-Rives, J. The OPLS Potential Functions for Proteins. Energy minimizations for crystals of cyclic peptides and crambin. *Journal of The American Chemical Society* **110,** 1657–1666 (1988).

225. Kann, Z. R. & Skinner, J. L. A scaled-ionic-charge simulation model that reproduces enhanced and suppressed water diffusion in aqueous salt solutions. *Journal of Chemical Physics* **141,** 1–7 (2014).

226. Gowers, R. J. *et al. MDAnalysis: A python package for the rapid analysis of molecular dynamics simulations* in *Proc. of the 15th Python in Science Conference* (2016), 98–105.

227. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **32,** 2319–2327 (2011).

228. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585,** 357–362 (2020).

229. Bankura, A., Carnevale, V. & Klein, M. L. Hydration structure of salt solutions from ab initio molecular dynamics. *Journal of Chemical Physics* **138,** 1–10 (2013).

230. Sun, L., Li, X., Tu, Y. & Ågren, H. Origin of ion selectivity at the air/water interface. *Physical Chemistry Chemical Physics* **17,** 4311–4318 (2015).

231. Jiménez-Ángeles, F., Harmon, K. J., Nguyen, T. D., Fenter, P. & De La Cruz, M. O. Nonreciprocal interactions induced by water in confinement. *Physical Review Research* **2,** 1–18 (2020).

232. Son, C. Y. & Wang, Z.-G. Image-charge effects on ion adsorption near aqueous interfaces. *Proceedings of the National Academy of Sciences* **118,** 1–8 (2021).

233. Benavides, A. L. *et al.* A potential model for sodium chloride solutions based on the TIP4P/2005 water model. *Journal of Chemical Physics* **147,** 1–15 (2017).

234. Leontyev, I. & Stuchebrukhov, A. Accounting for electronic polarization in non-polarizable force fields. *Physical Chemistry Chemical Physics* **13,** 2613–2626 (2011).

235. Kulik, H. J., Schwegler, E. & Galli, G. Probing the structure of salt water under confinement with first-principles molecular dynamics and theoretical X-ray absorption spectroscopy. *Journal of Physical Chemistry Letters* **3,** 2653–2658 (2012).

236. Pham, T. A. *et al.* Salt solutions in carbon nanotubes: The Role of Cation-π Interactions. *Journal of Physical Chemistry C* **120,** 7332–7338 (2016).

237. Aydin, F. *et al.* Ion solvation and transport in narrow carbon nanotubes: Effects of polarizability, Cation-π interaction, and confinement. *Journal of Chemical Theory and Computation* **17,** 1596–1605 (2021).

238. Leung, K. & Marsman, M. Energies of ions in water and nanopores within density functional theory. *Journal of Chemical Physics* **127,** 1–10 (2007).

239. Trick, J. L., Wallace, E. J., Bayley, H. & Sansom, M. S. Designing a hydrophobic barrier within biomimetic nanopores. *ACS Nano* **8,** 11268–11279 (2014).

240. Da Silva, L. B. Structural and dynamical properties of water confined in carbon nanotubes. *Journal of Nanostructure in Chemistry* **4,** 1–5 (2014).

241. Köhler, M. H., Bordin, J. R., da Silva, L. B. & Barbosa, M. C. Structure and dynamics of water inside hydrophobic and hydrophilic nanotubes. *Physica A: Statistical Mechanics and its Applications* **490,** 331–337 (2018).

242. Faucher, S. *et al.* Critical knowledge gaps in mass transport through single-digit nanopores: A review and perspective. *Journal of Physical Chemistry C* **123,** 21309–21326 (2019).

243. Faucher, S. *et al.* Diameter dependence of water filling in lithographically degmented isolated carbon nanotubes. *ACS Nano* **15,** 2778–2790 (2021).

244. Dellostritto, M., Xu, J., Wu, X. & Klein, M. L. Aqueous solvation of the chloride ion revisited with density functional theory: Impact of correlation and exchange approximations. *Physical Chemistry Chemical Physics* **22,** 10666–10675 (2020).

245. Powell, D. H., Neilson, G. W., Enderby, J. E. & Powellt, D. H. The structure of Cl-in aqueous solution: an experimental determination of gClH (r) and gClO (r). *Journal of Physics: Condensed Matter* **5,** 5723–5730 (1993).

246. Zhao, Z., Rogers, D. M. & Beck, T. L. Polarization and charge transfer in the hydration of chloride ions. *Journal of Chemical Physics* **132,** 1–10 (2010).

247. Leung, K., Rempe, S. B. & Von Lilienfeld, O. A. Ab initio molecular dynamics calculations of ion hydration free energies. *Journal of Chemical Physics* **130,** 1–11 (2009).

248. Muralidharan, A., Pratt, L. R., Chaudhari, M. I. & Rempe, S. B. Quasi-chemical theory for anion hydration and specific ion effects: Cl-(aq) vs. F-(aq). *Chemical Physics Letters: X* **4,** 1–9 (2019).

249. Enderby, J. E. Ionic hydration by neutron scattering. *Journal of Pure and Applied Chemistry* **57,** 1025–1030 (1985).

250. Rowley, C. N. & Roux, B. The solvation structure of Na+ and K+ in liquid water determined from high level ab initio molecular dynamics simulations. *Journal of Chemical Theory and Computation* **8,** 3526–3535 (2012).

251. Shi, Y. & Beck, T. L. Absolute ion hydration free energy scale and the surface potential of water via quantum simulation. *Proceedings of the National Academy of Sciences of the United States of America* **117,** 30151–30158 (2020).

252. Varma, S. & Rempe, S. B. Coordination numbers of alkali metal ions in aqueous solutions. *Biophysical Chemistry* **124,** 192–199 (2006).

253. McCaffrey, D. L. *et al.* Mechanism of ion adsorption to aqueous interfaces: Graphene/water vs. air/water. *Proceedings of the National Academy of Sciences of the United States of America* **114,** 13369–13373 (2017).

254. Inakollu, V. S. S., Geerke, D. P., Rowley, C. N. & Yu, H. Polarisable force fields: What do they add in biomolecular simulations? *Current Opinion in Structural Biology* **61,** 182–190 (2019).

255. Yoshizawa, S. *et al.* Functional characterization of flavobacteria rhodopsins reveals a unique class of light-driven chloride pump in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **111,** 6732–6737 (2014).

256. Yun, J.-H. *et al.* Early-stage dynamics of chloride ion-pumping rhodopsin revealed by a femtosecond X-ray laser. *Proceedings of the National Academy of Sciences of the United States of America* **118,** 2021 (2021).

257. Mous, S. *et al.* Dynamics and mechanism of a light-driven chloride pump. *Science* **375,** 845–851 (2022).

258. Lukasheva, N., Tolmachev, D. ; Martinez-Seara, H. ; & Karttunen, M. Changes in the local conformational states caused by simple Na+ and K+ ions in polyelectrolyte Simulations: Comparison of seven force fields with and without NBFIX and ECC corrections. *Polymers* **14,** 1–22 (2022).

259. Klesse, G., Tucker, S. J. & Sansom, M. S. Electric field induced wetting of a hydrophobic gate in a model nanopore based on the 5-HT3 receptor channel. *ACS Nano* **14,** 10480–10491 (2020).

260. Lee, J. *et al.* CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *Journal of Chemical Theory and Computation* **12,** 405–413 (2016).

261. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **29,** 1859–1865 (2008).

262. Olsson, M. H., SØndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. *Journal of Chemical Theory and Computation* **7,** 525–537 (2011).

263. Søndergaard, C. R., Olsson, M. H., Rostkowski, M. & Jensen, J. H. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *Journal of Chemical Theory and Computation* **7,** 2284–2295 (2011).

264. Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi 1$ and $\chi 2$ Dihedral Angles. *Journal of Chemical Theory and Computation* **8,** 3257–3273 (2012).

265. Feroz, H. *et al.* Light-driven chloride transport kinetics of halorhodopsin. *Biophysical Journal* **115,** 353–360 (2018).

266. Huang, J., Lopes, P. E., Roux, B. & MacKerell, A. D. Recent advances in polarizable force fields for macromolecules: Microsecond simulations of proteins using the classical drude oscillator model. *Journal of Physical Chemistry Letters* **5,** 3144–3150 (2014).

267. Schwans, J. P. *et al.* Use of anion-aromatic interactions to position the general base in the ketosteroid isomerase active site. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 11308–11313 (2013).

268. Schottel, B. L. *et al.* Anion-pi interactions. *Chemical Society Reviews* **37,** 1–236 (2008).

269. Chifotides, H. T. & Dunbar, K. R. Anion-π interactions in supramolecular architectures. *Accounts of Chemical Research* **46,** 894–906 (2013).

270. Jackson, M. R. *et al.* A preference for edgewise interactions between aromatic rings and carboxylate anions: The biological relevance of anion-quadrupole interactions. *Journal of Physical Chemistry B* **111,** 8242–8249 (2007).

271. Afshinpour, M., Smith, L. A. & Chakravarty, S. AQcalc: A web server that identifies weak molecular interactions in protein structures. *Protein Science* **32,** 1–17 (2023).

272. Kocman, M., Pykal, M. & Jurečka, P. Electric quadrupole moment of graphene and its effect on intermolecular interactions. *Physical Chemistry Chemical Physics* **16,** 3144–3152 (2014).

273. Dhotel, A. *et al.* Molecular motions in functional self-assembled nanostructures. *International Journal of Molecular Sciences* **14,** 2303–2333 (2013).

274. Breberina, L. M., Milčić, M. K., Nikolić, M. R. & Stojanović, S. D. Contribution of anion-π interactions to the stability of Sm/LSm proteins. *Journal of Biological Inorganic Chemistry* **20,** 475–485 (2015).

275. Chen, Y. H. *et al.* Homologue structure of the SLAC1 anion channel for closing stomata in leaves. *Nature* **467,** 1074–1080 (2010).

276. Deng, Z. *et al.* Structural mechanism for gating of a eukaryotic mechanosensitive channel of small conductance. *Nature Communications* **11,** 1–9 (2020).

277. Beckers, M., Mann, D. & Sachse, C. Structural interpretation of cryo-EM image reconstructions. *Progress in Biophysics and Molecular Biology* **160,** 26–36 (2021).

278. Nierzwicki, Ł. & Palermo, G. Molecular dynamics to predict cryo-EM: Capturing transitions and short-lived conformational states of biomolecules. *Frontiers in Molecular Biosciences* **8,** 1–6 (2021).

279. Rao, S., Klesse, G., Stansfeld, P. J., Tucker, S. J. & Sansom, M. S. A BEST example of channel structure annotation by molecular simulation. *Channels* **11,** 347–353 (2017).

280. Johnson, A. A. *et al.* Bestrophin 1 and retinal disease. *Progress in Retinal and Eye Research* **58,** 45–69 (2017).

281. Kittredge, A. *et al. Calcium-activated chloride channels* (IOP Publishing, 2023).

282. Yang, T. *et al.* Structure and selectivity in bestrophin ion channels. *Science* **346,** 355–359 (2014).

283. Vaisey, G. & Long, S. B. An allosteric mechanism of inactivation in the calcium-dependent chloride channel BEST1. *Journal of General Physiology* **150,** 1484–1497 (2018).

284. Miller, A. N., Vaisey, G. & Long, S. B. Molecular mechanisms of gating in the calcium-activated chloride channel bestrophin. *eLife* **8,** 1–17 (2019).

285. Vickery, O. N. & Stansfeld, P. J. CG2AT2: An enhanced fragment-based approach for serial multi-scale molecular dynamics simulations. *Journal of Chemical Theory and Computation* **17,** 6472–6482 (2021).

286. Gumbart, J., Khalili-Araghi, F., Sotomayor, M. & Roux, B. Constant electric field simulations of the membrane potential illustrated with simple systems. *Biochimica et Biophysica Acta - Biomembranes* **1818,** 294–302 (2012).

287. Zhang, C. *et al.* AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *Journal of Chemical Theory and Computation* **14,** 2084–2108 (2018).

288. Song, W. *et al.* PyLipID: A Python Package for Analysis of Protein-Lipid Interactions from molecular dynamics simulations. *Journal of Chemical Theory and Computation* **18,** 1188–1201 (2022).

289. Maiorov, V. N. & Crippen, G. M. Size-independent comparison of protein three-dimensional structures. *Proteins: Structure, Function, and Bioinformatics* **22,** 273–283 (1995).

290. Hol, W. G., van Duijnen, P. T. & Berendsen, H. J. The alpha-helix dipole and the properties of proteins. *Nature* **273,** 443–446.

291. Sheridan, R. P. & Allen, L. C. The electrostatic potential of the alpha helix. *Biophysical Chemistry* **11,** 133–136 (1980).

292. Serna, M. Hands on methods for high resolution cryo-electron microscopy structures of heterogeneous macromolecular complexes. *Frontiers in Molecular Biosciences* **6,** 1–8 (2019).

293. Gibbs, E. *et al.* Conformational transitions and allosteric modulation in a heteromeric glycine receptor. *Nature Communications* **14,** 1–15 (2023).

294. Chaudhari, M. I., Vanegas, J. M., Pratt, L. R., Muralidharan, A. & Rempe, S. B. Hydration mimicry by membrane ion channels. *Annual Review of Physical Chemistry* **71,** 461–484 (2020).

295. Kan, X., Liu, H., Pan, Q., Li, Z. & Zhao, Y. Anion-π interactions: From concept to application. *Chinese Chemical Letters* **29,** 261–266 (2018).

296. Vaisey, G., Miller, A. N. & Long, S. B. Distinct regions that control ion selectivity and calcium-dependent activation in the bestrophin ion channel. *Proceedings of the National Academy of Sciences of the United States of America* **113,** e7399–e7408 (2016).

297. Lin, F. Y. & MacKerell, A. D. Improved modeling of cation-π and anion-ring interactions using the drude polarizable empirical force field for proteins. *Journal of Computational Chemistry* **41,** 439–448 (2020).

298. Kali, R., Mkandawire, W. D. & Milner, S. T. Optimizing the structure and performance of biomimetic water channels. *Molecular Systems Design and Engineering* **8,** 443–450 (2023).

299. Lisbjerg, M. *et al.* Biotin[6]uril esters: Chloride-selective transmembrane anion carriers employing C-H···anion interactions. *Journal of the American Chemical Society* **137,** 4948–4951 (2015).

300. Huang, W. L., Wang, X. D., Ao, Y. F., Wang, Q. Q. & Wang, D. X. Artificial chloride-selective channel: shape and function mimic of the ClC channel selective pore. *Journal of the American Chemical Society* **142,** 13273–13277 (2020).

301. Melcr, J. & Piquemal, J. P. Accurate biomolecular simulations account for electronic polarization. *Frontiers in Molecular Biosciences* **6,** 1–8 (2019).

302. Chen, M. *et al.* Ab initio theory and modeling of water. *Proceedings of the National Academy of Sciences of the United States of America* **114,** 10846–10851 (2017).

303. Montero de Hijes, P., Romano, S., Gorfer, A. & Dellago, C. The kinetics of the ice-water interface from ab initio machine learning simulations. *Journal of Chemical Physics* **158,** 1–7 (2023).

304. Mauger, N., Plé, T., Lagardère, L., Huppert, S. & Piquemal, J.-P. Improving condensed phase water dynamics with explicit nuclear quantum effects: the polarizable Q-AMOEBA force field. *Journal of Physical Chemistry B* **126,** 8813–8826 (2022).

305. Muniz, M. C., Car, R. & Panagiotopoulos, A. Z. Neural network water model based on the MB-Pol many-body potential. *The Journal of Physical Chemistry B* **127,** 9165–9171 (2023).

306. Plé, T., Lagardère, L. & Piquemal, J.-P. Force-field-enhanced neural network interactions: From local equivariant embedding to atom-in-molecule properties and long-range effects. *Chemical Science* **14,** 12554–12569 (2023).

307. Besaw, J. E. *et al.* The crystal structures of a chloride-pumping microbial rhodopsin and its proton-pumping mutant illuminate proton transfer determinants. *Journal of Biological Chemistry* **295,** 14793–14804 (2020).

308. Jaswal, S. S. & Sharma, T. P. Electronic polarisabilities of ions in alkali halide crystals. *Journal of Physics and Chemistry of Solids* **34,** 509–511 (1973).

309. Liu, H. *et al.* Structural insights into anion selectivity and activation mechanism of LRRC8 volume-regulated anion channels. *Cell Reports* **42,** 1–17 (2023).

310. Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods* **18,** 176–185 (2021).

311. Liu, X. *et al.* CryoFormer: Continuous heterogeneous cryo-EM reconstruction using transformer-based neural representations. *arXiv,* 1–15 (2023).

312. Nashed, Y. *et al.* Heterogeneous reconstruction of deformable atomic models in Cryo-EM. *arXiv,* 1–8 (2022).

313. Cheng, M. *et al.* A funnel-shaped chloride nanochannel inspired by ClC protein. *Nano Letters* **21,** 4086–4091 (2021).